CAN ICA HELP IDENTIFY BRAIN TUMOR RELATED EEG SIGNALS?

M. Habl, Ch. Bauer, Ch. Ziegaus, E. W. Lang*

Institute of Biophysics
University of Regensburg
93040 Regensburg, Germany
elmar.lang@biologie.uni-regensburg.de

F. Schulmeyer

Department of Neurosurgery
University Hospital
93042 Regensburg, Germany

ABSTRACT

Scalp EEG has been used as a clinical tool for the diagnosis and treatment of brain diseases. A probabilistic ICA algorithm modified by a kernel-based source density estimation is studied to separate EEG signals from tumor patients into spatially independent source signals. A statistical method to automatically identify artifactual and tumor related ICA components is also presented.

1. INTRODUCTION

Scalp electroencephalography is a routinely used clinical tool for the diagnosis of brain diseases. Early identification and characterization of various brain tumors with non-invasive, low cost methods like EEG recordings would be especially valuable in this respect. But EEG signals are mostly contaminated by artifactual signals like eye movement and blinks, cardiac signals, muscle noise etc. Removal of these and other artifactual signals is of utmost importance whenever tumor related EEG signals are to be analyzed.

Recently Makeigh et al. [1] demonstrated the ability of ICA algorithms to separate EEG signals collected with multiple electrodes into spatially stationary and temporally independent signals generated by independent neural sources at different brain locations. ICA generally assumes that (*i*) the mixing medium is linear, (*ii*) the time courses of the sources are independent and (*iii*) the number of sources M equals the number L of sensors. In practical applications, however, the number of sources is seldom known and, in EEG recordings, might exceed the number of sensors.

Lewicki and Sejnowski [2] further proposed an overcomplete ICA algorithm which represents a generalization of the technique of independent component analysis and might provide a method for identification when there are more sources than mixtures. We further generalize this algorithm by incorporating a kernel-based density estimation of the unknown prior of the source signals. This modification provids a more flexible and adaptive approach and avoids any *ad hoc* assumptions concerning the source model.

* Financial support by the Cusanus Foundation is greatfully acknowledged.

2. THE ICA MODEL

A data vector $\vec{x} = (x_1, \ldots, x_L)^{\mathrm{T}}$ where each component represents the time course of a sensor (scalp electrode) potential is described in a, possibly overcomplete, linear basis plus additive noise according to

$$\vec{x} = \mathbf{A}\vec{s} + \vec{\epsilon} = \sum_{m=1}^{M} s_m \vec{a}_m + \vec{\epsilon} \qquad (1)$$

where $\mathbf{A}$ is an $L \times M$ mixing matrix with $L \leq M$, the columns of which represent the basis vectors $\vec{a}_m$ and $\vec{s} = (s_1, \ldots, s_M)^{\mathrm{T}}$ denote the sources with each component representing an independent neural activity pattern or any of the artifactual signals mentioned above. The sources are assumed mutually independent, hence $P(\vec{s}) = \prod_{m=1}^{M} P(s_m)$. Assuming Gaussian additive noise with variance σ_{N}^2 the *log likelihood* of the sensor data is given by

$$\log P(\vec{x} \mid \mathbf{A}, \vec{s}) = \log P(\vec{\epsilon}) \propto -\frac{1}{2\sigma_{\mathrm{N}}^2}(\vec{x} - \mathbf{A}\vec{s})^2 \qquad (2)$$

With a probabilistic approach to estimating the basis coefficients (= sources) the most probable decomposition $\hat{\vec{s}}$ of the signal is found by maximizing the posterior distribution of $\vec{s}$

$$\hat{\vec{s}} = \max_{\vec{s}} P(\vec{s} \mid \vec{x}, \mathbf{A}) = \max_{\vec{s}} P(\vec{x} \mid \mathbf{A}, \vec{s}) P(\vec{s}) \qquad (3)$$

A suitable approach for optimizing $\vec{s}$ in the case of finite noise ($\vec{\epsilon} > 0$) and a non-Gaussian prior probability $P(\vec{s})$ is to perform *conjugate gradient ascent* on the log posterior $\log P(\vec{s} \mid \vec{x}, \mathbf{A})$.

