DECOMPOSITION ALGORITHMS FOR ANALYSING BRAIN SIGNALS

Klaus-Robert Müller[†*], *Jens Kohlmorgen*[†], *Andreas Ziehe*[†], *Benjamin Blankertz*[†]

[†]GMD FIRST.IDA, Kekuléstr. 7, 12489 Berlin, Germany
[*]University of Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany

ABSTRACT

Analyzing biomedical data – e.g. from the brain – we encounter fundamental problems that lie largely in the fields of signal processing and machine learning. The current paper presents at first a method to deal with non-stationary signals, subsequently the signal processing technique of independent component analysis (ICA) is reviewed. We use EEG recordings of continuous auditory perception as illustration for the discussed algorithms.

1. INTRODUCTION

De-noising and artifact removal of biomedical data, e.g. brain signals is essential for providing a sound basis for neurophysiological model building. Brain data is inherently very noisy – technical noise sources, e.g. sensor noise, and biological noise sources, e.g. heart beat or eye blinks, interfere. So the machine learning and signal processing community faces an exciting challenge and testbed to apply state-of-the-art techniques for projection, prediction, classification, artifact reduction and de-noising, and to further develop them.

The following sections give short overviews about typical problems and challenges that are encountered during the analysis of biomedical data. One EEG data set of continuous auditory perception (described in section 1.1) serves as “red thread” to demonstrate the use of the data analysis techniques. Section 2 deals with the problem of treating non-stationary data under the assumption that it originates from a multimodal switching or drifting dynamical system (very much in the spirit of [15, 23]). We show an unsupervised segmentation of single EEG channels into dynamical modes that corresponds to an external stimulus. The subsequent section uses the same EEG, but now all 23 channels to illustrate ICA type projection methods for artifact removal. The final section discusses and points out open problems and future challenges.

A.Z. was partly funded by DFG under contracts JA 379/52 and JA 379/71. G.N. and G.C. were supported by DFG grant MA 1782/3-1. We thank Jörn Rittweger, Gabriel Curio and Gunnar Rätsch for valuable discussions. Correspondence to Klaus@first.gmd.de.

1.1. An Example EEG Data Set

The EEG data used throughout the paper consists of 23 channels (electrodes placed at prominent positions in the 10/20 system) sampled at 1000 Hz. It was recorded with a Neuroscan device. The subject was in a resting position in an armchair and had his eyes closed during the whole measurement session of approximately 11 minutes. The room in which the experiment took place is not soundproof and the subject was positioned in about 2 meters distance from the electronic devices for the recording of the EEG data.

We played an auditory stimulus based on the first eight bars of the variation 30 (quodlibet) of Bach’s so called Goldberg Variations (interpreted by Glenn Gould, 1981). The stimulus consists first of an enlarged period of silence (90 seconds) followed by 10 alternating sections of music and silence of the same length ($\sim$ 26 seconds). The subject heard the music biaurally (but monophonically) over earphones from a battery driven discman. An envelope of the music was synchronously fed into the 24-th channel of the EEG head box for reference.

2. NON-STATIONARITY AND SEGMENTATION

For a better understanding of a biological system it is desirable to learn about the dynamics of the measured signal components. A useful way of description of such time-series is to predict them and therefore construct a model of their dynamics. Typically, biomedical data has strong intrinsic non-stationarities. In such cases (see e.g. [15]), it is very helpful to first resolve the non-stationarities by a segmentation into stationary parts and then to identify the dynamical system inherent to the data. Among other techniques (see [15] for references) the ACE algorithm has shown to be a particular powerful data analysis technique, if the data is not purely noise driven but contains some deterministic components.

