Mining Electroencephalographic Data Using Independent Component Analysis

Tzyy-Ping Jung and Scott Makeig
Swartz Center for Computational Neuroscience
Institute for Neural Computation
University of California, San Diego
La Jolla, CA 92093-0961, U.S.A.
{jung.scott}@scn.ucsd.edu

Corresponding author:
Tzyy-Ping Jung, Ph.D.
Swartz Center for Computational Neuroscience
Institute for Neural Computation
University of California, San Diego
9500 Gilman Dr. DEPT 0961
La Jolla, Ca 92093-0961
Tel: 858-458-1927x12
Fax: 858-458-1847
E-mail: jung@sccn.ucsd.edu

Running Title: Mining EEG Data using ICA

Keywords: Blind source separation; Independent Component Analysis; ICA; ERP; Single-trial ERP; artifact removal; Event-related Brain Dynamics; Source localization

Number of words in the abstract: 250
Number of pages text pages: 58 (including acknowledgments and references)
Number of figures: 8
Number of tables: 0
ABSTRACT (250 words)

Long dominated by event-related potential (ERP) averaging methods, current electroencephalographic (EEG) research is beginning to make use of modern signal processing advances including time/frequency analysis, 3-D and animated visualization, source localization onto magnetic resonance images of the head, and independent component analysis (ICA) applied to collections of unaveraged continuous or single-trial event-related data. Research in our own and many other laboratories over the last decade has shown that much valuable information about human brain dynamics contained in event-related EEG data may be revealed by these new methods. New insights about brain function beginning to emerge from this research would have been difficult or impossible to obtain without first separating and identifying distinct brain processes contained in the data. As these processes have overlapping scalp projections, time courses, and spectra, their distinctive features cannot be separated by simple averaging or spectral filtering. Though ICA now allows nearly any EEG laboratory to assess more of their data than standard analysis procedures, the application of ICA to EEG and ERP data is still in its infancy. This article examines the assumptions underlying ICA and the data selection criteria necessary for successfully applying ICA to unaveraged or averaged event-related EEG data, through analysis of sample data sets collected in continuous and event-related EEG experiments. We suggest that to take full advantage of ICA requires a paradigm shift from modeling the data via trial averaging to modeling the event-related dynamics of independent cortical EEG processes, within and between cortical areas, trials, subjects, and experimental conditions.
1. INTRODUCTION

EEG is a non-invasive record of brain electrical activity measured as changes in potential difference between pairs of electrodes placed on the human scalp. The strength and complexity of scalp EEG means it must arise from partial coherence, in both time course and spatial orientation, of local field activities within several or many cortical source areas or domains. Because of volume conduction through brain tissue, cerebrospinal fluid, skull, and scalp, EEG data collected anywhere on the scalp mixes signals from multiple suitably-oriented cortical areas. Unfortunately, the general problem of determining the spatial distribution of contributing brain sources from an electromagnetic field pattern recorded on the scalp surface is mathematically underdetermined. These facts make it difficult to relate distinct EEG patterns, originating in specific brain areas, to behavior or pathology, or to identify the brain origins of distinct EEG sources.

Nearly all previous research on EEG sources used direct biophysical modeling approaches, either assuming known or simple spatial source configurations (Scherg & Von Cramon, 1986), or restricting the generating dipole elements to lie within and point outward from the cortical surface modeled from magnetic resonance (MR) images (Liu et al., 1998). Such biophysical inverse solutions are reliable only when the number of active sources is very small (e.g., near one), producing activity characterized by at most a few regions of maximum variance on the scalp (Baillet et al., 2001). Since scalp data rarely fits this description, preprocessing the data to minimize the number of active sources is desirable.

The attraction of response averaging methods is based in large part on the hope that averaging similar data trials time locked to sensory input will minimize the influence of cortical EEG sources other than those located in the stimulated sensory cortex. However, the rapid dispersion of stimulus event information through the brain (Hupe et al., 2001) and its early coordination of the phase of local field activities (Klopp et al, 2000) means that only the very earliest ERP features generated in primary sensory areas can be truly isolated by ERP averaging. In general, both multi-channel scalp EEG data and event-related averages of such data sum or linearly mix the projections of a large number of concurrently brain sources, making direct source modeling difficult or impossible.

A newer and radically different approach using independent component analysis (ICA) to separate the conjoined problems of source identification (what) and source localization (where). First formulated by Comon (1994), with early advances by Cichocki et al. (1994), Bell and Sejnowski (1995), Karhunen et al., (1995) Cardoso and Laheld (1996), and many others, ICA refers to so-called ‘blind’ methods for separating and recovering the activities of sources that are mixed linearly in recorded signals, even when their individual natures and dynamics are not known in advance. An easy to grasp (though not technically straightforward) example is the problem of separating so-called ‘cocktail party noise’ recorded simultaneously by a number of microphones, using no prior knowledge of what kinds of sound sources (voices, musical instruments, air conditioning, distant sirens, etc.) produced the recorded signals. The absence of a priori modeling of the sources to be separated cause this to be known as a ‘blind’ source separation problem, to distinguish it from previous methods that mostly involved matched filtering of source models to data. A wide and growing range of applications in signal processing domains as diverse as multi-spectral satellite imaging, wireless communications, acoustic separation, and fetal electrocardiogram (ECG) monitoring attest to the power of the simple idea underlying ICA, that source signals may be identified in linearly mixed multi-channel recordings by the unique information they contribute to the data rather than by their contributions to signal variance.

Applied first to multi-channel EEG data by the co-authors and colleagues (Makeig et al. 1996a, 1997), ICA derives statistically independent sources from highly correlated EEG signals and without regard to the physical location or configuration of the source generators. That is, rather than modeling the EEG as a unitary output of a multidimensional dynamical system, or as a “roar” of a crowd of
independent microscopic field generators, ICA assumes that the EEG is the linear mixture of a finite number of statistically independent but spatially fixed signal generating processes whose cortical distributions may be either spatially restricted or distributed. Over the past decade, blind source separation of EEG data using ICA has provided the co-authors and their colleagues, was well as an increasing number of other research groups, new insights into EEG dynamics and their behavioral and cognitive correlates. ICA has been successfully applied, as well, to many other types of bioimaging data, notably electrocardiographic (ECG) (Cardoso, 1994), magnetoencephalographic (MEG) (Vigário, 1997), electromyographic (EMG) (McKeown & Radtke, 2001), and functional magnetic resonance imaging (fMRI) data (McKeown et al., 1998a; Duann et al., 2002, etc.). In this article we outline the assumptions made in most applications of ICA to EEG data, explore their validity for EEG signals, examine criteria and considerations for preparing data for ICA, and review the contributions to date of this signal-decomposition framework to EEG research.

2. INDEPENDENT COMPONENTS OF EEG SIGNALS

2.1 History
A decade ago, the authors first read in preprint the seminal paper on infomax ICA of Bell & Sejnowski (1995) and set out at once with those colleagues to test its usefulness for EEG data analysis. In the resulting publication, submitted a week later, Makeig et al. (1996a) demonstrated the first applications of ICA to multiple-channel EEG and averaged ERP data recorded from the scalp for separating joint problems of source identification and source localization. Our first journal submission on this topic was refused on the basis of a complaint by an anonymous reviewer who confused independence with decorrelation, despite considerable effort on our part to explain the difference (Makeig, et al., 1996b). Continuing work and publications over the last decade by us (Makeig et al., 1997, 1999a, 1999b, 2002, 2004a, 2004b; Jung et al., 1998a, 1998b, 1999, 2000a, 2000b; 2001a; 2001b; Delorme & Makeig, 2004; Onton et al., 2005) and others (Vigário, 1997; Vigário et al., 2000; Joyce et al., 2004; Kobayashi et al., 1999; 2001; Pritchard et al., 1999; Nam et al., 2002; Melissant et al., 2005; Olbrich et al., 2005, etc.) has demonstrated that ICA does indeed open a new window into EEG analysis. Researchers can now study the activities of temporally and, more interestingly, functionally distinct brain sources surrounding cognitive events and/or subject responses. Applied to EEG data, ICA reveals what distinct, e.g. temporally independent activities compose the observed scalp recordings, separating this question from the question of where exactly in the brain (or elsewhere) these activities arise. ICA also facilitates answers to this second question by determining the scalp projection of each component alone. In the following sections, we examine the assumptions underlying the application of ICA to EEG/ERP data and demonstrate its application to a variety of electrical recordings from the human brain.

2.2 Assumptions of ICA applied to EEG/ERP data

2.2.1 What are EEG sources?

Though typically used without definition, the term “EEG source” may be defined in several ways. In the absence of comprehensive invasive measurements, the brain network or networks that produce the activity of a single EEG “source” must be defined primarily by the distributed field measurements they produce at the available scalp sensors. Here, an EEG source may be defined as the sum of local electromagnetic field activities that are synchronously active and fully or partially phase-coherent across some laminar cortical source region, thereby projecting a single far-field electromagnetic signal to the recording electrodes. In the simplest case, the source region might be a single cortical patch of some size and shape. By biophysics, the characteristic orientation of cortical pyramidal cells, mainly perpendicular to the cortical surface, in such a source domain or region will produce a detectable far-field signal on the scalp with a characteristic ‘dipolar’ spatial distribution (Scherg & Von Cramon, 1986). This distribution
will correspond to the projection of a single tiny oriented patch or battery, the equivalent source dipole typically located below the actual cortical generator area (Baillet et al., 2001).

Precisely synchronized activities in two or more cortical patches, supported by dense spike-mediated coupling, might also be defined to comprise a single EEG source, since these patches should also create a synchronous and spatially stable signal at the scalp sensors. The term EEG source, defined in this way, could also apply to synchronous activity in more widely distributed neural networks. However, the very high concentration of local (< 500 µm) coupling in the cortical neuropile (Salinas & Sejnowski, 2001), together with the spatial fixity constraint, make more diffuse source configurations difficult to imagine. The term EEG “source” might also be applied to more complex local field phenomena, such as moving waves of activity on the cortical surface (Huber et al., 2004). However, the projections of activity of such processes to the scalp would, in general, not be fixed or easy to characterize or to separate from other signals in the scalp data. Here, therefore, we will use the term EEG source to imply the time-varying modulation of activity with a fixed scalp distribution.

Our own interest in using ICA in EEG research was heightened when we noticed that the topographic projections of independent EEG components to the scalp often closely matched the projections of synchronized activity that would be generated within a single spatially compact cortical area. As mentioned above, such patch-like source activity should produce a ‘dipolar’ pattern on the scalp, i.e. matching the projection of a single, vanishingly small ‘equivalent dipole’ located (in most instances) beneath the actual cortical source patch (Makeig et al, 2004a). That many ICA component maps have a near-purely dipolar form -- although ICA decomposition itself incorporates no information about the relative locations of the electrodes or the composition of the head -- strongly suggests that ICA finds patch-like EEG sources which biophysical reasoning (above) and some invasive measurements (Freeman, 1975) suggest should be the simplest and biophysically best supported source configuration.

Other ICA component scalp maps match the projections of two spatially distinct source regions whose coherent field activities must therefore be largely synchronous. Support for this coupling should require that the two regions are densely connected. In practice, such ICA component maps are well fit by dual symmetric equivalent dipole models assuming two source patches that are near bilaterally symmetric across the midline, compatible with left and right source patches densely connected via the corpus callosum. Mathematically, the generator domains of EEG sources obtained by ICA need be limited to single or dual-symmetric cortical patches, or to artifact generators (eyes, muscle, etc.). Ongoing research using ICA as pre-processing to cortex-based source inversion algorithms, as well as ICA applications to intracranial data, should lead to further insights about the spatiotemporal properties of synchronous cortical field activity.

