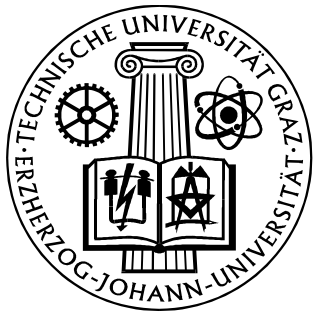
Dipl.-Ing. Alois Schlögl

THE ELECTROENCEPHALOGRAM AND THE ADAPTIVE AUTOREGRESSIVE MODEL: THEORY AND APPLICATIONS

Dissertation

vorgelegt an der Technischen Universität Graz



zur Erlangung des akademischen Grades "Doktor der Technischen Wissenschaften"
(Dr. techn.)

durchgeführt am Institut für Elektro- und Biomedizinische Technik

Graz, April 2000

PREFACE

The human species as conscious creatures seem to have something special, namely a particular organ - the brain - which can connect matter (physical entity) and mind (purely non-physical) to each other in both directions. For example, humans can assign a meaning to a physical entity; and they can also transform ideas into facts of the physical world. Through the brain, humans seem to be a kind of transformer between both worlds. However, a lot of mechanisms involved in this transformation have not been illuminated, yet.
The field of computer science (with its implications from information theory, neural networks and the discussion about artificial intelligence) and biomedical engineering (with in vivo measurements of body reactions) can contribute to solving the unanswered questions in neuroscience. Especially the EEG, as a completely non-invasive technique, can be easily applied in real world situations of humans. Moreover, it can give an image of the whole brain, not only parts of it. For these reasons, EEG is a well-suited method for analyzing the functions of the human brain in real world situations.

The English language, which I chose for this work, is not my native language. Despite extensive proofreading, mistakes as well as 'false friends' are probably not completely eliminated. I hope this does not constrain the understanding of the meaning of this work.

Alois Schlögl
Graz, January 2000

ACKNOWLEDGEMENT

Ich danke meinen Eltern, Sabine, und allen meinen Freunden welche mir besonders in den schierigen Zeiten geholfen haben. Ich danke auch Univ.-Prof. Pfurtscheller für die Ermöglichung der Dissertation an seiner Abteilung.

Thanks to all the people I worked with, especially mentioning Georg Dorffner, Bob Kemp, Alpo Värri, Thomas Penzel, Steve Roberts, Peter Anderer, Peter Rappelsberger, Iead Rezek, Gerhard Klösch, Oliver Filz, Jon Wolpaw, Dennis McFarland, Theresa Vaughan, Prof. Kolb und Prof. Dourdoumas.
Thanks to my co-workers at the institute, especially I want to mention Christa Neuper, Gernot Florian, Martin Pregenzer, Klaus Lugger, Herbert Ramoser, Günter Edlinger, Britta Ortmayr, Colin Andrew, Fr. Brigitte Wahl, Sabine Ferstl and all other co-workers.
This work was supported partly by the Austrian *Fonds zur Förderung der wissenschaftlichen Forschung,* Projects 11208 and P11571 and partly by the Biomed-2 Project BMH4-CT97-2040 (SIESTA).

Dedicated to truth and love.

CONTENTS

ABSTRACT

The problem of time-varying spectral analysis of the human EEG on a single trial basis is addressed. For this purpose the adaptive autoregressive model was investigated, which has the advantage to be well suited for on-line analysis, and requires a minimal number of parameters for describing the spectrum. The principle of uncertainty of the resolution between time- and frequency domain was reformulated into terms of a time-varying AR model, whereby the model order and the update coefficients correspond to the frequency- and time-resolution, respectively. Eighty-eight versions of AAR estimation algorithms (Kalman filtering, RLS, LMS and recursive AR techniques) were investigated. A criterion based on the one-step prediction error for the goodness-of-fit was defined. This criterion proved to be useful for determining free parameters like the model order, update coefficients and estimation algorithm.

The AAR model was applied to the single trial classification of off-line BCI data, for the on-line analysis of EEG patterns, for detecting transient events like muscle and movement artifacts and for sleep analysis. The AAR parameters were combined and further analyzed with a multivariate classifier; whereby linear discriminant analysis showed to be very useful. Accordingly, one obtains a time course of an error rate and the average and variance of the classification output. This shows well the dynamics of the EEG spectra. Furthermore, the classification output can be calculated on-line and provides the BCI feedback for subject training. This was used to develop a new paradigm in the Graz BCI system. The amount of (useful) information of the classification output is also estimated.

The method of inverse autoregressive filtering for detecting transient events was modified to adaptive inverse autoregressive filtering. It was used to detect artifacts within the sleep EEG. This is useful because the spectrum of sleep EEG changes with time and many artifacts are observed in EEG recordings. It is demonstrated how an AAR estimation algorithm reacts in case of a transient event; the detection quality for various types of artifacts was evaluated with the method of ROC-curves. The area-under-the-curve (AUC) was larger than 0.85 for muscle and movement artifacts. The AAR parameters were also used for the automated analysis of sleep EEG. In combination with a linear discriminant analysis, time courses were obtained which can be related to the expert scoring of sleep stages.

In summary, it is shown that the AAR method is convenient; no instabilities of the estimation algorithms occur, at least when the update coefficient and the model order are properly chosen. Two physiological models for EEG rhythms are discussed, and it is shown that there is a corresponding ARMA model. It can be expected that the criterion, which was introduced for comparing AAR models, is also useful for comparing other models (e.g. bi- and multivariate, non-linear, physiological reasoned, etc.) for time-varying spectrum including higher order spectra.

KURZFASSUNG

Das Problem der zeitvarianten Spektralanalyse beim menschlichen Elektroenzephalogramm ohne Anwendung von Mittelung wird behandelt. Dazu wird ein adaptives autoregressives Modell untersucht, welches gut für die Echtzeitberechung geeignet ist und eine minimale Anzahl von Parametern zur Beschreibung des Spektrums benötigt. Die Unschärferelation zwischen Zeit- und Frequenzauflösung wird auf das Gebiet der zeit-veränderlichen autoregressiven Modelle übertragen. Dabei entsprechen die Ordnung des Modells der Frequenzauflösung und der Adaptionsfaktor der Zeitauflösung. Insgesamt wurden 88 unterschiedliche Varianten von Schätzalgorithmen (Kalman filtern, RLS, LMS und rekursive Verfahren) untersucht. Ein Kriterium, basierend auf dem Vorhersagefehler, wurde als Maß für die Qualität eines AAR Schätzers eingeführt. Dieses Kriterium wurde zur Bestimmung der Modellordung, des Adaptionskoeffizienten, sowie zum Vergleich der verschiedenen Schätzverfahren, angewandt.

Das AAR-Modell wurde zur Offline-Klassifikation von BCI-Daten (ohne Verwendung von Mittelungsverfahren), zur Online-analyse von EEG-Mustern, zur Detektion von Muskel- und Bewegungsartefakten und zur Schlafanalyse verwendet. In Kombination mit einem Klassifikator - die Linear Diskriminanzanalyse zeigte sich brauchbar - wurden die AAR - Parameter zusammengefaßt und weiter analysiert. Daraus konnte ein Zeitverlauf der Fehlerrate, des Durchschnitts und die Varianz des Klassifikationsergebnisses bestimmt werden. Dies stellt deutlich den zeitlichen Verlauf der Dynamik von EEG-Spektren dar. Das Klassifikationsergebnis kann auch online berechnet werden um Personen eine Rückkopplung beim Trainieren mit einem BCI zu geben. Daraus wurde ein neues Paradigma im Grazer BCI-System entwickelt. Schließlich wurde auch eine Schätzung des Informationsgehaltes des Klassifikationsergebnisses durchgeführt

Die Methode von "Inversen Autoregressiven Filter" zur Detektion von transienten (kurzfristigen) Ereignissen, wurde zu "Adaptiven Inversen Autoregressiven Filter" modifiziert. Dieses Verfahren wurde zur Detektion von Artefakten im Schlaf-EEG eingesetzt. Dies ist sinnvoll, da sich das Spektrum des Schlaf-EEG's ändert und Artefakte enthält. Dabei wurde auch untersucht wie sich die Schätzalgorithmen im Falle eines solchen transienten Ereignisses verhalten. Die Qualität der Detektion wurde mittels ROC-Kurven analysiert, dabei zeigte sich daß die Fläche unter der ROC-Kurve für Muskel- und Bewegungsartefakte jeweils größer als 0.85 ist. Weiters wurden die AAR-Parameter auch für die automatische Analyse des Schlaf-EEG's angewandt. Unter Verwendung der Linearen Diskriminanzanalyse wurden Verlaufskurven ermittelt, welche in Beziehung zur Analyse durch Experten gesetzt wurden.

Zusammenfassend kann gesagt werden, daß die Methode der AAR-Parameter brauchbar ist; es traten keine Instabilitäten bei der Berechnung auf, sofern die Modellordnung und der Adaptionskoeffizient geeignet gewählt wurde. Zwei physiologisch begründete Modelle wurden diskutiert, wobei gezeigt werden konnte, daß ein entsprechendes ARMA - Modell existiert. Dasselbe Kriterium, welches zur Wahl des optimalen AAR - Modells verwendet wurde, könnte auch für den Vergleich von anderen Modellen wie z.B. bi- und multivariaten, nicht-linearen bzw. physiologisch begründeten Modellen verwendet werden.

PART I: BACKGROUND

1. Introduction

1.1 Generation and dynamics of EEG

Measurement and analysis of the human electroencephalogram (EEG) can be traced back to Berger (1929). The EEG measures the electrical activity of the brain, and is recorded on the head surface. It has the advantage of being a non-invasive technique and showed to be interesting in many fields related to neuroscience (physiology, psychology, neurology, psychiatry, etc.) (Niedermeyer and Lopes da Silva, 1999). The neuro-physiological basis of the EEG is the electrical field potential. The field potential is caused by secondary ionic currents, which stem from potential gradients of action potentials. (Speckmann and Elger, 1994). These action potentials are pulses of membrane depolarization travelling along the axons of neurons. The pulses can exhibit or inhibit through synapses the depolarization of other, postsynaptic, neurons. A series of pulses, or spike trains, can be seen as the coded information processes in the neural network. Rieke et al. (1997) provide a mathematical framework for the information encoding and emphasizes that each single spike is substantial.

An important link is the relation between these spike trains and the observed EEG patterns. Several works show the relationship between spike trains and EEG patterns like sleep spindles, k-complexes, alpha waves, delta, theta and beta rhythms (for review see Steriade et al. 1990, Lopes da Silva 1991, Steriade 1999). They discuss the different types of EEG oscillations, to which underlying spike train patterns they are related and which brain areas are involved.

EEG oscillations and rhythms were investigated in many fields, in cognitive and memory performance (Klimesch 1999), in terms of their relationship to the sleep process (Achermann and Borbely, 1997). Spectral changes are also used for modeling the sleep process (Merica and Fortune, 1997). Moreover, oscillations play an important role in the brain wave theory of Nunez (1995) which considers the 'global' (whole head) EEG dynamics. Furthermore, the relationship between functional meaning and rhythms (Basar et al. 1999), e.g. synchronization and desynchronization (Pfurtscheller and Lopes da Silva 1999a,b) seems to be very interesting and important.

In order to identify the spectrum, the EEG patterns must be averaged or, otherwise, very long stationary segments are needed. The assumption of long stationary segments is difficult to obtain. Stationarity can not be always assumed; and even if stationarity can be assumed, the time resolution is lost. For this reason, averaging (of an ensemble of similar trials) is often used as a method for increasing the time resolution. In case of averaging, one must take care about various factors. Maturation, age, sex, state-of-consciousness (e.g. sleep and wake), psychiatric and neurological disorders (epilepsy, parkinsonism, metabolic brain disorders, etc.), brain injury, stroke, drug effects, mental tasks, etc. as well as other subject-specific (individual) factors influence the EEG patters. It is the aim of EEG research to enlighten the relationship between these factors and the EEG patterns. Often, the need for averaging hides these relationships and thus it is a limiting factor in EEG analysis.

In this work, methods for the analysis of time-varying EEG spectra will be investigated which do not need averaging. In order to determine the frequency spectra of the EEG, autoregressive (AR) parameters are used (Lustick et al. 1968, Fenwick et al. 1969, 1979, 1971, Zetterberg, 1969, Gersch, 1970), Pfurtscheller and Haring, 1972, Florian and Pfurtscheller, 1995). The main reasons are: firstly, efficient AR estimation algorithms are available, which can be also

used on-line, secondly, an AR model considers quite well the stochastic behavior of the (ongoing) EEG and has some tradition in EEG analysis; and thirdly, the AR-spectrum is a maximum entropy spectral estimator (Burg 1967, 1975, Priestley, 1981, Haykin, 1996). In other words, a small number of parameters describe the spectrum most accurately without the need for averaging. Two applications for time-varying spectral analysis of the EEG are the automated sleep analysis system and the EEG-based Brain Computer Interface (BCI).

1.2 Sleep analysis

EEG-based sleep analysis can be traced back to Loomis et al. (1937, 1938), who already observed different sleep stages. After the discovery of the rapid-eye-movement (REM) by Aserinsky and Kleitman (1953), the different sleep stages were classified into Wake (W), drowsiness (1), light sleep (2), deep sleep (3), very deep sleep (4) and REM (Dement and Kleitman, 1957). Standardized scoring rules were set up by Rechtschaffen and Kales (1968) (R&K) which is still the only generally accepted standard for sleep scoring.

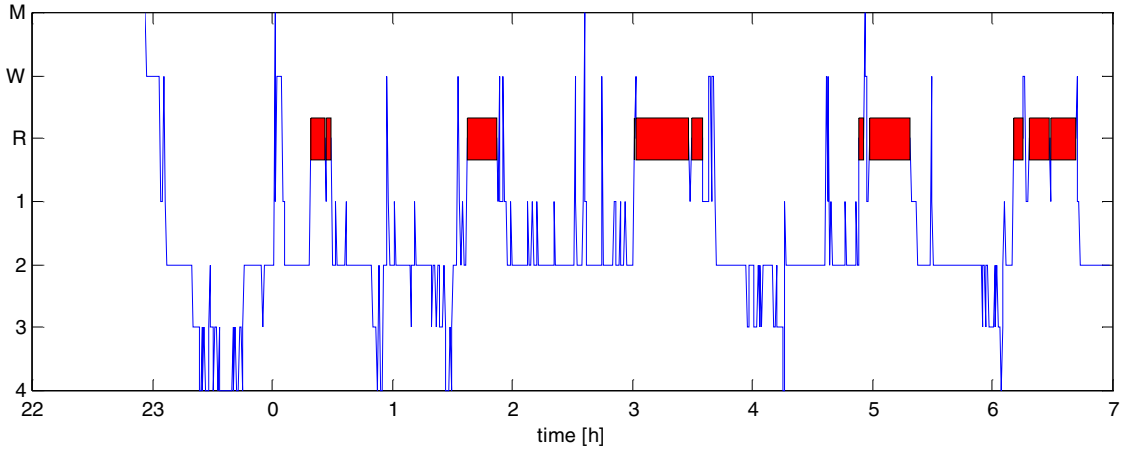


Figure 1: Hypnogram, describing the different sleep stages during night. The recording lasts from ca. 11pm to 7am. W, R, and 1-4 described the wake state, REM, and the four sleep stages, respectively; M indicates movement.

Many attempts were performed in order to automate sleep stage scoring (Larsen and Walter 1970, Smith and Karacan, 1971, Smith et al. 1975, Principe and Smith 1986, Kemp et al. 1987, Principe et al. 1989, Jansen and Dawant 1989, Jobert et al. 1989, Roberts and Tarasenko 1992, Nielsen, 1993, Nielsen et al. 1994, Kubat et al. 1993, 1994, Kemp, 1993). Although, the R&K manual has been criticized (Kubicki et al. 1982, Kubicki and Hermann, 1996) and it was shown (Haustein et al. 1986, Stanus et al. 1987, Hasan et al. 1993) that variability of human scorers is as large as the differences between a human scorer and an automated analysis system, no generally accepted standard for an automated sleep analysis system was obtained, yet. Some promising approaches (Kemp 1993, Nielsen 1993, Roberts and Tarasenko 1992, Hasan 1996) consist of basically two processing steps, a feature extractor and a combiner. The feature extractors are basically signal processing methods, including preprocessing like artifact processing; the combiner which might be a neural network-based classifier or a statistical quantifier, etc. These concepts were considered in the European research project on "A new Standard for Integrating polygraphic sleep recordings

into a comprehensive model of human sleep and its validation in sleep disorders" (SIESTA, Dorffner, 1998).

1.3 Brain computer interface

An EEG-based brain computer interface (BCI) is a device, which enables people to control devices by their EEG patterns. This would allow a direct way to transform mental activity into physical effects without using any muscles. It might be helpful for patients with severe motor disabilities (Vidal 1973, Wolpaw et al. 1994) e.g. amyotrophic lateral sclerosis (ALS) (Elder et al. 1982) or the locked-in syndrome. Bauby (1996) describes his own case by dictating his experience with his eye lid movements only. A further application is to use the EEG for biofeedback. This would be an alternative in the treatment of headache and pains without using sedative drugs (Birbaumer, 1999).

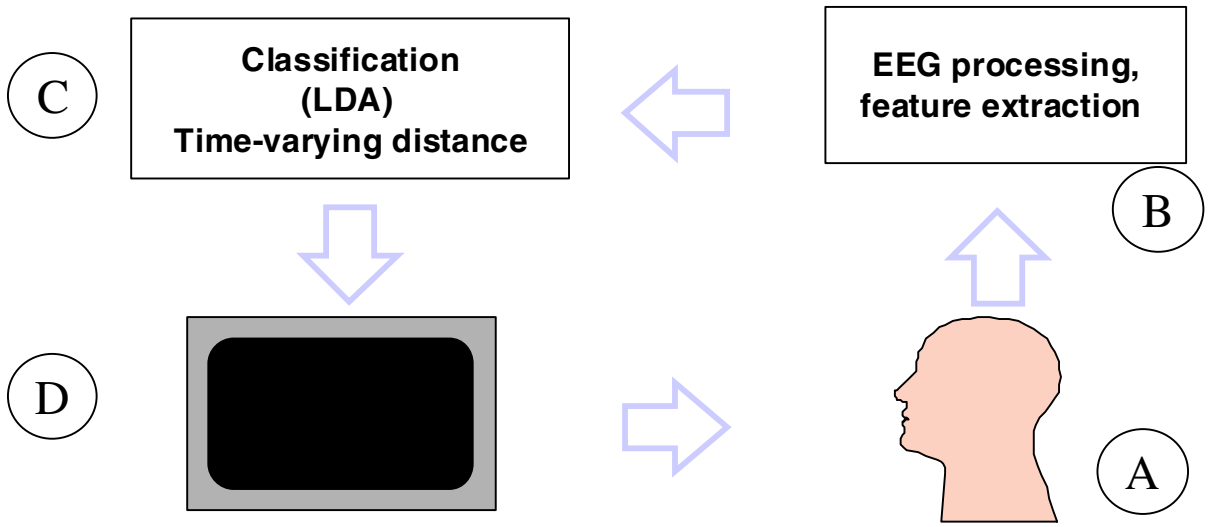


Figure 2: Scheme of an EEG-based BCI with feedback. The EEG from the subject's scalp is recorded (A); then it has to be processed on-line (B); next the extracted features are combined and classified (C); the output can be used to control a device, e.g. a cursor on a computer screen (D); simultaneously, the output provides feedback to the subject.

The idea of a BCI (see Fig. 2) is that different thoughts result in different EEG patterns on the surface of the scalp. Various attempts were made to analyze these patterns e.g. the power in specific frequency bands (McFarland et al. 1993, 1997, Wolpaw et al. 1991, 1994) or slow potential changes (Birbaumer et al. 1981) were used. The Graz BCI system mainly used the power on selected frequency bands (Flotzinger et al. 1994, Pregenzer et al., 1994, Kalcher, et al. 1996, Pfurtscheller et al. 1996, 1997, 1998)

The feedback is important for the subject in order to learn what to concentrate on to fulfill the task. Several issues must be addressed. Firstly, many different feature extraction methods with several degrees of freedom are available. An prerequisite is also that the method can be used on-line and in real-time. Thus, one issue is to decide which signal processing method should be used for EEG feature extraction. Secondly, the number of extracted features must be combined and reduced to one single parameter. Even with an autoregressive model, which uses only a few coefficients for spectrum representation, p parameters per channel are obtained. Other methods yield even more parameters. Therefore, one has to think about

reducing the number of parameters. For this purpose statistical or a neural network-based classifier can be used. In order to apply the classifier online, it must be available in advance. Usually, the classifier has to be learned (trained) from previous BCI recordings (e.g. without feedback), which is a third issue. Fourthly, the kind of feedback and the presentation of the classification output to the subject is important, because the feedback should enable the subject to learn how to control the EEG patterns. The feedback must contain useful information in order be supportive for this process. All these issues must be addressed in a BCI approach.

1.4 Goal

Theoretical considerations let assume that adaptive autoregressive (AAR) parameters are very appropriate for time-varying spectral analysis of the EEG. Because of the adaptive behavior, the estimation algorithms are very suitable for on-line application. Time-varying spectral analysis of the EEG has an important role in sleep analysis, too. In this work, the estimation algorithms of AAR model parameters and the difficulties that arise in AAR modeling are investigated. The AAR method is applied to the sleep EEG and the BCI approach; it will be discussed whether AAR parameters are useful in these applications.

This work is divided into four parts. The first part contains an introduction and presents some, selected, EEG models. The second part deals with the adaptive autoregressive (AAR) model. A criterion for the goodness-of-fit of the model estimates is introduced and the idea of inverse filtering is extended to adaptive inverse filtering. One chapter analyses different estimation algorithms, based on the mentioned criterion.
The third part contains applications of the AAR model. It is shown that artifacts frequently occur in the sleep EEG and must be considered seriously. Quality control of the available data is performed; some artifact processing methods are discussed and it is shown how adaptive inverse filtering can be used for the detection of muscle and movement artifacts. One chapter compares different EEG parameter for single-trial analysis (STA), which is followed by the off-line analysis of data from EEG-based Brain-Computer-Interface (BCI). This is also the basis for introducing a new BCI paradigm. The final part summarizes the results of AAR modeling and its applications. The appendix contains a list of abbreviations, an index list and a listing of the data used. Also, a solution for the problem that AR parameters of different sampling rates can be usually not compared, is presented.

2. EEG models

Many models about the functioning of the brain are available. They range from hippocampal networks, compartment models, lumped circuit models (Lopes da Silva et al. 1974, Zetterberg et al. 1978, Suffczinsky et al. 1999) to global models of brain dynamics (Nunez, 1995). In the following sections, some selected models are discussed in more detail. Firstly, the autoregressive (AR) and autoregressive moving average (ARMA) models are pure mathematical models of the EEG. Therefore, the AR (and ARMA) analysis methods are often denoted as parametric methods. However, these are very powerful models because they are able to describe completely the second order statistics of a time series which includes spectral analysis. Burg (1967) showed that an AR-spectrum is a maximum entropy spectrum and, hence, describes the spectral density function most accurately with a minimum number of parameter.

Two examples of more physiological models are also presented. One is the feedback loop model which was introduced by Kemp and Blom (1981). It is argued that this model is more realistic than a purely mathematical model like an AR model, because it models a feedback loop similar to the feedback loop between thalamus and cortex. The feedback loop model "*departures from physiological reality, but … maintains … main common properties (of the EEG)*" (Kemp and Blom, 1981).

Another model is the lumped circuit model which is used for describing the alpha rhythm of the thalamus (Lopes da Silva et al. 1974). The parameters are derived from biological measurements and describe the interaction between thalamo-cortical (TCR) and reticular nucleus (RE) neuronal populations. The model can show linear as well as non-linear behavior (Zetterberg et al. 1978) and is able to explain the interaction between different spatial populations (Suffczynski et al. 1999). This is the physiologically most-reasoned model.

2.1 The autoregressive model

The use of autoregressive models in EEG analysis can be traced back to Lustick et al. (1968), Fenwick et al. (1969, 1979, 1971), Zetterberg, (1969) and Gersch, (1970). An attraction of AR modeling was that efficient algorithms for parameter estimation are available (Levinson, 1947, Durbin, 1960, Pfurtscheller and Haring 1972). Nowadays, the efficient estimation algorithms are still an important advantage of autoregressive modeling. Zetterberg (1969), Isaksson and Wennberg (1975), and Isaksson et al. (1981) showed the usefulness of autoregressive models (Fig. 3) for spectral parameter analysis (SPA) of the EEG.

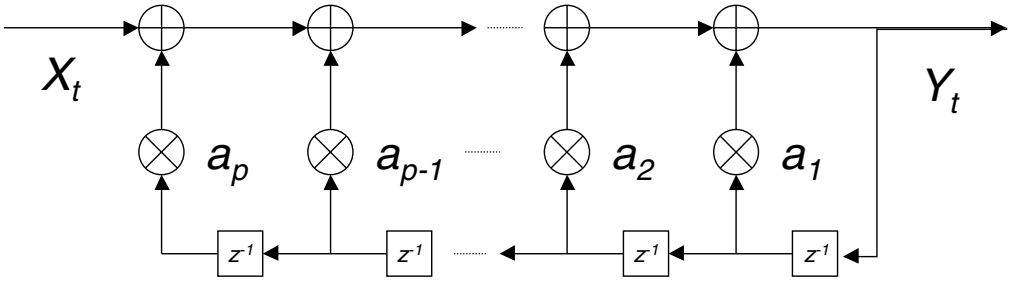


Figure 3: Scheme of an autoregressive model. It is assumed that the observed EEG Y_t can be described by white noise X_t filtered with the AR model. The AR method is also denoted as parametric method, because the parameters of a model are used to characterize the EEG.

An AR model is quite simple and useful for describing the stochastic behavior of a time series. This is described by the following equation

$$y_k = a_1 * y_{k-1} + \ldots + a_p * y_{k-p} + x_k \qquad (2.1)$$

with

$$x_k = N\{0, \sigma_x^2\}. \qquad (2.2)$$

x_k is a zero-mean-Gaussian-noise process with variance σ_x^2; the index k is an integer number and describes discrete, equidistant time points. The time t in seconds is $t = k / f_0 = k * \Delta T$ with the sampling rate f_0 and the sampling interval $\Delta T = 1/f_0$. y_{k-i} with $i = 1..p$ are the p previous sample values, p is the order of the AR model and a_i are the AR model parameter. For simplicity, the vector notation is introduced for the AR parameters and the vector $\mathbf{Y}_{k-1}$ consists of the past p samples

$$\mathbf{a} = [a_1, \ldots, a_p]^T \qquad (2.3)$$
$$\mathbf{Y}_{k-1} = [y_{k-1}, \ldots, y_{k-p}]^T \qquad (2.4)$$

Accordingly the AR model can be written as

$$y_k = \mathbf{a}^T * \mathbf{Y}_{k-1} + x_k \qquad (2.5)$$

and the transfer function in the Z-domain is

$$Y(z)/X(z) = 1/(1 - a_1 * z^{-1} - \ldots - a_p * z^{-p}). \qquad (2.6)$$

The AR model can also be seen as a linear filter with random noise input. The output process is stationary iff all poles (i.e. roots of the denominator) of the transfer function are inside the unit circle. While random noise has a flat spectrum, the spectrum of the model output (i.e. the observed EEG) is determined completely by the AR parameter. The AR model also explains the spectral composition of a signal (shortly the AR spectrum)

$$S(f) = Y(z, z = e^{j\,2\pi f \Delta T})$$
$$S(f) = \sigma_x / (1 - \Sigma_i\, a_i\, e^{-j\,2\pi i f \Delta T}) \quad i = 1..p \qquad (2.7)$$

with the sampling interval $\Delta T = 1/f_0$. Based on the works of Burg (1967, 1975), Haykin (1996) refers to the formula of equation (2.7) as the maximum entropy method (MEM) for spectral estimation. Basically, the AR model is able to describe weak stationary processes. In case of an AR model, the EEG is considered as a filtered white noise process.

Once the AR parameters are identified, they can be applied inversely to the observed process. In this case, transient events can be identified within colored background noise. Bodenstein and Praetorius (1977), Praetorius et al. (1977) and Lopes da Silva et al. (1977) applied it in order to identify spikes and transient phenomena in the EEG. The basic idea is that in a first step, the AR parameters of some EEG segments are identified. In the second step, the AR filter parameters are applied inversely to the EEG (see Fig. 9). Transient phenomena are detected from this inverse filtered process. In chapter 8, this idea is extended to adaptive inverse autoregressive modeling in order to detection transient events in the (non-stationary) sleep EEG.

An important issue in AR modeling is the selection of the model order. Many different criteria are defined; Haring (1975) describes 12 methods for model order selection and compared these methods in EEG data. Nowadays, the most common criteria are final prediction error (FPE, Akaike, 1969), and the Akaike information criterion (AIC, Akaike, 1974), the Bayesian AIC (BIC, Akaike, 1979), the minimum description length (MDL, Rissanen 1978, 1983), the CAT criterion (Parzen, 1977), Schwartz' (1978) bayesian criterion and the Phi-criterion (Hannan 1980, Hannan and Quinn, 1979). The various model order selection criteria are discussed in Priestley (1991), Marple (1987), Wei (1990) and Pukkila (1988).

Despite the great number of criteria, less conclusive results about the optimal AR model order in EEG analysis are available. Jansen et al (1981) found that a fifth order model is sufficient in 90% of the cases. But he stated further, *'visual inspection of the resulting spectra revealed that the order indicated by the FPE criterion is generally too low and better spectra can be obtained using a tenth-order AR model.*'. Vaz et al. (1987) reports that the most consistent model order estimation was provided by the MDL criterion of Rissanen (1978). The study showed that *'a 5th order AR model represents adequately 1- or 2-s EEG segments with the exception of featureless background, where higher order models are necessary*'. Shinn Yih Tseng et al. (1995) found that the average model order on 900 segments of 1.024s is 8.67. Florian and Pfurtscheller (1995) used an order of p=11 and found no differences for p=9..13. One reason for the difficulties of order selection in EEG modeling might be that the optimal model order depends on the length of the observed (investigated) segment, the sampling rate and EEG-specific properties like the number of frequency components. Because of these difficulties, the model order was often chosen twice the expected number of frequency components plus one for the low-frequency component.

