ERP Features and EEG Dynamics: An ICA Perspective

Scott Makeig and Julie Onton

smakeig@ucsd.edu

julie@sccn.ucsd.edu

Swartz Center for Computational Neuroscience

Institute for Neural Computation

University of California San Diego

La Jolla, California, USA

An invited chapter for:

S. Luck & E. Kappenman (2009). *Oxford Handbook of Event-Related Potential Components*.

New York, Oxford University Press

Prefatory remarks

Following the advent of averaging computers in the early 1960s, event-related potential (ERP) averaging became the first functional brain imaging method to open a window into human brain processing of first sensory and then cognitive events, and the first to demonstrate statistically reliable differences in this processing depending on the *contextual significance* of these events – or their unexpected absence. Yet the same response averaging methods, now easily performed on any personal computer, may have encouraged the separation of electrophysiological brain research into two camps. For nearly half a century now, researchers mainly in psychology departments have recorded human scalp electroencephalographic (EEG) data and studied the features of human average ERPs time-locked to events and behavior, while researchers mainly in physiology departments measured averaged event-related changes in the number of spikes emitted by single neurons in animals, captured from high-pass filtered local field potential (LFP) recordings from microelectrodes. Since the spatial scales of these phenomena are so different, these two groups have had little to say to one another. In fact, relationships between these quite different average brain response measures can be learned only by studying the spatiotemporal complexities of the whole EEG and LFP signals from which they are respectively extracted, and by understanding not only their average behavior, but the complexities of their moment-to-moment dynamics as well.

Many open questions remain about the nature and variability of these electrophysiological signals, including the functional relationships of their dynamics to behavior and experience. This investigation is now beginning, with much more remaining to be discovered about the distributed macroscopic electromagnetic brain dynamics that allow our brains to support us to optimize our behavior and brain activity *to meet the challenge of each moment.* From this point of view, key open questions for those interested in understanding the nature and origins of average scalp ERPs are how to identify the brain sources of EEG and ERP dynamics, their locations, and their dynamic inter-relationships. To adequately address these questions, new analysis methods are required and are becoming available.

Our research over the past dozen years has convinced us, and an increasing number of other researchers, that using independent component analysis (ICA) to find spatial filters for *information sources* in scalp-recorded EEG (and other) data, combined in particular with trial-by-trial visualization and time/frequency analysis methods, are a powerful approach to identifying the *complex spatiotemporal dynamics* that underlie both ERP averages and the continually unfolding and varying brain field potential phenomena they index. In essence, ICA is a method for training or learning spatial filters that, when applied to data collected from many scalp locations, each focus on one source of information in the data. Characterizing the information content of data rather than its variance (the goal of previous signal processing methods) is a powerful new approach to analysis of complex signals that is becoming ever more

important for data mining of all sorts. Applied to EEG data, ICA tackles 'head on' the major confounding factor that has limited the development of EEG-based brain imaging methods, namely the broad spread of EEG source potentials through the brain, skull, and scalp and the mixing of these signals at each scalp electrode.

I. EEG sources and source projections

In this chapter, we consider the relationship between ongoing EEG activity as recorded in event-related paradigms, and trial averages time locked to some class of experimental events, known as event-related potentials (ERPs). We first discuss the concept of the ERP as averaging potentials generated by spatially coherent activity within a number of cortical EEG source areas as well as non-brain sources typically treated as data artifacts. We use "ERP-image" plotting to visualize variability in EEG dynamics across trials associated with events of interest using an example data set. We then introduce the concept and use of independent component analysis (ICA) to undo the effects of source signal mixing at scalp electrodes and to identify EEG sources contributing to the averaged ERP. We note that "independent components" are brain or non-brain processes more or less active throughout the dataset, and thus represent a quite different use of the term "component" in the title and elsewhere in this volume. After introducing some basic time/frequency measures useful for studying trial-to-trial variability, we take another look at trial-to-trial variability, now focusing on the contributions of selected independent component processes to the recorded scalp signals. We hope the chapter will help the reader interested in event-related EEG analysis to think carefully about trial-to-trial EEG stability and variability. The latter we suggest largely reflects not "ERP noise" but instead the brain's carefully constructed response to the highly individual, complex, and context-defined "challenges" posed by unfolding events.

What is an EEG source?

A fundamental fact about electrophysiological signals recorded at any spatial scale is that they reflect and index emergent partial coherence (in both time and space) of electrophysiological events occurring at smaller scales. Brain electrophysiological signals recorded by relatively large and/or distant electrodes can be viewed as phenomena emerging from the possibly one quadrillion synaptic events that occur in the human brain each second. These events, in turn, arise within the still more vast complexity of brain molecular and sub-molecular dynamics. The synchronies and near-synchronies, in time and space, of synaptic and non-synaptic neural field dynamics precipitate not only neural spikes, but also other intracellular and extra-cellular field phenomena – both those measurable only at *near field* (e.g., within the range of a neural arbor), and those recorded only at *far field* (in particular, as electrically far from the brain as the human scalp). The emergence of spatiotemporal field synchrony or near-synchrony across an area of cortex is conceptually akin to the emergence of a galaxy in the plasma of space. Both are

spontaneously emergent dynamic phenomena large enough to be detected and measured at a distance – via EEG electrodes and powerful telescopes, respectively.

The emergence of synchronous or near-synchronous local field activity across some portion of the cortical mantle requires that cells in the synchronized cortical area be physically coupled in some manner. A basic fact of cortical connectivity is that cortico-cortical connections between cells are highly weighted toward local (e.g., shorter than 0.5-mm) connections, particularly those coupling nearby inhibitory cells whose fast gap-junction connections support the spread of near-synchronous field dynamics through local cortical areas (Murre and Sturdy, 1995; Stettler et al., 2002). Also important for sustaining rhythmic EEG activity are thalamocortical connections that are predominantly (though not exclusively) organized in a radial, one-to-one manner (Frost and Caviness, 1980). EEG is therefore likely to arise as emergent *mesoscopic* patterns (Freeman, 2000) of local field synchrony or near-synchrony in compact thalamocortical networks. Potentials arising from vertical field gradients associated with pyramidal cells arrayed orthogonal to the cortical surface produce the local field potentials recorded on the cortical surface (Luck, 2005; Nunez, 2005). Synchronous (or near-synchronous) field activity across a cortical patch produces the far-field potentials that are conveyed by volume conduction to scalp electrodes. Both scalp and direct cortical recordings agree that in nearly all cognitive states, such locally coherent field activities arise within many parts of human cortex, often with distinctive dynamic signatures in different areas. Direct observations in animals report that cortical EEG signals are indeed associated with sub-centimeter sized cortical patches whose spatial patterns resemble ‘phase cones’ (like ‘pond ripples,’ (Freeman and Barrie, 2000)) or repeatedly spreading ‘avalanche’ events (Beggs and Plenz, 2003), though more adequate multi-resolution recording and modeling are needed to better define their spatiotemporal geometry and dynamics.

In this chapter, we will use the term *EEG source* to mean a compact cortical patch (or occasionally, connected patches) within which temporally coherent (or partially-coherent) local field activity emerges, thereby producing a far-field potential contributing appreciably to the EEG signals recorded on the scalp. While non-cortical brain sources may contribute to the recorded EEG (as for example, auditory brain stem potentials), their contributions are typically small compared to cortical potentials and we will therefore assume for this chapter that resolvable EEG signals of interest are of cortical origin. We will use the phrase *source activity* to refer to the varying far-field potential arising within an EEG source area and volume-conducted to the scalp electrodes. Recorded EEG signals are then, in this view, the sum of EEG source activities, contributions of non-brain sources such as scalp muscle, eye movement, and cardiac artifacts, plus (ideally small) electrode and environmental noise.

