

Information-based modeling of event-related brain dynamics

Julie Onton
Scott Makeig

Swartz Center for Computational Neuroscience
University of California San Diego
La Jolla CA 92093-0961
{julie,scott}@sccn.ucsd.edu

For a volume in the Elsevier *Progress in Brain Research* book series, Eds. Neuper & Klimesch

February, 2006

1. Introduction

As human neuroimaging techniques have matured, the experimental questions about human brain dynamics that can be posed and answered have become more refined. Recent research in hemodynamic imaging has produced software that allows straightforward comparison of metabolic and in particular blood-oxygen dependent activities in equivalent brain areas of subjects whose brains may have different sizes and shapes. On the other hand, most current electroencephalographic (EEG) research studies still separately analyze electrical potential differences recorded between single electrodes on the scalp surface and a distant scalp reference electrode – both located relatively far from the brain itself. By basic biophysics, if temporally coherent activity in any brain area creates far-field potentials on the scalp, these potentials must be distributed widely, by passive volume conduction, across the scalp surface. Recorded EEG data sum the activities of all such EEG source areas, as well as electrical artifacts from muscles, eyes, electrodes, movements, and the electrical environment.

An additional complexity for electromagnetic brain imaging, as compared to metabolic imaging, is that electromagnetic field activity within a source area that contributes to scalp potentials has a net source orientation as well as location. Thus, typical differences in the number and configuration of cortical sulci between individual brains may lead to large difference in the *orientations* of spatially equivalent source areas in different brains, and thus in the scalp potential distributions produced by activity in functionally equivalent brain locations.

For this reason, comparing EEG activities as recorded at equivalent scalp locations from different subjects cannot achieve comparable accuracy to comparing hemodynamic activities of equivalent 3-D locations in their magnetic resonance (MR) images. The mixture of scalp source activities reaching a given scalp location, for example the vertex (Cz), in different subjects may depend on the relative amplitudes, distances, and orientations of subject cortical source areas across most of the cortex. Therefore, a first signal processing step in using scalp EEG (or its magnetic equivalent, magnetoencephalographic or MEG) data for dynamic brain imaging should be to *spatially* filter the recorded data so that the outputs of each spatial filter may be identified with activity in a particular cortical source area or domain.

Unfortunately, the widespread projection of source activities across the scalp surface also means that the ‘EEG inverse problem’ of locating the brain sources of the recorded data is mathematically ill-posed and is in fact not resolvable without additional constraints and assumptions. The EEG inverse problem is even more difficult than the corresponding inverse problem for magnetoencephalographic (MEG) data, since propagation of electrical potential through the brain by volume conduction is in part anisotropic (meaning it varies with the direction of propagation). However, direct non-invasive measures of this anisotropy are not currently available. For these reasons, as well as the considerable computational complexity involved, adequate brain location-based spatial filtering of EEG signals has not been thought possible, leading to the widespread but largely inaccurate perception that EEG brain imaging is doomed to low spatial resolution, allowing only poor separation and localization of a small number of source-level signals.

In recent years, an alternative approach has been developed for generating spatial filters that allow separate monitoring of field activities in different cortical areas, using a recent signal processing approach known as independent component analysis (ICA) . Eleven years ago, the senior author and colleagues first discovered that ICA is a useful tool for decomposing EEG signals into maximally independent activity patterns that now in many cases known to be compatible with activity in a single active cortical area . Thus, the ICA approach to dynamic brain imaging is, first, to

separate the independent EEG activities in each subject's data, not by direct spatial filtering for activities generated in a set of pre-defined cortical locations, but by using the information content of the data itself to separate portions of the recorded scalp data from each active cortical and artifact source area, based on the deceptively simple but statistically and physiologically plausible assumption that over time, these activities should be nearly independent of each other.

The major advantage of this approach is that the locally coherent activity constituting a single EEG *source* will be grouped together into a single independent component (IC) that includes its projections to *all* the scalp channels, while the activities of unrelated EEG sources will be rejected from this IC and isolated into other ICs. In this way, under favorable circumstances ICA will transform the recorded high-density scalp data into a set of cortical and artifact source recordings – thereby discovering *what* distinct signals are contained in the data before asking directly *where* these signals are generated.

This indirect or even (to some) backwards-seeming approach to spatial source filtering has made the ICA approach difficult to accept for some researchers with physical science backgrounds, while other researchers accustomed to using seemingly simpler measures of activity at single scalp channels might balk at the apparent increase in complexity of ICA-based analysis. Although skepticism among (at least) these two classes of researchers may have slowed widespread investigation of the utility of ICA for EEG analysis, as more and more students and investigators take advantage of both freely available and commercial software tools for performing the analysis, its benefits as well as its pitfalls are becoming more widely recognized. Here, we first review the promise, theory, and practical use of ICA decomposition of EEG data, then exemplify its benefits for event-related time/frequency analysis, and finally describe a problem ICA poses, of identifying equivalent components across subjects and sessions, and discuss our current approach to its solution through ICA component clustering..

2. ICA history

The concept of ICA was first developed in the field of signal processing around 1990 as part of a larger class of ‘blind source separation’ problems that aim to separate individual source signals from multi-dimensional data in which they are mixed. ICA applied to high-density scalp EEG data produces a strictly linear and invertible decomposition of the data, meaning that the activity of every resulting independent component (IC) is simply a weighted sum of the signals recorded at all of the input scalp channels, and every scalp channel signal is simply a weighted sum of the projected activities of all the independent components (ICs). Technically, ICA finds a set of *fixed spatial filters* that together constitute the most distinct (i.e., temporally near-independent) signals available in the input data.

Several related algorithms for performing ICA have been developed for which Matlab (The Mathworks, Inc.) and other implementations are readily available on the Internet. These include JADE, , infomax ICA , and so-called FastICA , as well as variants of second-order blind identification (SOBI) , that also factor in relationships between multiple time points using autoregressive models. We have found that infomax ICA, in particular, gives reliable results for data of sufficient quantity and quality having almost any number of channels. In its original formulation, infomax ICA could only find sources that have super-gaussian activity distributions, meaning roughly source processes that are only intermittently active. Over sufficient time, this well describes most EEG phenomena. However, ‘extended’ mode infomax, introduced by Lee et al. , can also learn filters for sources such as line noise that have sub-gaussian activity distributions (roughly speaking,

activities that are mostly ‘on’). This may be important when EEG data are obscured by 50-Hz or 60-Hz line noise from environmental AC power sources. Matlab and binary implementations of infomax ICA, in particular, as well as many other tools for analyzing EEG data with this or other ICA algorithms, are freely available in the open source EEGLAB analysis programming environment (<http://sccn.ucsd.edu/eeglab>).

In our hands, infomax ICA produces useful results from decomposition of EEG data sets with 31 to 256 channels. Decomposing data with fewer channels is also possible and should be useful for some purposes. Following, we present several examples showing how ICA may clarify the dynamics of the underlying cortical sources of observed EEG phenomena by comparing results of ICA decomposition with traditional analysis of the same scalp channel data. Finally, we will briefly address the issue of comparing independent component activities across one or more groups of subjects and/or sessions.

3. ICA theory

ICA model assumptions. A general and physiologically plausible assumption underlying most EEG analysis is that most of the far-field potentials detected at the scalp are generated not in the scalp itself, but within spatial domains (or most simply, patches) of similarly oriented cortical pyramidal neurons. EEG recordings are time series of measured potential differences between two scalp electrodes, usually thought of as potential differences between an ‘active’ electrode and a second (‘passive’) reference electrode – though in fact both electrodes are equally receptive to nearly all cortical and artifact source signals.

A given cortical patch can only produce a far-field potential and thus become an EEG source domain, however, if the local fields surrounding its pyramidal cells become partially (*not* necessarily completely) synchronized. There are several biophysical properties of the brain that encourage such synchronization. These include inherent rhythmic proclivities of pyramidal and other cortical cells, the high-speed, non-synaptic, electrotonic gap junction connections between (mainly) non-pyramidal inhibitory cells, bidirectional coupling between inhibitory and excitatory (including pyramidal) cells, and between cortex and thalamus. However, both the strength and frequency content of the local synchronization is highly variable and may not normally have sufficient coherency and/or may not extend over a sufficient area to produce appreciable far-field scalp potentials. Many biophysical systems and properties modulate the emergence and time course of local field synchronization. As a result of all this physiological complexity, EEG signals themselves have high spatiotemporal complexity. The simple biophysical fact remains, however, that potentials recorded between *any* electrode pair will sum activities from nearly all the active cortical EEG source domains as well as from nearly all the active artifact signal sources.

Temporally, EEG signals are comparatively wide band ($\ll 1$ Hz to $\gg 50$ Hz, more than six-octaves) and highly variable in amplitude, frequency content, and time course. Recent studies have even reported that amplitude modulation of the posterior resting alpha activity exhibits self-similar or fractal complexity over many octaves (seconds to hours) . In addition to exhibiting complex patterns of amplitude modulation, EEG signals often exhibit both small frequency shifts and large frequency jumps – phenomenally (if not functionally) akin to a European driver up-shifting and down-shifting while negotiating tight corners on a mountain road . The static ICA modeling discussed in this chapter does not itself attempt to model the temporal properties of EEG source signals. Here, we focus on specific examples of EEG processes and dynamics revealed by ICA-based spatial filtering.

The spatially static ICA model discussed here assumes that the ('far-field') activities recorded at the scalp are produced in cortical EEG source domains that project near-instantly to the scalp electrodes via volume conduction. This means that the scalp electrode signals sum potentials originating in nearly all cortical and artifact source domains. For electrical frequencies in the EEG range, the basic ICA assumptions that volume conduction is linear and practically instantaneous are confirmed by biophysics. Thus, scalp electrode signals can be modeled as instantaneous linear mixtures of cortical source plus non-brain artifact signals. The mechanisms by which the local synchronies appear in the cortical source domains and are modulated across time are of course highly nonlinear. ICA only seeks to cancel the volume conduction and linear summation of distinct cortical (plus artifact) signals at the scalp electrodes – opening the problem of characterizing the (highly nonlinear) generation and modulation of the EEG source activities themselves to further analysis.

