

MODEL ORDER ESTIMATION FOR INDEPENDENT COMPONENT ANALYSIS OF EPOCHED EEG SIGNALS.

Peter Mondrup Rasmussen, Morten Mørup, Lars Kai Hansen

Informatics and Mathematical Modelling, Technical University of Denmark

Richard Pedersens Plads, bld. 321, DK-2800 Kgs. Lyngby, Denmark

peter.mondrup@gmail.com, mm@imm.dtu.dk, lkh@imm.dtu.dk

Sidse M. Arnfred

Cognitive Research Unit, Department of Psychiatry, University Hospital of Copenhagen, Hvidovre

Brøndbyøstervej 160, DK-2605 Brøndby, Denmark

Sidse.Arnfred@hvh.regionh.dk

Keywords: EEG; Event related potentials; Independent component analysis (ICA); Molgedey Schuster; TDSEP; Model selection; Cross validation.

Abstract: In analysis of multi-channel event related EEG signals independent component analysis (ICA) has become a widely used tool to attempt to separate the data into neural activity, physiological and non-physiological artifacts. High density electrode systems offer an opportunity to estimate a corresponding large number of independent components (ICs). However, too large a number of ICs leads to overfitting of the ICA model, which can have a major impact on the model validity. Consequently, finding the optimal number of components in the ICA model is an important problem. In this paper we present a method for model order selection, based on a probabilistic framework. The proposed method is a modification of the Molgedey Schuster (MS) algorithm to epoched, i.e. event related data. Thus, the contribution of the present paper can be summarized as follows: 1) We advocate MS as a low complexity ICA alternative for EEG. 2) We define an epoch based likelihood function for estimation of a principled unbiased 'test error'. 3) Based on the unbiased test error measure we perform model order selection for ICA of EEG. Applied to a 64 channel EEG data set we were able to determine an optimum order of the ICA model and to extract 22 ICs related to the neurophysiological stimulus responses as well as ICs related to physiological- and non-physiological noise. Furthermore, highly relevant high frequency response information was captured by the ICA model.

1 INTRODUCTION

The electroencephalogram (EEG) is a recording of electrophysiological brain activity and the major benefit of EEG relative to other brain imaging modalities is a high temporal resolution. The basic electrophysiology of the EEG signal implies that it may be modelled as a linear mixture of multiple sources of neural activity, non-brain physiological artifacts such as eye blinks, eye movements, and muscle activity, and non-physiological artifacts such as line noise, and electrode movement (Onton et al., 2006; Hesse and James, 2004). By electrical conductance these source signals instantaneously project to the scalp electrodes used for acquisition (Onton et al., 2006). Assuming linear addition of these relatively independent source signals at the scalp electrodes motivates the use of instantaneous independent component analysis (ICA) as a technique for extracting a set of under-

lying sources from the recorded EEG signals (James and Hesse, 2005; Makeig et al., 2002; Hyvarinen and Oja, 2000). The EEGLAB software is widely used for decomposing EEG using ICA (Delorme and Makeig, 2004). More accurate modeling of the signal component(s) including residual delayed correlations can be achieved using so-called convolutive ICA in a subspace of components extracted by the initial instantaneous ICA (Dyrholm et al., 2007).

Epochs extracted from an EEG experiment are described by the data matrix $X \in \mathbb{R}^{M \times N}$, where M is the number of electrode channels and N is the number of sampling time points. In the following N is the total time consisting of a certain number of epochs, i.e., individual experiment. The epochs may be separated by variable time intervals according to the specific experimental design. It is a specific point in the following, where we are going to invoke temporal correlation based models, that we do not compute temporal

correlations across epoch boundaries.

In general the ICA model can be written as

$$X = AS \quad X_{n,t} = \sum_{k=1}^K A_{n,k} S_{k,t}, \quad (1)$$

where $X_{n,t}$ is the signal at the n' th sensor at t' th time point and K is the number of sources or independent components (ICs). $A \in \mathbb{R}^{M \times K}$ is denoted the mixing matrix and $S \in \mathbb{R}^{K \times N}$ the source matrix. In this model the sources as well as the mixing coefficients are unknown. The random signal X is observed, and from this A and S are estimated. It is impossible to determine the variance (energy) of the sources, since any scalar multiplier in one of the sources could be cancelled by dividing the corresponding column in A with the same multiplier. Therefore, the sources are often assumed to have unit variance, which can be achieved by normalizing the source signals and multiply the corresponding column of the mixing matrix. Specifically, there is a sign ambiguity, if we change the sign of a source signal and change the sign of the corresponding column in the mixing matrix, the same reconstructed signal is obtained by multiplication. Finally, the ordering of components is arbitrary. We may order independent component according to variance of their contribution to the reconstructed signal.