Note that the activity contributed by a cortical source to the recorded EEG typically does *not* comprise all the local field activity within the cortical source domain, since potentials

recorded with small cortical electrodes at different points in a cortical source domain may only be *weakly* coherent with the far-field activity that is partially coherent across the domain and is therefore *not* projected to the scalp electrodes. That is, only the portion of the local field activity in a source domain that is synchronous across the domain will contribute appreciably to the net source potentials recorded by scalp electrodes. Thus, cortical electrophysiology is by its nature multiscale, its properties differ depending on the size of the recording electrodes and their distance from the source areas in ways that are currently far from adequately observed or modeled. Scalp EEG recordings predominantly capture the sum of locally coherent source activities within a number of cortical source domains, plus non-brain artifact signals.

Roles of EEG source activities

The primary function of our brain is to organize and control our behavior 'in the moment' so as to optimize its outcome. For many neurobiologists, field potential recordings have been considered of possible interest at best only as passive, indirect, and quite poor statistical indices of changes in neural spike rates, their primary measure of interest. In fact, however, the variations in electrical potential recorded by EEG or LFP electrodes better reflect variations in concurrent dendritic synaptic input to neurons, input that may or may not provoke action potentials. Action potential generation is provoked by receipt of sufficient dendritic input within a brief (several-millisecond) time window. The emergence of synchronized local field potentials across a cortical area may therefore reflect changes in occurrence of *joint* spiking events across groups of associated neurons in that area. Some recent experimental results also suggest that local field potentials may also *actively* affect spike timing and degree of synchrony between neurons within a partially-synchronized source domain, biasing their joint spike timing towards (or away from) concentration into brief, potent volleys (Voronin et al., 1999; Francis et al., 2003; Radman et al., 2006). By this means, small statistical adjustments in joint spike timing of neurons with common axonal targets effected by spatially synchronized local field potentials might be associated with large changes in effective neural communication, and thence with behavior and behavioral outcome (Fries et al., 2007).

According to recent reports, the timing and phase of extracellular fields may also enhance or weaken the effects of concurrent input on future cell and areal responsivity, by affecting the amount of long-term synaptic potentiation (LTP) produced by that input (Dan and Poo, 2004). Thus, locally synchronous (or near-synchronous) field activity arising within compact cortical source areas may not only weakly index neural dynamics on spatial scales larger than a single neuron, but may also play a more direct *and active* role in organizing the distributed brain dynamics that support experience, behavior, and changes in psychophysiological state. The spatiotemporal dynamics of cortical field synchrony, and their relationship to neural spike timing have still been relatively little studied (Logothetis et al., 2001; Bollimunta et al., 2008) and,

contrary to standard presumption, there is likely much more to be discovered about relationships between extra-cellular fields and intracellular potentials in living brains.

Spatial source variability

The concept that an EEG source represents the emergence of synchronous field activity across a cortical patch is undoubtedly a simplification of the actual more complex, multi-scale dynamics that produce *near*-synchronous activity within an EEG source domain. Although the concept that EEG is produced by synchronous field activity in small cortical domains is supported qualitatively by fMRI results that are generally dominated by roughly cm-scale or smaller pockets of enhanced activity, and by a few reports of direct field potential grid recordings (Bullock, 1983; Freeman and Barrie, 2000), when cortical activity in animals is viewed at a smaller (sub-mm) scale using optical imaging, smaller-scale moving waves of electromagnetic activity are observed (Arieli et al., 1995). However, simple calculation of the phase difference between the edges and the center of a 'pond ripple' pattern at EEG frequencies, based on the estimated (cm^2-scale) domain sizes and observed traveling wave velocities (1-2 m/s), suggests that the spatial wavelength of the radiating 'ripples' is considerably larger than the size of the domain, meaning that the topographic scalp projection of a cortical 'phase cone' is *close* to that of totally synchronous activity across the patch, as in our simplified EEG source model[1].

However, larger-scale travelling or meandering waves at slow (1-3 Hz) delta or infraslow (<1 Hz) EEG frequencies have also been observed in epilepsy, sleep, and migraine (Massimini et al., 2004), and (near 12-Hz) sleep spindles may also 'wander' over cortex in concert with spatially varying activity in coupled regions of the thalamic reticular nucleus (Rulkov et al., 2004). Sufficiently detailed recordings from high-density, multi-resolution arrays are not yet available to allow models of the relationship between these moving activity patterns and apparently stable EEG source dynamics in waking life.

Temporal source variability

A hallmark of EEG is that its temporal dynamics are highly non-stationary and exhibit continuous changes on all time scales (Linkenkaer-Hansen et al., 2001). Changing EEG dynamics index changes in and between local synchronies that are driven or affected by a variety of mechanisms including sensory information as well as broadly projecting brainstem-based arousal or 'value' systems identified by their central neurotransmitters – dopamine, acetylcholine, serotonin, neurepinephrine, etc. These 'neuromodulatory' systems based in brainstem areas project to widespread cortical areas and are very likely an important source of temporal variability in the spatiotemporal coherence that produces far-field signals recorded at the scalp, variability that gives *flexibility and individuality* to our distributed brain responses so as to respond most appropriately to the particular *challenge of the moment.* Ranganath and

Rainer (Ranganath and Rainer, 2003) have reviewed what is known and still unknown about these systems and their interactions with cortical field potentials.

Volume conduction and source mixing

Experimental neurobiology suggests that the spontaneous emergence of partial coherence of complex rhythmic temporal patterns of local field activity across compact cortical 'phase cone' or 'avalanche' domains a few mm or larger in diameter produce the scalp EEG and therefore ERP signals. The differences between distinct parts of neurons within the partially-synchronized and nearly-aligned pyramidal cell domain sum coherently both in local recordings and as measured at any distance after passing by *volume conduction* through intervening conductive media including cortical grey and white matter, cerebral-spinal fluid (CSF), skull, and skin (Akalin-Acar and Gencer, 2004; Gencer and Akalin-Acar, 2005). The very broad cortical source field patterns (each generally resembling the double-lobed pattern iron filings take when placed around a bar magnet) are attenuated by the partial resistance of these media, and their propagation patterns are spatially distorted at tissue boundaries where conductance changes. The broadly projecting, spatially distorted, and severely attenuated signals are then summed within conductive EEG electrodes attached to the scalp (Nunez, 1977)[2].

Forward and inverse modeling

A first question for EEG/ERP researchers, therefore, is or should be how to separate the recorded EEG activities recorded at all the scalp channels into a set of activities originating within different spatial source domains. Finding the appropriate spatial filters is, unfortunately, technically difficult, and were any arbitrary 3-D arrangement of source configurations physiologically possible, any number of them could be found that would produce the same scalp potentials (Grave de Peralta-Menendez and Gonzalez-Andino, 1998). A biophysical solution to this so-called inverse problem must begin with construction of a forward head model specifying (1) where in the brain the electromagnetic sources may be expected to appear, and (2) in which orientations, and (3) how their electromagnetic fields propagate through the subject's head to the recording electrodes (Akalin-Acar and Gencer, 2004). Fortunately, the well-grounded assumption that the brain EEG sources are cortical source patches whose field patterns are oriented near-perpendicular to the local cortical surface allows more physiologically plausible estimates since these assumptions allow a fair MR image-derived model of the shape, location, and local orientation of subject cortex (Fischl et al., 2004). Constructing individualized cortical models for EEG analysis requires extensive computation and expensive MR head images, and thus the process is still rarely carried out for routine EEG/ERP experiments. While adequate head models are needed to develop EEG into a 3-D functional brain imaging modality, different analysis goals may require various degrees of anatomic accuracy, and use of standard head models may suffice for many analysis purposes.

