Psychophysiology, 37 (2000), 163–178. Cambridge University Press. Printed in the USA.

Removing electroencephalographic artifacts by blind source separation

TZYY-PING JUNG,[a,b] SCOTT MAKEIG,[b,c] COLIN HUMPHRIES,[a] TE-WON LEE,[a,b]
MARTIN J. MCKEOWN,[a] VICENTE IRAGUI,[b] AND TERRENCE J. SEJNOWSKI[a,b]

[a]Howard Hughes Medical Institute and Computational Neurobiology Laboratory, The Salk Institute, San Diego, California, USA
[b]University of California San Diego, La Jolla, USA
[c]Naval Health Research Center, San Diego, California, USA

Abstract

Eye movements, eye blinks, cardiac signals, muscle noise, and line noise present serious problems for electroencephalographic (EEG) interpretation and analysis when rejecting contaminated EEG segments results in an unacceptable data loss. Many methods have been proposed to remove artifacts from EEG recordings, especially those arising from eye movements and blinks. Often regression in the time or frequency domain is performed on parallel EEG and electrooculographic (EOG) recordings to derive parameters characterizing the appearance and spread of EOG artifacts in the EEG channels. Because EEG and ocular activity mix bidirectionally, regressing out eye artifacts inevitably involves subtracting relevant EEG signals from each record as well. Regression methods become even more problematic when a good regressing channel is not available for each artifact source, as in the case of muscle artifacts. Use of principal component analysis (PCA) has been proposed to remove eye artifacts from multichannel EEG. However, PCA cannot completely separate eye artifacts from brain signals, especially when they have comparable amplitudes. Here, we propose a new and generally applicable method for removing a wide variety of artifacts from EEG records based on blind source separation by independent component analysis (ICA). Our results on EEG data collected from normal and autistic subjects show that ICA can effectively detect, separate, and remove contamination from a wide variety of artifactual sources in EEG records with results comparing favorably with those obtained using regression and PCA methods. ICA can also be used to analyze blink-related brain activity.

Descriptors: Independent component analysis, ICA, EEG, Artifact removal, EOG

Eye movements, eye blinks, muscle noise, heart signals, and line noise often produce large and distracting artifacts in electroencephalographic (EEG) recordings. Asking subjects to fixate a visual target may reduce voluntary eye movements (blinks and saccades) in cooperative subjects during brief EEG sessions, but fixation does not eliminate involuntary eye movements and cannot be used when the subject is performing a task requiring eye movements. Rejecting EEG segments with artifacts larger than an arbitrarily preset value is the most commonly used method for dealing with artifacts in research settings. However, when limited data are available, or blinks and muscle movements occur too frequently, as in some patient groups, the amount of data lost to artifact rejection may be unacceptable.

Several proposed methods for removing eye-movement artifacts are based on regression in the time domain (Gratton, Coles, & Donchin, 1983; Hillyard & Galambos, 1970; Verleger, Gasser, & Möcks, 1982) or frequency domain (Whitton, Lue, & Moldofsky, 1978; Woestenburg, Verbaten, & Slangen, 1983). However, simple regression in the time domain for removing eye artifacts from EEG channels tends to overcompensate for blink artifacts and may introduce new artifacts into EEG records (Weerts & Lang, 1973; Oster & Stern, 1980; Peters 1967). The cause of this overcompensation is the difference between the electrooculographic (EOG)-to-EEG transfer functions for blinks and saccades. Saccade artifacts arise from changes in orientation of the retinocorneal dipole, whereas blink artifacts arise from alterations in ocular conductance produced by contact of the eyelid with the cornea (Overton & Shagass, 1969). The pickup of blink artifacts on the recording electrodes decreases rapidly with distance from the eyes, whereas the transfer of saccade artifacts decreases more slowly, so that at the vertex the effect of saccades on the EEG is about double that of blinks (Overton & Shagass, 1969).

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.

This report was supported in part by a Swartz Foundation grant to Drs. Jung and Sejnowski, a Howard Hughes Medical Institute grant to Dr. Sejnowski, a grant ONR.Reimb.30020.6429 from the Office of Naval Research to Dr. Makeig, and the Heart and Stroke Foundation of Ontario to Dr. McKeown.

We are grateful to discussions with Dr. Robert Galambos. We thank Drs. Eric Courchesne and Jeanne Townsend for providing ERP data from normal and autistic subjects.

Address reprint requests to: Tzyy-Ping Jung, Ph.D., Institute for Neural Computation, UCSD, 9500 Gilman Dr., Dept. 0523, La Jolla, CA 92093-0523, USA. E-mail: jung@inc.ucsd.edu.

Regression in the frequency domain (Whitton et al., 1978; Woestenburg et al., 1983) can account for frequency-dependent transfer function differences from EOG to EEG, but is acausal and thus unsuitable for real-time applications. Kenemans, Molenaar, Verbaten, and Slangen (1991) proposed a time domain multiple-lag regression method capable of taking into account frequency- and phase-dependent differences in EOG-to-EEG transfer functions. Their method can be viewed as a causal time-domain equivalent of frequency-domain methods. However, the method requires considerably more computation than its frequency-domain counterpart, and was not found to be better than simple time-domain regression in tests on actual EEG data (Kenemans et al., 1991). Regression methods in either time or frequency domain depend on having a good regressing channel (e.g., EOG), and share an inherent weakness that spread of excitation from eye movements and EEG signals is bidirectional. Therefore, whenever regression-based artifact removal is performed, relevant EEG signals contained in the EOG channel(s) are also cancelled out in the "corrected" EEG channels along with the eye movement artifacts. The same problem complicates removal of other types of EEG artifacts. For example, good reference channels for each of the muscles making independent contributions to EEG muscle noise are not usually available.

Berg and Scherg (1991a) have proposed a method of eye-artifact removal using a spatiotemporal dipole model that requires a priori assumptions about the number of dipoles for saccade, blink, and other eye movements, and assumes they have a simple dipolar structure. The major limitations of this method are that the inaccuracies in the dipole model might lead to inaccuracies in the locations of the sources and in the contributions from EOG to EEG (Lins, Picton, Berg, & Scherg, 1993). Berg and Scherg (1991b) then proposed another technique for removing ocular artifacts using principal component analysis (PCA). First, they collected EEG and EOG signals simultaneously while the subject performed some standard eye movements and blinks. Then, a PCA of the variance in these "calibration signals" gave major components representing blinks and horizontal and vertical eye movements. "Corrected" EEG data could be obtained by removing these components through the simple inverse computation. They showed that ocular artifacts can be removed more effectively by the PCA method than by regression or by using spatiotemporal dipole models. However, Lagerlund, Sharbrough, and Busacker (1997) showed that PCA methods cannot completely separate some artifacts from cerebral activity, especially when they both have comparable amplitudes.

Most EEG correction techniques focus on removing ocular artifacts from the EEG, and relatively little work has been done on removing other artifacts such as muscle activity, cardiac signals, electrode noise, and so on. Regressing out muscle noise is impractical because signals from multiple muscle groups require different reference channels. Line noise is most commonly filtered out in the frequency domain. However, when the 50-Hz or 60-Hz line frequency overlaps the spectrum of high-frequency EEG phenomena of interest, some other approach is needed.

Makeig, Bell, Jung, and Sejnowski (1996) proposed an approach to the analysis of EEG data based on a new unsupervised neural network learning algorithm, independent component analysis (ICA) of Bell and Sejnowski (1995). They showed that the ICA algorithm can be used to separate neural activity from muscle and blink artifacts in spontaneous EEG data and reported its use for finding components of EEG and event-related potentials (ERP) and tracking changes in alertness (Makeig et al., 1996; Jung, Makeig, Bell, & Sejnowski, 1997). Subsequent independent work (Vigário, 1997) based on a related approach also verified that different artifacts can be detected from multichannel magnetoencephalographic (MEG) recordings. However, this study did not try to remove the identified artifacts.

We present here a generally applicable method for isolating and removing a wide variety of EEG artifacts by linear decomposition using a recently developed extension of the ICA algorithm (Bell & Sejnowski, 1995). The extended algorithm (Lee & Sejnowski, 1997) separates sources that have either super-Gaussian or sub-Gaussian amplitude distributions, allowing line noise, which is sub-Gaussian, to be focused efficiently into a single source channel and removed from the data. ICA methods are based on the assumptions that the signals recorded on the scalp are mixtures of time courses of temporally independent cerebral and artifactual sources, that potentials arising from different parts of the brain, scalp, and body are summed linearly at the electrodes, and that propagation delays are negligible. The method uses spatial filters derived by the ICA algorithm, and does not require reference channels for each artifact source. Once the independent time courses of different brain and artifact sources are extracted from the data, "corrected" EEG signals can be derived by eliminating the contributions of the artifactual sources. We analyze experimental data containing a wide variety of artifacts to demonstrate the effectiveness of the method, and compare results with those of regression and PCA.

ICA

ICA (Comon, 1994) was originally proposed to solve the *blind source separation* problem, to recover independent source signals, $\mathbf{s} = \{s_1(t), \ldots, s_N(t)\}$, (e.g., different voice, music, or noise sources) after they are linearly mixed by an unknown matrix **A**. Nothing is known about the sources or the mixing process except that there are N different recorded mixtures, $\mathbf{x} = \{x_1(t), \ldots, x_N(t)\} = \mathbf{As}$. The task is to recover a version, $\mathbf{u} = \mathbf{Wx}$, of the original sources, **s**, identical save for scaling and permutation, by finding a square matrix, **W**, specifying spatial filters that invert the mixing process linearly. Bell and Sejnowski (1995) proposed a simple neural network algorithm that blindly separates mixtures, **x**, of independent sources, **s**, using information maximization (infomax). They showed that maximizing the joint entropy, $H(\boldsymbol{y})$, of the output of a neural processor minimizes the mutual information among the output components, $y_i = g(u_i)$, where $g(u_i)$ is an invertible bounded nonlinearity and $\mathbf{u} = \mathbf{Wx}$. Recently, Lee, Girolami, and Sejnowski (1999) extended the ability of the infomax algorithm to perform blind source separation on linear mixtures of sources having either sub- or super-Gaussian distributions (for further details, see the Appendix).

Applying ICA to Artifact Correction

The ICA algorithm is highly effective at performing source separation in domains where (1) the mixing medium is linear and propagation delays are negligible, (2) the time courses of the sources are independent, and (3) the number of sources is the same as the number of sensors; that is, if there are N sensors, the ICA algorithm can separate N sources (Makeig et al., 1996). In the case of EEG signals, we assume that the multichannel EEG recordings are mixtures of underlying brain and artifactual signals. Because volume conduction is thought to be linear and instantaneous, assumption (1) is satisfied. Assumption (2) is also reasonable because the sources of eye and muscle activity, line noise, and cardiac signals are not generally time locked to the sources of EEG activity which is thought to reflect synaptic activity of cortical neurons. Assumption (3) is questionable, because we do not know the effective

number of statistically independent signals contributing to the scalp EEG. However, numerical simulations have confirmed that the ICA algorithm can accurately identify the time courses of activation and the scalp topographies of relatively large and temporally independent sources from simulated scalp recordings, even in the presence of a large number of low-level and temporally independent source activities (Makeig, Jung, Ghahremani, & Sejnowski, in press).