2.2.2 ICA Assumptions

Standard, so-called complete and instantaneous ICA algorithms are effective for performing source separation in domains where: (1) the summation of different source signals at the sensors is linear, (2) the propagation delays in the mixing medium are negligible, (3) the sources are statistically independent, 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 we can separate N sources (Makeig et al., 1996a).

Linear mixing. The first two assumptions above, that the underlying sources are mixed linearly in the electrode recordings without appreciable delays, is assured by the biophysics of volume conduction at EEG frequencies (Nunez, 1981) and is the basis for any type of linear decomposition method including those based on principal component analysis (PCA). That is, the EEG mixing process is fortunately linear, though the processes generating it may be highly nonlinear. In current applications, ICA attempts only to ‘undo’ the linear mixing produced by volume conduction and linear summation of fields at the electrodes.
Independence. Assumption (3), of independence or near independence of the underlying source signals, is compatible with physiological models that emphasize the role of anatomically dominant local, short-range intracortical and radial thalamocortical coupling in the generation of local electrical synchronies in the EEG (Salinas and Sejnowski, 2003). As well, corticothalamic connectivity supporting oscillatory field activity is mainly radial, not dispersive (Lopes da Silva & Storm van Leeuwen, 1978). These facts suggest that synchronous field fluctuations should arise within compact cortical source domains, though they do not in themselves determine the spatial extent of these domains. If we assume, therefore, that the complexity of EEG dynamics can be modeled, in substantial part at least, as summing activities of a number of very weakly linked, and therefore nearly statistically independent brain processes, each produced and supported by dense short-scale coupling of excitatory and inhibitory networks within one or more patches of cortical neuropile, EEG data should satisfy assumption (3).

Synchrony. However, even if cortical source activities are largely generated by synchronous activity in compact spatial domains, as we hypothesize for many of the sources identified by ICA, the concept that precise and exact synchrony of field activity prevails across each domain is an idealization. Walter Freeman has reported the formation of small gamma band ‘phase cones’ under his small intracortical electrode arrays, which he likens to ripples spreading out from a local disturbance in a pool of water (Freeman & Baird, 1987). However, at the reported speed of propagation of local fields across cortex, near 2 meters per second (Lopes da Silva & Storm van Leeuwen, 1978), the phase delay of a 10-Hz alpha oscillation from the center to one edge of a radius 1-cm source patch (5 ms) would be only 18 degrees, meaning alpha activity in the patch would be nearly synchronous and the alpha activity projected by the patch on the scalp should strongly resemble the activity of a completely synchronous source patch. The actual spatial scale of frequency-domain coherence on rabbit cortex was observed by Bullock et al (1995) to be less than 1 cm. At that size, the near synchrony of activity within each source patch would be still more exact at alpha frequencies, and equally as strong at 20 Hz (beta). Others, however, would model the generation of EEG activity in cortex by global standing or traveling wave processes (Nunez, 1981; Linas & Ribary, 1993). So far, simultaneous electrical measurements at a range of spatial scales that would allow testing of competing EEG source models are not yet available.

Meaning of independence. One common source of confusion about ICA comes from the fact that the term ‘independence’ may itself have many meanings, while each ICA algorithm uses a strictly limited definition of this concept. In common usage, one would not think of a chorus of $N$ voices singing the birthday song, “Happy birthday to you....” in unison as being independent of each other, since several features of one of the voices would predict features of each of the others (e.g., their pitch, phonemic content, etc.). Yet, to the most commonly applied ICA algorithms each of the singers’ voices will be independent of the others. How is this so?

These ICA algorithms define independence in the sense of instantaneous prediction only. They ask, in effect, do the sound pressure levels of one or more voices, recorded at the same time instant, predict the simultaneously recorded sound pressure level of any of the other voices? This would only be the case if the voice signals were phase-coupled – in real life this should in fact occur only in extremely rare or artificial situations. Therefore, to such ICA algorithms these voice signals are (near) independent and could theoretically be separated from $N$ recordings, although ICA separation of real acoustic data is strongly complicated by the finite speed of propagation of sound in air and the addition of echoes from solid surfaces-- neither a problem for EEG data decomposition.

In this sense, EEG data are near optimal for application of the relatively simple and powerful ICA algorithms that maximize instantaneous independence. However, the distinction between independence in general and instantaneous independence must be clearly understood to avoid falling into a trap of feeling that ICA does not use the ‘correct’ meaning of term independence. ICA algorithms built on other,
somewhat difference measures of independence also exist and may also be useful for EEG analysis (Tang et al., 2000, Annemuller et al., 2003; Mads Dyrholm et al., 2004a; 2004b).

Limits to independence. Another ready stumbling block for those wanting to understand the application of ICA to EEG data is the fact that absolute independence of even two source signals is impossible to measure or establish from any finite amount of data. Therefore, any ICA algorithm given a finite sample of data from two or more sources can at best only separate sources that are maximally independent in the particular sense of independence used in the algorithm. Rather than being a ‘bug’ of ICA, this can be seen as a desirable ‘feature’ that allows useful estimation of separate sources from relatively small data sets or separation of nearly or nearly always independent sources from EEG data. In either case, it is important to remember that the sources identified by ICA can only be said to be maximally independent in a particular sense of the term.

*How many sources can be recovered?* ICA assumption (4), that N-channel EEG data mix the activities of N or fewer sources, is certainly questionable, since we do not know in advance the effective number of statistically independent brain signals contributing to the EEG recorded from the scalp. Hypothetically, field activities at various spatial scales (even around single synapses!) could be largely independent. This should mean that EEG scalp signals may sum some relatively small number of relatively large sources with numerous (or numberless) smaller and yet smaller sources.

To test the resilience of infomax ICA, in particular, to the presence of more small sources than recording channels, Makeig and colleagues (1996b, 2001) performed a number of simple 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. 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 when EEG training data consists of a mixture of fewer large source components than channels, plus many more small source components, as might be expected in actual EEG data, large source components are accurately separated into separate output components, with the remaining output components consisting of mixtures of smaller source components. In this sense, performance of the infomax ICA algorithm they used was found to degrade gracefully as the number of sources in the data increases or alternatively as the number of scalp channels decreases.

*How much data can be decomposed?* There is no definite rule to the number of points needed for each ICA decomposition. However, given, say, 64 channels and only 64 points, ICA may well give a trivial and degenerate solution in which one component is fit to each time point, providing no additional information or insight into the data. Experience suggests that the number of time points required for ICA decomposition is at least several times the square of the number of scalp sensors (data channels), since ICA decomposition amounts to learning an ‘unmixing’ matrix of size channels-by-channels. For very large numbers of channels (> 128), the ratio of time points to matrix weights may need to be large (> 30), in our experience, though this ratio may be strongly affected by data quality and details of the decomposition. The maximum number of time points useful for the decomposition is theoretically unlimited. However, as mentioned above, for ICA decomposition of very large data sets, the number of data channels (and scalp sensors) should exceed the number of expected independent data components. Decomposing too much data from different subject tasks and states, with relatively few channels, might therefore require ICA to return mixtures of sources that appear in different tasks or states.

*How many channels can be decomposed?* Our studies have successfully analyzed EEG data including 31, 71, 128 and 256 channels. We have not yet demonstrated a successful 256-dimensional decomposition, though wide availability of 64-bit systems with large amounts of random access memory (> 2GB) may make decomposition of very large dense-array data sets straightforward. Alternatives are to remove some
data channels or to reduce the dimensionality of the data by removing its smallest principal components before decomposition. Neither strategy is without drawbacks, though we recommend that dimensionality reduction by principal component analysis be used only as a last resort.

Relaxing ICA assumptions. More recently, several approaches have been suggested for further relaxing the assumptions of ICA. For example, overcomplete representations (Lewicki and Sejnowski, 2000) can separate more sources from the data than the available number of sensors (Lee, 1998), and Bayesian techniques may be used to infer the number of sources in addition to estimating the source signals (Chan et al., 2002). The simple instantaneous mixing model can be extended to include time delayed and convolved source signals (Torkkola, 1996; Yellin and Weinstein, 1996; Lambert, 1997; Lee, 1998; Visser et al., 2003; Dyrholm & Hansen, 2004). ICA researchers are also trying to further weaken the core assumptions of ICA, including the linearity of the mixing and the assumptions that every pair of sources are independent of one another (Taleb & Jutten, 1999; Hyvarinen & Pajunen, 1999; Lee et al., 2000; Taleb, 2002; Almeida, 2003; Jutten & Karhunen, 2003). However, these algorithms involve other assumptions about the data that may prove unrealistic for EEG data, and/or may not be applicable to large, high-density data for which instantaneous ICA algorithms, in particular infomax ICA, may shine.

2.3 Basic ICA Math

This section focuses on the technical aspect of ICA decomposition, based on the derivation of infomax ICA by Bell & Sejnowksi (1995). Further mathematical details are contained in the Appendix which is recommended for readers with sufficient background. Independent component analysis was originally proposed to solve the blind source separation problem of recovering \( N \) source signals, mathematically a set of vectors of positive or negative activity values at each time point, \( \mathbf{s}(t) = \{s_1(t), \ldots, s_{N(t)}\} \), after they are linearly mixed by multiplying by an unknown mixing matrix, \( \mathbf{A} \), while assuming as little as possible about the natures of \( \mathbf{A} \) or of the source signals themselves. \( \mathbf{A} \) gives the relative strength with which each source propagates to each scalp channel. In general, this should depend on the distance between the source and the electrode, the angle of the cortical surface to the line connecting the two, and the sensitivity of the individual electrode. The problem can be formulated as follows. Given observed data \( \mathbf{x} \), find matrices \( \mathbf{A} \), \( \mathbf{s} \) and \( \mathbf{n} \) such that

\[
\mathbf{x} = \mathbf{A} \mathbf{s} + \mathbf{n}
\]  

(1)

Above, \( \mathbf{A} \mathbf{s} \) (‘\( \mathbf{A} \) times \( \mathbf{s} \)’) are the mixed source signals, and \( \mathbf{n} \) is (generally small) additive noise, such as electrode channel noise. Due to its generality, there are many physical phenomena that can be described by the above equation. In acoustic environments, for example, \( \mathbf{x} \) might be recordings from an array of microphones in which the acoustic signal sources \( \mathbf{s} \) are mixed with weights \( \mathbf{A} \). Although the general solution of (1) is computationally intractable, there are certain reasonable assumptions or constraints that can lead to practicable solutions.

