ELSEVIER

Clinical Neurophysiology xx (xxxx) 1–9

www.elsevier.com/locate/clinph

EEG filtering based on blind source separation (BSS) for early detection of Alzheimer's disease

Andrzej Cichocki[a,b,*], Sergei L. Shishkin[a], Toshimitsu Musha[c], Zbigniew Leonowicz[a,d], Takashi Asada[e], Takayoshi Kurachi[c]

[a]*Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan*
[b]*Warsaw University of Technology, Warsaw, Poland*
[c]*Brain Functions Laboratory Inc., KSP Building E211, Sakado, Takatsu Kawasaki, Kanagawa, 213-0012, Japan*
[d]*Wroclaw University of Technology, Wroclaw, Poland*
[e]*Department of Neuropsychiatry, Tsukuba University, Tennoudai, Tsukuba-shi, 305-8575, Japan*

Accepted 16 September 2004

Abstract

Objective: Development of an EEG preprocessing technique for improvement of detection of Alzheimer's disease (AD). The technique is based on filtering of EEG data using blind source separation (BSS) and projection of components which are possibly sensitive to cortical neuronal impairment found in early stages of AD.

Method: Artifact-free 20 s intervals of raw resting EEG recordings from 22 patients with Mild Cognitive Impairment (MCI) who later proceeded to AD and 38 age-matched normal controls were decomposed into spatio-temporally decorrelated components using BSS algorithm 'AMUSE'. Filtered EEG was obtained by back projection of components with the highest linear predictability. Relative power of filtered data in delta, theta, alpha1, alpha2, beta1, and beta 2 bands were processed with Linear Discriminant Analysis (LDA).

Results: Preprocessing improved the percentage of correctly classified patients and controls computed with jack-knifing cross-validation from 59 to 73% and from 76 to 84%, correspondingly.

Conclusions: The proposed approach can significantly improve the sensitivity and specificity of EEG based diagnosis.

Significance: Filtering based on BSS can improve the performance of the existing EEG approaches to early diagnosis of Alzheimer's disease. It may also have potential for improvement of EEG classification in other clinical areas or fundamental research. The developed method is quite general and flexible, allowing for various extensions and improvements.

Keywords: Alzheimer's disease; Diagnosis; EEG; Blind Source Separation; AMUSE; Filtering

1. Introduction

Alzheimer' s disease (AD) is one of the most frequent disorders among the elderly population (Jeong, 2004). Recent studies have demonstrated that AD has a presymptomatic phase, likely lasting years, during which neuronal degeneration is occurring but clinical symptoms not yet appear. This makes preclinical discrimination between people who will and will not ultimately develop AD critical for early treatment of the disease which could prevent or at least slow down the onset of clinical manifestations of disease (Blennow and Hampel, 2003; DeKosky and Marek, 2003; Rapoport, 2000; Wagner, 2000). Moreover, early diagnostic tools could significantly facilitate the development of drugs for the treatment at the early stage of AD: without preclinical diagnosis, many times more subjects (potential patients with huge percentage of those who actually would never develop AD) should be involved for testing of these drugs (DeKosky and Marek, 2003). A diagnostic method should be relatively inexpensive, to make possible screening of many individuals who are at risk of developing this dangerous disease

* Corresponding author. Tel.: +81 48 467 9668; fax: +81 48 467 9686.
E-mail address: cia@brain.riken.jp (A. Cichocki).

1388-2457/$30.00 © 2004 Published by Elsevier Ireland Ltd. on behalf of International Federation of Clinical Neurophysiology.
doi:10.1016/j.clinph.2004.09.017

(DeKosky and Marek, 2003). The electroencephalogram (EEG) is one of the most promising candidates to become such a method.

To date, many signal processing techniques were applied for revealing pathological changes in EEG associated with AD (see Jeong, 2004, for review). For example, combination of linear and nonlinear measures improved the classification accuracy of AD versus normal subjects up to 92% (Pritchard et al., 1994). Using principal component analysis (PCA) as a postprocessing tool for compressing linear and nonlinear EEG features over channels and age as a moderator variable in a study with rigorous validation procedure (jack-knifing), Besthorn et al. (1997) obtained 89% correct classification. However, high classification accuracy was obtained for patients who already developed serious cognitive impairment (e.g. Mini Mental State Examination (MMSE) score was 11.5 ± 7.9 in the study of Besthorn et al. (1997)).

Finding a method for identification of patients who have no clinical signs of AD at the moment of EEG registration but later progress to AD is the main challenge in this field. The studies of this kind are very rare. Huang et al. (2000) obtained 87% classification accuracy for discrimination between patients with mild cognitive impairment (MCI) who later progressed and not progressed to AD, however, without reporting the use of cross-validation. Musha and co-authors demonstrated, in a computer simulation, that local cortical neuronal impairment should lead to lower dipolarity (goodness-of-fit for dipole localizations) of alpha EEG frequency components (Hara et al., 1999), and then, based on these results, developed a technique for estimation of cortical impairment in AD using a single index of dipolarity (Musha et al., 2002). Alpha dipolarity was able to differentiate MCI patients who showed no clinical signs of AD at the time when EEG was recorded but developed AD later, as diagnosed in the follow-up, from normal controls with high probability; it also correlated with the degree of cortical neuronal impairment, estimated by SPECT (Musha et al., 2002).

However, in spite of all of the achievements made in the above cited studies, the problem of preclinical diagnosis of AD using EEG is not yet solved and further improvement of the methodology is necessary.