For EEG analysis, the rows of the input matrix **x** are the EEG signals recorded at different electrodes, the rows of the output data matrix $\mathbf{u} = \mathbf{Wx}$ are time courses of activation of the ICA components, and the columns of the inverse matrix $\mathbf{W^{-1}}$ give the projection strengths of the respective components onto the scalp sensors. The scalp topographies of the components provide information about the location of the sources (e.g., eye activity should project mainly to frontal sites, etc.). "Corrected" EEG signals can then be derived as $\mathbf{x}' = (\mathbf{W})^{-1}\mathbf{u}'$, where $\mathbf{u}'$ is the matrix of activation waveforms, **u**, with rows representing artifactual components set to zero. The rank of corrected EEG data is less than that of the original data.

Relation to PCA

Singular value decomposition (SVD) (Golub & Kahan, 1965; Golub & Van Loan, 1989) is used to derive the principal components of EEG signals. Multichannel EEG recordings can be expressed by a P (time points) $\times$ N (channels) matrix, **E**, and decomposed as a product of three matrixes, $\mathbf{E} = \mathbf{USV}^{\mathrm{T}}$, where **U** is an $P \times N$ matrix such that $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}$, **S** is an $N \times N$ diagonal matrix, and **V** is an $N \times N$ matrix such that $\mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{VV}^{\mathrm{T}} = \mathbf{I}$. If **E** is an EEG epoch of N channels and P time points, **U** contains its N normalized principal component waveforms that are *decorrelated* linearly and can be remixed to reconstruct the original EEG. **S** contains the N amplitudes of the N principal component waveforms. We can define the "non-normalized" principal component waveforms as the columns of $\mathbf{P} = \mathbf{US}$. The eigenvector matrix, **V**, is essentially a set of topographic scalp maps, similar to the columns of the $\mathbf{W^{-1}}$ matrix found by ICA.

PCA finds orthogonal directions of greatest variance in the data, whereas ICA component maps may be nonorthogonal. In general, there is no reason why neurobiologically distinct EEG sources should be spatially orthogonal to one another. Therefore, PCA should not in general effectively segregate each EEG source such as brain, cardiac, and eye movement generators, into a separate component (Lamothe & Stroink, 1991).

Figure 1 illustrates schematically the differences between ICA and PCA decompositions of simulated EEG signals recorded at two electrodes (A and B), each of which sums the activities of two temporally independent response sources (#1, EOG; #2, EEG) that have arbitrary but nonidentical spatial distributions. A phase plane plot of the potentials recorded at the two electrodes shows the observed EEG data as a trajectory in the two-dimensional space. In this plot, activity of EOG source #1 alone would lie on a near-vertical axis (ICA-1), whereas activity of EEG source #2 alone would lie on a near-horizontal (but not orthogonal) axis (ICA-2). If the time courses of activation of the two brain networks are independent of one another, the summed output of sources #1 and #2 will, over time, fill the dashed parallelogram, although not necessarily with uniform density. The first principal component of the data (PCA-1) indicates the direction of maximum data variance, but neither this nor the second principal component orthogonal to it matches either of the two nonorthogonal independent component axes.

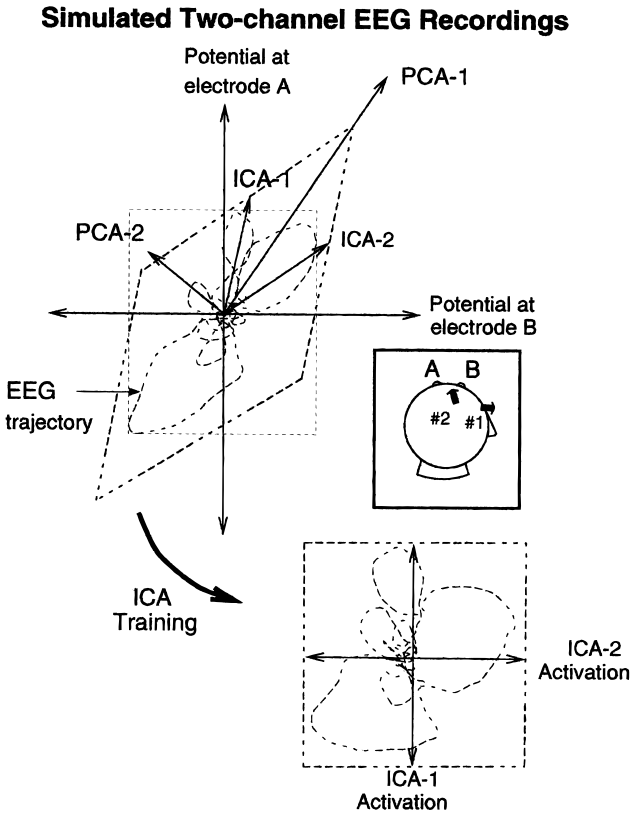


Figure 1. Schematic illustration of independent component analysis (ICA) and principal component analysis (PCA) decompositions of electroencephalogram (EEG) signals (upper left) recorded at two electrodes (A and B), summing the activity of two temporally independent electrooculogram (EOG) (#1) and EEG (#2) sources with differing spatial distributions. PCA finds orthogonal directions of maximum variance in the data. These have no particular relationship to either of the independent components composing the recordings. The ICA algorithm finds the directions of the two axes (ICA-1, ICA-2) by maximizing the entropy of the data transformed linearly to the ICA component axes by a weight matrix and transformed nonlinearly using a compressive nonlinearity (lower right). Maximizing entropy amounts to making the density of the data within the rectangle enclosing the data as uniform as possible, for example, eliminating the empty spaces in the upper left and lower right of the enclosing dotted rectangle (upper left).

The ICA algorithm finds the directions of the two axes (ICA-1, ICA-2) by maximizing the entropy of the data transformed linearly into the ICA component axes and compressed nonlinearly. Hence, the distribution density in the square enclosing the transformed data (lower right of Figure 1) is more uniform than that of the untransformed data (upper left of Figure 1), whose enclosing rectangle contains a larger amount of empty space. If true EEG and EEG artifacts arise through activations of independently active sources, then ICA is more appropriate than PCA for isolating them.

Methods and Materials

The first EEG data set used in the analysis was collected from 20 scalp electrodes placed according to the International 10-20 System and from 2 EOG placements, all referred to the left mastoid. The sampling rate was 256 Hz.

A second data set was recorded at 29 scalp electrodes and 2 EOG placements from an adult autistic subject in an ERP para-

digm. The subject participated in a 2-hr visual selective attention task in which he was instructed to attend to circles flashed in random order at one of five locations laterally arrayed 0.8 cm above a central fixation point. Locations were outlined by five evenly spaced 1.6-cm blue squares displayed on a black background at visual angles of 2.7° and 5.5° from fixation. Attended locations were highlighted through entire 90-s experimental blocks. The subject was instructed to maintain fixation on the central cross and press a button each time he saw a circle in the attended location (see Makeig et al., 1999, for details).

A third EEG data set contained 13 EEG channels (no EOG channel) and was recorded at a sampling rate of 312.5 Hz.

ICA decomposition was performed on 10-s EEG epochs from each data set using MATLAB 4.2c on a DEC 2100A 5/300 processor. The learning batch size was 90, and initial learning rate was 0.001. The learning rate was gradually reduced to 5×10^{-6} during 80 training iterations requiring 6.6 min of computer time (MATLAB toolbox for performing the analyses can be obtained from http://www.cnl.salk.edu/~scott/ica.html).

Regression Analysis

The multiple-lag regression model of Kenemans et al. (1991) was implemented to compare the relative effectiveness of ICA for artifact removal. In this model, the effect of the EOG on the EEG at each sampling time t is given by:

$$eeg(t) = EEG(t) - \sum_{g=0}^{T} \beta_g eog(t-g), \text{ where } \beta_g = SS^{-1} sp_g.$$

Here EEG denotes the "true" EEG minus eye artifacts, whereas $eeg()$ and $eog()$ are the recorded EEG and EOG signals and T is the maximum time lag. The sequence of lagged regression coefficients, β_g, describes the instantaneous and delayed effects of the EOG on the EEG. The vector, sp_g of length $(T + 1)$, contains the inner products of $\text{eeg}(t)$ and $\text{eog}(t - g)$ $(g = 0,\ldots,T)$, while SS is the $(T + 1) \times (T + 1)$ matrix of inner products of $\text{eog}(t - g)$. Note that this method takes into account the frequency- and phase-dependent differences in EOG-to-EEG transfer functions (Kenemans et al., 1991).

Results

Example 1: Removing Eye Movement and Muscle Artifacts

Figure 2A shows a 5-s portion of the recorded EEG time series collected from 20 scalp and 2 EOG electrodes, all referred to the left mastoid. Figure 2B shows the derived ICA component activations and the scalp topographies for five selected ICA components. The eye movement artifact between 2 and 3 s was isolated to ICA components 1 and 4. Components 12, 15, and 19 evidently represent muscle noise from temporal and frontal muscles. The "corrected" EEG signals obtained by removing the five selected (EOG and muscle noise) components from the data are shown in Figure 2C. The scalp maps indicate that components 1 and 4 account for the spread of EOG activity to frontal sites. After eliminating these five artifactual components, by zeroing out the corresponding rows of the activation matrix **u** and projecting the remaining components onto the scalp electrodes, the "corrected" EEG data (Figure 2C) were free of both EOG and muscle artifacts. The "corrected" data also revealed underlying EEG activity at temporal sites T3 and T4 (Figure 2C) that was masked well by muscle activity in the raw data (cf. Figure 2A).

Figure 3A compares the results, at frontal site Fp1, of correcting for eye movement artifacts by ICA and multiple-lag regression. Here, regression was performed only when the artifact was detected (e.g., in the 2-s period surrounding the EOG peak), because otherwise a large amount of EEG activity also would have been regressed out during periods without eye movements. Note that the eye movement artifacts were largely removed (middle trace of Figure 3), but so were portions of theta activity (near second 2 and between seconds 4 and 4.5). In contrast, ICA correction (bottom trace of Figure 3) preserved the theta activity in the original record.