2.2 The autoregressive-moving-average model

An AR model can easily be extended to an ARMA model. An ARMA(p,q) model is described by equation (2.8) with

$$y_k = a_1 * y_{k-1} + \ldots + a_p * y_{k-p} + \\ + x_k + b_1 * x_{k-1} + \ldots + b_q * x_{k-p} \qquad (2.8)$$

In this case, the vector notation is

$$\boldsymbol{a} = [a_1, \ldots, a_p, b_1, \ldots, b_p]^T \qquad (2.9)$$

$$\boldsymbol{Y}_{k-1} = [y_{k-1}, \ldots, y_{k-p}, x_{k-1}, \ldots, x_{k-q}]^T \qquad (2.10)$$

The transfer-function in the Z-domain is

$$Y(z)/X(z) = (1 + b_1 * z^{-1} + \ldots + b_q * z^{-p}) / (1 - a_1 * z^{-1} - \ldots - a_p * z^{-p}). \qquad (2.11)$$

The time-varying ARMA(p,q) spectrum is

$$S(f) = \sigma_x * (1 + \Sigma b_l * e^{-j\,2\pi l f \Delta T}) / (1 - \Sigma a_i * e^{-j\,2\pi i f \Delta T}) \quad l=1..q,\ i=1..p \qquad (2.12)$$

with the sampling interval $\Delta T = 1/f_0$.

In principle, the same model order selection criteria as for AR models can be applied. However, two parameter, *p* and *q*, instead of one must be identified a-priori. Shinn Yih Tseng

et al. (1995) found that the optimal model order is on average 8.67 and 6.17 for an AR and ARMA model, respectively. However, they found that the AR model can efficiently represent about 96% of the 900 segments, and the ARMA model can efficiently represent only about 78% of them. The authors, therefore, conclude that the AR model is preferable for estimating EEG signals.
Refering to Isaksson (1975), Blechschmid (1982) states that an argument against the use of an ARMA model is that the power spectrum can be negative. The meaning of such negative power would be unknown. Haykin (1996) writes "a*n important feature of the MEM spectrum is that it is nonnegative at all frequencies, which is precisely the way it should be*". In other words, only the AR-spectrum (2.7) but not an ARMA spectrum (2.12) is a maximum entropy method (MEM).

2.3 Kemp's feedback loop model

The feedback loop model (Kemp and Blom, 1981, Kemp 1983) is a simple model, using a bandpass in a feedback loop and a low-pass filter for modeling the volume conduction effect (Fig 4). Similar to an AR model, the feedback loop model is a white noise driven, parametric model for the EEG. The model is able to explain variations of one frequency component by varying one parameter, the feedback gain factor *p(t).* This means the sychronization and desynchronization is explained solely by the feedback gain factor *p*. Kemp's model was used to measure alpha attenuation (Kemp 1983) and to analyze sleep EEG (Mortazaev, et al. 1995).

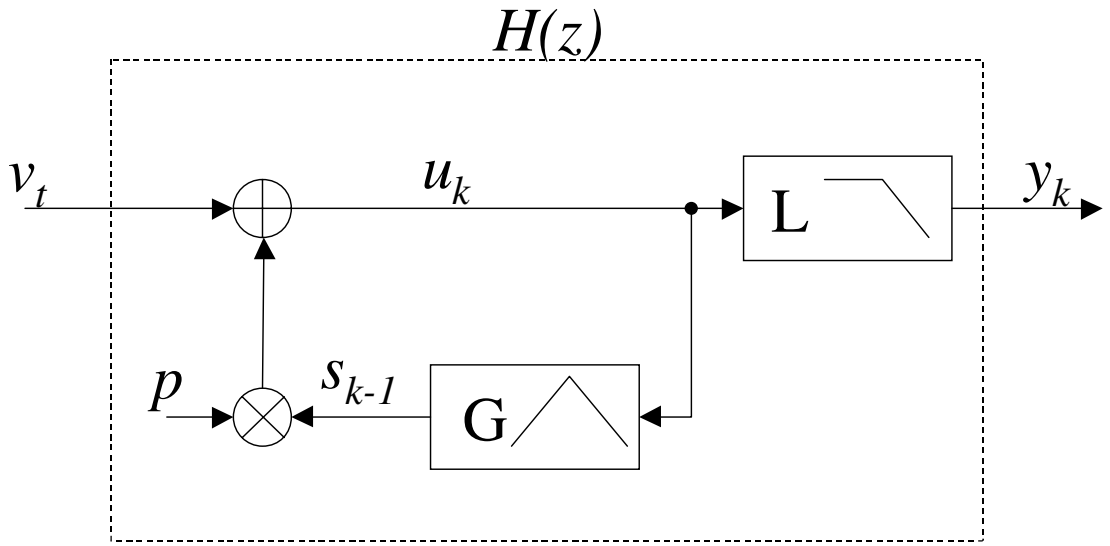


Figure 4: Feedback loop model (Kemp, 1983). v_k is random noise input, L is a low-pass that considers the volume conduction effect, G is a bandpass of the dominant frequency component, p is the feedback gain factor, y_k is the model output, i.e. the observed EEG (adapted from Schlögl et al. 1998b).

It is assumed that the EEG is generated by a white noise driven system, where the strength of the feedback is modulated by the gain factor *p*. According to Kemp and Lopes da Silva (1991), the model is described by the following equations:

$$u_k = c_0 y_k + c_1 y_{k-1} \qquad (2.15)$$

$$s_k = a_1 s_{k-1} + a_2 s_{k-2} + b_0 y_k + b_1 y_{k-1} + b_2 y_k \qquad (2.16)$$

$$u_k = v_k + p * s_{k-1} \qquad (2.17)$$

with

$$c_0 = 1/2+1/(2\pi)F_s / F_c \quad (2.18)$$
$$c_1 = 1/2-1/(2\pi)F_s / F_c \quad (2.19)$$

$$a_1 = (8-2.(2\pi F_0 / F_s)^2) / (4+(2\pi F_0 / F_s)^2+4\pi B / F_s) \quad (2.20)$$
$$a_2 = (-4+4\pi B / F_s-(2\pi F_0 / F_s)^2) / (4+(2\pi F_0 / F_s)^2+4\pi B / F_s) \quad (2.21)$$
$$b_0 = 4\pi B(1/\pi/F_c+1/F_s) / (4+(2\pi F_0 / F_s)^2+4\pi B / F_s) \quad (2.22)$$
$$b_1 = 4.\pi.B(-2/\pi/F_c) / (4+(2\pi F_0 / F_s)^2+4\pi B / F_s) \quad (2.23)$$
$$b_2 = 4.\pi.B(1/\pi/F_c - 1/F_s) / (4+(2\pi F_0/F_s)^2+4\pi B / F_s) \quad (2.24)$$

F_c is the low-pass cut-off frequency of L, F_0 and B are the mean frequency and the bandwidth of the band-pass G in the feedback loop. For time-discrete systems the frequency F_c should be pre-warped (Oppenheim and Schafer, 1975)

$$F_c = F_s * \arctan(\pi F_c'/F_s) / \pi \quad (2.25)$$

Using the Z-transformation, the convolution operator (2.16-17) can be replaced by the multiplication

$$U(z) = V(z) + p * S(z)* z^{-1} \quad (2.26)$$
$$S(z) = G(z) * U(z) \quad (2.27)$$
$$Y(z) = L(z) * U(z) \quad (2.28)$$

$L(z)$ and $G(z)$ can be derived from equations (2.15-16,) and (2.27-28)

$$U(z) = c_0 Y(z) + c_1 Y(z) . z^{-1} \quad (2.29)$$
$$L(z) = Y(z) / U(z) = 1 / (c_0 + c_1 z^{-1}) = 1 / C(z) \quad (2.30)$$

$$S(z) = a_1 S(z) z^{-1} + a_2 S(z) z^{-2} + b_0 Y(z) + b_1 Y(z) z^{-1} + b_2 Y(z) z^{-2} \quad (2.31)$$

$$Y(z) / S(z) = L(z) G^{-1}(z) = $$
$$= (1 - a_1 z^{-1} - a_2 z^{-2}) / (b_0 + b_1 z^{-1} + b_2 z^{-2}) \quad (2.32)$$

$$G(z) = (b_0 + b_1 z^{-1} + b_2 z^{-2}) / [(1 - a_1 z^{-1} - a_2 z^{-2}) * (c_0 + c_1 z^{-1})]$$
$$= B(z) / [A(z) C(z)] \quad (2.33)$$

The feedback loop has the following (forward) transfer function

$$U(z) / V(z) = 1 / [1 - p z^{-1} G(z)] \quad (2.34)$$

The input-output analysis gives a transfer function of

$$H(z) = Y(z) / V(z) = [Y(z) / U(z)] . [U(z) / V(z)] = \quad (2.35)$$
$$= L(z) / [1 - p z^{-1} G(z)] = [1 / C(z)] / [1 - p z^{-1} B(z) / [A(z) C(z)]]$$

$$H(z) = A(z) / [A(z)C(z) - p z^{-1} B(z)] \quad (2.36)$$
$$= (1- a_1 z^{-1}-a_2 z^{-2}) / [(1-a_1 z^{-1}- a_2 z^{-2}) (c_0+ c_1 z^{-1}) - p(b_0 z^{-1} + b_1 z^{-2} + b_2 z^{-3})]$$

$$H(z) = c_0^{-1} (1 -a_1 z^{-1}-a_2 z^{-2}) / \ldots$$

$$\ldots / [1-(a_1-c_1/c_0-p\ b_0/c_0)\ z^{-1} + (a_2 - a_1\ c_1/c_0 - pb_1/c_0)\ z^{-2} + (a_2\ c_1/c_0 - p\ b_2/c_0)\ z^{-3}] \quad (2.37)$$

Equation (2.37) represents the transfer function of the feedback loop model and exhibits that it is an ARMA(3,2) model with a normalizing factor of $1/c_0$

$$y_k -(a_1-c_1/c_0-p\ b_0/c_0)\ y_{k-1} + (a_2 - a_1\ c_1/c_0 - p\ b_1/c_0)\ y_{k-2} + (a_2\ c_1/c_0 - p\ b_2/c_0)\ y_{k-3} =$$
$$= 1/c_0\ (v_k - a_1\ v_{k-1} - a_2\ v_{k-2}) \quad (2.38)$$

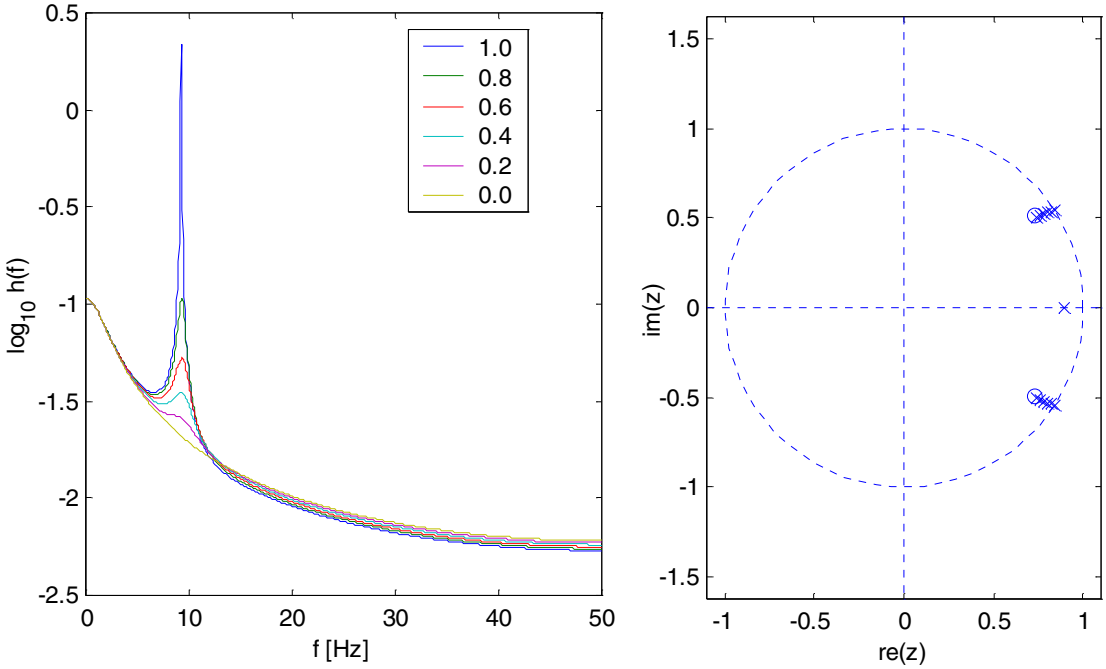


Figure 5: Transfer function of the feedback loop model for different gain factors *p*. (adapted from Schlögl et al. 1998b). The sampling rate is F_s = 100Hz, the low pass with cut-off frequency F_c = 1.8Hz, center frequency F_0 =10Hz, bandwidth B = 4 Hz. The filter coefficients can be calculated by equations (2.18-24); the frequency response *h(f) is* obtained by *H(z)* with *z=exp(j2π/Fs*f)* (left side). The pole-zeros diagram (in the complex Z-plane) is displayed at the right side. There is one pole (x) on the real axis (representing the low-pass) and a conjugate complex pair of zeros and poles. In case p = 0, the poles and zeros are equal and diminish. Increasing the feedback gain p results in moving of the poles towards the unit circle, while the zeros are constant.

Fig. 5. shows that an increase of the feedback gain p also increases the 10Hz-component. A variation in time of the gain factor modulates the amplitude of this spectral component. The model explains variations of one frequency component in the EEG by a variation of a gain factor in a feedback loop. The low-pass cut-off frequency F_c, the center frequency F_0 and bandwidth B must be known or have do be determined a-priori. In any case, variations of these terms are neglected. This is a disadvantage; an AR model does not need such assumptions. Individual variations are considered within the estimated parameter. However, the feedback loop model explains how an internal state-dependant variable can modulate the EEG spectrum. The above derivation shows that an EEG that is generated with Kemp's feedback loop model can be equally described by an ARMA(3,2) model. (Schlögl et al. 1998b).

2.4 The lumped circuit model

The lumped circuit model describes the behavior of a large population of neurons (Lopes da Silva et al. 1974, Zetterberg et al. 1978). The lumped model can be positioned between models of single cells and global models of EEG dynamics (Suffczinski et al. 1999). The lumped circuit model explains how the feedback loops between TCR and RE cells can generate the EEG alpha rhythm (ca. 8-13 Hz) (Lopes da Silva et al. 1974) and non-linear dynamics. Suffczinsky et al. (1999) extended the lumped circuit model to two and more modules and could explain the antagonistic ERD/ERS behavior of neighboring modules.

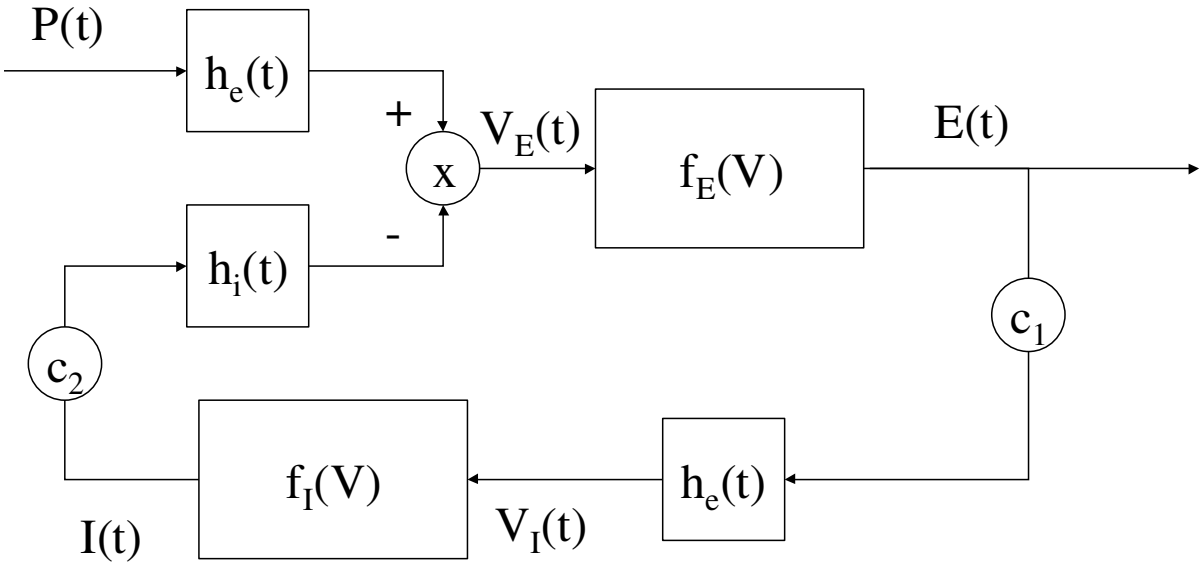


Figure 6: Block diagram of the lumped circuit model for a simplified alpha rhythm model. The thalamo-cortical relay (TCR) cells are represented by two input devices $h_e(t)$ and $h_i(t)$. These generate potentials which stimulate an excitatory and an inhibitory postsynaptic potential (EPSP and IPSP). $h_e(t)$ and $f_I(V)$ model the RE cells in the feedback loop. $f_E(V)$ and $f_I(V)$ represent the spike generating process; $E(t)$, $I(t)$ and $P(t)$ are pulse train densities at the TCR output, the RE output and the excitatory TCR input, respectively. The constant c_1 represents the number of RE cells to which one TCR cell projects and c_2 is the number of TCR neurons to which one RE projects. $V_E(t)$ and $V_I(t)$ represent the average membrane potential at the excitatory and inhibitory population, respectively. (Adapted from Lopes da Silva et al. 1974)

Figure 6 shows a lumped circuit model for an alpha rhythm model. The average excitatory potential $V_E(t)$ is the potential which can be measured on the head surface as EEG. The transfer function of the model in the Laplace-domain is given by (Lopes da Silva et al. 1974, Suffczinski et al. 1999)

$$V_E(s)/P(s) = A\,(a_2-a_1)\,(s+b_2)\,(s+b_1) / ((s+a_1)\,(s+a_2)\,(s+b_2)\,(s+b_1)+K) \qquad (2.39)$$

with the feedback gain

$$K = c_1\,c_2\,q_e\,q_i\,(a_2-a_1)\,(b_2-b_1)\,AB \qquad (2.40)$$

The values of the parameter are provided in Suffczinski et al. (1999). The transfer function is linear up to a critical value of the feedback gain K_c = 3.74E8 s^{-4}. The transfer function in the

Laplace domain $V(s)/P(s)$ can be transformed into the Z-domain by applying the bilinear transformation

$$s = 2F_s(z-1)/(z+1) \qquad (2.44)$$

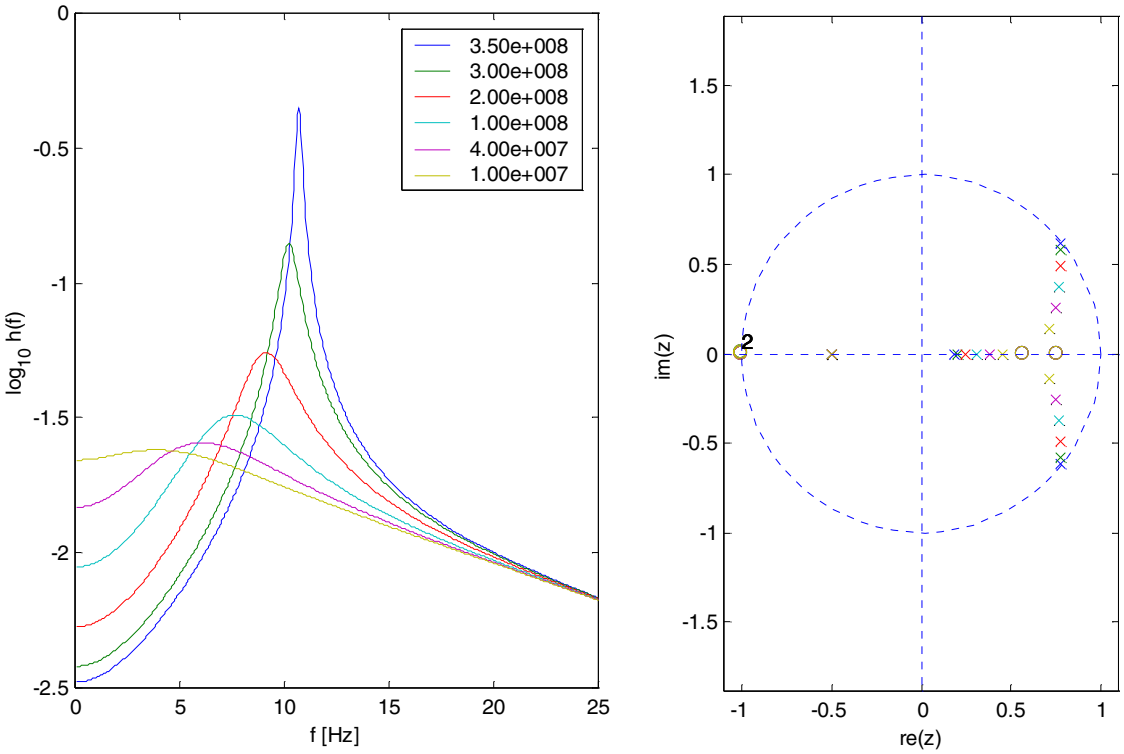


Figure 7: Transfer function of the lumped alpha rhythm model. The left figure shows the spectral density function of the transfer function; the right figure shows the pole-zeros diagram (for a sampling rate of 100Hz). The parameter were chosen accordingly to Suffczinsky et al. (1999). The pole-zeros diagram shows that all poles and zeros (except one pole pair) are on the real axis. The conjugate complex pair of poles moves with increasing K towards the unit circle. Simultaneously, this pole-pair also changes the angle, which corresponds to the shift in the center frequency.

The result of the bilinear transformation shows that the transfer function $V'(z)/P'(z)$ is an ARMA(4,4) model. Fig. 7 shows the spectral density function and the pole-zero diagram for various feedback gains K. In case of a random noise input $P(t)$ (which is actually assumed) the frequency spectrum of V_E is the same as the spectrum of the transfer function. In summary can be said, while the feedback gain K of the lumped model is below a critical value, an equivalent ARMA model can be found and the alpha rhythm of the EEG can be explained.

PART II: THE ADAPTIVE AUTOREGRESSIVE MODEL

3. Theoretical considerations

3.1 Non-stationarity

A classical approach for estimating time-varying AR-parameter is the segmentation based approach. In this case, the data is divided into short segments and the AR parameters are estimated from each segment. The result is a time-course of the AR parameters that describes the time-varying characteristics of the process. The segment length determines the accuracy of the estimated parameters and defines the resolution in time. The shorter the segment-length, the higher is the time resolution but this has the disadvantage of an increasing error of the AR estimates.

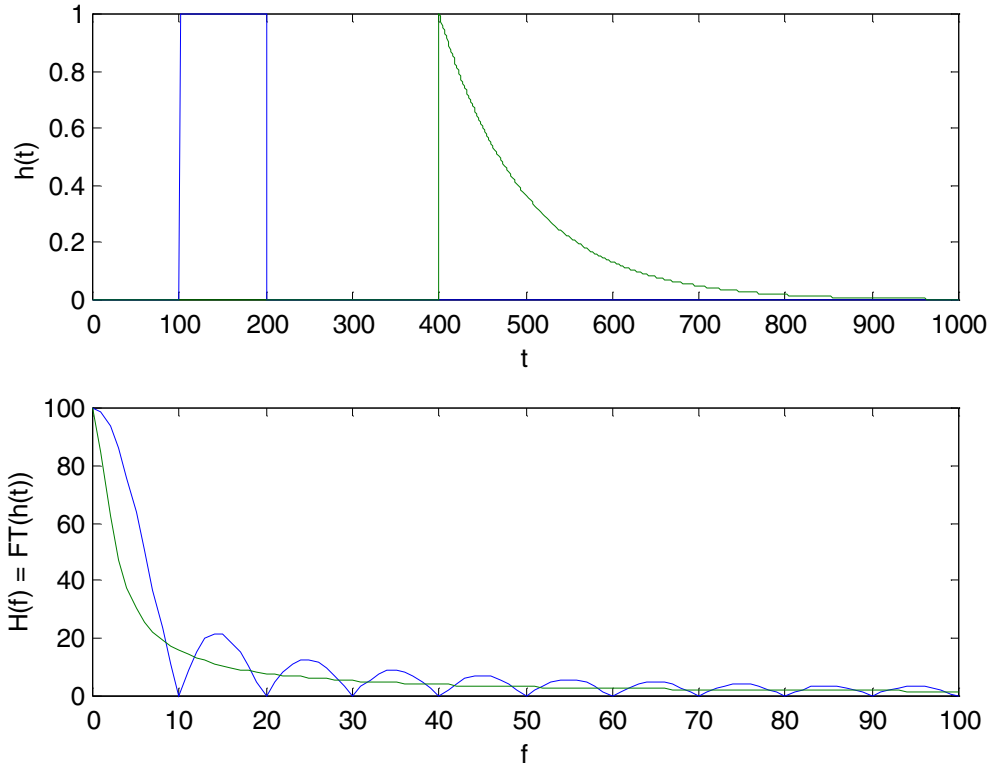


Figure 8: Exponential and a rectangular window. The upper panel displays the two types of windows in the time domain. The rectangle is 100 samples long, the exponential window has a time constant of 100 samples, and the area under both windows is 100. The lower panel displays the Fourier transform of both window functions. The smooth curve corresponds to the exponential window.

Basically, the segmentation based methods (Jansen et al. 1981, Florian and Pfurtscheller, 1995) implicitly use a rectangular window for the convolution; sometimes also a triangular window is used. The adaptive algorithms (LMS, RLS, etc., Akay, 1994, Patomäki et al. 1995, 1996, Bianchi et al. 1997) applied implicitly a one-sided window similar or equal to an exponential window. Beside these two major groups of algorithms, also time-varying AR estimation with some optimize windowing function can be found (Kaipio and Karjaleinen, 1997a,b, Bianchi et al. 1997). The window function is important because it determines the resolution in the time- and frequency domain.

Comparing the different approaches requires the consideration of many factors. Segmentation-based approaches can select segments subsequently, otherwise an overlap can be considered. If the overlap is the segment length minor 1, we speak of a sliding window approach. Only this type of algorithm provide estimates with the same time resolution as the adaptive algorithms. However, in this case, the computational effort can grow with the window length. Only in case of a rectangular window the algorithm can be fastened if the update is made in the following manner:

$$a(t) = a(t\text{-}1) + \alpha(t) - \alpha(t\text{-}L) \qquad (3.1)$$

In any case, the past L samples (L is the segment length) must be stored in memory. Alternatively, the adaptive algorithms can perform calculation concurrent to the data acquisition, where no buffering is required and the update can be made in the form

$$a(t) = \lambda\, a(t\text{-}1) + \alpha(t) \qquad (3.2)$$

requiring low computational effort. For this reason, the adaptive algorithms are well suited for on-line analysis.

At this point, the principle of uncertainty between time- and frequency domain (POU) must be mentioned. The POU means that "*The more accurately we try to determine Y(t,f) as function of time, the less accurately we determine it as a function of frequency, and vice versa*"; in other words "*In determining evolutionary spectra, one cannot obtain simultaneously a high degree of resolution in both the time domain and the frequency domain*" (Priestley, 1981). This statement addresses a problem of non-stationary spectral analysis which also has to be considered when time-varying AR parameters are used for spectral estimation. Because of this consideration, we assume an upper limit for the adaptation speed of an AAR(p)-model. For this reason, the type of smoothing window seems to be of less importance. However, the computational time-resolution must be distinguished from the actual time-resolution of the model estimates.
Another problem is addressed by Haykin (1999) '*The issue whether there will be one ultimate time-frequency resolution that effectively describes all time-varying signals, or whether there will be a number of densities tailored for individual applications is a major issue in the field* (of signal processing) *and unsettled'*. Also this issue should be taken into account if a method for time-varying spectral estimation is used.

3.2 Adaptive autoregressive model

It was shown how an autoregressive model can be used to describe the EEG in the stationary case. In order to consider the non-stationarity (i.e. the variation in time), the AR parameters (2.1) are allowed to vary in time. This is described by the following equation

$$y_k = a_{1,k} * y_k + \ldots + a_{p,k} * y_{k\text{-}p} + x_k\,, \qquad x_k = N\{0,\ \sigma_x^2(k)\} \qquad (3.3)$$

x_k is a zero-mean-Gaussian-noise process with variance $\sigma_x^2(k)$; the index k is usually an integer and describes discrete, equidistant time points. The time t in seconds is $t = k / f_0 = k*\Delta T$ with the sampling rate f_0 and the sampling interval $\Delta T = 1/f_0$. $y_{k\text{-}i}$ with $i = 1..p$ are the p previous sample values, p is the model order; $a_{i,k}$ are the time-varying AR model parameters.

Later, it will be shown how the time-varying AR parameters can be estimated in an adaptive way. In this case, the parameters are called adaptive autoregressive (AAR) parameters. Accordingly, an AAR model is an autoregressive model whose parameters are estimated adaptively. Emphasizing the model order can be done by the notion of an AAR(p) model.

The difference to the conventional AR model is 'only' that the AAR parameters (as well as the variance σ_x^2) are allowed to vary in time. It is assumed that only a small non-stationarity takes place; this will be called nearly stationary. Some upper limit of the adaptation rate can be assumed, this means the AR parameters change only '*slowly*' with time. It is assumed that the changes of the AAR-parameters within one interation are smaller than the estimation error. If the assumptions are fulfilled, the process is nearly stationary. Then, an AAR-model can be used to describe the time-variation (i.e. non-stationarity) in the data. If those assumptions are not fulfilled, the process is highly non-stationary; some transient event occurs which can not be described by the AAR parameter. This case will be investigated in chapter 7 in more detail.

