INDEPENDENT COMPONENT ANALYSIS OF ELECTROENCEPHALOGRAPHIC AND EVENT-RELATED POTENTIAL DATA

Tzyy-Ping Jung[1], Scott Makeig[2], Anthony J. Bell[1]
and Terrence J. Sejnowski[1]

[1]Computational Neurobiology Lab, The Salk Institute, P.O. Box 85800, San Diego, CA 92186, U.S.A.
[2]Naval Health Research Center, P.O. Box 85122, San Diego, CA 92186, U.S.A.

INTRODUCTION

The *electroencephalogram* (EEG) is a non-invasive measure of brain electrical activity recorded as changes in potential difference between points on the human scalp. Because of volume conduction through cerebrospinal fluid, skull and scalp, EEG data collected from any point on the scalp includes activity from processes occurring within a large brain volume.

Event-related potentials (ERP) are the portion of the EEG both time- and phase-locked to experimental events. Through time-locked averaging across a set of equivalent experimental events, EEG activity, but not both time- and phase-locked to event onsets, is removed by phase cancellation. What remains is called the event-related potential. In general, the problem of determining brain electrical sources from potential patterns recorded on the scalp surface is mathematically underdetermined. For several decades, ERP researchers have proposed a number of techniques to localize the sources of stimulus-evoked potentials, either by assuming that they have a known or simple spatial configuration (Scherg and Von Cramon, 1986), or that generators are restricted to a small subset of possible locations and orientations (Dale and Sereno, 1993).

Recently, we reported a method for separating joint problems of source identification and source localization by applying an information-theoretic neural network algorithm, Independent Component Analysis (ICA), to multiple-channel EEG and ERP activity recorded from the scalp (Makeig et al., 1996a; S. Makeig, T-P Jung, A.J. Bell, and T.J. Sejnowski, unpublished observations). The algorithm tells **what** temporally independent activations compose the observed scalp recording, without specifying **where** in the brain these activations arise. Here, we report an application of the ICA algorithm to the analysis of 14-channel EEG and ERP data recorded during eyes-closed performance of a sustained auditory detection task (Makeig and Inlow, 1993).

Central Auditory Processing and Neural Modeling
Edited by Poon and Brugge, Plenum Press, New York, 1998

INDEPENDENT COMPONENT ANALYSIS

The goal of *blind source separation* in signal processing is to recover independent source signals, $s_1(t),\ldots, s_N(t)$, (e.g., different people speaking, music etc.) after they are linearly mixed by an unknown matrix **A**, and recorded at N sensors, $x_1(t),\ldots, x_N(t)$. The blind source separation problem has been studied by researchers in the neural network (Bell and Sejnowski, 1995a; Amari et al., 1996, Cichocki et al., 1994; Girolami and Fyfe, 1996; Karhunen et al; 1996; Pearlmutter and Parra, 1997; Roth and Baram, 1996) and statistical signal processing communities (Cardoso and Laheld, 1996; Comon, 1994; Lambert, 1996; Pham, 1996; Yellin and Weinstein, 1996). Comon (1994) defined the concept of independent component analysis (ICA) as maximizing the degree of statistical independence among outputs using contrast functions approximated by the Edgeworth expansion of the Kullback-Leibler divergence. In contrast with decorrelation techniques such as Principal Component Analysis (PCA), which ensures that output pairs are uncorrelated:

$$\langle u_i u_j \rangle = 0, \forall i \neq j,$$

ICA imposes the much stronger criterion that the multivariate probability density function (p.d.f.) of **u** factorizes:

$$p(\mathbf{u}) = \prod_{i=1}^{N} p(u_i)$$

Finding such a factorization requires that the mutual information between the u_i go to zero: $I(u_i, u_j)=0$, for all pairs u_i and u_j. Mutual information depends on all higher-order statistics of the u_i while decorrelation only takes account of 2nd-order statistics.

Infomax Algorithm

Recently, Bell and Sejnowski (1995a) have proposed a simple neural network algorithm, or *infomax* algorithm, for carrying out ICA. The algorithm finds an *unmixing* matrix for the observed data by maximizing the joint entropy of the random vectors resulting from a linear transformation of the mixed signals, $x_i(t)$, followed by a non-linearity, $g()$, which in general allows minimization of statistical dependencies among the input variables. There are, however, cases where the ICA algorithm will not reach the minimum (see Bell and Sejnowski, 1995a). It has been shown that the joint entropy of the linearly unmixed and nonlinearly squashed variables has its maximum value when the non-linearity is the cumulative density function of the sources (Nadal and Parga, 1994; Bell and Sejnowski, 1995b).