To separate EEG source signals, ICA makes a key assumption: that the far-field signals produced by the cortical and non-cortical EEG sources are temporally distinct and, over sufficient input data, near temporally independent of one another. Is this a physiologically plausible assumption? Several factors suggest that in many cases cortical and artifactual source signals may indeed be nearly independent. It is important to realize, first, that establishing the absolute independence of a number of signals would require infinite data. Thus, independence measured by whatever approximation always means near independence. That said, why should the far-field signals produced by different cortical and artifactual sources be near independent?

A simple constraint on cortical signal dependence comes from cortical connectivity, which is very highly weighted toward short (< 500 μm) connections. The largest class of inhibitory cells in cortex, for example, has only short-range connections. Thus, synchrony (or partial synchrony) between local field activities should very likely spread through a contiguous cortical area, rather than jumping between distant and very weakly connected cortical territories. In fact, this is generally what has been observed in recordings from closely-spaced cortical electrode and optical grids.

Finally, the spatially static ICA model assumes that the cortical domains or patches, as well as the artifact sources that together constitute EEG 'sources,' remain spatially fixed over time and therefore project to the scalp channels with fixed weights or proportions. Note, however, that the concept that local field activities within a cortical source domain act in perfect synchrony may be simplistic. Invasive optical and electrical recordings using closely spaced sensors reveal moving sub-millimeter scale potential gradients with traveling wave patterns, leading some to speculate that traveling waves occur regularly in cortex at larger spatial scales. However, if the cortical 'patch' that produces a synchronous far-field EEG source signal is on the cm² scale, progressive or circular traveling wave activity *within* the source domain will produce a signal that appears nearly spatially constant on the scalp.

For example, consider Freeman's model of EEG source dynamics, based on his observations of mammal brains with small (sub-mm spaced) electrode grids, of circular wave patterns that spread across small areas of cortex like pond ripples produced by throwing a small rock into a pond. What field dynamics on the scalp should be produced by such activity active at, for example, 10 Hz. At a nominal traveling velocity of 2 m/s, and assuming a cortical domain diameter of as much as 3 cm, the 10-Hz phase difference between the focal center of the 'pond rippling' potentials and the edge of the active 'ripple' area (1.5 cm from the center) would be only

$$\frac{1.5\mu}{0.002\mu / \mu \sigma^* 100\chi\mu / \mu} / 100\mu \sigma^* \chi\chi\chi\epsilon^* 360 \text{ deg} / \chi\chi\chi\epsilon = 27 \text{ deg}$$

Thus the outer edge of the pond-ripple pattern would lead (or follow) the center by less than a thirteenth of a 10-Hz cycle, and mean local field potentials within the patch (and at the scalp electrodes) would change from positive to negative and back again *nearly* synchronously. Unless the cortical domain involved were (a) quite close to the skull and (b) folded (for example, straddling the edge of a sulcus), the projection pattern on the scalp would be very difficult to distinguish from the far-field potential of a cortical source domain behaving as an ideal, fully-synchronous cortical source. Thus, for resolving signals from most cm-scale source domains similar to those observed by Freeman, ICA may indeed be adequate, at least for EEG frequencies lower than the (30-Hz and above) gamma band. However, more advanced ICA approaches including complex ICA, if performed on Fourier or wavelet transformed EEG data at 10 Hz, might indeed be able to recover, in some cases, evidence of near cm-scale potential flow patterns within individual cortical alpha source domains .

Certain macroscopic EEG phenomena are also known to exhibit large-scale traveling wave properties – including epileptic seizures, slow spreading depressions associated with migraine headaches , and sleep spindles . At best, ICA can only model such phenomena as active within a component subspace, i. e., a set of components each accounting for a spatial (and temporal) phase of the moving activity pattern. One can think of individual component activities in such cases as ‘overlapping movie frames’ that, in combination, can capture spatially shifting phenomena that reoccur in the data. In many cases, the resulting separation of the moving signal into a relevant component subspace separate from the near-independent activities of other, spatially-static EEG sources may be useful for many analysis purposes. There are, however, many possible extensions of the static ICA model to more general blind source separation methods that may also be able to identify and isolate particular sorts of spatially fluid patterns in scalp EEG data – if and when the particular moving-source model being used does in fact fit the phenomena of interest .

Questions about spatial EEG source stability also extend over longer time scales and, as well, across changes in task and state. In general, it is not known if in the normal awake brain, major sources of synchronized activity tend to move *between* cortical source domains, or whether cortical source domains slowly change shape or location, for example during performance of a single task. The relatively stability of the results of infomax ICA decomposition applied to waking EEG data, both within and between subjects suggests that many major sources of EEG activity are sufficiently spatially stable to be reproducibly resolved into temporally near-independent sources with fixed scalp maps that can be used to estimate the spatial locations of the underlying source areas.

Preliminary analysis, however, suggests that the set of cortical source domains separated by ICA may well change when subjects perform a new task . Nonetheless, the spatially static ICA model, applied to data from two tasks performed consecutively in one session, might still find components specific to both task periods, weighting their activities during each task period appropriately. In summary, the subject of spatial non-stationarity of EEG cortical source domains is still a largely open, interesting and potentially important frontier of EEG research.

ICA model basics. The data submitted to ICA are simply the EEG channel recordings arranged in a matrix of n channels (rows) by t time points (columns) data values. Unlike direct spatial filtering methods, *no* channel location information is used in the analysis. ICA performs a blind separation of the data matrix (X) based only on the criterion that resulting source time courses (U) are maximally independent. Specifically, ICA finds a component ‘unmixing’ matrix (W) that, when multiplied by the original data (X), yields the matrix (U) of IC time courses

$$U = WX \quad (1)$$

where X and U are $n \times t$ matrices, and W is $n \times n$. By simple matrix algebra, equation (1) implies that

$$X = W^{-1}U \quad (2)$$

Here, W^{-1} (the inverse of W) is the $n \times n$ component ‘mixing’ matrix whose columns contain the relative weights with which the component projects to each of the scalp channels, i.e. the IC scalp map. The portion of the original data (X) that forms the i^{th} IC (X_i) is the (outer) product of two vectors, the i^{th} column of W and the i^{th} row of U ,

$$X_i = W^{-1}_i U_i \quad (3)$$

and the whole data (X) are the sum of the ICs (X_i)

$$X = \sum X_i \text{ where } i = 1, 2, \dots, n \quad (4)$$

Again, each column of the (W^{-1}) mixing matrix represents, for a single component source, the relative projection weight at each electrode. Mapping these weights to corresponding electrodes on a cartoon head model allows visualization of the scalp projection or scalp map of each source. The source locations of the components are presumed to be stationary for the duration of the training data. That is, the brain source locations and projection maps (W^{-1}) are assumed to be spatially fixed, while their ‘activations’ (U) reveal their activity time courses throughout the input data. Thus, the IC activations (U) can be regarded as the EEG waveforms of single sources, although obtaining their actual amplitudes at the scalp channels requires multiplication by the inverse of the unmixing matrix (W^{-1}) to return to μV units.

Neither the IC scalp maps nor the IC activations are themselves in original recorded units. Rather, the original activity units (μV) and polarities (+/-) are distributed between the two factors – the IC scalp map and activation time series. For example, reversing the polarities of the activation and inverse weight matrices, then back-projecting by multiplying these two matrices (as in Eq. 3 above) recovers the original component activities in their native μV units. Thus, neither the sign of the scalp maps nor the sign of the activations are meaningful in themselves, but it is their product which determines the sign of the potential accounted for at each scalp channel. However, IC activation magnitudes may be normalized by multiplying each by the root-mean square (RMS) amplitude of the corresponding IC scalp map. The activation units are then RMS μV across the scalp array.

The style of ICA decomposition considered here is said to be *complete*, i.e. a decomposition in which the number of ICA components recovered is the same as the number of channel inputs. Thus, 30-channel data will be decomposed by ICA into 30 ICs, whereas 60-channel data will be decomposed into 60 ICs. Methods for overcomplete ICA decomposition also exist, though these require additional assumptions. Frequently asked questions about ICA include: (1) Are there really only a fixed number of data sources? (2) What are the effects of recording and decomposing different numbers of data channels? Although full answers to these questions are mathematically difficult and possibly intractable, in general the number of near-independent brain sources of EEG data should theoretically be nearly unlimited, although our power to resolve them from scalp data from any fixed number of channels is limited.

Results of ICA decomposition of high-density (e.g., 256-channel) data acquired from normal subjects during performance of cognitive tasks suggest that some dozens of distinct EEG sources are large and/or distinct enough to be separated into ICs with physiologically interpretable scalp maps and activations. The remainder of the (e.g., 200+) ICs found by ICA in such data must be either ICs that clearly account for non-brain artifacts, or else mixtures of lower energy sources that are combined to satisfy the requirement that the component activities sum to the whole data. In our experience, applying ICA decomposition to 31-channel data typically yields five to fifteen physiologically plausible components comparable to those obtained from high-density recordings.