The recovery of the mixing matrix and the sources is not possible from the covariance matrix alone, hence, by principal component analysis (PCA). Additional information is needed. ICA is often based on a non-Gaussianity assumption of the sources (Bell and Sejnowski, 1995) or by assumed differences in source auto-correlation (Molgedey and Schuster, 1994).

The number of EEG channels M may be different from the number of sources K , thus it is relevant to estimate K . Estimation of the correct number of sources can have a major impact on the validity of the ICA solution and prevents overfitting (James and Hesse, 2005). One approach to prevent overfitting is based on pre-processing by PCA, where the number of sources is determined by the number of dominant eigenvalues which account for a high proportion of the total variance in the data set. However this procedure has been criticized for sensitivity to noise (James and Hesse, 2005). Another approach is based on step-wise extraction of sources until a specified accuracy is achieved (James and Hesse, 2005). However this method is highly dependent on the choice of the accuracy level. In this paper we present a method for model order selection, based on a probabilistic framework. This approach was earlier proposed in a multimedia contexts (Kolenda et al., 2001). However, the approach requires large amount of memory for long

signals and is inapplicable to EEG signals that are epoched due to temporal discontinuities where epochs are merged. Here we present a method that is customized to epoched data with the additional benefit of reducing the memory requirement. In our method PCA leads to a number of model hypotheses, of which an ICA model is estimated using a modified version of the Molgedey Schuster (MS) algorithm (Molgedey and Schuster, 1994). The MS algorithm is chosen because it is based on source autocorrelation, which is very relevant to EEG, and because of its relative low computational complexity. We take further advantage of the epoched nature of the signals, and split the data set into a training- and a test set. Model selection, i.e., estimating K , is then based on evaluating the likelihood of each model hypothesis using the test set in order to ensure generalization.

The paper is organized as follows. First we give a description of our method and we compare by simulation study the modified MS algorithm with the currently used ICA methods for EEG TDSEP (Ziehe et al., 2000) and infomax ICA (Bell and Sejnowski, 1995). We then test our model selection scheme within the simulated data and apply our method on real event related EEG data from an experiment involving visual stimulus.

2 METHODS

In the following, a description of PCA and the MS algorithm will be given. This is followed by a description of the probabilistic modelling. Finally the procedures for model order selection is presented.

2.1 PCA

Using PCA it is possible to reduce the dimensionality of the ICA model. In EEG $M \ll N$ which leads to the singular value decomposition (SVD) $X = UDV^T$, where $U \in \mathbb{R}^{M \times M}$, $D \in \mathbb{R}^{M \times N}$, and $V \in \mathbb{R}^{N \times N}$. By selecting the first K eigenvectors in U as a new basis, the signal space \mathcal{S} is reduced to K dimensions. ICA is performed in \mathcal{S} , where A is the ICA basis projected onto the PCA subspace. The mixing matrix in the original vector space and the source signals are then given by

$$\tilde{A} = UA \quad (2)$$

$$S = A^{-1}DV^T. \quad (3)$$

The noise space \mathcal{E} is spanned by the remaining $M - K$ eigenvectors.

2.2 Molgedey Schuster separation

The Molgedey Schuster approach is based on the assumption that the autocorrelation functions of the independent sources are non-vanishing, and can be used if the source signals have different autocorrelation functions (Molgedey and Schuster, 1994; Hansen et al., 2000; Hansen et al., 2001). Time shifted data matrices X_τ and S_τ are defined followed by the definition of the cross-correlation function matrix for the mixture signals

$$C(\tau) \equiv \frac{1}{N_e} X_\tau X^T, \quad (4)$$

where $C \in \mathbb{R}^{M \times M}$ and N_e is the epoch length. For $\tau = 0$ the usual cross-correlation matrix is obtained. Due to the epoched nature of the signals $C(\tau)$ is estimated within each epoch and averaged, since the cross-correlation is not valid over epoch boundaries. Now we define the quotient matrix $Q(\tau) \equiv C(\tau)C(0)^{-1}$ which is rewritten, using the relation $X = AS$, as

$$\begin{aligned} Q(\tau) &= C(\tau)C(0)^{-1} \\ &= \frac{1}{N_e} X_\tau X^T \left(\frac{1}{N_e} X X^T \right)^{-1} \\ &= (AS_\tau)(AS)^T ((AS)(AS)^T)^{-1} \\ &= AS_\tau S^T A^T A^{-T} (SS^T)^{-1} A^{-1} \\ &= AD(\tau)D(0)^{-1}A^{-1}, \end{aligned} \quad (5)$$

where $D(\tau) \equiv \frac{1}{N_e} S_\tau S^T$ in the limit $N_e \rightarrow \infty$ is the diagonal source cross-correlation matrix at lag τ . It is now seen, that the eigenvalue decomposition of the quotient matrix

$$Q\Phi = \Phi\Lambda \quad (6)$$

leads to $A = \Phi$ and $\Lambda = C(\tau)C(0)^{-1}$. τ is estimated as described in (Kolenda et al., 2001).