A key assumption in ICA is the statistical independence of the source signals. Statistical independence means the joint probability density function (p.d.f.) of the output factorizes:

\[
p(\mathbf{s}) = \prod_{i=1}^{M} p(s_i)
\]  

(2)

where \( i = 1, \ldots, M \) are the \( M \) source signals. This implies that each source signal contributes distinct information to the data, i.e. the total amount of information in the mixed data equals the sum of the information contents of the separate source signals, since the source signals are independent and not redundant. Note that decorrelation, commonly used in statistical analysis to identify principal directions of variance in data, means only that \( \langle \mathbf{s} \mathbf{s}^\top \rangle \), the covariance matrix of \( \mathbf{s} \), is diagonal (here \( \langle \cdot \rangle \) means
average. Decorrelation is much easier to achieve than independence, since it says nothing about relationships among sets of more than two component activities in s, which in fact may cooperate in uncorrelated data. Also, decorrelation is not unique, since any orthogonal rotation matrix B applied to s will give a mixing matrix Bs of signals that are also uncorrelated, by

\[ x = As + n = A(B^{-1}B)s + n = (AB^{-1})(Bs) + n \]  

However, if x is truly a mixture of independent sources s, then applying any matrix B to s (other than one that simply rescales and/or transposes the order of the signals s) will yield new signals Bs that are not independent. ICA algorithms exploit this fact to iteratively locate and refine an unmixing matrix (defined below) that identifies ever more independent putative source signals in the data.

Low-noise solution. For EEG decomposition, the sensor noise matrix n is usually assumed to be negligible or else adequately accounted for by a few of the recovered ICA sources. In fact, in good EEG acquisition systems, acquisition noise on most channels may be minimal. If in addition it can be assumed that N = M, then the blind source separation problem reduces to recovering a version u of the original source signals, s, identical to s except for source order and scaling. This involves finding a square matrix, W, specifying a spatial filter of the observed data x, that linearly inverts the mixing process and makes the output source signals, u, statistically independent,

\[ u = Wx \]  

If the mixing matrix A is square (i.e., N = M) and known, then the unmixing matrix W sought is simply A^{-1} since then

\[ u = Wx = A^{-1}x = A^{-1}(As) = (A^{-1}A)s = s \]  

When A is unknown (as is normally the case), deriving an algorithm to learn W from the signals x alone involves formulating an initial guess W0, a cost function specifying the distance of each estimated set of source signals Wkx from a set of independent signals, and an optimization process to find an estimate, Wk+1, that is closer to the ideal u. There are many possible cost functions and many more optimization processes. Thus there are many somewhat different algorithmic approaches to solving the blind source separation problem (see Appendix). In infomax ICA, in particular (Bell and Sejnowski, 1995), the cost function penalizes outputs whose joint entropy is less than maximal. Intuitively, if the output N-channel data points are spread out evenly (like molecules of a gas) throughout their range (an N-dimensional hypercube), then knowing the value of a data point on any number of the hypercube axes gives no statistical information about its value on the other axes, and maximum independence has been achieved.

Dimensionality. The numbers of source signals, M, and sensor signals, N, are very important in determining the degree of difficulty of the ICA solution. In standard ICA algorithms, N (the number of recording channels) is assumed to be greater than or equal to M (the number of maximally independent sources). This means that the matrix A (when N = M, a square matrix) is invertible and the sources can be recovered. If N is smaller than M, the problem is called underdetermined and is more difficult. Linear mixing of the source signals (As) is another important assumption that in general may or may not hold depending on the type of physical mixing process.

ICA as beamformer. Recently so-called beamforming methods first developed for radar have been applied to magnetoencephalographic (MEG) data analysis for estimating activity at arbitrary cortical locations from scalp-recorded data. In standard beamforming (Oostenveld et al., 2004; Hillebrand et al., 2005), the signals from each brain voxel are considered to be uncorrelated with each other, and the spatial filter for each voxel activity is constructed to minimize the leakage of other sources of variance in the data into the filtered data. Beamforming approaches begin with a forward model of how activity at each brain voxel should project to the scalp. Since the forward head model for EEG is, at least at first order, more difficult
to construct accurately than for MEG, beamforming methods have so far been applied more to MEG than to EEG.

Application of temporal ICA to EEG data can also be viewed as an alternate beamforming approach. ICA begins with the statistically more powerful assumption that the activities of the cortical sources, whatever their spatial distributions, are temporally independent. This allows use of a forward head model only after ICA decomposition has identified the scalp map projections of the sources, for estimating (by any inverse source localization method) the cortical locations of the identified sources.

2.4 What Are Independent EEG Components?

For EEG/ERP data, the rows of the input matrix $\mathbf{x}$ in (4) are EEG/ERP signals recorded at different electrodes and the columns are measurements recorded at different time points. ICA finds an ‘unmixing’ matrix, $\mathbf{W}$, that decomposes or linearly unmixes the multi-channel scalp data into a sum of temporally independent and spatially fixed components, $\mathbf{u} = \mathbf{Wx}$. The rows of this output data matrix, $\mathbf{u}$, called the *component activations*, are the time courses of relative strengths or levels of activity of the respective independent components through the input data. The columns of the *inverse* of the unmixing matrix, $\mathbf{W}^{-1}$, give the relative projection strengths of the respective components onto each of the scalp sensors. These may be interpolated to show the *scalp map* associated with each component. These scalp maps provide very strong evidence as to the components' physiological origins (for example, vertical eye movement activity projects principally to bilateral frontal sites) and may be separately input into any inverse source localization algorithm to estimate the actual cortical distributions of the cortical area or areas generating each source. Note that each independent component of the recorded data is specified by *both* a component activation and a component map – neither alone are sufficient.

*Noise components.* For some component maps, the best-fitting simple single or dual-symmetric equivalent dipole model, in particular, may account for little of the spatial variance in the component map. In practice, such ‘noisy-appearing’ components typically contribute little to the observed event-related dynamics (for instance, ERP and event-related spectral perturbation or ERSP measures of interest) and together may account for some activity in the recorded signals that does not well fit the simplified model (4), since the implied source distribution for these components, e.g. synchronous field activity sparsely distributed throughout cortex, is usually implausible. Other components may be active at only a single scalp channel. Comparing their activations to the original data recordings usually reveals that these components account for (and reveal) single channel noise (e.g., during periods when the affected electrode has dried out or become loose from the scalp).

*Relation to PCA decomposition.* ICA component activations are similar to the *factor weights* produced by spatial principal component analysis (PCA), while ICA scalp maps are similar to spatial PCA eigenvectors or *factor loadings*. Unlike components produced by PCA, however, component scalp maps found by ICA are not constrained to be both spatially and temporally orthogonal, and are thus free to accurately reflect the actual scalp projections of the functionally distinct and, across time, near-independent sources, if they are successfully separated. In essence, the goals of PCA and ICA are nearly opposite. The goal of PCA is to include as much source *variance* in each successive component as possible, whereas the goal of ICA is to separate independent sources of temporal *information* in the data as cleanly as possible. Thus we can say that applied to EEG data, as far as possible *PCA lumps* while *ICA splits* the data into temporally distinct EEG processes. The primary usefulness of PCA for EEG analysis is to reduce the dimensionality of the data, although for most accurate ICA decomposition the use of PCA should be limited as much as possible.
Figure 1. (A) A scatter plot of ten target responses (middle panel) at two scalp electrodes, Fz and Pz, contains two strongly radial (i.e. spatially-fixed) features. The dashed lines show the directions associated with components P3b and Pmp in these data, as determined by the values of their respective component scalp maps (black dots on the scalp maps). ICA found the two component directions by maximizing joint entropy (i.e. the evenness of the data distribution) of a nonlinear transform of the (31-channel) unmixed data (center right insert). [Adapted from Makeig et al., 1999 by permission] (B) The schematic illustration of the back-projection of a selected component onto the scalp channels.

Illustrative example. Figure 1A provides geometric insight into how the infomax ICA algorithm decomposes EEG/ERP data. The figure (adapted from Makeig et al., 1999) shows 10 detection-task target responses (here, response averages from different experimental conditions) at two midline scalp electrodes (Fz and Pz). In this scatter plot format, time is not represented explicitly. The data time series generally follow a cyclic trajectory. Moving radially out from the origin in any direction in this graph represents intensifying activity in a particular scalp map. For example, moving out along the Fz axis represents increasing intensity in a map projecting only to Fz, with 0 projection to Cz. The plot shows the two radial directions (dashed lines) corresponding to the scalp maps (here reduced to only two channels) of two ICA components of a set of 25 averaged condition ERPs from this experiment, including the 10 shown here. The two component directions align with the most nearly radial portions of the data trajectories. These represent periods when the scalp maps of the responses at the two channels were changing only in amplitude but not in scalp distribution.

The spatial structure of the data scatter plot (Fig. 1A) resembles an oblique parallelogram rather than a Gaussian data cloud. Infomax ICA decomposition here found its periods of strongest spatial stationarity,
and in so doing found the axes that could best transform the irregular shape of the input data scatter plot into a near-evenly filled square (Fig 1A, right plot insert), thereby maximizing its joint entropy. In contrast, PCA, applied to these data, would in effect fit a 2-D Gaussian distribution to the whole data cloud, returning only its major and orthogonal minor axes. In this case, the first principal component (not shown) would point in a direction resembling but not matching that of the independent component labeled P3b, and the second principal component, orthogonal to it, would ignore the sizable contribution to the data along the direction of the independent component labeled Pmp. Note that the two independent component scalp maps (Fig 1A, bottom plots) are well correlated (r = 0.888), whereas the PCA axes (and hence, principal component maps) would be constrained to be orthogonal.

Component projections. The projection of the ith independent component back into the original data channels, called the back-projection of the component to the data, is given by the outer product of the ith row of the component activation matrix, u, with the ith column of the inverse unmixing matrix, W, or in matrix terms, \( W^{-1} u \). Back-projected activity is at the original channel locations and in the original recording units (e.g., \( \mu V \)). Typically, neither the scalp maps nor the activation time series found by ICA algorithms are themselves normalized. In this case, scaling information about the actual strength in \( \mu V \) and polarity (+ or -) of their activity in the original data is distributed between them. The true magnitude and polarity of the component activity is given only by its back-projection.

Thus, brain activities of interest accounted for by single or by multiple components can be obtained by projecting selected ICA component(s) \( k \) back onto the scalp, \( x_k = W^{-1}_k u_k \), where \( u_k \) is set of the activation matrix rows for components in set \( k \) and \( W^{-1}_k \) is the scalp map matrix columns \( k \). Figure 1B schematically depicts the projection of the P3b component onto the scalp channels. This is also easily computed by setting the irrelevant component activations to zero. In this case, columns of the inverse unmixing matrix, \( W^{-1} \), associated with these components become non-factors in the back-projection, whereas the column of the inverse unmixing matrix associated with the P3b component determines the amplitude distribution of the P3b across scalp channels, as interpolated and plotted in Fig. 1 (lower left scalp maps).

For each component, the distribution of current across the scalp electrodes is fixed over time, but the actual potential values (including their polarities) are modulated by the corresponding time course of component activation (red traces in Fig. 1B), the relevant row of the output data matrix, \( u \), depicted in the lower panels as the intensity fluctuations of the scalp maps over time.

2.5 Applying ICA to EEG Signals

2.5.1 Independent components of averaged ERP data

Averaged event-related potentials (ERPs) capture the often small portion of the recorded EEG data time-locked and partly phase-locked to some set of experimental events. Through averaging across a set of epochs time-locked to a set of experimental events considered to be equivalent, EEG activity not both time-locked and at least partially phase-locked to the events is removed by phase cancellation. What remains is known as the event-related potential (ERP).