Figure 3B shows that the signal from site T3 contained eye and muscle activity from components 1, 3, and 19 along with underlying EEG activity. Spectral analysis of the original and "corrected" records shows a large amount of overlap between their power spectra, hence bandpass filtering could not have been used to separate them. If, alternatively, the EEG record at T3 were used as a reference to regress out its contributions to signals at adjacent sites, the EEG activity at T3 would also be subtracted from each site and T3 would become silent. ICA, on the other hand, uses spatial filtering to separate and preserve the spectra of all the constituent components.

Figure 4 shows the principal component waveforms from PCA/SVD performed on the EEG data shown in Figure 2, and the scalp topographies of five selected principal components or basis vectors. The eye movement artifact between 2 and 3 s in the EEG data was mostly contained in components 1 and 3, and the left and right temporal muscle activity in the data was concentrated in principal components 4, 5, and 8. "Corrected" EEG signals (Figure 4C) were obtained by removing these five principal components from the data. Note that the eye movement artifact between 2 and 3 s was largely reduced but not completely removed. In particular, this procedure ignored the EOG signals also contained in the second principal component (Figure 4B), which also contained a large amount of EEG activity. If this component were eliminated along with the five selected components, the EEG record would have become nearly silent. In contrast, ICA effectively removed the eye movement artifacts (Figure 2C) with less loss of the EEG signals.

Figure 5 shows the waveforms and spectrograms of the data at one frontal electrode, Fp1 (top panel), before and after correction of an eye movement artifact by ICA and PCA. The waveforms show that ICA was better at removing the low-frequency activity produced by the eye movement. The spectrograms show that ICA removed only the low-frequency activity, whereas PCA also removed a large portion of the theta activity (4–6 Hz). PCA also induced some spurious alpha activity (8–10 Hz), especially near 2 s and 6 s. In contrast, ICA better preserved the theta, alpha, and beta band rhythmic activities in the original record.

Example 2: Removing Eye Blink and Muscle Artifacts

Figure 6 shows a 3-s portion of the recorded EEG time series and its ICA component activations, the scalp topographies of four selected components, and the "corrected" EEG signals obtained by removing four selected EOG and muscle noise components from the data. The eye movement artifact at 1.8 s (left side of Figure 6) was isolated to ICA components 1 and 2 (left middle of Figure 6). Their scalp maps (right middle of Figure 6) indicate that they accounted for the spread of EOG activity to frontal sites. After eliminating these two components and projecting the remaining components onto the scalp channels, the corrected EEG data (right side of Figure 6) were free of these artifacts.

Removing EOG activity from frontal channels revealed alpha activity near 8 Hz that occurred during the eye movement but was

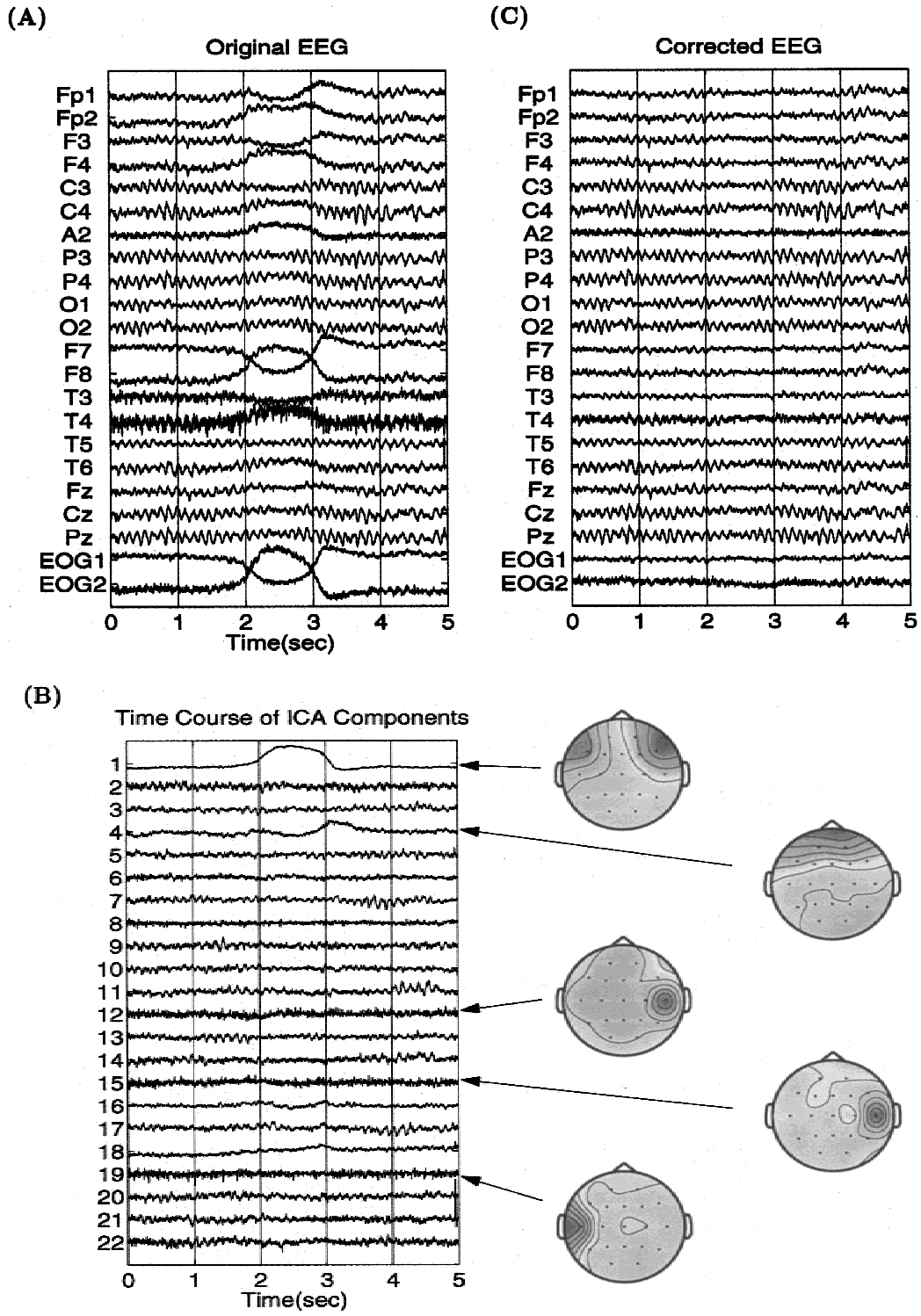


Figure 2. Demonstration of electroencephalogram (EEG) artifact removal by independent component analysis (ICA). (A) A 5-s portion of an EEG time series containing a prominent slow eye movement. (B) Corresponding ICA component activations and scalp maps of five components accounting for horizontal and vertical eye movements (top two) and temporal muscle activity (lower three). (C) EEG signals corrected for artifacts by removing the five selected ICA components in (B).

obscured by the eye artifact in the original EEG traces. Close inspection of the EEG records (Figure 6B) confirmed its existence in the raw data. ICA also revealed the EEG present in the EOG signals (right). In contrast, the corrected EEG at site Fp1 produced by multiple-lag regression contained no sign of this 8-Hz activity (Figure 6B).

Left and right temporal muscle activity in the data was concentrated in ICA components 14 and 15 (Figure 6A). Removing

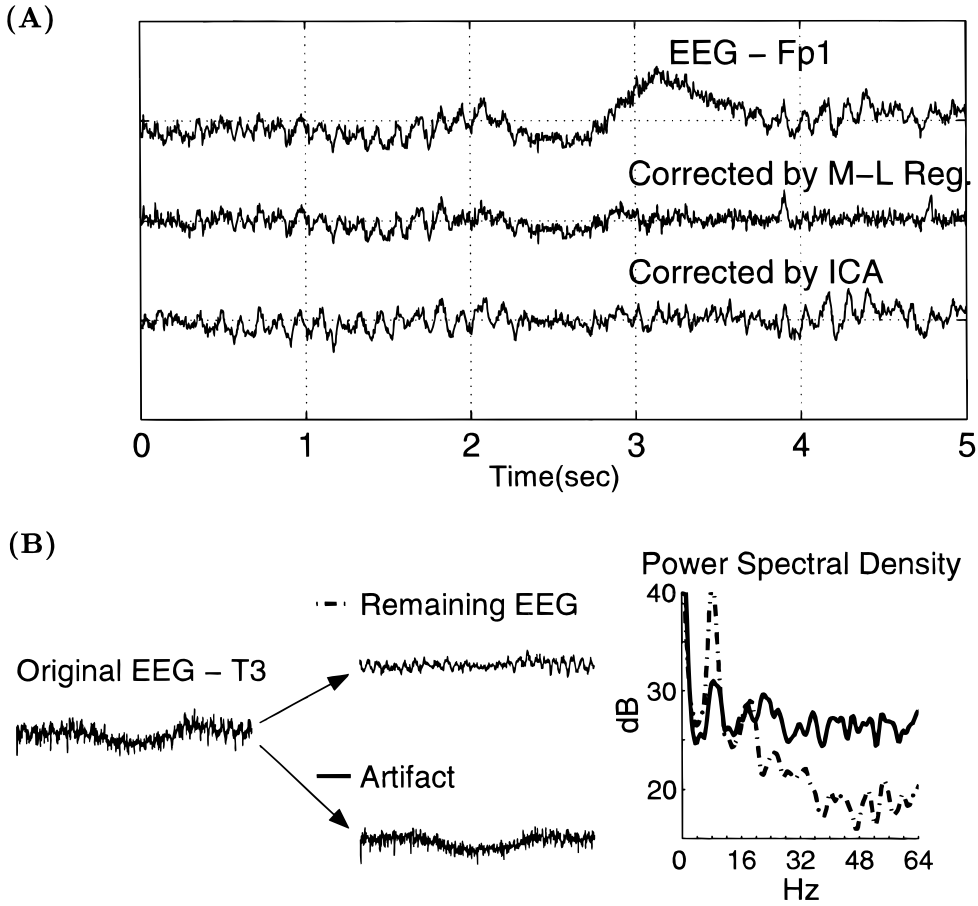


Figure 3. (A) Comparison of results at frontal site Fp1 of multi-lag regression and independent component analysis (ICA) eye-movement correction methods applied to the 5-s electroencephalogram (EEG) epoch of Figure 2. ICA removed only the eye movement artifacts (between 2 and 3 s), whereas the regression method also removed portions of alpha activity (near 8 Hz at second 2 and seconds 4–4.5). (B) The EEG record at left temporal site T3 (cf. Figure 2) is the sum of underlying EEG activity and muscle activity occurring near the electrode. Below 20 Hz, the spectra of the remaining EEG (dash-dotted line) and muscle artifact (solid line) overlapped strongly. ICA separated them by spatial filtering, which preserved their individual spectra.

them from the data (right) revealed underlying EEG activity at temporal sites T3 and T4 that was highly masked by muscle activity in the raw data (left). ICA component 13 (Figure 6A, left middle) also revealed the presence of small periodic muscle spiking in right frontal channels (e.g., F4 and F8) that was obscured in the original data.