2.1. The ACE Framework

In the following, we briefly outline the Annealed Competition of Experts (ACE) method (see [15, 23] for a detailed description). ACE is a framework for the analysis of time series from switching or drifting dynamics, in which adaptive prediction experts specialize on the dynamics of individual operating modes hidden in the data. An ensemble of experts f_i, $i = 1, ..., N$, is trained in order to maximize the likelihood L that the ensemble might have generated a given time series. This is accomplished by using a gradient method (cf. [15]). The derivative of the log-likelihood with respect to the output of an expert is given by

$$\frac{\partial \log L}{\partial f_i} \propto \left[\frac{e^{-\beta(y-f_i)^2}}{\sum_j e^{-\beta(y-f_j)^2}} \right] (y - f_i), \qquad (1)$$

where y is a data point to be predicted and β is a scaling factor which controls the degree of competition between the experts. Eq. (1) is a special case of the well-known mixtures of experts approach [12], in so far as the input-gating network is simply omitted. This is because the ACE method aims to identify operating modes also in cases where the current input to the ensemble is not sufficient to distinguish between different modes. Instead of selecting an expert based on the input, ACE uses a moving average of the expert's prediction errors as selection criterion. In this way, memory is introduced into the expert selection scheme to exploit the low mode switching frequency compared to the sampling rate (see [15, 23] for details).

2.2. Limits and Problems

An assumption of the ACE framework is that mode changes occur infrequent, i.e. between two mode changes the dynamics is expected to operate more or less stationary in one mode for a certain number of time steps. A second prerequisite is that the individual dynamics can be modeled to some extent by a time-invariant mapping of past data points to future data points. Thus, it requires some functional dependence in the data. In the case of EEG data, the ACE experts typically still have a large prediction error after training. Although the performance is better than just predicting the global or local mean, the EEG dynamics can not be reconstructed by means of iterated prediction. We found that the predictors trained on EEG can only be used to *discriminate* between different modes, but they do not capture the individual dynamics properly [15]. This result, however, is not surprising, since the EEG is a very complex signal.

2.3. Results

To give an example, we applied ACE to the EEG recording from section 1.1. The idea was to find structure in EEG that corresponds to the two phases. For EEG position Cz, the result of ACE is depicted in Fig. 1. The data recorded from Cz was first subsampled from 1 kHz to 100 Hz and then the first-order differences were taken as training data. As predictors we used 8 radial basis function (RBF) networks with 6 Gaussian basis functions. The embedding dimension was $d = 4$ and the time lag $\tau = 2$ (cf. [15]). The resulting ACE segmentation nicely corresponds to the phases of music and silence (Fig. 1). Note that the segmentation was done purely data driven: only a single channel of unlabeled EEG data and *not* the music signal was given for training the experts.

3. ICA PROJECTION TECHNIQUES

Blind source separation (BSS) methods have been successfully applied for a variety of problems (see e.g. [1, 10, 3, 5, 18, 20, 27, 29, 30]). The source separation problem is stated as follows. Consider M unknown sources that generate M statistically independent time series $s_i(t)$ $i = 1, ..., M$, $t = 1, ..., T$ that are spatially uncorrelated but have a 'non-delta' temporal autocorrelation function. A sensor array consisting of M sensors $x_j(t)$ measures a stationary linear superposition

$$x_j(t) = \sum_i A_{ji} s_i(t) \qquad (2)$$

(in matrix notation $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$). The goal is to identify $\mathbf{A}$ in this model and to blindly reconstruct $\mathbf{s}(t)$ given only $\mathbf{x}(t)$. This decomposition approach is suited for the analysis of multichannel recordings of brain signals, like EEG or MEG, and can be used for post-processing of the measured data. The spatial structure of the recorded magnetic/electric fields is condensated in the columns of the mixing matrix $\mathbf{A}$ and the temporal information is preserved in the components $s_i(t)$. The most appealing advantage is – as in the previous section – the unsupervised ("blind") functioning of this method, i.e. no reference or template signals are needed.