Infomax derives, by stochastic gradient ascent, a matrix, **W**, which maximizes the entropy (Cover and Thomas, 1991), $H(y)$, of an ensemble of 'sphered' input vectors $\mathbf{x}_{sph}$, linearly transformed and sigmoidally smoothed ($\mathbf{u}=\mathbf{W}\mathbf{x}_{sph}$, $\mathbf{y}=g(\mathbf{u})$). The 'unmixing' matrix **W** performs component separation, while the sigmoidal nonlinearity $g()$, provides higher-order statistical information through its Taylor series expansion. Initial sphering of the (zero-mean) input data (Bell and Sejnowski, 1996), ($\mathbf{x}_{sph} = \mathbf{Z}\mathbf{x}$, where $\mathbf{Z} = 2\langle \mathbf{x}\mathbf{x}^T \rangle^{-1/2}$) speeds convergence. **W** is initialized to the identity matrix (**I**) and iteratively adjusted using small batches of data vectors (normally 10 or more) drawn randomly from $\mathbf{x}_{sph}$ without substitution, according to:

$$\Delta \mathbf{W} = \varepsilon \frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} \mathbf{W}^T \mathbf{W} = \varepsilon (\mathbf{I} + \mathbf{y}' \mathbf{u}^T) \mathbf{W}$$

where e is the learning rate (normally < 0.01) and vector $\mathbf{y}'$ has elements:

$$y' = \frac{\partial}{\partial u_i} \ln(\partial y_i / \partial u_i).$$

More details appear in Bell and Sejnowski, (1995a, 1995b). The ($\mathbf{W}^T \mathbf{W}$) 'natural gradient' term in the update equation (Amari et al., 1996; Cichocki et al., 1994) avoids matrix inversions and speeds convergence by normalizing the variance in all directions.

We use the logistic nonlinearity, $g(u_i) = (1+exp(-u_i))^{-1}$, which gives a simple update rule, $y_i' = 1-2y_i$ and biases the algorithm toward finding sparsely-activated or super-Gaussian independent components with high kurtosis (Olshausen, 1996).

The ICA algorithm is easily implemented and computationally efficient. Because the algorithm uses parametric probability density estimation, the number of data points needed for the method to converge may be as few as several times the number of input channels, which in turn must be at least equal to the number of components to be separated.

Assumptions of ICA Applied to EEG/ERP Data

The aforementioned ICA algorithm appears very effective (Bell and Sejnowski, 1995a, 1996) for performing source separation in domains where, (1) the sources are independent, (2) the propagation delays in the mixing medium are negligible, (3) the sources have super-Gaussian (kurtosis > 0) distributions, and (4) the number of independent signal sources is the same as the number of sensors, meaning if we employ N sensors, the ICA algorithm can separate N sources.

In the case of EEG signals, N scalp electrodes pick up correlated signals and we would like to know what effectively independent components generated these mixtures. If we assume that the complexity of EEG dynamics can be modeled, at least in part, as a collection of a modest number of statistically independent brain processes, the EEG source analysis problem satisfies ICA assumption (1). Since volume conduction in brain tissue is effectively instantaneous, assumption (2) is also satisfied. Assumption (3) is also plausible, but assumption (4), that the EEG is a linear mixture of exactly N sources, is questionable, since we do not know the effective number of statistically independent brain signals contributing to the EEG recorded from the scalp.

We have run a number of numerical simulations in which 600-point signals recorded from the cortex of a patient during preparation for operation for epilepsy were projected to simulated scalp electrodes through a three-shell spherical head model (Makeig et al., 1996b). We used electrocorticographic (ECoG) data in these simulations as a plausible best approximation to the temporal dynamics of the unknown ERP brain generators. Results confirmed that the ICA algorithm could accurately identify the activation waveforms and scalp topographies of relatively large and more temporally-independent ECoG signals from the simulated scalp recordings, even in the presence of multiple low-level and temporally independent sources (synthesized from ECoG data, or from uniformly-distributed or Gaussian noise).