Example 3: Separating Blink and Blink-Related Activities

The underlying assumption in applying ICA to EEG artifact removal is that the time courses of true EEG activity and artifacts are statistically independent. However, some true EEG activity might be correlated temporally with particular artifacts. For example, in some ERP experiments, blinks tend to follow significant stimuli and be superimposed on late evoked-response components. In particular, removal of eye artifacts is a significant problem for research on the P300. Could the independent components accounting for blinks also account for some stimulus-evoked brain activity? ICA can be used to investigate the possible coupling between blink-evoked brain and extra-brain activities that may be temporally correlated.

EEG data were recorded at 29 scalp electrodes and 2 EOG placements from an adult autistic subject in a 2-hr visual selected attention ERP experiment. To display all single-trial EEG records, we used a recently developed visualization tool, the "ERP image" (Jung et al., 1999), to illustrate the intertrial variability. Figure 7 (left panel) shows all 641 single-trial ERP epochs recorded at the vertex (Cz) and time-locked to onsets of target stimuli (left vertical line). Single-trial event-related responses are plotted as color-coded horizontal traces (see color bar) sorted by the subject's reaction time in each trial (thick black line). The ERP average of these trials is plotted below the ERP image. ICA, applied to all 641 31-channel EEG records, isolated the blink artifact to a single component whose projections to site Cz are shown in Figure 7 (center). Note that blinks indeed tend to follow the visual target stimuli, as is evident from the poststimulus occurrences of blinks in most of the trials. However, the evoked P300 activities are isolated into different components and remain in the artifact-corrected single-trial ERP epochs (Figure 7, right panel) obtained by subtracting the blink activity (Figure 7, center) from the raw data (Figure 7, left panel). Note that the contributions of the stimulus-induced blink artifacts were mainly on the second peak of the P300 features (Figure 7, bottom trace, center panel), and were removed from the raw data (Figure 7, bottom trace, left).

To investigate the possible coupling between blinks and blink-evoked EEG activities, we extracted trials containing blinks from

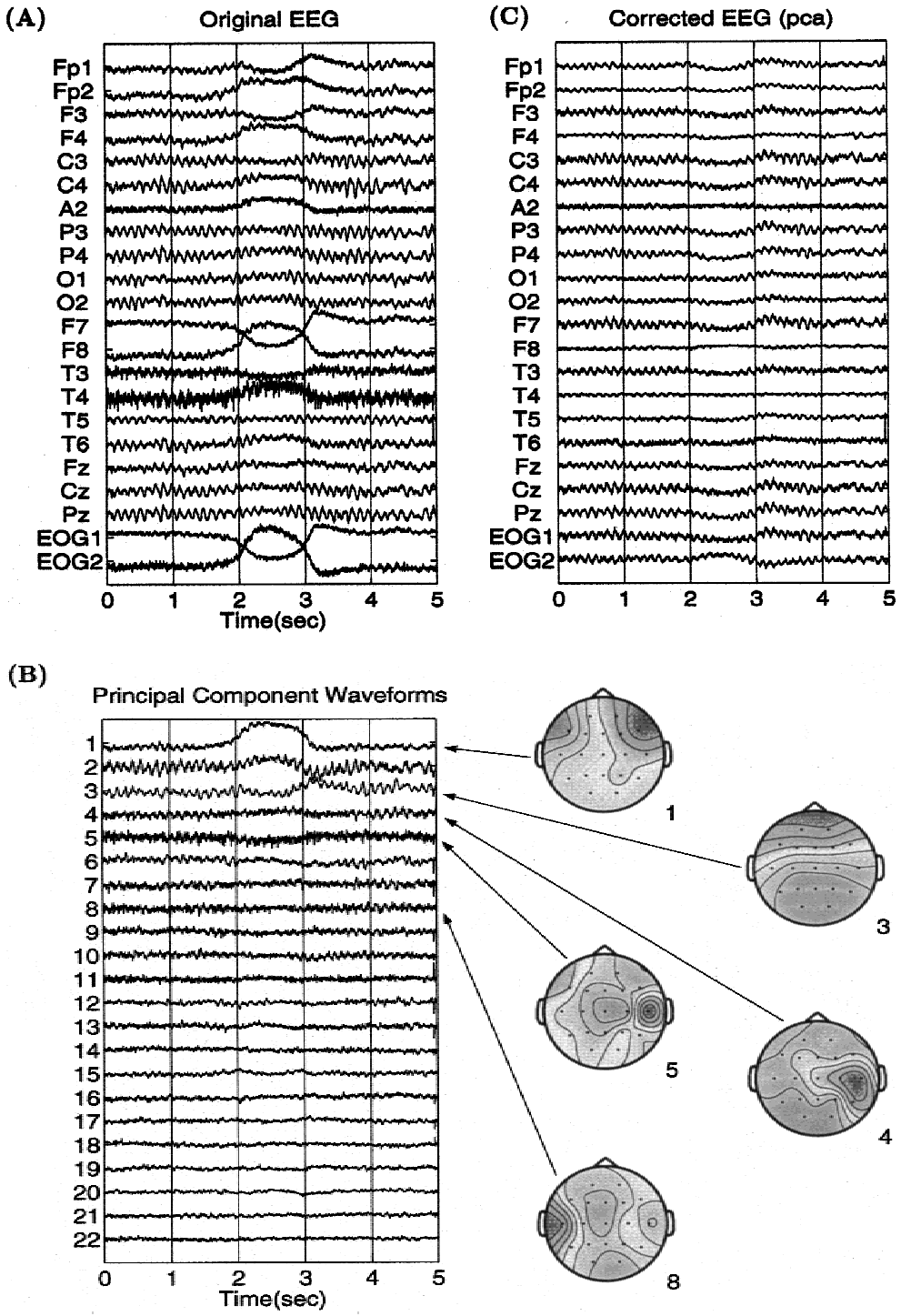


Figure 4. Demonstration of electroencephalogram (EEG) artifact removal by principal component analysis (PCA). (A) The 5-s EEG epoch shown in Figure 2. (B) Principal component waveforms and scalp maps for five selected components. (C) The same epoch corrected for artifacts by PCA by removing the five selected principal components.

all the 641 trial epochs, and realigned all the single-blink epochs to the peak of the blink component excursion.

Figure 8A shows all 185 of these blink epochs at sites EOG1, Fz, Cz, and Pz (Note the different vertical scales in the averages shown below the single trials). Blink epochs are plotted as horizontal colored lines (see color bar). Peak blink amplitude is aligned at time 0 (dashed vertical line). For visibility, epochs are smoothed (top to bottom) with a 10-trial moving window. The blink-triggered average of these trials is plotted in the bottom panel. Note that blink peak amplitude is successively smaller in more posterior channels, and that some blink-related activity occurred 120 ms or longer after the blink peaks. This was most visible at posterior sites.

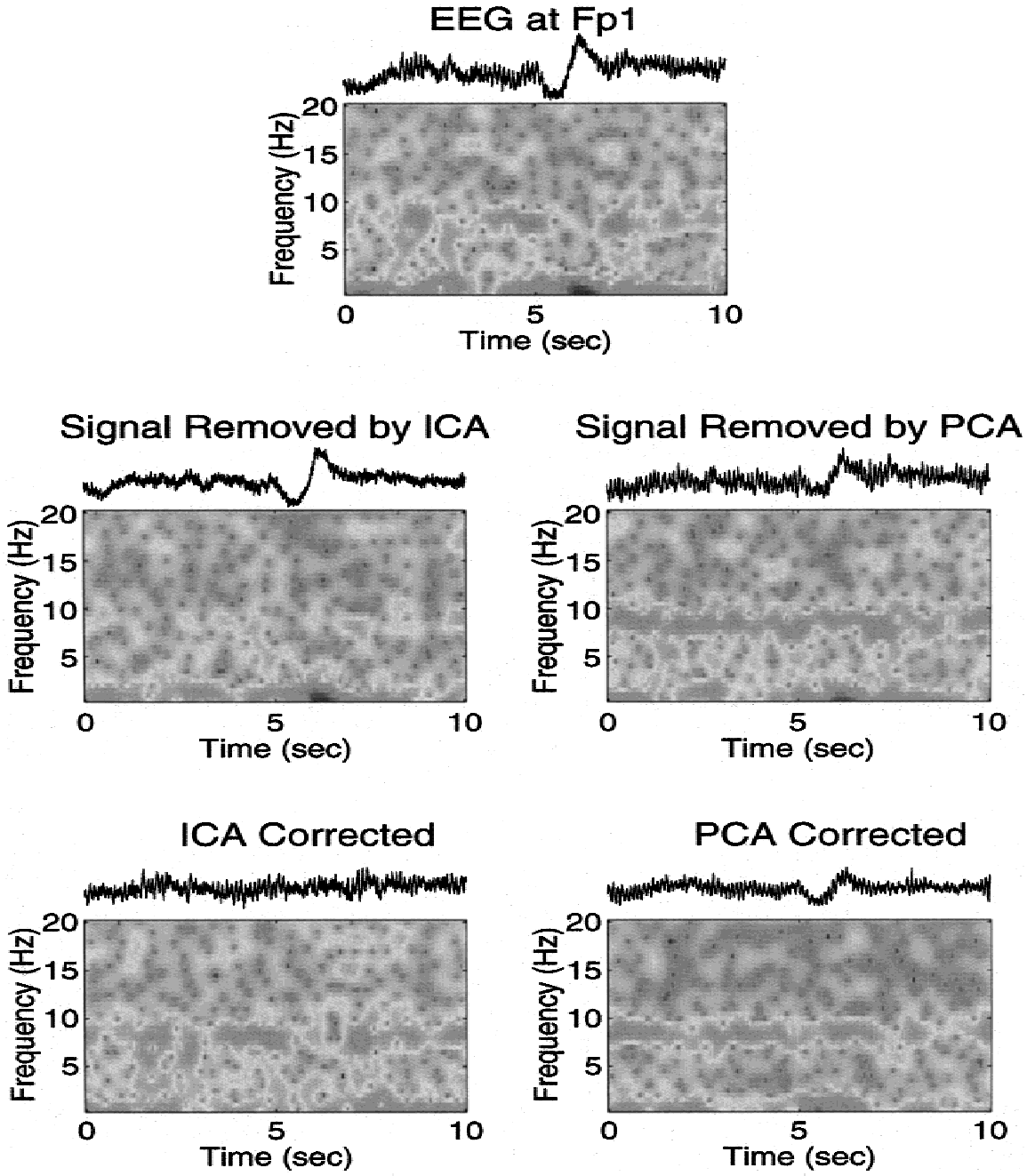


Figure 5. Comparison of eye movement artifact removal by independent component analysis (ICA) and principal component analysis (PCA) techniques. (Top panel) Waveforms and spectrograms of the electroencephalogram (EEG) signals at site Fp1 (cf. Figure 2). (Middle panels) The signals removed using ICA and PCA. (Lower panels) The corrected EEG records produced by both methods.