Although many algorithms [1, 3, 13, 16, 10] utilize higher-order statistics to exploit the non-Gaussian distribution of the sources to achieve a separation, a decomposition of neuro-physiological signals relying on second-order statistics only was shown to be useful

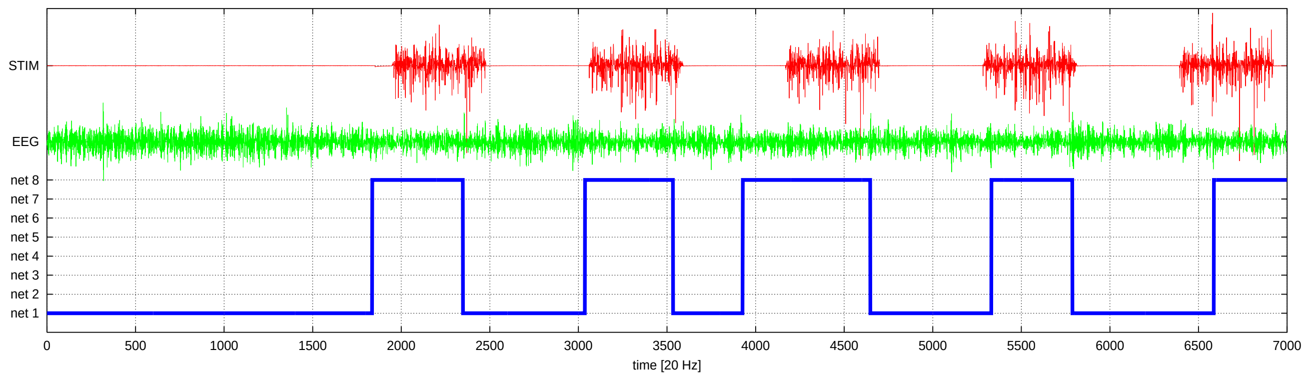

Figure 1: Segmentation of EEG data recorded from a human subject during alternating phases of music and silence (signal STIM). The resulting ACE segmentation into dynamical modes, given only the unlabeled EEG data, nicely corresponds to the two phases. Note, however, that the first and third transition towards prediction expert 8 (net 8) are 6 resp. 12 s before the next music interval starts, which might be attributed to the varying attention in the experiment.

[30, 28]. The success of the second-order approach is owed mainly to the fact that neuro-physiological signals possess an inherent time structure.

The advantage of second-order methods is their computational simplicity and efficiency. They are also more robust against outliers and for a reliable estimate of covariances only comparably few samples are needed. In the following we give a detailed description of one particular implementation of a second-order BSS algorithm.

Let us recall that for mutual independent signals the cross-correlation function vanishes. If the signals have a temporal structure resulting in a non-delta autocorrelation function we can define so called time-delayed correlation matrices $R_{\tau(\mathbf{s})}$, which should be in diagonal form. This knowledge is used to calculate the unknown mixing matrix in Eq. (2) as follows. Let us consider time-lagged correlation matrices of the form

$$R_{\tau(\mathbf{x})} = \langle \mathbf{x}(t)\mathbf{x}^T(t-\tau)\rangle = \begin{bmatrix} \phi_{x_1,x_1}(\tau) & \cdots & \phi_{x_1,x_n}(\tau) \\ \phi_{x_2,x_1}(\tau) & \cdots & \phi_{x_2,x_n}(\tau) \\ \vdots & \ddots & \vdots \\ \phi_{x_n,x_1}(\tau) & \cdots & \phi_{x_n,x_n}(\tau) \end{bmatrix},$$

where $\phi_{x_i,x_j}(\tau) = \langle x_i(t)x_j(t-\tau)\rangle$ denotes the respective auto- or cross-correlation functions.