2.3 Probabilistic modeling

The ICA model is defined in terms of the model parameters i.e. the mixing matrix A . Using Bayes theorem the probability of specific model parameters given the observed data $P(A|X)$ can be written as

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)}, \quad (7)$$

where $P(X|A)$ is the likelihood function, and $P(A)$ is the prior probability of a specific model. This likelihood function is rewritten as

$$\begin{aligned} P(X|A) &= \int P(X, S|A) dS \\ &= \int P(X|S, A) P(S) dS \\ &= \int \delta(X - AS) P(S) dS \end{aligned} \quad (8)$$

Evaluating the integral in (8) gives

$$P(X|A) = P(A^{-1}X) \frac{1}{||A||}, \quad (9)$$

where $||A||$ is the absolute determinant of A .

In order to write the likelihood function we need the likelihood for the reduced signal space \mathcal{S} as well as for the noise space \mathcal{E} . Since the sources are statistically independent we have $P(S) = \prod_{i=1}^K P(s_i)$, where s_i denotes the i 'th source. If the sources are assumed stationary, independent, have zero mean, possess time-autocorrelation and are Gaussian distributed, then the source distribution is given by (Hansen et al., 2001; Hansen et al., 2000)

$$P(S) = \prod_{i=1}^K \frac{1}{\sqrt{|2\pi\Sigma_{s_i}|}} \exp\left(-\frac{1}{2}s_i^T \Sigma_{s_i}^{-1} s_i\right), \quad (10)$$

where $\Sigma_{s_i} = E[s_i s_i^T] = \text{Toeplitz}([\gamma_{s_i}(0), \dots, \gamma_{s_i}(N_e - 1)])$ and γ_{s_i} are the source autocorrelation function values. The autocorrelation function values are estimated in each epoch and averaged. This estimate of the source distribution leads to a formulation of the likelihood for the signal space as

$$\begin{aligned} P(S|A) &= \prod_{i=1}^K \frac{1}{\sqrt{|2\pi\Sigma_{s_i}|}} \left(\frac{1}{||A||} \right)^{N_e} \\ &\quad \times \exp\left(-\frac{1}{2}s_i^{-1} \Sigma_{s_i}^T s_i\right). \end{aligned} \quad (11)$$

The noise space \mathcal{E} is assumed to be isotropic with noise variance $\sigma_{\mathcal{E}}^2 = (M - K)^{-1} \sum_{i=K+1}^M D_{ii}^2$. It can be shown (Kolenda et al., 2001; Minka, 2001) that

$$\begin{aligned} P(\mathcal{E}|\sigma_{\mathcal{E}}^2) &= (2\pi\sigma_{\mathcal{E}}^2)^{-\frac{N_e(M-K)}{2}} \\ &\quad \times \exp\left(-\frac{N_e(M-K)}{2}\right). \end{aligned} \quad (12)$$

The signal and noise space are assumed independent which leads to the likelihood function

$$P(X|A) = P(S|A)P(\mathcal{E}|\sigma_{\mathcal{E}}^2). \quad (13)$$

2.4 Model order selection

PCA reduction of dimensionality leads to a set of M model hypotheses. Since the data set consists of a large number of epoch e.g. 105, we have the opportunity to split the data set into a training set \mathcal{D}_{train} and a test set \mathcal{D}_{test} . Using \mathcal{D}_{train} the model parameters A and Σ_{s_i} in (11) are estimated. The negative logarithm of the likelihood function (13) is then evaluated using

\mathcal{D}_{test} , where (11) is rewritten as

$$\begin{aligned} -\log(P(\mathcal{S}|A)) &= N_e \log(|A|) \\ &+ \frac{1}{2}N_e + K \log(2\pi) + \frac{1}{2} \sum_{i=1}^K \log(|\Sigma_{s_i,train}|) \\ &+ \frac{1}{2} \sum_{i=1}^K \text{Tr}(\Sigma_{s_i,test} \Sigma_{s_i,train}^{-1}), \quad (14) \end{aligned}$$

where N_e is the number of samples in each epoch, $|A|$ is the absolute determinant of A estimated from \mathcal{D}_{train} , K is the dimension of \mathcal{D} and $\Sigma_{s_i,train}$ and $\Sigma_{s_i,test}$ are estimated from \mathcal{D}_{train} and \mathcal{D}_{test} respectively. By observing (13) model order selection is performed by identifying the model order having minimal generalization error.

3 EXPERIMENTAL EVALUATION

Simulation experiments were conducted to investigate the performance of the MS algorithm and the test set procedure for model order selection. The data sets are constructed from three sources s_1 , s_2 , and s_3 which show bursts at frequencies of 14, 19, and 11 Hz respectively. The simulated source signal matrix S consists of 80 epochs of bursts with random intra epoch interval. A 50 Hz noise source s_4 is included after generation of the 80 epochs. Electrode signals are created by mixing the simulated source signals with a specified mixing matrix A , and Gaussian noise E is added to the electrode signals leading to a specific signal-to-noise ratio (SNR) (a description of noise generation is found in Appendix). Epochs are extracted from the mixed signals using EEGLAB (Delorme and Makeig, 2004).

