Biomed. Technik
42 (1997), 162-167

A. Schlögl[1]
D. Flotzinger[2]
G. Pfurtscheller[2]

Adaptive Autoregressive Modeling used for Single-trial EEG Classification

Verwendung eines Adaptiven Autoregressiven Modells für die Klassifikation von Einzeltrial-EEG-Daten

[1]*Ludwig Boltzmann Institute for Medical Informatics and Neuroinformatics*
[2]*Department of Medical Informatics, Institute for Biomedical Engineering, University of Technology, Graz*

Key words: Adaptive Autoregressive model – single trial EEG analysis – Kalman filtering – RLS and LMS algorithm – event-related EEG

An adaptive autoregressive (AAR) model is used for analyzing event-related EEG changes. Such an AAR model is applied to single EEG trials of three subjects, recorded over both sensorimotor areas during imagination of left and right hand movements. It is found that discrimination between both types of motor-imagery is possible using linear discriminant analysis, but the time point for optimal classification is different in each subject.
For the estimation of the AAR parameters, the Least-mean-squares and the Recursive-least-squares algorithms are compared. In both methods, the update coefficient plays a key role: it determines the adaptation ratio as well as the estimation accuracy. A new method, based on minimizing the prediction error, is introduced for determining the update coefficient.

Schlüsselwörter: Adaptives Autoregressives Modell, Einzeltrial-EEG-Analyse, Kalman-Filterung, RLS- und LMS-Algorithmus, ereignisbezogene EEG

Ein Adaptives Autoregressives (AAR-)Modell wird zur Analyse von ereignisbezogenen Veränderungen im EEG verwendet. Das AAR-Modell wird auf nicht-gemittelte EEG-Daten, die von drei Personen während der Vorstellung von linker und rechter Handbewegung über dem sensomotorischen Kortex aufgezeichnet wurden, angewandt. Die beiden Arten von Bewegungsvorstellung können mit einer linearen Diskriminanzanalyse unterschieden werden, wobei jedoch der optimale Zeitpunkt für die Klassifikation von Person zu Person variiert.
Zur Schätzung der AAR-Parameter wurden der Least-mean-squares- und der Recursive-least-squares-Algorithmus verglichen. In beiden Methoden spielt der Adaptionskoeffizient eine wichtige Rolle: Er bestimmt sowohl die Adaptionsrate als auch die Schätzgenauigkeit. Eine neue Methode zur Bestimmung des Adaptionskoeffizienten, welche auf dem Minimieren des Vorhersagefehlers basiert, wird vorgestellt.

Introduction

Single trial analysis is a very important task in EEG analysis, yet it is also very difficult due to EEG's high intra- and inter-subject variability. Such an analysis is important for automatic sleep staging [8] as well as for the realization of EEG-based communication systems [14, 3, 6]. Furthermore, single-trial EEG analysis is important for studying sensory, motor and cognitive processing within an event-related paradigm.

In many methods of analysis, segmentation of the data is necessary, where stationary segments are assumed (e.g. calculation of frequency spectra, AR parameters [4]). Here, an adaptive (i.e. time-varying) autoregressive model (AAR) is used for analyzing event-related EEG. The estimation methods are well-established and have been developed in linear estimation theory (originally applied on orbit determination of planets and satellites), in adaptive filtering algorithms and in adaptive signal processing tasks [2].

The goals of this paper are:

- to introduce an AAR modeling technique for the analysis of single trial event-related EEG changes and to compare two estimation algorithms,
- to demonstrate selecting of control parameters, with special emphasis on determining the update coefficient,
- to investigate the optimal time point for discrimination of EEG data recorded during two different brain states.

For the investigation, a large number of single EEG trials from 3 subjects was available. Those subjects had participated in several sessions of Brain-Computer-Interface (BCI) experiments, performing imagination of left and right hand movement [6].

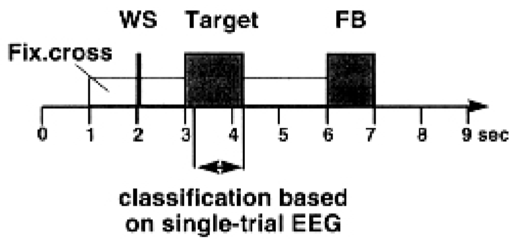

Figure 1. Experimental paradigm: at second 2, the subject received an auditory warning stimulus („WS"); after presentation of a visual cue at second 3 („target"), the subject performed an imaginary left or right hand movement; from second 3.25 to 4.25 the power of two selected frequency bands at 2 EEG channels was used for on-line classification; the feedback („FB") was presented visually from second 6 to 7. Throughout the trial a fixation cross was visible on the monitor.