2.1. Learning the basis vectors

Suitable basis vectors can be learnt by maximizing the probability of the data given the model

$$P(\vec{x}_1, \ldots, \vec{x}_K \mid \mathbf{A}) = \prod_{k=1}^{K} P(\vec{x}_k \mid \mathbf{A}) \qquad (4)$$

where the likelihood is obtained as an average over all possible states of the model

$$P(\vec{x}_k \mid \mathbf{A}) = \int P(\vec{s}) P(\vec{x}_k \mid \mathbf{A}, \vec{s}) \, \mathrm{d}\vec{s} \qquad (5)$$

For a noisefree model and a complete basis set, ($L = M$), this integral is solvable and leads to the standard ICA learning algorithm [3]. In case of finite noise and an overcomplete basis, performing gradient ascent on the objective function $\log P(\vec{x}_1, \ldots, \vec{x}_K \mid \mathbf{A})$ and using a saddle point approximation of (5) around the inferred mode $\hat{\vec{s}}$ yields the learning rule given by [2]

$$\Delta \mathbf{A} \propto \mathbf{A}\mathbf{A}^{\mathrm{T}} \frac{\partial}{\partial \mathbf{A}} \log P(\vec{x} \mid \mathbf{A}) \approx -\mathbf{A}(\Phi(\hat{\vec{s}}) \, \hat{\vec{s}}^{\mathrm{T}} + \mathbf{I}) \qquad (6)$$

where $\Phi(\hat{s}_m) = \partial \log P(\hat{s}_m) / \partial \hat{s}_m$ is called the *score function*, $\mathbf{I}$ is the identity matrix and $\hat{\vec{s}}$ has to be evaluated at each iteration via (3). The crucial point, then, is the estimation of the unknown source density.

2.2. Estimating the source density

The proper choice of the prior probability of the basis coefficients $P(\vec{s})$ seems to be critical to any success of the decomposition. It is generally assumed that the basis coefficients are sparsely distributed, hence a Laplacian prior is used. If the assumed prior density does not fit the unknown source density reasonably well, i. e. if the latter is sub-Gaussian, for example, a reliable extraction of some independent source signals is no longer possible.

Instead, an appropriate prior can be estimated from the data during the training process. We use a kernel-based density estimation with either Gaussian or Laplacian kernels [4]. The kernel estimator is given as

$$\hat{P}(s_m) = \frac{1}{h_m \, n} \sum_{i=1}^{n} K\left(\frac{s_m - \hat{s}_m^{(i)}}{h_m} \right) \qquad (7)$$

where h_m represents the *bandwidth* of each kernel $K(\ldots)$. It plays the role of a smoothing parameter and may be estimated as [5] $h_m = 1.06 \sigma_m n^{-1/5}$ with σ_m^2 representing the variance of the basis coefficient $s_m(t)$. Starting with an initial Laplacian prior

$$P(s_m) \propto \exp(-\lambda |s_m|) \qquad (8)$$

a new estimate can be obtained from the last n inferred source signals $\hat{s}_m^{(i)}$ $(i = 1, \ldots, n)$. To obtain a good estimate to be used in equation (6), $n = 200$ has been used.

3. EXPERIMENTAL RESULTS

EEG data sets were recorded from 21 scalp electrodes placed according to the International 10-20 System with a sampling rate of 167 Hz. Additional signals from the ECG or EOG placements were not used for this analysis. ICA decomposition was performed on 40–50 sec EEG epochs from the data sets forming the 21 rows of the input matrix. Each of these 21 rows of the input matrix consisted of 7000 data points then. Learning was performed by presenting the columns of the input matrix in consecutive order until 200000 updates using (6) were finished. The learning rate was fixed to 0.00045.