4. ICA practice

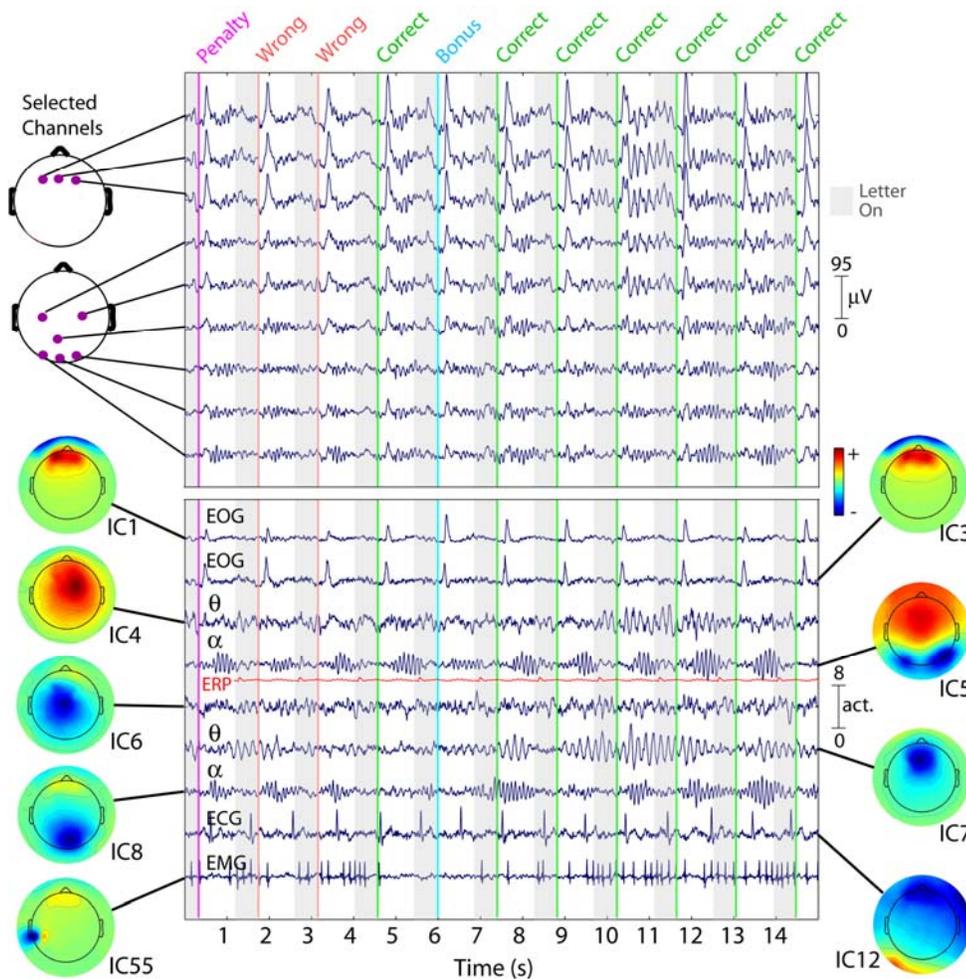


Figure 1. Fifteen seconds of EEG data at 9 (of 100) scalp channels (top panel) with activities of 9 (of 100) independent components (ICs, bottom panel). While nearby electrodes (upper panel) record highly similar mixtures of brain and non-brain activities, ICA component activities (lower panel) are temporally distinct (i.e. maximally independent over time), even when their scalp maps are overlapping. Compare, for example, IC1 and IC3, accounting for different phases of eye blink artifacts produced by this subject after each visual letter presentation (grey background) and ensuing auditory performance feedback signal (colored lines). Compare, also, IC4 and

IC7, which account for overlapping frontal (4-8 Hz) theta band activities appearing during a stretch of correct performance (seconds 7 through 15). Typical ECG and EMG artifact ICs are also shown, as well as overlapping posterior (8-12 Hz) alpha band bursts that appear when the subject waits for the next letter presentation (white background) For comparison, the repeated average visual evoked response of a bilateral occipital IC process (IC5) is shown (in red) on the same (relative) scale. Clearly the unaveraged activity dynamics of this IC process are not well summarized by its averaged response – a dramatic illustration of the independence of phase-locked and phase-incoherent activity.

Example data. In the following sections we will illustrate the advantages of using ICA to isolate EEG source activities in multi-channel data. In these examples, we will use data from a single subject performing a standard ‘two-back’ working memory task. The subject was presented with a letter (B, H, J, C, F, or K) at roughly 1500-ms intervals, and responded to each letter by pressing one of two buttons using their right or left thumb, respectively, to indicate whether the current letter was the same as (match) or different from (non-match) the letter that had been presented two back in the sequence. At each letter offset, an auditory feedback stimulus indicated whether the subject response was correct or incorrect. Letter presentation duration was adjusted after each trial block to induce performance to be as close as possible to 75% correct.

Correct responses each earned the subject 1 point, while incorrect responses cost 1 point. At the end of the experiment, the volunteer subjects were paid a 1¢ (US) bonus for every point accrued. Total bonus money earned was typically in the range of \$10 (US) in addition to their regular hourly rate of compensation. To introduce occasional heightened experiences of reward and punishment, on 10% of the correct and incorrect trials, respectively, upward and downward gliding tones, respectively, were delivered to indicate that the number of points earned or lost on that trial were five times as large as usual. Thus, infrequent ‘bonus’ signals added 5 points, while infrequent ‘punishment’ signals lost 5 points from the subject point total, which was displayed on the subject screen between task bouts.

Figure 1 illustrates typical results of ICA decomposition performed on 1,917 seconds of 100-channel EEG data digitized with 24-bit resolution at 256 Hz. The top half of the figure displays EEG data from a subset of the 100 electrodes over the course of 15 sec, while the bottom half of the figure shows the (‘activation’) time courses of several independent components during the same period. The gray bars show when a letter was displayed on the computer screen. Vertical colored lines indicate the type of auditory performance feedback signal delivered at each letter offset.

During the illustrated data period, the subject blinked after each performance feedback signal (as clearly revealed by the time courses and scalp maps for IC1 and IC3, and as visible in several frontal scalp channels). Other types of EEG artifacts were also isolated by ICA, including left temporal scalp muscle activity (IC55) and cardiac pulse artifact (IC12). ICA found several ICs predominantly projecting to posterior scalp with a peak in the (8-12 Hz) alpha band (e. g., IC5 and IC8), as well as at least three ICs with spectral peaks in the (4-7 Hz) theta band (IC4, IC6, and IC7). Alpha peak components tend to be associated with scalp maps suggesting projection of one, or else of two symmetric equivalent dipole(s) (see Section 4.1) in posterior brain, while frontal components typically have a mean spectral peak in the theta range.

Fig. 1 also shows the concatenated average event-related potential (ERP) waveform time (the mean for all 308 letter presentations) locked to visual letter onsets for IC5. Although the ERP is simple to summarize, comparison of its waveform with that of the unaveraged IC5 activity provides a dramatic example of how the average ERP captures little of the actual process activity. In particular, the IC5 alpha bursts that regularly follow letter offsets in this time span are nearly absent from the ERP because, though they clearly follow most stimulus presentations, they are not precisely phase locked to letter onsets (i. e., the phase of their alpha activity, relative to stimulus onset, varies near randomly). Thus, these large alpha activity bursts are not prominent in the average ERP waveform, having been removed from it by phase cancellation.

Component source modeling. As we have seen, EEG recordings do not contain the experimenters know about the physical locations of the electrodes on the scalp. The electrode locations, however, are at best quite crude indicators of the locations of even the strongest of the underlying cortical

sources. Thus, EEG recordings are typically and traditionally considered to have ‘low spatial resolution.’ The standard approach to EEG source localization is to directly attempt inverse modeling of EEG source locations from one or more observed scalp potential distributions. However, since nearly all recorded EEG scalp maps sum activities from multiple brain and non-brain sources, this has appeared to be a nearly intractable problem.

Originally, some researchers hoped that the number of active EEG sources could be minimized by computing the *average* scalp distribution at some latency or latencies following significant task events, reasoning that all but one EEG process might be cancelled out during response-locked averaging. Unfortunately, more than one cortical EEG process reacts to a stimulus, and further research has shown that such event-related potential (ERP) averaging can only isolate activity from a single brain area if it is the earliest responder to a sensory event. In general, the changes in EEG dynamics induced by significant events quickly involve several cortical areas .

ICA decomposition, however, offers a new and more promising approach to solving the EEG inverse problem – namely, performing inverse modeling of the individual IC scalp maps themselves. The simplest ‘realistic’ EEG inverse model attempts to match one or more observed scalp maps with the best-fitting projection pattern of one or more active equivalent dipoles placed in a 3-D head model at some location(s) and orientation(s) . Applying such standard inverse source modeling methods to the IC scalp maps from the example session of Fig. 1, using a best-fitting four-shell spherical head model, we found single equivalent dipole models for about 20 ICs whose scalp projection patterns adequately fit the observed IC maps (e. g., with no more than 15% residual variance between the IC scalp map and the dipole scalp projection).