Method and material

Subjects and experimental paradigm

Data from 3 subjects (F3, F5 and F7), who took part in a Brain-Computer-Interface (BCI) experiment, were available. The subjects were asked to imagine left or right hand movements in response to a visual cue stimulus in the form of an arrow on a computer screen (see Fig. 1). Each subject participated in 3–4 training sessions during which EEG was recorded to obtain subject-specific examples of left and right hand movement imagery patterns. EEG was recorded from two bipolar EEG channels over the left and right sensorimotor hand areas (around C3 and C4 of the international 10–20 system) and was sampled at 128 Hz. Each session consisted of 160 trials; after visual artifact rejection at least 143 trials remained. For the BCI experiments, subject-specific frequency bands were defined according to the data and a classifier was constructed which was applied in further sessions to give feedback. In these sessions, the EEG of a predefined 1-second time window (3.25–4.25 s, see Fig. 1) was classified online and the classification result was presented in the interval from second 6 to 7. The on-line classification accuracy varied between 64 % and 86 %. For a more detailed description of the data, the reader is referred to Pfurtscheller et al. [6].

The Adaptive Autoregressive model

The AAR model can be used to describe time-variations in the signal characteristics: the attribute „adaptive" means that the AR parameters are not constant but rather time-varying. Just like an AR model can be used to describe stationary EEG, event-related EEG can be characterized by an AAR model.
An AAR model describes the signal Y_t in the following form:

$$Y_t = a_{1,t}Y_{t-1} + a_{2,t}Y_{t-2} + \dots + a_{p,t}Y_{t-p} + E_t \tag{1}$$

where, in the ideal case, E_t is a purely random or white noise process with zero mean and variance σ^2_E. The difference to an AR model is that the parameters $a_{1,t} \dots a_{p,t}$ can vary with time. However, it is assumed that the parameters just change slowly. For a detailed discussion of non-stationary time series see Priestley [7].

AR models do not model any trend in the signal, i.e. the mean value of the signal has to be zero. In AAR models, the model parameters represent the signal characteristics only of a short period of time. A Blackman window with a width of 128 samples was used to remove the DC component from the signal; this procedure is equivalent to high-pass filtering with a cut-off frequency of approximately 2 Hz.

A rough estimate of the model order can be derived from the number of peaks in the spectrum. For each peak, two AR parameters are needed. Examination of EEG spectra shows a prevalent existence of 2–3 peaks, giving an approximate model order of 6. Since the DC component of the signal has been removed, no additional coefficient is required.

Estimation of AAR parameters

For the estimation of the AAR parameters, the Least-mean-square (LMS) [2, 12] and the Recursive least-square (RLS) [2] methods can be used: the RLS algorithm is derived from Kalman filtering for an AR model; LMS has been applied to biosignals e.g. by Schack et al. [9] for time-varying ARMA-models.

For the following sections, the AR parameters and the past p samples of the time series are defined as vectors:

$$\mathbf{a}_t = [\, a_{1,t} \dots a_{p,t}]^T \tag{2}$$

$$\mathbf{Y}_{t-1} = [Y_{t-1} \dots Y_{t-p}]^T \tag{3}$$

where p is the order of the autoregressive model and ^{T} denotes the transpose of the vector; bold letters indicate vectors.

As can be seen in equation (1), the prediction error E_t, also called the error process, depends on past values of $\mathbf{a}_i$ and $\mathbf{Y}_{i-1}$ with $i \le t$ only. Furthermore, we can utilize the fact that with a smaller error process, the EEG signal is described more accurately by the AAR model. Thus we define the Mean Square Error (MSE)

$$MSE = N^{-1} \sum_{t=1}^{N} E_t^{\,2} \tag{4}$$

and the total power of the EEG

$$MSY = N^{-1} \sum_{t=1}^{N} Y_t^{\,2} \tag{5}$$

Furthermore, we define a ratio between the mean square error and the total power of the signal:

$$REV = MSE/MSY \tag{6}$$

The relative error variance (REV) is a measure for the goodness-of-fit, i.e. how well the AAR model describes the EEG signal Y_t. This measure will be used for optimizing the adaptation ratio.