Due to the time consuming density estimation using (7) in conjunction with the optimization of the basis coefficients using (3) learning would take at least several weeks of CPU time on a SUN Ultra 10 workstation. Hence, in this preliminary analysis, simulations within a noise-free model have been performed. The results presented below refer to the square case $L = M$ as well as to the overcomplete case $L \leq M$. In the former $\hat{\vec{s}} = \mathbf{A}^{-1}\vec{x}$ could be used to quickly infer the most probable sources (3). For a non-square mixing matrix $\mathbf{A}$ this had to be replaced by

$$\hat{\vec{s}} = \mathbf{A}^{\dagger}\vec{x}, \qquad (9)$$

where $\mathbf{A}^{\dagger}$ denotes the Moore-Penrose inverse of $\mathbf{A}$. This relation gives the correct value for $\hat{\vec{s}}$ in the case of $\vec{\epsilon} = 0$ and a Gaussian prior, because the solution not only satifies (1) then but has minimal L_2-norm [6], too. Simulations then took between 20 min. and several hours of CPU time.

3.1. ICA Components

Figure 1 shows a 4 sec portion of the recorded EEG time series and figure 2 its corresponding ICA component activations obtained with square case ($L = M$) simulations. The EEG was recorded from a patient with a cancerous Meningeom which has been identified in the left frontal brain region. Three artifactual and two tumor related ICA components may be identified from their signal shapes or characteristic frequencies. Component nr. 4 reveals periodic muscle spiking sampled at the left scalp area mainly (figure 4). Eye blink artifacts in the EEG data are isolated to ICA component nr. 7. Any cardiac contamination in the EEG data

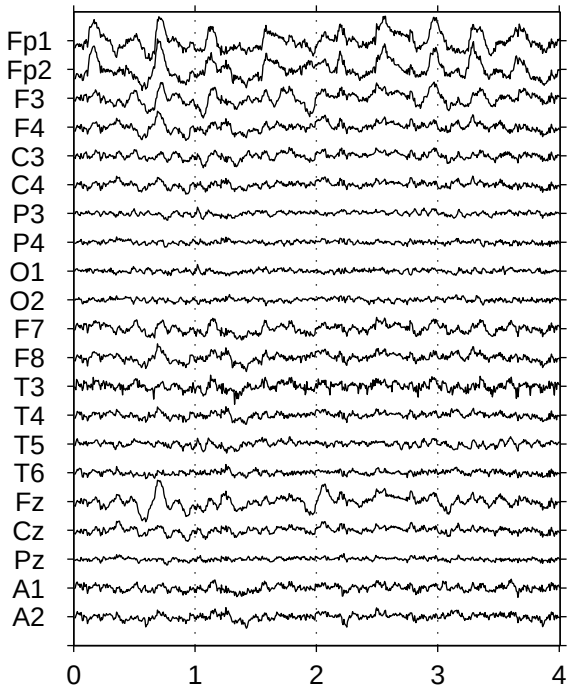

Figure 1: Original EEG signals

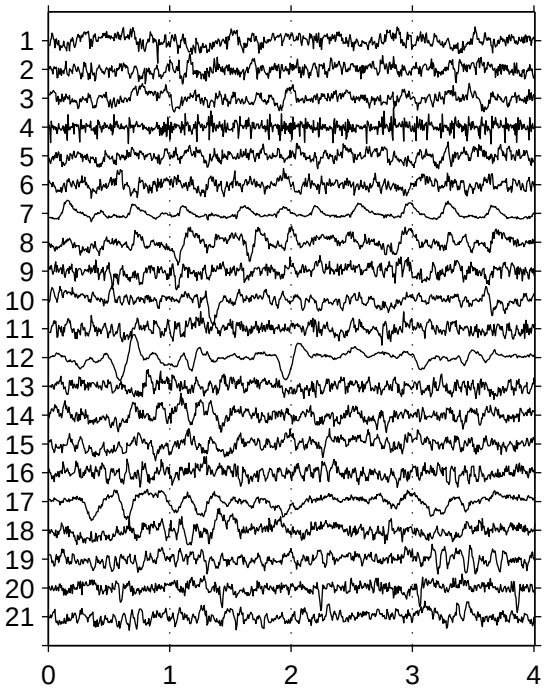

Figure 2: ICA components (square case)

is concentrated in ICA component nr. 20 and shows up mostly in the sensor signals close to both ears.