The AR model also explains the spectral composition of a signal (shortly the AR spectrum); Accordingly, AAR parameters describe a 'time-varying' or 'evolutionary' spectrum $S(\omega,t)$:

$$S(f,t) = \sigma_x(k)\,(1+\Sigma_l\, b_{l,k}\, e^{-j\,2\pi l f \Delta T})\,/\,(1-\Sigma_i\, a_{i,k}\, e^{-j\,2\pi i f \Delta T})$$
$$l=1..q,\ i=1..p, \qquad (3.4)$$

with the sampling rate $f_0 = 1/\Delta T$ is the time t

$$t = k/f_0 = k*\Delta T \qquad (3.5)$$

At this point, it should be noted that the AR parameters of a given spectral density also depend on the sampling rate. As a consequence, AR parameters of different sampling rates cannot be compared directly. This became a problem especially in multi-center data (data set D1 and D4). For this reason, a method of re-sampling was developed which is described in Appendix B.

3.3 Innovation process, prediction error and Goodness-of-fit

An AR-model can be used to describe the EEG as filtered white noise. In case the filter input is zero, also the output is flat. In other words, x_k "drives" the output y_k. Furthermore, x_k is a measure for the new information, since the predictable part is already determined by the past p samples. For this reason, x_k is called the "innovation process" and is orthogonal to all past values y_{k-i}, $i>0$ (Haykin, 1996, pp. 303-4, 307-8)

$$x_k = y_k - \boldsymbol{a}_{k-1}^T * \boldsymbol{Y}_{k-1}. \qquad (3.6)$$

Because the predicted value $\boldsymbol{a}_{k-1}^T * \boldsymbol{Y}_{k-1}$ is uncorrelated to the innovation process x_k, the variances are related by the following equation

$$var\{x_k\} = var\{y_k\} - var\{(\boldsymbol{a}_{k-1}^T * \boldsymbol{Y}_{k-1})\} \qquad (3.7)$$

In practice, the AAR parameters $\boldsymbol{a}_k$ are only estimated values $\hat{\boldsymbol{a}}_k$. If the estimates are near the true values, the prediction error e_k will be quite close to the innovation process.

$$e_k = y_k - \hat{\boldsymbol{a}}_{k-1}^T * \boldsymbol{Y}_{k-1} =$$

$$= y_k - \boldsymbol{a}_{k-1}^T * \boldsymbol{Y}_{k-1} + (\boldsymbol{a}_{k-1}-\hat{\boldsymbol{a}}_{k-1})^T * \boldsymbol{Y}_{k-1} = \quad (3.8)$$
$$= x_k + (\boldsymbol{a}_{k-1}-\hat{\boldsymbol{a}}_{k-1})^T * \boldsymbol{Y}_{k-1}$$

The AAR estimates at time k-1 depend only on previous samples. They do not depend on samples at time $t = k * \Delta T$ or larger. Actually, this is the case in all estimation algorithms investigated in this work. Exceptions would be Kalman smoother algorithm (Grewal and Andrews 1993, Haykin, 1996). Hence, the prediction error e_k is independent of all previous samples y_{k-i}, i>0; the one-step prediction process $(\hat{\boldsymbol{a}}_{k-1}^T * \boldsymbol{Y}_{k-1})$ is uncorrelated to the prediction error process e_k; both processes are orthogonal. Hence, the total variance of the signal is

$$var\{y_k\} = var\{(\hat{\boldsymbol{a}}_{k-1}^T * \boldsymbol{Y}_{k-1})\} + var\{e_k\} \quad (3.9)$$

Alternatively, the terms *MSE* (mean squared error) and *MSY* (mean squared signal Y) might be used

$$MSE = var\{e_k\} = mean[e_k^2] \quad (3.10)$$

$$MSY = var\{y_k\} = mean[y_k^2] \quad (3.11)$$

Equation (3.8) shows that the prediction error consists of the innovation process and the prediction error due to the estimation error of the parameter. This estimation error $\boldsymbol{a}_{k-1}-\hat{\boldsymbol{a}}_{k-1}$ is caused by the estimation algorithm and changes of the 'true' model parameter $\boldsymbol{a}_k$. In the optimal case, when $\boldsymbol{a}_k=\hat{\boldsymbol{a}}_k$, the one-step prediction error e_k is equal to the innovation process x_k, and the mean squared error would be minimal. The mean square of e_k (*MSE*, equation 3.10) increases with bad estimates. Hence the *MSE* is a measure for the goodness-of-fit of the AAR estimates $\hat{\boldsymbol{a}}_k$.

In terms of neural network learning, it can be said that the past values y_{k-i} are used for estimating the model parameters (at time $k*\Delta T$) and the present value y_k is used for evaluation of the model validity. For this reason, the mean square of the prediction error (*MSE*) can be used as a criterion for measuring the goodness of fit. The smaller the *MSE* is, the better the estimated AAR parameters $\hat{\boldsymbol{a}}_k$ describe the process.

$$REV = MSE/MSY \quad (3.12)$$

For comparing the results of different data sets, the *MSE* is normalized by the variance of the signal (*MSY*) for obtaining the "relative error variance" *REV*. Because *MSY* is constant for a given data series, the REV can be used (instead of *MSE*) as a measure for the goodness-of-fit.

REV provides the ratio how much of the signal is not explained by the model estimates. *REV*=1 means that the model parameters are zero and the signal is white noise; *REV* = 0 means that the signal can be explained completely by the model (a theoretical consideration only). If 0 < *REV* < 1, *REV* tells us how much of the signal is contributed by the white noise process and how much stems from the one-step prediction process. Consequently, *REV* expresses how much of the signal is not explained by the model parameters. The smaller the *MSE* (and *REV*) is, the larger is the part of the signal variance (*MSY-MSE*) that is explained by the estimated AAR model parameter. In this sense, 1-*REV* is a measure for the goodness-of-fit and describes how well the AAR estimates explain the observed signal y_k.

3.4 Adaptive inverse autoregressive filtering

Previously, it was described that an AAR model can be used to model slowly varying, nearly stationary stochastic processes. In this case the estimates are close to the parameters and reliable AR parameters and AR spectra can be obtained. However, artifacts (like muscle activity) can cause transient, highly non-stationary, patterns in EEG recordings. Bodenstein and Praetorius (1977), Praetorius et al. (1977) and Lopes da Silva et al. (1977) applied (stationary) autoregressive models for identifying spikes and transient phenomena in the EEG. The basic idea of inverse AR filtering is that in a first step, the AR parameters of some EEG segments are identified. In the second step, the AR filter parameters are applied inversely to the EEG. Transient phenomena are detected from this inverse filtered process (see Fig. 9).

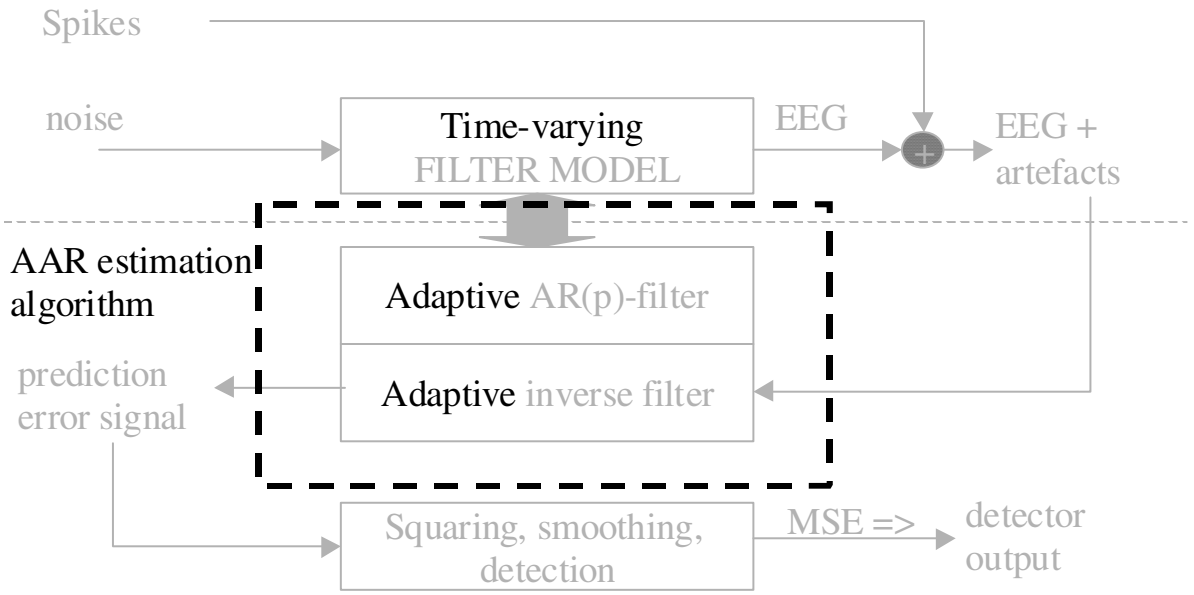


Figure 9: Principle of adaptive inverse filtering. The gray parts show the principle of inverse filtering for detection of spikes or other transient events (Lopes da Silva et al. 1977). The AAR estimation algorithm identifies the AR-filter parameter and simultaneously calculates the one-step prediction error process. The black parts indicate the differences to the (stationary) inverse autoregressive filtering.

The scheme of inverse filtering can be modified to a scheme of adaptive inverse filtering. It will be shown in later chapters, that all AAR estimation algorithms calculate the filter coefficients and the one-step prediction error. In other words, both steps, model estimation and inverse filtering, are performed simultaneously. The principle of adaptive inverse filtering can also be used to detect transient events. An example for this approach is provided in chapter 7, where the AAR model was applied for detecting muscle and movement artifacts in the sleep EEG (Schlögl et al. 1998c, 1999f). The behavior of the AAR estimation in case of transients in the EEG and the contradiction to the basic assumption of near stationarity is discussed in more detail in section 8.2 .

4. Estimation algorithms

AAR estimation algorithms have the advantage of providing a high (computational) time resolution with low computational effort and they are also well suited for on-line analysis. Such methods are the Least-Mean-Squares (LMS) approach (Schack et al. 1993, Haykin, 1996, Widrow and Stearns, 1985), the Recursive-Least-Squares (RLS) approach (Akay, 1994, Patomäki et al. 1995, 1996, Mainardi et al. 1995), Recursive AR (RAR) techniques (Bianchi et al. 1997) as well as Kalman filtering (Isaksson, 1975, Mathieu, 1976, Duquesnoy, 1976, Blechschmid, 1977, 1982, Jansen, 1979, 1981). Penny and Roberts (1998) used Kalman filtering based on Jazwinski (1969) and Skagen (1988).

All AAR estimation algorithms have in common the calculation of the one-step prediction error e_k

$$e_k = y_k - \hat{\boldsymbol{a}}_{k-1}^T * \boldsymbol{Y}_{k-1} \qquad (4.1)$$

as the difference between the prediction from the past p samples $\hat{\boldsymbol{a}}_{k-1}^T * \boldsymbol{Y}_{k-1}$ and the actual sample value y_k. In a further step, e_k is used to update the AAR estimates from $\hat{\boldsymbol{a}}_{k-1} \rightarrow \hat{\boldsymbol{a}}_k$.

At first, some simple algorithms, like the Least Mean Squares (LMS) algorithm and a recursive AR algorithm, are presented. The following section deals with Kalman filtering (KF) applying KF to an AR model, also the mathematical derivation is provided. It is shown that different assumptions lead to different variants of the KF algorithm for AAR estimation. This phenomenon explains why different algorithm, all named KF, are used for time-varying AR estimation. A summary of differences is provided; it is also shown that the Recursive-Least Squares (RLS) algorithm is a special case of Kalman filtering.

4.1 Least mean squares and recursive AR methods

The LMS algorithm (Haykin 1986, Widrow and Stearns, 1985, Akay, 1994) is described by the following update equation:

LMS I:

$$\hat{\boldsymbol{a}}_k = \hat{\boldsymbol{a}}_{k-1} + UC/MSY * e_k * \boldsymbol{Y}_{k-1} \qquad (4.2)$$

$\hat{\boldsymbol{a}}_k$ is the estimate of the time-varying AR vector, e_k is the (one-step prediction) error process (4.1). MSY is the variance of the signal y and is used instead of normalizing y to variance 1. The update coefficient UC determines the speed of adaptation, the time-resolution as well as the smoothing of the AR estimates.

LMS II:

The adaptation equation (4.2) uses a constant adaptation rate UC/MSY. Schack et al. (1993) used time-varying adaptation rate in the way that also the normalization factor is estimated adaptively

$$V_k = (1-UC) * V_{k-1} + UC * e_k^2 \qquad (4.3)$$

and the update equation is

$$\hat{\boldsymbol{a}}_k = \hat{\boldsymbol{a}}_{k-1} + UC/V_k * e_k * \boldsymbol{Y}_{k-1}. \qquad (4.4)$$

This algorithm considers a time-varying variance $\sigma_x^2(k)$ of the innovation process, whereby V_k is the estimate of the variance X_k calculated from e_k^2 using the exponential smoothing window with a time constant of *1/UC*. Schack et al. (1993) used different update coefficients in equation (4.3) and (4.4). However, no advantage was found to select different *UC*. Moreover, it was difficult to select an additional free parameter. For these reasons, the same update coefficient was used in both equations (4.3-4).

RAR I:

$$\boldsymbol{A}_k = (1-UC)*\boldsymbol{A}_k + UC*\boldsymbol{Y}_k*\boldsymbol{Y}_k^T \qquad (4.5)$$
$$\boldsymbol{k}_k = UC*\boldsymbol{A}_k*\boldsymbol{Y}_k / (UC*\boldsymbol{Y}_k^T*\boldsymbol{A}_k*\boldsymbol{Y}_k+1) \qquad (4.6)$$
$$\hat{\boldsymbol{a}}_k = \hat{\boldsymbol{a}}_{k-1} + \boldsymbol{k}_k^T * e_k \qquad (4.7)$$

Bianchi et al. (1997) presented a recursive method for estimating the time-varying AR parameter. The difference to the RLS algorithm is that $\boldsymbol{A}_k$ is not the state error correlation matrix (see next section) but rather the covariance matrix of the signal itself. This method will be called in short recursive AR technique (RAR).

RAR II:

The second version of this algorithm, which was proposed by Bianchi et al. (1997), estimated the covariance matrix (4.5) using a 'whale forgetting function'.

$$\boldsymbol{A}_k = c_1*\boldsymbol{A}_{k-1} + c_2*\boldsymbol{A}_{k-2} + c_3*\boldsymbol{Y}_k*\boldsymbol{Y}_k^T \qquad (4.8)$$

Bianchi et al. (1997) used coefficients c_1, c_2 and c_3 which are 1.74, -0.7569, and $(1-c_1-c_2)$, respectively. Here, the whale function is extended to a general approach in order to apply different update coefficients *UC*. The coefficients c_1, c_2 and c_3 were determined such that the zeros $p_{1,2}$ of the polynomial (4.9) are both $1/(1-2*UC)$.

$$1 - c_1*z^{-1} - c_2*z^{-2} = (1- p_1*z^{-1})*(1- p_2*z^{-1}) = \qquad (4.9)$$
$$= (1-(1-2*UC)*z^{-1})*(1-(1-2*UC)*z^{-1}) =$$
$$= 1 - (p_1+p_2)*z^{-1} + p_1*p_2*z^{-2}$$
$$c_3 = 1- c_1 - c_2 \qquad (4.10)$$

It is a simple exercise to derive that the adaptation time constant is approximately $1/UC$. The AAR estimates are adapted by the equations (4.6-8).

4.2 The Kalman filtering and the state-space model

This section introduces the basics about KF, as far as it is important for AAR estimation. Kalman (1960) and Kalman and Bucy (1961) presented the original idea of KF. Meinhold and Singpurwalla (1983) provided a bayesian formulation of the method. Grewal and Andrews (1993) and Haykin (1996) are the main references used in this investigation.

Kalman filtering is an algorithm for estimating the state (vector) of a state space model with the system equation

$$z_k = \boldsymbol{G}_{k,k-1} * z_{k-1} + \boldsymbol{w}_k \qquad (4.11)$$

and the measurement (or observation) equation

$$y_k = \boldsymbol{H}_k * z_k + v_k \qquad (4.12)$$

z_k is the state vector and depends only on the past values of $\boldsymbol{w}_k$ and some initial state z_0. The system is defined for all $k>0$. The output signal y_k, which can be measured, is a combination of the state vector and the measurement noise v_k with zero mean and variance $V_k = \mathrm{E}\{v_k v_k\}$. $\boldsymbol{w}_k$ is the process noise with zero mean and covariance matrix $\boldsymbol{W}_k = E\{\boldsymbol{w}_k * \boldsymbol{w}_k^T\}$. The state transition matrix $\boldsymbol{G}_{k+1,k}$ and the measurement matrix $\boldsymbol{H}_k$ are known and may or may not change with time. Kalman filtering is a method that estimates the state z_k of the system from measuring the output signal y_k with the prerequisite that $\mathbf{G}_{k,k-1}$, $\boldsymbol{H}_k$, V_k and $\boldsymbol{W}_k$ for $k>0$ and z_0 are known.

The inverse of the state transition matrix $\boldsymbol{G}_{k+1,k}$ exists and

$$\boldsymbol{G}_{k,k+1} * \boldsymbol{G}_{k+1,k} = \boldsymbol{I} \qquad (4.13)$$

is the unity matrix $\boldsymbol{I}$. Furthermore, $\boldsymbol{K}_{k,k-1}$, the a-priori state-error correlation matrix and $\boldsymbol{Z}_k$, the a posteriori state-error correlation matrix are used; $\boldsymbol{K}_{1,0}$ is known. In this case, Q_k is the estimated prediction variance that can be calculated by the following formula:

$$Q_k = \boldsymbol{H}_k * \boldsymbol{K}_{k,k-1} * \boldsymbol{H}^T{}_k + V_k \qquad (4.14)$$

It consists of the measurement noise V_k and the error variance due to the estimated state uncertainty $\boldsymbol{H}_k * \boldsymbol{K}_{k,k-1} * \boldsymbol{H}_k^T$. Next, the Kalman gain $\boldsymbol{k}_k$ is determined by

$$\boldsymbol{k}_k = \boldsymbol{G}_{k,k-1} * \boldsymbol{K}_{k,k-1} * \boldsymbol{H}^T{}_k / Q_k \qquad (4.15)$$

Using the next observation value y_k, the one-step prediction error e_k is calculated,

$$e_k = y_k - \boldsymbol{H}_k * \hat{z}_k \qquad (4.16)$$

the state vector z_{k+1} is updated,

$$\hat{z}_{k+1} = \boldsymbol{G}_{k,k-1} * \hat{z}_k + \boldsymbol{k}_k * e_k \qquad (4.17)$$

the a posteriori state vector can be estimated

$$z'_k = \boldsymbol{G}_{k,k'+1} * \hat{z}_{k+1} \qquad (4.18)$$

z'_k is the estimate of the state if y_k is already known. Finally, the a posteriori state-error correlation matrix $\boldsymbol{Z}_k$ and

$$\boldsymbol{Z}_k = \boldsymbol{K}_{k,k-1} - \boldsymbol{G}_{k,k+1} * \boldsymbol{k}_k * \boldsymbol{H}_k * \boldsymbol{K}_{k,k-1} \qquad (4.19)$$

the a-priori (y_{k-1} is unknown) state-error correlation matrix $\boldsymbol{K}_{k+1,k}$ for the next iteration

$$\boldsymbol{K}_{k+1,k} = \boldsymbol{G}_{k+1,k} * \boldsymbol{Z}_k * \boldsymbol{G}_{k+1,k}^T + \boldsymbol{W}_k \qquad (4.20)$$

is updated. Equations (4.14-20) describe one iteration k-1->k of the Kalman filtering algorithm for the state space model (4.11-12). This is the very general form of Kalman filtering. In case of univariate signal processing (one channel), the output signal y_k is one-dimensional; hence also the (co-) variance matrix V_k reduces to a scalar, namely the variance of the signal. For estimating AR parameter with KF, the AR model has to be embedded into the state space model. The AR (2.1) and ARMA model (2.9) are fitted into the state space model as follows:

The state transition matrix $\boldsymbol{G}_{k,k-1}$ as well as its inverse matrix $\boldsymbol{G}_{k-1,k}$ is the identity matrix, which is time-invariant

$$\boldsymbol{G}_{k,k-1} = \boldsymbol{G}_{k-1,k} = \boldsymbol{I} = constant. \qquad (4.21)$$

The time-varying AR parameters (3.3) are identical to the state vector z_k in (4.11)

$$z_k = \boldsymbol{a}_k = [a_{1,k}, \ldots a_{p,k}]^T, \qquad (4.22)$$

the measurement matrix $\boldsymbol{H}_k$ consists of the past p sampling values

$$\boldsymbol{H}_k = \boldsymbol{Y}_{k-1} = [y_{k-1}, \ldots, y_{k-p}]^T, \qquad (4.23)$$

the AR parameters follow a multivariate random walk model

$$\boldsymbol{a}_k = \boldsymbol{a}_{k-1} + \boldsymbol{w}_k \qquad (4.24)$$

with a multivariate zero mean random noise with covariance $\boldsymbol{W}_k$ such that

$$\boldsymbol{w}_k = N(0, \boldsymbol{W}_k). \qquad (4.25)$$

The measurement noise and one-step prediction error is

$$x_k = e_k \qquad (4.26)$$
$$x_k = N(0, \boldsymbol{V}_k) \qquad (4.27)$$

Next, substituting $\boldsymbol{Z}_k$ (4.20) in equation (4.19), assuming that $\boldsymbol{W}_k = UC * \boldsymbol{Z}_k$ and replacing $\boldsymbol{K}_{k,k-1}$ by $\boldsymbol{A}_k$ gives

$$\begin{aligned} \boldsymbol{A}_k &= (1+UC) * (\boldsymbol{A}_{k-1} - \boldsymbol{k}_k * \boldsymbol{Y}_{k-1}^T * \boldsymbol{A}_{k-1}) = \\ \boldsymbol{A}_k &= \boldsymbol{A}_{k-1} - \boldsymbol{k}_k * \boldsymbol{Y}_{k-1}^T * \boldsymbol{A}_{k-1} + \\ &\quad + UC * \boldsymbol{A}_{k-1} - UC * \boldsymbol{k}_k * \boldsymbol{Y}_{k-1}^T * \boldsymbol{A}_{k-1} \end{aligned} \qquad (4.28)$$

which is the Recursive-Least-Squares (RLS) algorithm (Haykin, 1996, p509, Patomäki et al. 1995, 1996). For the general case of KF, the equation with explicit $\boldsymbol{W}_k$ is used

$$\boldsymbol{A}_k = \boldsymbol{Z}_k + \boldsymbol{W}_k \qquad (4.29)$$

Basically, $\boldsymbol{Z}_k$ is the a-posteriori correlation matrix, which is smaller than the a-priori correlation matrix $\boldsymbol{A}_{k-1}$; $\boldsymbol{W}_k$ increases $\boldsymbol{Z}_k$, in order to prevent the 'stalling phenomenon'

(Haykin, 1996, p. 756). Furthermore, a state space model is 'observable' (in terms of system theory) if the observability matrix (Grewal and Andrews, 1993 p.44) has full rank. Because the system matrix is the identity matrix and the observation matrix is a vector of the past p samples, the observation matrix $\Sigma\, \boldsymbol{Y}_k \boldsymbol{Y}_k^T$ has a toeplitz form and is, therefore, non-singlular. Hence, the state space model for the AR parameters is 'observable'. It is important to note that $\boldsymbol{W}_k$ and V_k are not determined by the Kalman equations, but must be know or have to be assumed. Accordingly, different assumptions can result in different algorithms.

4.3 Kalman filtering and the AAR model

Adaptive autoregressive (AAR) parameters are estimated with Kalman filtering as for the iteration from k-1->k. The update equations of Kalman filtering for an AAR model can be summarized by the following equations.

$$e_k = y_k - \boldsymbol{a}_{k-1} * \boldsymbol{Y}_{k-1} \qquad (4.30)$$

$$\boldsymbol{AY}_{k-1} = \boldsymbol{A}_{k-1} * \boldsymbol{Y}_{k-1} \qquad (4.31)$$

$$Q_k = \boldsymbol{Y}_{k-1}^T * \boldsymbol{A}_{k-1} * \boldsymbol{Y}_{k-1} + V_k \qquad (4.32)$$

$$\boldsymbol{k}_k = \boldsymbol{A}_{k-1} * \boldsymbol{Y}_{k-1} / Q_k \qquad (4.33)$$

$$\hat{\boldsymbol{a}}_k = \hat{\boldsymbol{a}}_{k-1} + \boldsymbol{k}_k^T * e_k \qquad (4.34)$$

$$\boldsymbol{X}_k = \boldsymbol{A}_{k-1} - \boldsymbol{k}_k * \boldsymbol{Y}_{k-1}^T * \boldsymbol{A}_{k-1} \qquad (4.35)$$

$$\boldsymbol{A}_k = \boldsymbol{X}_k + \boldsymbol{W}_k \qquad (4.36)$$

$\boldsymbol{Y}_{k-1}^T = [y_{k-1} \ldots y_{k-p}]$ are the past p values, e_k is the one step prediction error, V_k is the variance of the innovation process and x_k, $\boldsymbol{k}_k$ is the Kalman Gain vector. $\boldsymbol{A}_{k-1}$ and $\boldsymbol{X}_k$ are the a-priori and the a posteriori state error correlation matrix. The initial values $\boldsymbol{a}_0$ and $\boldsymbol{A}_0$ are a zero vector and the identity matrix $\boldsymbol{I}$ of order p, respectively.

The product $\boldsymbol{A}_{k-1} * \boldsymbol{Y}_{k-1}$ is required three times, in equations (4.32, 4.33, 4.35) within each iteration. Storing the result in an intermediate variable (4.31) makes the algorithm faster and computationally more efficient. Because of the symmetric properties of the state error correlation matrix $\boldsymbol{A}^T_k = \boldsymbol{A}_k$ for real valued time series, applies the following formula:

$$\boldsymbol{AY}_{k-1} = \boldsymbol{Y}_{k-1}^T * \boldsymbol{A}_{k-1} = (\boldsymbol{A}_{k-1} * \boldsymbol{Y}_{k-1})^T \qquad (4.37)$$

$\boldsymbol{AY}_{k-1}$ is the intermediate variable used in (4.32, 4.33, 4.35) for improving the speed of the calculation.

The remaining unknown parameters are the variance V_k of the innovation process (measurement noise, one-step prediction error) and the covariance matrix $\boldsymbol{W}_k$ which describes the adaptation or speed-of-change of the AR parameter. Isaksson et al. (1981) assumed that the covariance matrix is $\boldsymbol{W}_k = \mu^2 * \boldsymbol{I}$ and the term V_k in (4.32) V_k is 1; Abraham and Leodolter (1983) showed that in case $\boldsymbol{W}_k / V_k = const$, $\boldsymbol{A}_k$, $\boldsymbol{X}_k$, $\boldsymbol{W}_k$, Q_k can be replaced by $\boldsymbol{A'}_k = \boldsymbol{A}_k / V_k$, $\boldsymbol{W'}_k = \boldsymbol{W}_k / V_k$, $Q'_k = Q_k / V_k$, and $V'_k = V_k / V_k = 1$. Accordingly, the algorithm is the same as in (4.30-36) with $V_k = 1$. However, if $\boldsymbol{W}_k / V_k$ is not constant, V_k should be estimated separately. Another version of Kalman filtering (Roberts and Tarasenko, 1992; Roberts, 1997; Penny and Roberts, 1998) turned out to be very different to the former approaches. It can be traced back to Jazwinski (1969) and contains, in essence, a decision whether the prediction error is due to the state uncertainty or due to the innovation process.

The state error correlation matrix is increased only if the prediction error is larger than expected due to the state uncertainty. A consequence is that the covariance $\boldsymbol{W}_k$ is most of the time equal the zero matrix, only at some time points it is larger. Patomäki et al. (1995, 1996) used the Recursive-Least-Squares (RLS) algorithm, which is a special case of Kalman filtering using equations (4.30-36) with $V_k = \mu = 1\text{-}UC$ and $\boldsymbol{W}_k = UC * \boldsymbol{Z}_k$. (Note, for $0<UC<<1$ is $\mu = 1/(1\text{-}UC)$ approx. $\mu = 1+UC$). Looking at the earliest algorithm (Kalman, 1960; Kalman and Bucy, 1961) it can be found that the term V_k does not appear in the update equations, hence $V_k = 0$.

Tables 1 and 2 provide a systematic overview of the different variants of Kalman filtering. It has been shown that the different Kalman filtering algorithms are due to different assumptions about the covariance matrix $\boldsymbol{W}_k$ and the variance V_k. Consequently, the differences can be reduced to differences how to estimate the $\boldsymbol{W}_k$ and V_k. The various versions (totally 7*12 = 84 permutations) of Kalman filtering have been implemented and tested.

Jansen et al. (1979), Blechschmid (1982) and Haykin et al. (1997) describe other variants. Jansen et al. (1979) refers to Duquesnoy (1976), Isaksson (1975) and Mathieu M. (1976) who suggested $\boldsymbol{W}_k = 0$. He used Kalman filtering for identifying the (stationary) AR parameters of short EEG segments. Blechschmid (1982) also used $\boldsymbol{W}_k = 0$, but has investigated earlier periodic re-initialization (Blechschmid, 1977). Haykin et al. (1997) proposed a state transition matrix $\boldsymbol{G} = q*\boldsymbol{I}$ with a q-value slightly smaller than 1. All these methods were implemented, too. The version with $\boldsymbol{W}_k = 0$ worked well for a certain time before it became unstable. Periodic re-initialization and the version with $\boldsymbol{G} = q*\boldsymbol{I}$ showed no further advantages. First attempts also gave a much larger variance of the residual process (MSE) than many of the methods from Table 1 and 2. Therefore, periodic re-initialization was not investigated further.

Table 1: Estimation of the variance of the innovation process V_k in Kalman filtering. The variance V_k is usually not known, it must be estimated from the available data. The following 7 variants of selecting V_k in the Kalman filtering approach were investigated.