The rows of the output data matrix, $\mathbf{u}$, are the activation waveforms of the ICA components, while the columns of the inverse matrix, $(\mathbf{WZ})^{-1}$, of the overall transformation, $\mathbf{WZ}$, give the projection strengths of the respective components onto the scalp sensors. The data accounted for by the ith component is the outer product of the ith component activation, u_i, with the ith column of the inverse matrix, $(\mathbf{WZ})_i^{-1}$. Scaling information is distributed between the two, hence relative component strengths can only be compared via their projections $(\mathbf{WZ})_i^{-1} u_i$.

METHODS

EEG and behavioral data were collected to develop a method of objectively monitoring the alertness of operators listening for weak signals (Makeig and Inlow, 1993). Ten adult volunteers participated in three or more half-hour sessions during which they pushed one button whenever they detected an above-threshold auditory target stimulus (a brief increase in the level of the continuously-present background noise). To maximize the chance of observing alertness decrements, sessions were conducted in a small, warm, and dimly-lit experimental chamber, and subjects were instructed to keep their eyes closed. Auditory targets were 350 ms increases in the intensity of a 62 dB white noise background, 6 dB above their threshold of detectability, presented at random time intervals at a mean rate of 10/min, and superimposed on a continuous 39-Hz click train evoking a 39-Hz *steady-state response* (SSR). Short, and task-irrelevant probe tones of two frequencies (568 and 1098 Hz) were interposed between the target noise bursts at 2-4 s intervals. EEG was collected from thirteen electrodes located at sites of the International 10-20 System, referred to the right mastoid, at a sampling rate of 312.5 Hz. A bipolar diagonal electrooculogram (EOG) channel was also recorded for use in eye movement artifact correction and rejection. Hits were defined as targets responded to within a 100-3000 ms window, while Lapses were targets not responded to (because of drowsiness or loss of vigilance). Two sessions each from three subjects were selected for analysis based on their containing at least 50 response Lapses. A continuous performance measure, local error rate, was computed by convolving the irregularly-sampled performance index time series (Hit=0/Lapse=1) with a 95-sec smoothing window advanced through the data in 1.64 sec steps.

The ICA algorithm in (1) was applied to the 14-channel, 28 minute EEG data. The time index was permuted before each training step to ensure signal stationarity, and the 14-dimensional input vectors were presented to a 14→14 ICA network. To speed convergence, we first pre-whitened the data to remove first- and second-order statistics. The learning rate was annealed from 0.03 to 0.0001 during convergence. After each pass through the training data, we checked the remaining correlation between the ICA output channels and change in the weight matrix, and stopped the training procedure when the correlation among all channel pairs was below 0.05 and the ICA weights had stopped changing appreciably.

For the ERP analysis, 154 hit and 107 lapse responses were averaged. The ICA algorithm was used to decompose simultaneously these two 1-sec (312 points) ERPs into 14 ICA components.

RESULTS

EEG Results

A 6-sec portion of the EEG time series and corresponding ICA components activations and scalp maps and shown in Fig. 1. As expected, correlations between the ICA components are close to zero (range $0 < |r| < 0.05$). The ICA components were little changed by retraining using different learning rates, data presentation orders, or initial conditions. Several observations on the ICA components in Fig. 1 are of interest:

- ICA component 1 (*right panel*) contains near-DC changes associated with slow eye movements captured in the EOG and frontal EEG channels (note the pre-frontal scalp distribution of this component).
- The eye movement artifact at 0.4 sec in the EEG data (*left panel*) is isolated to component 1 (*right panel*) leaving all the other ICA components free of this artifact.

- The dominant 8 Hz waves at 2 sec and 5.5 sec are spread across many EEG channels, but are more or less isolated to ICA components 3, 7, and 12, each having a different latency, phase, and/or duration, and a different scalp distribution (*head plots*). The activation of component 7 leads that of component 12 throughout the session.
- Alpha activity (near 10 Hz) not obvious in the EEG data is revealed in ICA components 2 and 14, which both here and throughout the session contain alpha bursts interspersed with quiescent periods.
- ICA components 6, 8 and 10 contain mostly line noise (60 Hz), while ICA components 4 and 5 have a broader high frequency (50-100 Hz) spectrum. The spectral characteristics and scalp distributions of these components (*head plots*) suggest that they most likely represent high-frequency activity generated by scalp muscles.
- Other ICA components (9 & 11) contain mixed oscillatory activity which is not easy to characterize.