The Least Mean Square Algorithm (LMS)

The LMS estimation method can be characterized by the following two formulas:

$$E_t = Y_t - a_{1,t-1} Y_{t-1} - \dots - a_{p,t-1} Y_{t-p} \tag{7}$$

$$a_{i,t} = a_{i,t-1} + c E_t Y_{t-i}; \quad i = 1 \dots p \tag{8}$$

with E_t being the prediction error and c an update ratio or the „... *gain constant that regulates the speed and stability of adaptation*" ([12], p.100); Haykin [2] called it the *step size parameter*. We can denote

$$c = UC/MSY \tag{9}$$

where UC is called the update coefficient. This has the advantage that the adaptation ratio can be defined independently of the signal power, which is equivalent to normalizing the signal to a power of unity.

The choice of an appropriate UC is essential for AAR models. A value between 0 and 2 has been recommended by Widrow and Stearns ([12], p. 106). However this suggestion is far too rough, at least the order of magnitude should be fixed. It is one aim of this paper to determine an appropriate UC.

Another important issue is the initial values a_0. Two possibilities, initialization with zero and with average values (estimated by Yule-Walker equation, [11]), will be examined (see Table 1).

Table 1. Initialization methods: (i) with zero and (ii) with average AR parameters, the latter were calculated from the autocovariance function of the whole signal (mean autocovariance function over all trials) with the Yule-Walker method. $\mathbf{I}_{p \times p}$ denotes a p×p identity matrix.

Algorithm	Initialization
LMS04	$\mathbf{a}_0 = [a_{1,0} \dots a_{p,0}]^T = [0 \dots 0]^T$
LMS03	$\mathbf{a}_0 = [a_{1,0} \dots a_{p,0}]^T$ estimated with Yule-Walker method
RLS04	$\mathbf{a}_0 = [a_{1,0} \dots a_{p,0}]^T = [0 \dots 0]^T$ A0 = Ip¥p
RLS05	$\mathbf{a}_0 = [a_{1,0} \dots a_{p,0}]^T$ estimated with Yule-Walker method
	$\mathbf{A}_0 = 0.1 * \mathbf{I}_{p \times p}$

Kalman filtering or Recursive Least Squares Algorithm (RSL)

The RLS parameter estimation method is characterized by the following set of equations:

$$E_t = Y_t - \mathbf{a}_{t-1}{}^T \mathbf{Y}_{t-1} \tag{10}$$

$$\mathbf{r}_t = \lambda^{-1} \mathbf{A}_{t-1} \mathbf{Y}_{t-1} \tag{11}$$

$$\mathbf{k}_t = \mathbf{r}_t / (\mathbf{Y}_{t-1}{}^T \mathbf{r}_t + 1) \tag{12}$$

$$\mathbf{a}_t = \mathbf{a}_{t-1} + \mathbf{k}_t E_t \tag{13}$$

$$\mathbf{A}_t = \lambda^{-1} \mathbf{A}_{t-1} - \mathbf{k}_t \mathbf{r}_t^T \tag{14}$$

with, again, E_t being the prediction error and $\mathbf{k}_t$ the Kalman Gain Vector. The term λ^{-1}, a value slightly larger than one, has a similar function as the update coefficient UC in the LMS algorithm: it controls the adaptation ratio. A UC with a similar order of magnitude as for LMS is obtained with the following transformation:

$$\lambda^{-1} = 1/(1 - UC) \tag{15}$$

or

$$UC = 1 - \lambda \tag{16}$$

Classification System

For single trial classification purposes, the current AAR model parameters can be read at a specific time point within each trial. By concatenating the model parameters estimated for each of the available EEG channels (here, C3 and C4), a high-dimensional data vector per trial is formed.

The classifier used is Fisher's linear discriminant analysis (LDA) [1]. This method divides the 12-dimensional feature space into two half-spaces (one per category) such that the classification error rate on the training data is minimized. To achieve an estimate of the generalization ability of the LDA classifier, 10-times 10-fold cross-validation was performed on each data set (i.e. each session of each subject). For more details on cross-validation see Bishop [1].