The main idea of this paper can be formulated as 'filtering based on Blind Source Separation (BSS)', that is, filtering of EEG by selection of most relevant components followed by reconstruction of the relevant part (subspace) of EEG signal using back projection of only these components. We propose a preprocessing technique based on this idea for improving EEG-based AD diagnosis (possibly useful also in other fields of EEG analysis). Its usefulness was evaluated in combination with standard procedures, namely the linear discriminant analysis (LDA) applied to spectral power in several frequency bands. To make comparison clear and fair, we used only most reliable but simple procedures. However, more sophisticated analysis based on recent advances in techniques for EEG processing and data classification may provide, in combination with proposed preprocessing, further significant improvement of early AD diagnosis, and some relevant emerging techniques will be mentioned in Discussion.

2. Methods

2.1. Blind source separation filtering for EEG classification

Intuitively, one can expect that some hidden components of such a complex signal like EEG can be more sensitive to Alzheimer's disease and the related disorders than others. These more sensitive components can be considered as useful 'signal', and the other components of EEG as 'noise' or 'unwanted signals'. Improving the 'signal-to-noise ratio' by filtering off the 'noise' could enhance the performance of subsequent feature extraction and data classification. Blind Source Separation (BSS) algorithms (see Cichocki and Amari, 2003, for extensive review) can be used for the purpose of such filtering.

BSS, in its application to EEG analysis, assume that EEG signal is composed of a finite number of components (signals from the brain and other sources), $s(t) = [s_1(t),\ldots,s_n(t)]T$. Here t is a discrete time index, n is the number of components and $[\ldots]^T$ means transpose of row vector. Components are mixed through unknown linear mixing process (described by $n \times n$ mixing matrix $\mathbf{A}$), and n sensors (EEG electrodes) record the mixed signals $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$. Each of the components may change in time, but has a fixed weight for each channel. BSS algorithm finds an unmixing (separating) $n \times n$ matrix $\mathbf{W}$ consisted of coefficients with which the electrode signals should be taken to form, by summation, the estimated components: $\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$. (In more general case, the number of components can be not equal to the number of sensors.) The entries of the estimated mixing matrix $\hat{\mathbf{A}} = \mathbf{W}^{-1}$ are components' weights in the mixing process; in other words, they indicate how strongly each electrode picks up each of individual components. *Back projection* of some selected components $\mathbf{x}_r(t) = \mathbf{W}^{-1}\mathbf{y}_r(t)$ (where $\mathbf{x}_r(t)$ is a vector of reconstructed sensor signals and $\mathbf{y}_r(t)$ is the vector obtained from the vector $\mathbf{y}(t)$ after removal of all the undesirable components (i.e. by replacing them with zeros)) allows us to filter the EEG data.

In strict sense, BSS means estimation of true (original) sources, though exactly the same procedure can be used for separation of two or more subspaces of the signal without estimation of true sources. One procedure currently becoming popular in EEG analysis is removing artifact-related BSS components and back projection of components originating from brain (e.g. Jung et al., 2000; Joyce et al., 2004; Vorobyov and Cichocki, 2002). In this procedure, components of brain origin are not required to be separated from each other exactly, because they are mixed again by

back projection after removing artifact-related components. But by the same procedure we can filter off the 'noise' also in wider sense, improving the relative amount of any types of useful information in the signal. Specifically, we can try to increase the relative amount of signals content related to AD (i.e. to improve signal to noise ratio—SNR).

Finding the rules or fundamental principles for identification of *relevant and irrelevant components* is critical for the proposed approach and, in general, may require extensive studies. In the case of removing artifact-related components, such components typically can be easily identified by visual inspection, but in more general case exact discrimination of relevant and non-relevant components is more difficult. In this paper we attempt to differentiate clusters or subspaces of components with similar properties or features. For the purposes of EEG classification the estimation of individual components corresponding to separate and meaningful brain sources is not required, unlike in other applications of BSS to EEG processing (including its most popular variant, Independent Component Analysis (ICA)). The use of clusters of components is especially beneficial when the data from different subjects are compared: similarity between individual components in different subjects is usually low, while subspaces formed by similar components are more likely to be sufficiently overlapped. Differentiation of subspaces with high and low amount of diagnostically useful information can be made easier if components are separated and sorted according to some criteria which, at least to some extent, correlate with the diagnostic value of components. BSS algorithm 'AMUSE', in our opinion, can be relevant for this task.

2.2. AMUSE algorithm and its properties

AMUSE (Cichocki and Amari, 2003; Szupiluk and Cichocki, 2001; Tong et al., 1991, 1993)) is a BSS algorithm which arranges components not only in the order of decreasing variance (that is typical for the use of singular value decomposition (SVD) which is implemented within the algorithm), but also in the order of their decreased linear predictability. Low values for both characteristics can be specific for many of EEG components related to high frequency artifacts, especially electromyographic signal (which cannot be sufficiently removed by usual filtering in frequency domain, see Goncharova et al., 2003). Thus, a first attempt of selection of diagnostically important components can be made by removing a range of components separated with AMUSE (below referred to as 'AMUSE components') with the lowest linear predictability. Automatic sorting of components by this algorithm makes it possible to do this simply by removing components with indices higher than some chosen value.

AMUSE algorithm belongs to the group of second-order-statistics spatio-temporal decorrelation (SOS-STD) BSS algorithms. It provides similar decomposition as the well known and popular SOBI algorithms (Belouchrani et al., 1997; Tang et al., 2002). AMUSE algorithm uses simple principles that the estimated components should be spatio-temporally decorrelated and be less complex (i.e. have better linear predictability) than any mixture of those sources. The components are ordered according to decreasing values of singular values of a time-delayed covariance matrix. As in Principal Component Analysis (PCA) and unlike in many ICA algorithms, all components estimated by AMUSE are uniquely defined (i.e. any run of algorithms on the same data will always produce the same components) and consistently ranked. Fig. 1 illustrates typical components obtained by decomposing EEG using AMUSE algorithm.