Type	Estimation of V_k	References
v1	$V_k = (1\text{-}UC) * V_{k-1} + UC * e_k^2$	Schack et al. (1993)
v2	$V_k = 1$	Abraham and Leodolter (1983), Isakson et al. (1981).
v3	$V_k = 1\text{-}UC$	Akay, (1994), Blechschmid (1982), Haykin (1996), Patomäki et al. (1995, 1996)
v4	$V^+_k = (1\text{-}UC) * V^+_{k-1} + UC * e_k^2$ $V_k = V^+_{k-1}$	cf. (v1), one-step delayed
v5	$q_k = \boldsymbol{Y}_{k-1}^T * \boldsymbol{A}_{k-1} * \boldsymbol{Y}_{k-1}$ $V^+_k = (1\text{-}UC)*V^+_{k-1} +$ $+ UC*(e_k^2 - q_k)$ if $e_k^2 > q_k$ $V^+_k = V^+_{k-1}$ if $e_k^2 \le q_k$ $V_k = V^+_k$	Jazwinski (1969) see (v6)
v6	Same as v5 except $V_k = V^+_{k-1}$	Penny and Roberts (1998)
v7	$V_k = 0$	Kalman (1960) Kalman and Bucy (1961)

Table 2: Variants for estimating the covariance matrix $\boldsymbol{W}_k$. The version a1 is most similar to the RLS-algorithm; all elements of $\boldsymbol{W}_k$ are different to zero. In the versions a2-12, only the diagonal elements are unequal (larger than) zero. The versions a2-a5 are the same as a9-a12 respectively, except for the term (1+*UC*). The latter stemmed from a special formulation of the RLS algorithm. Due to different implementation details in Roberts (1997) and Penny and Roberts (1998), several variants (a6, a7 and a8) of Jazwinski's (1969) versions were implemented.

Type	Estimation of $\boldsymbol{W}_k$	References
a1	$\boldsymbol{W}_k = UC * \boldsymbol{Z}_k$	Akay (1994), Haykin (1996)
a2	$\boldsymbol{Z}_k = (\boldsymbol{I} - \boldsymbol{k}_k * \boldsymbol{Y}_{k-1}^T) * \boldsymbol{A}_{k-1}$ $\boldsymbol{W}_k = \boldsymbol{I} * UC * trace(\boldsymbol{Z}_k)/p$	
a3	$\boldsymbol{Z}_k = (\boldsymbol{I} - \boldsymbol{k}_k * \boldsymbol{Y}_{k-1}^T) * \boldsymbol{A}_{k-1}$ $\boldsymbol{W}_k = \boldsymbol{I} * UC * trace(\boldsymbol{A}_{k-1})/p$	
a4	$\boldsymbol{Z}_k = (\boldsymbol{I} - \boldsymbol{k}_k * \boldsymbol{Y}_{k-1}^T) * \boldsymbol{A}_{k-1}$ $\boldsymbol{W}_k = UC * \boldsymbol{I}$	
a5	$\boldsymbol{Z}_k = (\boldsymbol{I} - \boldsymbol{k}_k * \boldsymbol{Y}_{k-1}^T) * \boldsymbol{A}_{k-1}$ $\boldsymbol{W}_k = UC^2 * \boldsymbol{I}$	Isaksson et al. 1981
a6	$\boldsymbol{Z}_k = \boldsymbol{A}_{k-1} * V_{k-1} / Q_k$ $q_k = (1-UC) * q_{k-1} +$ $+UC * (e_k^2 - Q_k) / (\boldsymbol{Y}_{k-1}^T * \boldsymbol{Y}_{k-1})$ $\boldsymbol{W}_k = q_k * \boldsymbol{I}$ if $q_k > 0$ $\boldsymbol{W}_k = 0;$ if $q_k \le 0$	personal communications (S.J. Roberts, 1997)
a7	$\boldsymbol{Z}_k = \boldsymbol{A}_{k-1} * V_k / Q_k$ $q_k = (1-UC) * q_{k-1} +$ $+ UC * (e_k^2 - Q_k) / (\boldsymbol{Y}_{k-1}^T * \boldsymbol{Y}_{k-1})$ $\boldsymbol{W}_k = q_k * \boldsymbol{I}$ if $q_k > 0$ $\boldsymbol{W}_k = 0;$ if $q_k < 0$	same as (a6) but V_k used instead of V_{k-1}
a8	$Q2_{,k} = \boldsymbol{Y}_{k-1}^T * \boldsymbol{Z}_{k-1} * \boldsymbol{Y}_{k-1} + V_k$ $q_k = (1-UC) * q_{k-1} +$ $+ UC * (e_k^2 - Q2_k) / (\boldsymbol{Y}_{k-1}^T * \boldsymbol{Y}_{k-1})$ $\boldsymbol{W}_k = q_k * \boldsymbol{I}$ if $q_k > 0$ $\boldsymbol{W}_k = 0;$ if $q_k \le 0$	Jazwinski (1969) Penny and Roberts (1998)
a9	$\boldsymbol{Z}_k = (\boldsymbol{I} - (1+UC) * \boldsymbol{k}_k * \boldsymbol{Y}_{k-1}^T) * \boldsymbol{A}_{k-1}$ $\boldsymbol{W}_k = \boldsymbol{I} * UC * trace(\boldsymbol{Z}_k)/p$	
a10	$\boldsymbol{Z}_k = (\boldsymbol{I} - (1+UC) * \boldsymbol{k}_k * \boldsymbol{Y}_{k-1}^T) * \boldsymbol{A}_{k-1}$ $\boldsymbol{W}_k = \boldsymbol{I} * UC * trace(\boldsymbol{A}_{k-1})/p$	
a11	$\boldsymbol{Z}_k = (\boldsymbol{I} - (1+UC) * \boldsymbol{k}_k * \boldsymbol{Y}_{k-1}^T) * \boldsymbol{A}_{k-1}$ $\boldsymbol{W}_k = UC * \boldsymbol{I}$	
a12	$\boldsymbol{Z}_k = (\boldsymbol{I} - (1+UC) * \boldsymbol{k}_k * \boldsymbol{Y}_{k-1}^T) * \boldsymbol{A}_{k-1}$ $\boldsymbol{W}_k = UC^2 * \boldsymbol{I}$	

5. Comparison of the AAR estimation algorithms

The previous section showed that many different algorithms (LMS, RAR, RLS, Kalman filtering) for estimating AAR model parameters are available. Some free parameters like the update coefficient UC and the model order p must be selected in advance. Moreover, additional a-priori assumptions must be made in Kalman filtering, Especially the variance of the innovation process V_k (time-varying or not) as well as the covariance matrix $\boldsymbol{W}_k$ of the multivariate random walk of the AR parameters are required. Still, we do not know how the covariance matrix $\boldsymbol{W}_k$ should be selected, we only know it should have the form of a covariance matrix (symmetric, positive definite, main diagonal elements must be positive and the largest elements).

In section 3.3 , a criterion for the goodness-of-fit was introduced, namely the relative error variance REV (i.e. normalized *MSE*). It was said that the REV measures the part of the variance that is not explained by the model estimates. Hence, the smaller REV is, the better the AAR estimates describe the observed process Y_k. In the following, the REV will be used to compare the AAR estimation algorithms. These algorithms were applied to five real world data sets (data set D5, Appendix E); the model order was fixed for each data set (Table 6). An ARMA-model was also estimated with each algorithm. Furthermore, the update coefficient was varied over 10 decades in order to find the optimal value for each algorithm.

Twelve versions for estimating $\boldsymbol{A}_k = \boldsymbol{Z}_k + \boldsymbol{W}_k$ have been implemented. These versions can be grouped into 3-4 types. The first form is $\boldsymbol{W}_k = UC^* \boldsymbol{A}_{k-1}$. This is most similar to a RLS approach and relates to the proposal of $\boldsymbol{W}_k = q^*\boldsymbol{A}_k$ version (a1). In the second form $\boldsymbol{W}_k$ is a diagonal matrix, which is under certain conditions (q_k>0) larger than zero (a6-8). That means, $\boldsymbol{W}_k$ is only in some (rare) cases larger than zero. The third types utilize a diagonal matrix of $\boldsymbol{W}_k = q^*\boldsymbol{I}$ with q_k = trace($\boldsymbol{Z}_k$)/p*UC (a2), UC*UC (a5), trace($\boldsymbol{A}_{k-1}$)/p*UC (a3) and $q_k = UC$ (a4). The forth group (a9 - a12) is the same as the third type (a2-a5), except for the calculation of $\boldsymbol{Z}_k$ (see Table 2). The results are very similar to the results of (a2-a5) and are, therefore, not shown in Fig. 10.

Figures 10 and 11 show the REV for the various AAR estimation algorithms with 40 different update coefficients applied to five different data sets. The algorithms and the update coefficients UC with the lowest *REV* provide the best estimates. In cases without converging estimates (a*v7: $V_k = 0$, RAR1 and RAR2), the *REV* is larger, in some cases it is even infinite. These cases are not shown. It should be noted that also in these cases, the *REV* is never smaller than the best estimates. In Figure 11, the dependencies of *REV* from UC of each algorithm are shown . The markes 'o' indicate the following versions and the update coefficients within the parenthesis: a1v1 (10^{-8}), a1v2 (10^{-7}), a1v3 (10^{-6}), a8v4 (0.1), a7v1 (10^{-3}), a1v2 (10^{-2}) are indicated in Fig. 11. In some cases (e.g. a8v4 in S4 UC = 0.1 gives REV = 0.251), the marker was out of the displayed scope.

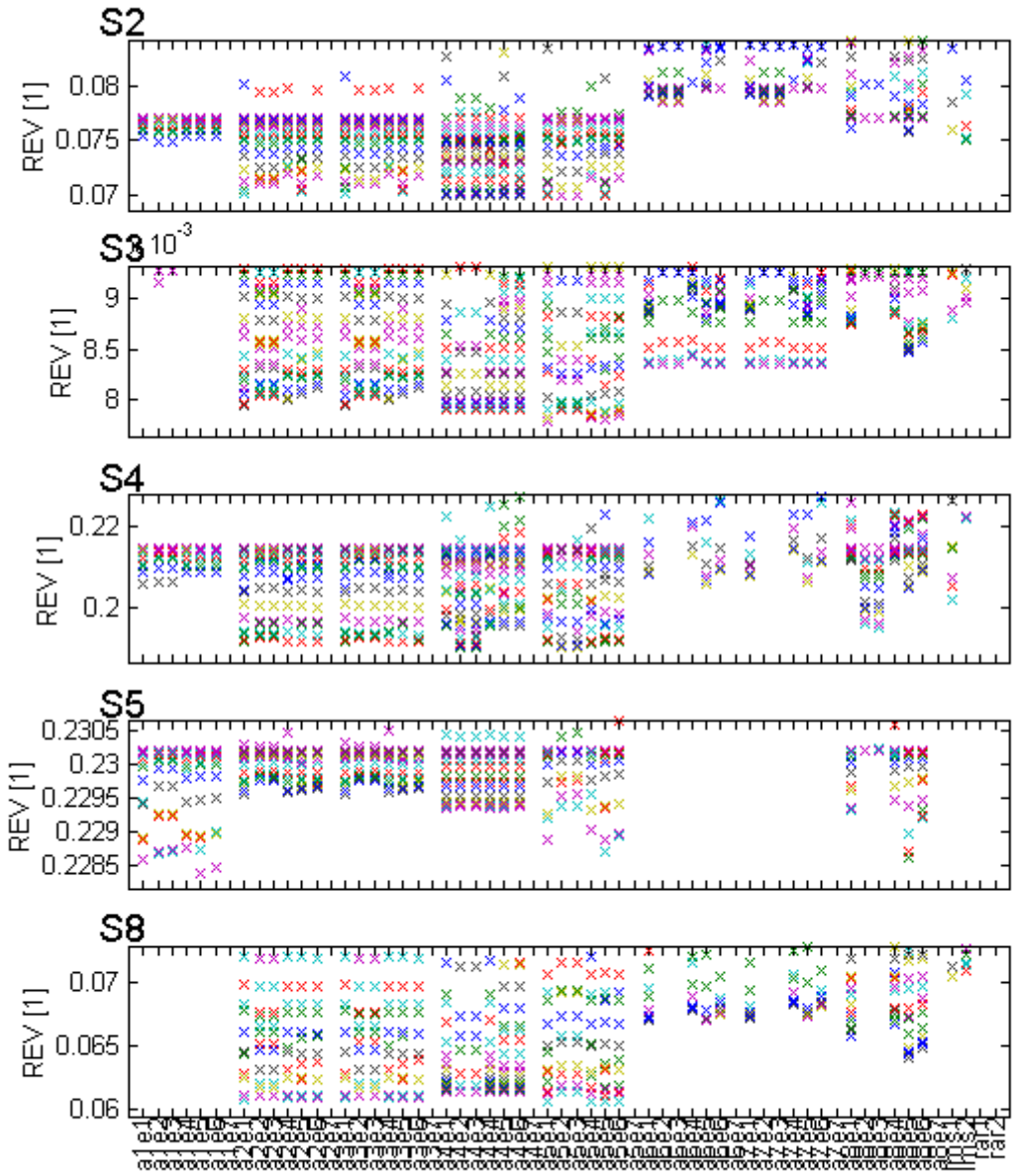


Figure 10: Comparison of different AAR estimation algorithms. Different algorithms are grouped into 8 blocks of 7 KF methods and 4 alternative methods. The 8 blocks correspond to 8 versions for estimating the covariance matrix W_k. Each block contains the 7 versions of estimating the variance of the observation noise process V_k. The update coefficient *UC* was varied in a range over 10 decades and zero. Each cross represents the relative error variance *REV* for a certain UC.

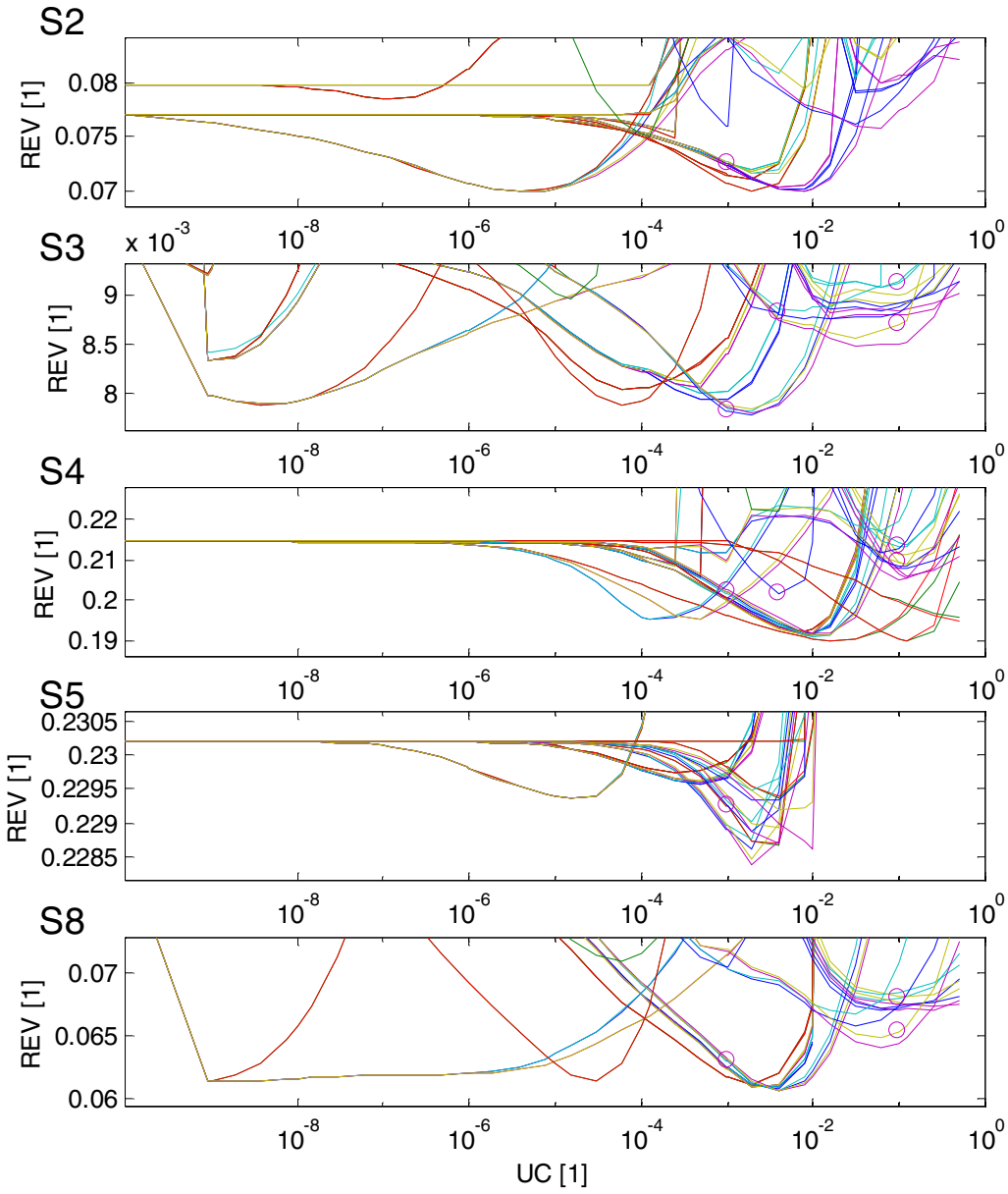


Figure 11: Dependency of *REV* on *UC* for some selected algorithms. The circles 'o' indicate certain update coefficients as described in the text. In some cases, *REV* was too large, in other cases no values were available, which means that the algorithm did not converge and *REV* was infinite. The x-axis has a logarithmic scale except for the leftmost value of *UC* which is zero.

5.1 Estimation of the measurement variance V_k

At first, we look at the estimation methods for V_k. It can be seen that (v2) and (v3) as well as (v1) and (v4) yield quite similar results. The similarity between (v2) and (v3) is due to V_k equal and close to 1, respectively. The difference between (v1) and (v4) is mostly very small, but in some cases (e.g. a8v1 vs. a8v4 in S2, S3, S5; a7v1 vs. a7v4 in S5; a3v1 vs. a3v4 in S2) a difference can be observed leading to the conclusion that v1 is preferable. The question, whether the variance V_k of the innovation process should be chosen as a constant (v2, v3) or should be estimated adaptively (v1), can not be answered generally from these results.

5.2 Estimation of the covariance $\boldsymbol{W_k}$

We can find several good solutions for estimating $\boldsymbol{W}_k$, but for some it can be definitely said that they perform worse. E.g. (a1) works only well for one data set (S4), while for all other data sets it does not; (a8) is good for S4 but not for S1, S2, S3 and S5. It should be taken into account, that data set S4 is the shortest one (less 1000 samples); for this reason, the results from S4 might be not that representative.

In general, it can be said that the version (a1), (a6), (a7) and (a8) do not perform well, (a2) - (a5) seem to be favorable. The versions (a9)-(a12) (not displayed in Fig. 11) give almost the same results as (a2)-(a5), respectively. (a1) uses the estimate $\boldsymbol{W}_k = UC^*\boldsymbol{A}_k$ whereby all elements of the matrix (not only the diagonal elements) are utilized. (a6, a7, a8) use $\boldsymbol{W}_k$ being a diagonal matrix but only in some iteration steps $\boldsymbol{W}_k$ is non-zero, mostly $\boldsymbol{W}_k$ is a zero matrix.

According to these results, it can be said that $\boldsymbol{W}_k$ should be a diagonal matrix larger than zero, which should be added at each iteration to the a posteriori error covariance. The assumption of an uncorrelated, multivariate random walk model for the variation of the AAR parameter (i.e. state vector of the state space model) seems to perform best in all large data sets (S1, S2, S3, S5).

6. Model order and update coefficient

The REV is a measure for the goodness-of-fit; a *UC* can be identified that yields the lowest REV for a fixed order *p* (Fig. 11). The value of this *UC* is called the optimal update coefficient UC_{opt}. Note that in almost all cases, a *REV* smaller than 1 can be identified. E.g. for a6v1 $UC = 10^{-3}$ is optimal in S2 and S4 and $UC = 10^{-2}$ in S1, S3 and S5. In the same way, an optimal update coefficient UC_{opt} for each version can be identified. This means, that each algorithm was stable and gave useful results, not any algorithms became unstable, if only a proper update coefficient was used. However, a fixed model order was used in each case. Next, it will be investigated how the model order *p* and the update coefficient UC_{opt} are related. For this purpose, S3 from data set D5 with various *p* and *UC* was analyzed .

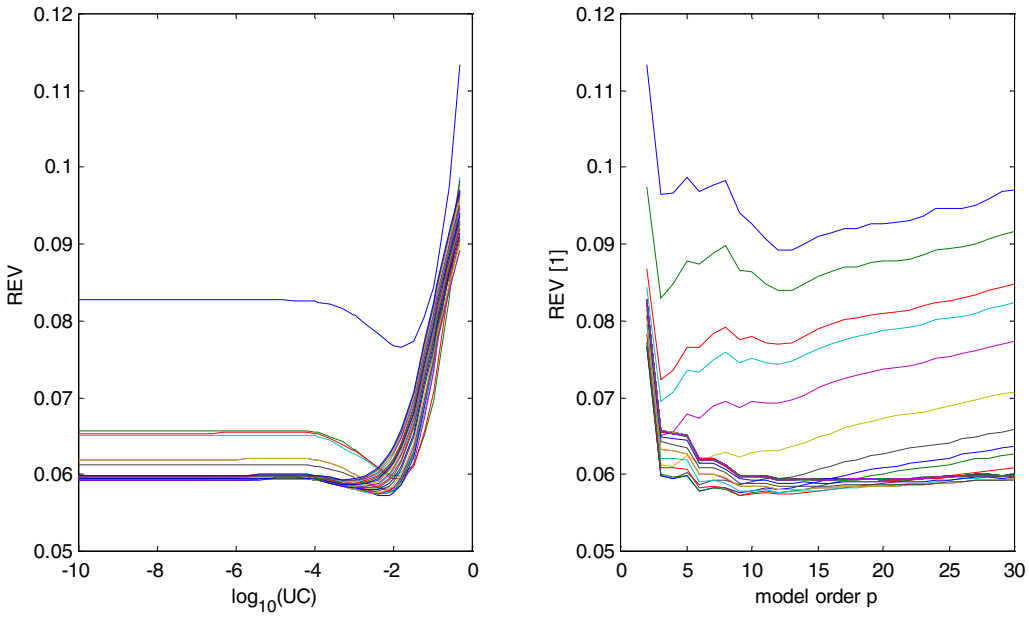


Figure 12: Relative error variance (REV) depending on the update coefficient (UC) and the model order p. The algorithm a5v1 was applied to EEG with time-varying spectrum (data is described in Schlögl et al. 1997a,b, Pfurtscheller et al. 1998) sampled with 128Hz and a length of 407.5s. The EEG was derived from electrode position C3 during repetitive imaginary left and right hand movement. The model order was varied from 2 to 30. The left part shows REV(UC) for different model orders p; the right figure shows REV(p) for different update coefficients UC. The minimum REV can be identified for p=9 and $UC = 2^{-8} = 0.0039$.

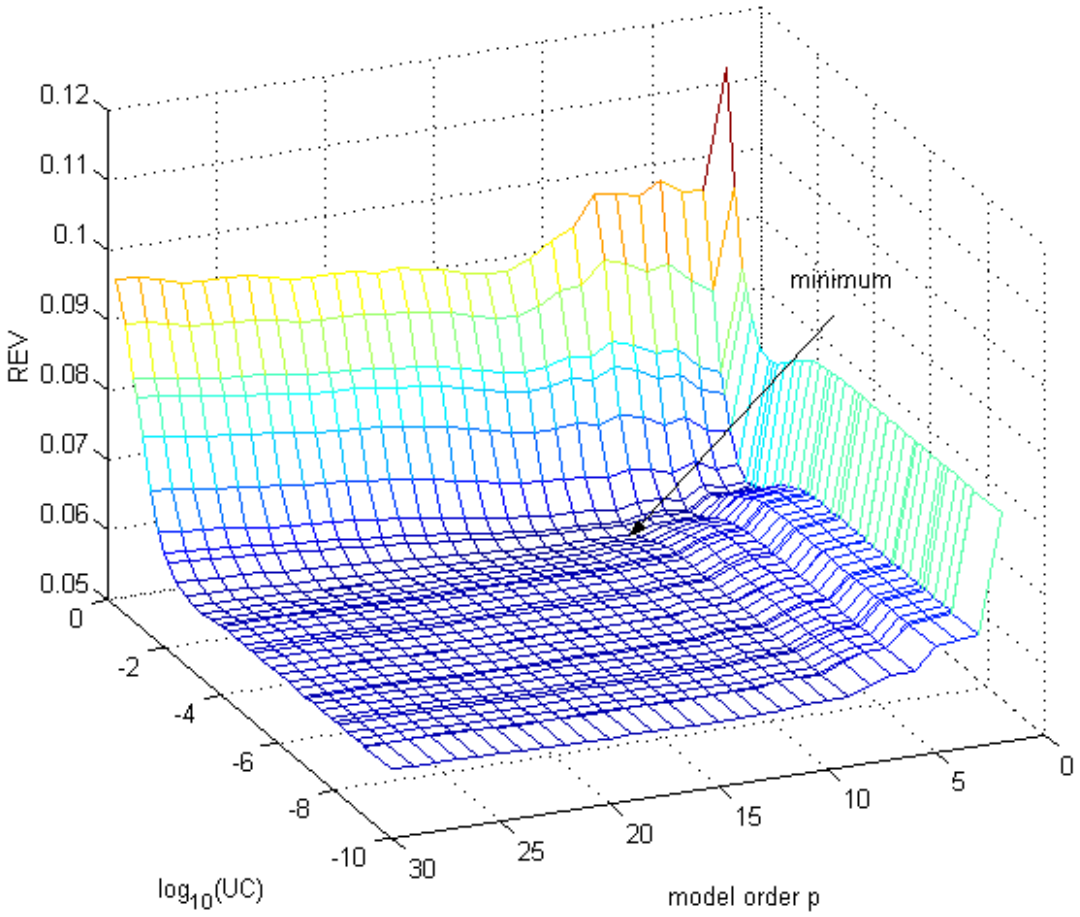


Figure 13: (Relative) error (REV) surface depending on model order p and update coefficient UC. The model order was varied from 2 to 30; UC was varied 2^{-k} with k=1..30 and 10^{-k} with k=1..10. The same results of Fig. 12 are displayed in three dimensions (adapted from Schlögl et al. 2000).

Fig. 12 and 13 display the relative error variance depending on the model order and the update coefficient. The *REV(p,UC)* is quite smooth, at p=9 and UC=2^{-8}, a minimum with a value of REV_{min} = 0.0572 can be identified. In summary, the optimal setting of estimation algorithm, model order and update coefficient can be identified with the REV-criterion and depends on the investigated data only. In general, slow changes require a smaller update coefficient. Also, with a smaller update coefficient, the time considered window is larger. Thus, the frequency resolution can be increased with the increase of the model order. Generally, one can expect that a larger model order requires a smaller update coefficient and vice versa. To a certain extend, this effect can be also identified in Fig. 12 and 13.

Of course, the task of identifying the optimal model order with Kalman filtering is quite laborious because the KF algorithm has to be applied for each combination of p and UC. A more efficient way is to use order-recursive filters (e.g. adaptive lattice filters, Sayed and Kailath, 1994, Haykin 1996). The advantage is that by calculating REV(p,UC), also all REV(i,UC), i<p are obtained. Thus, the overall computational effort is much lower .

PART III: APPLICATIONS

7. Brain Computer Interface

In the introductory chapter, the various issues in developing an EEG-based BCI were already mentioned. These are extracting the EEG features, obtaining a classifier, generation and presentation of feedback, and the subjects' motivation. The subjects' motivation must be supported by reliable feedback in order to support the subject in learning to control the EEG patterns. Therefore, EEG processing, classification and feedback generation will be discussed in the following. For this purpose, the results of different EEG parameters and classification methods are compared off-line. A new BCI paradigm is developed; furthermore, an attempt to estimate the amount of information in the feedback is made.

A common measure for the performance of STA is the classification error rate. For this purpose, the EEG for (at least) two classes is processed. In other words, signal processing methods are applied for extracting features (like bandpower, AR-parameter etc.) from the EEG. These features are applied to a classification system. Such a classification system can be a neural network, e.g. Learning Vector Quantization (LVQ, Kohonen, 1995, Flotzinger et al. 1992, 1994, Pregenzer et al. 1994, Pregenzer, 1997, 1998) or a statistical classifier like Liner Discriminant Analysis (LDA). Usually, a certain number of experiments (trials) is performed; the number of trials must be much larger than the number of extracted features. Each trial belongs to one out of two or more distinguished classes. The classification system is used to evaluate the separability of the features or to generate the on-line feedback.

Reliable values of error rates can be obtained for large data sets (5000 vs. 1000 samples, Michie et al. 1994). If this huge amount of data is not available, cross-validation (X-V, Stone 1978) must be applied in order to prevent over-fitting. If not stated otherwise, the following results were obtained by 10times 10fold X-V. A random permutation of the data set is divided into 10 subsets; 9 subsets are the training set, one subset is the test set. The classifier is calculated from the training set and applied to the test set. This procedure is repeated 10 times, with every subset being once a test set. The percentage of false classified trials is used as the error rate. The whole procedure is repeated for 10 different, random permutations of the data set. Thus, error rates from 100 different combinations of training and test sets are obtained. Note that training and test set were different in each permutation. The average of these error rates is ERR10, the 10fold10times X-V error rate.

7.1 Comparison of different EEG parameters

In this section, the average error rate from cross-validation is used to compare different feature extraction methods and two classification methods. EEG data from three subjects were investigated (data set D2). The following feature extraction methods were used:
Hjorth (1970), Barlow (Goncharova and Barlow, 1990), Hilbert transformation (Medl et al. 1992), AR(6) parameter estimated with the Burg algorithm (Kay, 1988) and with the Yule-Walker method (Kay, 1988), bi-variate AR(6) parameter, bandpower (Pfurtscheller and Arinabar, 1977) with predefined fixed frequency bands, and bandpower with optimized frequency bands using DSLVQ (Pregenzer et al. 1997, 1998). In some cases (Hjorth, Barlow), the data was pre-filtered with a bandpass (5-15 and 5-25Hz). All these parameters were calculated from the segment 4-5s.