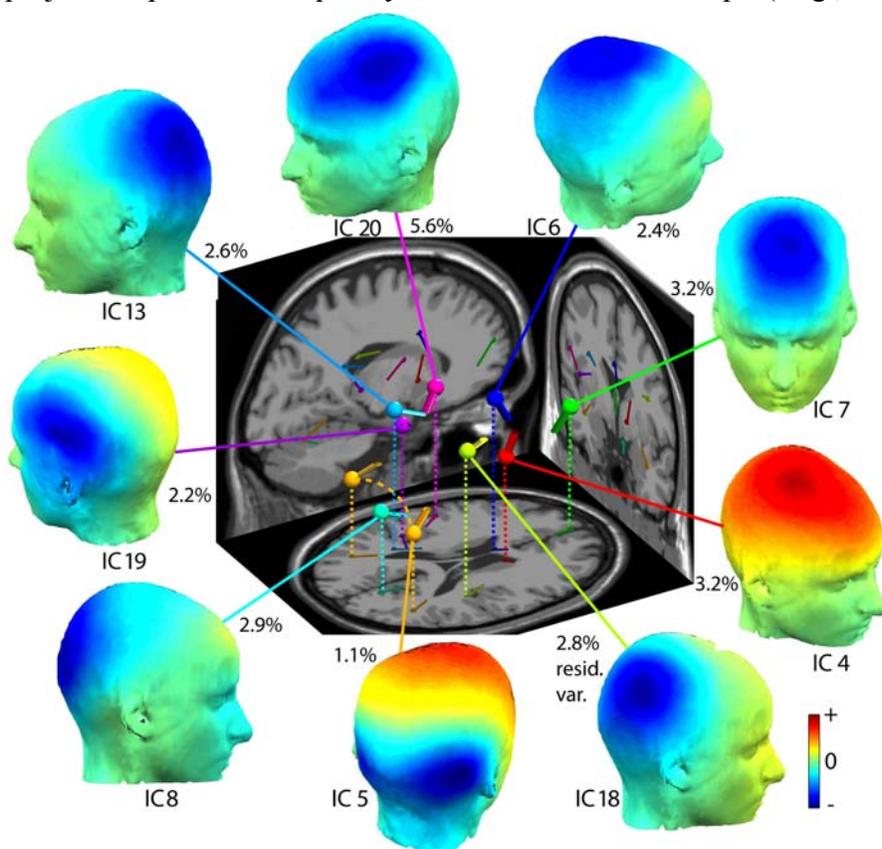


Figure 2. Independent component scalp maps modeled as single dipolar sources. The location and orientation of dipoles within the head model determine the theoretical scalp projection of the dipole’s electric field. For the nine IC scalp maps pictured here (from the decomposition shown in Fig. 1), there is a median 2.8% residual variance between the IC scalp map (W^1) and the model projection of the best-fit equivalent dipole, likely close to the error inherent to fitting sources within a spherical head model. Thus, each of the IC

scalp maps pictured (except IC5) are highly compatible with a cortical source domain consisting of a single cortical patch of unknown extent. IC5 is well modeled by two dipoles located symmetrically across the occipital midline and is likely tightly coupled through the corpus callosum.

Figure 2 illustrates the scalp projections (W^T) of nine ICs from one subject linked to the locations of their respective best-fitting equivalent dipoles in a common 3-D (Talairach) brain space. The indicated percent residual variance specifies the percent difference between the scalp projection of the model dipole and the actual IC scalp map. For all nine pictured ICs, median residual variance was 2.8%. This is near to the minimum level of error expected from imperfectly modeling the head as a set of conductive spheres (representing brain, pial surface, skull, and scalp). Each dipole is associated with a particular orientation that determines the pattern of its scalp distribution. For example, IC7 is oriented radial to the scalp surface (see the sagittal brain projection) and thus projects most strongly to the forehead. The other end of this dipole (e. g., the red end, in this depiction) might project near the throat, but was not recorded by the electrode montage used in this experiment. In contrast, IC5 can be accurately modeled (with residual variance near 1%) only using two occipital dipoles located symmetrically across the midline and oriented tangential to the posterior scalp and thus simultaneously contributing both positive and negative potentials to the scalp recordings.

If an IC scalp map cannot be well modeled by a single dipole, we typically do not attempt to model it. When an IC scalp map appears to be bilaterally symmetrical, it may be reasonably modeled by two dipoles symmetrically located in left and right hemispheres, though possibly with differing dipole orientations. An example of this situation is illustrated in by IC5 in Figure 2. This source configuration might arise from, 1) synchronous evoked responses in left and right visual cortices in response to central visual stimuli, and/or 2) synchronized alpha band activity bi-directionally coupled through dense callosal connections. In the latter case, it might be possible to observe ICs generated in other pairs of cortical areas also bi-directionally coupled by dense white matter tracts, for example the arcuate fasciculus that connects frontal cortex to the ipsilateral temporoparietal junction, though we do not yet have convincing evidence confirming this possibility.

In general, ICs tending to account for most event-related dynamics are likely to be dipolar (or eye artifacts), whereas non-dipolar, noisy-appearing components tend to make little contribution to event-related dynamics. Non-dipolar ICs may represent activity that does not fit the ICA spatial stationarity assumption, or may represent mixtures of small, inconsistent sources. Given the reasonable presumption that the total number of quasi-independent brain and non-brain sources (mostly quite small) may be unlimited, whereas the number of scalp channels is limited, some ICs will contain information from more than one source. However, in our experience, such components tend to account for relatively little of the recorded EEG signals, and even less of the signals of most interest for cognitive neuroscience.

Modeling ICs using single or dual equivalent dipole models not only helps localize activities of interest to a particular brain region; it is also a convenient way to assess and visualize the spatial homogeneity of clusters of functionally similar or equivalent IC processes across subjects (see Section 6.1). However, more advanced methods of distributed source modeling incorporating structural information from subject magnetic resonance head images, should be still more informative.

Practical considerations. Decomposing EEG data by ICA is relatively straightforward, though the quality of the results is highly dependent on two major factors. First, the number of time points of n -channel data used in the decomposition must be sufficient to learn the n^2 weights in the $(n \times n)$ ICA unmixing matrix. As a rule of thumb, the amount of data needed is related to the number of electrodes squared (or the number of weights) times a factor, k . In our experience, k should be 25 or larger, particularly when the number of electrodes is relatively high. For example, using 256 electrodes, for $k=25$, $256^2 \times 25 = 1,638,400$ or more data points may then be required to perform a full-rank (256-component) decomposition. At a 256-Hz sampling rate, this would require nearly two hours of data. For smaller numbers of channels, the amount of data required should be much smaller. For example, for one quarter the number of channels (64), only a sixteenth the amount of data (~7 minutes) would be required to give the same k factor. However, ICA decompositions using still more data ($k > 25$) tend to be more regular and produce more dipolar component maps. Thus, more data is better – so long as the EEG source locations do not shift. For example, jointly decomposing data from awake and sleeping conditions might not be optimal, if the EEG source locations in these portions of the data differed.

Second, the universal rule of signal processing, ‘*garbage in, garbage out*’ (GIGO), applies to ICA decomposition as well. Two classes of artifacts must be considered to decide what constitutes undesirable (‘garbage’) data for ICA. EEG artifacts arising from eye movements, eye blinks and muscle tension have stereotyped scalp projections (since the positions of the eyes and muscles do not change throughout the session), although eye movements in different directions, or blinks of each eye separately, etc., may introduce multiple scalp projection patterns into the data and thus require more than one IC to account for their artifacts.

Another class of artifacts is more problematic for ICA decomposition. These include large muscle movements such as clenching the jaw, talking, swallowing, clearing the throat, or scratching the scalp. Because these activities involve many muscle groups and possible electrode movements, the potential maps they create in the data are likely to be unique, so their inclusion in data submitted to ICA may require many ICs to be devoted to them, to account for the unique scalp patterns they introduce. Even when a large number of channels and ICA dimensions, are available, relatively few ICs will then be left to account for independent cortical sources of interest. Therefore, it is important to remove such ‘irregular’ artifact periods from the data *before* final decomposition.

Artifact prototypes. One of the simplest applications of ICA to EEG data is to remove stereotyped artifacts from the data. As mentioned above, eye blinks are highly spatially stereotyped and are usually separated into one or more ICs that together neatly account for the entire eye-blink artifact. Once a component has been identified as artifactual, it may be removed from the data by reversing the ICA linear unmixing process. To remove the activity of the k^{th} IC from the data, simply replace the entire k^{th} row of the component activation matrix U with zeros and then multiply the modified activation matrix, U_{k0} , by the ICA mixing matrix W^T . The resulting back-projected data will be the same size and units as the original data matrix, though its intrinsic dimensionality (or ‘rank’) will have been reduced by 1 and the data will no longer contain activity from IC k .

The process of determining whether an IC accounts for an artifact or a cortical process may in many cases be simple for the trained eye, but typically requires some evaluation of its activity pattern. Figure 3 demonstrates two types of data we use to determine whether an IC accounts for cortical or artifact activity. Each panel shows the activity power spectrum (left) and the actual activity time course in the single experimental trials (right) of four ICs from the same session as Figs. 1-3. Panel A shows a typical spectrum of eye blink activity, consisting of relatively high power

at low frequencies and no spectral peaks. This subject blinked (A, right) consistently following each stimulus pair. While this is unusual, and so cannot be considered a criterion for eye-blink identification, it was not unreasonable behavior for a subject in this demanding task. Eye blinks are one of the easiest types of components to recognize by scalp map alone because (particularly when the electrode montage includes sites below as well as above the eyes) the scalp map of a blink-related IC will be nearly identical to that shown in panel A.

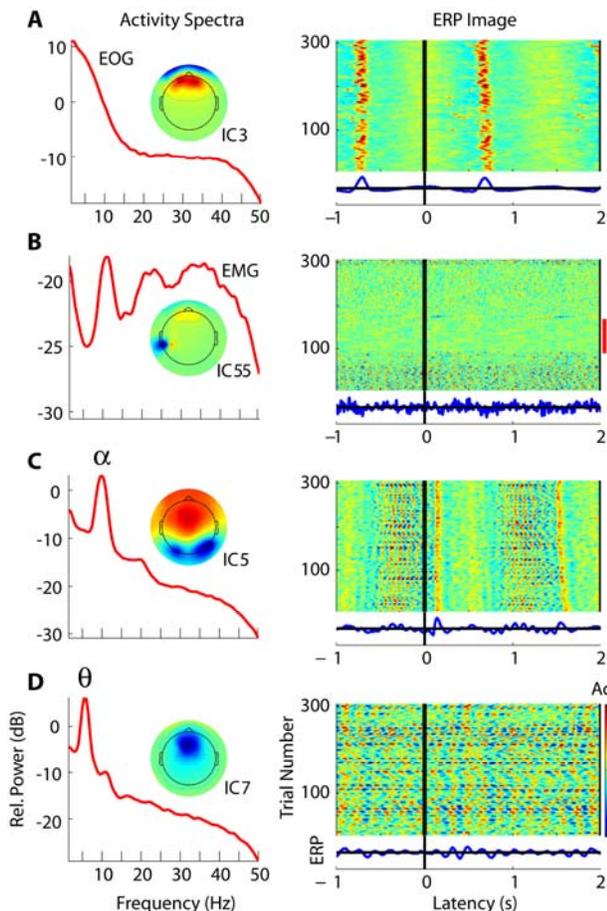


Figure 3. Power spectra and ERP-image plots of single-trial activities time locked to auditory feedback stimuli for four selected independent components (ICs). (A) Eye blinks create high power at low frequencies and characteristic blink-like deviations in single trials that may (as here) or may not be regularly time-locked to experimental stimuli. (B) Muscle tension (here, in a left temporal muscle) is associated with high power above 20 Hz. The roll-off of power above 40 Hz here was produced by 50-Hz low pass filtering of the original data. Note the absence of this muscle’s activity during the middle portion of the experiment (red bar on right). (C and D) IC processes accounting for cortical EEG activity during cognitive tasks typically show a mean spectral peak in either the (8-12 Hz) alpha band (posterior ICs) or (4-7 Hz) theta band (anterior ICs), and produce both phase-locked evoked and phase-random induced spectral responses to significant stimuli (C). They may also exhibit partial phase resetting of their ongoing activity at their characteristic frequencies (D).