ICA components nr. 12 and 17 represent δ-waves with a characteristic frequency of 1.5–3.5 Hz as can be seen from estimations of their respective power spectra shown in figure 5. We used a fast fourier transform in conjunction with Welch's method. These oscillations are slightly phase shifted with respect to each other. δ-waves are not observed with normal objects and result from swollen brain regions induced by the spreading malignant tumor. Backprojection to the scalp electrodes indicates (see fig. 4), that at least ICA component nr. 17 stems from the identified cancerous Meningeom, whereas component nr. 12 originates from the central scalp region and is not yet characterized clearly.

The ICA component activations obtained with simulations using an *overcomplete* basis set ($L = 21 \leq M = 26$) are shown in figure 3. Again five distinct ICA components can be identified which correspond to those obtained with a complete basis set. However, one component, the eye blink artifact, appears identically in six different channels. Hence, an overcomplete basis set repeatedly lead to a multiple degenerate solution with no further independent source signals resolved.

Figure 6 presents corresponding results obtained with a normal object for comparison. Using the method of identification described in 3.2 the only ICA com-

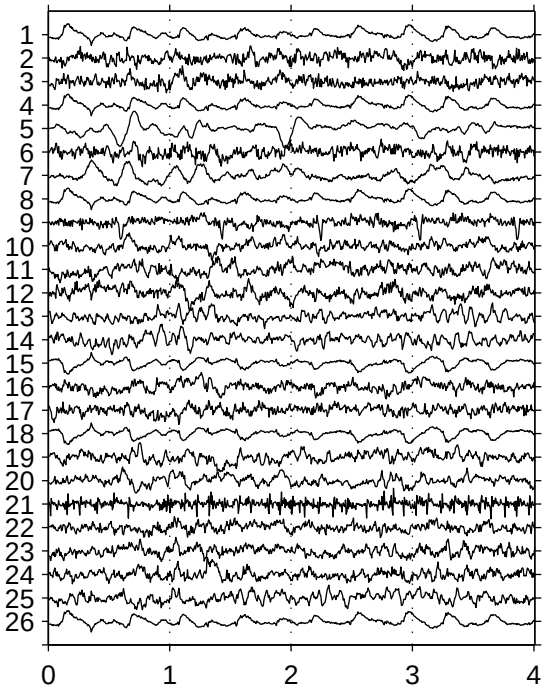

Figure 3: Overcomplete case (26 dim)

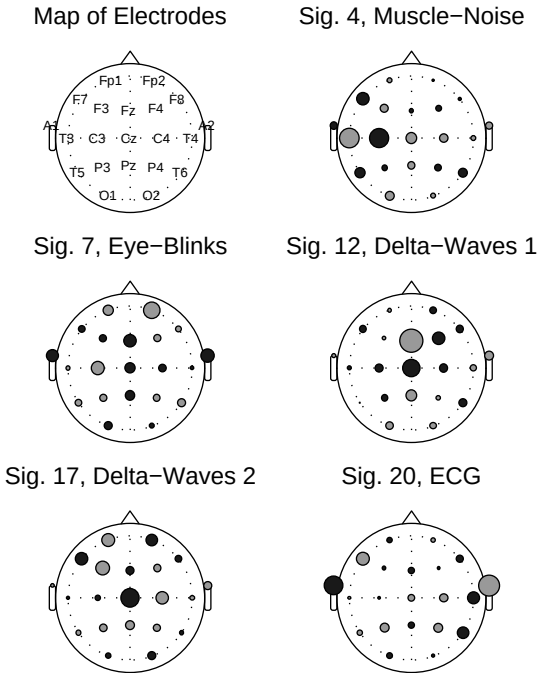

Figure 4: Scalp topography of five source signals. The area of the circles designates the corresponding signal intensities and different gray values encode different signs.