All computations were performed with a precision of 8 byte floating point numbers, using Matlab®4.2 from Mathworks Inc. on an Alpha Workstation.

Results

Estimation of the AAR parameters

Selection of Update Coefficient UC

The adaptation ratio plays an essential role in AAR modeling. Here the optimal update coefficient was found by minimizing MSE (4), which is eqivalent to minimizing REV (6). First the order of magnitude was determined, then the interesting range was studied in detail. For that purpose, update coefficients from the set {0.0010; 0.0013; 0.0018; 0.0024; 0.0032; 0.0042; 0.0056; 0.0075; 0.0100; 0.0133; 0.0178; 0.0237; 0.0316} were investigated.

As can be seen in Fig. 2, there is one UC per data set where REV is minimal. This is the optimal update coefficient UC_{opt} of the AAR model which describes the EEG data best.

The method of minimizing the MSE to find the optimal UC stems from the idea that the EEG can best be described by an AAR model with a certain adaptation ratio. This idea is supported by the fact that in all in-

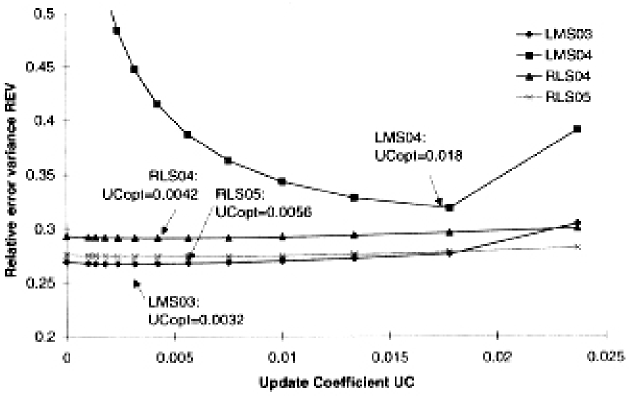

Figure 2. The relative error variance REV of the signal depending on the update coefficient UC. The different curves show the REV of F35-1 (subject F3, session 5, channel C3) for the estimation methods LMS04 and RLS04 (zero initialization) and LMS03 and RLS05 (initialization with average AR parameters). The UC where REV is minimal, is indicated by an arrow for each curve.

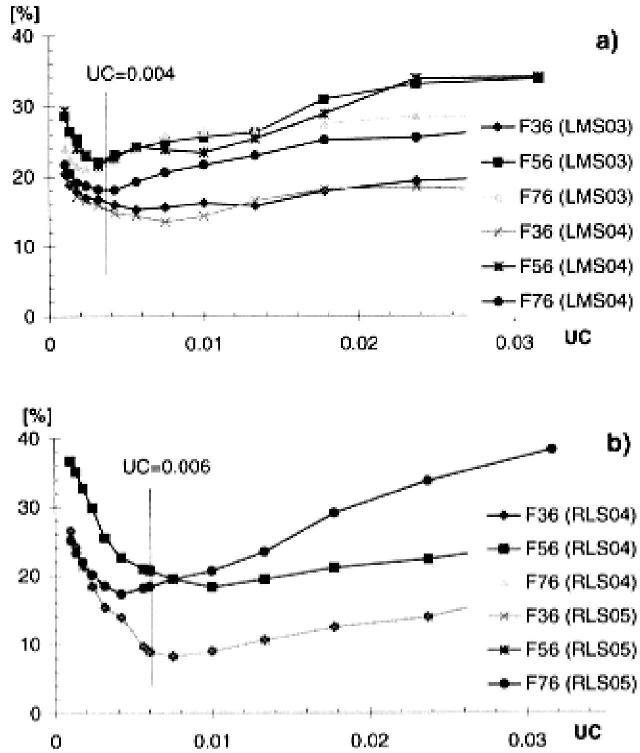

Figure 3. Error rate (10-times 10-fold cross-validation of a Linear Discriminant) depending on the update ratio UC for the classification time second 5 for different data sets; a) results of LMS03 and LMS04 on the 6[th] session for subjects F3, F5 and F7; b) same as a) but for RLS algorithms.

vestigated cases, only one minimum MSE was found. Naturally, UC_{opt} depends on many factors, e.g. initialization, model order, speed of event-related variation of the underlying EEG data, the sampling rate, etc. However, with the method provided here, UC_{opt} can be determined more accurately than discussed in the literature.[2, 12]

In Table 2, the optimal UC for twelve signals (EEG signals recorded at electrodes C3 and C4 from the 5[th] and 6[th] session of subjects F3, F5 and F7) are shown for the LMS and RLS algorithms. From these results, an UC of 0.004 (LMS) and 0.006 (RLS) was chosen for the investigation of data from other sessions.