The application of ICA to ERP data was first systematically tested using two averaged 14-channel 1-s (312-point) ERPs time-locked to detected and undetected targets, respectively, presented in an experiment in which the subject responded by pressing a button each time he heard a weak, slow-onset noise burst (6 dB sensation level) embedded in a continuous 62-db noise background (Makeig et al., 1996a, 1997). These ERPs contained an auditory steady-state response as well as standard auditory evoked response features. Subsequent work (Makeig et al., 1999a, 1999b) demonstrated the efficacy of ICA for decomposing functionally distinct components of early and late features of responses to flashed visual stimuli from sets of 25 to 75 averaged ERPs (31 channels, 512 time points) time locked to onsets of target
or non-target stimuli presented at each of 5 stimulus locations in 5 spatial attention conditions (Makeig et al., 1999a, 1999b).

Subsequently, several researchers reported exploring the applicability of ICA to their averaged ERP data without obtaining comprehensible results (personal communication). The majority of these had attempted to decompose a single ERP average from one task condition rather than a collection of ERP averages from several or many task conditions, as suggested by Makeig et al. (1996a, 1997, and 1999a, 1999b). A first issue needs to be considered when decomposing a set of averaged ERPs is the number of time points used in the decomposition. The assumption of temporal independence used by ICA cannot be satisfied when the training data set is too small. The number of time points required for training is proportional to the number of variables in the unmixing matrix (the square of the number of channels), although in well-averaged ERP data with relatively few strong components, fewer time points may be needed to achieve a useful result than with unaveraged EEG data.

However, it is important to consider which EEG processes may express their independence in the ERP training data. As mentioned above, ICA assumes that the component time courses are temporally near independent. In the case of ERP components whose active periods are temporally overlapping in each response average, this may be achieved or approximated only by sufficiently and systematically varying the experimental stimulus and task conditions, creating an ERP average for each stimulus/task condition, and then decomposing the concatenated collection of resulting ERP averages. The component activities separated by ICA in this case will be those whose independence, across response latencies and/or conditions, is best expressed in the input data averages.

However, simply varying stimuli and tasks does not always guarantee that all the spatiotemporally overlapping EEG processes contributing to the averaged responses will be independently activated in the ensemble of input data. Independent control of temporally overlapping component activities may often be difficult or impossible to produce in ERP averages. Examples of processes unlikely to be separated by ICA are parallel ERP features generated in left and right auditory cortices by auditory stimuli. In this case, ICA would very likely fuse both activations into a single component, unless appropriate experimental interventions were developed to block or delay each component contribution independently in one or more of the training conditions. Another example is the bilaterally symmetric ERP component labeled Pmp (for post-motor positivity) in Fig. 1A. Subsequent application of ICA to the concatenated single trials from the same experiment accounted for this ERP component largely as summing concurrent contributions from independent left and right mu-rhythm EEG processes, whose summed effects were mixed in the ERP-based component.

These issues imply that results of ICA decomposition of averaged ERPs must be interpreted with caution, and converging behavioral or other evidence must be obtained before concluding that spatiotemporally overlapping ICA components measure neurophysiologically or functionally distinct activities.

2.5.2 Independent components of event-related EEG epochs

Response averaging, used to create average ERPs, ignores the fact that the EEG activity averaged in the ERP may vary widely across trials in both time course, polarity or phase, and scalp distribution, reflecting the differing activities of several or many underlying EEG processes that contribute to each scalp channel recording. The temporal and spatial variability may reflect trial-to-trial differences in subject performance or state possibly linked to changes in attention, arousal, task strategy, or other factors. Trial-to-trial activity of ongoing EEG processes may become partially phase locked following salient events (Makeig et al., 2002). Hence, there need be no clear physiological separation between evoked (phase-locked) and induced (non-phase locked) activity. Indeed, except for some early sensory ERP features, this appears to be the rule (Makeig et al., 2004a). Hence, conventional averaging methods may not be suitable for investigating brain dynamics arising from intermittent changes in subject state...
and/or from complex interactions between task events, i.e. EEG changes linked to the brain’s response to trial-to-trial differences in event meaning or context.

In this case, trial-by-trial analysis of event-related trial epochs may reveal more information about event-related brain dynamics than simple response averaging. Note that applying ICA to single event-related data epochs should suffer from the same problems obtaining in ICA decomposition of a single ERP average, and therefore should be avoided. Rather, ICA decomposition should be more usefully applied to a large set of concatenated single-trial data epochs. Systematic application of ICA to unaveraged event-related EEG data was first demonstrated for a set of EEG epochs time locked to target and non-target stimuli (Jung et al., 1999; 2001a; 2001b; Makeig et al., 2002, 2004a). In these analyses, 1-s epochs (from 100 ms before to 900 ms after stimulus onset) were extracted either from 300-700 target trials or from over 3,000 non-target trials for each subject. These data sets were then concatenated and decomposed by ICA. In this case, the columns of the input data matrix are measurements recorded at different latencies in each trial. The total number of time points in the matrix is the number of trials times the number of time points in each trial.

Applied to a set of event-related data epochs, ICA algorithms can separate the most salient concurrent EEG processes active within the trial time windows. Simultaneous analysis of a set of hundreds of single-trial EEG epochs gives the concurrently active processes that contribute to the response and/or the response baseline a far better chance of expressing their temporal independence and thus being available for identification by ICA.

The authors have often been asked whether there is any problem with decomposing discontinuous or partially overlapping data epochs. Since standard ICA algorithms consider only instantaneous independence, the presence of temporal discontinuities in the training data has no effect on it. In fact, the infomax ICA decomposition method we use regularly shuffles the order of the input data points during training for statistical purposes (Makeig et al., 1997; Delorme & Makeig, 2004). However, some blind source separation algorithms such as SOBI (Tang et al., 2000) make use of temporal correlations in the data. Such algorithms may need to be adjusted to deal with discontinuous data epochs. Overlapping epochs, meanwhile, may be influenced more by the data at overlapping time points than by data represented only once in the epoched data. If the sources of the data are truly independent and spatially fixed, this may have no effect on the decomposition.

Our studies (Jung et al., 2001a; 2001b; Makeig et al., 2002, 2004a, see examples below) have shown that relatively small numbers of independent components exhibited robust event-locked and/or non-phase locked activities near stimulus presentation and/or the subject motor response. These components tended to have near-dipolar scalp maps, compatible with a compact cortical source area and suggesting that the brain areas exhibiting responses to these experimental events were indeed spatially stable across epochs and latencies, as assumed for ICA. This supposition is concordant with repeated observations in functional brain imaging experiments that discrete, spatially restricted areas of cortex are activated during task performance (Friston et al., 1998). In section 3.2 below, we give examples of several classes of EEG processes separated by ICA from collections of task event-related data epochs.

2.5.3 Independent components of continuous EEG data

ICA is also applicable to the analysis of continuous EEG data as first reported by Makeig and colleagues (1996a) and explored by Jung et al. (2000a). In contrast to the decomposition of averaged ERPs, the amount of training data available from continuous EEG records is usually sufficient to decompose 32 or more data channels. For example, two minutes of 32-channel recorded at a 250 Hz sampling rate will include 30,000 data points, before artifact data rejection, or just over 29 time points per weight in the learned 32x32 ICA unmixing matrix.

However, ICA assumes that the physical sources of artifactual and neural activity contributing to EEG signals are spatially stationary through time. EEG source areas need not necessarily remain fixed, particularly through changes of state of consciousness or arousal. Examples of spatially non-fixed sources
appear to include spreading sleep spindles (McKeown et al., 1998b) and K-complexes (Huber et al., 2004) as well as epileptic seizures (McKeown et al., 1999).

Statistically, it is best to use all available data to reliably derive spatial filters characterizing the time courses and spatial scalp projections of artifacts in the EEG. However, this is only true when the physical sources of artifacts and neural activity are spatially stationary through time, and the total number of such large and separable sources is less than or equal to the number of data channels. Thus concatenating data from many tasks and states may be problematic. A possible solution is to perform separate decompositions of subsets of the recorded data, each consisting of periods during which the sources may be stationary (e.g., without marked changes in subject state).

It is also important to note that ICA is quite sensitive to changes in channel locations. For one data set we analyzed, ICA returned two sets of quite similar components. Studying their activations, we found a point in the data when the first set became inactive and the second set active. The component maps suggested that at that moment the electrode montage had shifted slightly on the head! Therefore, decomposition of concatenated data from multiple sessions, even from the same subject, is unlikely to return optimum results. Possibly, physiologically-based interpolation of electrode activities in multiple sessions to a common subject montage could allow ICA to find a common set of independent components, though we have not yet attempted this and believe it may prove difficult.

3 CONTRIBUTIONS OF ICA TO EEG/ERP ANALYSIS AND INTERPRETATION
3.1 ICA-based EEG Artifact Detection and Removal

One of the most pervasive problems in EEG analysis and interpretation is the interference in the data produced by often large and distracting artifacts arising from eye movements, eye blinks, muscle noise, heart signals, and line noise. Asking subjects to fixate a visual target may reduce voluntary eye-movements (blinks and saccades) in cooperative subjects during brief EEG sessions, but it does not eliminate involuntary eye-movements and cannot be used when the subject is performing a task that requires eye movements. Therefore, artificial tasks in which the subject is asked to stare at a fixation cross have been standard in EEG studies, and data from frontal and temporal electrodes located near the eyes or scalp muscles have often been ignored or discarded since these scalp areas are more heavily contaminated by artifacts than central and posterior scalp channels.

Another common strategy for avoiding EEG artifacts is to reject all EEG epochs containing artifacts larger than some arbitrarily selected EEG voltage value. However, when limited data are available, or when blinks and muscle movements occur too frequently as in children and some patient groups, the amount of data lost to artifact rejection may be unacceptable. For example, Small (1971) reported a visual ERP experiment conducted on autistic children who produced electrooculographic (EOG) artifacts in nearly 100% of the trials. In this case, the presence of large background EEG signals not time- and phase-locked to experimental events may make ERP averages of the few artifact-free trials too unstable to permit useful analysis. In addition, artifacts smaller than the arbitrarily set threshold may also adversely affect the data analysis.

Three examples. Figure 2 shows a 5-s portion of continuous EEG time series data collected from 20 scalp electrodes placed according to the International 10-20 System and from two EOG electrode placements, all referred to the left mastoid. The sampling rate was 256 Hz. In this example, ICA was trained with 10 s of spontaneous EEG data (5.2 time points per weight). Figure 2B shows component activations and scalp topographies for five of the independent components. The eye movement artifact (between seconds 2 and 3) was isolated by ICA to components IC1 and IC4. The scalp maps indicate that components IC1 and IC4 account for the spread of EOG activity to frontal sites. Components IC12, IC15, and IC19 evidently represent muscle noise from temporal and frontal muscles. After eliminating these five artifact components, by zeroing out the corresponding rows of the activation matrix $u$ and projecting the remaining components to the scalp electrodes, the 'corrected' EEG data (Fig. 2C) are free of both EOG
and muscle artifacts. The artifact-corrected data reveal underlying EEG activity at temporal sites T3 and T4 (Fig. 2C) that was well masked by muscle activity in the raw data (cf. Fig. 2A).

![Figure 2](image)

**Figure 2.** Demonstration of EEG artifact removal by ICA. (A) A 5-sec 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). [From Jung et al, 2000a by permission].