Panel B displays properties of another common type of EEG artifact, sporadic muscle tension in scalp or neck muscles. This is usually well-modeled by ICA. Unlike eye blink components, muscle component scalp maps vary depending on the muscle they represent. The equivalent dipole for a muscle component is oriented parallel to, and ideally within the scalp, and the component scalp map exhibits a sharp polarity reversal at the muscle’s point of insertion into the skull (e.g., the shift from red to blue in the scalp map in panel B). Muscle component spectra typically have highest power at frequencies above 20 Hz. Other frequencies may be present (as, here, near 10 Hz), though the most prominent feature of electromyographic (EMG) activity is the dominant power in the gamma range. Muscle tension is typically not maintained throughout an entire experiment, instead tending to switch high or low for stretches of time (likely without explicit awareness of the subject). As an illustration, a red bar to the right of the ERP-image plot in panel B indicates a block of trials when this muscle’s activity was relatively quiet. Thus, ICs accounting for muscle activity may be identified using multiple criteria and subsequently removed from the rest of the data, if desired.

For comparison, panels C and D display data from two types of cortical IC processes with strong alpha (C) and theta (D) rhythms, respectively, as indicated by the strong spectral peaks in their activity spectra. Furthermore, the ERP-image plots (right) of the activities of these components in single trials show consistent patterns of activity time-locked to experimental stimuli of interest. Other criteria can be used to determine which ICs are putative cortical components, such as the fit of each IC scalp map to a best-fitting single (or sometimes bilateral) dipole model (see Section 4.1). The estimated location of an IC can be another helpful tool for discovering whether or not a component explains artifact activity. For example, eye blink and muscle ICs will typically localize outside of the brain volume (given a reasonably good co-registration between the head model and the actual electrode locations).

5. ICA applications

Time-domain analysis of independent component activities. Traditionally, EEG analysis has focused on the averaged scalp potentials in relation to a time-locking event. These average event-related potentials (ERPs) are interpreted as if they captured the relevant event-related EEG activity. It is important to realize that evoked potentials, even at relatively early time points, sum projections to the scalp of the net activities in many EEG sources that survive averaging. In addition, ERPs capture only the portions of the single-trial signals that are both time-locked and phase-locked (e.g., usually positive or negative) to the set of time-locking events. Often, this may be only 1% of the single-trial activities. It cannot be concluded from this that the other 99% of the recorded signal that does not survive averaging is unaffected by the time-locking events. In particular, event-related modulation of the power spectrum in one or more parts of cortex may occur which is generally time-locked to the time-locking events but leaves the distribution of signal phase at all relevant frequencies random with respect to event onsets. Measuring such event-related modulations in signal spectral power is a major purpose of time/frequency analysis.

Time/frequency analysis of independent component activities. As the red trace in Fig. 1 shows, and contrary to naïve assumptions often gradually accrued in the minds of researchers who come to view data averages as if they were the data themselves, the averaged evoked response typically retains little of the complexity of the event-related EEG dynamics that follow (or, sometimes, anticipate) significant events in cognitive task paradigms. As Pfurtscheller (1977) realized over 25 years ago, changes in amplitude of EEG oscillations time locked but *not* phase locked to a set similar events are not evident in ERP averages. Pfurtscheller's event-related desynchronization (ERD) and later synchronization (ERS) measures capture mean event-related dynamics of oscillatory activity in a selected narrow band (originally, the alpha band).

In 1993, Makeig generalized this analysis to consider a wider range of frequencies at once, producing a mean latency-by-frequency image, the event-related spectral perturbation (ERSP) image. The time/frequency measures and equivalent 2-D images we call ERSPs reveal the frequencies and latencies when mean changes in log power occur, relative to some mean power baseline, time-locked to a class of experimental events. Subtracting log mean baseline power (or equivalently, dividing mean power by mean baseline power and then taking the log of the result) measures how strongly mean, event-related power at different frequencies either increased or decreased relative to the baseline mean. This normalization models spectral perturbations as results of multiplicative influences by which ongoing activity is either augmented or reduced, time locked to events. Note the use of mean baseline and event-related power. Converting to (log) dB scale after taking means

regularizes the measure, avoiding extreme negative values that would be produced by taking the log of individual power spectra containing near-zero values. The normalization also removes the dominance of low frequencies in cortical EEG (see Fig. 3, left column), making a common color scale applicable for all frequencies.

When *interpreting* ERSP images, it should be remembered that an ERSP is a statistical measure. EEG activity at both source and scalp levels is typically highly variable from second to second and trial to trial. Thus, an event-related increase at some frequency by 6 dB should not be imagined to mean that a steady train of oscillatory activity at that frequency continues following the event, only doubled in amplitude. More likely, either the frequency of occurrence of bursts at that frequency doubles following the events in question, or the mean amplitude of the bursts doubles – or some combination of the above. In general, mean measures (including ERPs and ERSPs) do not characterize the variability of the quantity measured, nor more generally, its distribution – for this, further analysis is required.

To determine which ERSP features are significant, we have implemented non-parametric testing using data permutation methods to create distributions of surrogate data whereby statistical significance of the observed mean changes can be evaluated. Specifically, for each time/frequency point we generate a distribution of pseudo or surrogate data points in which all features of the actual data except one are preserved. For example, to assess the significance of a non-zero ERSP value at a given time/frequency point, we could repeatedly shuffle the latency of the time window from each trial used in the mean ERSP computation, thereby collecting a distribution of some hundreds of surrogate mean ERSP values representing random variations in mean spectral power at the selected frequency that occurred during the trials. If the actual mean ERSP value at the time/frequency point is found to be outside the distribution of surrogate values (e.g., for $p < .01$, outside its 99% percentile), then the observed mean ERSP difference at this time/frequency point may be said to differ significantly from expected power fluctuations at this frequency in the data. That is, there is likely to be some influence at the selected latency that increases (or decreases) power at this frequency. It is important, of course, to recognize the problem of multiple comparisons when interpreting results of significance testing. For example, at a pixel significance threshold of $p \leq .01$, one percent of even random data transforms should be expected to be marked as ‘significant.’

To test the significance of differences between mean ERSPs in two conditions across a group of subjects, one may simply shuffle the assignments of ERSPs to the two conditions and make a distribution of surrogate mean difference ERSPs whose limits at each time/frequency point define the bounds of expected variation of the ERSP contrast of interest. For comparing ERSP features across subjects, a computationally somewhat simpler though possibly less reliable method of evaluating the reliability of the observed ERSP differences is to match the observed number of significant power perturbations at a given time/frequency point for the individual subjects, and then to reject observed differences whose significance across subjects is less than a pre-defined binomial probability level. This procedure often yields strong significance levels that minimize the chance of significant results arising simply from the problem of multiple comparisons.

Analysis of both scalp channel data in single trials and IC activities shows that in most cases, the cortical sources contributing to ERP activity may be the same as those contributing to concurrent event-related changes in channel power spectra. Further, ICA spatial filtering shows that oscillatory activity of EEG source processes contributing to average ERPs is usually only *partially* phase reset (or, alternatively, phase locked or phase constrained) relative to the time locking events. That is, mean spectral amplitude shifts and degrees of phase locking at various frequencies and latencies with respect to events of interest form a multi-dimensional space of possibilities for event-related

dynamics of cortical source processes . That is, ERP features and event-related spectral power shifts or perturbations are usually not, in fact, distinct phenomena. Instead, each present a limited view of more complex event-related brain dynamics that involve changes in both spectral power and phase at one or more frequencies . Naturally, their application to data recorded at single scalp channels confounds the multiple and often partially cancelling contributions of the underlying EEG sources, complicating the interpretation of event-related phase consistency in scalp channel records.

Another specific pitfall of both ERP and ERSP analysis is that the typically flat pre-stimulus ‘baseline’ of average ERPs and ERSPs suggests that the activity captured in the ERP or ERSP is produced by cortical sources that are silent (ERP) or active in some steady state (ERSP) before the events of interest. Neither average ERPs nor average ERSPs are sufficient measures of event-related dynamics unless all the single-trial data exhibit the same source dynamics plus other (‘noise’) processes that are unaffected by the events of interest -- an assumption that a recent detailed analysis, like inspection of Fig. 1, shows to be inadequate . In general, the sources of ongoing EEG activity isolated by ICA show highly variable activity patterns in pre-stimulus periods, and complex transformations of their joint amplitude and phase statistics following significant stimuli in event-related paradigms.