Since the mixing model in Eq. (2) is just a linear transformation we can substitute $\mathbf{x}(t)$ by $\mathbf{As}(t)$ and get:

$$\begin{aligned} R_{\tau(\mathbf{x})} &= \langle \mathbf{x}(t)\mathbf{x}^T(t-\tau)\rangle \\ &= \langle \mathbf{As}(t)\ (\mathbf{As}(t-\tau))^T\rangle \\ &= \mathbf{A}R_{\tau(\mathbf{s})}\mathbf{A}^T. \end{aligned} \quad (3)$$

Obviously, the temporal de-correlation algorithm can be used successfully only if the signals have non-identical spectra i.e. distinctive autocorrelation functions, since otherwise the eigenvalues would be degenerate. Hence the quality of the signal separation depends strongly on the very choice of τ, therefore it is better to try to diagonalize a larger set $\{R_\tau(\mathbf{x})\}$ of delayed correlation matrices simultaneously [29]. To achieve an approximate simultaneous diagonalization of several matrices one proceeds in two steps: (1) whitening and (2) a number of Jacobi rotations [5, 11]. First a whitening transformation $\mathcal{W} = R_{(\mathbf{x})}^{-\frac{1}{2}} = (\mathbf{V}\Lambda\mathbf{V}^T)^{-\frac{1}{2}} = \mathbf{V}\Lambda^{-\frac{1}{2}}\mathbf{V}^T$ achieves a white basis $\mathbf{z}(t) = \mathcal{W}\mathbf{x}(t)$ on a unit sphere [8]. The remaining set of time delayed correlation matrices $R_{\tau(\mathbf{z})}$ can be diagonalized subsequently by a unique orthogonal transformation $\mathbf{Q}$, since in the white basis all degrees of freedom left are rotations [5]. For several matrices, that share a common Eigen-structure, a Jacobi-like algorithm proposed by Cardoso can be used to find a satisfying solution [11, 6]. The basic idea is that one can approximate the rotation matrix $\mathbf{Q}$ by a sequence of elementary rotations $Q_k(\phi_k)$ in a two dimensional subspace each trying to minimize the off-diagonal elements

$$\min \sum_{\tau, i\neq j} |(R_\tau)_{ij}|$$

of the respective $R_{\tau(\mathbf{z})}$ matrices, where the rotation angle ϕ_k can be calculated in closed form (see [6] for details). The final rotation, which diagonalizes $R_{\tau(\mathbf{z})}$ up to a certain level of accuracy, is then obtained by $\mathbf{Q} = \prod_k Q_k(\phi_k)$. Concatenation of both transforms (whitening $\mathcal{W}$ and rotation $\mathbf{Q}$) yields an estimate of the mixing matrix $\hat{\mathbf{A}} = \mathcal{W}^{-1}\mathbf{Q}$, which has to be inverted to get the demixing matrix $\mathbf{W} = \hat{\mathbf{A}}^{-1}$ of our

TDSEP algorithm [29]. Further second order source separation algorithms are e.g. [20, 26, 14, 4].

3.1. Limits and Problems

While using ICA (or other projection) algorithms one has to be aware of their assumptions (see above), general limits and difficulties and we will give a checklist of those possible problems in the following.

(a) A particularly hard practical problem is the availability of only few data points in combination with a high-dimensional sensor input, the latter being a problem of computational complexity that can be overcome by e.g. TDSEP or Fast ICA algorithms [10], while the former is a ubiquitous systematic statistical problem ("curse of dimensionality").

(b) Channel noise is potentially a rather serious harm to ICA algorithms as it effectively doubles the number of independent sources. Often, however, the application problem allows to construct an approximate noise model and projections to signal spaces orthogonal to the noise space can be performed [21, 9].

(c) A further difficulty comes from the independence assumption: any projection algorithm can only retrieve and denoise signals *within* the subspace of the linear space of all components that we define by certain a priori assumptions. Generally speaking in data analysis we are always interested in finding a proper basis that is describing the relevant characteristics of the data. So we aim for a linear component analysis (generative model) where the components (latent variables) are meaningful with respect to the application in mind [2]. An orthogonality assumption leads to principal component analysis (PCA), positivity constraints on the linear decomposition yields non-negative matrix factorization [17], orthogonality in some feature space gives rise to non-linear PCA (cf. [25]) and enforcing mutual independence of the components defines ICA.