The second example (Fig. 3) demonstrates that ICA can also be used to remove stimulus-induced eye artifacts from unaveraged event-related EEG data through analysis of a sample data set collected during a selective attention task. The subject performed a visual spatial selective attention task during which he/she covertly attended one of five squares continuously displayed on a back background 0.8 cm above a centrally located fixation point. Four squares were outlined in blue, while one, marking the attended location, was outlined in green. The location of this green square was counterbalanced across 72-s trial blocks. The subject was asked to press a right-hand held thumb button as soon as possible following target stimulus presentations in the attended location (green box), and to ignore the similar (non-target) stimuli presented in the other four boxes. Stimuli were white disks, presented in one of the five boxes at random. EEG data were collected from 29 scalp electrodes mounted in a standard electrode cap (Electrocap, Inc.) at locations based on a modified International 10-20 System, and from two periorcular electrodes placed below the right eye and at the left outer canthus. Data were sampled at 512 Hz (downsampled to 256 Hz) with an analog pass band of .01-50 Hz. Although the subject was instructed to fixate the central cross during each block, he tended to blink or move his eyes slightly towards target stimuli presented at peripheral locations.

After ICA training on 555 concatenated trials (147 time points per matrix weight), independent components accounted for blinks and eye-movements were identified by the procedures detailed in (Jung et al., 2001a) based on the characteristics of time course of component activations, the component scalp topographies and the locations and orientations of equivalent dipoles obtained from BESA (Scherg, 1990). Here, ICA successfully isolated blink artifacts to a single independent component (Fig. 3A) whose
contributions were removed from the EEG record by subtracting its component projection from the data (Jung et al., 1998a, 1998b, 2000a, 2000b).

Though the subject was instructed to fixate the central cross during each block, the technician watching the video monitor noticed that the subject's eyes also tended to move slightly towards target stimuli presented at peripheral locations. A second independent component accounted for EEG artifacts produced by these small horizontal eye-movements (Fig. 3B). Its scalp pattern is consistent with that expected for lateral eye movements. Note the overlap in scalp topography between the two independent components accounting for blinks (Fig. 3A) and for lateral eye movements (Fig. 3B). ICA component maps need not be orthogonal and may even be nearly spatially coincident.

A standard approach to ERP artifact rejection is to discard eye-contaminated trials containing maximal potentials exceeding some selected value (e.g., ≥ 60 µV) at periorcular sites. For this dataset, this procedure rejected 78 of 555 trials, or 14% of the subject's data. Fig. 3C shows ERP averages of relatively uncontaminated target trials (red traces) and of the contaminated target trials (blue traces) that would have been rejected by this method. These averages differ most at frontal electrodes. Fig. 3D shows averages of the same uncontaminated (red traces) and contaminated (blue traces) trials after the independent components accounting for the artifacts were identified and removed, and the summed activities of the remaining components projected back to the scalp electrodes. The two ICA-corrected averages were almost completely coincident, showing that ICA-based artifact removal did not change the neural signals that were not contaminated. Note that the ICA-corrected averages of these two trial groups are remarkably similar to the average of the uncontaminated trials before artifact removal (Fig. 3D, green trace). This implies that the corrected recordings contained only event-related neural activity and were free of artifacts arising from blinks or eye movements.

**Figure 3.** Elimination of eye-movement artifacts from ERP data. (A) The scalp topography of an ICA component accounting for blink artifacts. This component was separated by ICA from 555 target response trials recorded from a normal subject in a visual selective attention experiment. Note: As this scalp map interpolation was based on only very few frontal electrodes, it is not a representative depiction of an eye blink component map. (B) The scalp map of a second component accounting for lateral eye movements. (C) Averages of (N=477) relatively uncontaminated (red) and (N=78) contaminated (blue) single-trial target response epochs from a normal control subject. (D) Averages of ICA-corrected ERPs for the same two trial subgroups overlapped on the average of uncorrected uncontaminated trials (green trace).

Figure 4 shows muscle artifacts separated from 253-channel whole-head EEG data by ICA decomposition of approximately three-quarters of an hour of event-related data first reduced to its first 160 principal components (29 points per matrix weight). As shown in Fig. 4A, the muscle component
activity spectra exhibit wide-band activity between 20 and 50 Hz (the low pass filter cutoff) typical of muscle components. The 3-D scalp maps of these components reverse polarity at or near the insertion points for several muscles identified in standard anatomic atlases. The depicted directions of these muscles also matched the principal orientations of the component scalp maps.

**Figure 4.** Scalp projections of six independent components accounting for activities of individual scalp muscles. ICs ranked 23rd to 69th by variance in the data from 160-dimensional decomposition of 253-channel scalp EEG data by infomax ICA. The component activity spectra exhibit typical broadband muscle activity at 20 Hz and above, here limited by the 50-Hz low pass filter applied to the data before processing. The projections to the scalp electrodes reverse polarity over the muscle insertion, here co-registered approximately on the 3-D subject scalp image from a structural MR image.

**Advantages of ICA-based artifact correction.** ICA-based artifact correction has at least four advantages compared to other artifact removal methods: (1) ICA simultaneously separates EEG signals including artifacts into independent components based on the characteristics of the data, without relying on the availability of one or more ‘clean' reference channels for each type of artifact, as used in regression-based methods. This avoids the problem of mutual contamination between regressing and regressed channels. (2) ICA can be used to remove a wide variety of EEG artifacts. Separate analysis methods are not required for removing different classes of artifacts. Once the data decomposition is complete, artifact-free EEG records in all channels can be derived by simultaneously eliminating the contributions of various identified artifactual components in the EEG record. (3) ICA-based artifact removal may preserve all of the recorded trials, a crucial advantage over rejection-based methods when relatively little data are available, or when blinks and muscle movements occur too frequently, as in some patient groups. (4) Unlike regression methods, ICA-based artifact removal can preserve data at all scalp channels, including frontal and periocular sites.

**Cautions concerning ICA-based artifact removal.** ICA-based artifact removal also has some shortcomings. First, it is important to distinguish between artifacts produced by processes associated with stereotyped scalp maps – for example eye movements, single muscle activity, and single-channel noise. These may be well-accounted by a single independent component, if sufficient data are used in the decomposition. At the other end of the scale, 'paroxysmal' artifacts that produce a long series of non-stereotyped artifact maps into the data – for example, artifacts produced by the subject vigorously
scratching their scalp! – defy the standard ICA model. Here, each artifact time point may be associated with a unique and novel scalp map, posing a severe problem for ICA decomposition. *It is by far preferable to eliminate episodes containing non-stereotyped artifacts from the data before decomposition.*

Since such artifacts can negatively affect the ICA decomposition even at small amplitudes, we normally perform ICA decomposition iteratively. First, we remove any portions of the data with obvious non-stereotyped artifacts. Then we decompose the remaining data. Next, we use the component activities to locate periods of incomplete independence (e.g., common disjunctions in several component activities), and finally decompose the data a second time after removing those episodes. (We are now exploring whether and to what extent this process might be automated).

More ordinary artifacts may also be associated with moving scalp maps and therefore be imperfectly modeled by ICA. For example, the large electrocardiographic (ECG) (or ‘ballistocardiographic’, BCG) artifacts often recorded in a strong static magnetic field, as during or preceding fMRI scanning, have a regular pulsatile pattern beginning in the right neck and spreading over the scalp. (Ordinary ECG signals recorded in the laboratory appear to be more nearly stationary). Artifact activities that move in a stereotyped path, while not directly accounted in the standard ICA model, may be separated by ICA into a small ‘subspace’ of components whose activities together account for the observed artifact. In this case, removal of all of the affected components can effectively remove the artifact.

As well, components accounting for frontotemporal muscle activities may, in our experience, may also exhibit event-related response activity and spectral peaks that suggest they do not account only for electromyographic (EMG) artifact. This might result from competition among frontotemporal EEG and muscle activity sources for independent components, particularly using relatively sparse electrode montages (e.g., 64 channels). Finally, decompositions using insufficient data may return components accounting for mixtures of cortical and artifact data. Thus, although ICA-based artifact correction has been shown to be preferable to other methods for sample data sets (Jung et al. 1998a, 1998b, 2000a, 2000b), it should not be assumed to be ‘perfect’ and its effects on the data should be examined carefully, particularly by researchers not experienced in the use of ICA.

ICA has also been applied to artifact identification and removal from magnetoencephalographic (MEG) recordings, with similar results (Vigario et al., 1997; Barros et al., 2000; Ikeda & Toyama, 2000; Tang et al., 2000; Vigário et al., 2000, etc.).

### 3.2 Event-Related Dynamics of Independent EEG Processes

For the last forty years, the fields of psychophysiology and electrophysiology have been dominated by measures of time-locked averages: scalp-recorded event-related potentials. Typically, psychologists using ERP methods have relied on limited measures of event-related EEG dynamics -- peak amplitudes and latencies in ERP averages at selected scalp channels. The adequacy of time-domain ERP averaging for modeling macroscopic brain dynamics depends on the assumption that the cortical sources of EEG activity contributing to and eliminated from average ERP waveforms are somehow distinct. The scalp topographies of unaveraged EEG and averaged ERP data may, however, be quite similar (Makeig et al., 2002), strongly suggesting that areas contributing to ongoing EEG signals may also contribute to ERP averages.

Often, active brain responses to significant events or external stimuli may involve synchronized oscillations in local field potentials in a number of brain regions (Basar, 1980; Klopp et al., 2000; Makeig et al., 2002, 2004a). Furthermore, EEG processes not contributing to response averages may also be affected by experimental events, and several types of dynamic EEG response processes are not reflected in ERP averages (Pfurtscheller and Aranibar, 1977; Makeig, 1993; Makeig et al., 2002). This increasing tide of evidence supporting the concept of an intimate relationship between features of ERP averages and ongoing EEG processes demonstrates that methods for analyzing unaveraged event-related EEG data need to be developed.
Analysis of trial-to-trial variations in collections of unaveraged single event-related trial epochs may reveal more information about event-related brain dynamics than simple response averaging, but raises three signal processing challenges: (1) Difficulties in identifying and removing artifacts associated with blinks, eye-movements and muscle noise, which are a serious problem for EEG interpretation and analysis; (2) Poor signal-to-noise ratio arising from the fact that non-phase locked EEG activities often are larger than phase-locked response components; (3) Significant trial-to-trial variability in latencies, amplitudes, and even frequency content of both event-related responses and endogenous EEG components (Onton et al., 2005). Although many attempts have been made to analyze single trial data from event-related EEG experiments, no method has been widely adopted.

**ERP-image visualization.** Several years ago, we developed a new method for visualizing trial-by-trial EEG dynamics time-locked to events that we called 'ERP-image' plotting (Jung et al., 1999; Makeig et al., 1999; Jung et al., 2000; Delorme & Makeig, 2004). ERP-image visualization (see below) allows relationships between any trial-dependent variables (e.g., reaction time, alpha component phase, gamma band amplitude, etc.) and time-locked EEG channel or component dynamics to be visualized. Flexible Matlab routines for this and other analysis and visualization methods discussed in this paper are freely available in the open source EEGLAB toolbox (http://sccn.ucsd.edu/eeglab).

Applying ERP-image visualization to the activities of ICA components allows the distinct dynamics of ICA components to be visually recognized and/or characterized, making it possible to optimally combine multidimensional information obtained from an array of scalp electrodes and to detect event-related dynamics in cortical source areas without first averaging over all trials or subjects. In practice, rather than blindly averaging all of the event-related data trials of a given type, the ERP-image approach first sorts the trial epochs in order of some relevant behavioral or physiologic trial variable, and then visualizes, as a color-coded 2-D image, a moving-average image of neighboring trials in the given sorting order. Optionally, the single-trial time course of power at a given frequency or frequency band may be color coded instead of raw potential (Onton et al., 2005). However, as there are many possible relevant trial variables, many different ERP-image plots may be constructed from the same data, no one of which visualizes the multidimensional complexities of the EEG variations.