A next important step in ICA analysis of EEG data is to analyze the IC activity (or ‘activation’) time courses themselves. This involves, first, performing an ICA decomposition of data of adequate quantity and quality to yield physiologically plausible components. Next, physiologically plausible ICs should be selected for detailed analysis. Typically a workable criterion is that the IC be reasonably well-fit by a single (or bilateral) equivalent dipole(s) (e. g., within ~15% residual variance; see Section 4.1) and that this dipole be localized inside the head volume. If a dipole cannot be fit to an IC scalp projection, then the homogeneity of its activity may be questioned. As mentioned in Section 4.3, ICs that also express physiological rhythms such as alpha or theta activities are likely to represent both anatomically and functionally distinct cortical source activities (though Fig. 3B shows a counterexample).

Note, however, that IC activities represent the result of instantaneous and therefore broad-band filtering of the scalp data. Each IC activation time series apparently represents the synchronous portion of activity within one (or sometimes two linked) patches of cortex, and should therefore have a broad colored-noise spectrum with or without single or multiple spectral peaks. If multiple peaks are present, higher peaks may represent harmonics of lower-frequency peaks (e.g., harmonics of alpha or theta rhythms arising from the non-sinusoidal, more triangular wave shape of the oscillatory activity), and/or bursts of unrelated higher-frequency activity interspersed between periods of dominant frequency activity .

Figure 4 (below) directly compares time/frequency analysis for a mid-frontal scalp channel and for three contributing ICs. It makes visible the ambiguities inherent in time/frequency analysis of scalp channel data alone. Panels A and E show the ERSP and inter-trial coherence (ITC) images, respectively, for a frontal midline scalp channel indicated by the grey disc on the model head. Here, artifacts including vertical and lateral eye movements, pulse artifact, and muscle activity were removed from the data before time/frequency transformation. The side panels display ERSP and ITC images for three dipolar components contributing to the channel ERSP, as shown by the relation of the channel location (grey disc) to each of the IC scalp maps.

The ERSP transform of the frontal channel data (A) is quite similar in pattern to that of IC5 activity (D), though the blocking of (10-Hz) alpha band activity in this bilateral occipital component is stronger (~12 dB versus ~6 dB). In contrast, the small event-related alpha power increase (ERS) of midfrontal component IC7 (B) is not visible in the channel data transform (A). Further, the post-

stimulus increase in theta band activity of midfrontal IC4 (C) is not visible in the scalp channel data transform (A). Either another theta-band component activity projecting to this channel may have decreased during this time interval, thus balancing the increase in IC4, or the summed mixture of many source activities at the scalp channel had more trial to trial variability, making the alpha increase from IC7 insignificant in the channel data. Figure 4 illustrates that not only may scalp channel ERSP measures be blind to some of the event-related spectral perturbations in the underlying cortical sources, but they may also suggest incorrect conclusions about source locations. Here, the strongest IC contributor to this *frontal* electrode ERSP is in fact from an *occipital* source. Thus, results of time/frequency analysis applied to single-channel data should also be interpreted with caution.

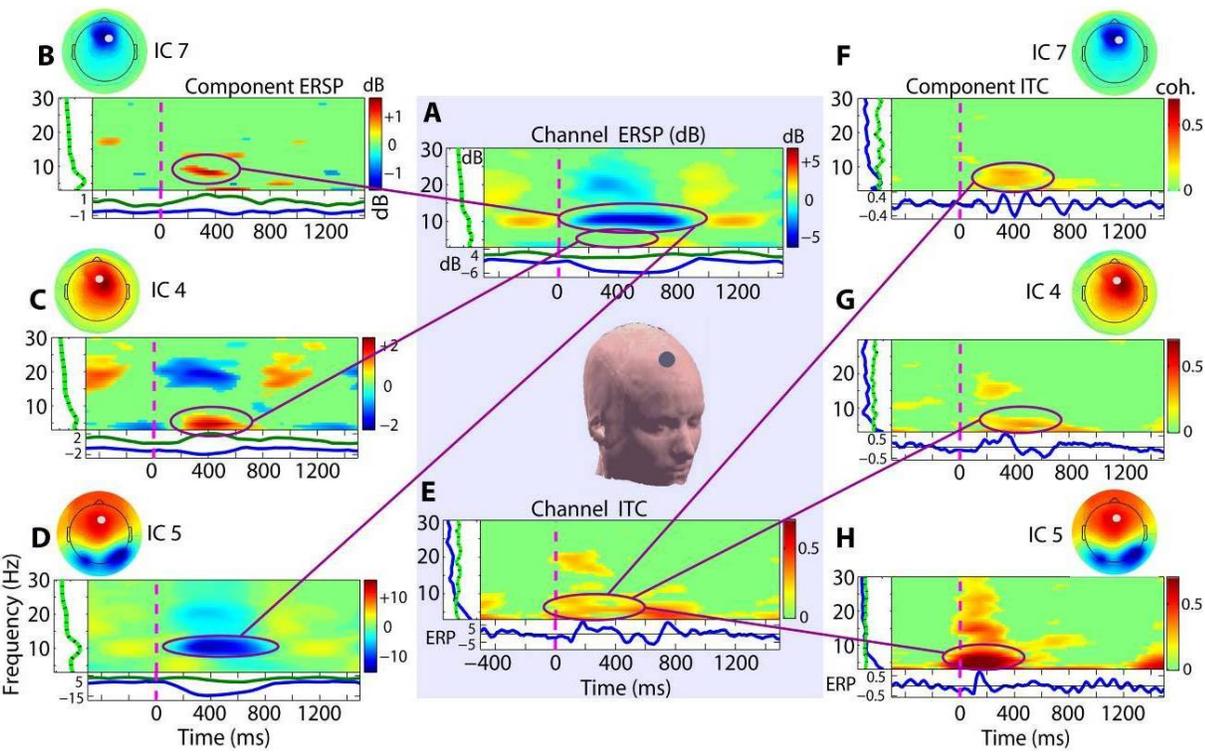


Figure 4. Comparison of average event-related spectral power (ERSP) and inter-trial coherence (ITC) measures time-locked to letter onsets in the two-back task, at a single frontal midline scalp channel (A and E, center) and for three ICs projecting to this channel (ERSP: B-D; ITC: F-H). Green portions of the ERSP and ITC images are non-significant ($p > .01$) by surrogate data testing. ERSP and ITC measures of event-related IC signal dynamics represent different aspects of event-related perturbations of ongoing oscillatory activity. Component ERSP and ITC features may conceal differences in the summed contributions of different sources to individual scalp channels, as is the case here, both for theta and low alpha band power and phase locking. The far left panels of A-D (green traces) show the mean component activity spectra, and the lower panels (blue and green traces) show the maximum and minimum spectral power perturbations. The green and blue traces of the left panels in E-H show the mean and maximum ITC values at each frequency, while the lower panels in F-H show the activity ERPs

In itself, the average ERP says nothing about the portion of the energy in the whole EEG signal that it captures. Inter-trial coherence (ITC), first introduced as ‘phase-locking factor’ by Tallon-Baudry et al. (1996), is a measure of trial-to-trial phase consistency at each frequency and latency relative to a set of time locking events. Its theoretic range is from 0 (uniform phase distribution across trials) to 1 (identical phase in each trial). Most EEG processes contributing to ERPs do so via partial rather than complete phase locking, i.e. with an ITC \ll 1. In Fig. 4, the marked theta-band ITC differences between the contributing ICs are unavailable in the scalp channel ITC. As with scalp channel ERSP analysis, scalp channel ITC measures may mask differences in underlying phase consistency of the distinct cortical source activities whose projections are mixed at the scalp electrodes.

6. Comparing IC activities across subjects

Do different subjects have equivalent ICs? In exchange for the benefits that ICA offers to EEG analysis in both spatial and temporal resolution of separable source-level activities, it also introduces a new level complexity into EEG analysis. In traditional scalp channel signal analysis, clustering of event-related EEG phenomena across subjects is straightforward, as each scalp electrode is assumed to be comparable with results from equivalently placed electrodes for the all subjects. Comparing ICA results across subjects, on the other hand, requires that, if possible, subject ICs from different subjects should likewise be grouped into clusters of ICs that are functionally equivalent despite differences in their scalp maps.

As we have seen, however, data recorded at a single scalp channel *within* each subject is heterogeneous, so the idea of grouping channel activity *across* subjects may actually be a risky proposition. In particular, clear physical differences between subjects in the locations and, particularly, the orientations of cortical gyri and sulci mean that even exactly equivalent cortical sources may project, across subjects, with varying relative strengths to any single scalp channel location, no matter how exactly reproduced across subjects. Thus, the basic assumption in nearly all EEG research, that activity at a given scalp location should be equivalent in every subject, is itself questionable. On the other hand, changing the basis of EEG evaluation from scalp channel recordings to IC activities necessitates an extra step compared to channel analysis – that of combining and/or comparing results across subjects through identifying equivalent IC processes, if any, in their data.

Approaches to IC clustering. The process of identifying sets of equivalent ICs across subjects, or even across sessions from the same subject, can proceed in many ways depending on the measures and experimental questions of interest. An appealing approach to clustering ICs is by their scalp map (W^I) characteristics. Such clustering can be attempted by eye, by correlation, or by an algorithm that searches for common features of IC scalp maps. The disadvantage of this method is that, as shown in Fig. 2, slight differences across subjects in the orientation of equivalent dipoles for a set of equivalent ICs can produce quite different IC scalp maps.

Clustering ICs based on the 3-D locations of their equivalent dipoles may avoid this problem. Using this method, it is possible to describe typical event-related or other activities in cortical areas of interest, or at least in cortical areas with sufficient density of IC equivalent dipoles across subjects or sessions. Common clustering algorithms such as K-means and other distance-based algorithms can be used to cluster ICs based on the 3-D locations of their equivalent dipoles quickly and easily. However, clustering on estimated cortical location alone may introduce similar confounds as

clustering by scalp channel location, since subjects may have multiple types of IC processes in the same general cortical regions.