3.1 Algorithm performance results

By PCA the dimensionality of the simulation was reduced to 4 dimensions, and the ICA model estimated by TDSEP, the infomax ICA implementation of EEGLAB, and our modified MS algorithm. To evaluate the separation performance of our algorithm, we use the correlation between original- and estimated sources as well as the source-to-interference ratio (SIR) (Fevotte et al., 2005; Vincent et al., 2006) measure

$$SIR = 10 \log_{10} \frac{\|s_{target}\|^2}{\|e_{interf}\|^2}, \quad (15)$$

where s_{target} represents the target source or true source and e_{interf} represents interferences of unwanted sources. The SIR was calculated using the

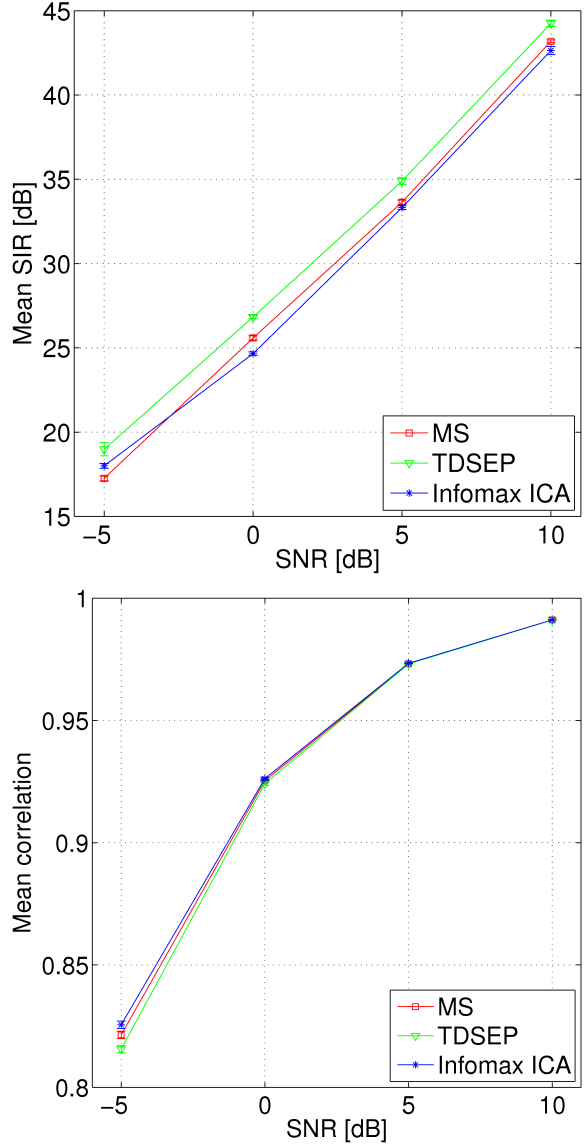


Figure 1: Simulation experiment. Results for source separation for TDSEP (Applied with default timelags 0,1), EEGLAB’s implementation of infomax ICA, and our epoch modified MS algorithm. The simulation data is constructed from four signal sources mixed out in 32 channels, Gaussian noise is added. Dimensional reduction to four dimensions by PCA. Experiment repeated 10 times, error bars indicate three standard deviations of the mean. Top: Performance of source estimation measured in terms of mean SIR. Bottom: Performance measured in terms of mean correlation between true sources and estimates.

BSS_EVAL toolbox (Fevotte et al., 2005), where the performance measure is computed for each estimated source \hat{s}_i by comparing it to the true source s_i and other unwanted sources $(s_j)_{j \neq i}$. In general SIR levels below 8-10 dB indicate failure in separation (Bos-