Given an accurate forward head model, the inverse problem is still underdetermined if multiple sources contribute to the observed scalp potential distribution whose sources are to be determined. In contrast, the solution is much simpler, and its answer better determined, if the scalp maps whose source projections they sum are *simple maps* representing the activity of only one source. Both EEG and MEG researchers have long attempted to consider scalp maps of amplitude peaks in average ERPs to be simple maps. That is, they attempt to use ERP averaging, a purely temporal filtering method, as a means of, in effect, *spatial filtering*, attempting to eliminate from each channel the projections of EEG sources not directly affected by the time-locking events.

In many cases, taking the difference between two average ERPs time-locked to related sets of experimental events may further restrict the number of strongly contributing brain source areas. Unfortunately, trial averaging or differencing is rarely completely effective for this purpose, since meaningful sensory (as well as purely cognitive) events rapidly perturb the statistics of many cortical EEG source areas as well as subcortical areas (Halgren et al., 1998; Schroeder et al., 1998). Therefore, except for very early sensory brainstem and cortical potentials, ERP maps sum activities that arise, typically, in many source domains. This means that scalp maps of ERP or ERP difference-wave peaks are only rarely *simple maps* representing potentials projected to the scalp from a single source. To optimally estimate the source distribution responsible for EEG or ERP data, it is desirable to find a better way to isolate simple maps representing the projection of single sources contributing to the data, a subject we will return to in Section III.

II. ERP trial averaging and trial variability

Figure 3.1A shows a single-subject ERP for all 238 scalp channels averaged over 500 data epochs time-locked to onsets (a latency 0) of an infrequently presented visual target disc in a visuospatial selective attention task (Makeig et al., 1999b; Makeig et al., 1999a; Makeig et al., 2002; Makeig et al., 2004b). ERP traces for all 238 channels are overlaid on the same plot axis. Interpolated scalp maps show the ERP scalp distribution at four indicated latencies. The bottom panel shows the ERP average of the same 500 epochs, but now time-locked to the subject's button press in each trial. In both panels, the data were averaged after removing artifacts produced by eye movements, eye blinks, electrocardiographic (ECG) activity, and electromyographic (EMG) from scalp and neck muscles using independent component analysis (ICA), as explained in Section III[3,4]. The late positive complex (LPC or 'P300') feature of this averaged response to the infrequent attended visual targets is attenuated by the highpass filtering above 1 Hz applied to the data; it is most evident in the motor response-locked data average (Fig. 3.1B). We will use these data through this chapter to explore relations between the average ERPs as shown in this figure and the single EEG data epochs that were averaged to produce them.

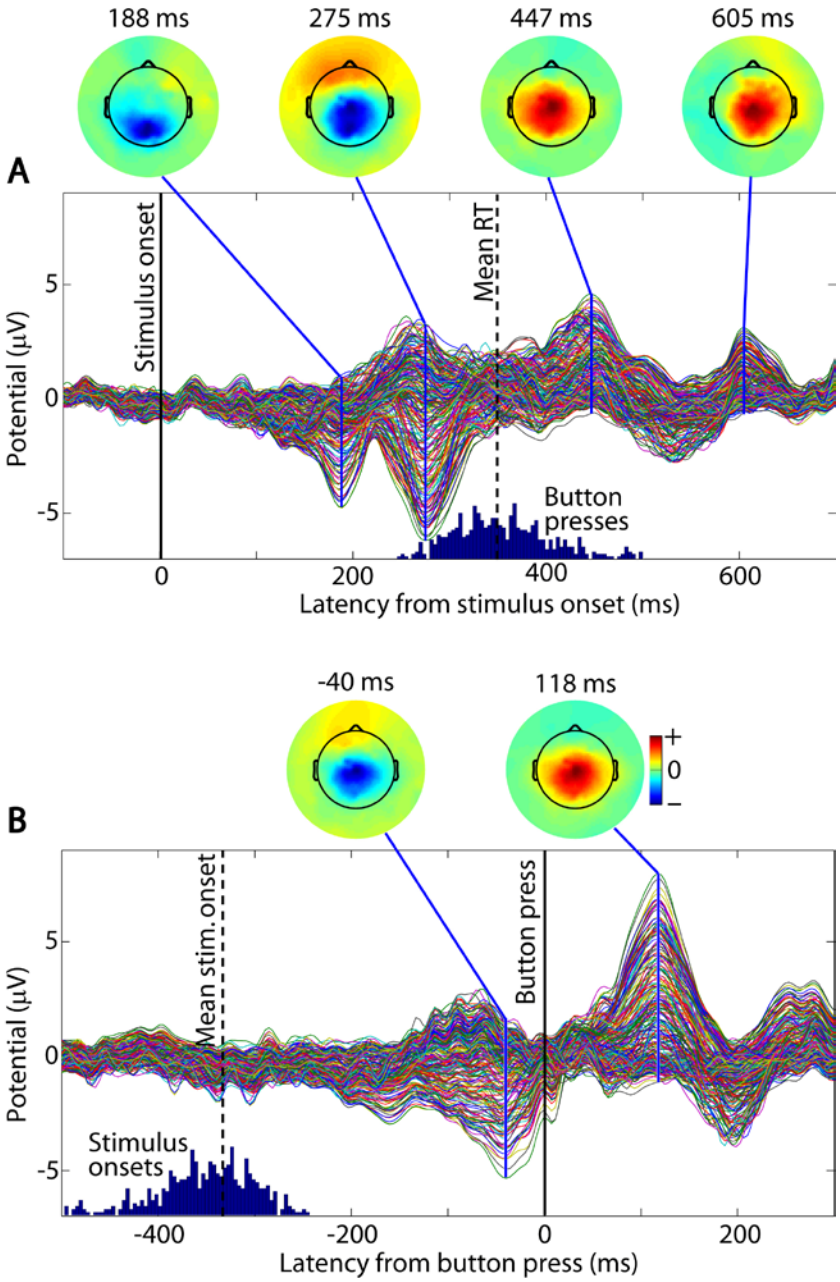

Figure 3.1. *ERP traces from each of 238 scalp channels averaged over 500 EEG epochs in a single subject, time locked to (A) continually anticipated but infrequently presented visual target stimuli in a visuospatial selective attention task, and (B) immediately following speeded subject button presses cued by target presentation4 Solid lines cutting vertically through the ERP channel trace bundles lead to cartoon heads showing the interpolated scalp potential distribution at the moment indicated.*

The trial-averaging model

Over the past near half-century, the predominant method for reducing the complexity of event-related EEG data collected in sensory and cognitive paradigms is to form event-related potential (ERP) averages of trial records time-locked to sets of experimental events assumed by the experimenter to generate the same or essentially similar brain responses. To gain a realistic understanding of the features of ERP trial averages and their relationship to the underlying brain

dynamics, it is important to understand both the strengths and limitations of ERP averaging. The physiological model underlying ERP averaging is that cortical processing of sensory (or other) event information follows a fixed spatiotemporal sequence of source activities, and that this processing produces a fixed sequence of deviations in scalp potentials whose distribution reflects the locations of their cortical generators. However, these traces of the cortical processing sequence are obscured in single response epochs by typically much larger ongoing EEG activities generated in many brain areas, as well as artifacts generated in non-brain structures. Crucially, these ongoing activities are assumed to be unaffected by the time-locking events of interest.