For one, comparing cortical locations across subjects raises the same spatial normalization questions as arise in functional magnetic resonance imaging (fMRI) analysis. Since brain shapes differ across subjects, true comparison of 3-D equivalent dipole locations should be performed only after spatially normalizing each set of subject IC locations to his or her normalized individual structural magnetic resonance (MR) brain image. This requires MR images be obtained for each EEG subject, a requirement that may greatly increase the resources required for EEG data acquisition.

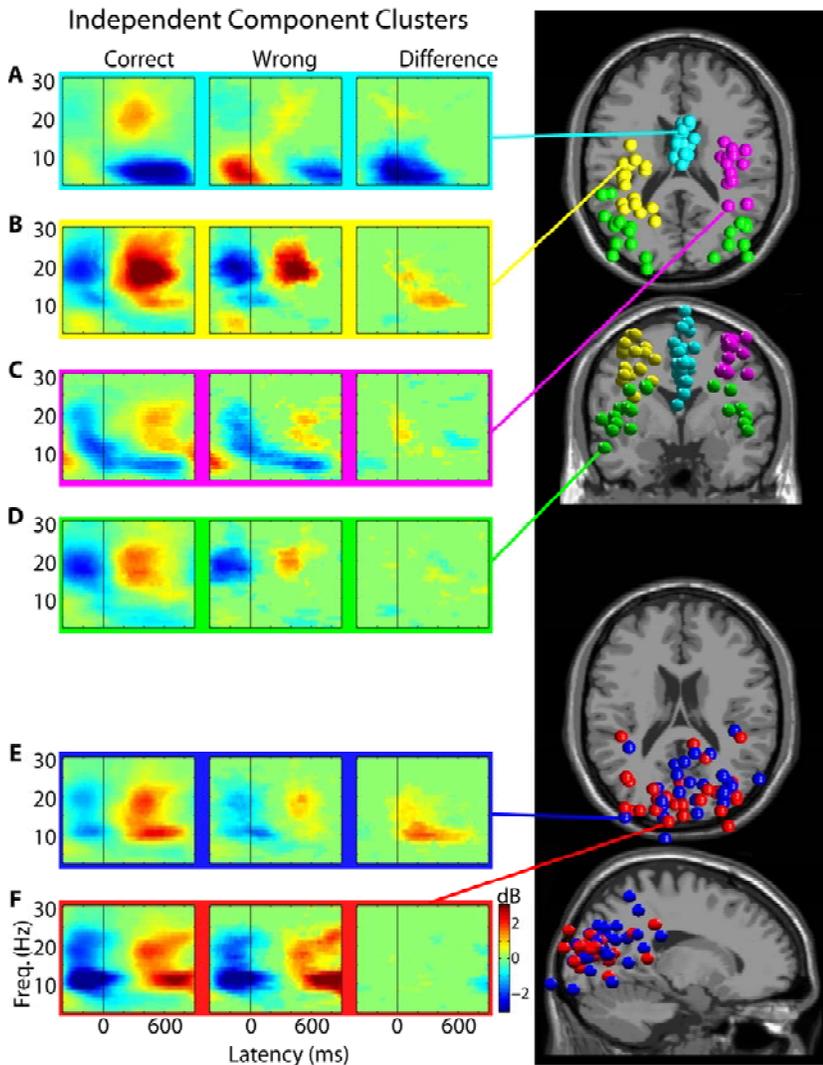
A simpler method normalizes the 3-D equivalent dipole locations via normalizing the subject head shape, as learned from the recorded 3-D locations of the scalp electrodes, to a standard head model. When 3-D electrode location information is not available, the expected functional specificity of equivalent dipole clusters based on estimated equivalent dipole locations in a standard head model must be reduced. In this case, some IC processes estimated to be located in the same cortical area may not express the same functional activities. Despite this drawback, our results show that clustering component dipole locations in a standard spherical head model still allows meaningful conclusions about differences in regional EEG activities across one or more subject groups, if sufficient statistical testing is applied to the data, and if the limitations of the analysis are acknowledged.

Because homogeneity of an IC cluster is most accurately assessed and characterized by the activities of its constituent ICs, a more direct route to obtaining *functionally* consistent clusters may be to group ICs from experimental event-related studies according to their event-related activity patterns. For example, a recent EEG/fMRI study clustered components contributing most strongly to the event-related negativity (ERN) feature of the average ERP time locked to incorrect button presses in a speeded choice manual response task. Remarkably, the authors showed that trial-to-trial variations in the strength of the activity underlying the ERN correlated with changes in the fMRI BOLD signal only in the immediate vicinity of the equivalent dipole source for the component cluster. In some cases, therefore, clustering ICs on similarities in their ERP contributions can be a simple but powerful approach to discovering sources of well-documented ERP peaks.

If the measure of primary interest is not the average ERP but, instead, event-related fluctuations in spectral power of the ongoing EEG across frequencies and latencies, as measured by average ERSPs, then component ERSP characteristics may similarly be used as a basis for IC clustering. Given a small number of subjects and a simple experimental design, it might be possible to group component ERSPs across subjects by eye, though this quickly becomes discouraging as the number of subjects and/or task conditions rise. In any case, an objective approach is more desirable.

As an example, let us consider data from the same two-back task described earlier and illustrated for one subject. Assume there are 20 subjects, each with a mean of 15 dipolar cortical ICs. To prepare the data, each 2-D (*latencies, frequencies*) component ERSP image for one or more task conditions (correct, incorrect, etc.) must be concatenated and then reshaped into a 1-D (1, *latencies* × *frequencies* × *conditions*) vector. Thereafter, the vectorized component ERSPs from the 20 subjects can be concatenated to form a large 2-D matrix of size (#*ICs*, *latencies* × *frequencies* × *conditions*) or, in this case, (300, *latencies* × *frequencies* × *conditions*). A number of options are now available. A simple approach is to compute a distance measure between the rows of the matrix (e.g., using the *pdist* function in the Matlab Signal Processing Toolbox) and then to cluster on the distances between them. Alternatively, PCA and ICA can be used to prepare the data for K-means clustering.

Figure 5. Clustering ICs from 29 subjects by common properties of their mean event-related activity time courses can be an efficient method for finding homogeneous groups of independent processes across sessions or subjects. Here, ICs were clustered by similarities in 3-D dipole location as well as



features of their mean ERSPs time-locked to auditory performance feedback signals within two task conditions (following Correct and Wrong button presses) and by the significant ERSP difference (when any) between them (this significance estimated by non-parametric binomial statistics, $p < 1e-5$). Colored spheres show the locations of the equivalent dipoles for the clustered components. Colored lines connect these clusters to the respective cluster-mean ERSP and ERSP-difference images. Although IC equivalent dipole locations were only a portion of the data used in the clustering algorithm, the equivalent dipole models for four of the obtained clusters (A-D) are spatially distinct. Two other, spatially intermingled clusters (E-F) illustrate how activity-based clustering can differentiate spatially similar components that would not be separated in clustering based on location alone.

It is also possible to combine dissimilar IC activity and/or location measures in computing component ‘distance’ measures. The current release (v5.0b) of our Matlab-based open source EEG analysis toolbox (EEGLAB, <http://scn.ucsd.edu/eeglab>) includes a clustering interface that uses the latter technique. The ICACLUST facility enables clustering on a variable set of IC features including ERPs, ERSPs, scalp maps, mean spectra, and/or dipole locations across subjects or sessions.

Figure 5 illustrates preliminary clustering results on 368 near-dipolar ICs from 29 subjects performing the two-back task described earlier. In the figure, equivalent dipoles of the same colors were clustered by computing a Euclidean ‘distance’ measure between the concatenated average component ERSPs time-locked to auditory feedback tones signaling ‘correct’ and ‘wrong’ responses, as well as the significant ERSP difference between the two, and also 3-D dipole location of each IC. The ERSPs plotted for each cluster represent the means over all the cluster components, after masking spectral perturbations not significant ($p < 0.00001$) by binomial probability across the set of clustered components.

Note that although IC equivalent dipole locations were here *only a portion of the data used in the clustering*, ICs with similar event-related activity patterns proved to be naturally associated with distinct cortical regions. From the ‘Difference’ activity the (A, light blue) central midline cluster, it is clear that this cortical area produced a different activity pattern following wrong responses, namely a 400-ms theta band burst that began before the auditory feedback during the period of the motor response. This result is in line with our previous findings, and neatly reproduces the recent result of Debener et al. (2005) who used time-domain analysis of simultaneous EEG and functional magnetic resonance imaging (fMRI) data to show that trial-to-trial variations in post-error activity of a very similarly located IC cluster were correlated with trial-to-trial variations in fMRI blood oxygen level-dependent (BOLD) signal only directly below the cortical projection of the component cluster, and highly coincident with the location of the equivalent dipole cluster in Fig. 5.

ERSP and equivalent dipole locations for three other activity-derived but spatially ‘tight’ component clusters are shown in Fig. 5. IC clusters located in/near left and right hand somatomotor cortex (B, yellow and C, magenta, respectively) exhibited significantly stronger alpha activity for a half-second after receiving Correct feedback than after receiving Wrong feedback, as confirmed by permutation-based statistical testing. Analysis of responses to matching versus non-matching letters (not shown) revealed the expected dominance of spectral perturbations contralateral to the actual response hand. Finally, event-related perturbations in spectral power in the bilateral occipital (D, green) IC cluster differed little following Correct versus Wrong feedback.

In contrast, the same clustering of ICs on dipole locations, condition ERSPs, and ERSP differences produced two more spatially diffuse and highly overlapping occipital dipole clusters (red and blue). Although the spatial distributions of these two clusters cannot be distinguished, the difference ERSPs between the two clusters do differ. Much like the left somatomotor (yellow) cluster, components in the blue cluster exhibit increased alpha activity following Correct auditory feedback. Of course, *interpretation* of the clustered component activities is a problem separate from clustering. However, once successful clustering of independent component activities has been accomplished, meaningful conclusions about brain function may be approached with more confidence.