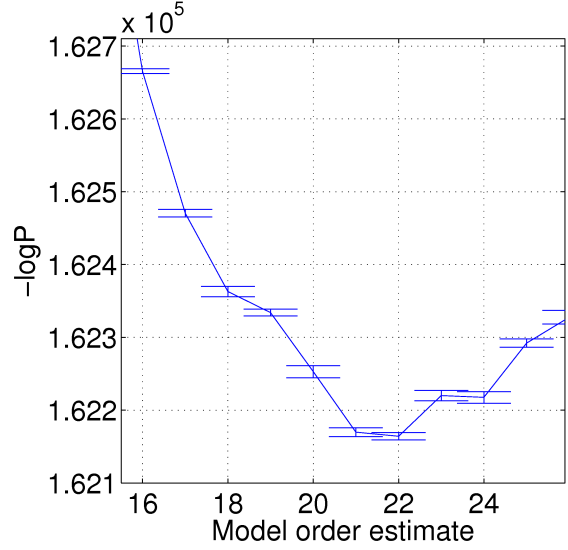
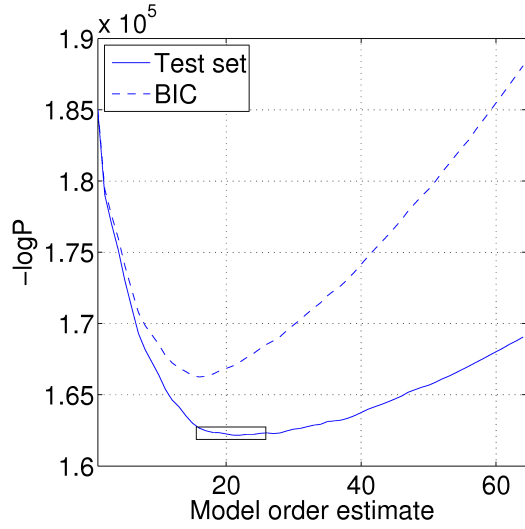


Figure 3: Left: Negative log likelihood for each of the 64 model hypotheses. Test set curve is averaged over 10 experiments, and minimum is found at 22 dimensions, suggesting an ICA model with 22 ICs. BIC is more conservative and estimates 16 ICs. Right: Zoom of the minimum region in the test set curve indicated by the box on left plot. Errorbars indicate three standard deviations of the mean.

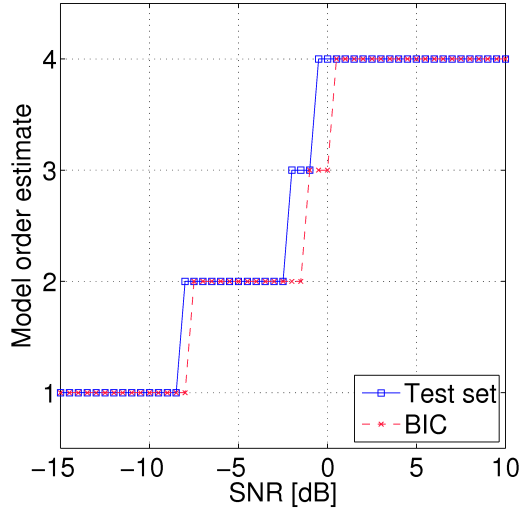


Figure 2: Simulation experiment. Results for model order selection for the test set procedure and for BIC estimation. The simulation is constructed from four signal sources mixed out in 32 channels, and Gaussian noise is added. BIC underestimates the number of sources at 0 dB whereas the test set procedure remains stable until -1 dB.

colo et al., 2004). Figure 1 shows that source estimates achieved with the modified MS algorithm are comparable with results from the alternative ICA al-

gorithms. The MS algorithm has the advantages that it is fast compared to infomax ICA and TDSEP. Furthermore, there exists a heuristic for estimation of the time lag parameter τ .

3.2 Model estimation

Model order selection is performed using the test set likelihood function (13) as evaluated using 10-fold cross-validation. Figure 2 shows model order estimates for a wide range of SNR. Here the proposed cross-validation procedure is compared to the Bayesian Information Criterion (BIC) (MacKay, 1992). The experiment indicates, that the cross-validation procedure is more robust than BIC estimation but it also has a tendency to underestimate the number of sources at low SNR.

4 APPLICATION ON EEG DATA SET

Our model selection procedure was applied on a data set from a visual stimulation experiment with experimental details described in (Mørup et al., 2006) and paradigm described in (Herrmann et al., 2004). EEG was recorded with 64 scalp electrodes arranged according to the International 10-10 system, sampling frequency 2048 Hz, band pass filter 0.1-760 Hz.