Also, the LMS algorithm was applied to the same data in order to estimate AAR(6) parameters. For this purpose, the adaptive algorithms were initialized with second zero; the AAR estimates between 4-5s were averaged in order to obtain one feature vector for each trial. However, this means the AAR parameters, due to the adaptive estimation, take into account a longer data segment. This fact has to be considered when comparing the results.

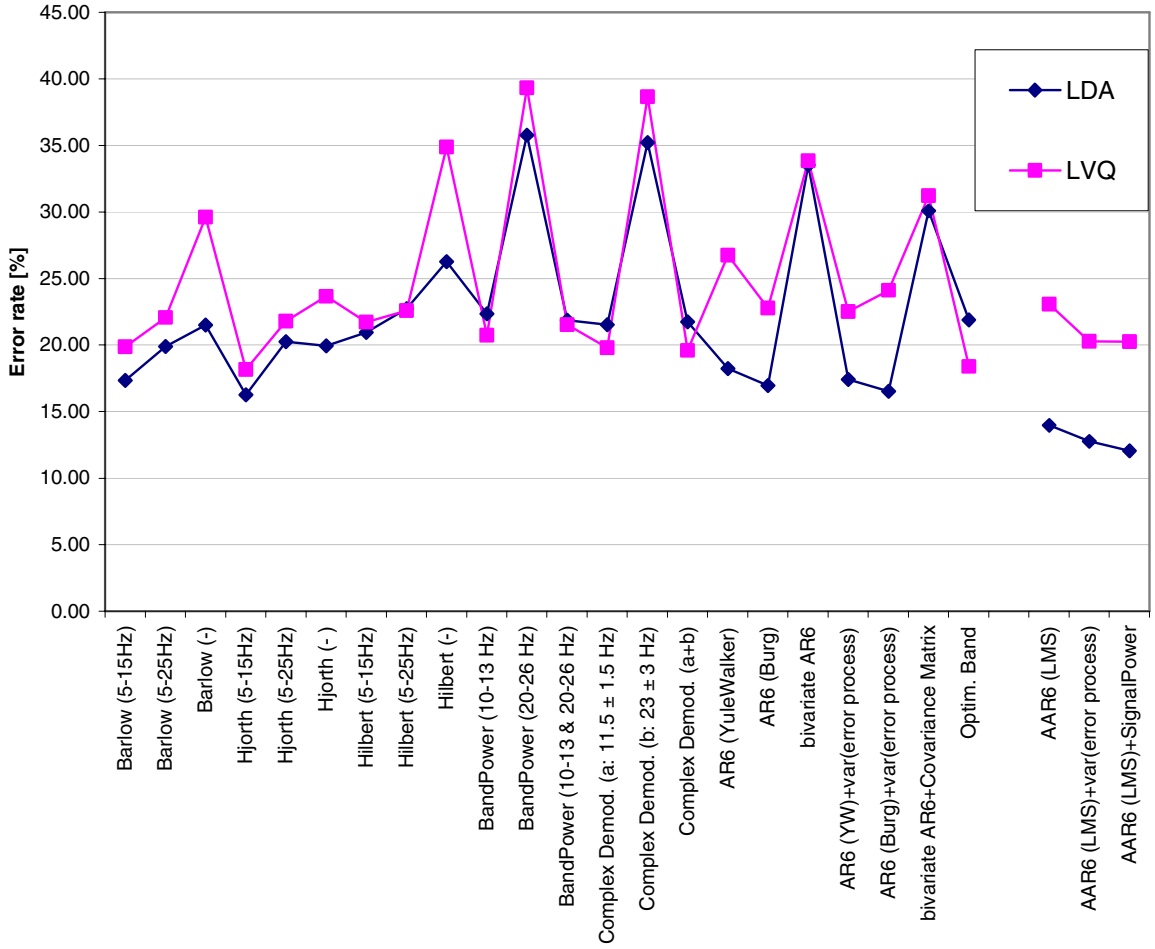


Figure 14: Comparison of different feature extraction and classification methods. The LVQ-algorithm (Flotzinger et al. 1994, Pregenzer, 1998) as well as LDA (see Appendix C) were applied for classification. Always, a 10-times-10-fold cross-validation was used within each subject. The error rates from all three subjects were averaged.

Fig. 14 shows that in cases of bandpower (10-13, Hz, Complex demodulation and optimized Band), the LVQ classifier is superior to LDA. However, the lowest error rates were obtained by the AAR + LDA method. Beside filtered Hjorth and Barlow parameters, also the AR-parameters, estimated with Yule-Walker and Burg, provide good results in combination with LDA. The consequence of these results was that LVQ should be used in combination with bandpower values, whereby AR parameters should be classified with LDA. Furthermore, the AR parameters in general (adaptive and the classical Burg and Yule-Walker approach) showed very good results.

7.2 Optimal classification time

Previously, it was already mentioned that in order to provide feedback, a classifier is needed. This classifier has to be calculated from earlier BCI recordings. From the previously performed ERD analysis, the period of time is known, which has high discriminative power. However, the AAR parameters were calculated for every time point; furthermore, the time window for estimating the AAR was clearly not rectangular and has, due to the update coefficient, a different length. Hence, the optimal time points for the classification is not known.

The solution for this problem was to apply the single trial classification for different time points; in practice, 8 time points per second were investigated. From each time point, an error rate was obtained; accordingly a time course of the error rate was yielded. Several subjects performed BCI-experiments; the timing is shown in Fig. 16(a). The EEG data of one session (160 trials) from four different subjects were used to calculate the time course of the error rate as shown in Fig. 15(a) (first column). The thick line shows the time course of the cross-validated error rate, calculated 8 times per second; the thin line displays the error rate for each sample time (128 per second) without cross-validation. The computational effort of LDA is not high, and the AAR parameters were calculated only once. Therefore, the computational effort was feasible.

From these time courses, the optimal classification time point can be easily identified. The classifier from this time point can be used for the on-line classification. A side effect of this investigation was that an idea of the time course of the EEG patterns was obtained. E.g. the time course of subject f76 has a peak up to 50% at about 6s. This peak is not found in f310 and f56. The relationship between the time course of the ERD and the time course of the error rate was discussed in Pfurtscheller et al. (1998). Clearly, subject-specific differences can be identified.

The time courses of the error rate $ERR10_t$ and ERR_t, the mean and standard deviation of the distance function for both classes, and the entropy obtained with 4 subjects are shown in Fig. 15. The first column in (a) shows the time courses of the error rate. In all 4 cases, the decrease of the error rate starts during presentation of the cue stimulus (3.0 - 4.25s). This means that the EEG patterns become better separable. For every subject, a time point with the best discrimination can be found. These time points were used for calculating the classifier $\boldsymbol{w}_{Tc}$. The classifier, obtained by applying LDA, is a weight vector w. It can be applied to the AAR parameter of both EEG channels in the following way.

$$D_t = [1, a_{1,t}^{\#1}, \ldots, a_{p,t}^{\#1}, a_{1,t}^{\#2}, \ldots, a_{p,t}^{\#2}] \cdot \boldsymbol{w}_{Tc} \qquad (7.1)$$

The offset (i.e. threshold) is incorporated in $\boldsymbol{w}$ and is considered by the element '1' in the data vector. Once the classifier is fixed, it can be applied to the AAR parameters at any time point t, and the distance D_t becomes also time-varying. Thus, D_t is called the time-varying signed distance (TSD) function because D_t varies in time; the sign of D_t describes whether the classification is left or right and the absolute value of D_t expresses the distance to the separating hyperplane described by $\boldsymbol{w}_{Tc}$.

The advantage of this procedure is that all AR parameter (also from several channels, thus considering different spatial locations) were reduced to one dimension. Now, D_t was a one-dimensional, time-varying function that can be calculated for every time point t on a single-trial basis. Equation (7.1) is the most informative linear projection - with respect to the class

relation - from multiple to one feature; the weight vector $\boldsymbol{w}_{Tc}$ incorporates the class information and expresses what the observer is interested in. In other words, the output is a (linear) combination of AAR parameters and produces a new one-dimensional feature.

This might seem to be trivial. However, the TSD can be calculated on each single trial; it can be averaged, also the standard deviation can be analyzed. The differences between classes are displayed in the second column of Fig. 15(b). It can be clearly seen, that after presenting the cue, the TSD is different for the two classes. Again, a different temporal behavior for every subject can be found. However, now we can see that in session f76, a crossing of the TSD can be observed. This crossing also explains why the error rate (Fig 15(a), 1st column) increases at 6s and decreases again afterwards.

A novelty is that also the standard deviation (S.D.) of the TSD is included in Fig. 15(b). The S.D. of the TSD shows the within-class variability of the output. It is caused, mainly, by the inter-trial variability of the EEG patterns. The ratio between the average TSD of the two classes on the one hand and the S.D. of the TSD on the other hand yields an impression of the ratio between useful information (i.e. signal) and noise (SNR). The SNR can be used to estimate the amount of information from the classification output.

7.3 Entropy of the BCI output

The distance $D_t^{(i)}$ is the distance function as defined in (7.1), whereby the superscript $^{(i)}$ denotes the number of the trial. In an ideal case $D^{(i)} > 0$ for all left trials and $D_t^{(i)} < 0$ for all right trials. However, it was shown that the distance $D^{(i)}$ does not only depend on the class relationship, but also on many other factors like background EEG activity, noise from the amplifier and the analog-digital converter, numerical errors etc. These factors are the noise term and can be summarized in the part that is uncorrelated to the class information. In the TSD (Fig. 15b), this term is visualized as the within-class standard deviation of the TSD. The noise variance is

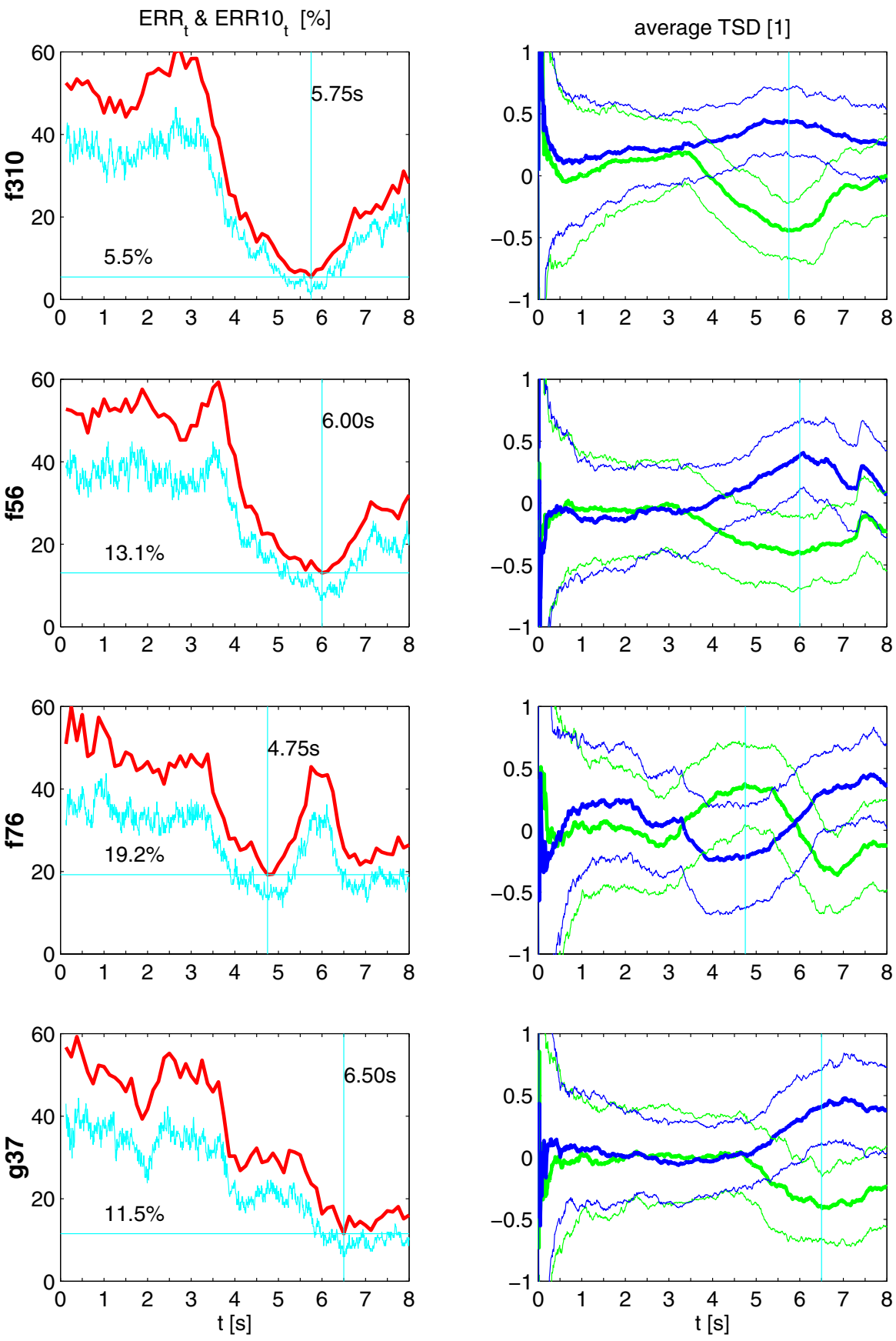
$$N_t = \tfrac{1}{2}\left(\operatorname*{var}_{i\in\{L\}}\left\{D_t^{(i)}\right\} + \operatorname*{var}_{i\in\{R\}}\left\{D_t^{(i)}\right\} \right) \tag{7.2}$$

where ***{L}*** and ***{R}*** are the sets of left and right trials, var{.} is the variance over all trials *i*. The between-class standard deviation is shown as the difference between the average TSD for both classes. This part is correlated to the class and can be seen as the term which carries the useful information, hence, it is the signal. Note that the signal and the noise term are not correlated, hence the total variance is the sum of both, the signal and the noise variance.

$$S_t + N_t = \operatorname*{var}_{i\in\{L,R\}}\left\{D_t^{(i)}\right\} \tag{7.3}$$

Assuming that the signal and noise is uncorrelated (which it is per definitionem), it can be shown that

$$S_t = \operatorname*{var}_{i\in\{L,R\}}\left\{D_t^{(i)}\right\} - N_t = \left(\mu_1 - \mu_2\right)^2 \tag{7.4}$$


ERR$_t$ & ERR10$_t$ [%]
average TSD [1]
f310
5.75s
5.5%
f56
6.00s
13.1%
f76
4.75s
19.2%
g37
6.50s
11.5%
t [s]
t [s]

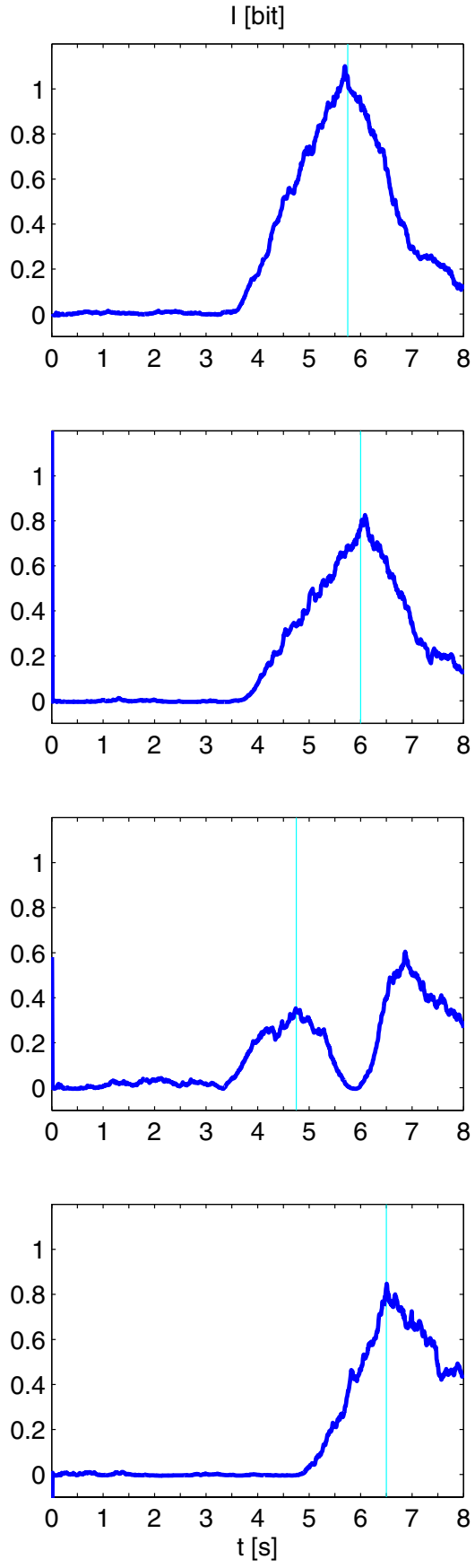


Figure 15: Time courses displaying the separability between two classes obtained in four subjects. The vertical lines indicate the time point used to generate the weight vector. The time point t with the lowest error rate ERR10 was used. The timing scheme is shown in Figure 16(a) (adapted from Schlögl et al. 1999g).

(a) In the first column (previous page), the time courses of the error rates $ERR10_t$ (thick line) and ERR_t (thin line) are shown. ERR_k gives the classification error with LDA of the EEG-channels C3 and C4 at time t. AAR(10) parameters were used as EEG features. The thick line shows $ERR10_t$ calculated 8 times per second; the thin line shows the time course of ERR_t calculated at every sample. The numbers indicate the lowest ERR10 and the corresponding classification time.

(b) The second column (see previous page) shows the averaged TSD for the left and right trials. The TSD is calculated as linear combination of AAR(10)-parameters of the EEG channel C3 and C4. The average TSD curves (thick lines) clearly show a different behavior during imagined left and right hand movement. The thin lines represent the within-class standard deviation (SD) of the TSD and indicate the inter-trial variability of the EEG patterns.

(c) The third column shows the mutual information between the TSD and the class relationship. The entropy difference of the TSD with and without class information was calculated for every time step. This gives (a time course of) the mutual information in bits/trial.

One can say $D_t^{(i)}$ consists of two processes. One is correlated to the movement imagery task containing useful information (signal) while the other process is uncorrelated to the task

(noise). The Signal-to-Noise ratio (SNR) (for any fixed time point t within a trial) is defined as follows:

$$SNR_t = \frac{S_t}{N_t} = \frac{(\mu_1 - \mu_2)^2}{\frac{1}{2}\left(\underset{i\in\{L\}}{\text{var}}\{D_t^{(i)}\} + \underset{i\in\{R\}}{\text{var}}\{D_t^{(i)}\} \right)} \tag{7.5}$$

and

$$SNR_t + 1 = \frac{S_t + N_t}{N_t} = \frac{\underset{i\in\{L,R\}}{\text{var}}\{D_t^{(i)}\}}{\frac{1}{2}\left(\underset{i\in\{L\}}{\text{var}}\{D_t^{(i)}\} + \underset{i\in\{R\}}{\text{var}}\{D_t^{(i)}\} \right)} \tag{7.6}$$

The mutual information between the BCI output D and the class relationship can be determined by calculating the entropy difference between noise and signal (see also chapter 2.1, Shannon, 1948). This leads to the following formula for the mutual information for each possible classification time t:

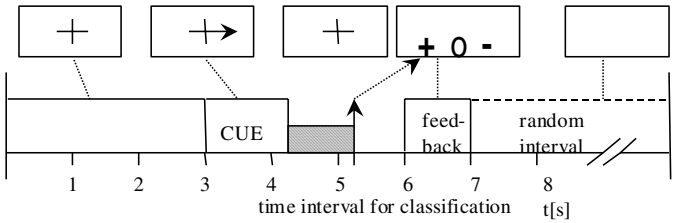
$$I_t = 0.5 * log_2 (1 + SNR_t) \tag{7.7}$$

Figure 15(c) displays the time courses of the mutual information between the distance TSD and the target cue (left or right). The time course of the mutual information shows at which point of time the patterns are most separable. As expected, no information about the class relationship can be obtained prior to cue presentation; the entropy starts to increase at 3.5s. In general, the mutual information displays a maximum when the error rate is minimal. However, there are also differences between the time courses of the error rate and the mutual information curves. In the data set f310, the mutual information is larger than 1 bit per trial, although the error rate is not zero. The explanation is, that the entropy analysis is based on the variance, however, there are some outliers which make a complete separability of the two classes impossible. Data set f76 shows a larger mutual information at about t=7s; this is surprising because the error rate is smaller at t=5s. The unexpected result in g37 is that (the time course of) the mutual information does not increase prior to t=5s although the error rate already decreases at 4s. One explanation for these differences might be that the EEG patterns are different at 4-5s and 6-8s and the former patterns are not "represented" by the specific weight vector.
This is a first attempt to estimate the amount of information of a BCI output, in order to approach the required information rate of 20 words per minutes (Robinson, 1999)

7.4 The continuous-feedback-paradigm

A typical task for a subject was to imagine a left or a right hand movement according to a target cue. The timing is shown in Fig. 16. The target cue was an arrow on a computer screen pointing either to the left or the right side (3.0-4.25s). Depending on the direction, the subject was instructed to imagine a movement of the left or right hand. From 6.0 – 7.0s feedback was given. The feedback was provided in five classes, '++', '+','0','-','--'. The '+' and '-' sign indicated a correct and false classification, respectively. The double signs were presented as larger sign on the screen (Pfurtscheller, et al. 1996). The feedback was calculated from the band power of the most reactive frequency bands. These frequency components were found by using the Distinction Sensitive-LVQ algorithm (Pregenzer et al. 1994) from previous recordings of the same subject.

a) Paradigm BCI4c



b) Paradigm BCI4e

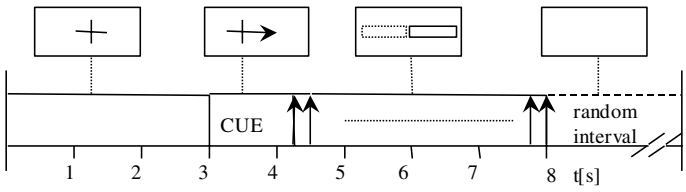


Figure 16: Timing of one trial in a BCI experiment. (a) BCI4c paradigm, (b) BCI4e paradigm (adapted from Schlögl et al. 1997b,c, Pfurtscheller et al. 1998, Neuper et al. 1999)

However, the AAR+LDA method offered even more possibilities. Once, a classifier is obtained, it can be applied continuously in time. Since the AAR parameters are updated with the same sampling rate, also the output can be calculated with the same rate. The display rate could be increased from one per trial to a feedback, which is continuous in time. Applying the LDA classifier is a simple weighted sum and requires very little computational effort; at least as compared to the AAR estimation algorithms. Furthermore, the form of the feedback was changed; instead of qualitative feedback (correct or wrong), the result of the distance function (9.1) was represented by the length of a bar. Thus, a quantitative feedback was provided.

This new type of feedback generation was developed according to the timing scheme in Fig 16b and was called BCI4e (Schlögl et al. 1997c). This should enable to learn better how to control the EEG patterns. It was used to enhance EEG differences (Neuper et al. 1999) and a new online BCI system (Guger et al. 1999). Also, it is an approach towards a more continuous analysis of the EEG patterns, beyond a fixed timing scheme as in Fig. 16a. In the meantime, several hundred BCI sessions (e.g. Neuper et al. 1999, Guger et al. 1999, Pfurtscheller and Guger 1999) were recorded, using AAR+LDA for the feedback generation.

The system was also applied successfully to a tetraplegic patient (T.S., 22 years, male). The subject had an accident in April 1998, which caused a lesion (at C4/C5) and results in a severe paralysis of lower and upper extremities, except for his left biceps muscle. The goal is to restore the hand grasp function of his left hand with an electrically powered hand orthesis. The control of the orthesis is performed with the BCI system (Pfurtscheller and Guger, 1999).

8. Analysis of sleep EEG

The second application of the AAR method is related to a research project on sleep research. The aim of the project is to develop an automated sleep analysis system. Eight different European sleep laboratories provided the data. For this reason, also a quality control of the data was performed which is described in the first section. Another important topic in automated EEG analysis is artifact processing. In this section, it will be shown how the AAR method can be used for adaptive inverse autoregressive filtering. The method is validated and the influences of the transient patterns to the AAR estimates are discussed. Finally, the AAR parameters were used to perform an all-night sleep analysis.

8.1 Quality control of EEG recordings

Typically, the resolution of analog-to-digital conversion is 12 or 16 bit in EEG analysis. The data is stored in a 16bit-integer format (e.g. EDF, Kemp et al. 1992, and BKR-format[1]). This means that $2^{16} = 65536$ different amplitude values are possible. Fig. 17 shows the histograms of the EEG channel C3-A2 from several all-night sleep recordings. The 'sidelobes' are due to saturation effects of the input amplifier and/or ADC. The results are described in detail in Schlögl et al. (1999).

Various parameter can be obtained from the histogram: the total number of samples N (8.1), the entropy of information (8.2), the mean value μ (8.3), the variance σ^2 (8.4), the skewness (8.6) and the kurtosis (8.7) of the data can be calculated. Note that once the histogram is available, the computational effort is quite low, even for an all-night EEG recording.

$$N = \Sigma_i H(i) \qquad (8.1)$$

$$I = \Sigma_i (H(i)/N * log_2(H(i)/N) \qquad (8.2)$$

The mean μ and variance σ^2 of signal Y can be obtained from the histogram

$$\mu_Y = E\{Y_k\} = \Sigma_i (i\, H_Y(i))/N \qquad (8.3)$$

$$\sigma^2_Y = E\{(Y_k - \mu_Y)^2\} = \Sigma_i((i - \mu_Y)^2 * H_Y(i))/N_Y \qquad (8.4)$$

The Gaussian distribution with the mean μ, variance σ^2 and the number of samples N is

$$p_Y(x) = N/\sqrt{(2\pi\sigma_Y^2)} * exp(-(x-\mu_Y)^2/(2\sigma_Y^2)) \qquad (8.5)$$

Furthermore, the skewness γ_3 and kurtosis γ_4 of the data series Y are defined (Nikias and Petropulu, 1993) as follows:

$$\gamma_{3,Y} = \Sigma_i ((i - \mu_Y)^3 * H_Y(i)) / N_Y \qquad (8.6)$$

$$\gamma_{4,Y} = \Sigma_i((i - \mu_Y)^4 * H_Y(i))/N_Y - 3*(\sigma_Y^2)^2. \qquad (8.7)$$

[1] BKR-Format V2.07, Department for Medical Informatics, Institute for Biomedical Engineering, University of Technology Graz

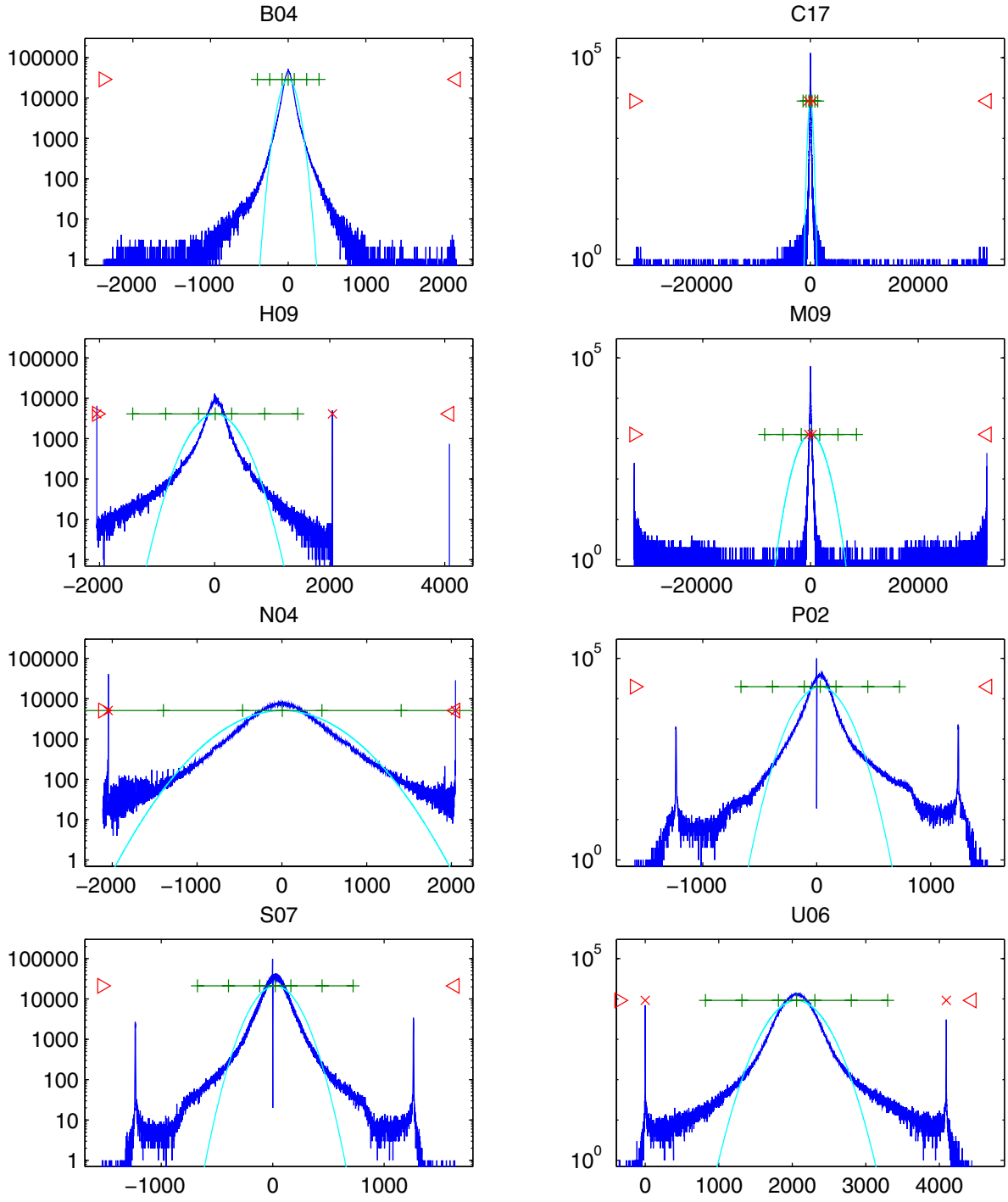


Figure 17: Histograms of EEG channel C4-A1 from 8 all-night recordings. The units at the horizontal axes are digits and can range from -32 768 to 32 767. The horizontal line with the vertical ticks displays the mean ± 1, 3 and 5 times of the standard deviation. The markers |> and <| indicate the real maxima and minima found in the recordings. The markers x indicate the digital minima and maxima as stored in the header information; if invisible, the values are outside the scope. The light parabolic lines display the Gaussian distribution with the same mean and variance as the data (adapted from Schlögl et al. 1999b).