Nonstationarity

Figure 2 demonstrates nonstationarity in the ICA decomposition of the EEG data with changes in dynamic brain state. ICA training minimizes mutual information among transformed input data, and therefore also minimizes correlations between them. However, when an ICA

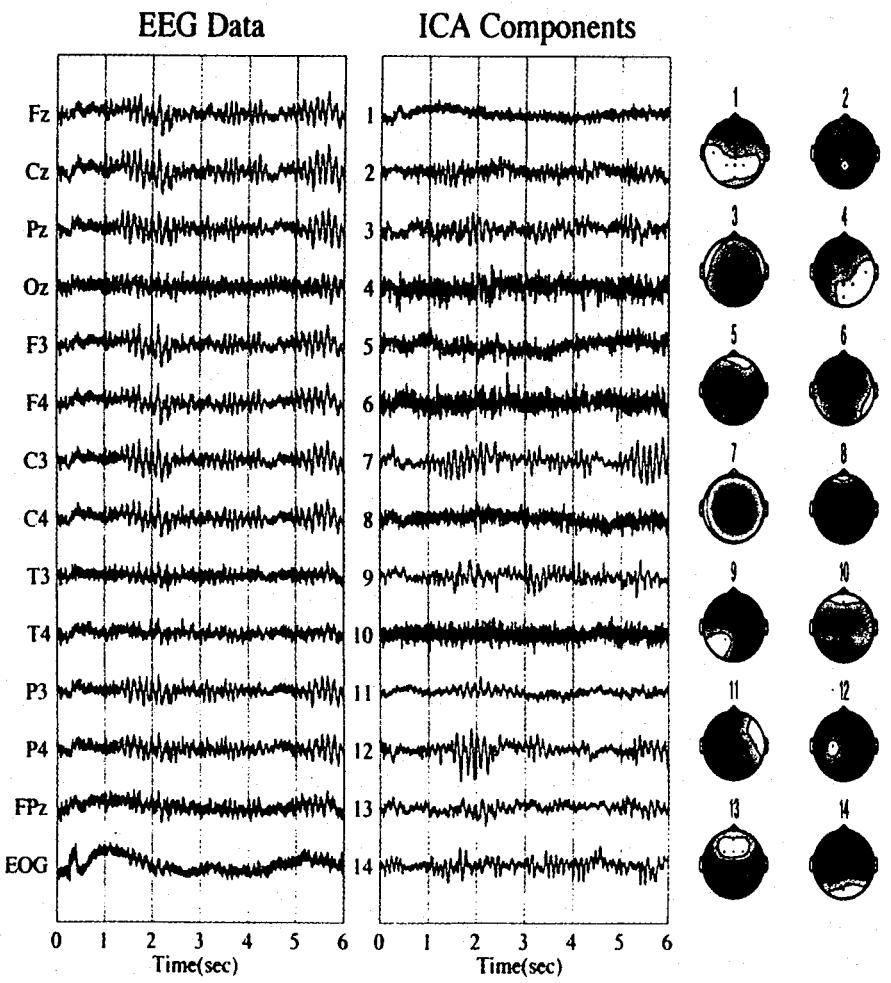

Figure 1. Left: 6 seconds of 14-channel EEG data. Right: ICA component activations and scalp maps for the same data.