Thus, in the ERP averaging model, EEG epochs are assumed to sum (1) a temporally consistent *event-related* activity sequence ("the evoked response"), plus (2) *event-unrelated* ongoing or spontaneous EEG activities (not contributing to the ERP). Under these circumstances, averaging a sufficient number of event-locked epochs subtracts, cancels, or spatially filters out the unrelated brain activities, leaving a single average response epoch dominated by the consistent event-related activity sequence, recorded on the scalp as the flowing ERP field 'movie.' The amplitudes of EEG (or other non-brain) activities unaffected by the time-locking events that remain in the average will be approximately $1/N^{1/2}$ of the amplitude of those activities in the single trials, where N is the number of epochs averaged. Thus, achieving a faithful representation of the actual (and typically relatively small) evoked response sequence usually requires averaging a relatively large number of event-related epochs known or assumed to contain the same time-locked ERP sequence.

To understand how ERP averaging leads to a reduction in event-*un*related EEG activity, we first need to define the *phase* of an EEG source signal. The simplest sense of the term might be the sign or polarity of the recorded potential at a given time point, either positive or negative in relation to some baseline potential (typically established by averaging the potentials recorded in some period of the recording assumed to be unaffected by the events of interest). A more specific meaning for the 'phase' of a signal at some time point and frequency is as the phase of a best-fitting brief, tapered sinusoid at the given frequency centered on that time point. Thus the phase of an EEG source or scalp signal, at any given time point and frequency, is defined by the relation of its value at that time point to its values in an enclosing window of time points. Note that at a given time point a signal has a different phase at each frequency. Also, since EEG signals are relatively smooth, EEG phase differences between neighboring time points cannot vary freely but must change smoothly.[5]

ERP averaging can remove the contributions of those source activities unrelated to the time-locking events by means of *phase cancellation*, which works as follows. If a given source signal is unaffected by the time-locking events, and if the timing of the experimental events is not based on the ongoing EEG signals, their phase at each latency and frequency will differ

randomly across trials. Mathematically, the sum of random-phase signals at a given frequency tends to become smaller and smaller (at that frequency) as the number of summed trials increases. We can see this most easily by considering the signs of the signals (+ or -) instead of their phases. If the signs of a set of signal epoch values at a given latency are random, then in the average of those epochs the positive-phase and negative-phase values in different epochs will partially cancel each other, and the magnitude of the average epoch at that latency will be smaller than the average of the same values were they all of the same sign.

Similarly, if at a given analysis frequency (say for example 9 Hz), the single-trial EEG signals have a random phase distribution when measured in a time window centered at some latency (say, 200 ms following the time-locking event), then the vectors that can be used to represent their amplitudes and phases at that frequency will be more or less evenly distributed around the phase circle, and the expected length of the vector average of these vectors will become smaller as the number of trial vectors averaged increases, provided that the exact timing of experimental events cannot be predicted by the brain. On average, the length of the average phase vector will decrease as the square root of the number of trials averaged. In this way, trial averaging filters out all features of the data that are not wholly or at least partially *phase-locked* to the time-locking events at any frequency and latency.

It is important to understand how event-related phase-locking and time-locking differ. For example, imagine a set of EEG trial epochs, time-locked to a particular type of event, that each contain a burst of 10-Hz alpha band activity centered 500 ms after the time-locking event. Further, imagine that these alpha bursts, while undeniably *time*-locked to the experimental events of interest, may exhibit any phase at 500 ms (ascending, descending, etc.). These bursts, therefore, are not *phase*-locked to the events. An ERP average of enough such epochs would therefore contain little trace of 10-Hz activity at 500 ms, even thought this is a striking feature of the single-trial data. This is because trial averaging filters out all activity that is not *both time-locked and phase-locked* to the time-locking event (see discussion of time- and phase-locking below). Thus, scalp ERPs do *not* capture all of the consistent event-related dynamics in the averaged EEG epochs, but only those dynamic processes that affect the phase distribution of their signals at some analysis frequencies and trial latencies. We will consider this question again in Section IV (see also Chapter 2, this volume).

However, if a given brain source contributes a fixed activity sequence to a given scalp channel signal in every trial, at each analysis frequency and epoch latency the phase of its contributions to the scalp signals will be consistent across trials, and the ERP average at that scalp channel will contain *all* of that source's activity sequence, without diminution. If the phase of its single-trial activity at a given frequency and latency is variable and only weakly consistent, relative to a true random phase process, then the source will contribute only weakly to the scalp

ERP. If the source activity has truly random phase at the given frequency and latency, then its ERP contribution will be minimal and further decreasing as more trials are averaged.

If all the EEG sources that project to a scalp channel have fixed evoked activity sequences, then their collective contribution to the channel in each trial will be the sum of all their source activities, and the average ERP at that channel will be the average of the summed source activities at each trial latency. Thus source mixing that occurs at the scalp electrode could decrease (or increase) the apparent ERP magnitude through the same process of phase cancellation. But again, only those signals that are phase-locked to the time-locking events from trial to trial will be retained in the average. And conversely, for activity at a given frequency and latency to be removed from the signal by averaging, only the signal *phase* in the single trials need be random.

Limitations of event-related averaging

The mean of any distribution is simply one statistical measure of the distribution – a statistic that, if provided apart from other statistics, may be informative and/or misleading. For example, telling a New Guinean unacquainted with Americans that the *average* adult American height is 5’6” (1.68 m) might give him or her an adequate concept of the distribution of American adult heights – assuming the shapes and the widths of the (near-normal) height *distributions* in the two cultures are not dissimilar. However, sending Martian scientists the arithmetically equally correct information that the average human is half male and half female might well engender quite incorrect ideas about human biology and society. The problem here is that human sexual physiology has not one but two quite distinct modes (female and male), information that is not captured in or conveyed by the average. Thus, the *average* of a distribution may or may not in itself provide or suggest a useful and realistic model of the underlying distribution or its features. This may be even more problematic for time series averages that sum disparate activities of many distinct brain and non-brain sources whose detailed features are of primary interest, including their spatial and temporal trial-to-trial variability.

Spatial variability of event-related activity

Note how the scalp topographies of the ERPs in both panels of Figure 3.1 differ slightly before and after the button press. If ERP spatial variations were generated within or directly under the scalp itself, such changes would reflect potential changes occurring directly below the most strongly affected electrodes. Such an interpretation, while having naïve appeal, is however contrary to the anatomic and biophysical facts about volume-conducted cortical field potentials that actually produce scalp EEG signals, as summarized above. The very broad ‘point-spread’ pattern of potentials propagating out by volume conduction from each cortical source area means

that each source contributes to some extent to the signals recorded at nearly all of the scalp electrodes, and contributes appreciably to many of them.

Further, if each source area is spatially fixed (or nearly so), by itself it cannot produce a moving topographic pattern of field activity on the scalp – it can only produce proportional and *simultaneous* changes across all the electrodes in its projection pattern. Therefore, changes in the scalp map of average ERP data (as in Fig. 3.1) must reflect sums of time-varying potentials projected in the broad and highly-overlapping scalp patterns from several spatially-fixed EEG source activities, each contributing to the ERP in spatially and temporally overlapping time windows. This view is compatible with fMRI results showing that cortical activations and deactivations mainly occur within compact cortical domains – though direct high-resolution, multi-scale observations of electrocortical activity that could constitute 'ground truth' evidence for this assumption are not yet available.

Two main points here are, first, that basic biophysical knowledge is *not* consistent with an interpretation of ERP potentials as exclusively (or even principally) reflecting activity generated directly below each electrode, a fact it is easy to lose sight of when focusing on the details of single-channel ERP or EEG waveforms. Second, although when animated, changes in high-density ERP time series appear to flow across the scalp, features of most ERPs recorded in cognitive experiments are much more likely produced by sums of time-varying activities of relatively small, spatially-fixed cortical generator domains, each with a broad 'point-spread' pattern of projection to the scalp surface. Looking ahead a bit: Although the exact size distribution of these domains is not yet known, fairly precise indications of their centers can be obtained by methods that spatially filter the scalp EEG data to focus on single sources (see Section III).