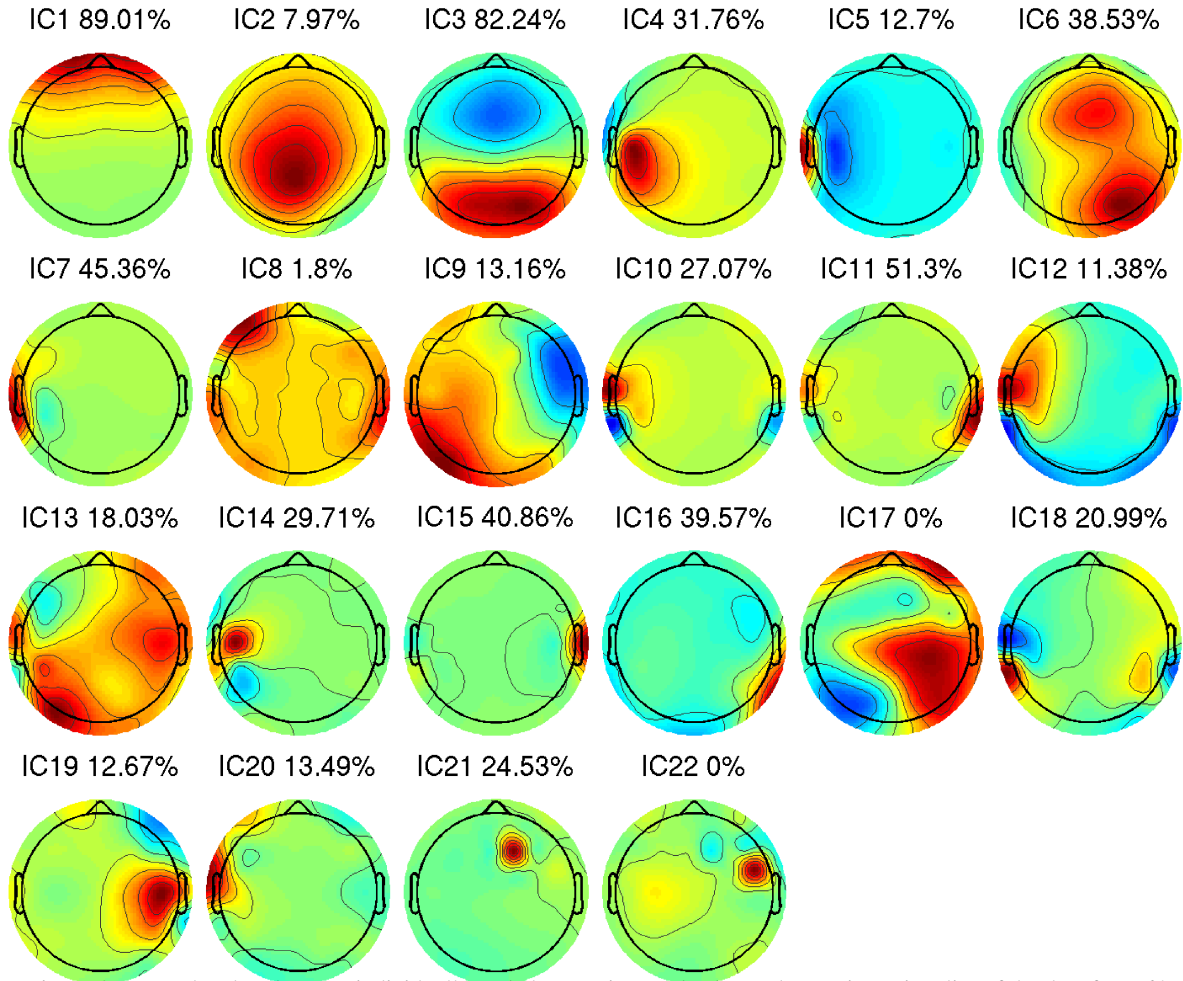


Figure 4: Interpolated scalp maps individually scaled to maximum absolute values. Dimensionality of the data from 64 scalp electrodes reduced to 22 by PCA. ICs estimated by MS algorithm, and components are sorted according to variance. The percentage at each IC indicates how much variation is explained by the respective IC of the average ERP at the electrode, where the respective IC project the strongest, calculated as $(\|X_k\|_F^2 - \|X_k - P_k\|_F^2) / \|X_k\|_F^2$, where X_k is the ERP at electrode k and P_k is the projection ERP of the respective IC onto electrode k . Estimated ICs represents different types of sources, for example, IC1 reflects eye artifacts, IC2, IC3, IC4, IC6 reflect brain sources and IC21, IC22 reflect electrode noise.

Data was high pass filtered at 3 Hz in EEGLAB, and line noise removed using a maximum likelihood 50 Hz filter. The data were referenced to digitally linked earlobes, down sampled to 256 Hz and cut into 105 epochs (-500 to 1500 ms).

PCA leads to a set of 64 model hypotheses. For each hypothesis the negative logarithm of the likelihood function (13) was evaluated using 10-fold cross-validation. The experiment was repeated 10 times with different splits of training- and test sets. Figure 3 shows model order estimation by the test set procedure and BIC estimation.

According to model order estimation the dimensionality of the data set was reduced to 22 by PCA.

ICs were estimated by the MS algorithm and sorted according to variance. Figure 4 shows all IC scalp maps. To categorize components each scalp map and averaged event related potentials (ERPs) were examined, where for example IC2, IC3, IC4 and IC6 reflect brain sources, IC1 reflects physiological eye artifacts, and IC21 and IC22 reflect electrode noise.