The results of ICA decomposition of all 185 blink epochs are shown in Figures 8B and 8C. Figure 8B shows the "envelopes" (the most positive and most negative single-channel data values, across 31 scalp channels) of the projected activity of the 4 most-active of the 31 blink-related components (red traces), superimposed on the envelope of the blink-locked data average (black traces). Envelope plots allow the time courses, strengths, latencies, and predominant polarities of ICA components to be visualized in relation to the envelope of original scalp data average (Makeig et al., 1999). The major portion of the large blink artifact was isolated to ICA component 1 (IC1, Figures 8B and 8C, leftmost panel), which was silent outside the main lobe. A second blink-related component (IC3, Figure 8C) appeared in nearly every epoch, mainly after the blink peak. Component IC7 accounted for alpha activity whose phase was reset after blinks, as evident by the larger amplitudes in the blink-locked average near 120 ms after the blink peak. Another distinct component, IC8, (Figure 8C) accounted for additional blink-related brain activity peaking 150 ms after the blink peak in most epochs. Figure 8 shows that ICA, rather than mixing together all blink-related activity into a single

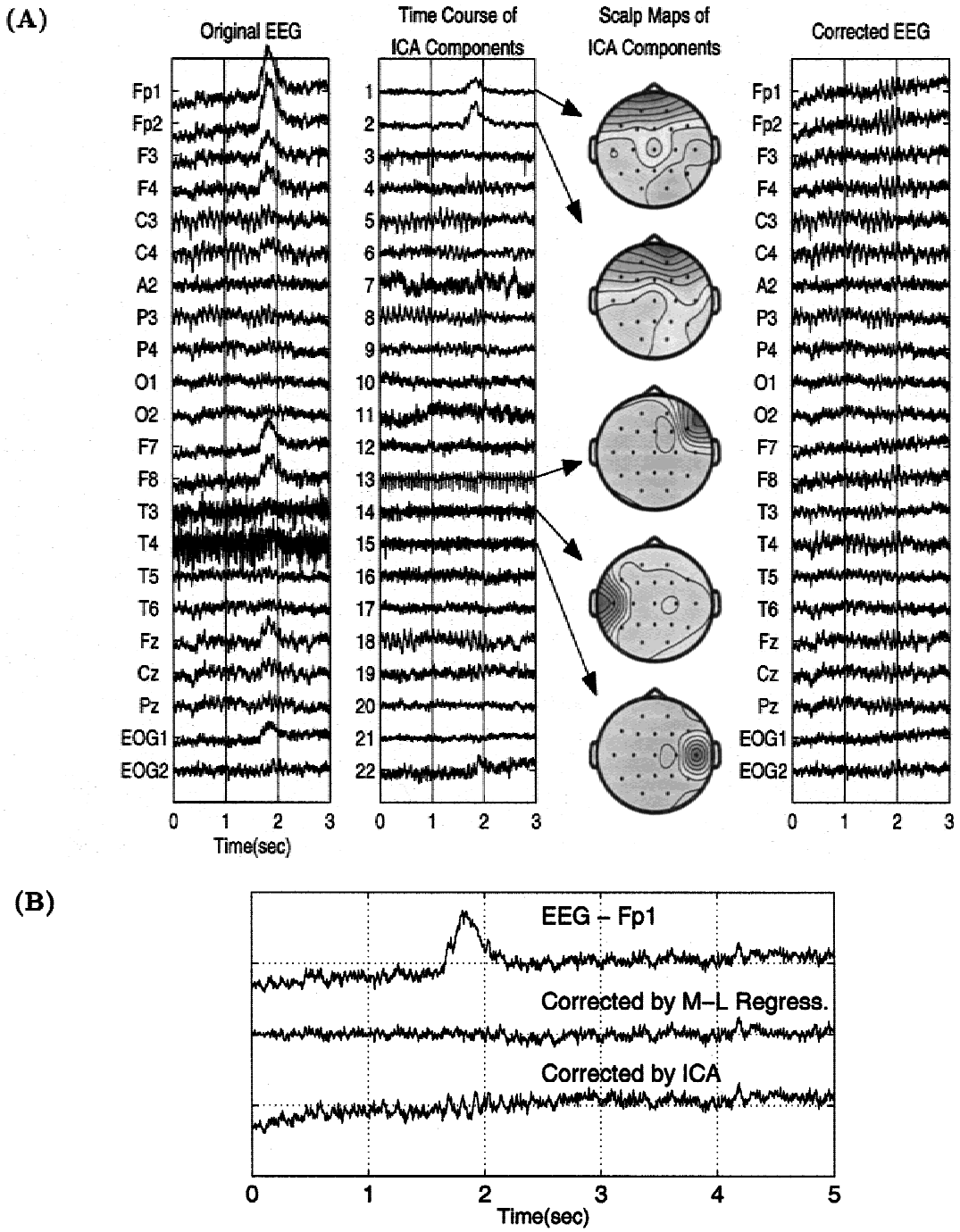


Figure 6. Comparison of artifact removal by independent component analysis (ICA) and multiple-leg regression techniques. (A) A 3-s portion of an electroencephalogram (EEG) time series (left), the corresponding ICA component activations (left middle), scalp maps of five of the ICA components (right middle), and the same EEG signals corrected for artifacts by removing the five selected ICA components (right). (B) Comparison of artifact removal at frontal site Fp1 by ICA and multiple-lag regression. ICA can be used to cancel multiple artifacts in all the data channels simultaneously.

component, derived components whose dynamics were affected by blinks in distinct ways.

Example 4: Removing Line Noise

Figure 9A shows a 10-s portion of an EEG time series collected from 13 scalp electrodes that were heavily contaminated by line noise. Its ICA component activations and principal component waveforms are shown in Figures 9B and 9C, respectively. The top panel of Figure 9D shows the distribution of line noise power near 60 Hz in the EEG channels. The line noise power accounted for by each ICA and PCA component was calculated by averaging power near 60 Hz in the projections of each component all 13 scalp

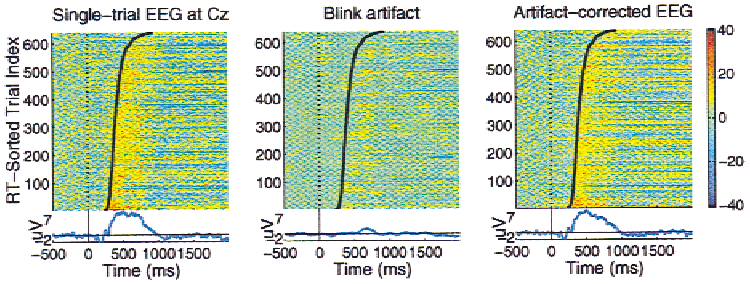
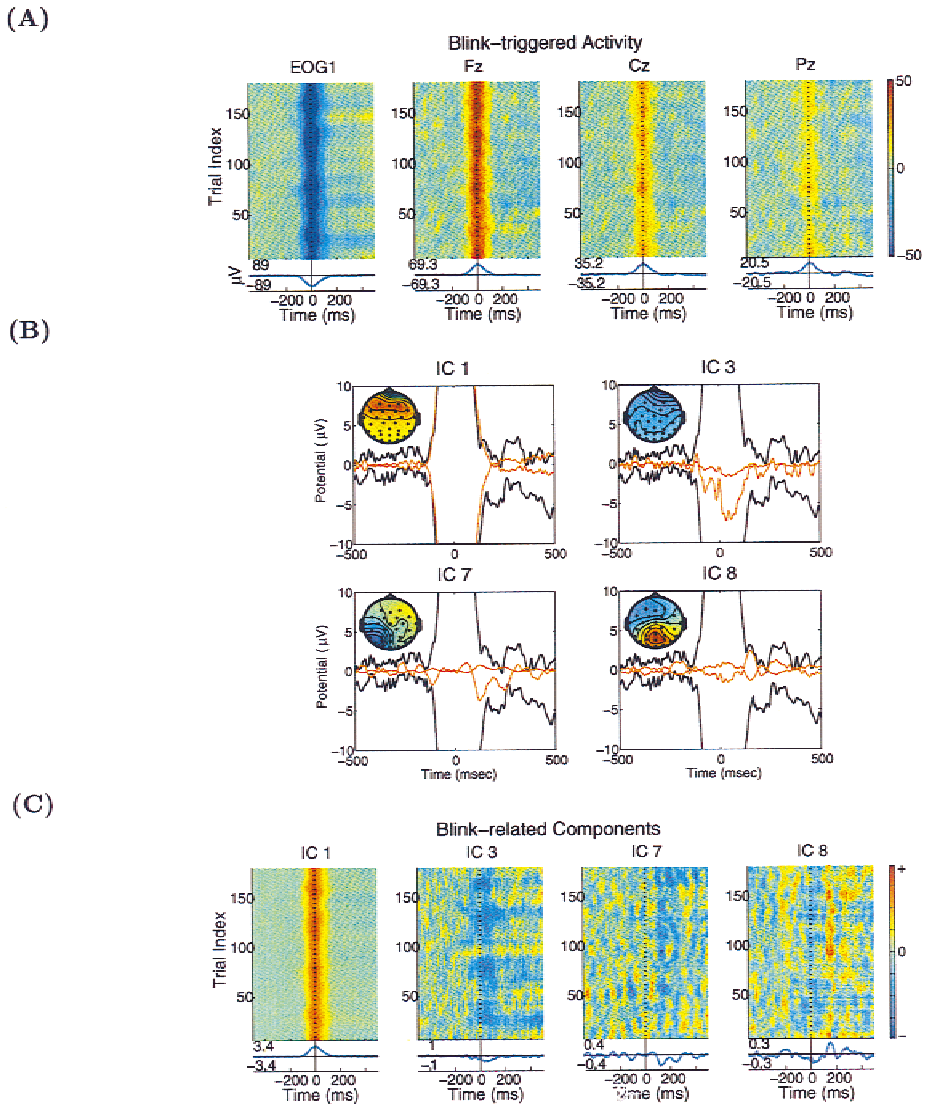


Figure 7. Eye blink artifact removal from single-trial event-related potentials (ERPs) with independent component analysis (ICA). (Left) ERP images of single-trial ERPs at site Cz from one autistic subject EOG2, time locked to 641 targets presented at all five attended locations, and sorted by response time (thick black line). (Center) Projection of ICA component 1 identified as blink artifacts. (Right) Corrected single-trial ERPs obtained by subtracting the artifacts (center) from the original data (left). For visibility, epochs are smoothed (top to bottom) with a three-trial moving window.

electrodes. ICA effectively isolated the line noise power into component 3, which accounted for 75.1% of line noise in the data, whereas PCA concentrated the line noise into the first principal component, which accounted for only 57.4% of the line noise power in the data. Furthermore, the first principal component also contained a large portion of the cerebral activity. Hence, some portions of the relevant brain signals would be removed if this principal component were eliminated to remove line noise arti-



facts. This result is similar to the report of Lagerlund et al. (1997) that large EEG activity with a spatial distribution somewhat similar to that of a principal component may be combined in the same PCA components as the artifacts. Because line noise is sub-Gaussian, the original ICA algorithm (Bell & Sejnowski, 1995), without the extension to sub-Gaussian sources, did not coalesce the line noise in the data into a single component (Lee et al., 1999).