ERPs as spatial filters?

Early ERP analysis attempted to deal with the difficulty in interpretation posed by volume conduction and scalp mixing by assuming that event-related averaging provides sufficient spatial filtering of the many source signals reaching the scalp electrodes so that activity from only *one* affected source area contributes to each ERP amplitude peak. That is, some early ERP researchers hoped that the sequence of peaks comprising ERP waveforms would each spatially filter out all activities not generated in a single cortical area.

However, subsequent research clearly suggested that soon after sensory signals arrive in cortex following meaningful events, *multiple* EEG sources begin to contribute to ERP averages. It has now been shown that in animals coordination of activities between early visual areas (Grinvald et al., 1994) and between primary visual and auditory cortex (Foxe and Schroeder, 2005) begins as early as 30 ms after stimulus onsets, and invasive recordings from epileptic

patients for clinical purposes show that by at most 150 ms after presentation of meaningful visual stimuli, the phase statistics of local field processes are altered in many parts of the brain, both cortical and subcortical (Klopp et al., 2000). Thus, the somewhat different scalp distributions in the single-channel ERP scalp field maps of Fig. 3.1 in fact represent differently weighted mixtures of mean event-related activities, time-locked to subject button presses, from several to many cortical sources having broad and strongly overlapping scalp projections. Also, direct cortical recordings from both animals and humans show that, through activity cycles through thalamocortical and other network connections involving significant delays, single cortical areas often produce multi-peaked complexes instead of single response peaks in response to single stimulus presentation events (Swadlow and Gusev, 2000), adding to the spatial overlapping of source activities in ERP waveforms.

Thus, response averaging is not an efficient method for spatial filtering of event-related EEG data since it typically does not produce a sequence of *simple* maps each reflecting the projection to the scalp of a single cortical source. Computing the ERP 'difference wave' at each scalp channel between ERPs in two contrasting conditions may more effectively isolate condition differences in the activity of a single generator or set of generators, although in general the effectiveness of this approach cannot be guaranteed. Finding more effective methods for spatial filtering of EEG data into EEG source activities is therefore of urgent importance to human electrophysiology research.

Temporal variability of event-related activity

Another problem with modeling event-related EEG dynamics based on ERP averages alone is that ERP averaging collapses and thus conceals both the *orderly* (event-, task-, or context-related) as well as the *disorderly* (event-, task-, and context-unrelated) trial-to-trial variability in the recorded EEG scalp and source signals – giving no way for the user to determine the relative proportions or types of these two classes of signal variations that are present in the single trials. A model of brain activity built solely on an ERP average of scalp activity in event-related epochs must fail to include many aspects of the brain processes that produce it. If the single-trial EEG epochs each sum activity and time-varying activities from multiple spatial sources whose dynamics are tightly linked to multiple task or context-related factors, focusing solely on their average may discourage study of their orderly trial-to-trial variations.

Averaging itself simplifies not only the spatial pattern, but also the temporal patterns of the signals averaged, retaining only that portion of the signals (typically, a small portion) that are both time-locked and phase-locked to the time-locking signals (as explained above and in Chapter 2, this volume). Far too often trial-to-trial variability of EEG signals is simply dismissed by researchers (either explicitly or implicitly), as representing irrelevant brain 'noise' – without sufficient consideration or evidence for this assumption. To consider this point more carefully,

let us examine some aspects of the trial-to-trial temporal variability in scalp EEG activity before and after target presentations in the selective visual attention task session for which Fig. 3.1 showed the ERP averages.

A first step towards understanding trial-by-trial variability in EEG epochs time-locked to some class of time-locking events is to find useful ways to visualize their variability. Jung and Makeig have developed a method of sorting the order of single trials by some criterion and then plotting them as horizontal color-coded lines in a rectangular image they called the ERP image (Makeig et al., 1999b; Jung et al., 2001). In ERP-image plots, single-trial traces drawn as horizontal colored lines, with color (instead of vertical placement) encoding potential. The colored lines can be fused into a rectangular image and smoothed, if desired, with a vertical moving average to bring out trends. Crucially, the order in which the trials are sorted (bottom to top) need not follow the actual time order of the trials, but can be based on any other criterion. Unlike, the average ERP, which is fixed, the ERP images resulting from different trial sorting orders can differ dramatically from each other, each bringing out one or more aspects of trial-to-trial variability in the data.

Panel A of Figure 3.2 uses this method to visualize nearly 500 single trials time locked to visual target stimulus presentations at a scalp site near the vertex (Cz) – the same stimulus-locked trials averaged to form the ERP shown in Fig. 3.1A, with the (vertical) order of trials here sorted in their original time order. For easier visualization of event-related patterns across trials, in each panel of Fig. 3.2 we have smoothed the single-trial data (vertically) with a 20-trial moving window. The solid vertical line in panel A marks the onset of the visual target, the vertical dashed line the subject's median button press latency. The average ERP (shown below the panel) appears to have only a small and temporally diffuse late positive complex (LPC or, for historical reasons, "P300") feature, here peaking near 400 ms after stimulus onset.

Panel B shows the same data trials, but here sorted by the latency of the subject's button press (dashed trace), again smoothing with a 20-trial moving average. We now see that in most trials a (much more distinct) LPC follows the subject's button press by about 120 ms, with LPC amplitude smaller in long-RT trials (near the top of the ERP-image panel). This panel shows which features of the post-stimulus ERP are primarily time-locked to the stimulus onset itself (e.g., the negative (blue) pre-response N2 peak), and which to the subject behavioral response (the following LPC and two sparser ensuing positive wave fronts). The trial-to-trial latency variability of the LPC cannot be labeled as irrelevant trial-to-trial 'noise,' and cannot be deduced from the stimulus-locked trial average (blue trace below panels A and B) in which trial-to-trial variability time-locked to the button press rather than the stimulus onset is temporally "smeared out."