AMUSE algorithm can be considered as two consecutive PCAs: first, PCA is applied to input data; second, PCA (SVD) is applied to the time-delayed covariance matrix of the output of previous stage. In the first step standard or robust prewhitening (sphering) is applied as a linear transformation $\mathbf{z}(t)=\mathbf{Q}\mathbf{x}(t)$, where $\mathbf{Q}=\mathbf{R}_x^{-\frac{1}{2}}$ of the standard covariance matrix $\mathbf{R}_x=E\{\mathbf{x}(t)\mathbf{x}^T(t)\}$ and $\mathbf{x}(t)$ is a vector of observed data for time instant t. Next, SVD is applied to a time-delayed covariance matrix of pre-whitened data: $\mathbf{R}_z=E\{\mathbf{z}(t)\mathbf{z}^T(t-1)\}=\mathbf{U}\mathbf{S}\mathbf{V}^T$, where $\mathbf{S}$ is a diagonal matrix with decreasing singular values and $\mathbf{U}$, $\mathbf{V}$ are matrices of eigenvectors. Then, an unmixing matrix is estimated as $\mathbf{W}=\hat{\mathbf{A}}^{-1}=\mathbf{U}^T\mathbf{Q}$ or $\hat{\mathbf{A}}=\mathbf{Q}^T\mathbf{U}$.

AMUSE algorithm is much faster than the vast majority of BSS algorithms (its processing speed is mainly defined by the PCA processing within it) and is very easy to use, because no parameters are required. It is implemented as a part of package 'ICALAB for signal processing' (Cichocki et al., online) freely available online and can be called also from current version of EEGLAB toolbox (Delorme and Makeig, 2004) (which is freely available online at http://www.sccn.ucsd.edu/eeglab/) if both toolboxes are installed.

2.3. Subjects and EEG recording

We used EEG recordings collected in the previous study (Musha et al., 2002). In that study, patients who complained only for memory impairment, but had no apparent loss in general cognitive, behavioral, or functional status, were recruited. Fifty-three patients of this group met the following criteria for Mild Cognitive Impairment (MCI): MMSE score 24 or higher, Clinical Dementia Rating (CDR) scale score of 0.5 with memory performance less than one standard deviation below the normal reference (Wechsler Logical Memory Scale and Paired Associates Learning subtests, IV and VII, ≤ 9 (Wechsler, 1987), and/or ≤ 5 on the 30 min delayed recall of the Rey-Osterreith figure test (Hodges, 1993). These patients were followed clinically for 12–18 months. Twenty-five of them developed probable or possible AD according to NINDS-ADRDA criteria (McKhann et al., 1984). Normal age-matched controls were recruited from family members of the patients

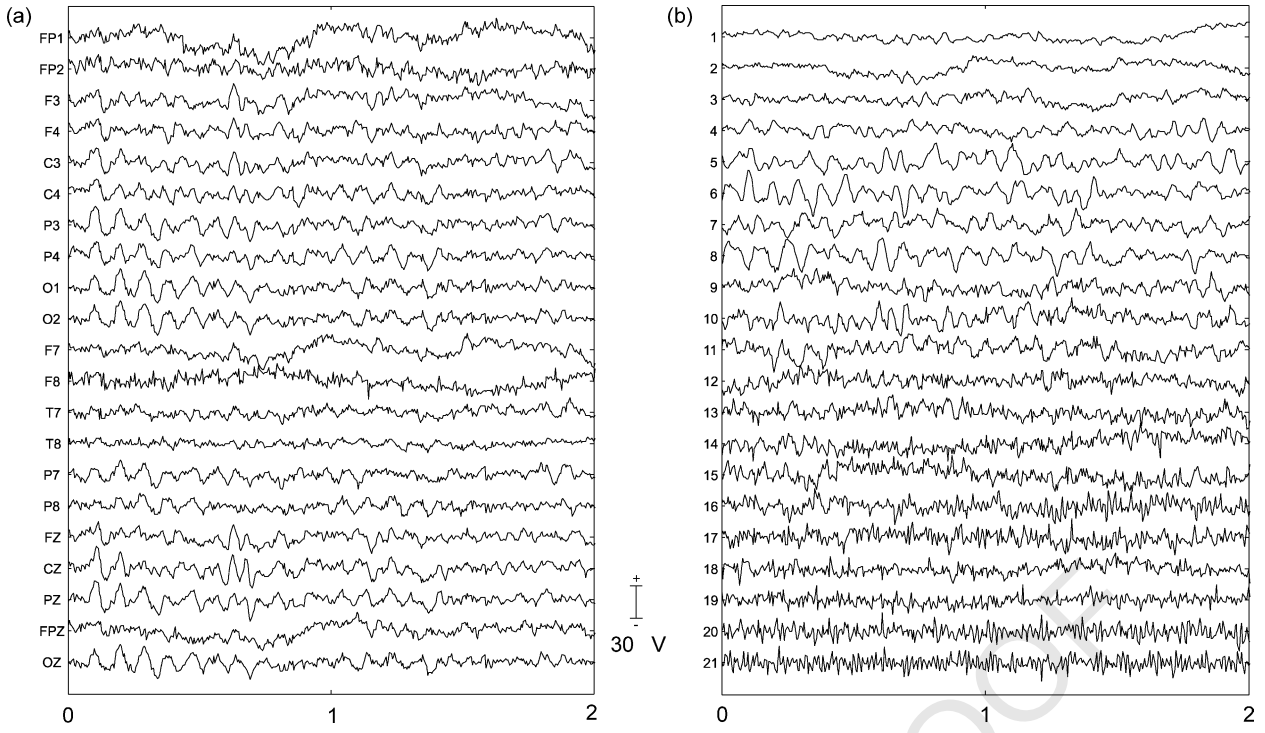

Fig. 1. Example of raw EEG (a) and its components separated with AMUSE algorithm (b) for a patient with MCI who later progressed to AD (MildAD002). AMUSE was applied to 20 s artifact-free interval of EEG, but only 2 s are shown. The scale for the components is arbitrary but linear. Note that the components are automatically ordered according to decreasing linear predictability (increasing complexity).