The Gaussian distribution (8.5) can be related to the histogram (see Fig. 17). The Gaussian distribution appears as a parabolic line due to the logarithmic scaling. It can be clearly seen that the histogram deviates from a Gaussian distribution. Due to this fact, the higher order moments, like skewness and kurtosis are non-zero, which would mean the EEG is non-linear.

Priestley (1988) showed by means of state-dependent models (SDM), that a non-linear system can be represented by a non-stationary system and vice versa. In theory, the non-linear form is equivalent to the non-stationary form. It is known that the sleep EEG varies during the different sleep stages. The spectral density function and the amplitude (variance) changes. Solely, a varying variance of the EEG can explain the deviation from normal distribution of the sleep EEG. In other words, the deviation from Gaussian distribution can be explained by the non-stationarity in the sleep EEG. This is an example for how a non-linearity might be caused by a non-stationarity. It can be concluded that the observed non-linearity can be described also by a time-varying (i.e. non-stationary) model.

8.2 Artifact processing with the AAR parameters

Artifacts are one major problem in EEG recordings. EEG artifacts are all parts of the signal which are caused by non-cortical activity. They are caused by other biomedical sources, like eye movement superimpose the electrooculogram (EOG), the electrical activity of the heart (electrocardiogram ECG), or muscle activity in the face or due to chewing. Technical artifacts are failing electrode, power line interference (50Hz) or simply a saturation of the amplifier. Anderer et al. (1999) gives a review of artifacts and artifact processing in the sleep EEG. In the sleep research project SIESTA, nine types of artifacts were distinguished. Experts scored a database (Data set D3) visually with these nine artifact types. The results are shown in Table 3 (Schlögl et al. 1999d).

Table 3: Artifact in the sleep EEG. In total, 563 192 1s-epochs were scored for nine different artifact types. In some epochs several artifact types were found simultaneously, therefore the sum is 108.5%

Type	Epochs [1s]	%
no artifact	345 254	**61.3**
EOG	51 427	**9.1**
ECG	123 886	**22.0**
muscle	43 861	**7.8**
movement	20 209	**3.6**
Failing elect.	10 804	**1.9**
sweat	4 135	**0.7**
50/60 Hz	11 608	**2.1**
breathing	2	**0.0**
pulse	-	**0.0**
SUM	611 186	**108.5**

Different artifact processing methods were validated (Schlögl et al. 1999e). The most frequent artifact is ECG; it can be minimized by regression analysis which is quite simple, by adaptive filtering (Sahul et al. 1995), or by a method on removing a template of the ECG artifacts (Harke et al. 1999). The eye movement artifact can be removed e.g. by principle component analysis (Ille et al. 1997, Lagerlund 1997 et al.); a 50Hz notch filter can remove the line interference. Failing electrode artifacts cause a flat input at the upper or lower saturation value of the amplifier. This artifact type can be detected with an over/underflow check of the input signal whereby the saturation thresholds must be known (Schlögl et al. 1999b). Movement

artifacts are a mixture of failing electrodes and muscle artifacts. In the following, it is shown how the AAR estimation algorithms can be used for the detection of these muscle and movement artifacts by means of adaptive inverse filtering.

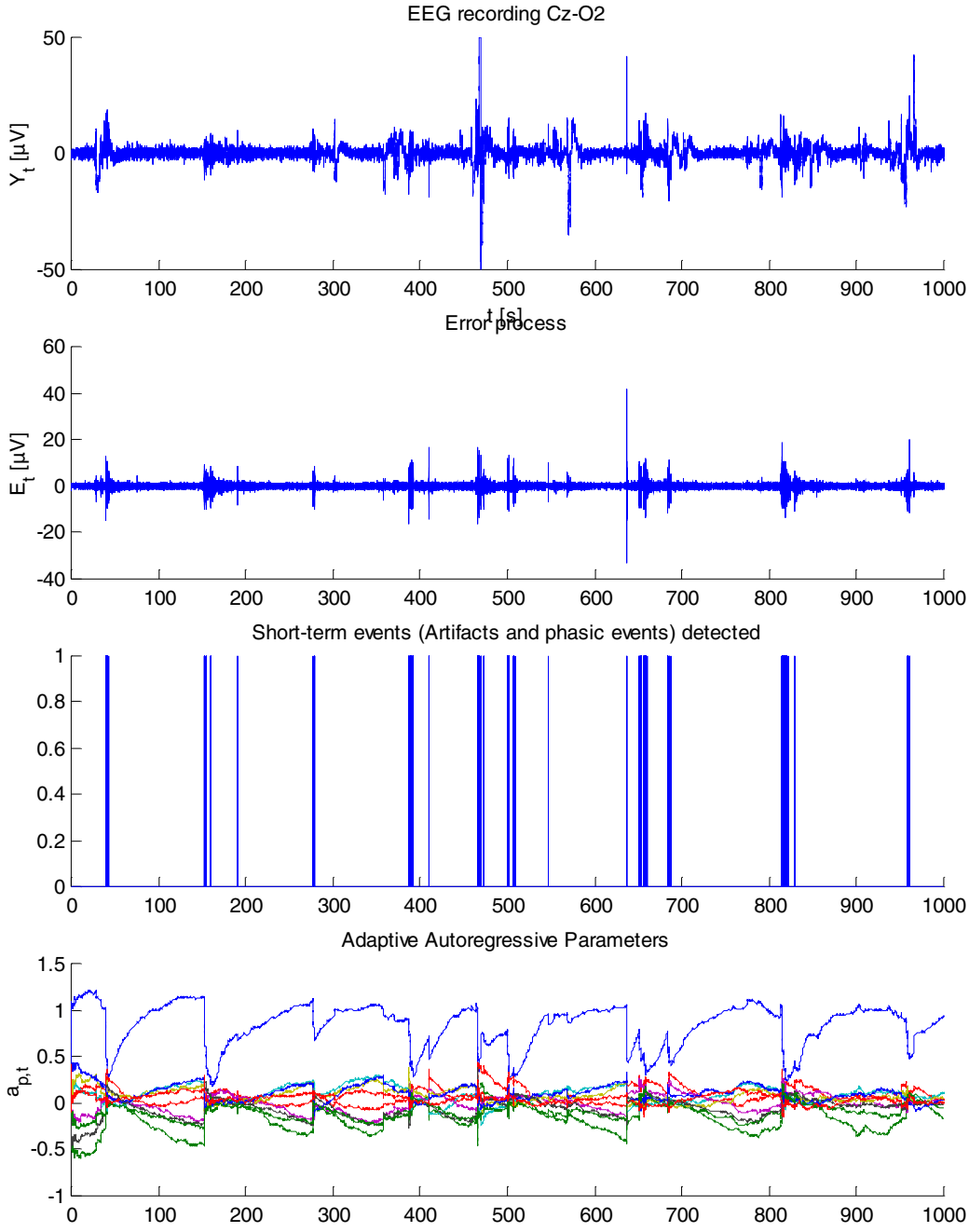


Figure 18: AAR parameter of a segment of sleep EEG. a) shows the raw EEG, (b) is the prediction error process, (c) the detection threshold $e_t^2 > 3*var\{y_t\}$ and (d) shows the AAR estimates (adapted from Schlögl et al. 1998c).

In section 3.4 the principle of adaptive inverse filtering as a modification of inverse filtering was explained as described by Praetorius et al. (1977), Bodenstein and Praetorius (1997), and Lopes da Silva et al. (1977). The idea is that (the variance of) the prediction error process can be used to detect transient events in the EEG. In this case, muscle artifacts are such transient events. Fig. 18(a) shows an example of a sleep EEG contaminated with muscle activity (fast spikes) and eye movements (slow, smooth spikes). The data was sampled with a rate of 100Hz and is 1000s (16min 40s) long. The Kalman filtering algorithm (a12v2) was applied with p=10 and UC = 0.001. The one-step prediction error process is shown in Fig. 18(b). It can be seen that the transients are more pronounced. The detection function in Fig. 18(c) is obtained by applying a threshold of $3*var\{y_t\}$ to the squared prediction error process e^2_k. Note that the threshold determines the specificity and sensitivity of detector.

Earlier, when discussing the premises of an AAR model, it was already stated that transient events contradict the assumption of a slowly varying AAR process. In other words, the characteristics of transients can not be characterized by the AAR model parameter. Fig.18(d) displays how transients influence the corresponding AAR estimates; all AAR parameter change towards zero. Note, that an AR-spectrum (2.7) with zero coefficients gives a flat spectrum; also the spectrum of a single spike (Dirac impulse) is flat. It can be concluded that in the case of transients, the AAR estimates still describe the spectral density distribution of the signal. The increase of the absolute power is considered in the variance of the error process. It is interesting that even in such cases, the spectral estimation moves towards the theoretically expected spectral density function.

After the transient has vanished, the estimation algorithms require about the same time to adapt in order to obtain estimates that accurately describe the signal characteristics. During this time, the poles (roots of the AR polynomial in the complex Z-plane) move from the origin towards the unit circle. Moreover, it means that the spectral peaks, which have diminished during the transient, become now more and more pronounced again.

Figure 18 shows that the mean squared error can be used to detect muscle artifacts. In the next step, the detector quality is evaluated using the artifact database (Table 3). For this purpose, Kalman filtering (version a5v1, UC=0.001, p=10, $\boldsymbol{a}_0$ =0, $\boldsymbol{A}_0=\boldsymbol{I}$) was applied (Data set 5) and the prediction error was calculated. The mean squared prediction error process MSE of each 1s-epoch was analyzed by the receiver-operator-characteristics (ROC) curve in order to investigate the quality of the detector (see also Schlögl et al. 1999f).

The ROC-curve was calculated for each artifact type with sufficient examples (pulse and breathing artifacts were not considered, see Table 3). A ROC-curve displays the sensitivity vs. 100%-specificity for different detection thresholds. If the ROC curve is diagonal (sensitivity = 100% - specificity), the area under the ROC curve (AUC) is 0.5, this means no correlation between expert scoring and output can be observed. The ROC curve can be used to identify the optimal threshold if the costs (for false positive and false negative decisions) are known. Furthermore, the area-under the ROC curve (AUC) is a measure for the detector quality (Judy et al. 1992), whereby no threshold has to be selected in advance.

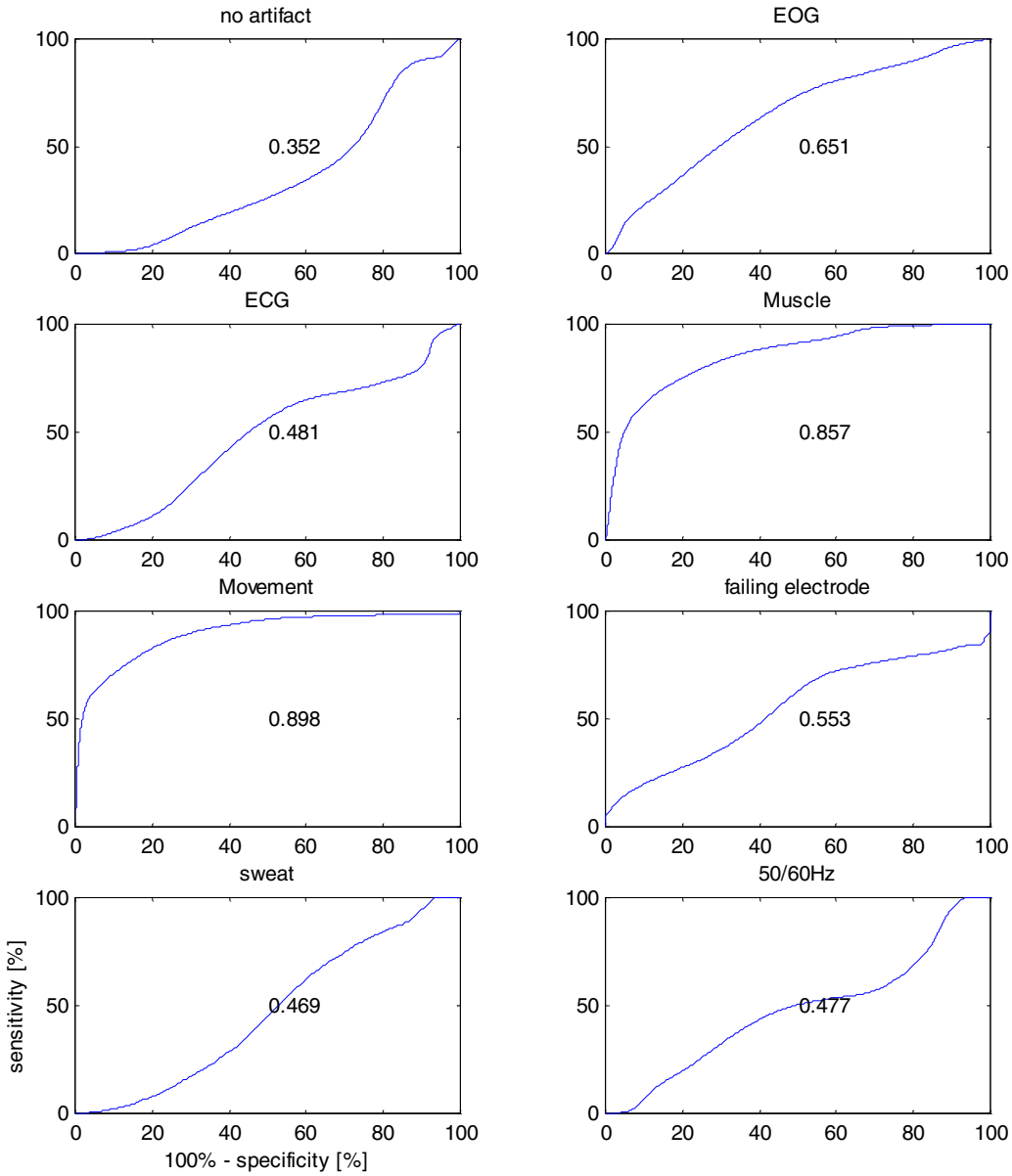


Figure 19: ROC curves of the *MSE* versus the expert scoring for: 'no artefact' (top left), 'EOG' (t.r.), 'ECG', 'muscle', 'movement', 'failing electrode', 'sweat', '50/60Hz' (b.r.).

In Fig. 19, the ROC curves for each artifact type are shown. The AUC is displayed in the center of each curve. It can be seen that the AUC is 0.857 and 0.898 for muscle and movement artifacts, respectively. An AUC of 0.870 was obtained, when both artifact types were combined to one type. Other artifact types were rarely detected by the *MSE*; probably, those artifact types are closer to slow variations than to transients.

However, an AUC of 0.87 also means that a perfect separation of artifacts from artifact-free epochs, at least according to the definition of the expert, is not possible. The basic principle of adaptive inverse autoregressive filtering is that transient events can be clearly distinguished from nearly stationary changes. This might not always be the case. E.g. the (muscle) artifact

can last several seconds up to minutes. E.g. some subjects grit their teeth during sleep, causing muscle artifacts in the EEG. This artifact type is clearly not a short event, but can last much longer. It was also shown that very short transients require a longer adaptation period afterwards. On the other hand, EEG patterns might be transient, short-term phenomena, e.g. sleep spindles, or K-complexes which have a typical length of 1s. All these phenomena limit the accuracy of the *MSE* as a detector of transient events.

Despite these facts, some ability to detect muscle artifacts could be observed. Summarizing, it can be said the proposed method of adaptive inverse filtering is not a very specific method for detecting transients. But it a useful byproduct of the AAR estimation algorithm. It is obtained without additional computing effort and can easily be applied on-line.

8.3 Sleep analysis

It was shown (equation 3.6) that the AAR method can be used to describe time-varying spectra. It is also known that the spectrum of the sleep EEG is an important measure for determining the sleep stages. In the following is investigated how AAR parameters can be used to determine sleep stages. For this purpose, the AAR parameters from 38 all-night sleep recordings (from data set 1) were estimated. The model order p=10, update coefficient UC=0.001, the Kalman filtering algorithm a5v1, a 50Hz notch filter, a low pass filter with an edge frequency of about 35Hz and regression analysis for minimizing the ECG artifact were used. The re-sampling algorithm as described in the Appendix B was used to obtain AAR parameter for 100Hz, even if the sampling rate was 200 or 256Hz. The AAR estimates at the end of each second were stored for further processing. No smoothing or averaging of the estimates was performed.
These AAR estimates were used to determine a classifier based on LDA, similar as described in the BCI approach. For each of the three classes (Wake, REM and deep sleep), a weight vector was obtained. These weights can be used to calculate the time-varying distance function for each classifier. The results of one recording are shown in Fig. 20.

In the second plot of Fig.20, the *MSE* of the one step prediction error for each 1s-epoch is shown. The increase of the *MSE* appears in bursts and is related to sleep stages. In case of deep sleep and REM without interruptions the *MSE* is relatively small. In these cases, the EEG contains no transient events, thus it can be assumed that the EEG is nearly stationary and the spectrum changes slowly. In section 8.2 was shown that an increased *MSE* is an indicator for muscle and movement artifacts or other transient phenomena. Hence, it can be assumed that the increase of the *MSE* represents arousal reactions, which often cause muscle and movement artifacts in the EEG. The fact that several *MSE* spikes coincide with short awakenings supports this assumption.

Below the *MSE* curves (in Fig. 20) is the output of the three classifiers (Wake, REM and deep sleep) displayed . The scaling is arbitrary; but it can be seen that the larger the values are, the higher is the probability of the corresponding process. The Wake and REM processes increase during Wake and REM. However, it can be seen, with respect to this scaling, that in case of Wake scoring, the WAKE process is larger than REM and vice versa. The weights for the deep sleep process were obtained from the classes 2 versus 3+4. Despite the fact that the sleep curve is superimposed by the awakenings and REM into the negative direction, it can be seen that a value of approx. zero corresponds stage 2; the highest values correspond to the sleep stages 3 and 4.

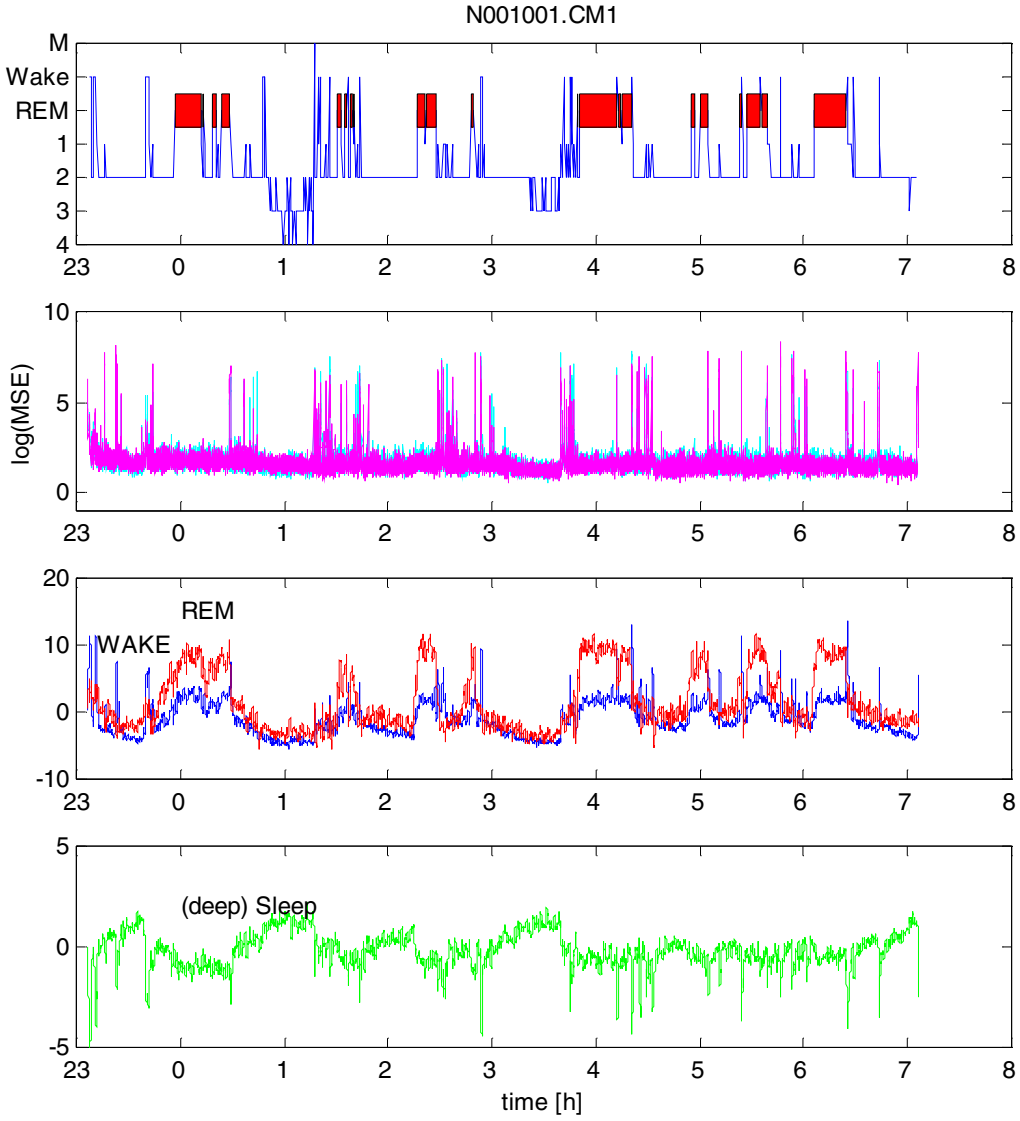


Figure 20: Sleep analysis with AAR method. The first plot displays the hypnogram scored by an expert. The second part contains the *MSE* (variance of the inverse filtered process) of the channels C3-M2 and C4-M1. The third and forth plot contain time courses for Wake, REM and Sleep. These were obtained by a linear combination of 10 AAR parameter of two EEG channels (C3 and C4). The curves were smoothened with a rectangle window of length 30.

This result is only a first approach, which has to be imporved in the future. For example, in case of a transient, not only the *MSE* is increased, but also the AAR estimates tend to zero and need some time afterwards for adaptation. One idea would be to implement an artifact detector (e.g. with a threshold for the *MSE*). In case of a detected artifact, the adaptation of the AAR parameter should stop until the *MSE* of the prediction error process is below this threshold. This would limit the variation of the AAR estimates and it can be expected that the classification output shows a better resolution. Furthermore, also non-linear classifier like neural networks might be useful in order to improve the classification output.

It is important to note that this sleep staging is based on EEG only. One can clearly identify variations according to the R&K sleep stages. Although this is only a first result, it seems to be a possible approach towards an automated sleep analysis based on spectral variations of the EEG.

PART IV: CONCLUSIONS

9. Comments on AAR modeling

9.1 Historical notes

Bohlin (1972), Mathieu (1976), Dusquesnoy (1976), Blechschmid (1982) and Jansen et al. (1979, 1981) already applied Kalman filtering to an AR model for analysing EEG. Despite these works and the theoretical advantages (optimal filter, non-stationary spectral analysis), Kalman filtering was not very much used for EEG analysis. It can be assumed that the unsolved problem of selecting the model order and the unstable estimation algorithms were reasons for this.

Furthermore, it can be speculated that the instabilities were caused by neglecting the covariance matrix $\boldsymbol{W}_k$. Often, the covariance $\boldsymbol{W}$ was set to zero or not considered at all. In these cases, the algorithm adapts from some initial value towards the 'true' AR parameter of a stationary time series; the longer the observation was, the smaller was the estimation variance. This leads to the stalling phenomenon and numerically unstable estimation algorithms. The solutions for these problems was to increase the covariance at some (rare) time points (e.g. Blechschmid, 1977, Roberts, 1997 and Penny and Roberts, 1998 who refers to Jazwkinski, 1969) or to use the Kalman filter as (stationary) AR estimation algorithms of short segments (Jansen, 1979, 1981, Blechschmid, 1982).

These approaches were very different from a random walk approach, which means the use of a non-zero covariance matrix $\boldsymbol{W}_k$ (4.24-25) at every iteration. It was shown in this work that a random walk approach gives a lower one-step prediction error than the other assumptions. Furthermore, if the update coefficient and the model order were chosen properly, no stability problem was observed. It can be concluded that the random walk model is the best model but it was not used in the past.

9.2 Model order, update coefficient and the time-frequency resolution

Since AR models are used, the order of the model is of immanent importance. Various information criteria (AIC, BIC, FPE, CAT, etc.) (Priestley, 1981) can be used in the stationary case. These criteria are not applicable in the time-varying case. Hence, it was suggested to use the *MSE* (REV) criterion also for model order selection. The advantage of the REV criterion for the model order selection is that no penalty term is needed.

The problem of selecting the model order and the update coefficient can be seen as a reformulation of the principle of uncertainty between time and frequency domain (Priestley, 1981) in terms of stochastic time series analysis with an autoregressive model. The model order corresponds to the frequency resolution and the update coefficient determines the time resolution. Hence, for a certain UC, one optimal model order exists and vice versa.

The question of Haykin (1999) '*…whether there will be an ultimate time-frequency distribution …or a number of densities tailored for individual applications…* ' can be reformulated whether to look for the global minimum of *REV(p,UC)* or for one of several local minima which yield locally optimal solutions for REV*(p,UC)*.

9.3 Alternative and related methods

In this work, only adaptive algorithms with forward prediction were investigated. Other, related algorithms are time-varying autoregressive (TVAR) modeling (Kaipio and Karjalainen, 1997a,b) as well as Kalman smoothing algorithms (Grewal and Andrews, 1993). They might be interesting alternatives for off-line analysis or when a delayed estimation of the AAR parameters is acceptable. Both approaches were not investigated because the REV-criterion is based on an un-correlated prediction error process (i.e. residual process) which would be not the case in these methods.

In this work, the spatial relationship was considered in the classification. The spatial relationship can be addressed by bi-variate (Schack et al. 1995, Andrew 1997, Arnold et al. 1998) and multivariate time-varying AR models (Gersch, 1987). In these cases, it should be considered that the number of coefficients increases with the square of the number of components. A larger number of parameters also increases the estimation variance of the model parameter. Here, the test of the residual process might be a useful tool in order to prevent overparametrization.

In this work were also discussed two more or less physiological models of the EEG. The lumped circuit model explains only one rhythm (alpha), the Kemp's feedback loop model is able to describe the low frequency components (delta) and one high frequency component (alpha). These models do not explain an EEG spectra with more components e.g. delta alpha and beta as described by Isaksson and Wennberg (1975). An AR model is a linear model, which is able to describe the spectral composition of the EEG signal. An AAR model also considers the time-variation of the spectrum, assuming that the characteristic change only gradually. However, non-linear EEG analysis (like the lumped circuit model with a feedback gain larger than the critical value) can not be represented by an AR model. But recently, an extended Kalman filtering method was applied (Valdes et al. 1999) to the lumped model (Lopes da Silva et al. 1974, Zetterberg et al. 1978). The state-space model was constructed directly from the neural mass model. The advantage of this method is that the model parameters in equation (2.39-40) may not need to be assumed as in Suffczinki et al. (1999). They are rather part of the state vector and can, therefore, be estimated with the Kalman filter. The work of Valdes et al. (1999) shows one possible direction how non-linear EEG analysis, based on physiological models, can be performed with Kalman filtering.

All models that address the problem of time-varying spectral analysis (including higher order spectra) without averaging can be tested by means of the REV criterion. It is only important that the prediction error $e(t)$ at time t is uncorrelated to all previous sample values $y(t-i)$, $i>0$. This can be ensured if $y(t)$ is used firstly for calculating the prediction error $e(t)$ at each time point t; afterwards, $y(t)$ can be used for all further processing steps.

9.4 Is an AAR model a useful model for EEG?

The ERD/ERS and the AAR parameters have in common that both describe time-varying EEG spectra. The ERD/ERS describes changes of power in distinct frequency bands; all AAR parameters together describe the time-varying spectral density functions. A single AAR-parameter does not mean very much and all AAR parameters are more difficult to deal with. This is clearly a disadvantage of the AAR parameters as compared to the ERD method. However, in combination with the LDA, the results can also be visualized. Furthermore, AR modeling does not require the selection of specific frequency bands. Individual variations of the mean frequency, which might be important for the significance of a study (Klimesch et al.

1998, Klimesch, 1999), can be considered more easily. The temporal changes of the ERD (Pfurtscheller, 1999) are also considered by the time-variation of the AAR parameters. Compared to the segmentation-based AR estimation (Florian and Pfurtscheller 1995, Pfurtscheller 1999), the computational effort of the adaptive estimation algorithms is reasonable; moreover, the time-resolution of the parameters is the same as the sampling rate.

In summary, the AAR method is appropriate for on-line and single trial analysis of the time-varying EEG spectrum. The advantages of the AAR method, compared to other methods, are most significant in cases where no averaging of an ensemble of recordings can be applied.

REFERENCES

Abraham B. and Leodolter J. (1983) *Statistical Methods for Forecasting*, Wiley, New York.

Akaike H. (1969) Fitting Autoregressive models for prediction. Ann. Inst. Statist. Math. 21, 243-247.

Akaike H. (1974). A new look at the statistical model identification, *IEEE Trans. Autom.* Control, vol. AC-19, pp.716-723.

Akaike, H. (1979) A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika,* 66, 2, 237-42.

Akay M. (1994) *Biomedical Signal Processing*, Acad. Press, San Diego.