(d) The number of sources that can be unmixed has to be assumed to be smaller or equal than the number of sensors. However, in biomedical measurements a multitude of microscopic sources contributes to the recorded signal. How these sources can be collapsed into fewer macroscopic sources depends on the particular biological system under study.

(e) The mixing model as defined in Eq.(2) might be too simple-minded and models that include noise terms (see discussion above) or cope with convolutive (e.g. EMG) or even non-linear mixtures would be more appropriate. For MEG/EEG recordings a linear model is sufficient, due to the linearly superimposing magnetic/electric fields.

(f) Outliers can strongly decrease the performance of ICA algorithms involving higher-order statistics, nevertheless second-order algorithms are more robust against outliers.

3.2. Results

To apply ICA algorithms to this data we have to make sure that the criteria of the checklist from section 3.1 are fulfilled. The criterion (a) is easy to meet since we have 23 channels and abundant data points per channel. Additive channel noise (b) is an issue due to the general experimental set-up of an EEG in a non-shielded environment. Our assumption of temporal decorrelation/independence and a linear mixing model (c) holds as we are looking for signals with high temporal structure. Also the number of sources (d) has to be less than the number of sensors. Even though the exact number of sources is unknown, at least the eigenvalue spectrum of the covariance matrix decayed rapidly. Finally, as we see from the occasional spikes in various channels, outliers (f) can pose a problem in this data set.

The spectra of the EEG data set all look rather similar (strong α rhythm, weaker β rhythm and 50 Hz noise) with but subtly different mixtures of the rhythms. A decomposition seems therefore promising.

We applied TDSEP ($\tau = \{0, \dots, 50\}$) to the EEG channels resulting in 23 (approximately) independent components. Fig. 2 shows three selected components to demonstrate the decomposition properties of the ICA approach: IC 1 in Fig. 2(a) can be clearly perceived as artifact as it consists mainly of a 50 Hz power line interference. Fig. 2(b) depicts component IC 12 with strong α-rhythm (and comparably weak β contribution), while the component IC 17 in Fig. 2(c) is largely β dominated. Note that this physiologically useful decomposition was found in an unsupervised manner. Other components have mostly rather flat and unstructured spectra.

Further information is contained in the demixing matrix $\mathbf{W}$ computed by the ICA algorithm. For each component, say the jth, the corresponding unmixing row $w_j = (W_{j1}, \dots, W_{jM})$ is the weighting of the sensor data. Since the sensors have fixed known positions on the head, we can compute the activity of the corresponding (latent) source on the scull surface. Further physiological reasoning goes beyond the scope of this contribution.

4. DISCUSSION AND OUTLOOK

By means of a standard EEG recording we demonstrated the use of two interesting decomposition techniques. ACE can distinguish between two dynamical modes (music vs. silence) using a single, highly noisy