ICA decomposition may be useful as well for observing fine details of the spatial structure of ongoing EEG activity in multiple brain areas or neural populations (Jung et al., 1997; Makeig, Jung, Bell, Ghahremani, & Sejnowski, 1997). For example, in this decomposition, ICA components 1 and 7 accounted for low-frequency alpha activity occurring between 2 and 5 s. Spectral analysis (Figure 10) showed that their peak frequencies were near 7 and 8 Hz, respectively. The two EEG components also had different scalp topographies. Thus, although the ICA algorithm used no explicit temporal sequence or frequency-domain information, alpha activity in this record was separated into two different components, probably arising in different parts of the brain, with distinct frequency contents.

Example 5: Recovering Information From Corrupted Data

In this example, ICA was used to recover useful information from corrupted EEG recordings collected from a normal subject performing a compensatory tracking task. In this session, the low-pass filter was off when the recordings were made, so the data were heavily contaminated not only by line noise but also by harmonics that were aliased into the recordings at irregularly spaced frequencies. Figure 11A shows a 5-s portion of the 7 most contaminated channels chosen from an EEG time series collected from 1 EOG and 22 scalp electrode placements. After ICA was performed on these 23-channel data, the six components accounting for most of the aliased line noise artifact were eliminated from the records (Figure 11B). ICA revealed the presence of alpha activity near 10 Hz between 0.5 and 2 s (Figure 11C) that was highly obscured in the original data. Spectral analyses of the original and corrected EEG records (Figure 11D) shows that the amplitudes of line noise and its harmonics signals were reduced significantly (96–99.9% in the different channels), whereas signal amplitudes at other frequencies remained intact.

Discussion

Although the neural mechanisms that generate EEG are not fully known, the assumptions of the ICA algorithm are generally compatible with a widely assumed model that EEG data recorded at multiple scalp sensors are a linear sum at the scalp electrodes of activations generated by distinct neural and artifactual sources. The algorithm derives spatial filters that decompose EEG data recorded at multiple scalp sensors into a sum of components with fixed scalp distributions and maximally independent time courses. Our confidence in ICA decomposition of EEG signals is strengthened by the fact that topographic projections (scalp maps) of ICA components tend to have few spatial maxima, suggesting a few localized brain sources (Figures 2C, 6A, and 10), whereas those of most principal components derived by PCA and SVD have more complex spatial patterns (Silberstein & Cadusch, 1992), probably due to the spatial orthogonality imposed on the component maps by PCA. Although ICA also imposes a strong criterion (temporal statistical independence) on the components, ICA does not impose any condition on the spatial filters. As a result, spatial filters derived by ICA are not affected by each other and can collect concurrent activity arising from any spatially overlapping source distributions.

Limitations of ICA

Although the ICA method appears to be generally useful for EEG analysis, it also has some inherent limitations. First, like PCA, ICA can decompose, at most, N sources from N data channels. The effective number of statistically independent signals contributing to the scalp EEG is generally unknown, but brain activity probably arises from effectively more physically separable sources than the available number of EEG electrodes. To explore the effects of a larger number of sources on the results of the ICA decomposition from a limited number of channels, we performed a number of numerical simulations in which selected signals recorded from the cortex of an epileptic patient during preparation for operation for epilepsy were projected to simulated scalp electrodes through a three-shell spherical model. We used electrocorticographic data in these simulations as a plausible best approximation to the temporal dynamics of the unknown EEG brain generators. Results confirmed that the ICA algorithm can accurately identify the time courses of activation and the scalp topographies of relatively large and temporally independent sources from simulated scalp recordings, even in the presence of a large number of simulated low-level source activities (Makeig et al., in press).

Second, like PCA, ICA is based on statistical analysis of the data, hence its results will not be meaningful if the amount of data given to the algorithm is insufficient. In principle, it is best to use all available data to reliably derive spatial filters characterizing the appearance and spread of artifacts in the EEG. However, this is only true when the physical sources of artifacts and cerebral activity are spatially stationary through time, and the total number of these sources is less than the number of data channels. In general, there is no reason to believe that the cerebral and artifactual sources remain stationary over time. The goal then should be to use the maximum

Figure 8. (*facing page*) Separation of blink and blink-related activity by independent component analysis (ICA). (A) Single-trial blink episodes, recorded at sites EOG1, Fz, Cz, and Pz and time-locked to peaks of blinks (vertical center line), averaged using a 10-trial moving window advanced (top to bottom) in one-trial increments. The blink-triggered average of these trials is plotted in the bottom of each panel. (Note different vertical scales.) (B) The 185 blink episodes were decomposed by ICA, and four of the components are shown here. For each component (panel), the "envelope" (the most positive and most negative single-channel data values, across 31 scalp channels) of the projected activity of the blink-related component (red traces) was overplotted on the envelope of the blink-locked data average (black traces). The scalp maps of the components IC1 and IC3 indicate that they accounted for the spread of electrooculogram (EOG) activity to frontal sites. Synchronization of ongoing activity in components IC7 and IC8 following the blinks created small temporally overlapping evoked responses. (C) Event-related potentials (ERP) images of the activations of the same four selected ICA components accounting for blink-related brain and extra-brain activities. Note that each component exhibits distinct reactivities to the blinks.

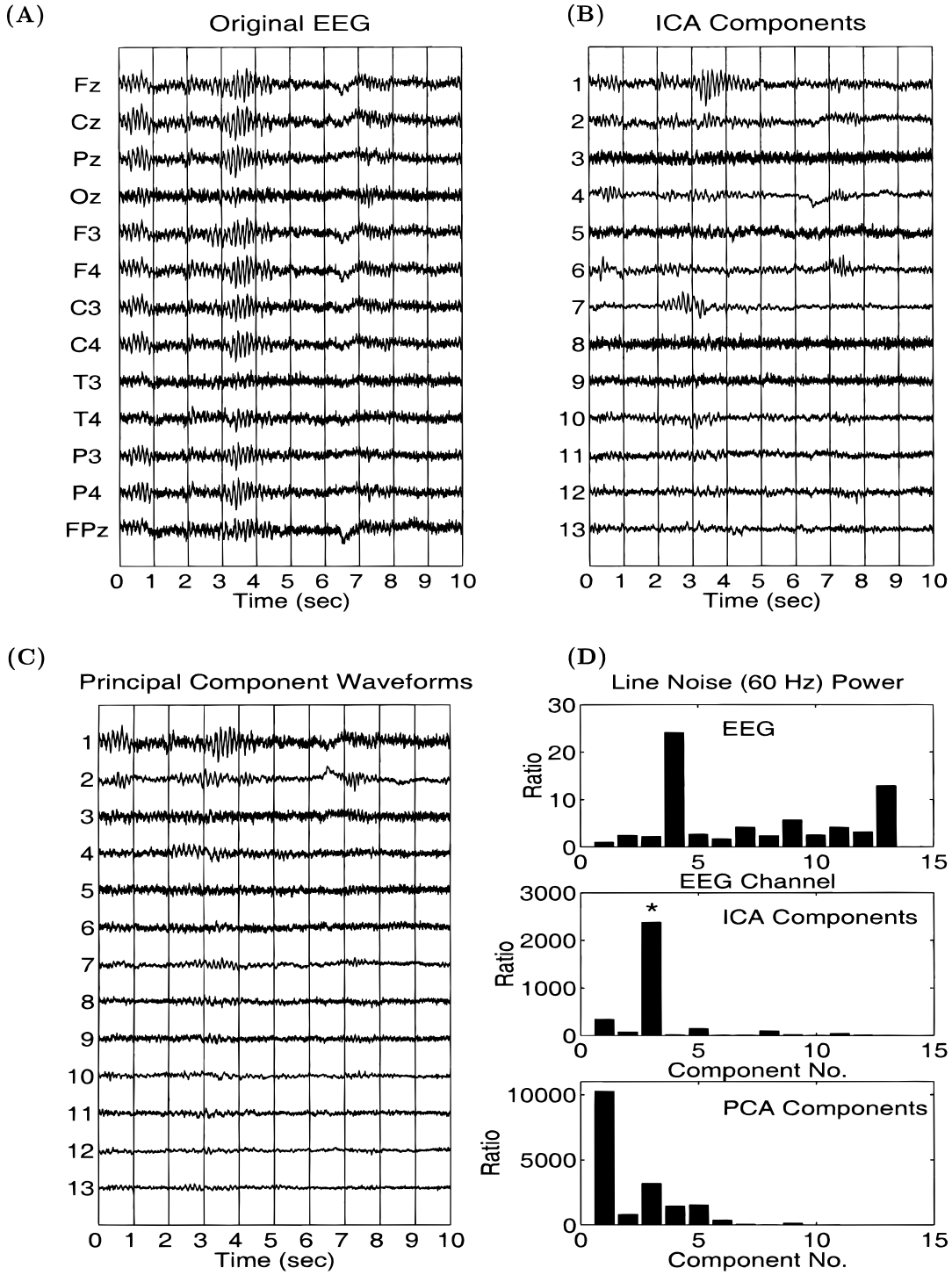


Figure 9. Comparison of line noise (60 Hz) removal by independent component analysis (ICA) and principal component analysis (PCA). (A) A 5-s portion of another electroencephalogram (EEG) time series, (B) its ICA component activations, and (C) its principal component waveforms. (D) The ratio of power at the line frequency (60 Hz) in the EEG channels (top panel), in the ICA components (middle panel), and in the principal components (bottom panel). Note the differences in ratio scale between the three panels. *The ICA algorithm isolates most of the line noise into a single component.

amount of data during which the sources are reasonably stationary. Experience suggests that 10-s epochs usually give good results.