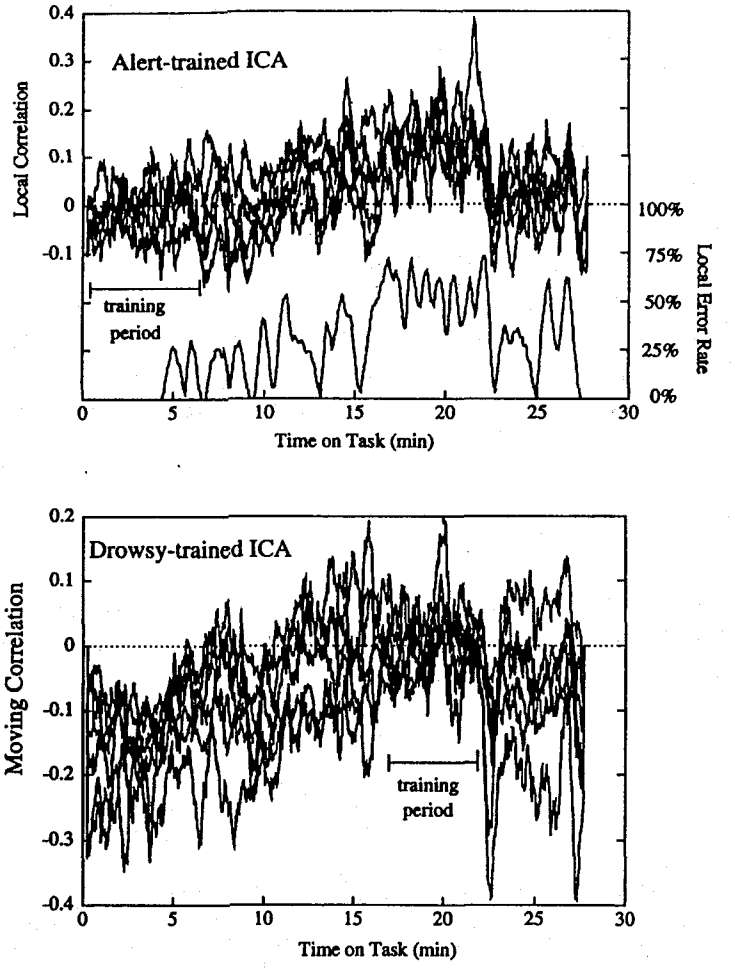

Figure 2. *Upper panel*: Moving correlations (in a 34-s square window) between 6 selected ICA output channel-pairs throughout a 28-minute session in which the subject performed a continuous auditory detection task. The ICA weight matrix was trained on a 6.5 min period from the beginning of the session. Changes in the subject's local detection error rate during the session are shown in the bottom trace. Note the low initial correlations between the ICA channels, the gradual introduction of correlation during the central drowsy portion of the session, and the return to low correlations at minute 23, concurrent with a sharp drop in the subject's error rate. *Lower panel*: Moving correlations between 7 selected ICA channel pairs filtered using an ICA weight matrix trained on 6.5 min from the drowsy (high error) portion of the run. Note the sharp increases in residual correlation during minutes 23 and 27, when the subject's error rate fell sharply, and the parallels between changes in correlation in the left and right panels.

unmixing matrix was trained on data from the first 6 minutes of one session, a period during which the subject's performance was nearly perfect, and this matrix was then used to filter EEG data from the remaining portion of the session, a period in which the subject became drowsy and began to fail to respond to targets, the ICA-filtered outputs became more correlated. When the subject regained alertness (at min 23), the residual correlations between the ICA outputs reverted to their initial levels of (de)correlation. Conversely, filtering data from the whole session using an ICA weight matrix trained on the drowsy portion of the session produced component activations that were more correlated during the alert portions of the session than during the drowsy training period. Presumably, these changes in residual correlation between ICA output channels reflect changes in the dynamics and topographic structure of the EEG signals between alert and drowsy brain states, and would be used to predict the level of vigilance in the subject.