Further analysis of IC3 is performed by creating ERP images (Delorme and Makeig, 2004) as shown in Figure 5 top, from the PO4 electrode signal and IC3 projected onto electrode PO4. Generally the electrode signal has a larger amplitude than the projection of IC3, however, the major dynamics of the ERP seems to be captured by IC3. Another common analy-

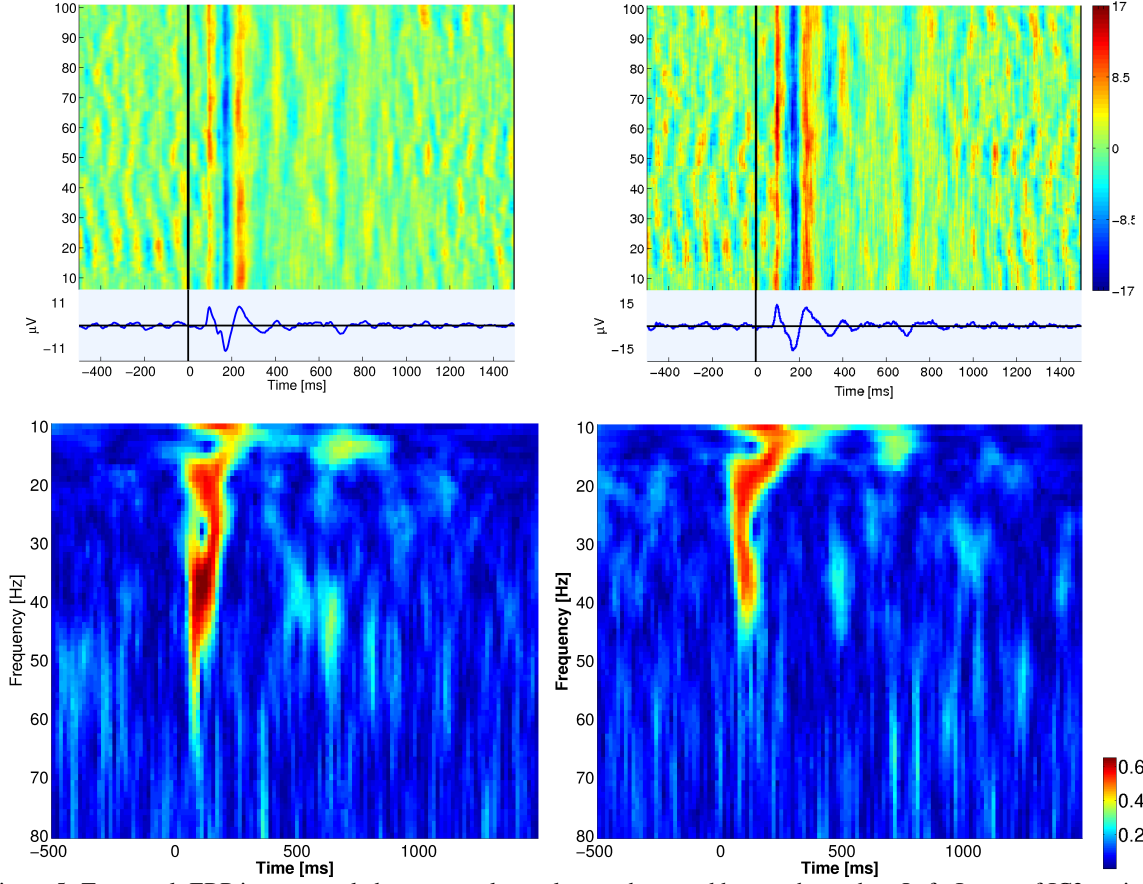


Figure 5: Top panel; ERP images, scaled to same color scale, epochs sorted by epoch number. Left; Image of IC3 projected onto electrode PO4. Right; Electrode signal at PO4. The electrode signal has a larger amplitude than the projection of IC3, however, the major dynamics of the ERP seems to be captured in IC3. Bottom panel; Time-frequency plots of the ITPC scaled to same color scale. Left; IC3 projected onto electrode PO4. Right; Electrode signal at PO4. IC3 reveals prominent evoked activity in the gamma band around 40 Hz compared to the raw electrode signal.

sis tool is time-frequency analysis of ERPs (Delorme and Makeig, 2004; Mørup et al., 2007), where different time-frequency measures exist. By ERPWAVE-LAB (Mørup et al., 2007) we wavelet transformed the data using the complex Morlet wavelet and calculated the inter-trial phase coherence (ITPC)

$$ITPC(c, f, t) = \frac{1}{N} \sum_{n=1}^N \frac{X(c, f, t, n)}{|X(c, f, t, n)|}, \quad (16)$$

where $X(c, f, t, n)$ denotes the time-frequency coefficient at channel c , frequency f , time t and epoch n . ITPC measures phase consistency over epochs. Figure 5 bottom shows time-frequency plots of ITPC for the PO4 electrode signal and IC3 projected onto electrode PO4. It is evident that IC3 reveals prominent evoked activity in the gamma band around 40 Hz compared to the raw electrode signal. Gamma band activity is consistent with earlier findings (Mørup

et al., 2006; Herrmann et al., 2004). Accordingly, relevant high frequency response information is captured in IC3, whereas noise contributions are isolated in other ICs.