(mainly spouses) participated in the study as control group. Both patients and controls underwent general medical, neurological, psychiatric, and neuroimaging (SPECT, CT and MRI) investigation for making the diagnosis more precise.

EEG was recorded within 1 month after entering the study from all patients and controls, but only EEG recorded from the patients who progressed to AD ($n=25$; below: MCI group) and age-matched controls ($n=56$) was used for the analysis. No patient or control subject received psychotropic medication at the period when EEG was recorded. Mean MMSE score was 26 ± 1.8 in MCI group and 28.5 ± 1.6 in control group; age 71.9 ± 10.2 and 71.7 ± 8.3, respectively. EEG recording was done in an awake resting state with eyes closed, under vigilance control. Ag/AgCl electrodes (disks of diameter 8 mm) were placed on 21 sites according to 10–20 international system, with the reference electrode on the right ear-lobe. EEG was recorded with Biotop 6R12 (NEC San-ei, Tokyo, Japan) using analog filtering bandpass 0.5–250 Hz and sampling rate 200 Hz.

2.4. EEG data analysis

All computations were done using MATLAB (The MathWorks, Inc.). EEGLAB (Delorme and Makeig, 2004) was used for visual analysis of EEG recordings, and AMUSE algorithm implemented in ICALAB (Cichocki et al., online) was used for BSS processing.

Out of the EEG database described above (from the study of Musha et al., 2002), we selected 25 MCI patients (later progressed to AD) and 47 age-matched controls who had relatively little artifacts. Their EEGs were visually inspected by an experienced EEG researcher and the first continuous artifact-free 20 s interval of each recording was chosen for the analysis. Due to the lack of such interval in some recordings, the number of patients and controls were reduced to 22 and 38, correspondingly. The reason for selecting artifact-free intervals was that most of the artifacts produced amplifier blocking (saturation) due to its low amplitude range, which lead to strongly nonlinear distortion of the signal. AMUSE, as most of BSS methods, assumes a linear model of summation of source signals, and amplifier blocking should be excluded from the data.

Each EEG was decomposed into 21 decorrelated components by BSS algorithm AMUSE (see above). Some of the components (see Results) were selected for back projection, which formed preprocessed ('AMUSE filtered') EEG data. Spectral analysis based on Fast Fourier Transform (Welch method, Hanning 1 s window, 2 s epochs overlapped by 0.5 s) was applied to raw data, to the components and to the projections of selected components. Relative spectral powers were computed by dividing the power in delta (1.5–3.5 Hz), theta (3.5–7.5 Hz), alpha 1 (7.5–9.5 Hz), alpha 2 (9.5–12.5 Hz), beta 1 (12.5–17.5 Hz) and beta 2 (17.5–25 Hz) bands by the power in 1.5–25 Hz band. These values were normalized for better fitting

the normal distribution using the transformation $\ln(\mathbf{x}/(1-\mathbf{x}))$, where $\mathbf{x}$ is the relative spectral power (Gasser et al., 1982). To reduce the number of variables used for classification, we averaged band power values over all 21 channels.

Linear discriminant analysis (LDA) (using publicly available software for both linear classical and robust discriminant analysis, by Croux and Dehon, 2001) was used for discriminating MCI and control groups on the basis of log-transformed relative spectral power in the six frequency bands, averaged over channels. To improve validation of the classification results, discriminant analysis was applied in combination with jack-knifing, a procedure which typically produces lower discrimination rate than, e.g. cross-validation based on using part of a sample for learning and other part for classification, but is statistically more correct and enables increased reproducibility in other samples (Besthorn et al., 1997). Jack-knifing means that each case is classified using individual discriminant function trained with all cases except this one. Results of this procedure was used for computing sensitivity (the number of MCI subjects who were classified as MCI divided by the number of all subjects in MCI group) and specificity (the number of normal subjects who were classified as normal divided by number of all normal subjects).

3. Results

Averaged power spectra of each AMUSE component for patients and control subjects are presented in Fig. 2. As expected, components with lower indices (corresponding to lower linear predictability) had higher relative power at lower frequencies, while components with higher indices had higher relative power at highest frequencies. What is especially important is that the difference between patients and control subjects was clearer in the components with lower indices (i.e. components with highest linear predictability and highest variance of their projections). Thus, in further analysis we used combination of components with lowest indices.

To estimate how many components with highest linear predictability provides optimal classification rate, we applied LDA without jack-knifing (the latter requires much more computation time) to all projected components with indices from 1 to 2, from 1 to 3 and so on. Overall misclassification rate was computed each time by applying obtained discriminant function to the same 60 subjects (22 patients + 38 controls). Results are presented in Fig. 3. The best classification was obtained for projection of the first five components (with indices from 1 to 5); however, performance was also high when the number of components