Combining ICA decomposition with ERP-image visualization can dramatically increase the range, quantity, and quality of cognitive state information that can be gleaned from event-related EEG data over blind averaging of the available single trials. For example, we have demonstrated that brain responses to stimuli and events during cognitive task performance involve perturbations in spectral power and resetting of spectral phase in a variety of EEG processes occurring in many parts of cortex, and that these changes may parsimoniously account for much or all of the activity captured in ERP data sets after event-related signal averaging (Makeig et al., 2002, 2004a).

Figure 5 shows sample results of ERP-image visualization of concatenated event-related EEG epochs from the spatial selective attention task described above (cf. Fig. 3). Figure 5A shows data recorded from a normal adult subject comprising 652 1-sec trials time-locked to onsets of target stimuli (left vertical line). Stimulus event-locked EEG epochs recorded at four midline scalp electrodes: frontal (Fz), vertex (Cz), central parietal (Pz), and occipital (Oz), all referred to right mastoid, are plotted as color-coded horizontal traces (see color bar) sorted in order of subject reaction time latency (thick black line). The ERP average of these trials is plotted below each ERP-image panel. Note the significant inter-trial variability and the strongly overlapping activity recorded at these four widely-spaced scalp sites.
Figure 5. ERP-image plots of correctly responded target response data from a visual selective attention experiment and typical independent components. (A) Single-trial ERPs recorded at a frontal (Fz), a central (Cz), a parietal (Pz) and an occipital (Oz) electrode from a normal subject and time-locked to onsets of visual target stimuli (left thin vertical line) with superimposed subject response times (RT). (B-G) Single-trial activations of sample independent components accounting for (B) eye blink artifacts, (C) non-phase locked, (D) stimulus-locked, (E) response-locked ERP activities, (F) stimulus phase-reset alpha and (G) response-blocked oscillatory mu activities.
ICA decomposition, applied to these concatenated 1-sec EEG epochs at 31 scalp channels, separated at least six categories of independent component processes:

(1) Artifact component processes accounting for EEG artifacts (Fig. 5B) arising from eye blinks or eye movements whose contributions can be removed from the raw EEG records by subtracting the component activity from the data (Jung et al, 1998b, 2000b, 2001a).

(2) EEG or muscle component processes (Fig. 5C) whose activities appear unaffected by experimental events, at least in this simple ERP-image view.

(3) Stimulus-related EEG component processes (Fig. 5D) exhibiting potential fluctuations time- and phase-locked to stimulus onsets with consistent polarity in many to all trials, and thus contributing to the stimulus-locked ERP.

(4) Motor-response related EEG component processes (Fig. 5E) exhibiting potential fluctuations co-varying in latency with the subject motor response. These components contribute activity to the stimulus-locked ERP that is ‘smeared out’ in time; their mean event-related response would be better measured by the ERP average of the same epochs time-locked to the motor response.

(5) EEG component processes (Fig. 5F) exhibiting prominent alpha band (8-12 Hz) activity whose inter-trial phase coherence (ITC) in the alpha frequency band (bottom trace), measuring the mean degree of consistent phase-locking to stimulus onsets, became significant after stimulus presentation (Makeig et al., 2002), even in the absence of any mean alpha band power increase (middle trace).

Such processes may be referred to as exhibiting stimulus-induced ‘phase resetting,’ by analogy with circadian rhythms and model non-linear dynamic systems in which a perturbing event (in the case of circadian rhythms, a light burst) can reset the phase though not the power or frequency of an endogenous oscillation (Tass, 1999). Their signature in the ERP is a train of waves at one frequency (or in some cases more than one) that seem to arise from a ‘flat’ baseline. In actuality, in this case and others, the appearance of rhythmic post-stimulus ERP activity does not imply a mean post-stimulus increase in EEG energy.

While the appearance of rhythmic activity in an ERP does not mathematically rule out the hypothetic possibility that a long fixed-latency ERP wave sequence (as in Fig. 5F) was actually evoked in every trial, though obscured in many trial records by ‘other’ activity (Yeung et al., 2004). However, ICA spatial filtering (as in Fig. 5F) strongly reduces the chances that such fixed-latency activity is obscured by activity from other spatial EEG sources, as may well be the case for scalp channel data (Fig. 5A). In general, parsimonious interpretation of ERP images like Fig. 5F increase doubt that a simple decomposition of event-related data into (a) a fixed-latency ERP plus (b) ‘everything else’ is physiologically well founded. Sorting trials by phase at a given frequency and trial latency (not shown) can be used to clearly visualize the phase distribution of a process. In most instances, such phase-sorted ERP images of independent component activities exhibit partial phase locking to stimulus or other events (Makeig et al., 2004b).

(6) EEG component processes (Fig. 5G) exhibiting ‘nu-rhythm’ peaks in their single-trial mean power spectra, typically near 10 Hz and 20 Hz, that were blocked (middle trace) following the subject button press, but exhibited no significant inter-trial phase coherence (bottom trace) (Makeig et al, 2001). Note, in this example, the two negative (blue) potential dips separated by about 200 ms that follow the subject motor response. These constitute a complex transient theta band event explored in more depth elsewhere (Makeig et al., 2004b).
The process taxonomy summarized in Fig. 5 could not have been obtained from signal averaging or conventional frequency-domain filtering approaches. Because the scalp maps associated with these processes were clearly not orthogonal to each other, this separation of the data into temporally near-independent and functionally distinct processes could not be produced by PCA, although other non-orthogonal decomposition methods applied to principal components of the single-trial data might separate some of the independent processes to some extent.

Better understanding of trial-to-trial differences in macroscopic brain dynamics recorded as EEG may allow a better understanding of the variability of normal human performance and brain function typical in both repetitive and non-repetitive tasks. Often, this variability has been interpreted as evidence for ‘noise’ in EEG signals that is best removed by response averaging. However, there is no evidence that behavior, organized by brain activity, is based solely on mean brain dynamics, as appropriate behavioral responses to each event and circumstance need to be made promptly, not ‘on average.’ An alternate view suggested by the trial-to-trial variations in EEG dynamics visualized in ERP-image plots is that these variations reflect brain responses both to stimulus events per se and to the differing neurobehavioral challenges posed by each event in light of the preceding sensory, behavioral, and associational context (Onton et al., 2005).

Consistency of independent component dynamics across subjects. More detailed study of differences in EEG dynamics of normal, brain-damaged, diseased, aged, or genetically abnormal individuals is of interest in both clinical and basic research. ICA-based analysis allows identification and segregation of spatially-overlapping event-related EEG activities exhibiting variety of distinct relationships to task events. ERP-image plotting of independent component processes allows assessment of interactions between diverse features of ERP records and ongoing EEG processes. As we have seen, contrary to the common supposition that ERP time series are necessarily brief, stereotyped, latency- and polarity or phase-consistent responses evoked by some events, independent of ongoing background EEG activity, many ERP features may be generated in part or large part by dynamic changes in ongoing EEG processes induced by events (Makeig et al., 2002, 2004a). ICA thus may help researchers to take fuller advantage of what until now has been an only partially-realized strength of event-related paradigms - the ability to examine systematic relationships between single trials within and between subjects (Jung et al., 2001a, 2001b; Makeig et al., 2002, 2004a). However, identifying component similarities between subjects, a key first step before attempting to characterize subject or group component differences, is not a simple process.

To study the cross-subject stability of independent components derived by ICA applied to single-trial epochs, we applied ICA decomposition to data from each of 15 normal subjects in the visual selective attention experiment described above, to parse the dynamics of the unauged 31-channel EEG signals into maximally independent processes. We then clustered the resulting processes across subjects by similarities in their scalp maps and activity power spectra, using a symmetrical Mahalanobis distance measure (see Jung et al., 2001a and Makeig et al., 2004a for details). Component clustering, applied to the 31×15 = 465 component processes, identified at least 13 component clusters having similar power spectra and scalp maps. Cluster membership was in a few cases further adjusted by hand for uniformity. The component clusters showed functionally distinct activity patterns. Four distinct component clusters accounted for eye blinks, horizontal eye movements and left and right temporal muscle noise respectively. These activities were effectively separated from the activity of the other component clusters by the ICA decomposition (and are not discussed further here).

Mean dynamics properties of the remaining nine non-artifact component clusters are summarized in Figure 6, each of which shows the mean scalp map and response-locked ERP image of epochs time-locked to subject responses instead of to stimulus presentations, as well as the contributions (red shaded areas) of the component cluster to the cluster-subject grand-mean ERP (here visualized via its envelope, showing the most positive and negative channel values at each latency), and the mean cluster event-
related spectral perturbation (ERSP) (Makeig, 1993). ERSP images visualize changes (in dB) from baseline in spectral power across a broad frequency range (here, 3-50 Hz).

Figure 6. Nine non-artifact clusters of independent components from 15 subjects performing the visual selective attention task. (A) Far-frontal component cluster accounting for the pre-response (P3f) positivity. (B) Broad parietal component cluster accounting for part of the post-response (P3b) positivity. (C) Central midline component cluster projecting maximally to the vertex. (D) Frontal midline cluster of components often exhibiting a theta-band peak in their activity spectra. (E) Left mu rhythm component cluster with mu characteristic 10-Hz and 22-Hz peaks in the activity spectrum. Following the button press, this activity is blocked. (F) Right mu rhythm component cluster with mu characteristic 10-Hz and 22-Hz peaks in the activity spectrum. Following the button press, this activity is blocked. (G) Left posterior alpha component cluster. (H) Central posterior alpha component cluster with characteristic trapezoidal scalp projection, consistent with a bilateral, peri-calcarine equivalent dipole source model, and contributing prolonged phase resetting following stimulus onset (curving dashed trace). (I) Right posterior alpha component cluster. Each set of four panels shows: (Top left) The mean component scalp map. (Top right) The whole-data (black traces) and cluster-accounted (red filled) ERP envelopes (min and max voltage channel values at each time point). (Lower left) ERP-image plot of the color-coded single trials time locked to the response (solid vertical line) and sorted by subject response time from stimulus onset (dashed line). Trials normalized by dividing by the standard deviation of component activity in the 1-s pre-stimulus baseline. (Lower right) The component mean ERSP showing mean event-related changes in (log) spectral power across data trials time locked to the response (solid line). Here, median stimulus delivery time is indicated by the dashed line. [Adapted from Makeig et al., 2004a].
Identified component clusters. Figure 6A plots an inferior frontal (P3f) cluster of 10 independent component processes (from 10 of the 15 subjects) that together largely accounted for an ERP feature whose time course was similar to the peak labeled ‘P3f’ (for P3-frontal) found earlier in ICA decomposition of 25 grand-mean condition ERPs from this experiment and illustrated in Fig. 1 (Makeig et al., 1999a). This component cluster accounted for nearly all the early (P3f) positivity occurring before the subject button press, particularly in shorter response-latency trials (Fig. 6A, lower left), also visualized as the reddened pre-stimulus contribution of the cluster processes to the response-locked grand-mean ERP envelope (upper right).