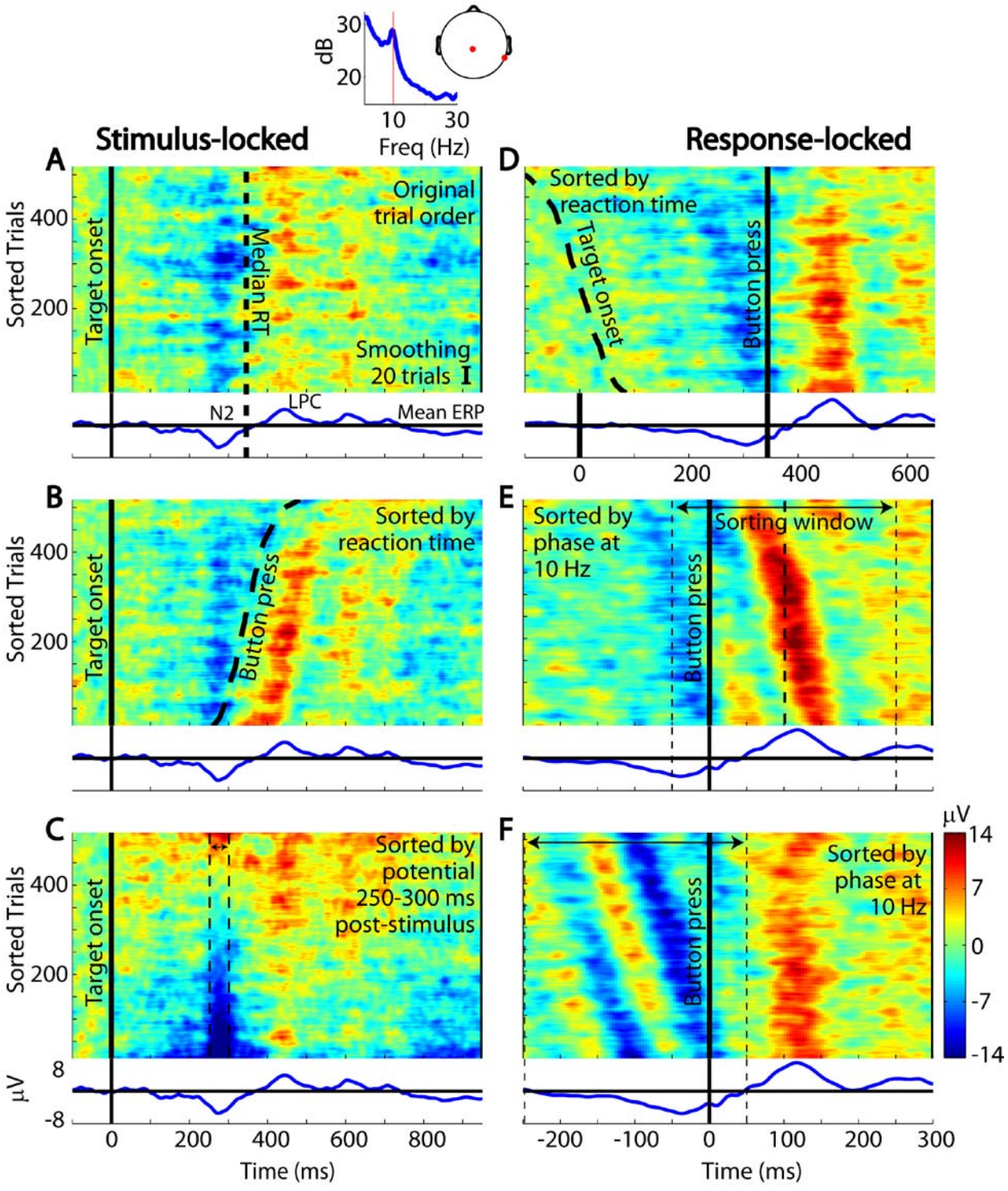

Figure 3.2. *Six ERP-image plots of nearly 500 visual target response epochs (same data as in Fig. 3.1) at a scalp electrode near the vertex (referred to right mastoid). In each panel, trial potentials are sorted (from bottom to top) in the order indicated, then smoothed (vertically) with a 20-trial moving window, and finally color-coded (see color bar on lower right). In left panels (A-C), the trials are aligned to the moment of stimulus onset in each trial. In right panels (D-F),*

they are aligned to the moment of the subject button-press response. The trace below each image shows the ERP average of the trials. Stimulus-locked ERP peaks N2 and LPC (late positive complex) are labeled in A. The trial sorting criteria are indicated in each ERP-image panel. Horizontal arrows show the value (C) and phase (E,F) sorting windows. The ERP images illustrate the wide variety of trial-to-trial differences in the data.

Panel C shows again the same trials, here sorted instead by mean potential in the N2 response ERP latency window indicated by the dashed lines. We see that only in about the bottom half of the trials is the mean potential in this window actually negative. But these include about a (bottom) third of the trials in which the negativity is relatively large, thereby "outweighing" the contributions of the trials in which the single-trial value is positive, thus producing a negative peak in the average ERP (again shown below the ERP-image panel).

The curving post-RT positive (red-orange) wave fronts in the ERP-image plot in panel B reveal that the evoked LPC can be more accurately represented by an ERP average of the same data epochs time-aligned to the subject button press rather than to stimulus onset – as in panel D. In particular, the average motor response-aligned ERP (below panel D) better reflects the abrupt onset, slope, and duration of the post-motor LPC in the single trials than the stimulus-aligned ERP (below panel A). As in panel B, panel D shows that the post-button press positivity is generally stronger in (lowermost) trials with short button press latencies, and is weak or even absent in (uppermost) trials with long response latencies. Also, the slightly curving (red) response column in panel D suggests that the time locking of the LPC peak to the button press is only relatively constant, the LPC appearing slightly wider and centered slightly later in shortest-latency trials, relative to other trials.

These scalp data, measuring the potential difference between an "active" electrode near the vertex (Cz) and a "reference" electrode on the right mastoid (see cartoon head above panel A), actually sum broadly spreading projections of field activities projecting by volume conduction from several to many cortical sources. The mean power spectrum for these trials (above panel A) contains a strong alpha-band peak at 10 Hz. Panel E shows the same response-aligned trials as in D, now sorted according to the best-fitting 10-Hz phase in the indicated three-cycle (300-ms) wide trial sorting window centered on the LPC. In this ERP-image view of the data, the central positive (red) LPC peak of the activity in the single trials forms a red curving wave front near the center of the sorting window. The peak latency difference (from bottommost to topmost trials in the image), clearly visible in panel E, is hidden in panel D using a different trial-sorting order.

Panel F shows again the same response-aligned trial data, but now sorted by alpha phase in the *pre-response* (but post-stimulus) period between the dotted lines. Here, we see that

ongoing alpha activity that is random phase before the button press (as reflected in the perfectly diagonal positive and negative wave fronts in the sorting window), but the ensuing LPC positivity peaking in this view about 120 ms after the button press is nearly vertical in this view, and therefore appears to be independent of the phase of the the preceding alpha band activity (which might well have a different set of sources than those generating the LPC). Note that the LPC positivity is again wider than the tighter peak obtained in the alpha-sorted view of the same data in panel E, though panels (D- F) all visualize aspects of the same data and have the same motor response-locked ERP.

What conclusions can we draw from these six quite different ways of plotting the same data? Note that each panel highlights a different way of separating the ERP trial average into a set of single-trial activities, followed by moving-average smoothing to bring out trial-to-trial trends in the trial-sorted data images. Each panel represents, in a way, a decomposition of the single-trial data highlighting some foreground features and smoothing other types of trial-to-trial variability to form a kind of background "noise" (as it were). Which of these decompositions in this sense – if any of them – is the most physiologically realistic or 'correct' decomposition? Arithmetically, there is no difference between them; in each case the same trial data, aligned as in panels A-C or D-F, have the same ERP average, no matter how the trials are sorted – just as 5 pennies are equally the sum of 3+1+1 or 1+3+1 coin subsets.

The standard model underlying ERP analysis is that the EEG data essentially sum, (1) contributions to the ERP occurring in each trial, and (2) other (undefined) variable activities unaffected by the time-locking events and therefore not contributing to the ERP. Is this standard ("ERP + background") decomposition of the single-trial signals more physiologically "realistic," in any sense, than the different implied "decompositions" of the data into the quite different features that are highlighted (plus those obscured) in these six quite different ERP-image panels? In particular – do any of these views parse the data into physiologically distinct source contributions?

From these ERP-image representations of the data, we can at least see some ways in which an average ERP of the trials is simply one statistical measure of them, a measure that does not reveal their orderly (though complex) temporal variations or multiple spatial sources. In each trial, the subject performed the same task – attempting to produce quick button press responses to target stimuli while withholding responses to non-targets. Panel B clarifies at least one aspect of trial-to-trial EEG variability directly related to subject behavior (e.g., their manual response latency). What other trial-to-trial differences, either in stimulus features (here, in target location), and/or in the trial *context* (here, the history of preceding stimuli and manual responses) may have altered the nature of the 'challenge' posed to the subject's brain – and thus affected aspects of trial-to-trial EEG variations? Neither the ERP averages nor these six ERP-image panels answer these questions. There are many other possible decompositions into putative underlying EEG

source activities that examination of its trial average ERP cannot rule in or out as reflecting physiologically valid distinctions among spatial sources and their event-related temporal patterns.