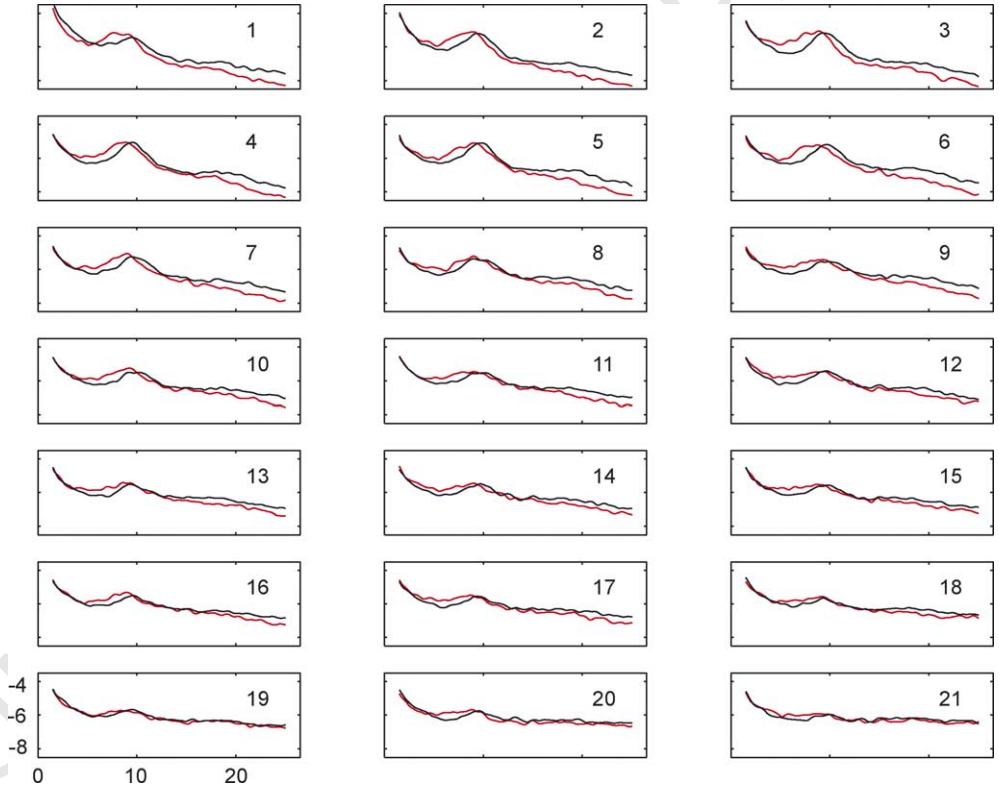

Fig. 2. Averaged power spectra of AMUSE components 1–21. x-axis: frequency, Hz. y-axis: transformed relative spectral power. Relative spectral power was obtained by dividing the absolute values in each frequency bin by total power in the range 1.5–25 Hz. Before averaging, the power values were normalized using transformation $\log(x/(1-x))$ (negative values appear because of this transformation). Red: MCI patients later progressed to AD ($n=22$). Black: control subjects ($n=38$).

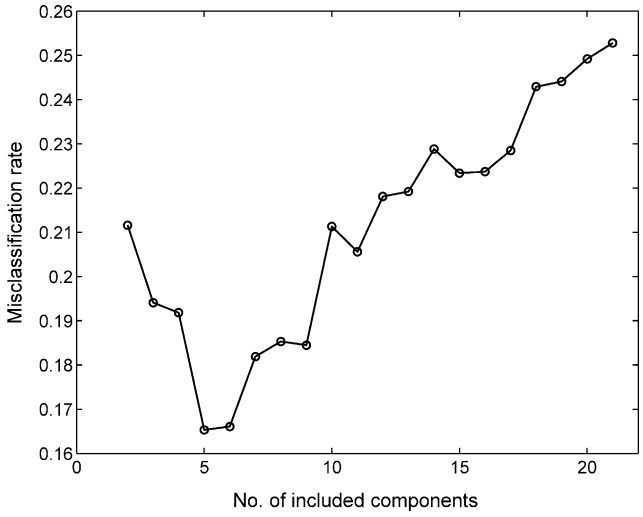

Fig. 3. LDA approximate (computed without cross-validation) misclassification rate for different number of projected components. Only components with highest linear predictability were used, thus, data points correspond to the following combinations of components: 1,2; 1–3; 1–4;…1–20, 1–21.

was in a rather wide range between 3 and 9. Thus, the method appeared to be robust in respect to the number of selected components.

Classification with jack-knifing procedure was applied to projections of several combinations of components, including 1–5 which appeared to be optimal according to Fig. 2. As follows from Table 1, results of classification were better if preprocessing included selection of AMUSE components with lower indices (1–5, 1–7, 1–10), comparing to raw data. When components with higher indices (6–21, 8–21, 11–21) were selected in preprocessing, the results were worse than in the case of raw data. Best results were obtained with components 1–5 and 1–7 (improvement by 14% over the raw data for classification of MCI and by 8% for control subjects), while components 11–21 gave the worst results. More detailed classification results for two combinations of components (1–5 and 1–10) and for the raw data, presented as Relative Operating Characteristic (ROC) curves in Fig. 4, confirm that use of components 1–10 only slightly improved the classification (Fig. 4(a)), while improvement of classification with components 1–5 over raw data was substantial (Fig. 4(b)). Best classification performance after preprocessing using 1–5 components was obtained in the range of approximately 0.6–0.8 for sensitivity and 0.7–0.9 for specificity. Selection of components with high indices was clearly not good for classification: for components 11–21 classification performance was almost at random level (Fig. 4(a)).

Table 1
Number of subjects who were correctly and incorrectly classified by discriminant analysis applied to relative power in six frequency bands after selection and back projection of certain AMUSE components (AMUSE filtering). Results were obtained using jack-knifing

AMUSE components selected in preprocessing	Misclassified		Correctly classified %		
	MCI $n=22$	Controls $n=38$	MCI $n=22$	Controls $n=38$	All $n=60$
No preprocessing	9	9	59	76	70
Components 1–5	**6**	**6**	**73**	**84**	**80**
Components 1–7	**6**	**6**	**73**	**84**	**80**
Components 1–10	6	9	73	76	75
Components 6–21	9	11	59	71	67
Components 8–21	9	11	59	71	67
Components 11–21	**12**	**12**	**45**	**68**	**60**

4. Discussion

With EEG preprocessing proposed in this paper, we obtained 80% rate of correct classification (Table 1) for MCI using only 20 s artifact-free interval of EEG recording from each patient or control subject. While groups of patients and controls were relatively small (22 and 38, correspondingly), it should be noted that the classification performance was estimated using the rigorous jack-knifing cross-validation procedure, which reduce the risk of overstating the results. The jack-knifing procedure was applied only to LDA but not to approximate optimization of the choice of components for back projection. Optimization of the choice of components was made for the whole dataset on the basis of components' spectra and preliminary run of LDA. Nevertheless, Figs. 2 and 3 suggest that the dependence of the difference between patients' and controls' spectra on component index and dependence of LDA results on the number of selected components were systematic; thus, it is unlikely that we simply picked up some random variations in LDA performance dependent on details of preprocessing and that improvement of LDA performance by preprocessing with the same parameters will be not reproducible in other groups of patients and controls.