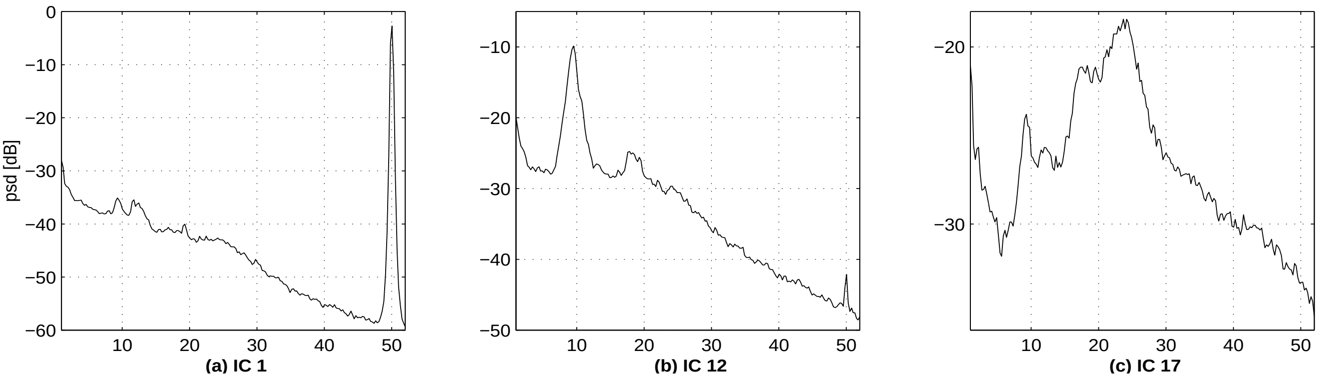

Figure 2: Spectra of selected independent components (see text).

EEG channel as input. The TDSEP source separation approach – here we reviewed a particular blind source separation approach out of many other possible ones (e.g. [1, 3, 13, 16, 10, 7, 2, 24]), constructed a linear projection from mixed multi-channel EEG data by enforcing temporal independence using only second order statistics. This projection technique gives a useful decomposition into artifacts (e.g. 50Hz noise) and several components that represent typical EEG bands (e.g. α, β activity). The decompositions (segments, ICA components) obtained can then serve as a basis for neuro-physiological model building, which might involve further steps as for example: source localization, identification or a detailed mathematical modeling e.g. in terms of differential equations. Our emphasis in this overview paper was to discuss general problems (interesting to the signal processing community) that are encountered in such a typical biomedical data analysis set-up rather than to provide new algorithms or detailed neuro-physiological insights.

So far segmentation algorithms like ACE had their strength in off-line data analysis. Further research is required to obtain on-line segmentation algorithms. A combination of ACE and ICA techniques to a multichannel framework appears promising.

Clearly, future directions for the ICA methodology have to consider the practical cases where strong noise is present or a-priori knowledge is available [22], or where the underlying components might have hidden dependencies that do not match the standard ICA model assumptions or the limits discussed in section 3.1. In the context of using prior knowledge, in particular a combination of beam-forming and ICA methods seems auspicious. Furthermore, it would be interesting to see in a biomedical context whether nonlinear ICA models [19] provide useful decompositions beyond the linear ones.

From the biomedical point of view it is highly important to find a measure that allows to assess the reliability of a decomposition result.

5. REFERENCES

[1] S. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, editors, *Advances in Neural Information Processing Systems (NIPS 95)*, volume 8, pages 882–893. The MIT Press, 1996.

[2] H. Attias. Independent factor analysis. *Neural Computation*, 11:803–851, 1998.

[3] A. J. Bell and T. J. Sejnowski. An information maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[4] A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique based on second order statistics. *IEEE Trans. on SP*, 45(2):434–44, Feb 1997.

[5] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non Gaussian signals. *IEE Proceedings-F*, 140(6):362–370, 1993.

[6] J.-F. Cardoso and A. Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM J.Mat.Anal.Appl.*, 17(1):161 ff., 1996.

[7] P. Comon. Independent component analysis, a new concept? *Signal Processing, Elsevier*, 36(3):287–314, 1994.

[8] G.H. Golub and C.F. van Loan. *Matrix Computation*. The Johns Hopkins University Press, London, 1989.

[9] A. Hyvärinen. Independent component analysis in the presence of gaussian noise by maximizing joint likelihood. *Neurocomputing*, 22:49–67, 1998.

[10] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997.

[11] C.G.J. Jacobi. Über ein leichtes Verfahren, die in der Theorie der Säcularstörungen vorkommenden Gleichungen numerisch aufzulösen. *Crelle J. reine angew. Mathematik*, 30:51–94, 1846.

[12] R. A. Jacobs, M.I. Jordan, S.J. Nowlan, and G.E. Hinton. Adaptive mixtures of local experts. *Neural Computation*, 3:79–87, 1991.