At the least, these panels illustrate the fact that trial-to-trial variations in EEG data are not simply “EEG noise.” Rather, they include several types of orderly trial-to-trial variability linked to several EEG source and/or behavioral and task parameters. Panels C-F, in particular, pose an interesting question. Is the LPC activity at this channel, time-aligned to the subject button presses dominated by positive-phase alpha activity (as panel E suggests) or by a broader fixed-latency, central LPC peak (as in panel F)? Or perhaps, by both types of activity arising in different cortical domains and both projecting to the vertex to differing relative extents in different trials?

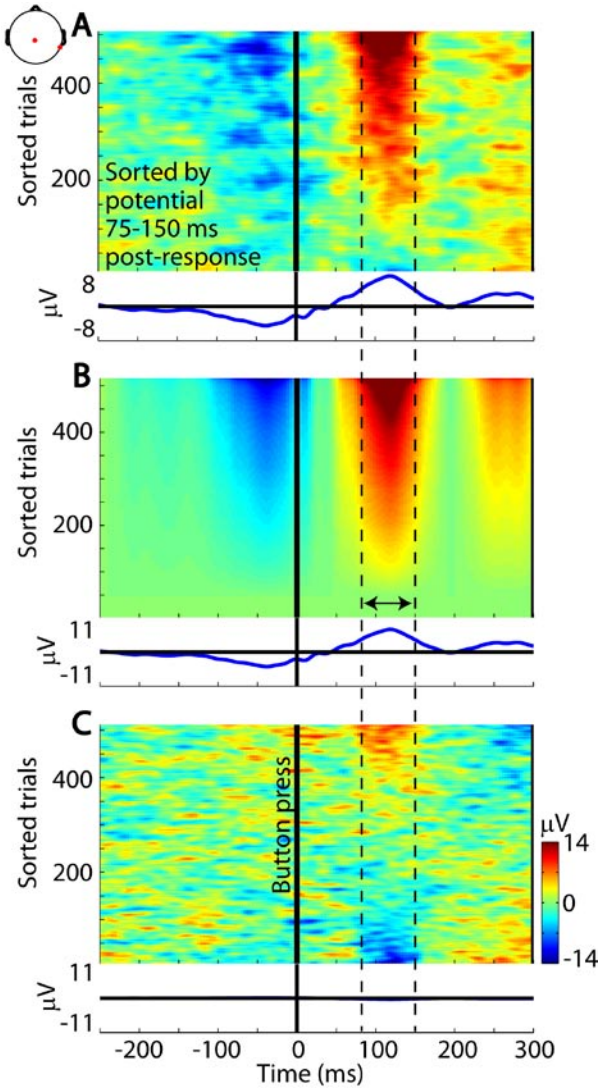

Figure 3.3. *ERP-image plots separating the ERP image in panel A, in which the trials are sorted by mean potential between the two dashed lines surrounding the post-response ERP peak, into the sum of (B) an ERP-image of the best-fitting (non-negative) mean-ERP contribution to each trial, and (C) the remaining unexplained portion of each trial. As in Fig. 3.2, the mean of the single-trial data in each panel is shown below the panel. The ERP has been largely (though not completely) removed from the lower panel data. This decomposition is compatible with the assumption that the single-trial data sum a single ERP response of variable amplitude (B) plus unrelated EEG activity (C). However, many EEG sources may make separate contributions to the data and to the ERP average. Regression of the whole average ERP trace on the single data trials does not take into account the spatiotemporal variability of the independent sources of trial-to-trial variation, for example the partial sorting by value remaining within the sorting window and near 300 ms. (Vertical smoothing window width 20 trials).*

A simple ERP model of the trial-to-trial variability in EEG data represents the non-ERP portion of the data as summing contributions of brain source activities that are unaffected by the

time-locking events. In an extreme version of this model, the amplitude of the ERP might be assumed to be identical in every trial, with any trial differences reflecting additional task-irrelevant EEG (or non-brain artifact) activity. But Figure 3.2 suggests that such a strict version of the model is unlikely to adequately capture the trial-to-trial variability in the event-related dynamics in these trials.

Panel A of Figure 3.3 shows the same trial data as in Fig. 3.2, but now sorted by mean potential in the indicated post-response LPC data window. Panels B and C visualize an ERP-based decomposition of the data in panel A, all three panels using the same trial sorting order (sorting by LPC amplitude). Panel B shows the estimated contribution of the mean ERP to each trial, as determined by finding the best least-square fit of the mean ERP to each single-trial epoch but not allowing negative ERP trial weights in the few lowest trials. Panel C shows the remaining (non-ERP) data for each trial. Thus, the sum of the values in panels B and C are the whole-trial data as shown in panel A. At least two latency windows in Panel C (90-150 ms and 200-300 ms) exhibit systematic differences from a random trial distribution, indicating trial-to-trial variability of potentials in those windows is partially independent of the overall amount of ERP-like activity in the trial. But is this attempted decomposition of the single-channel data (in panel A) into ERP and non-ERP data portions (in panels B and C) a physiologically valid separation between completely motor response-locked (ERP) and motor response-independent (non-ERP) cortical source processes in the data?

How can we begin to find answers to this question? To model the nature of the highly variable signals occurring during these recorded data epochs, ideally we should first find a way to separate the whole scalp EEG data into a set of functionally and physiologically distinct source activities.

III. Separating EEG sources using Independent Component Analysis (ICA)

In 1995 the first author and colleagues at Salk Institute (La Jolla, CA) performed the first decomposition of multi-channel EEG data into its maximally independent components (Makeig, 1996) using a then-new and elegant ‘infomax’ algorithm (Bell and Sejnowski, 1995) that followed insights, a few years earlier, that weighted sums of independent source signals should be separable ‘blindly’ into the individual source signals *without* advance knowledge of the nature of the source processes, as had been thought necessary (Jutten and Herault, 1991; Comon, 1994). The prototypical example of this problem is the ‘cocktail party problem’ in which an array of microphones records mixtures of the voices of several people talking at once at a cocktail party. Individually, the recordings sound like indecipherable ‘cocktail party noise.’ The blind source separation problem is to determine how to combine the recorded signals so as to separate out each speaker’s voice, “blind” to any knowledge of the nature or properties of individual sources.

At root, the insight allowing the solution to this problem is that the individual speakers' voices are the only sources of independent *information* in the recorded data. By adapting randomly weighted sums of the recorded signals in such a way as to make the weighted-sum signals more and more *temporally independent* of each other, the unmixing process must finally arrive at producing the individual voice signals. In signal processing terms, the joint microphone data is separated into its maximally independent signal *components*, which must be the original voice sources since they are the only possible independent sources in the recorded mixtures.

When the process proceeds without relying on any knowledge of the qualities of the individual source signals, the unmixing process is called *blind source separation*. Using temporal independence to separate the source signals is a form of blind source separation called *independent component analysis* (ICA). Intuitively, two time series are maximally independent when their waveforms are maximally distinct from each other. More technically, independent time series have no *mutual information*, meaning that knowing the value of one process at a given time gives no information at all (even partial or probabilistic) about the concurrent value of the other process.

Reading about the ICA solution to the cocktail problem in the influential paper of Bell and Sejnowski before its publication in 1995, the first author suspected that the same approach should be applicable to EEG data. The results of the first EEG decomposition (Makeig, 1996) were highly promising, and subsequent work over the next dozen years or more has confirmed the ability of ICA to identify both temporally and functionally independent source signals in multi-channel EEG or other electrophysiological data. ICA in effect creates a set of spatial filters. Each *independent component* (IC) filter cancels out the contributions of all but one of the distinct source signals that contribute to the multi-channel data. ICA can be thought of as a method of data information-driven spatial filtering or beamforming (Iyer et al., 1990) that learns spatial filters that each focus on a distinct physical source of EEG signals – separating out distinct brain generator processes as well as different non-brain (artifact) signals.