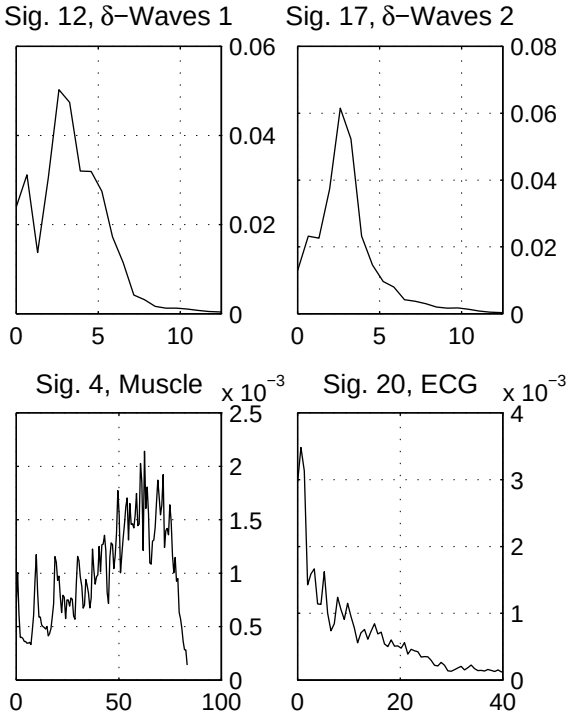

Figure 5: Power spectral densities of some sources (see figure 2). The figures on the abscissae are given in Hertz.

ponents selected robustly by the identification procedure and shown in Figure 6 were the cardiogram (EKG signal nr. 1) and the three artifacts refered to above. Signal nr. 2 contains a prominent potential oscillation caused by an eye blink. The time courses nrs. 3 and 4 stem from muscle noises at two distinct locations. No further characteristic signals corresponding to δ-waves have been identified.

3.2. A Method of Identification of Artifactual and Tumor Related ICA Components

Only few of the ICA components extracted may be identified as artifactual or tumor related signals unequivocally. Means to automatize this identification will be presented next.

After learning has converged starting from randomly initialized matrizes $\mathbf{A}$ only a subset of the basis vectors (columns of the matrix $\mathbf{A}$) remains reproducibly constant, all others vary from trial to trial. Still calculating their mutual information using the estimated densities confirms their statistical independence. After performing the ICA analysis 10–15 times the reproducible basis vectors $\vec{a}_i$ were identified and normalized. To this end the matrix $\mathbf{S} = \mathbf{A}_i^{\mathrm{T}} \cdot \mathbf{A}_j$ has been calculated for all combinations $i < j$. The components of the matrix $\mathbf{S}$ are the dot products of the basis vectors of the matrizes $\mathbf{A}_i$ and $\mathbf{A}_j$. These matrix elements provide a measure of similarity of the corresponding basis vectors. These have been judged similar if their dot product exceeded a threshold value of 0.94.

Table 1 summarizes the results whereby each row corresponds to a matrix $\mathbf{A}_i$ ($i = 1, \ldots, 10$). Within the expression '`B:N`' B gives the position of the column vector and N counts the number of similar basis vectors found in other matrices. The entries in row '`a`' give the suitably normalized number of similar basis vectors extracted. The entries in row '`b`' give the averaged dot product and row '`c`' contains the average mutual information divided by the entropy of one of the corresponding signals. Ideally each of these entries should be close to 1, which is best fulfilled with basis vectors nr. 1, 2, 3, 4, and 7. The example discussed in (3.1) corresponds to the matrix $\mathbf{A}_3$.

Still better results were obtained if, in the square case simulations, the ICA components $\vec{u}^i(t) = \mathbf{A}_i^{-1}\vec{x}(t)$ were compared directly via their related correlation co-

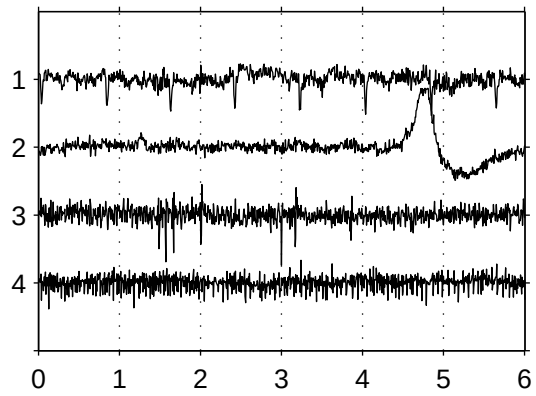

Figure 6: Some of the ICA components of a normal object. A representative epoch of 6 sec is presented.