Table 2. Update coefficient obtained by minimizing the relative error variance for LMS and RLS algorithms with different initialization and for different signals. SIGNAL indicates the EEG-channel of different recordings, e.g. F35-1 denotes subject F3, session 5 and channel 1 (C3) whereas F35-2 denotes the same for channel 2 (C4). The UC chosen for further analysis is given for each algorithm.

SIGNAL	LMS04	LMS03	RLS04	RLS05
f35-1	0.018	0.0032	0.0042	0.0056
f35-2	0.018	0.0032	0.0042	0.0042
f36-1	0.018	0.0024	0.0042	0.0042
f36-2	0.024	0.0024	0.0042	0.0042
f55-1	0.018	0.0032	0.0056	0.0056
f55-2	0.018	0.0042	0.0075	0.0075
f56-1	0.018	0.0032	0.0075	0.0056
f56-2	0.010	0.0056	0.0100	0.0100
f75-1	0.032	0.0042	0.0056	0.0056
f75-2	0.024	0.0042	0.0075	0.0056
f76-1	0.024	0.0024	0.0032	0.0032
f76-2	0.024	0.0032	0.0042	0.0042
chosen	–	**0.004**	**0.006**	**0.006**

Initialization of LMS and RLS algorithm

In Table 2 it can be seen that in the case of the RLS method, UC_{opt} hardly depends on the initialization. In comparison, the variation across subjects (F3, F5 and F7) and across sessions (F75 and F76) is much larger Therefore we can say that the kind of initialization is of minor importance in the case of the RLS algorithm.

This is not the case with the LMS algorithm; zero initialization (LMS04, see Fig. 2) requires a much bigger UC_{opt} (0.018 instead of 0.0042) than initialization with average AR parameters (LMS03). The value of UC_{opt} found with LMS04 is near the limit for consistent estimates: if UC were chosen only slightly larger, the error process would increase substantially and the AAR parameters would not be reliably estimated. For some other recording, an UC of 0.018 would already be too large. Consequently, we cannot use zero initialization for the LMS algorithm but have to use average initialization values to find a proper update coefficient.

Classification

Influence of the update coefficient

In Fig. 3a the error rates using the LMS algorithm for different initializations and update coefficients are shown. It can be seen that the optimum of UC = 0.004, found by minimizing MSE, has the same order of magnitude as the UC (0.0024 - 0.0075) that gives the best classification.

Furthermore, zero initialization (LMS04) seems to give as good results as LMS03. That is surprising because we found that the AAR parameters estimated with LMS04 do not describe the EEG signal very well (see Fig. 2, for the LMS04 algorithm REV is much higher than for LMS03). One explanation is that for the current classification task (imagined left vs. right hand movement) only the asymmetries between the two channels are evaluated.

It can be seen in Fig. 3b that the classification result using the RLS algorithm does not depend on the initialization method (graphs are overlapping). Furthermore, it can be seen that the optimal update coefficient of UC = 0.006 has the same order of magnitude as the UC with best classification (0.0042, 0.0075 and 0.001).

Summarizing, it can be stated that the values of UC_{opt} as determined by minimizing REV are also good choices for single trial classification.

Averaging (smoothing) of parameters

It is of interest whether smoothing of the AAR parameters, as done by Roberts and Tarassenko [8], has some advantage. Therefore, the average parameters over 2, 4, 8, 16, 32 and 128 samples were also used for classification. It was found that sometimes the error rate was lower while sometimes it was higher. In general, no advantage was found by averaging the estimated AAR parameters. Hence, smoothing of the AAR parameters seems unnecessary.

Comparing LMS and RLS estimates

It is known [2] that RLS converges faster and/or estimates the AAR parameters more accurately than LMS. For that reason it can be expected that the classification error is lower with the RLS algorithm. We compared the classification results of all sessions (5–11) for subjects F3, F5 and F7 at a fixed classification time point (second 5) and for a fixed UC_{opt} (RLS05 UC = 0.006; LMS04 UC = 0.004). The RLS algorithm was better in 17 (of 21) cases and gave, on average, 3.7 % better classification results.