The P3f cluster-mean response-locked positivity began as early as 150 ms after stimulus onset, consistent with direct neurophysiological evidence that by 150 ms after stimulus onset, visual information is spread throughout the brain by a complex web of afferent and efferent connections (Klopp et al., 2000; Hupe et al., 2001). Subtracting the button travel time (~25 ms, roughly estimated from EMG recording during one experimental session) and the neuromuscular conduction time (~15 ms) suggested the P3f peak, centered at 39 ms before the button press, occurred very near the moment of the subcortical motor command (Makeig et al., 2004a). It is thus tempting to speculate that the P3f process should originate in frontal structures involved in motivated decision making and response selection, such as orbitofrontal cortex (Ikeda et al., 1996), though the sparse spatial sampling of the present data did not allow more specific conclusions. The cluster-mean ERSP (lower right) shows the positive P3f peak and the succeeding post-response negativity reflect a weak transient increase in low theta band activity (near 4 Hz) time locked to the button press.

Fig. 6B shows the mean scalp map and activity patterns of a bilaterally distributed cluster of 15 component processes (from 9 of the subjects) that projected most strongly to posterior and central scalp sites and made substantial contributions to the slow post-motor ‘P300’ or ‘P3b’ positivity in the target stimulus response. The ERP image of the normalized single-trial component activity (lower left) includes an early series of small, positive and negative, wave fronts following the (sorted) time of stimulus delivery (dashed curve) by fixed delays. These are followed by a large response-locked positivity (red area) accounting for 62% of post-response ERP variance at 300 ms for site Pz. The P3b cluster positivity is clearly smaller in late-response trials. The mean cluster ERSP (Fig. 6H) reveals a significant (3-dB) low-theta power increase centered just after the button press.

The dynamics of a frontal midline (FM) component process cluster is shown in Fig. 6C. In the RT-sorted response-locked ERP image (lower left), the two negative wave fronts follow the curving dashed line marking stimulus onsets, the second of these merging with the first post-motor response negativity (left arrow). The two post-motor negative troughs (two arrows above the ERP image, also the two lumps in the cluster contribution to the response-locked ERP, upper right) also constitute a low theta-band event (lower right). Though the vertex-centered central midline (CM) component cluster (Fig. 6D) also exhibited the post-motor theta feature (red arrows) with the mid frontal and mu rhythm clusters (see below), it contributed little to the broader (P3b) positivity.

Two left and right mu-rhythm component clusters shown in Fig. 6E-F exhibited the characteristic features of mu-rhythm processes – distinct spectral peaks near 10 Hz and, here, 22 Hz that are strongly blocked following movements. Equivalent dipoles fit to the scalp maps of these components were located roughly in hand somatomotor cortex, and oriented roughly orthogonal to the usual directions of the central sulci. Both the ERP and ERSP peaks were larger in the left mu cluster (contralateral to the subject response hand) than in the right. In common with the FM and CM clusters (Fig. 6C-D), the mu component clusters also exhibited a two-cycle post-motor theta pattern, concurrent with a mean low theta power increase. They also made slower, positive-going contributions to the parietal ERP, particularly to the late ‘slow wave’ phase of the stimulus-locked P300 complex that, unlike the main (P3b) peak, exhibits a polarity reversal over the central scalp (Simpson et al., 1977).
The dynamics of three component clusters shown in Fig. 6G-I project primarily to the posterior scalp, though cluster G and I also have weak projections of opposite polarity to the frontal scalp. Each had a distinct near 10-Hz alpha frequency peak in their activity spectrum, most pronounced in components of the central alpha cluster (Fig. 6H). The contributions of the two lateral occipital alpha clusters (Fig. 6G, 6I) to the stimulus-locked ERP (not shown here), appearing as sloping wave fronts in Fig. 6G, I included an early stimulus-locked peak accounting for most of the early (before 150 ms) P1 ERP peak, and for a portion of the succeeding N1 peak (peaking at 150-200 ms in different scalp channels).

In the central alpha component cluster (Fig. 6H), the initial stimulus-locked response feature was followed by a train of phase-aligned circa 10-Hz stimulus-locked waves most likely produced by partial phase resetting of the intermittent alpha activity of these components following stimulus onsets, since they were accompanied by no mean increase in alpha power. The central alpha cluster also made an appreciable slow triangular wave contribution to the P300 positivity (lower left trace and upper right), while the contributions of the lateral alpha clusters (Figs. 6I, J) to the response-locked ERP were smaller.

Together, the nine component clusters shown in Fig. 6 accounted for 91.1% of the variance of the response-locked grand mean ERP for all 15 subjects, as well as 90.8% of the variance of the stimulus-locked grand mean ERP (Makeig et al., 2004a). ICA cleanly separated the EEG contributions of these processes to the ERP, allowing exploration of their individual and joint event-related dynamics across the single trials. The clear separation of ‘alpha-ringing’ of the central alpha cluster in the stimulus-locked response from the other alpha activities accounted for by the left and right mu and alpha clusters (Fig. 6E-I), each having highly overlapping alpha-band peaks in their power spectra, illustrates the ability of ICA to separate temporally and functionally distinct activities generated in different brain areas that project to the same (whole scalp) range of scalp channels. This example also reveals a variety of event-related dynamics (spectral power perturbations, phase modulation or resetting, etc.) may occur consistently across sets of trials with distinct reactives, facts that would be nearly impossible to observe by blind response averaging of the single scalp channel data (Fig. 5A).

### 3.3 Localization of Independent EEG Sources

As mentioned above, EEG data collected from any point on the scalp typically includes activity projected by volume conduction from multiple EEG processes in different cortical regions. This has made it difficult to localize the sources of the EEG signals. In general, the problem of determining the distribution of brain electrical sources from electromagnetic field patterns recorded on the scalp surface is mathematically underdetermined. It is thus necessary to employ model-based procedures that impose constraints on the solution that take into account auxiliary sources of information. ICA offers an important benefit for source localization by separating the question of source identification from that of source localization. From this point of view, ICA acts as efficient spatial filtering for source activities that contribute distinct temporal information to the data, using the temporal information contained in the unaveraged data to constrain and identify the scalp-map projections of these source activities.

**Physiologic constraints on ICA sources.** Independent EEG sources must each be produced by partially synchronous field activity within some spatial cortical domain. The high density of local excitatory and inhibitory interconnections in cortex support synchronous oscillations at EEG frequencies, as well as the predominantly radial thalamocortical coupling, should support independent component generation within compact local domains of unknown extent and distribution. By separating the data into maximally independent domains of partial synchrony based on the temporal information contents of the entire training data, ICA identifies scalp maps associated with strong, synchronous field activity in compact domains (Makeig et al., 2002, 2004a). If the source geometry is not too complex, EEG source localization methods based on single equivalent dipoles using realistic constraints derived from single-subject MRI images should produce useful inverse models.
Without ICA spatial filtering of source activities, researchers must attempt to find time points in the raw or averaged data in which one source of interest may dominate. In the raw data, there is no established metric indicating time points at which all EEG sources but one are by chance inactive or at zero crossings. Most often, researchers hope that averaging data epochs time-locked to a class of similar events will cancel out EEG sources other than a small set of successively activated sources that successively produce separable peaks in the average ERP waveform. However, much evidence (as in Figs. 5-6) shows that this hope is unfounded (Makeig et al., 2002). In general, response averaging fails to isolate the scalp distribution of single EEG sources (with the exception of very early activations in primary sensory areas), whereas given adequate numbers of time points and data channels, ICA can do so (Zhukov et al, 2000).

Figure 7. Scalp maps, ERP-image and ERP plots, mean activity spectra, and equivalent dipole locations of two maximally independent EEG components (ICs 9 and 11, ordered by decreasing EEG variance accounted for) from the same 253-channel experimental session as in Fig. 4. Vertical black lines in the ERP-image plots: Onsets of visual letter stimuli. Although the two components have highly similar spectra and ERPs, and nearly symmetrically located equivalent dipole locations, they were determined by ICA to be temporally near independent. Note the apparent phase resetting of these active alpha band processes in the ERP-image panels, and the resulting ERP features. The residual variances of the single dipole models (shown in blue and green on a mean MR image) across all the electrodes, was near quite low (near 2%), consistent with their possible origin in partially synchronized local field activity across a single patch of cortex.

Figure 7 shows examples of equivalent dipole source modeling of maximally independent EEG component processes identified by infomax ICA from the same 253-channel data set that produced the muscle-related components shown in Fig. 4. The locations and orientations of single equivalent dipole models of two ICA components accounting for stimulus event-related activities resembling alpha phase...
resetting were separately estimated by least-squares equivalent dipole modeling (Scherg & von Cramon, 1986; Oostendorp & Oosterom1991) and plotted on a mean MR image using EEGLAB (Delorme & Makeig, 2004). Although the activity spectra and target event-locked ERPs of these two processes were highly similar, and their equivalent dipole locations near bilaterally symmetric, ICA identified the two processes being temporally independent both of one another and also of all other brain and non-brain activity sources in the data. The scalp topographies of both components were well fit by a single equivalent dipole with a residual scalp map variance of under 3% across all 253 scalp channels, suggesting (though not proving) that these activities were generated within compact patches of occipital cortex.

4. ICA APPLICATIONS: PAST, PRESENT AND FUTURE

Early reports on applying ICA to EEG data showed that ICA can be used to segregate clear artifactual EEG components (line and muscle noise, eye movements) from other sources (Makeig et al., 1996, 1997; Jung et al., 1998a, 1998b, 2000a; 2000b). Subsequent studies showed that ICA may have several advantages compared to other methods of artifact removal (Jung et al., 1998a; Joyce et al., 2004). EEG researchers also began exploring the application of ICA for separating maximally independent event-related information from sets of averaged ERPs (Makeig et al., 1997, 1999a, 1999b; Olbrich et al., 2005) and from clinical EEG data (McKeown et al., 1998b; Kobayashi et al., 1999; 2001; Pritchard et al., 1999; Nam et al., 2002; Melissant et al., 2005).

In recent studies, the co-authors and others have begun to apply ICA to the analysis of collections of unaveraged EEG epochs in combination with time/frequency modeling, to explore the complex event-related brain dynamics associated with cognitive task performance (Jung et al., 1999, 2001, 2001b; Johnson et al., 2001; Makeig et al., 2002, 2004a; Delorme et al., 2002, 2003; Contreras-Vidal & Kerick, 2004; Debener et al., 2005; Onton et al., 2005). Rather than focusing on peaks evoked in averaged scalp-channel ERPs, or on changes induced in the scalp channel EEG power spectrum by experimental events, these studies propose that sensory and cognitive event-related EEG dynamics are better modeled as time/frequency perturbations of the activities of underlying source processes (Makeig et al., 2004a).

ICA, in this model, minimizes the both the correlations and the higher-order redundancies produced by volume conduction of source activities to the same scalp electrodes, and identifies temporally independent activities of individual cortical sources based on the unique temporal information they contribute to the recorded signals at the single trial or continuous data level. These extracted source dynamics often appear to be more physiologically simpler, and linked to behavior both more tightly and more distinctly than measures of the single scalp channel data (Makeig et al., 2004b). The temporal and functional distinctions uncovered by ICA make trial-by-trial analysis more promising, since single trials may be characterized by measures of the activities of multiple or many independent EEG processes.

Certainly, more research will be required to fully assess the strengths, weakness and implications of the combined ICA spatial filtering and time/frequency analysis approach. However, because of the evident promise and new insights we and others have gained into macroscopic brain dynamics through first applications, we believe this approach to analysis of multidimensional EEG (and also MEG) data will continue to grow and may even become a major avenue of research in event-related brain dynamics.