	1	2	3	4	5	6	7
1	13:9	12:9	7:9	14:7	2:1	0:0	0:0
2	19:9	8:9	9:9	11:9	0:0	18:2	13:6
3	7:9	17:9	12:9	4:9	0:0	10:2	20:4
4	17:9	15:9	18:9	3:9	0:0	14:2	5:4
5	21:9	1:9	19:9	15:9	11:1	0:0	3:4
6	2:9	8:9	17:9	18:8	0:0	0:0	6:4
7	14:9	15:9	3:9	21:9	0:0	0:0	17:6
8	10:9	21:9	12:9	18:9	0:0	0:0	3:6
9	9:9	7:9	16:9	18:8	0:0	0:0	0:0
10	9:9	6:9	3:9	12:9	0:0	0:0	0:0
a	1.00	1.00	1.00	0.96	0.02	0.07	0.38
b	0.99	0.99	0.98	0.98	0.9	0.96	0.96
d	0.81	1.17	1.16	1.19	0.11	0.29	1.18

Table 1: Searching similar vectors in ten different matrices.

efficients

$$C^{ij}_{mm'} = \frac{|\sum_t u^i_m(t) u^j_{m'}(t)|}{\sqrt{\sum_t (u^i_m(t))^2 \sum_t (u^j_{m'}(t))^2}} \qquad (10)$$

Two signals $u^i_m(t)$ and $u^j_{m'}(t)$ may be considered similar if $C^{ij}_{mm'} > 0.9$ holds. After calculating $C^{ij}_{mm'}$ for all combinations $i < j$ $(i, j = 1, \ldots, 15)$ and $m, m' = 1, \ldots, M$ a suitable search algorithm classifies all similar signals $u^i_m(t)$ into different classes. With the example given above 5 different classes were obtained containing 11–15 ICA components each. All others of the 15×21 signals have found no or just by accident only one similar partner. The reason why only 5 ICA components could be isolated reproducibly can be found in the densities of the non-reproducible components which were almost Gaussian. Hence these near Gaussian components could not be separated into independent components by the algorithm.

The upper limit of the entropy of a signal with unit variance is $\frac{1}{2}(1 + \log(2\pi)) \approx 1.42$ which is exactly obtained with a normally distributed signal. The entropy

$$\begin{aligned} H(u(t)) &= -\int \mathrm{d}u\, p(u) \log(p(u)) \\ &\approx -\frac{1}{\sum_t 1} \sum_t \log(p(u)) \end{aligned} \qquad (11)$$

is thus a suitable measure to judge the results obtained with the ICA analysis. Hence any strong deviation of $H(u^i_m(t))$ from $H_{max} = 1.42$ signals an independent source signal with non-Gaussian density. The corresponding calculation of the entropy of the 21 signals presented in Figure 2 yielded the result collected in table 2. Figure 7 exemplifies the estimated densities of two identified source signals (nr. 7 and 12) and of two others (nr. 2 and 6), which seem to be almost normally distributed.

1	2	3	**4**	5	6	**7**
1.42	1.44	1.43	**1.34**	1.43	1.42	**1.22**
8	9	10	11	**12**	13	14
1.42	1.43	1.40	1.43	**1.25**	1.43	1.43
15	16	**17**	18	19	**20**	21
1.42	1.43	**1.31**	1.42	1.42	**1.33**	1.42

Table 2: The odd and even rows refer to m and $H(u^3_m) \pm 0.02$, respectively.

Note that if the data are prewithened, i. e. the covariance matrix of $\vec{x}(t)$ is diagonal, and if the estimated mixing matrices $\mathbf{A}_i$ are orthogonal, then both approaches to automatically identify independent source signals become equivalent.