Selection of Classification time

Fig. 4 displays the classification error for every 1/4 s between second 3.5 and 8. Note that the AAR parameters provide information about the signal characteristics with every sample. For each subject a different classification time course is obtained which can be explained by the different imagery strategies employed by subjects. For subjects F3, F5 and F7 the optimal classification time points are second 5.25, 6.0 and 4.75, respectively. Overall a classification time at second 5 seems reasonable for this paradigm.

Discussion

Classification of different EEG patterns recorded during different imagined movements is possible through applying non-linear classifiers as for example LVQ [3], but also through applying linear discrimination methods as used by Wolpaw [13]. Both groups use band power values as input for the classifier assuming stationarity over a small time window.

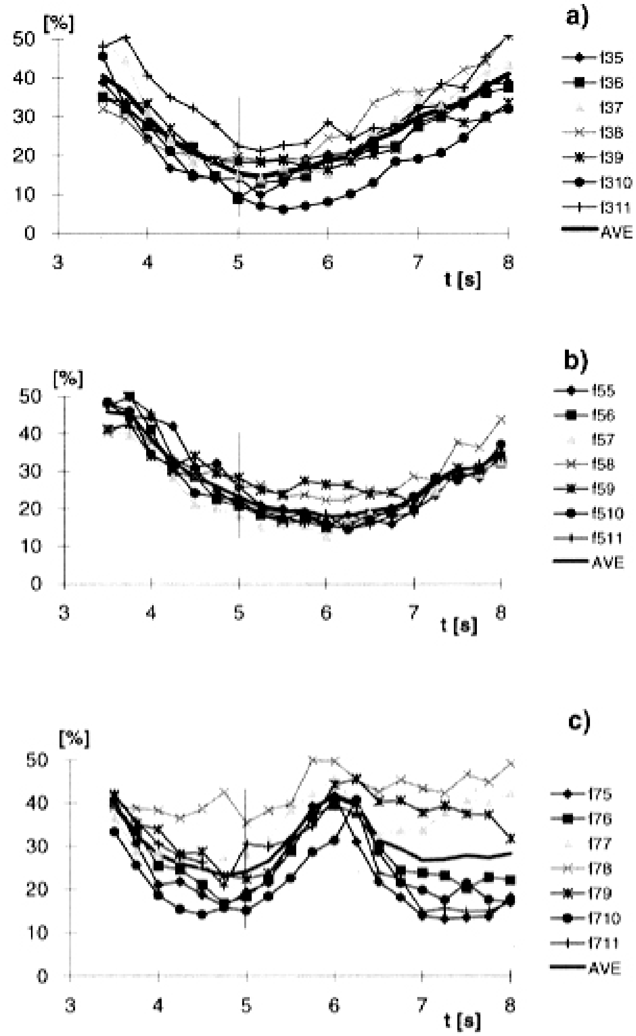

Figure 4. Error rate (10-times 10-fold cross-validation Linear Discriminant Analysis) over classification time for subjects F3, F5 and F7 and sessions 5–11. AVE denotes the average error rate across sessions. The AAR parameters of channel C3 and C4 were estimated with the RLS method with model order 6 and UC = 0.006.

In this paper, an adaptive AR model was used for single-trial analysis of event-related EEG recorded during imaginary hand movement. Analyses of data from a movement experiment have shown that for classification of single EEG trials, AAR parameters are features as good – or even superior – as traditionally used band-power values [10].

AAR parameters have several advantages compared to other EEG parameters:

- With AAR parameters it is possible to determine the optimal classification time point because the AAR parameters have already been computed for every time point.

- With AAR parameters no individual selection of frequency bands, as used in [5], is necessary .
- The estimation algorithms for AAR parameters are appropriate (i.e. fast enough) for on-line analysis and no segmentation is necessary.

Using an AAR model, several questions have to be addressed, e.g. the selection of the model order and the update coefficient, the initialization values (i.e. a priori knowledge about the signal), the elimination of the low frequency components and how to deal with artifacts in the EEG. The influence of the model order, the high-pass filtering and artifacts in the EEG have not been evaluated in this paper and are still open topics in AAR modeling.