The procedure of selection of artifact-free EEG intervals used in this study could introduce some bias in absolute values of discrimination results, because it was done by only one expert, and this expert did know to which group each EEG belongs. In fact, the proportion of the EEG recordings which were not analyzed due to the lack of a sufficiently long artifact-free interval was different in the groups of patients (12%) and controls (19%), and this difference was in the direction which can be expected if the criteria for selecting the analyzed interval were more strict for control group. This difference could be a result of random variations, and it should be noted that most of artifacts were easily identifiable (due to low amplifier range, any high amplitude artifact led to amplifier saturation), so it was rather unlikely that the subjective bias could strongly

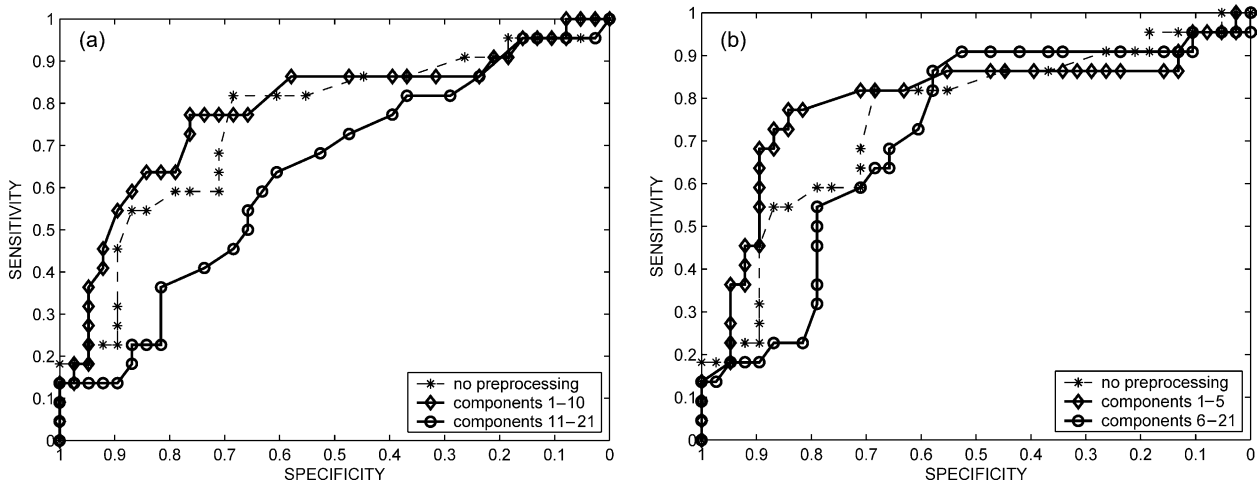

Fig. 4. Relative Operating Characteristic (ROC) curves obtained using jack-knifing for classification of MCI patients later progressed to AD ($n=22$) versus normal controls ($n=38$). LDA was applied to relative power in six EEG frequency bands. Comparison between data without preprocessing and data after selection and back projection of certain AMUSE components (AMUSE filtering). (a) Selection of first 10 components, compared to the rest of components and no preprocessing. (b) Selection of first five components, compared to the rest of components and no preprocessing.

influence the results. However, we cannot guarantee that the use of subjective criteria for selection of artifact free intervals did not affect classification results at all, and it is difficult to predict whether the obtained high values of specificity and sensitivity can be reproduced in other studies. We would like to emphasize, nevertheless, that our main claim is that the proposed preprocessing method increases the performance *relatively* to the level obtained without its use. This tendency could not be altered by subjective bias in search for artifact-free intervals.

We do not discuss here to which physiologically meaningful brain sources AMUSE components can correspond, because they can be a mixture of activity from many physical sources in the brain. This is clearly not critical for improving of EEG classification. The improvement of classification after AMUSE filtering comparing to non-preprocessed EEG data was probably caused by higher difference between patients and controls' spectra in the selected components than in the non-used (filtered off) components. Spectra computed for AMUSE components separated by BSS algorithm AMUSE (Fig. 2) demonstrate that the difference between patients and controls decreased with the index of component. Interestingly, this effect is visible at the same time in several frequency ranges: in theta range, where patients had an increase of relative power; in alpha range, where shift of the peak to slower frequencies was observed in patients; and in beta range, where power was lower for patients. All these differences in spectral power are typically found between AD patients and normal subjects. Components with the highest indices showed almost no difference between patients and controls, and it was not surprising that the performance of classification based on back projection of only these components was close to random level (Fig. 4(a), components 11–21). Thus, AMUSE components with higher indices can be considered as mainly representing 'noise' which makes difficult, in processing of raw EEG, to detect diagnostically important changes in characteristics of 'signal'. Note that 'signal' and 'noise' here are not labels for signal from brain sources and for artifacts: we refer to the 'signal' only as to diagnostically important (significant) part (subspace) of raw EEG signal, and to 'noise' as to the diagnostically not important part (non-significant subspace). AMUSE filtering, i.e. extraction of part of EEG reach with 'signal' by using only 'best' (here, most useful for diagnosis) components for back projection, naturally leads to the improvement of 'signal-to-noise ratio' and, as a result, to the improvement of EEG classification.