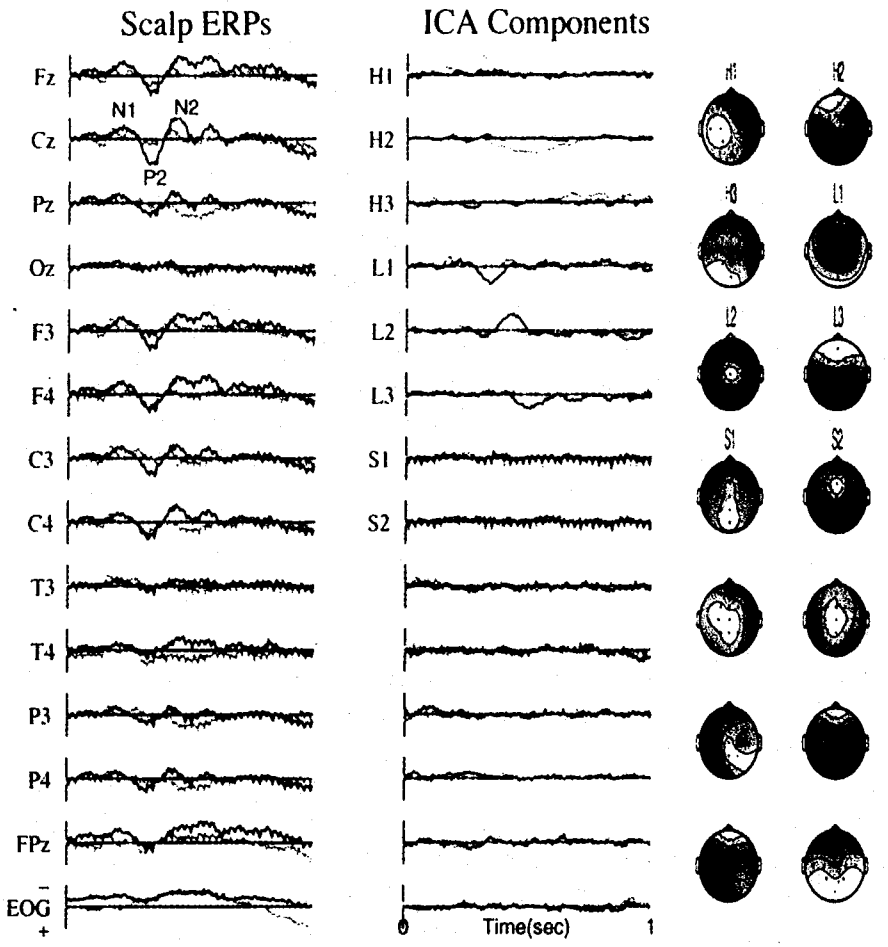

Figure 3. *Left panel*: One-second event-related potentials averaging responses to undetected (*bold traces*) and detected (*faint traces*) noise targets during two half-hour sessions from one subject. *Right panel*: ICA component activations and scalp maps for this data.

ERP Results

An important problem in human electrophysiology is to identify objectively overlapping ERP subcomponents. Figure 3 shows an ICA decomposition (*right panel*) of ERPs (*left panel*) to detected (Hit) and undetected (Lapse) targets by one subject. The ERPs contain the standard auditory response peaks N1, P2, and N2, although the N1 peak is indistinct, most probably because of the long rise time of the noise-burst stimulus and the variable noise background. As expected from sleep studies of auditory evoked responses (Van Sweden et al., 1994), the P2 and N2 peaks were larger and had longer latencies in response to undetected targets. The detected-target response also had a parietal P3 component (quite small in this subject), and both responses contained a robust 39-Hz SSR in all channels. The EOG channel showed some residual ocular activity spreading into frontal sites (e.g. Fpz).

The ICA algorithm was trained simultaneously on the two 1-sec ERPs averaging responses to detected and undetected targets, producing two components (S1 & S2) separating out the 39-Hz SSR induced by the continuous 39-Hz click stimulation during the session. Note the stimulus-induced perturbation in SSR amplitude previously identified (Makeig and Galambos, 1989). The scalp distribution of the SSR appears to sweep from the front to the back of the scalp every cycle. ICA decomposed these apparent movements into the sum of

spatially-fixed anterior- and posterior-bilateral components with different phase lags. ICA also accounted for the transient perturbations in SSRs induced by experimental events using the same components producing the SSR, supporting the hypothesis that these transient perturbations represent modulation of the ongoing response (Makeig and Galambos, 1989).

Six of the ICA components (H1, H2, H3, L1, L2, & L3) were active in a single 50-300 msec interval in one of the response conditions. Three channels (H1, H2, & H3) passed briefly-activated components of the detected target response, while three others (L1, L2, & L3) components of the (larger) undetected-target response. The activation time course and scalp map of ICA component H2 accounted for the parietal P3 activity. The scalp maps of the components contained one or two spatial extrema. We suggest these ICA components represent focal or distributed components of evoked brain response activity, and may represent a solution to the longstanding problem of objectively decomposing evoked responses into neurobiologically meaningful, temporally overlapping subcomponents (S. Makeig, T-P Jung, A.J. Bell, and T.J. Sejnowski, unpublished observations).

CONCLUSIONS

ICA decomposition opens a new and potentially useful window into complex EEG and ERP data that can complement other analysis techniques. ICA can isolate various kinds of artifacts to a few components while removing them from remaining components. These may in turn represent the scalp maps and time courses of activity in long lasting or transient brain processes on which the algorithm converges reliably. Measures of nonstationarity of the ICA unmixing matrix may be useful for observing brain state changes. Applied to two 1-sec averaged ERP data, the algorithm derived eight components that decomposed each of the major response peaks and the accompanying auditory steady-state response into one or more ICA components with relatively simple scalp distributions. Although it may be difficult to locate the generators of ICA components within the brain on the basis of their time courses and scalp projections, ICA decomposition might nonetheless prove useful for preprocessing data prior to applying source localization algorithms. Future work includes exploring the strengths and limitations of the method, determining the proper number of input channels, and interpreting the physiological and/or psychophysiological significance of the derived components.