4. DISCUSSION AND CONCLUSIONS

In order to use EEG recordings to detect and even characterize brain tumors and their respective location artifactual signals have to be removed from the recorded EEG signals. We have presented a flexible ICA algorithm which does not rely on *a priori* assumptions about the unknown source distribution. It simultaneously separates both the EEG and its artifacts into independent components based on the statistics of the data. Once the training is complete, artifact-free EEG records can then be derived by eliminating the contributions of the artifactual sources. Concerning the

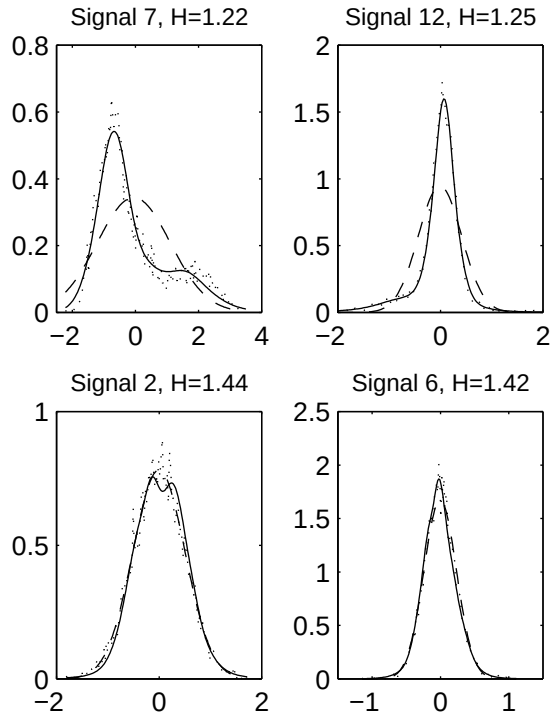

Figure 7: The estimated densities of some source signals (see figure 2). The dots denote the estimated source signal histograms and the solid and dashed lines denote an adaptive density estimation with Gaussian kernels and a fit to a normal distribution with equal variance and appropriate center position. H denotes the signal entropy calculated via (11).

reproducible identification of independent ICA components comparable results have been obtained with complete as well as overcomplete basis sets in a noise-free model.

Comparing results of the ICA analysis starting with randomly initialized mixing matrizes or by calculating the entropy of the isolated components within a single ICA analysis it was possible to identify consistently independent source signals, whether neural or artifactual. But most of the components isolated were almost normally distributed, hence cannot be considered independent source signals and as such are hard to interpret by any physician. Yet we have shown that tumor related EEG signals can be isolated into single independent ICA components. Such signals were not observed in corresponding EEG traces of normal patients. The backprojection of the consistently identified ICA components onto the scalp electrodes allowed a topographic assignment and thus a meaningful interpretation by an experienced physician.

An evaluation of the mutual information of the ICA components extracted from a square case simulation seemed to indicate that there are more sources than mixtures in normal EEG recordings. Yet in applying the generalized overcomplete ICA algorithm discussed above to the data repeatedly resulted in multiply degenerate artifactual ICA components and no further independent components could be identified. Hence an overcomplete representation of the sensor signals did neither result in additional nor in different significant source signals that could have been isolated into the output channels available in a noisefree model. Simulations with the inclusion of additive Gaussian noise in conjunction with a kernel-based density estimation rendering the optimization step (3) very time consuming are currently under way.

5. REFERENCES

[1] S. Makeigh, A. J. Bell, T.-P. Jung, T. J. Sejnowski *Independent Component Analysis of Electroencephalographic Data* Advances in Neural Information Processing Systems 8, D. Touretzky, M. Mozer and M. Hasselmo (Eds.), MIT Press, Cambridge MA, 145-151, 1996

[2] M. S. Lewicki, T. J. Sejnowski *Learning overcomplete representations*, Neural Computation, in press, 1999

[3] J.-F. Cardoso *Blind Signal Separation: Statistical Principles*, Proc.IEEE, 86(10):2009-2025, 1998

[4] B. W. Silverman, *Density Estimation for Statistics an Data Analysis*, Chapman and Hall, London, 1986

[5] A. Taleb, Ch. Jutten *Batch Algorithm for Source Separation in Postnonlinear Mixtures* in J. F. Cardoso, Ch. Jutten, Ph.Loubaton, eds., Proc. ICA '99, 155-159, 1999

[6] A. Ben-Israel, T. N. E. Greville *Generalized inverses: Theory an Applications*, John Wiley & Sons, New York, 1974