A BSS-based approach to improvement of signal-to-noise ratio in MEG signal by defining and removing noise subspace was already developed (Kawakatsu, 2003). More simple and already rather widely used technique is removing EEG and MEG artifact-related components with BSS using visual or automatic identification of such components one by one after decomposition (e.g. Jung et al., 2000). However, since in many kinds of EEG and MEG studies the goal is to extract the brain signal in possibly less distorted form, the existing techniques are limited to remove only such part of raw signal, which contain no or almost no components of brain origin but rather external artifacts and noise. In EEG classification tasks, such as diagnosis or Brain-Computer Interface (BCI), preserving the original signal is less important, noise can be defined not only as artifacts but also as any part of the signal which do not contribute to the difference between the classes of EEG which should be differentiated, and larger subspace with high percentage of such 'noise' can be removed. The existing techniques can only identify, by some a priori known characteristics, noise components (Barbati et al., 2004; Jung et al., 2000; Kawakatsu, 2003) and some very specific diagnostically important

components (epileptic spike separation: e.g. Kobayashi et al., 2002). Xu et al. (2004) recently suggested using a subspace approach for differentiating between task-related EEG patterns in BCI. They selected several ICA components related to P300 according to the a priori knowledge of P300 spatio-temporal pattern and reconstructed a clear P300 peak using back projection of these components. Like in the case of epileptic spikes, the components in this case were easily identifiable.

In a general case, however, significant and non-significant components are not easily identifiable. The task becomes especially challenging if EEG components from different subjects should be compared, because the sets of components produced by BSS in different subjects usually differ dramatically. In our approach, we rank components using some empirical rule, such as their linear predictability, and select those where difference between the pathological and normal EEG is most differentiated. This made possible to achieve substantial improvement in the discrimination between MCI patients who later progressed to AD and normal age-matched controls. To our best knowledge, no study till now investigated the application of BSS/ICA methods as preprocessing tools with possible application for AD diagnosis.

Dividing of components into two groups (or subspaces) as below or above some component's index (in the case of ranking) or using a threshold for some index computed for each component is not the only way. One may try to divide the sets of components at more than one level and, e.g. remove not only components with highest indices but also with the lowest indices. As one may suppose from Fig. 1(b) (example of individual data), the first two components could represent, to rather high extent, artifacts (roving eye movements). Fig. 2, however, shows that components #1 and #2 substantially differed between groups. We made an attempt to exclude 1 or 2 first components from the analysis and this, in fact, led to slightly lower discrimination results. However, it is possible that for other data (for example, including high amplitude low frequency artifacts) or other processing techniques dividing the set of components on more than one level could be beneficial.

Not only spectral but also other EEG features, such as measures of synchronization between channels, can be investigated for the possibility of improving contrast between pathological and normal data using the presented approach. Several studies indicated that synchronization between different brain areas is sensitive to AD. Such results were obtained for quite different techniques, including coherence (e.g. Adler et al., 2003; Jelic et al., 1996; Locatelli et al., 1998; Wada et al., 1998), mutual information (Jeong et al., 2001) and synchronization likelihood (a new measure combining estimation of linear and nonlinear coupling) (Stam et al., 2003). One may hypothesize that EEG components can be divided into two parts, one of which represents signal subspace with lower (or stronger) synchronization among some cortical areas in AD relative to normal EEG, and another one represents signal subspace which synchronization characteristics are not related to the disease. In this case, the general approach described in this paper also could appear to be useful. One may probably try to apply it also in the case of using nonlinear measures (see review in Jeong, 2004) or in combination with other advanced approaches.

There is obviously room for improvement and extension of the proposed method both in ranking and selection of optimal (significant) components, apparatus and post-processing to perform classification task. Especially, we can apply a wide variety of BSS methods, i.e. instead of the applied and investigated second order statistics spatio-temporal decorrelation, we can exploit other new types of BSS algorithms, such as higher order statistic ICA, sparse component analysis or smooth component analysis with a suitably ordered and ranked components. Furthermore, instead of standard LDA we can use more sensitive and robust methods, such as neural networks or support vector machine (SVM) classifiers. Classification can be probably strongly improved by supplementing the set of spectral power values which we used with much different indices, such as alpha dipolarity, a new index depending on prevalence local vs. distributed sources of EEG alpha activity, which was shown to be very sensitive to AD-related cortical impairment (Musha et al., 2002). Additional attractive but still open issue is that using the proposed approach, we can not only detect but also measure in consistent way the progression of AD and influence of medications. The proposed method can also be potentially useful and effective tool for differential diagnosis of AD from other types of dementia, and possibly for diagnosis of other diseases. Other areas of EEG analysis can be also possible field for the application of our preprocessing technique. For these purposes, more studies would be needed to asses of the impact of the proposed enhancement/filtering procedures on the EEG signal of interest.

5. Uncited references

Lindau et al. (2003), Wackermann (1996).

References

Adler G, Brassen S, Jajcevic A. EEG coherence in Alzheimer's dementia. J Neural Transm 2003;110(9):1051–8.

Barbati G, Porcaro C, Zappasodi F, Rossini PM, Tecchio F. Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals. Clin Neurophysiol 2004;115(5):1220–32.

Belouchrani A, Abed-Meraim K, Cardoso JF, Moulines E. A blind source separation technique using second order statistics. IEEE Trans Signal Process 1997;45(2):434–44.

Besthorn C, Zerfass R, Geiger-Kabisch C, Sattel H, Daniel S, Schreiter-Gasser U, et al. Discrimination of Alzheimer's disease and normal

aging by EEG data. Electroencephalogr Clin Neurophysiol 1997; 103(2):241–8.

Blennow K, Hampel H. CSF markers for incipient Alzheimer's disease. Lancet Neurol 2003;2(10):605–13.

Cichocki A, Amari S. Adaptive blind signal and image processing: learning algorithms and applications. New York, NY: Wiley; 2003.