A new method for determining UC has been introduced which is based on the principle of minimizing the mean squared error and the relative error variance. Hence, the time-variations of the event-related EEG are described by an AAR model with an optimal UC. The optimization can be interpreted as follows: UC determines the speed of adaptation; the larger UC, the faster the variation of the AAR parameters $\mathbf{a}_t$; but the faster the adaptation, the lower the accuracy of the estimated parameters, which means the model is badly estimated and the MSE increases. On the other hand if UC is zero, no adaptation takes place and the parameters are fixed to their initial values; one expects the MSE to be higher than in the case of modeling the event-related changes by varying parameters. Thus, the overall error should decrease if UC is greater than zero. This is always the case in non-stationary signals, even if the initial AR parameters are estimates of average values. In statistical terms, this is an optimization of bias and variance of the estimated parameters.

Furthermore, it has been shown that for UC obtained by minimized REV, the classification results are near the minimal error. Correspondingly, smoothing of the estimated AAR parameters has not improved the classification result notably.

The initialization of the RLS algorithm has a negligible influence on UC_{opt} as well as on the classification results. With LMS, the initialization has hardly any influence on the classification, but for determining the update coefficient, zero initialization gives an UC_{opt} that is too large; an initialization with some typical AR parameters is therefore recommended.

Summarizing, it can be stated that AAR parameter estimation offers an interesting opportunity for classification of single-trial data and can be used for the investigation of event-related EEG data.

Acknowledgments

Many thanks to Dr. Neuper and the staff of psychologists at the Department of Medical Informatics, University of Technology, Graz, for recording the data and performing artifact rejection. This work was supported by the Austrian Fonds zur Förderung der wissenschaftlichen Forschung, Project 11208, and by the Biomed Project PL931129 (ANNDEE).

References:

[1] Bishop, C. M.: Neural Networks for Pattern Recognition. Clarendon Press, Oxford, 1995.

[2] Haykin, S.: Adaptive Filter Theory. Prentice Hall, Englewood Cliffs, NJ, 1986.

[3] Kalcher, J.; D. Flotzinger, Ch. Neuper, S. Gölly, G. Pfurtscheller: 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 (1996), 382–388.

[4] Niedermeyer, E.; F. Lopes da Silva: Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Urban & Schwarzenberg, Baltimore, München, 1993.

[5] Pfurtscheller, G.; J. Kalcher, Ch. Neuper, D. Flotzinger, M. Pregenzer: On-line EEG classification during externally-paced hand movements using a neural network-based classifier. Electroenceph. clin. Neurophysiol. 99 (1996), 416–425

[6] Pfurtscheller G.; Ch. Neuper, D. Flotzinger, M. Pregenzer: EEG-based discrimination between imaginary right and left hand movement. Submitted to Electroenceph. clin. Neurophysiol., 1997.

[7] Priestley, M. B.: Spectral Analysis and Time Series. Academic Press, London, 1981.

[8] Roberts, S.; L. Tarassenko: Analysis of the sleep EEG using a multilayer network with spatial organisation. IEE Proceedings-F, 139/6 (1992), 420–425.

[9] Schack, B.; H. Witte, G. Grießbach: Parametrische Methoden der dynamischen Spektralanalyse und ihre Anwendung in der Biosignalanalyse. Biomedizinische Technik 38 (1993), 79–80.

[10] Schlögl, A.; B. Schack, G. Florian, K. Lugger, M. Pregenzer, G. Pfurtscheller: Classification of Single trial EEG: A comparison of different parameters. Proceedings of 3rd Hans Berger Congress, October 4–6, 1996, Jena.

[11] Wei, W.: Time series analysis; Univariate and multivariate methods. Addison Wesley, New York, 1990.

[12] Widrow, B.; S. D. Stearns: Adaptive Signal Processing. Prentice Hall, Englewood Cliffs, NJ, 1985.

[13] Wolpaw, J. R.; D. J. McFarland: Multichannel EEG-based brain-computer communication. Electroenceph. clin. Neurophysiol., 90 (1994), 444–449.

788

Address of correspondence:
Univ.-Prof. DI Dr. G. Pfurtscheller
Abteilung für Medizinische Informatik
Technische Universität Graz
Brockmanngasse 41,
A-8010 Graz, Austria