7. Summary/Conclusions

This review is intended to provide an overview of how ICA is currently applied to EEG data decomposition and to time/frequency analysis of EEG data in particular. As we have tried to illustrate through sample results and explanations, ICA is a powerful tool for EEG analysis. Because of the numerous cortical EEG sources, as well as the considerable variety of stereotyped non-brain EEG artifacts, experimental results for either averaged or unaveraged scalp channel data are inherently ambiguous and may become even more so when data are pooled across subjects.

Nearly eleven years after the utility of ICA for EEG analysis was first discovered by the senior author and colleagues, it is increasingly accepted by many EEG researchers that ICA is an effective method for removing stereotyped data artifacts including eye blinks and lateral eye movements, muscle activities, electrode or line noise, and pulse artifacts. Independent components accounting for and isolating these artifacts can be recognized easily. The value of separately studying the scalp maps and activity time courses of the *non-artifact* cortical components extracted by ICA from high-density EEG data is still less generally understood or adopted. We predict, however, that the benefits of this approach to EEG (and MEG) analysis will gradually become clear to the field, particularly as more and more students and younger professional researchers explore their own data using freely available analysis tools such as are being made available in EEGLAB and other publicly available toolboxes.

Here, we have demonstrated, first, how scalp channel average ERPs, long given a starring role in cognitive EEG research, are in fact composed of signals from many cortical and artifactual sources. Without a method for spatially filtering the scalp-channel data to reveal the underlying source signals, these signals are summed or confounded in scalp channel averages. In fact, a peak in a scalp channel ERP may not occur at a time point at which *any* of the contributing cortical signals actually reach a peak. This point is important to keep in mind when specific labels for ERP peaks (N100, P100, etc.) are supposed to represent latencies when some ‘thing’ occurred in the subject’s brain. For example, in the ERP shown in Fig. 3, several sources were active, though the average time course of no single contributing IC process mimicked the rise and fall of the averaged scalp channel potentials. Latencies of ERP peaks are thus only moments at which the *sum* of the underlying cortical source signals created a net peak in scalp potential. In addition, as shown by the long history of event-related time/frequency research, ERP peaks may not correspond to moments when EEG energy peaks, either at the scalp channels or in the cortical source signals.

While event-related average ERD and ERS, or more generally ERSP measures provide more information about event-related EEG dynamics than is available in average ERP measures, these spectral power are subject to similar spatiotemporal confounds as ERPs. A single scalp electrode receives projections from many cortical areas, thus spectral power measures are the result of arbitrary summation and cancellation of distinct source signals. Similarly, event-related average ITC measures can be highly affected by source mixing, as was the case illustrated in Fig. 5 where the strength of event-related ITC in at least one of the contributing IC processes was underestimated in the ITC computed at a supervening scalp channel.

Inevitably, with the added benefits of ICA come some additional complexity and inconvenience. This complexity, however, reflects the actual and likely irreducible complexity both of the brain itself and of the recorded EEG signals. This complexity cannot be reduced, but instead only hidden and/or confounded by modeling the data using simpler measures.

Chief among the problems introduced by ICA decomposition is the issue of how to cluster ICs across subjects and/or sessions. This is an evolving research area that will likely become more widely explored as researchers discover new ways of working with independent signal components. Our own results in this direction have convinced us that component clustering, when successfully accomplished, can increase the amount, consistency, and utility of information about macroscopic event-related brain dynamics that can be extracted from high-dimensional EEG (or other electromagnetic) brain signals . We look forward to the results that will come from next decade of research in this direction and beyond.

Acknowledgments

The authors acknowledge the support of the Swartz Foundation and the National Institute of Mental Health, USA, as well as the value of collaborations with Terrence Sejnowski, Tony Bell and Tzyy-Ping Jung to this research, plus the work of Arnaud Delorme and the many contributors to the EEGLAB programming environment.

References

- Anemuller J., Sejnowski T.J., Makeig S. (2003) Complex Independent Component Analysis of Frequency-Domain Electroencephalographic Data. *Neural Netw* 16:1311-1323.
- Arieli A., Sterkin A., Grinvald A., Aertsen A. (1996) Dynamics of Ongoing Activity: Explanation of the Large Variability in Evoked Cortical Responses. *Science* 273:1868-1871.
- Bell A.J., Sejnowski T.J. (1995) An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Comput* 7:1129-1159.
- Blair R.C., Karniski W. (1993) An Alternative Method for Significance Testing of Waveform Difference Potentials. *Psychophysiology* 30:518-524.
- Budd J.M., Kisvarday Z.F. (2001) Local Lateral Connectivity of Inhibitory Clutch Cells in Layer 4 of Cat Visual Cortex (Area 17). *Exp Brain Res* 140:245-250.
- Cardoso J.-F., Laheld, B. (1996) Equivariant Adaptive Source Separation. *IEEE Trans. on Signal Processing* 44:3017-3030.
- Comon P. (1994) Independent Component Analysis, a New Concept. *Signal Processing* 36:287-314
- Debener S., Ullsperger M., Siegel M., Fiehler K., von Cramon D.Y., Engel A.K. (2005) Trial-by-Trial Coupling of Concurrent Electroencephalogram and Functional Magnetic Resonance Imaging Identifies the Dynamics of Performance Monitoring. *J Neurosci* 25:11730-11737.
- Delorme A., Makeig S. (2003) EEG Changes Accompanying Learned Regulation of 12-Hz Eeg Activity. *IEEE Trans Neural Syst Rehabil Eng* 11:133-137.
- Delorme A., Makeig, S. (2004) Eeglab: An Open Source Toolbox for Analysis of Single-Trial EEG Dynamics. *Journal of Neuroscience Methods* 134:9-21.
- Freeman W.J. (2004a) Origin, Structure, and Role of Background EEG Activity. Part 1. Analytic Amplitude. *Clin Neurophysiol* 115:2077-2088.
- Freeman W.J. (2004b) Origin, Structure, and Role of Background EEG Activity. Part 2. Analytic Phase. *Clin Neurophysiol* 115:2089-2107.
- Hyvärinen A., Karhunen, J., Oja, E. . (2001) Independent Component Analysis
- Jung T.P., Makeig S., Westerfield M., Townsend J., Courchesne E., Sejnowski T.J. (2001) Analysis and Visualization of Single-Trial Event-Related Potentials. *Hum Brain Mapp* 14:166-185.
- Klopp J., Marinkovic K., Chauvel P., Nenov V., Halgren E. (2000) Early Widespread Cortical Distribution of Coherent Fusiform Face Selective Activity. *Hum Brain Mapp* 11:286-293.
- Lauritzen M. (1994) Pathophysiology of the Migraine Aura. The Spreading Depression Theory. *Brain* 117:199-210.
- Lee T.W., Girolami M., Sejnowski T.J. (1999) Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources. *Neural Comput* 11:417-441.
- Linkenkaer-Hansen K., Nikulin V.V., Palva S., Ilmoniemi R.J., Palva J.M. (2004) Prestimulus Oscillations Enhance Psychophysical Performance in Humans. *J Neurosci* 24:10186-10190.
- Luu P., Tucker D.M., Makeig S. (2004) Frontal Midline Theta and the Error-Related Negativity: Neurophysiological Mechanisms of Action Regulation. *Clin Neurophysiol* 115:1821-1835.
- Makeig S. (1993) Auditory Event-Related Dynamics of the EEG Spectrum and Effects of Exposure to Tones. *Electroencephalogr Clin Neurophysiol* 86:283-293.
- Makeig S., Bell, A.J., Jung, T.P., Sejnowski, T.J. (1996) Independent Component Analysis of Electroencephalographic Data. *Advances in Neural Information Processing Systems* 8:145-151.

- Makeig S., Debener S., Onton J., Delorme A. (2004a) Mining Event-Related Brain Dynamics. *Trends Cogn Sci* 8:204-210.
- Makeig S., Delorme A., Westerfield M., Jung T.P., Townsend J., Courchesne E., Sejnowski T.J. (2004b) Electroencephalographic Brain Dynamics Following Manually Responded Visual Targets. *PLoS Biol* 2:E176.
- Makeig S., Westerfield M., Jung T.P., Covington J., Townsend J., Sejnowski T.J., Courchesne E. (1999) Functionally Independent Components of the Late Positive Event-Related Potential During Visual Spatial Attention. *J Neurosci* 19:2665-2680.
- Makeig S., Westerfield M., Jung T.P., Enghoff S., Townsend J., Courchesne E., Sejnowski T.J. (2002) Dynamic Brain Sources of Visual Evoked Responses. *Science* 295:690-694.
- Massimini M., Huber R., Ferrarelli F., Hill S., Tononi G. (2004) The Sleep Slow Oscillation as a Traveling Wave. *J Neurosci* 24:6862-6870.
- Molgedey L., Schuster, H.G. (1994) Separation of a Mixture of Independent Signals Using Time Delayed Correlations. *Phys Rev Lett.* 72:3634-3637.
- Nunez P., Srinivasan, R. (2005) Electric Fields of the Brain : The Neurophysics of EEG
- Onton J., Delorme A., Makeig S. (2005) Frontal Midline EEG Dynamics During Working Memory. *Neuroimage* 27:341-356.
- Onton J., Makeig,S. (2005) Independent Component Analysis (ICA) Source Locations Vary According to Task Demands. *Organization for Human Brain Mapping Abstracts*
- Scherg M., Von Cramon D. (1985) Two Bilateral Sources of the Late AEP as Identified by a Spatio-Temporal Dipole Model. *Electroencephalogr Clin Neurophysiol* 62:32-44.