Another limitation of the proposed method is that artifact removal requires visual inspection of the ICA components and determination of which components to remove. This can be time-consuming and is not desirable for artifact removal in routine clinical EEG. However, the distributions of spectral power in some artifactual components were distinct, which suggests that it might

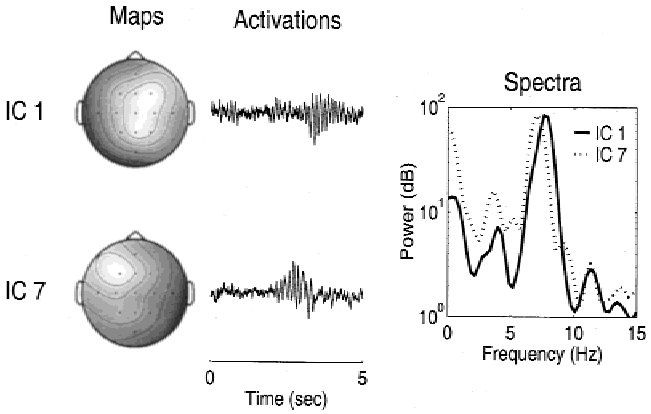


Figure 10. Separation of two overlapping electroencephalogram (EEG) sources by independent component analysis (ICA) from the data shown in Figure 9. The scalp maps of ICA components 1 and 7 (Figure 7) (left), the first 5 s of their activations (middle), and their power spectra (right). Note the spatial, temporal, and peak frequency differences between the two components.

be feasible to automate procedures for removing these artifacts from contaminated EEG recordings.

Separation of Artifact-Evoked EEG Sources by Single-Trial ICA

The underlying assumption in applying ICA to EEG artifact removal is that the time courses of true EEG activity and artifacts are statistically independent. However, EEG activity may be correlated temporally with particular artifacts. For example, in a visual ERP experiment, blinks may follow significant stimuli that also elicit particular types of brain activity (e.g., P300) with similar latency on average, especially in patient groups. However, blinks in ERP experiments are likely to also occur at times when target stimuli have not been presented and target-related brain activity is therefore not present. To illustrate this point, assume activities from EEG source A and EEG source B are both elicited in a certain condition (condition 1), but are sometimes active independently in the same or another condition (condition 2). If ICA were trained on data collected only in condition 1, one of the ICA components would likely combine sources A and B and treat them as a single

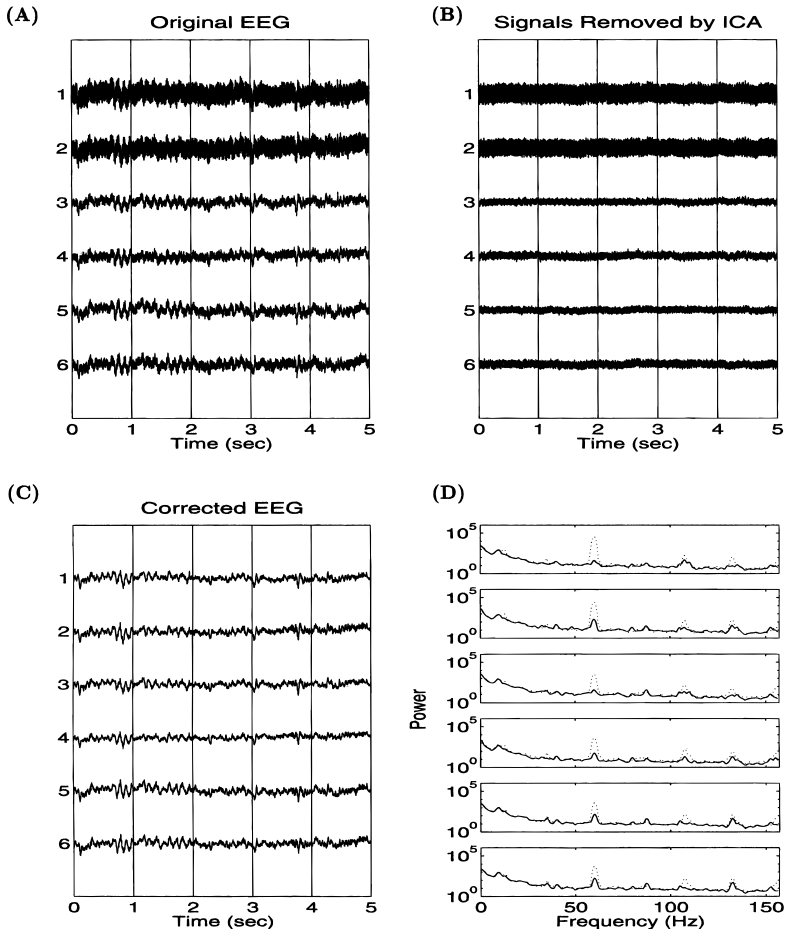


Figure 11. Removal of harmonic artifacts with independent component analysis (ICA). (A) A 5-s portion of a corrupted electroencephalogram (EEG) time series resulting from a poor data-acquisition setting; (B) noise components extracted by ICA (right panel). (C) The same EEG signals corrected for artifacts by ICA by removing the six selected components, and (D) spectral analysis of the original and corrected EEG recordings. Note that EEG activity is more visible than in (A), particularly in channels 1 and 2, and the line noise (60 Hz) and aliased line noise frequencies (near 12, 105, and 135 Hz) are reduced.

source—if A and B always, or nearly always, occurred simultaneously. However, if ICA were given data in which the functional independence of sources 1 and 2 were expressed, for example, data from both condition 1 and condition 2, ICA would separate the activities arising from the two sources based on their temporal independence in the input data as a whole. For this reason, ICA should be applied to single-trial EEG recorded during ERP experiments under a variety of related conditions, rather than to time-restricted single responses or averaged epochs time locked to a single class of experimental events. The separation of P300, blink, and blink-related EEG activity by ICA (Figures 7 and 8) provides strong evidence for this approach.

Separation of extra-brain and brain activity is not affected by the similarity in spatial distributions of these sources. ICA imposes a strong criterion (temporally statistical independence) on the temporal activity of components, but, unlike PCA, it does not impose any condition on the spatial filters or on the spatial projections of the components to the different EEG channels. As a result, spatial filters derived by ICA are not affected by each other and can separate independent (but often concurrent) activity arising from sources with similar spatial distributions (Makeig et al., in press).

Conclusions

ICA opens new and useful windows into many brain and non-brain phenomena contained in multichannel EEG records by separating data recorded at multiple scalp electrodes into a sum of temporally independent components. In many cases, the temporally independent ICA components are also functionally independent. In particular, ICA appears to be a generally applicable and effective method for removing a wide variety of artifacts from EEG records, because their time courses are generally temporally independent and spatially distinct from sources of cerebral activity. However, because ICA decomposition is based on the assumption that EEG data are derived from spatially stationary brain or extra-brain generators, further research will be required to fully assess the value and limitations of this new analytic method.

ICA has several advantages compared with other artifact removal methods: (1) The algorithm is computationally efficient and the computational requirements are not excessive even for fairly large EEG data sets. (2) ICA is generally applicable for removal of a wide variety of EEG artifacts. It simultaneously separates both the EEG and its artifacts into independent components based on the statistics of the data, without relying on the availability of one or more "clean" reference channels for each type of artifacts. This avoids the problem of mutual contamination between regressing and regressed channels. (3) Unlike regression-based methods, no arbitrary thresholds (usually variable across sessions) are needed to determine when artifact correction should be performed. (4) Separate analyses are not required to remove different classes of artifacts. Once the training is complete, artifact-free EEG records in all channels can then be derived by simultaneously eliminating the contributions of various identified artifactual sources in the EEG record. (5) The ICA artifact subtraction method preserves and recovers more brain activity than regression and PCA. (6) The same ICA approach should be equally applicable to other types of multichannel biomedical data for which linear summation can be assumed (e.g., MEG, ECoG, ECG, EMG, etc.). In addition to artifact removal, ICA decomposition can be highly useful for observing changes in the spatial structure of ongoing or averaged EEG activity in multiple brain areas, networks, or neural populations (Jung et al., 1997, 1999; Makeig et al., 1997, 1999).

REFERENCES

Amari, S., Chen, T.-P., & Cichocki, A. (1997). Stability analysis of adaptive blind source separation. *Neural Networks, 10*, 1345–1352.

Amari, S., Cichocki, A., & Yang, H. (1996). A new learning algorithm for blind signal separation. In D. Touretzky, M. Mozer, & M. Hasselmo (Eds.), *Advances in neural information processing systems* (Vol. 8, pp. 757–763). Cambridge, MA: The MIT Press.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation, 7*, 1129–1159.

Berg, P., & Scherg, M. (1991a). Dipole models of eye movements and blinks. *Electroencephalography and Clinical Neurophysiology, 79*, 36–44.

Berg, P., & Scherg, M. (1991b). Dipole models of eye activity and its application to the removal of eye artifacts from the EEG ad MEG. *Clinical Physics and Physiological Measurements, 12(Supplement A)*, 49–54.

Cardoso, J. -F. (1998). Blind signal processing: A review. *Proceedings of the IEEE, 86*, 2009–2025.

Cardoso, J. -F., & Laheld, B. (1996). Equivariant adaptive source separation. *IEEE Transactions on Signal Processing, 45*, 434–444.

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? *Signal Processing, 36*, 287–314.

Girolami, M. (1998). An alternative perspective on adaptive independent component analysis. *Neural Computation, 10*, 2103–2114.

Girolami, M., & Fyfe, C. (1997). Generalized independent component analysis through unsupervised learning with emergent Bussgang properties. In *Proceedings of the IEEE International Conference on Neural Networks*, 1788–1791. Houston, TX.

Golub, G. H., & Kahan, W. (1965). Calculating the singular values and pseudoinverse of a matrix. *SIAM J. Numer. Anal.* 2:205-224.

Golub, G. H., & Van Loan, C. F. (1989). *Matrix computations* (2nd ed.). Baltimore: John Hopkins University Press.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55*, 468–484.

Hillyard, S. A., & Galambos, R. (1970). Eye-movement artifact in the CNV. *Electroencephalography and Clinical Neurophysiology, 28*, 173–182.

Jung, T. -P., Makeig, S., Bell, A. J., & Sejnowski, T. J. (1997). Independent component analysis of electroencephalographic and event-related potential data. In P. Poon & J. Brugge (Eds.), *Central auditory processing and neural modeling* (pp. 189–197). New York: Plenum.

Jung, T. -P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (1999). Analyzing and visualizing single-trial event-related potentials. In M. Kearns, S. Solla, & D. Cohn (Eds.), *Advances in neural information processing systems* (Vol. 11, pp. 118–124). Cambridge, MA: The MIT Press.

Karhunen, J., Oja, E., Wang, L., Vigario, R., & Joutsensalo, J. (1997). A class of neural networks for independent component analysis. *IEEE Transactions on Neural Networks, 8*, 487–504.

Kenemans, J. L., Molenaar, P., Verbaten, M. N., & Slangen, J. L. (1991). Removal of the ocular artifact from the EEG: a comparison of time and frequency domain methods with simulated and real data. *Psychophysiology, 28*, 114–121.