Anderer P., Roberts S., Schlögl A., Gruber G., Klösch G., Herrmann W., Rappelsberger P., Filz O., Barbanoj M-J., Dorffner G., Saletu B. (1999) Artifact processing in computerized analysis of sleep EEG - a review, *Neuropsychobiology*, 40(3): 150-157.

Andrew C.M. (1997) Multivariate spectral analysis of movement-related EEG data. Shaker Verlag, Aachen, Germany.

Andrew C.M. (1999) Quantification of event-related coherence. in Pfurtscheller, G. and Lopes da Silva, F.H. (Eds.) *Event-Related Desynchronization.* Handbook of Electroenceph. and Clin. Neurophysiol. Revised Edition Vol. 6. Elsevier, Amsterdam.

Arnold M., Miltner W., Witte H., Bauer R., Braun C. (1998) Adaptive AR Modeling of Nonstationary Time Series by Means of Kalman filtering. *IEEE Trans Biomed Eng* 45(5): 553-562.

Aserinsky E., Kleitman N. (1953) Regularly occurring episodes of eye mobility and concomitant phenomena during sleep. *Science*, 118: 273-274.

Basar E., Basar-Eroglu C., Karakas S., Schürmann M. (1999) Oscillatory Brain Theory: A new trend in Neuroscience. *IEEE Eng. Med. and Biol.* 18(3): 56-66.

Bianchi A., Mainardi L., Meloni C., Chierchia S., Cerutti S. (1997) Continuous monitoring of the Sympatho-Vagal Balance through spectral analysis. *IEEE Engineering in Medicine and Biology*. 16(5): 64-73.

Bauby J-D. *Schmetterling und Taucherglocke,* Paul Zsolnay Verlag,Wien, 1997. (orig. franz.: Le scaphandre et le papillon, Edition Robert Laffont, S.A. Paris, 1997.)

Berger H. (1929) Über das Elektroenzephalogramm des Menschen, *Arch. Psychiat. Nervenkr*. 87:527-570.

Birbaumer N. (1999), Selbstkontrolle des Gehirns und die Beherrschung von Krankheiten. *Materie, Geist und Bewußtsein*, Forum Alpbach.

Birbaumer N., Elbert T., Rockstroh B. and Lutzenberger W. (1981) Biofeedback of event-related potentials of the brain. *Int. J. Psychophysiol.* 16: 389-415.

Birbaumer N., Ghanayim N., Hinterberger T., Iversen I., Kotchoubey B., Kübler A., Perelmouter J., Taub E., Flor H. (1999) A brain-controlled spelling device for the completely paralyzed. *Nature* 398: 297-298.

Blechschmid H. (1977) Die Analyse nichtstationärer Zeitserien mit Hilfe eines Kalman-filters. Diplomarbeit, Technische Universität Graz, Austria.

Blechschmid H. (1982) Die mathematische EEG-Auswertung mit einem schnellen online-fähigen Kalman-Filter. PhD-Thesis, University of Technology Graz, Austria.

Bodenstein G. and Praetorius H.M. (1977) Feature extraction from the electroencephalogram by adaptive segmentation, *Proc. IEEE*, 65: 642-657.

Bohlin T. (1972) A method of analyzing EEG-signals with changing spectra. Digest of the 3rd International Conference on Medical Physics, Including Medical Engineering. Chalmers Univ. Technol, Gothenburg, Sweden; 1972; xvi+317 pp. p.21-6.

Burg J.P. (1967) Maximum entropy spectral analysis, in *37th Ann. Int. Meet., Soc. Explor.Geophys*., Oklahoma City, Okla.

Burg J.P. (1975) Maximum entropy spectral analysis, PhD-thesis, Stanford, University, Stanford, Calif.

CEN (1995) Vital signs Information Representation Version 1.2, Interim Report - CEN/TC251/WG5/N95-3, European Committee for Standardisation, Brussels.

Dement W. and Kleitmann N. (1957) Cyclic variations in EEG during sleep and their relation to eye movemnts, body motility and dreaming. *Electroenceph. clin. Neurophysiol.* 9: 673-690.

Dorffner G. (1998) Towards a new standard of modeling sleep based on polysomnograms - the SIESTA project. Proc. ECCN 98, Ljubljana, *Electroenceph. and Clin. Neurophys*. 106(Suppl. 1001): 28.

Duda R.O. and Hart P.E. (1973), *Pattern classification and Scene Analysis,* John Wiley & Sons.

Durbin J. (1960) The fitting of time series models, *Rev. Int. Stat. Inst.*, 28, 233-244.

Duquesnoy A.J. (1976) Segmentation of EEG's by means of Kalman filtering. Progress Report No. PR5, pp.87-92, Institute of Medical Physics TNO, Utrecht.

Elder S.T., Lashley J.K. and Steck C.G. (1982) Amyotriphic Lateral Sclerosis: A challenge for biofeedback. *American J. of Clin. Biofeedback.* 5(2): 123-125.

Fenwick P.B., Mitchie P., Dollimore J., Fenton G.W. (1969) Application of the autoregressive model to E.E.G. analysis. Agressologie. 10:Suppl:553-64.

Fenwick P.B., Mitchie P., Dollimore J., Fenton G.W. (1970) The use of the autoregressive model in EEG analysis. *Electroencephalogr Clin Neurophysiol.* 29(3):327.

Fenwick P.B., Michie P., Dollimore J., Fenton G.W. (1971) Mathematical simulation of the electroencephalogram using an autoregressive series. *Int J Biomed Comput.* 2(4):281-307.

Florian G. and Pfurtscheller G. (1995) Dynamic spectral analysis of event-related EEG data. *Electroenceph. clin. Neurophysiol.* 95: 393-396.

Flotzinger D., Kalcher J., Pfurtscheller G. (1992) EEG classification by Learning Vector Quantization. *Biomed. Technik,* 37: 303-309.

Flotzinger D., Pfurtscheller G., Neuper Ch., Berger J., Mohl W. (1994) Classification of non-averaged EEG data by learning vector quantisation and the influence of signal preprocessing. *Medical & Biological Engineering & Computing,* 32: 571-576.

Gersch W. (1970) Spectral analyisis of EEGs by autoregressive decomposition of time series. *Math. Biosci.,* 7, 205-222.

Gersch W. (1987) Nonstationary mulitchannel time series analysis, pp. 261-296 in Gevins A.S. and Rèmond A, (eds) *Methods of Analysis of brain electrical and magnetic signals, EEG handbook* (revised series, vol. 1), Elsevier.

Goel V., Brambrink A.M., Baykal A., Koehler R.C., Hanley D.F., Thakor N.V. (1996) Dominant frequency analysis of EEG reveals brain's response during injury and recovery. IEEE-Transactions-on-Biomedical-Engineering. 43(11): 1083-92

Goncharova I.I. and Barlow J.S. (1990) Changes in EEG mean frequency and spectral purity during spontaneous alpha blocking. *Electroenceph. clin. Neurophysiol.,* 76: 197-204.

Grewal M.S. and Andrews A.P. (1993) *Kalman filtering: Theorie and Practice.* Prentice Hall, Englewood Cliffs, New Jersey.

Guger C., Schlögl A., Walterspacher D., Pfurtscheller G. (1999) Design of an EEG-based Brain-Computer Interface (BCI) from Standard Components running in Real-time under Windows. *Biomedizinische Technik,* 44: 12-16.

Guger C., Schlögl A., Neuper C., Walterspacher D., Strein T., Pfurtscheller G. (2000) Rapid prototyping of an EEG-based brain-computer interface (BCI), *IEEE Trans. Rehab. Eng.*, in press.

Hannan E.J. and Quinn B.G. (1979) The determination of the order of an autoregression, *J. Roy. Statist. Soc.*, Ser. B,41: 190-195

Hannan E.J. (1980) The estimation of the order of an ARMA process, *Ann. Statist,* 8: 1071-1081.

Haring G. (1975) Über die Wahl der optimalen Modellordung bei der Darstellung von stationären Zeitreihen mittels Autoregrressivmodels als Basis der Analyse vom EEG - Signalen mit Hilfe eines Digitalrechners. Habilschrift, Technische Universität Graz.

Harke K. C., Schlögl A., Anderer P., Pfurtscheller G. (1999) Cardiac field artifact in sleep EEG. *Proceedings EMBEC'99,* Vienna, Austria, Part I, 482-483.

Hasan J., Hirvonen K., Värri A., Häkkinen V., Loula, P. (1993) Validation of computer analysed polygraphic patterns during drowsiness and sleep onset. *Electroenceph. clin. Neurophysiol.* 87, 117-127.

Hasan J. (1996) Past and future of computer-assisted sleep analysis and drowsiness assessment. *J Clin Neurophysiol.* 13(4):295-313.

Haustein W, Pilcher J, Klink J, Schulz H. (1986) Automatic analysis overcomes limitations of sleep stage scoring. *Electroencephalogr Clin Neurophysiol.* 64(4):364-74.

Haykin S. (1996) *Adaptive Filter Theory*. Prentice Hall, Englewood Cliffs, NJ.

Haykin S., Sayed A.H. , Zeidler J.R., Yee P., Wei P.C. (1997) Adaptive Tracking of Linear Time-varinat Systems by Extended RLS Algorithms, *IEEE Trans. Signal Proc.* 45(5): 1118-1128.

Haykin S. (1999) Adaptive Filters, *IEEE Signal Processing Magazine*, 16(1): 20-22.

Hjorth B. (1970) EEG analysis based on time domain parameter. *Electroenceph. clin. Neurophysiol.*, 29: 306-310.

Isaksson A. (1975) SPARK - A sparsely updated Kalman filter with application to EEG signals. Technical Report 120, Department of Telecommunication Thearoy, Royal Insitute of Technology, Stockholm.

Isaksson A. and Wennberg A. (1975) Visual evaluation and computer analysis of the EEG - A comparison. *Electroenceph. clin. Neurophysiol.*, 38: 79-86.

Isaksson A., Wennberg A., Zetterberg L.H. (1981) Computer Analysis of EEG signals with Parametric Models, *Proc. IEEE*, 69(4): 451-463.

Ille N., Berg P., Scherg M. (1997) A spatial component method for continuous artifact correction in EEG and MEG. *Biomedizinische Technik (Ergänzungsband 1)* 42: 80-83.

Jazwinski A. H. (1969) Adaptive filtering, *Automatica,* 5(4): 475-485.

Jansen B.H., Hasman A., Lenten R., Visser S.L. (1979) Usefulness of autoregressive models to classify EEG-segments. *Biomedizinische Technik*. 24(9): 216-23.

Jansen B.H., Bourne J.R., Ward J.W. (1981) Autoregressive estimation of short segment spectra for computerized EEG analysis. IEEE Trans. Biomedical Engineering. 28(9).

Jansen B.H., Dawant B.M. (1989) Knowledge-based approach to sleep EEG analysis-a feasibility study. *IEEE Trans Biomed Eng.* 36(5):510-8.

Jobert M., Mineur J., Scheuler W., Kubicki S., Scholz G. (1989) System zur koninuierlichen Digitalisierung und Auswertung von 32 Biosignalen bei Ganznacht-Schlafableitungen. *EEG EMG Z Elektroenzephalogr Elektromyogr Verwandte Geb*. 20(3):178-84.

Judy P.F., Schaefer C.M., Green R.E., Oestmann, J-W. (1992) *Measuring Observer Performance of Digital Systems*. In eds. Green and Oestmann, Thieme Medical Publishers, New York.

Kaipio J.P. and Karjalainen P.A. (1997a) Simulation of nonstationary EEG, *Biol. Cybern*. 76(5): 349-56.

Kaipio J.P. and Karjalainen P.A. (1997b) Estimation of event-related synchronization changes by a new TVAR method. *IEEE Trans Biomed Eng*. 44(8): 649-56.

Kalcher J., Flotzinger D., Neuper Ch., Gölly S., Pfurtscheller G. (1996) Graz Brain-Computer Interface II - Towards communication between humans and computers based on online classification of three different EEG patterns, *Med. Biol. Eng. Comput.*, 34: 382-388.

Kalman R.E. (1960) A new approach to Linear Filtering and Prediction Theory, *Journal of Basic Engineering Trans. of ASME*, 82: 34-45.

Kalman R.E. and Bucy R.S. (1961) New Results on Linear Filtering and Prediction Theory, *Journal of Basic Engineering*, 83: 95-108.

Kemp B. (1983) Accurate measurement of flash-evoked alpha attenuation. *Electroenceph. Clin. Neurophysiol.* 56(2): 248-53.

Kemp, B. (1993) A proposal for computer based sleep/wake analysis (Consensus report). *J Sleep Res.*, 2: 179-185.

Kemp B. and Blom H.A.P. (1981) Optimal detection of the alpha state in a model of the human electroencephalogram. *Electroenceph. Clin. Neurophysiol.* 52(2): 222-5.

Kemp B. and Lopes Da Silva F.H. (1991) Model-based analysis of neurophysiological signals, Weitkunat, R. (ed.) *Digital Signal Processing,* Elsevier.

Kemp B., Värri A., Rosa A.C., Nielsen K.D., Gade J. (1992) A simple format for exchange of digitized polygraphic recordings. *Electroenceph. clin. Neurophysiol.,* 82: 391-393.

Kemp B, Groneveld EW, Janssen AJ, Franzen JM. (1987) A model-based monitor of human sleep stages. *Biol Cybern.* 57(6):365-78.

Klimesch W., Doppelmayr M., Russegger H. and Pachinger Th. (1998) A method for the calculation of induced band power; implications for the significance of brain oscillations. *Electroenceph. Clin. Neurophysiol.* 108: 123-130.

Klimesch W. (1999) Event-related band power changes and memory performance, in Pfurtscheller, G. and Lopes da Silva, F.H. (Eds.) *Event-Related Desynchronization*. Handbook of Electroenceph. and Clin. Neurophysiol. Revised Edition Vol. 6. Elsevier, Amsterdam.

Kohonen, T. (1995) *Self-organizing maps*, Springer.

Kubat M, Pfurtscheller G, Flotzinger D. (1994) AI-based approach to automatic sleep classification. *Biol Cybern.* 70(5):443-8.

Kubat M., Flotzinger D., Pfurtscheller G. (1993) Towards automated sleep classification in infants using symbolic and subsymbolic approaches. *Biomed Tech.* 38(4):73-80.

Kubicki S, Herrmann W.M. (1996) The future of computer-assisted investigation of the polysomnogram: sleep microstructure. *J Clin Neurophysiol.* 13(4):285-94.

Kubicki S, Holler L, Berg I, Pastelak-Price C, Dorow R. (1989) Sleep EEG evaluation: a comparison of results obtained by visual scoring and automatic analysis with the Oxford sleep stager. *Sleep*. 12(2):140-9.

Kubicki, St., Herrmann, W.M., Höller, L., Scheuler, W., (1982) Kritische Bemerkungen zu den Regeln von Rechtschaffen und Kales über die visuelle Auswertung von EEG-Schlafableitungen, *Z. EEG-EMG* 13: 51-60.

Lagerlund T.D., Sharbrough F.W., Busacker N.E. (1997) Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition. *J Clinic Neurophysiol.* 14: 73-82.

Larsen L.E. and Walter D.O. (1970) On automatic methods of sleep staging by EEG spectra. *Electroencephalogr Clin Neurophysiol.* 28(5):459-67.

Levinson N. (1947) The Wiener RMS (root-mean-square) error criterion in filter design and prediction, *J. Math. Phys*., 25, 261-278.

Loomis A.L., Harvey E.N., Hobart G.A. (1937) Cerebral states during sleep as studied by human brain potentials. *J. Exp. Psychol.* 21:127-144.

Loomis A.L., Harvey E.N., Hobart G.A. (1938) Distribution of disturbance-patterns in the human electroencephalogram with special reference to sleep. *J. Neurophysiol.* 1:413-430.

Lopes da Silva F.H. (1991) Neural mechanisms underlying brain waves: from neural membranes to networks. *Electroencephalogr Clin Neurophysiol.* 79(2):81-93.

Lopes da Silva F.H., Hoeks A., Smits H., Zetterberg L.H. (1974) Model of brain rhythmic activity. The alpha-rhythm of the thalamus. *Kybernetik*. 15(1):27-37.

Lopes da Silva, F.H., van Hulten, K., Lommen, J.G. Stomr van Leeuwen, W., von Veelen, C.W.M. and Vliegenthart, W. (1977) Automatic detection and localization of epiliptic foci, *Electroenceph. Clin. Neurophysiol.* 43: 1-13.

Lopes da Silva F.H. (1999) EEG analysis: theory and practice. In Niedermeyer E. and Lopes da Silva F.H. (Eds.) *Electroencephalography - basic prinicples, clinical Applications and related fields.* Urban & Schwarzenberg, 4th edition.

Lopes da Silva F.H. and Pfurtscheller G. (1999) Basic concepts of EEG sychronization and desychronization. in Pfurtscheller, G. and Lopes da Silva, F.H. (Eds.) *Event-Related Desynchronization.* Handbook of Electroenceph. and Clin. Neurophysiol. Revised Edition Vol. 6. Elsevier, Amsterdam.

Lugger K., Flotzinger D., Schlögl A., Pregenzer M., Pfurtscheller G. (1998) Feature extraction for on--line EEG classification using principal components and linear discriminants. *Med. Biol. Eng. Comput.*, 36: 309-314.

Lustick L.S., Saltzberg B, Buckley J.K., Heath R.G. (1968) Autoregressive model for simplified computer generation of EEG correlation functions. *Proceedings of the annual conference on engineering in medicine and biology*, Vol.10. IEEE, New York, NY, USA; 1968; 552+xxvii pp. 1 pp.

Mainardi L.T., Bianchi A.M., Baselli G., and Cerutti S. (1995) Pole-tracking algorithms for the extraction of time-variant heart rate variability spectral parameter. *IEEE Trans Biomed Eng.* 42(3): 250-9.

Marple Jr. A.L. (1987) Digital spectral Analysis with applications. Prentice-Hall, Englewood Cliffs, N. J.

Mathieu M. (1976) Anayse de l'electroencéphalogramme par prédiction linéartre. *These*, Université Pierre et Marie Curie, Paris.

McFarland D.J., Neat G.W., Read R.F., Wolpaw J.R. (1993) An EEG-based method for graded cursor control, *Psychobiol*., 21: 77-81.

McFarland D.J., Lefkowicz A.T., Wolpaw J.R. (1997) Design and operation of an EEG-based brain-computer interface with digital signal processing technology. *Behavior Research Methods, Instruments & Computers,* 29(3): 337-345.

McFarland D.J., McCane L.M., Wolpaw J.R. (1998) EEG-based communication and control: short-term role of feedback. *IEEE Trans Rehab Engng.* 6: 7-11.

Medl A., Flotzinger D., Pfurtscheller G. (1992) Hilbert-Transform Based Predictions of Hand Movement from EEG Measurements. *Proceedings of the 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Paris, France, 2539-2540.

Meinhold R. J. and Singpurwalla N.D. (1983) Understanding Kalman filtering. *The American Statistician*. 37(2): 123-127.

Merica H, Fortune RD. (1997) A neuronal transition probability model for the evolution of power in the sigma and delta frequency bands of sleep EEG. Physiol Behav. 62(3):585-9.

Michie, D., Spiegelhalter, D.J. and Taylor, C.C. (1994) *Machine Learning, Neural and Statistical Classification*, Ellis Horwood Limited, Englewood Cliffs, N.J..

Mourtazaev M.S., Kemp B., Zwinderman A.H., Kamphuisen H.A. (1995) Age and gender affect different characteristics of slow waves in the sleep EEG. *Sleep* 18(7): 557-64.

Neuper C., Schlögl A., Pfurtscheller G. (1999) Enhancement of left-right sensorimotor EEG differences during feedback-regulated motor imagery. *J. Clin. Neurophysiol.* 16(4): 373-82.

Niedermeyer E. and Lopes da Silva F.H. (1999) *Electroencephalography - basic prinicples, clinical Applications and related fields.* Urban & Schwarzenberg, 4th edition.

Nielsen K.D. (1993) Computer assisted sleep analysis, PhD-Thesis, Aalborg University, Denmark, ISBN 87-984421-0-4.

Nielsen K.D., Drewes A.M., Svendsen.L., Bjerregard K., Taagholt S.A. (1994) Ambulatory recording and power spectral analysis by autoregressive modelling of polygraphic sleep signals in patients suffering from chronic pain. *Methods of Information in Medicine.* 33(1): 76-80.

Nikias C.L. and Petropulu A.P. (1993) *Higher-order spectra analysis*. Prentice Hall, Englewood Cliffs, NJ.

Oppenheim A.V. and Schafer R.W. (1975) *Digital Signal Processing*, Prentice Hall, Englewood Cliffs, NJ.

Parzen E. (1977) Multiple time series modeling: Determining the order of approximating autoregressive schemes, in Multivariate Analysis IV In "Applications of Statistics (P.R. Krishnaiah, ed.) 283-295, North Holland, Amsterdam.

Patomäki L., Kaipio J.P., Karjalainen P.A. (1995) Tracking of nonstationary EEG with the roots of ARMA models. in *Proceedings of 1995 IEEE Engineering in Medicine and*

Biology 17th Annual Conference and 21 Canadian Medical and Biological Engineering. IEEE, New York, NY, USA 2: 887-8.

Patomäki L., Kaipio J.P., Karjalainen P.A., Juntunen M. (1996) Tracking of nonstationary EEG with the polynomial root perturbation. in *Proceedings of the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* IEEE, New York, NY, USA; 3: 939-40.

Penny W.D. and Roberts S.J. (1998) Dynamic linear models, Recursive Least Squares and Steepest descent learning, *Technical Report,* Imperial College London.

Pfurtscheller G. (1999) Quantification of ERD and ERS in the time domain. In Pfurtscheller, G. and Lopes da Silva, F.H. (Eds.) *Event-Related Desynchronization.* Handbook of Electroenceph. and Clin. Neurophysiol. Revised Edition Vol. 6. Elsevier, Amsterdam.

Pfurtscheller G and Haring G. (1972) The use of an EEG autoregressive model for the time-saving calculation of spectral power density distributions with a digital computer. *Electroencephalogr Clin Neurophysiol.* 33(1):113-5.

Pfurtscheller G. and Cooper, R. (1975) Frequency dependence of the transmission of the EEG from cortex to scalp. *Electroenceph. clin. Neurophysiol.* 38: 93-96.

Pfurtscheller G. and Aranibar A. (1977) Event-related cortical desynchronisation detected by power measurements of scalp EEG. *Electroenceph. clin. Neurophysiol.,* 42, 817-826.

Pfurtscheller G., Flotzinger D., Kalcher J. (1993) Brain-computer interface—a new communication device for handicapped persons, *J. Microcomp. Appl.,* 16: 193-299.

Pfurtscheller G. and Guger C. (1999) Brain-Computer Communicatio System: EEG-based Control of Hand Orthesis in a tetraplegic patient. *Acta Chir. Autriaca*, 31, Suppl. 159, 23-25.

Pfurtscheller G., Kalcher J., Neuper Ch., Flotzinger D., Pregenzer M. (1996) Online EEG classification during externally-paced hand movements using a neural network-based classifier, *Electroenceph. Clin. Neurophysiol.,* 99: 416-425.

Pfurtscheller G., Neuper Ch., Flotzinger D., Pregenzer M. (1997) EEG-based discrimination between imaginary right and left hand movement. *Electroenceph. clin. Neurophysiol.* 103(6): 642-651.

Pfurtscheller G., Neuper C., Schlögl A., Lugger K. (1998) Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameter. *IEEE Trans. on Rehab. Eng.* 6(3): 316-25.

Pfurtscheller G. and Lopes da Silva F.H. (1999a) *Event-Related Desynchronization and Related Oscillatory Phenomena of the Brain.* Handbook of Electroenceph. and Clin. Neurophysiol. Revised Edition Vol. 6. Elsevier, Amsterdam.

Pfurtscheller G. and Lopes da Silva F.H. (1999b) Event-Related EEG/MEG synchronization and desynchronization: basic prinicples. *Clin. Neurophysiol.* 110: 1842-1857.

Praetorius H.M., Bodenstein G. and Creutzfeld O. (1977) Adaptive Segmentation of EEG records: A new approach to automatic EEG analysis. *Electroenceph. Clin. Neurophysiol.* 42: 84-94.

Pregenzer M. (1998) DSLVQ - Distinct Sensitive Learning Vector Quantization, Shaker Verlag, Aachen, Germany.

Pregenzer M., Pfurtscheller G., Flotzinger D. (1994) Selection of electrode positions for an EEG-based Brain Computer Interface, *Biomed. Technik*, 39: 264-269.

Pregenzer M., Pfurtscheller G., Flotzinger D. (1996) Automated feature selection with a distinction sensitive learning vector quantizer. *Neurocomputing*, 11: 19-29.

Priestley M.B. (1981) *Spectral Analysis and Time Series.* Academic Press, London, UK.

Priestley M.B. (1988) *Non-linear and non-stationary Time Series Analysis.* Academic Press, London, UK.

Principe J.C., Smith J.R. (1986) SAMICOS - a sleep analyzing microcomputer system. *IEEE Trans Biomed Eng.* 33(10):935-41.

Principe J.C., Gala S.K., Chang T.G. (1989) Sleep staging automaton based on the theory of evidence. *IEEE Trans Biomed Eng.* 6(5):503-9.

Principe J.C., B. de Vries and P.G. Oliveira (1993) The gamma filter - a new class of adaptive iir filters with restricted feedback. IEEE Trans. Signal Proc., 41(2): 649-656.

Pukkila T., Krisnaiah P., (1988) On the use of autoregressive order determination criteria in multivariate white noise tests. *IEEE Trans. on ASSP* 36(9): 1396-1403.

Rechtschaffen A. and Kales A. (1968) *A manual of standardized terminology techniques and scoring system for sleep stages in human subjects*. U.S. Departmant of Health, Education and Welfare, Public Health Service, Washington, D.C.: U.S. Government Printing Office.

Rieke, F., Warland D., Rob de Ruyter van Steveninck and Bialek W. (1997) *Spikes - Exploring the neural code*. MIT Press, Cambridge.

Rissanen, J. (1978) Modeling by shortest data description. *Automatica* 14, 465-471.

Rissanen, J. (1983) Universal Prior for the Integers and Estimation by Minimum Description Length. *Ann. Stat.*, 11: 417-431.

Roberts S.J. (1997) Matlab source code of Kalman filtering , personal communications.

Roberts S.J. and Tarassenko L. (1992) New Method of Automated Sleep Quantification. *Medical and Biological Engineering and Computing*, 30(5): 509-517.

Roberts S. and Tarassenko L. (1992a) The analysis of the sleep EEG using a multi-layer Neural network with spatial organisation. *IEE Proceedings Part F*, 139(6): 420-425.

Robinson C.J. (1999) An Information Theory View of the Use of Brain-Control Interfaces in augmentative Commmunication. *First international meeting on Brain-Computer Interface technology: theory and Practice,* The Rensselaerville Institute, Rensselaerville, NY.

Sahul Z., Black J., Widrow B., Guilleminault C. (1995) EKG artifact cancellation from sleep EEG using adaptive filtering. *Sleep Research*, 24A: 486.

Sayed A.H. and Kailath, T. (1994) A state-space aproach to adaptive RLS filtering, *IEEE Signal Processing Magazine*, 18-60.

Schack B., Witte H., Grießbach G. (1993) Parametrische Methoden der dynamischen Spektralanalyse und ihre Anwendung in der Biosignalanalyse. *Biomedizinische Techni*k, 38, 79-80.

Schack B., Bareshova E., Grieszbach G., Witte H. (1995) Methods of dynamic spectral analysis by self-exciting autoregressive moving average models and their application to analysing biosignals, *Med. Biol. Eng. Comput*. 33: 492-8.

Schlögl A. (1995) Dynamic spectral analysis based on an Autoregressive Model with time-varying coefficients, *Proceedings of 1995 IEEE Engineering in Medicine and Biology 17th Annual Conference and 21 Canadian Medical and Biological Engineering*. IEEE, New York, NY, USA, 881-2.

Schlögl A., Schack B., Florian G., Lugger K., Pregenzer M., Pfurtscheller G. (1996) Classification of Single trial EEG: A comparison of different parameter. in *Qualitative and Topological EEG and MEG analysis* - Proc. Third Int Hans Berger Congress, eds. H. Witte, U. Zwiener, B. Schack, A. Doering; Druckhaus Mayer, Jena, pp.266-268.

Schlögl A., Flotzinger D., Pfurtscheller G. (1997a) Adaptive Autoregressive Modeling used for Single-Trial EEG Classification. *Biomed. Techn*. 42: 162-167.

Schlögl A., Neuper C., Pfurtscheller G. (1997b) Subject specific EEG patterns during motor imaginary. *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* IEEE, Piscataway, NJ, USA, pp.1530-1532.

Schlögl A., Lugger K., Pfurtscheller G. (1997c) Using adaptive autoregressive parameter for a brain computer-interface experiment. *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* IEEE, Piscataway, NJ, USA, pp.1533-1535.

Schlögl A. and Pfurtscheller G. (1998a) Considerations on Adaptive Autoregressive Modeling in EEG Analysis. *Proceedings of First International Symposium on Communication Systems and Digital Signal Processing,* Sheffield, UK, pp.367-370.

Schlögl A., Kemp B., Pfurtscheller G. (1998b) Parametric Models in EEG analysis. - analogies between Autoregressive and Kemp's model. Abstracts of the 9th European Congress of Clinical Neurophysiology Ljubljana, Slovenia June 3-7, 1998 - *Electroenceph. Clin. Neurophys*. 106(Suppl. 1001): 40.

Schlögl A., Penzel T., Conradt R., Ramoser H., Pfurtscheller G. (1998c) Analysis of Sleep EEG with Adaptive Autoregressive Parameter - preliminary results, Abstracts of the 9th European Congress of Clinical Neurophysiology Ljubljana, Slovenia June 3-7, 1998 - *Electroenceph. and Clin. Neurophys.* 106(Suppl. 1001): 28.

Schlögl A., Woertz M., Trenker E., Rappelsberger P., Pfurtscheller G. (1998d) Adaptive Autoregressive Modeling used for Single-trial EEG Classification, Proc. 14th Congress European Sleep Research Society ESRS'98 in Madrid. *Journal of Sleep Research* 7(Suppl.2): 242.