Future advances. Independent component analysis and blind source separation in general continue to be active research fields in engineering and applied mathematics. One direction this research is taking is to learn whether and how the assumptions used in standard ICA can be further relaxed or customized to particular types of signals. The studies reviewed here assumed the EEG signals to be linear and instantaneous mixtures of temporally independent and spatially fixed EEG source activities. However, some recent studies have made use of more complex ICA algorithms with slightly different assumptions. Dyrholm et al. (2004b), for example, showed how convolutive ICA can capture more complex...
spatiotemporal source behavior in a single component than is possible with conventional ICA. Possibly, this method could model and measure delayed interactions between some EEG source processes identified by standard ICA as independent in the instantaneous sense. Anemüller et al. (2003) demonstrated the applicability of complex spectral-domain ICA to EEG modeling and analysis, a method also allowing for more spatiotemporal flexibility in the modeled source processes. Other so-called overcomplete ICA or ‘sparse decomposition’ methods may be able to identify more sources than channels in the data (Lewicki and Sejnowski, 2000), though these methods may be less robust to other distributional assumptions and are more computationally demanding.

The near future should see a continual increase in the amount of information contained in high-dimensional EEG and MEG data, and made available for modeling of brain processes by these and other blind source separation methods. Meanwhile, a current impediment to adoption of ICA methods on the part of many EEG researchers is the unavailability of statistical tools for identifying component clusters and for performing condition and group comparisons of component cluster activities, although some relevant tools should soon be made available, both by open source software projects and by commercial EEG manufacturers.

5. APPLYING ICA TO EEG DATA: BALANCING CAUTION WITH ENTHUSIASM

This review has examined the assumptions underlying the application of ICA algorithms to EEG signals, and has reviewed some practical considerations involved in using ICA to decompose sets of averaged ERPs, continuous EEG records, and/or sets of event-related EEG data trials. For example data from a spatial attention experiment, we showed how ICA, applied to the concatenated single trials of each subject after removing trials containing out-of-bounds or non-stereotyped artifacts, decomposed the whole training EEG data into 31 spatially-fixed, temporally maximally independent component processes. As we recounted, the scalp maps associated with many of these processes resembled the scalp projections of synchronous activity in either one or sometimes two near-bilaterally symmetric cortical patches. This allowed several further steps in the analysis:

1) Removal of independent components accounting for non-neural artifacts (eye movements, blinks, muscle noise, electrocardiogram, etc.) to ‘clean up’ single-trial data for subsequent single-trial processing, without sacrificing e.g. frontotemporal brain dynamic information.

2) Identification and segregation of stimulus- and response-locked event-related activity in collections of single event-locked data trials.

3) Investigation of the interplay between ERP features and ongoing EEG processes.

4) Separation of spatially-overlapping EEG activities each projecting to the entire scalp and active frequency range into source activities exhibiting a variety of distinct reactivities to task events.

5) Time/frequency analysis of the activity time courses of independent components in single trials to measure and compare oscillatory dynamics of independent brain EEG sources within and between subjects and experimental conditions.

6) Investigation of event-related coherence (ERCOH) between independent EEG components to reveal transient event-related switching between states of independence and partial coupling of otherwise independent EEG sources.

7) Biophysical source modeling of temporally and functionally distinct EEG processes, many of whose scalp distributions are compatible with activity projecting from a single cortical patch.

We also demonstrated the utility of component-matching clustering analysis to semi-automatically group components with similar scalp maps and power spectra from 15 normal subjects (Jung et al, 2001a;
Makeig et al., 2002, 2004a). The resultant nine independent component clusters resembled classes of EEG phenomena long described by neurologists from observation of paper data displays – central and lateral alpha, left and right mu, and frontal-midline theta rhythms. By cleanly separating the EEG contributions of these processes, ICA allowed trial-by-trial exploration of their individual and joint event-related dynamics.

Limitations of ICA. Although ICA appears to be generally useful for EEG and MEG analysis, it is a so-called ‘blind’ decomposition method with some inherent limitations, and should therefore be used with care and some degree of caution.

First, temporal independence, used by ICA to perform the component separation, may not be sufficiently expressed when the training data set is too small, when the number of data channels is insufficient, or when topographically distinguishable phenomena always co-occur in the data, as may be the case in applying ICA to averaged ERP data. One might say that the ICA training data must express the independence of its different brain and non-brain source processes, for ICA to be able to extract their maximally independent brain signals. However, there is no separate measure of whether the training data are sufficient in this regard, and thus no guarantee that ICA is able to correctly separate the complete set of maximally independent sources contributing to the data.

Second, the robust ICA methods reviewed here can decompose at most \( N \) sources from data collected at \( N \) scalp electrodes. Usually, the effective number of statistically-independent signals contributing to the scalp EEG is unknown, and it is likely that observed brain activity arises from more physically separable effective sources than the available number of EEG electrodes. Although infomax ICA decomposition, in particular, appears relatively robust to additional small sources in the data (Makeig et al., 1996c), in general it is not known if or when the spatial distribution of EEG sources, as well as the character and relative strength of their activities, differ from moment to moment Applying ICA to some subset of the training data may alleviate this concern with possible source non-stationarity, but may also raise questions of accuracy of the decomposition, if the amount of training data is limited. Further, decomposing isolated parts of the data raises difficult questions of the ways in which the resulting decompositions are essentially the same or different. Another difficult question is whether the distributions of spatial sources generating EEG activity in high and low frequency bands, e.g. in the gamma (30 Hz and above) and infraslow (< 1 Hz) bands, may differ from the source distribution of activity in intermediate (2-30 Hz) bands.

In sum, although results of applying ICA to EEG and ERP data have already shown great promise and given new insights into event-related brain dynamics, the analysis method is still in its infancy. The plausibility and reliability of its results should in each case be validated, therefore, using convergent evidence, typically behavioral and/or other physiological measurements, before interpreting its functional significance.

Conclusion. The first results summarized in this review suggest that a wealth of information on active links between cognitive processes and event-related brain macrodynamics is available in non-invasive high-density EEG (and/or MEG) data. We believe that availability of more comprehensive analysis techniques including ICA and ERP-image visualization should make EEG and MEG imaging of increasing interest and relevance to both basic neuroscientists and clinical neurophysiologists. We foresee a day in which analysis of event-related EEG/MEG dynamics may routinely complement and even partially supplant imaging of slow-changing hemodynamics, while complementing and expanding the amount of information about distributed brain dynamics available from invasive and microelectrode recordings.
APPENDIX : ICA Cost Functions

Objective or cost functions that can be used in ICA algorithms include information maximization (Nadal and Parga, 1994; Bell and Sejnowski, 1996), maximum likelihood (Pearlmutter and Parra, 1996; Pham, 1996), and other adaptive filtering techniques that can approximate a solution for the source unmixing matrix W. The common factor between these different cost functions are that they are all related to mutual information (Lee et al., 2000; Hyvarinen et al., 2001). Mutual information measures the difference between the joint probability of the estimated source signals and its factorial version (Cover and Thomas, 1991). The objective of the Infomax ICA algorithm is to minimize redundancy between the outputs. This is a generalization of the mutual information:

\[ I(u) = \int p(u) \log \frac{p(u)}{\prod_{i=1}^{N} p_i(u_i)} du \]  

(5)

The redundancy measure has value 0 when the p.d.f. p(u) factorizes, as in (4). Mutual information, however, is a difficult function to minimize directly. The insight that led to the infomax ICA algorithm is that I(u) is related to the joint entropy, H(g(u)), of the output signals, u, passed through a set of sigmoidal non-linear functions, g_i:

\[ I(u) = -H(g(u)) + E \left[ \sum_i \log \frac{|g_i'(u_i)|}{p_i(u_i)} \right] \]  

(6)

Thus, if the absolute values of the slopes of the sigmoid functions, |g_i'(u_i)| are the same as the independent component p.d.f.’s, p_i(u_i), then Infomax, maximizing the joint entropy of the g(u) vector, will be the same as ICA (minimizing the redundancy in the u vector).

![Figure A1](https://example.com/figure.png)

**Figure A1.** Optimal information flow in sigmoidal neurons (left) Input x having probability density function p(x), in this case a gaussian, is passed through a non-linear function g(x). The information in the resulting density, p(x) depends on matching the mean and variance of x to the threshold, w_0, and slope, w, of g(x). (right) p(y) is plotted for different values of the weight w. The optimal weight, w_opt transmits most information (from Bell & Sejnowski, 1996 by permission).
The principle of ‘matching’ the $g'_i$’s to the $p_i$’s is illustrated in Figure A1, where a single Infomax unit attempts to match an input Gaussian distribution to a logistic sigmoid unit, for which:

$$g(u) = \frac{1}{1 + e^{-u}}$$  \hspace{1cm} (7)

The match cannot be perfect, but it does approach the maximum entropy p.d.f. for the unit distribution by maximizing the expected log slope, $E[\log|g'(x)|]$.

The generalization of this idea to $N$ dimensions leads to maximizing the expected log determinant of the absolute value of the Jacobian matrix $|\partial g_i(u)/\partial x_j|$. This optimization attempts to map the input vectors uniformly into the unit $N$-cube (assuming that the $g$-functions are still 0-1 bounded). Bell and Sejnowski (1996) showed that the stochastic gradient descent algorithm that maximizes $H(g(u))$ is:

$$\Delta W \approx W^{-T} + f(u)x^\top$$  \hspace{1cm} (8)

where $-T$ denotes inverse transpose, and the vector-function, $f$, has elements:

$$f_i(u_i) = \frac{\partial}{\partial u_i} \ln g'_i(u_i)$$  \hspace{1cm} (9)

When $g'_i(u_i) = p_i(u_i)$ for all $i$ then, according to (6), the ICA algorithm is exact. Unfortunately, this leaves a difficulty. Either one has to estimate the functions $g$ during training, or one needs to assume that the final term in (6) does not interfere with Infomax performing ICA. We have empirically observed a systematic robustness to mis-estimation of the prior, $\hat{p}_i(u_i | W, g) = |g'_i(u_i)|$, an observation that has been supported theoretically by Amari (1998).

There are many optimization methods that can be deployed to estimate $W$. Gradient ascent or descent techniques are common in information maximization or maximum likelihood approaches (Bell and Sejnowski, 1995; Lee et al., 1999) and fixed point optimization, (Hyvarinen and Oja, 1997), is also popular, due to their simple implementations. No comparison has yet been published of differences in results of different ICA algorithms applied to the same EEG data. The difficulty has been in selecting a figure of merit, since no gold-standard source decomposition exists. In simulations of data built to exactly match ICA assumptions, many ICA algorithms perform equally well. Any differences in results of actual data decomposition, therefore, are likely steered by (unknown) deviations of EEG generation from the idealized ICA model.

Acknowledgments

The authors wish to thank their many collaborators in the work reviewed here, foremost among them Terry Sejnowski, Tony Bell and Te-Won Lee at Salk Institute, La Jolla (1995-2001), and Marissa Westerfield, Arnaud Delorme, and Julie Onton at UCSD (2002-). The authors gratefully acknowledge the financial support of the Swartz Foundation (Old Field, NY) and of the National Institutes of Health USA (RO1-NS047293-01A1).
REFERENCES