Lagerlund, T. D., Sharbrough, F. W., & Busacker, N. E. (1997). Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition. *Journal of Clinical Neurophysiology, 14*, 73–82.

Lambert, R. (1996). *Multichannel blind deconvolution: FIR matrix algebra*

and separation of multipath mixtures. Thesis, University of Southern California, Department of Electrical Engineering.
Lamothe, R., & Stroink, G. (1991). Orthogonal expansions: Their applicability to signal extraction in electrophysiological mapping data. *Medical and Biological Engineering and Computing*, *29*, 522–528.
Lee, T. -W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources. *Neural Computation*, *11*, 606–633.
Lee, T. -W., & Sejnowski, T. J. (1997). Independent component analysis for sub-Gaussian and super-Gaussian mixtures. *Proceedings of the 4th Joint Symposium on Neural Computation*, *7*, 132–139. La Jolla, CA: University of California, San Diego.
Lins, O. G., Picton, T. W., Berg, P., & Scherg, M. (1993). Ocular artifacts in recording EEG and event-related potentials II: Source dipoles and source components. *Brain Topography*, *6*, 65–78.
Makeig, S., Bell, A. J., Jung, T. -P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. In D. Touretzky, M. Mozer, & M. Hasselmo (Eds.), *Advances in neural information processing systems* (Vol. 8, pp. 145–151). Cambridge, MA: The MIT Press.
Makeig, S., Jung, T. -P., Bell, A. J., Ghahremani, D., & Sejnowski, T. J. (1997). Blind separation of event-related brain responses into independent components. *Proceedings of the National Academy of Sciences Of the United States of America*, *94*, 10979–10984.
Makeig, S., Westerfield, M., Jung, T. -P., Covington, J., Townsend, J., Sejnowski, T. J., Courchesne, E. (1999). Independent components of the late positive response complex in a visual spatial attention task. *Journal of Neuroscience*, *19*, 2665–2680.
Makeig, S., Jung, T. -P., Ghahremani, D., Sejnowski, T. J. (in press). Independent component analysis of simulated ERP data. *Human Higher Function I: Advanced Methodologies*.
Nadal, J. P., & Parga, N. (1994). Non-linear neurons in the low noise limit: A factorial code maximises information transfer. *Network*, *5*, 565–581.
Oster, P. J., & Stern, J. A. (1980). Measurement of eye movement electrooculography. In I. Matin & P. H. Venables (Eds.), *Techniques in psychophysiology* (pp. 275–309). Chichester, UK: Wiley.
Overton, D. A., & Shagass, C. (1969). Distribution of eye movement and eye blink potentials over the scalp. *Electroencephalography and Clinical Neurophysiology*, *27*, 546.
Pearlmutter, B., & Parra, L. (1997). Maximum likelihood blind source separation: A context-sensitive generalization of ICA. In M. Jordan, M. Mozer, & T. Petsche (Eds.), *Advances in neural information processing systems* (Vol. 9, pp. 613–619). Cambridge, MA: The MIT Press.
Pearson, K. (1901). *Contributions to the mathematical study of evolution.* London: Dulau and Co.
Peters, J. F. (1967). Surface electrical fields generated by eye movement and eye blink potentials over the scalp. *Journal of EEG Technology*, *7*, 27–40.
Pham, D. T. (1997). Blind separation of instantaneous mixture of sources via an independent component analysis. *IEEE Transactions on Signal Processing*, *44*, 2768–2779.
Roth, Z., & Baram, Y. (1996). Multidimensional density shaping by sigmoids. *IEEE Transactions on Neural Networks*, *7*, 1291–1298.
Silberstein, R. B., & Cadusch, P. J. (1992). Measurement processes and spatial principal components analysis. *Brain Topography*, *4*, 267–276.
Verleger, R., Gasser, T., & Möcks, J. (1982). Correction of EOG artifacts in event-related potentials of EEG: Aspects of reliability and validity. *Psychophysiology*, *19*, 472–480.
Vigário, R. N. (1997). Extraction of ocular artifacts from EEG using independent component analysis *Electroencephalography and Clinical Neurophysiology*, *103*, 395–404.
Weerts, T. C., & Lang, P. J. (1973). The effects of eye fixation and stimulus and response location on the contingent negative variation (CNV). *Biological Psychology*, *1*, 1–19.
Whitton, J. L., Lue, F., & Moldofsky, H. (1978). A spectral method for removing eye-movement artifacts from the EEG. *Electroencephalography and Clinical Neurophysiology*, *44*, 735–741.
Woestenburg, J. C., Verbaten, M. N., & Slangen, J. L. (1983). The removal of the eye-movement artifact from the EEG by regression analysis in the frequency domain. *Biological Psychology*, *16*, 127–147.
Yellin, D., & Weinstein, E. (1996). Multichannel signal separation: Methods and analysis. *IEEE Transactions on Signal Processing*, *44*, 106–118.

Received February 9, 1998; ACCEPTED March 26, 1999

APPENDIX

ICA Algorithm

The blind source separation problem is an active area of research in statistical signal processing (Amari, Chen, & Cichocki, 1996; Amair, Cichocki, & Yang, 1997; Bell & Sejnowski, 1995; Cardoso & Laheld, 1996; Cichocki, Unbehauen, & Rummert, 1994; Comon, 1994; Girolami & Fyfe, 1997; Karhunen, Oja, Wang, Vigário, & Joutsensalo, 1996; Lambert, 1996; Nadal & Parga, 1994; Pearlmutter & Parra, 1997; Pham, 1997; Roth & Baram, 1996; Yellin & Weinstein, 1996). Comon (1994) defined the concept of ICA as maximizing the degree of statistical independence among outputs using contrast functions approximated by Edgeworth expansion of the Kullback–Leibler divergence. In contrast with decorrelation techniques such as PCA, which ensure that output pairs are uncorrelated ($\langle u_i u_j \rangle = 0$, for all i, j), ICA imposes the much stronger criterion that the multivariate probability density function (p.d.f.) of **u** factorizes:

$$\boldsymbol{f_u}(\boldsymbol{u}) = \prod_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{f_{u_i}}(\boldsymbol{u_i})$$

Statistical independence requires all higher-order correlations of the u_i to be zero, while decorrelation only takes account of second-order statistics (covariance or correlation).

Bell and Sejnowski (1995) derived a simple neural network algorithm based on information maximization ("infomax") that can blindly separate super-Gaussian sources (e.g., sources that are "on" less often than a Gaussian process with the same mean and variance). The important fact used to distinguish a source, s_i, from mixtures, x_i, is that the activity of each source is statistically independent of the other sources. That is, their joint probability density function (p.d.f.), measured across the input time ensemble, factorizes. This statement is equivalent to saying that the mutual information between any two sources, s_i and s_j, is zero:

$$\boldsymbol{I}(u_1, u_2, \ldots, u_N) = E\left[\ln \frac{\boldsymbol{f_u}(\boldsymbol{u})}{\prod_{i=1}^{N} \boldsymbol{f_{u_i}}(\boldsymbol{u_i})}\right] = 0$$

where E[.] denotes expected value. Unlike sources, s_i's, which are assumed to be temporally independent, the observed mixtures of sources, x_i's, are statistically dependent on each other, so the mutual information between pairs of mixtures, $I(x_i, x_j)$ is in general positive. The blind separation problem is to find a matrix, **W**, such that the linear transformation

$$\mathbf{u} = \mathbf{Wx} = \mathbf{WAs}$$

reestablishes the condition $I(u_i, u_j) = 0$ for all $i \neq j$.

Consider the joint entropy of two nonlinearly transformed components of **y**:

$$H(y_1, y_2) = H(y_1) + H(y_2) - I(y_1, y_2)$$

where $y_i = g(u_i)$ and $g(\,)$ is an invertible, bounded nonlinearity. The nonlinearity function provides, through its Taylor series expansion, higher-order statistics that are necessary to establish independence.

Maximizing this joint entropy involves maximizing the individual entropies, $H(y_1)$ and $H(y_2)$, while minimizing the mutual information, $I(y_1, y_2)$, shared between the two. Thus, maximizing $H(\boldsymbol{y})$, in general, minimizes $I(\boldsymbol{y})$. When this latter quantity is zero, the two variables are statistically independent.

The algorithm attempts to maximize the entropy $H(\boldsymbol{y})$ by iteratively adjusting the elements of the square matrix, **W**, using small batches of data vectors (normally 10 or more) drawn randomly from {**x**} without substitution, according to Bell and Sejnowski (1995):

$$\Delta W \propto \frac{\partial H(y)}{\partial W} W^T W = [I + \phi u^T]W, \text{ where } \phi_i = \frac{\partial}{\partial u_i} \ln \frac{\partial y_i}{\partial u_i}.$$

The $(W^T W)$ "natural gradient" term (Amari et al., 1996; Cardoso & Laheld, 1996) avoids matrix inversions and speeds convergence. The form of the nonlinearity $g(u)$ plays an essential role in the success of the algorithm. The ideal form for $g(u)$ is the cumulative density function (c.d.f.) of the distributions of the independent sources. When $g(u)$ is a sigmoid function (as in Bell & Sejnowski, 1995), the algorithm is then limited to separating sources with super-Gaussian distributions.

A way of generalizing the learning rule to sources with either sub-Gaussian or super-Gaussian distributions is to estimate p.d.f. of sources using a parametric density model. Sub-Gaussians can be modeled with a symmetrical form of the Pearson mixture model (Pearson, 1901) as proposed in Girolami (1998) and Lee et al. (1999), whereas super-Gaussians can be modeled as the derivative of the hyperbolic tangent (Girolami, 1998; Lee et al., 1999). For sub-Gaussians, the following approximation is possible: $\phi_i = +\tanh(u_i) - u_i$. For super-Gaussians, the same approximation becomes $\phi_i = -tanh(u_i) - u_i$. The two equations can be combined as

$$\Delta W \propto [I - K\tanh(u)u^T - uu^T]W \begin{cases} k_i = 1: & \text{super-Gaussian} \\ k_i = -1: & \text{sub-Gaussian} \end{cases}$$

where k_i are elements of the N-dimensional diagonal matrix **K**. The k_i's can be derived from the generic stability analysis (Cardoso, 1998; Cardoso & Laheld, 1996; Pham, 1997) of separating solutions. This yields the choice of k_is used by Lee et al. (1999),

$$k_i = \text{sign}(E[\text{sech}^2(u_i)]E[u_i^2] - E[(\tanh(u_i)u_i]),$$

which ensures stability of the learning rule.

Note that although a nonlinear function is used in determining **W**, once the algorithm converges and **W** is found, the decomposition is a linear transformation, **u = Wx**. This extended infomax algorithm was used to analyze the EEG recordings in this study.