Schlögl A., Kemp B., Penzel T., Kunz D., Himanen S.-L., Värri A., Dorffner G., Pfurtscheller G. (1999b) Quality control of polysomnographic sleep data by histogram and entropy analysis. *Clin. Neurophys.* 110(12): 2165 - 2170.

Schlögl A., Anderer P., Barbanoj M.-J., Dorffner G., Gruber G., Klösch G., Lorenzo J.L., Rappelsberger P., Pfurtscheller G. (1999d) Artifacts in the sleep EEG - A database for the evaluation of automated processing methods. *Sleep Research Online* 2(Supplement 1): 586.

Schlögl A., Anderer P., Barbanoj M.-J., Klösch G., Gruber G., Lorenzo J.L., Filz O., Koivuluoma M., Rezek I., Roberts S.J., Värri A., Rappelsberger P., Pfurtscheller G., Dorffner G. (1999e) Artifact processing of the sleep EEG in the SIESTA project. *Proceedings EMBEC'99*, Vienna, Austria, Part II, pp.1644-1645.

Schlögl A., Anderer P., Roberts S.J., Pregenzer M., Pfurtscheller G. (1999f) Artefact detection in sleep EEG by the use of Kalman filtering. *Proceedings EMBEC'99*, Vienna, Austria, Part II, pp.1648-1649.

Schlögl A., Neuper C., Pfurtscheller G. (1999g) Estimating the Mutual Information of EEG-based Brain Computer Communication, *IEEE Trans. Rehab. Eng.* submitted.

Schlögl A., Roberts S.J., Pfurtscheller G. (2000) A criterion for adaptive autoregressive models. *Proc. of the World Congress on Medical Physics and Biomedical Engineering.* Chicago, accepted.

Schwartz, G. (1978) Estimating the dimension of a model. *Ann. Statist.*, 6: 461-464.

Shannon, C.E. (1948) *The mathematical theory of communication*; Bell Syst. Tech. J., 27: 379-423, 623-656.

Shinn Yih Tseng, Rong Chi Chen, Fok Ching Chong; Te Son Kuo (1995) Evaluation of parametric methods in EEG signal analysis. *Medical Engineering & Physics.* 17(1): 71-8.

Skagen D.W. (1988) Estimation of running frequency spectra using a Kalman filter algorithm, *J. Biomed Eng*, 10: 275-279.

Smith J.R., Funke W.F., Yeo W.C., Ambuehl R.A. (1975) Detection of human sleep EEG waveforms. *Electroencephalogr Clin Neurophysiol.* 38(4):435-7.

Smith J.R. and Karacan I. (1971) EEG sleep stage scoring by an automatic hybrid system. *Electroencephalogr Clin Neurophysiol.* 31(3):231-7.

Stancak A. Jr, Pfurtscheller G. (1996) The effects of handedness and type of movement on the contralateral preponderance of mu-rhythm desynchronisation. *Electroencephalogr Clin Neurophysiol*. 99(2):174-82.

Stanus E., Lacroix B., Kerkhofs M., Mendlewicz J. Automated sleep scoring: a comparative reliability study of two algorithms. *Electroencephalogr Clin Neurophysiol.* 66(4):448-56.

Steriade M., Gloor P., Llinas R.R., Lopes da Silva F., Mesulam M. (1990) Report of IFCN Committee on Basic Mechanisms. Basic mechanisms of cerebral rhythmic activities. *Electroencephalogr Clin Neurophysiol.* 76(6): 481-508.

Stone M. (1978) Cross-validation: A review. *Mathematische Operationsforschung und Statistik, - Series Statistics.* 9(1):127-39.

Suffczynski P., Pijn J.P., Pfurtscheller G. and Lopes da Silva F.H. (1999) Event-related dynamics of alpha band rhythm: A neuronal network model of focal ERD/surrounded ERS. In: Pfurtscheller and Lopes da Silva (1999) (Eds.) *Event-Related Desynchronization.* Handbook of Electroenceph. and Clin. Neurophysiol. Revised Edition Vol. 6. Elsevier, Amsterdam.

Valdes P.A., Jimenez, H.C. Riera J., Biscay, R. Ozaki T. (1999) Nonlinear EEG analysis based on a neural mass model. Biological Cybernetics. 81, 415-424.

Van de Velde M., van den Berg-Lenssen M.M., van Boxtel G.J., Cluitmans P.J., Kemp B., Gade J., Thomsen C.E., Varri A. (1998) Digital archival and exchange of events in a simple format for polygraphic recordings with application in event related potential studies. *Electroencephalogr Clin Neurophysiol.* 106(6): 547-51.

Vaz F., Guedes De Oliveira P., Principe J.C. (1987) A study on the best order for autoregressive EEG modelling. *International Journal of Bio-Medical Computing*. 20(1-2): 41-50.

Wei W.W. (1990) *Time Series Analysis. Univariate and multivariate methods.* Redwood City, Addison-Wesley.

Widrow B. and Stearns S.D. (1985) *Adaptive Signal Processing*, Prentice Hall, Englewood Cliffs, NJ.

Woertz M., A. Schlögl A., Trenker E., Rappelsberger P., Pfurtscheller G. (1999) Spindle Filter Optimisation with Receiver Operating Characteristics-Curves *Proceedings EMBEC'99,* Vienna, Austria, Part I, 480-481.

Wolpaw J.R., McFarland D.J., Neat G.W., Forneris C. (1991) An EEG-based brain-computer interface for cursor control, *Electroenceph. Clin. Neurophysiol.,* 78, 252-259.

Wolpaw J.R. and McFarland D.J. (1994) Multichannel EEG-based brain-computer communication, *Electroenceph. Clin. Neurophysiol.,* 90, 444-449.

Zetterberg L.H. (1969) Estimation of parameter for linear difference equation with application to EEG analysis. *Math. Biosci.,* 5, 227-275.

Zetterberg LH, Kristiansson L, Mossberg K. Performance of a model for a local neuron population. *Biol Cybern.* 31(1):15-26, 1978.

APPENDIX

A. Notation

$\otimes$	convolution operation.
∞	infinity
ΔT	sampling interval
Σ	sum
Σ_i	sum over all i
$\sigma_x^2(k)$	variance of innovation (AR model)
$\boldsymbol{A}_k$	a-priori state error correlation matrix
$\boldsymbol{AY}_k$	$\boldsymbol{A}_k * \boldsymbol{Y}_k$, intermediate variable in order to increase the computational speed.
E{.}	expectation operator
$\boldsymbol{G}_{k,k-1}$	system matrix of the state space model
$\boldsymbol{H}_k$	measurement (observation) matrix of the state space model
$\boldsymbol{K}_{k,k-1}$	a-priori state error correlation matrix
$\boldsymbol{I}_{(pxp)}$	identity matrix (of order p)
Q_k	estimated prediction error variance
V_k	estimated variance of measurement noise
UC	update coefficient
X_k	variance of measurement noise (state space model)
$\boldsymbol{Z}_{k,k-1}$	a posteriori state error correlation matrix
$a_{i,k}$	i-th AAR parameter
$\hat{a}_{i,k}$	i-th AAR estimate
$\boldsymbol{a}_k$	vector of the AAR parameter
e_k	prediction error
f_0	sampling rate
k	time index, discrete
$\boldsymbol{k}_k$	Kalman gain
log	logarithmus
$\log_2$	logarithmus dualis (logarithm of basis 2)
p	AR model order, feedback gain factor in Kemp's feedback loop model
q	Moving average model order
^{T}	transposition operator
t	time, continuous
v_k	observation noise
$\boldsymbol{w}_k$	system noise
$y(t)$	time-continuous process
y_k	time-discrete process, observation at time $t = k*\Delta T$
z^{-1}	backshift operator, independent variable in the Z-transformed expression
$\mathbf{z}_k$	state vector at time k

B. Subsampling

In a multi-center sleep research project, EEG data with different sampling rates (100, 200 and 256Hz) were available. One aim was to apply Kalman filtering for calculating adaptive autoregressive (AAR) parameters. The AR parameters are in a one-to-one relationship to the autocorrelation function as well as to the spectral composition of the signal. It will be outlined how AAR parameters for a target sampling rate (e.g. 100Hz) different to the source sampling rate (e.g. 256. Hz) can be generated. The AR-equation (2.1) shows that a value $y(k*\Delta T)$ is composed of the linear combination of the past values $y((k-i) * \Delta T_1)$, $i=0,1,2,...,p$, plus some error term. The problem would be solved if one would be able to interpolate the values for the sampling times $y(k*\Delta T_2 - i*\Delta T_2)$, $i=1,...,p$.

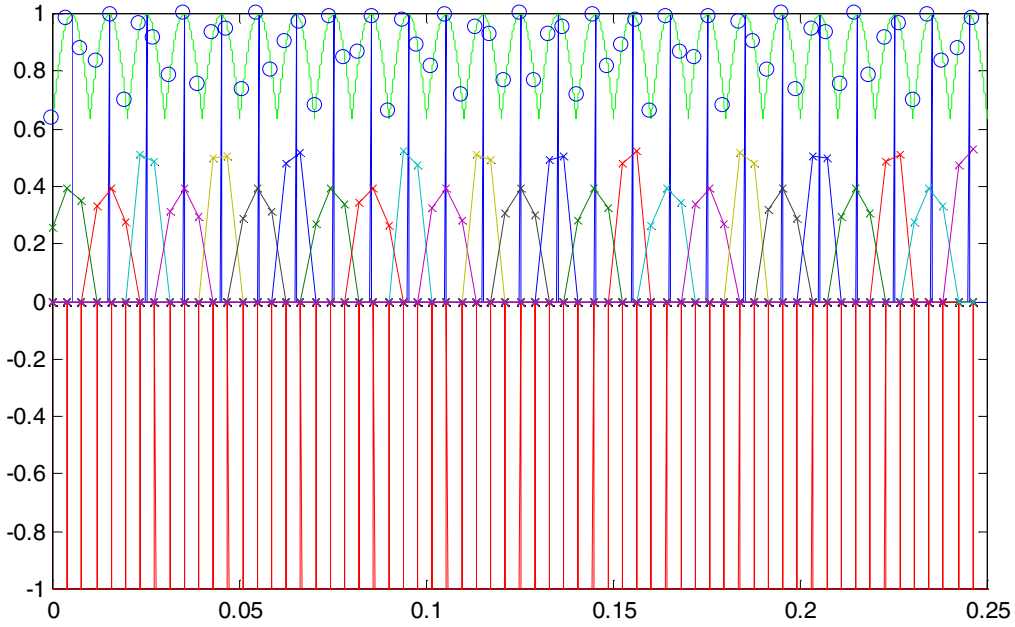


Figure 21: Scheme for subsampling from 256Hz to 100Hz for 1/4s. Each sample value of the target sampling is the "moving average" of 2 or 3 samples from the source. The weights were calculated using the sinc-function at the corresponding time point (o). These weights were normalized (x) such that the sum for each target sample is 1.

The weights of the matrix in Table 4 are obtained by the scheme in Fig. 21. The subsampling is easily performed by subsequently multiplying segments with the resampling matrix in Table 4. The resampling matrix for a source sampling rate of 200 Hz is the average of two succeeding samples; hence, the weights are all 0.5. This can be generalized for integer multiples of the target. If the ratio f_1/f_2 is integer, the weights are all f_2/f_1; f_1/f_2 succeeding samples are averaged. Furthermore, note the moving average has a low pass characteristic. However, in order to prevent aliasing, it should be ensured that the source signal contains only frequency components up to half the target sampling rate. If this is not the case, a digital anti-aliasing filter can be applied. This filter can be combined with a 50Hz Notch filter.

Table 4. Weights for the resampling matrix from 256 to 100 Hz. The matrix has 64 rows and 25 columns. Subsampling of data is performed by subsequently multiplying source vectors of length 64 with the matrix. Note that the matrix is sparse.

T	1/200	3/200	5/200	7/200	…	49/200	51/200
0	0.2563						
1/256	0.3935						
2/256	0.3502						
3/256		0.3310					
4/256		0.3942					
5/256		0.2748					
6/256			0.5135				
7/256			0.4865				
8/256				0.3104	…		
:				:	…	:	
62/256						0.4722	
63/256						0.5278	
64/256							0.2563

Until here, the sub-sampling can be performed on the raw data and afterwards, the AAR estimation algorithm can be applied. However, in the following it will be shown that the proposed sub-sampling method can be incorporated into the state space model. It can be seen from the AAR update equations (e.g. 4.1) that in each iteration p+1 sample values y_k , y_{k-1}, ... , y_{k-p} with the target sampling rate are needed. This means the interpolated values can be calculated "on the fly" meaning the iteration rate of the AAR estimation algorithm is the sampling rate of the source, while the AAR parameters consider a time distance of the target sampling rate.

Example: p=2, f_1=256Hz, f_2=100Hz.
The actual sample is

$$y_k = y(t+3/100) = 0.49*y(t+7/256)+0.51*y(t+6/256)$$

using the weights of Table 4, and the previous p samples (i.e. state vector) are

$$\mathbf{Y}_{k-1} = [y_{k-1}, \dots , y_{k-p}]^T = [y(t+2/100), y(t+1/100)]^T =$$
$$= [0.27*y(t+5/256)+0.39*y(t+4/256)+0.33*y(t+3/256),$$
$$= 0.26*y(t+2/256)+0.39*y(t+1/256)+0.35*y(t+0/256)]$$

The one-step prediction error would be

$$e_k = e(t+3/100) = y(t+3/100) - a_1*y(t+2/100) - a_2*y(t+1/100)$$
$$= 0.49 * y(t+7/256) + 0.51*y(t+6/256) +$$
$$- a_1*(0.27*y(t+5/256) + 0.39*y(t+4/256) + 0.33*y(t+3/256))$$
$$- a_2*(0.26*y(t+2/256) + 0.39*y(t+1/256) + 0.35*y(t+0/256))$$

Note, that the iteration k -> $k+1$ can be $\Delta t = \Delta T_1 = 1/f_1$ or $\Delta t = \Delta T_2 = 1/f_2$. The former case, $\Delta t = \Delta T_1 = 1/f_1$, means that firstly the data is sub-sampled and afterwards the AAR estimation algorithm is applied; in case $\Delta t = \Delta T_2 = 1/f_2$ sub-sampling is applied "on-the-fly" within the AAR estimation algorithm.

In practice, the target vector $[y'_k, \ldots, y'_{k-p}]^T$ can be simple generated by multiplying a source vector (of appropriate length m) with some part of the transformation matrix T (of appropriate size m×p+1) from Table 4. Note that p+1 elements must be generated in order to calculate the one-step prediction error. The transformation matrix is sparse and can, therefore, be implemented with O(p) computational effort.

$$[y'_k, \ldots, y'_{k-p}] = [y_l, \ldots, y_{l-m+1}] * T_{[m \times p+1]} \qquad \text{(B.1)}$$

In this section a solution for the sub-sampling problem of an autoregressive model was provided. It was shown how 100Hz-AAR-parameter can be calculated from 256Hz data.

C. Linear Discriminant Analysis

The principle of Linear Discriminant Analysis (LDA) will be outlined briefly. A more detailed view can be found in (Duda and Hart, 1973). LDA can be seen as a method to identify the regression coefficients for a certain target value in an n-dimensional feature space. Alternatively, LDA can be seen as a method for identifying the best discriminating hyperplane in an n-dimensional feature space. We assume an n-dimensional feature space with $N = N_1 + N_2$ examples; each example is represented by an n-dimensional feature vector $\boldsymbol{d}_i$. Each feature vector $\boldsymbol{d}_i$ is assigned to one of two classes C_1 and C_2, whereby N_1 and N_2 are the total number of examples within each class. It can be shown that the weight vector $\boldsymbol{w}$ is the best discriminating hyperplane between the features of both classes.

$$[w_1 \dots w_N]^T = \boldsymbol{S}_w^{-1} * (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \qquad (C.1)$$

μ_1 and μ_2 are the means of the classes C_1 and C_2 and $\boldsymbol{S}_w$ is the within-class scatter-matrix

$$\boldsymbol{S}_w = 1/N_1 * \Sigma_i (\boldsymbol{d}_{1,i} - \boldsymbol{\mu}_1)^T (\boldsymbol{d}_{1,i} - \boldsymbol{\mu}_1) + 1/N_2 * \Sigma_i (\boldsymbol{d}_{2,i} - \boldsymbol{\mu}_2)^T (\boldsymbol{d}_{2,i} - \boldsymbol{\mu}_2) \qquad (C.2)$$

Furthermore, the offset w_0 is

$$w_0 = [w_1 \dots w_N]^T * \mu \qquad (C.3)$$

with the overall mean μ (mean of the data from both classes). When $\boldsymbol{d}_i$ are the features, $\boldsymbol{w} = [w_1 \dots w_N]^T$ are the weighting factors that determine the best discriminating hyperplane. If we apply any feature vector $\boldsymbol{d}_i$, in the following way to the weight vector, we obtain a scalar value D_i that is the normal distance of $\boldsymbol{d}_i$ to the discriminating hyperplane.

$$D_i = \boldsymbol{w}^T * \boldsymbol{d}_i - w_o \qquad (C.4)$$

Applying a zero-threshold to the distance D_i results in a classification of the data $\boldsymbol{d}_i$ in the way that

$$\boldsymbol{w}^T * \boldsymbol{d}_i > w_o \text{ means } \boldsymbol{d}_i \subset C_1 \qquad (C.5)$$

$$\boldsymbol{w}^T * \boldsymbol{d}_i \leq w_o \text{ means } \boldsymbol{d}_i \subset C_2 \qquad (C.6)$$

In case of a classification task, the detection output is compared to the original class relationship and an error rate of incorrectly classified examples can be obtained. Usually, cross-validation with several permutations of training and test data set is applied. The average error rate from different permutations is a measure for the separability of the data.

D. Communication theory

Shannon (1948) introduced the concept of the "entropy of information" in communication theory (coding, information transmission). The entropy value is a measure for the variability, randomness, the average amount of choices or the average amount of information. The larger the variability, the higher is the entropy. For the purpose of estimating the entropy of the raw EEG, several definitions are applicable.
Firstly, coding theory uses the entropy of information in binary digits (bits). It is the entropy in discrete systems, which is defined as

$$I_Y = \Sigma_i \, (p_Y(i) * log_2(p_Y(i)). \qquad (D.1)$$

The index *Y* indicates the relation to the signal *Y*. For large *N*, the probability density function $p_Y(i)$ is the normalized histogram of the time series *Y* with the total number of samples

$$p_Y(i) = H_Y(i)/N_Y \qquad (D.2)$$

$$N_Y = \Sigma_i \, H_Y(i). \qquad (D.3)$$

Secondly, the entropy of a continuous distribution *x* of the signal *Y* is

$$I_Y = \int_{-\infty}^{\infty} p_Y(x) \log(p_Y(x)) dx \qquad (D.4)$$

In case of a Gaussian distribution is

$$p(x) = 1/\sqrt{(2\pi\sigma^2)} * exp(-(x-\mu)^2/(2\sigma^2)) \qquad (D.5)$$

whereby $-\infty < x < \infty$. The entropy of a continuous Gaussian process *Y* with the probability distribution *p(x)* is

$$I_Y = 0.5 * log_2(2\pi e * \sigma^2_Y) = log_2(\sqrt{(2\pi e)} * \sigma_Y) \qquad (D.6)$$

(*e* denotes Euler's constant 2.718…) and is determined solely by the variance σ^2_Y. One can also estimate the entropy of noise, for example the quantization noise of an ADC or the amplifier noise.

Thirdly, the entropy difference between a signal and noise is determined by the *SNR* (2.8) and vice versa. A larger value of the entropy difference means a better *SNR* and a better resolution of the signal. Assuming that signal and noise are uncorrelated Gaussian processes #which define the signal-to-noise ratio *SNR* and an entropy difference ΔI of

$$SNR = \sigma^2_Y / \sigma^2_N \qquad (D.7)$$

$$\Delta I = 0.5 * log_2 (SNR+1) \qquad (D.8)$$

Hence, the entropy difference ΔI between signal *Y* and noise *N* is determined by the *SNR*.

E. Data

Data set D1: Several hundreds of polygraphic all-night data from healthy subjects and patients according to the protocol of the SIESTA project (Dorffner 1998) with sixteen channels (6+1 EEG, 2 EOG, 2 EMG, 1 ECG, 3 respiration and 1 oxygen saturation (SaO2) -channels) were recorded. Sleep recorders from the following providers were used: Fa. Jaeger (SleepLap 1000P), Nihon Kohden/DeltaMed, Walter Graphtek (PL-EEG), Flaga (EMBLA) and Siemens. Sampling rates of 1, 8, 16, 20, 25, 100, 200, 256 and 400 Hz were used for the various channels. The EEG was sampled with 100, 200 or 256Hz. The sampling rates and filter settings were stored in the headers of the data files for further evaluation. The data set was used to perform the histogram and entropy analysis (chapter 2, Schlögl et al. 1999). If not stated otherwise, the #analysis results of the sleep EEG were obtained from this database.

Data set D2: The EEG data stem from three subjects (E08, E10, E15) who displayed pronounced 10 Hz and beta-band changes in response to movement. The data were extracted from a series of experiments which were performed in twelve right-handed subjects (details see Stancak and Pfurtscheller, 1996). The task of the subjects was to move the right (left) index finger with 10-12 second intervals. The movement consisted of a rapid finger extension and flexion. The subjects were trained to hold the movement time in the range of 0.13 to 0.21 sec which ensured that the movements were ballistic. About 80 movements were obtained for the left and right finger.

Artifacts were rejected visually; the number of artifact-free trials is shown in (see Table 5) . One trial consists of 5 pre-movement and 4 post-movement seconds. Each trial comprises two signals corresponding to electrodes C3 and C4 (10-20 system). The EEG signals represent reference-free data that were obtained by the Laplacian operator method. Since the sampling rate was 128 Hz, each trial contains 1152 x 2 values. The files have a BKR format, which is used at the Department of Medical Informatics. The data were triggered according to the onset of EMG burst; this is at time 5.0 sec of each trial. This data set was used for the comparison of different EEG parameters and different classification methods based on single trials.

Table 5: Data sets from 3 three subjects performing left and right hand movement.

Subject	right hand number of trials	left hand number of trials	trivial error
E08	50	41	45 %
E10	42	57	42 %
E15	35	53	40 %

Data set D3: This data set stems from the BCI4c study. The experiment is described in Pfurtscheller et al. (1997, 1998) and Schlögl et al. (1997a,b). Selected sessions of 4 subjects (subject f3, session 10; f5, session 6; f7, session 6; g3, session 7) with 80-79, 78-78, 80-80 and 76-75 (L)eft-(R)ight trials, respectively, were investigated. The data was used to show the time courses of the classification error , the time-varying distance, and the amount of information gained by single trial EEG data. The data was used in off-line analysis for inventing the AAR+LDA classification system which was applied in succeeding BCI-experiments with continuous feedback (paradigm BCI4e, Schlögl et al. 1997c, Neuper et al. 1999, Guger et al. 1999) .

Data set D4: This data set is part of the data set D1; fifteen polygraphic sleep recordings were selected randomly. Experts scored artifacts from 7 EEG channels of 90 minutes segments in each. Nine types of artifacts (EOG, ECG, EMG, movement, failing electrodes, sweat, 50Hz, breathing and pulse) were distinguished and scored on a 1 second resolution (in total 563 192 1s-epochs). (Nearly) no pulse and breathing artifacts were identified. These data were used for analyzing the artifact processing ability of different processing methods. The data is described in more detail in (Schlögl, et al. 1999b,e,f). The data was used to demonstrate the artifact processing in the sleep EEG and to calculate the histograms from the all-night sleep recordings.

Data set D5: The sets that were used for evaluation of various AAR estimations the algorithms and to find the optimal update coefficient. For each data set, a different model order was used (column 3). The selected model orders range from an AR(6) up to an ARMA(20,5) model. Different update coefficients *UC* (0 and in the range between 10^{-1} and 10^{-9}) were applied.

Table 6: Description of five real-world data sets used for the evaluation. The data is displayed in Fig. 22.

No	duration Fs [Hz] S.D. ($MSY^{1/2}$) samples	Model	Description
S2	1000s Fs = 100Hz S.D. = 4.025 [units] 100001 samples	AR(10)	Sleep EEG (channel C3-A2)
S3	90min Fs = 256Hz 40.3 µV 1 328400 samples	ARMA(20,5)	Sleep EEG, Fp1-A2, 2nd 90-min epoch (1h28m48s-2h58m48s a.m.) with saturation artifacts. 29053 (2.1%) of the samples reach the upper or lower saturation limit.
S4	8.1 min Fs=128Hz S.D. = 0.112 x 50µV 62279 samples	AR(6)	Bipolar recording from 2.5cm anterior to 2.5cm posterior over the left sensori-motor area, position C3, The subject was asked to perform an imaginary hand movement according to a cue on a computer screen in approx. 10s intervals
S5	836s S.D.=1.232 852 heart cycles	AR(12)	852 heart cycles (R-R intervals), the mean (66.9/min) was removed.
S8	90min 256Hz 165.2 1382400 samples	AR(15)	Sleep EEG, electrode position C3-A1, 2nd 90 min epoch, 1h31m33s - 3h01m33s a.m.,

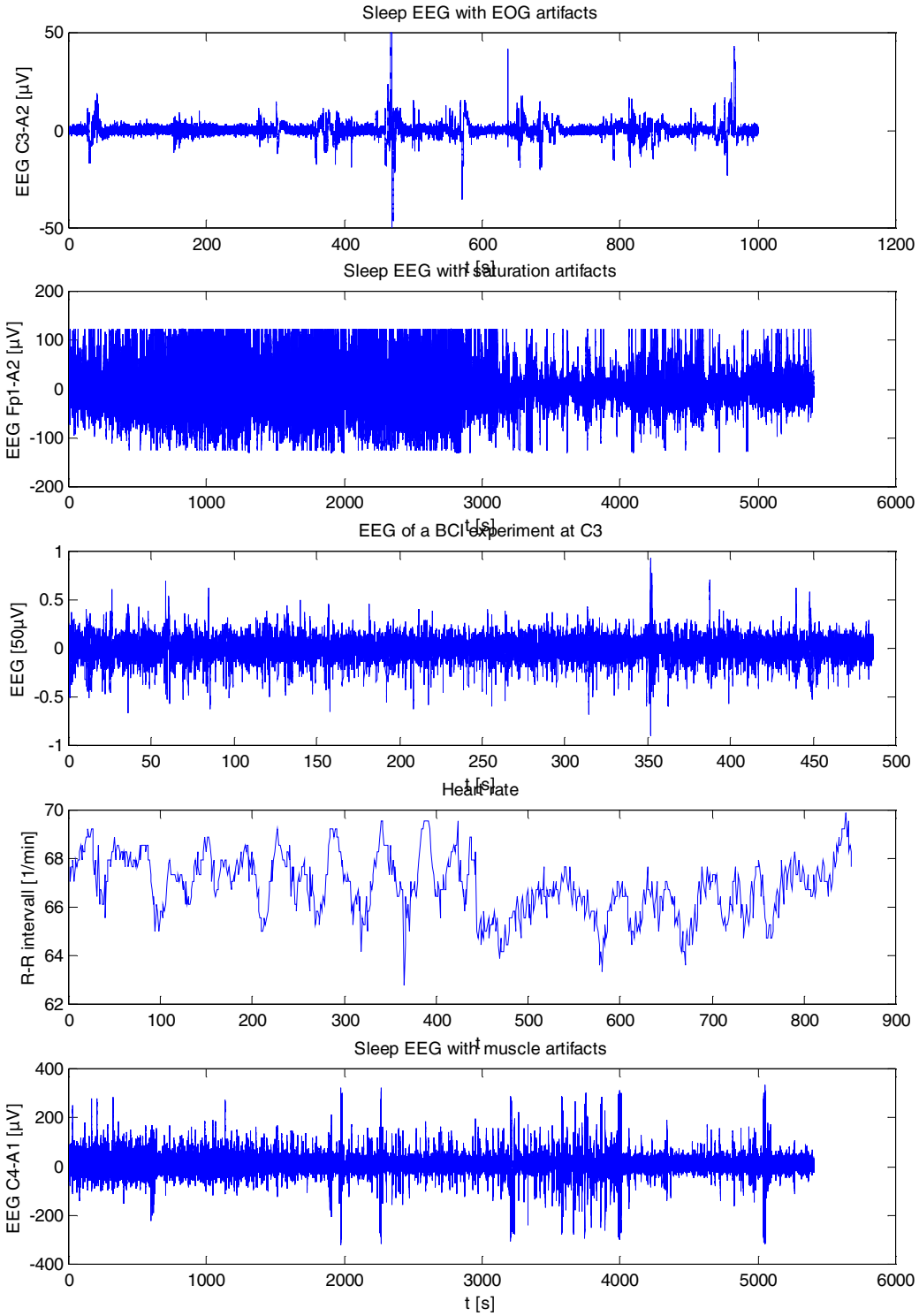


Figure 22: Five real word data sets. The first, second and fifth are sleep EEG recordings with eye movement artifact, muscle artifact, amplifier saturation artifacts, respectively. The third set is the EEG recording of a BCI experiment, and the forth set are the R-R intervals of heartbeats. A more detailed description is provided in Table 6. The data sets were selected from different biomedical research projects. The data are all real world data; no simulated data was used.

F. *List of figures*

G. Abbreviations

AAR	adaptive autoregressive
AIF	adaptive inverse filtering
AR	autoregressive
BKR	EEG data format from the Department for Medical Informatics, University of Technology Graz
EDF	European Data Format for Biosignals (Kemp et al. 1992)
KFR	Kalman filtering
LDA	Linear discriminant analysis
LMS	least mean squares
MEM	maximum entropy method
MSE	mean squared error
MSY	mean squared signal Y
RAR	recursive AR technique
REV	relative error variance
RLS	recursive least squares
R&K	Rechtschaffen and Kales (rules for sleep scoring)
STA	single trial analysis
TSD	time-varying signed distance

H. Index

T

U

W