ACKNOWLEDGMENTS

This research was supported by the Department of the Navy, Naval Research and Development Command, Bethesda, Maryland under *ONR.WR.30020(6429)* to Dr. Makeig and *ONR N00014-91-J-1674* to Dr. Sejnowski. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of Defense, or the U.S. Government.

REFERENCES

Amari S, Cichocki A, Yang H (1996) A new learning algorithm for blind signal separation. In: Advances in Neural Information Processing Systems (Touretzky D, Mozer M, Hasselmo M, eds), 8:757-763.

Bell AJ, Sejnowski TJ (1995a) An information-maximization approach to blind separation and blind deconvolution. Neural Computation 7:1129-59.

Bell AJ, Sejnowski TJ (1995b) Fast blind separation based on information theory. In: Proc 1995 Intern Symp on Nonlinear Theory and Applications (NOLTA) 1:43-47.

Bell AJ, Sejnowski TJ (1996) Learning the higher-order structure of a natural sound. Network: Computation in Neural Systems 7:261-7.
Cardoso JF, Laheld B (1996) Equivariant adaptive source separation. IEEE Trans Sig Proc 45:434-44.
Cichocki A, Unbehauen R, Rummert E (1994) Robust learning algorithm for blind separation of signals. Electronics Letters 30:1386-1387.
Comon P (1994) Independent component analysis—a new concept? Sig Proc 36:287-314.
Cover TM, Thomas JA (1991) Elements of Information Theory. New York: Wiley.
Dale AM, Sereno MI (1993) Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction - a linear approach. J Cogn Neurosci 5:162-176.
Girolami M, Fyfe C (1996) Negentropy and kurtosis as projection pursuit indices provide generalized ICA algorithms. In: Advances in Neural Information Processing Systems (Jordan M, Mozer M, Petsche T, eds) 9.
Karhunen J, Oja E, Wang L, Vigario R, Joutsensalo J (1995) A class of neural networks for independent component analysis. Report A28 , Helsinki Univ. of Technology.
Lambert R (1996) Multichannel blind deconvolution: FIR matrix algebra and separation of multipath mixtures. Thesis: Department of Electrical Engineering, University of Southern California.
Makeig S, Galambos R (1989) The CERP: Event-related perturbations in steady-state responses. In: Brain Dynamics: Progress and Perspectives (Basar E, Bullock TH, eds), pp. 375-400. Berlin:Springer-Verlag.
Makeig S, Inlow M (1993) Lapses in alertness: Coherence of fluctuations in performance and EEG spectrum. EEG Clin Neurophysiol: 86:23-35.
Makeig S, Bell AJ, Jung T-P, Sejnowski TJ (1996a) Independent Component Analysis of Electroencephalographic Data. In: Advances in Neural Information Processing Systems (Touretzky D, Mozer M, and Hasselmo M, eds) 8:145-51.
Makeig S, Jung T-P, Ghahremani D, Sejnowski TJ (1996b) Independent Component Analysis of Simulated ERP Data. Tech. Rep. INC-9606, Institute for Neural Computation, University of California, San Diego, CA.
Nadal JP, Parga N (1994) Non-linear neurons in the low noise limit: a factorial code maximises information transfer. Network 5:565-581.
Nunez PL (1981) Electric Fields of the Brain. New York: Oxford.
Olshausen B (1996) C.B.C.L. Paper 138. Dept. of Brain and Cognitive Sciences, MIT.
Pearlmutter B, Parra L (1997) Maximum likelihood blind source separation: a context-sensitive generalization of ICA. In: Advances in Neural Information Processing Systems (Touretzky D, Mozer M, and Hasselmo M, eds) 9:613-619.
Pham DT (1996) Blind separation of instantaneous mixture of sources via an independent component analysis. IEEE Trans Sig Proc 44:2768-2279.
Roth Z, Baram Y (1996) Multidimensional density shaping by sigmoids. IEEE Trans Neural Networks 7:1291-1298.
Scherg M, Von Cramon D (1986) Evoked dipole source potentials of the human auditory cortex. EEG Clin Neurophysiol 65:344-601.
Van Sweden B, Van Dijk JG, Caekebeke JF (1994) Auditory information processing in sleep: late cortical potentials in an oddball paradigm. Neuropsychobiology 29:152-156.
Yellin D, Weinstein E (1996) Multichannel signal separation: Methods and analysis. IEEE Trans Sig Proc 44:106-118.