[13] Ch. Jutten and J. Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.

[14] B.-U. Köhler and R. Orglmeister. Independent component analysis using autoregressive models. In *Proc. Int. Workshop on Independent Component Analysis and Blind Source Separation (ICA'99)*, pages 359–364, Aussois, France, January 11–15, 1999.

[15] J. Kohlmorgen, K.-R. Müller, J. Rittweger, and K. Pawelzik. Identification of nonstationary dynamics in physiological recordings. *Biological Cybernetics*, 2000. in press.

[16] B. Laheld and J.-F. Cardoso. Equivariant adaptive source separation. *IEEE Trans. on Signal Processing*, 44(12):3017–3030, 1996.

[17] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

[18] S. Makeig, T-P. Jung, D. Ghahremani, A.J. Bell, and T.J. Sejnowski. Blind separation of event-related brain responses into independent components. *Proc. Natl. Acad. Sci. USA*, 94:10979–10984, 1997.

[19] G.C. Marques and L.B. Almeida. Separation of nonlinear mixtures using pattern repulsion. In *Proc. Int. Workshop on Independent Component Analysis and Blind Source Separation (ICA'99)*, pages 277–282, Aussois, 1999.

[20] L. Molgedey and H.G. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters*, 72(23):3634–3637, 1994.

[21] K.-R. Müller, P. Philips, and A. Ziehe. $JADE_{TD}$: Combining higher-order statistics and temporal information for blind source separation (with noise). In *Proc. Int. Workshop on Independent Component Analysis and Blind Source Separation (ICA'99)*, pages 87–92, Aussois, 1999.

[22] L. Parra, C. D. Spence, P. Sajda, A. Ziehe, and K.-R. Müller. Unmixing hyperspectral data. In S.A. Solla, T.K. Leen, and K.-R. Müller, editors, *Advances in Neural Information Processing Systems 12*, pages 942–948. MIT Press, 2000.

[23] K. Pawelzik, J. Kohlmorgen, and K.-R. Müller. Annealed competition of experts for a segmentation and classification of switching dynamics. *Neural Computation*, 8:340–356, 1996.

[24] J.C. Principe and D. Xu. Information-theoretic learning using Renyi's quadratic entropy. In *Proc. Int. Workshop on Independent Component Analysis and Blind Source Separation (ICA'99)*, pages 407–412, Aussois, 1999.

[25] B. Schölkopf, A. Smola, and K.-R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10:1299 – 1319, 1998.

[26] L. Tong, V.C. Soon, and Y. Huang. Indeterminacy and identifiability of identification. *IEEE Trans. on Circuits and Systems*, 38(5):499–509, 1991.

[27] R. Vigário, V. Jousmäki, M. Hämäläinen, R. Hari, and E. Oja. Independent component analysis for identification of artifacts in magnetoencephalographic recordings. In Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, editors, *Advances in Neural Information Processing Systems*, volume 10. The MIT Press, 1998.

[28] G. Wübbeler, K.-R. Müller, A. Ziehe, B.-M. Mackert, L. Trahms, and G. Curio. Independent component analysis of non-invasively recorded cortical magnetic dc-fields in humans. *IEEE Transactions on biomedical Engineering*, 2000.

[29] A. Ziehe and K.-R. Müller. TDSEP – an efficient algorithm for blind separation using time structure. In L. Niklasson, M. Bodén, and T. Ziemke, editors, *Proc. Int. Conf. on Artificial Neural Networks (ICANN'98)*, pages 675 – 680, Skövde, Sweden, 1998. Springer Verlag.

[30] A. Ziehe, K.-R. Müller, G. Nolte, B.-M. Mackert, and G. Curio. Artifact reduction in magnetoneurography based on time-delayed second order correlations. *IEEE Trans. Biomed. Eng.*, 47(1):75–87, 2000. also GMD Technical Report No. 31, 1998.