Cichocki A, Amari S, Siwek K, Tanaka T, et al. ICALAB toolboxes. [Available online at http://www.bsp.brain.riken.jp/ICALAB]

Croux C, Dehon C. Software package for robust discriminant analysis; 2001. [http://www.econ.kuleuven.ac.be/public/NDBAE06/software/DA/matlab.htm]

DeKosky ST, Marek K. Looking backward to move forward: early detection of neurodegenerative disorders. Science 2003;302(5646): 830–4.

Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 2004;134(1):9–21.

Gasser T, Bacher P, Mocks J. Transformations towards the normal distribution of broad band spectral parameters of the EEG. Electroencephalogr Clin Neurophysiol 1982;53(1):119–24.

Goncharova II, McFarland DJ, Vaughan TM, Wolpaw JR. EMG contamination of EEG: spectral and topographical characteristics. Clin Neurophysiol 2003;114(9):1580–93.

Hara J, Shankle WR, Musha T. Cortical atrophy in Alzheimer's disease unmasks electrically silent sulci and lowers EEG dipolarity. IEEE Trans Biomed Eng 1999;46(8):905–10.

Hodges JR. Cognitive assessment for clinicians. Oxford: Oxford Medical Publications; 1993 p. 197–228.

Huang C, Wahlund L, Dierks T, Julin P, Winblad B, Jelic V. Discrimination of Alzheimer's disease and mild cognitive impairment by equivalent EEG sources: a cross-sectional and longitudinal study. Clin Neurophysiol 2000;111(11):1961–7.

Jelic V, Shigeta M, Julin P, Almkvist O, Winblad B, Wahlund LO. Quantitative electroencephalography power and coherence in Alzheimer's disease and mild cognitive impairment. Dementia 1996;7(6): 314–23.

Jeong J. EEG dynamics in patients with Alzheimer's disease. Clin Neurophysiol 2004;115(7):1490–505.

Jeong J, Gore JC, Peterson BS. Mutual information analysis of the EEG in patients with Alzheimer's disease. Clin Neurophysiol 2001;112(5): 827–35.

Joyce CA, Gorodnitsky IF, Kutas M. Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. Psychophysiology 2004;41(2):313–25.

Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. Clin Neurophysiol 2000;111(10):1745–58.

Kawakatsu M. Application of ICA to MEG noise reduction. Fourth international symposium on independent component analysis and blind signal separation (ICA2003), Nara, Japan; April 1–4, 2003, p. 535–41.

Kobayashi K, Akiyama T, Nakahori T, Yoshinaga H, Gotman J. Systematic source estimation of spikes by a combination of independent component analysis and RAP-MUSIC. II: preliminary clinical application. Clin Neurophysiol 2002;113(5):725–34.

Lindau M, Jelic V, Johansson SE, Andersen C, Wahlund LO, Almkvist O. Quantitative EEG abnormalities and cognitive dysfunctions in frontotemporal dementia and Alzheimer's disease. Dement Geriatr Cogn Disord 2003;15(2):106–14.

Locatelli T, Cursi M, Liberati D, Franceschi M, Comi G. EEG coherence in Alzheimer's disease. Electroencephalogr Clin Neurophysiol 1998; 106(3):229–37.

McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984; 34(7):939–44.

Musha T, Asada T, Yamashita F, Kinoshita T, Chen Z, Matsuda H, et al. A new EEG method for estimating cortical neuronal impairment that is sensitive to early stage Alzheimer's disease. Clin Neurophysiol 2002; 113(7):1052–8.

Pritchard WS, Duke DW, Coburn KL, Moore NC, Tucker KA, Jann MW, et al. EEG-based, neural-net predictive classification of Alzheimer's disease versus control subjects is augmented by non-linear EEG measures. Electroencephalogr Clin Neurophysiol 1994; 91(2):118–30.

Rapoport SI. Functional brain imaging to identify affected subjects genetically at risk for Alzheimer's disease. Proc Natl Acad Sci USA 2000;97(11):5696–8.

Stam CJ, van der Made Y, Pijnenburg YA, Scheltens P. EEG synchronization in mild cognitive impairment and Alzheimer's disease. Acta Neurol Scand 2003;108(2):90–6.

Szupiluk R, Cichocki A. Blind signal separation using second order statistics. Proceedings of SPETO; 2001, p. 485–8.

Tang AC, Pearlmutter BA, Malaszenko NA, Phung DB. Independent components of magnetoencephalography: single-trial response onset time estimation. NeuroImage 2002;17:1773–89.

Tong L, Soon V, Huang YF, Liu R. Indeterminacy and identifiability of blind identification. IEEE Trans CAS 1991;38:499–509.

Tong L, Inouye Y, Liu R. Waveform-preserving blind estimation of multiple independent sources. IEEE Trans Signal Process 1993;41(7): 2461–70.

Vorobyov S, Cichocki A. Blind noise reduction for multisensory signals using ICA and subspace filtering, with application to EEG analysis. Biol Cybern 2002;86(4):293–303.

Wackermann J. Beyond mapping: estimating complexity of multichannel EEG recordings. Acta Neurobiol Exp (Wars) 1996;56(1):197–208.

Wada Y, Nanbu Y, Koshino Y, Yamaguchi N, Hashimoto T. Reduced interhemispheric EEG coherence in Alzheimer disease: analysis during rest and photic stimulation. Alzheimer Dis Assoc Disord 1998;12(3): 175–81.

Wagner AD. Early detection of Alzheimer's disease: an fMRI marker for people at risk. Nat Neurosci 2000;3(10):973–4.

Wechsler D. Wechsler memory scale: revised manual. San Antonio, TX: Psychological Corp.; 1987.

Xu N, Gao X, Hong B, Miao X, Gao S, Yang F. BCI Competition 2003—Data set IIb: enhancing P300 wave detection using ICA-based subspace projections for BCI applications. IEEE Trans Biomed Eng 2004;51(6): 1067–72.