5 CONCLUSION

Based on a probabilistic framework, we have formulated a cross-correlation procedure for ICA and a model order selection scheme applicable to epoched EEG signals. Our procedure is an extension of the Molgedey Schuster approach to ICA and utilizes the epoched nature of the signals. The approach is based on assuming source autocorrelation, which is very relevant to EEG. In our model selection procedure we split data into a training- and a test set to obtain an unbiased measure of generalization. Based on the

unbiased test error measure we perform model order selection for ICA of EEG. Applied to a 64 channel EEG data set we were able to determine the order of the ICA model and to extract 22 ICs related to the neurophysiological stimulus responses as well as ICs related to physiological- and non-physiological noise. Furthermore, relevant high frequency response information was captured by the ICA model. In this study we have applied our model selection procedure to EEG signals. However, our approach may also be applicable to other types of signals.

REFERENCES

- Bell, A. and Sejnowski, T. (1995). Blind separation and blind deconvolution: an information-theoretic approach. *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, 5:3415–3418.
- Boscolo, R., Pan, H., and Roychowdhury, V. (2004). Independent component analysis based on nonparametric density estimation. *IEEE Transactions on Neural Networks*, 15(1):55–65.
- Delorme, A. and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21.
- Dyrholm, M., Makeig, S., and Hansen, L. K. (2007). Model selection for convolutive ica with an application to spatiotemporal analysis of eeg. *Neural Computation*, 19(4):934–955.
- Fevotte, C., Gribonval, R., and Vincent, E. (2005). Bss eval toolbox user guide. Technical Report 1706, IRISA, Rennes, France. http://www.irisa.fr/metiss/bss_eval/.
- Hansen, L., Larsen, J., and Kolenda, T. (2001). Blind detection of independent dynamic components. *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, 5:3197–3200.
- Hansen, L. K., Larsen, J., and Kolenda, T. (2000). *On Independent Component Analysis for Multimedia Signals*. CRC Press.
- Herrmann, C. S., Lenz, D., Junge, S., Busch, N. A., and Maess, B. (2004). Memory-matches evoke human gamma-responses. *BMC Neuroscience*, 5:13.
- Hesse, C. and James, C. (2004). Stepwise model order estimation in blind source separation applied to ictal eeg. *Engineering in Medicine and Biology Society, 2004. EMBC 2004. Conference Proceedings. 26th Annual International Conference of the*, 1:986–989.
- Hyvarinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks*, 13(4-5):411–430.
- James, C. J. and Hesse, C. W. (2005). Independent component analysis for biomedical signals. *Physiological Measurement*, 26(1):R15.
- Kolenda, T., Hansen, L. K., and Larsen, J. (2001). Signal detection using ICA: Application to chat room topic spotting. *Third International Conference on Independent Component Analysis and Blind Source Separation*, pages 540–545.
- MacKay, D. (1992). Bayesian model comparison and backprop nets. *Proceedings of Neural Information Processing Systems 4*, pages 839–846.
- Makeig, S., Westerfield, M., Jung, T.-P., Enghoff, S., Townsend, J., Courchesne, E., and Sejnowski, T. (2002). Dynamic brain sources of visual evoked responses. *Science*, 295(5555):690–694.
- Minka, T. P. (2001). Automatic choice of dimensionality for PCA. *Proceedings of NIPS2000*, 13.
- Molgedey, L. and Schuster, H. (1994). Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters*, 72(23):3634–3637.
- Mørup, M., Hansen, L., and Arnfred, S. (2007). Erpwave-lab. *Journal of Neuroscience Methods*, 161(2):361–368.
- Mørup, M., Hansen, L. K., Herrmann, C. S., Parnas, J., and Arnfred, S. (2006). Parallel factor analysis as an exploratory tool for wavelet transformed event-related eeg. *NeuroImage*, 29(3):938–947.
- Onton, J., Westerfield, M., Townsend, J., and Makeig, S. (2006). Imaging human eeg dynamics using independent component analysis. *Neuroscience and Biobehavioral Reviews*, 30(6):808–822.
- Vincent, E., Gribonval, R., and Fevotte, C. (2006). Performance measurement in blind audio source separation. *IEEE Transactions on Audio, Speech and Language Processing*, 14(4):1462–1469.
- Ziehe, A., Muller, K.-R., Nolte, G., Mackert, B.-M., and Curio, G. (2000). Artifact reduction in magnetoneurography based on time-delayed second-order correlations. *IEEE Transactions on Biomedical Engineering*, 47(1):75–87.

APPENDIX

5.1 Definition of SNR

Let N be the number of samples and M the number of electrodes. The signal to noise ratio is defined by $SNR = \frac{\|AS\|_F^2}{\|E\|_F^2}$, where $\|E\|_F^2 = NM\sigma^2$. Then the variance of the additive noise is $\sigma^2 = \frac{\|AS\|_F^2}{NM \cdot SNR}$. In decibels the signal to noise ratio is $SNR_{dB} = 10 \log_{10}(SNR)$.