More formally, a linear decomposition of a (channels by time points) signal matrix is its representation as any weighted sum of component signal matrices of the same size (the same numbers of channels and time points). Figure 3.4 schematically visualizes the simple matrix algebraic formulation of the linear signal decomposition used in ICA. The scalp data channel signals are formed into a matrix (top center). ICA decomposition finds an unmixing matrix (W) which, when multiplied by the data matrix, decomposes the data (downwards pointing arrow) into a matrix of independent component (IC) signals, called the IC activations (lower right), of the same size as the input scalp data. Multiplying the IC activations matrix by the inverse of the unmixing matrix W (lower middle) reconstitutes or back-projects the original scalp data channels (upwards pointing arrow).

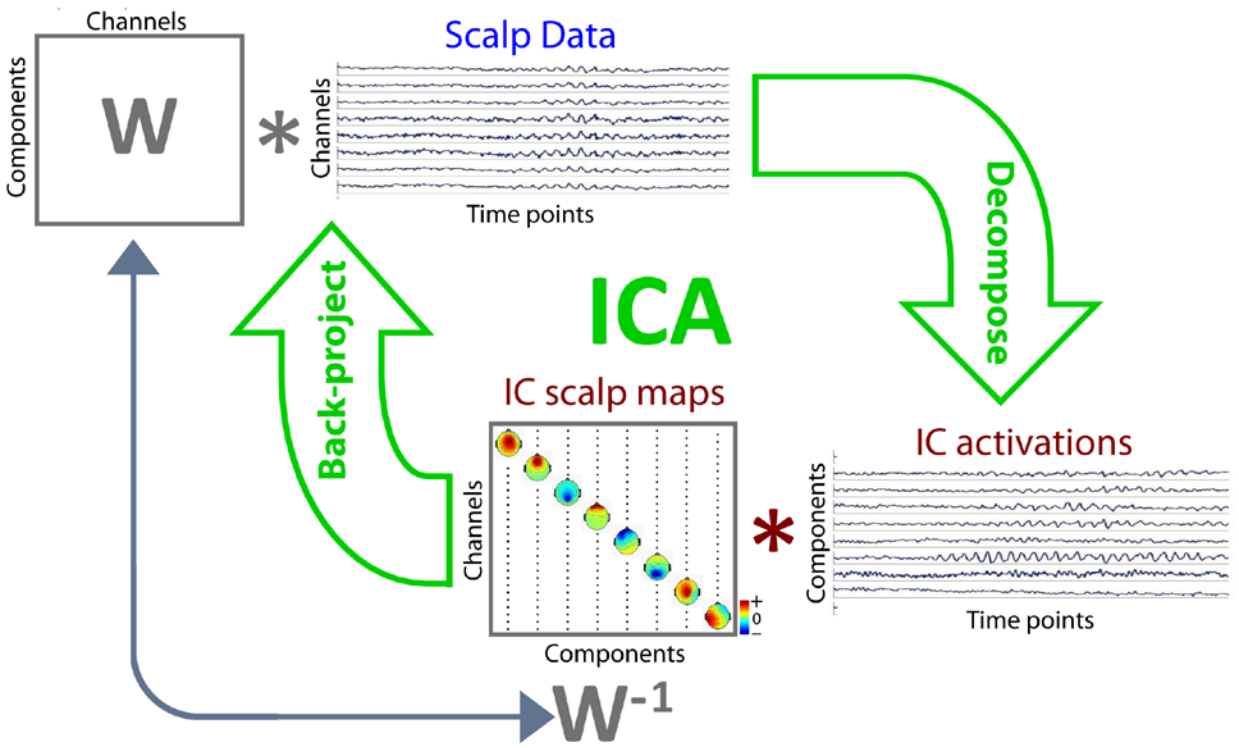

Figure 3.4. *Schematic flowchart for Independent Component Analysis (ICA) data decomposition and back-projection. ICA applied to a matrix of EEG scalp data (upper middle) finds an 'unmixing' matrix of weights (W, upper left) that, when multiplied by the (channels by time points) Scalp data matrix, gives a matrix of independent component (IC) activities or activations (lower right). This is the process of ICA decomposition (downward arrow) of the data into maximally temporally independent processes, each with its distinct time series and scalp map. The process of back-projection (upward arrow) recaptures the original scalp data by multiplying the IC activations matrix (lower right) by the matrix of independent component (IC) scalp maps (lower center) whose columns give the relative projection weights from each component to each scalp channel. The IC scalp map or 'mixing' matrix (W-1, lower center) is the inverse of the 'mixing matrix' (W, upper left). In simple matrix algebra form, if the indicated scalp data matrix is X and the component activations matrix U, then algebraically WX = U and X = W-1U. Here, W is a matrix of spatial filters learned by ICA from the EEG scalp data that, when applied to the data finds the activity projections of the underlying EEG source processes, and the IC activations (lower right). This general schematic holds for all "complete" linear decomposition methods returning as many components as there are data channels.*

The inverse of the unmixing matrix, W^{-1}, is the component *mixing* matrix (lower center) whose columns give the relative strengths and polarities of the projections of one component source signal to each of the scalp channels. In the figure, the values in the columns of the mixing

matrix are color coded and interpolated onto cartoon heads to visualize the topographic projection patterns or *scalp maps* associated with each of the sources.

The component scalp maps found by ICA decomposition are not constrained to have any particular relationship to each other (unlike in PCA decomposition). They may be highly (though not perfectly) correlated. They may also have any (simple or complex) spatial pattern, although in practice scalp maps for components truly accounting for a distinct source process (contributing independent temporal information to the data) must reflect the relative projections of the source process to the individual scalp channels. Thus, a source comprised of spatially coherent local field activity across a cortical patch must have a scalp map that matches that of a single tiny battery (dipole) placed in "the electrical center of mass" (as it were) of the source patch and called its equivalent dipole (Scherg, 1990). However, IC scalp maps may, again have any form, depending on the pattern of the source projection to the scalp electrodes, and on the degree of dominance of a (maximally) independent component by a single signal source.

What is an independent component?

Before going further, we must first discuss a basic terminological confound. ICA (like PCA and other linear decompositions) uses the term 'component' to mean something quite different than its use elsewhere in this volume (including its title), i.e. as a contraction of the term "ERP component feature" – some identifiable feature in an ERP waveform (typically associated with a single peak). By broader definition, an ERP component may be any functionally distinct feature or portion of an ERP waveform, i.e. a feature with a functionally distinct relationship to experimental parameters, and/or an ERP feature generated in a particular brain region (see Chapter 1, this volume).

In this chapter, however, we will use the term component process (or *component* for short) to mean some portion of an entire multi-channel recorded data set separated from the remaining recorded data by linear decomposition. To minimize confusion, we will substitute a terminological equivalent, '*ERP peak'* or '*ERP peak feature,'* for the more usual term 'ERP component.' Thus again, in this chapter *components* will not refer to ERP peaks or other features but to *EEG source processes,* each accounting for some portion of the continuous EEG activity (at all time points) forming a multi-channel data set. Each data set component naturally then also accounts for some portion (large, small, or negligible) of any ERP average of epochs drawn from the data set.

As shown in Fig. 3.4, an independent component (process) of an EEG data set (or IC for short) comprises *both* a fixed scalp map and a time series that gives its relative amplitude (or "activation") and polarity (positive or negative) at each time point. The scalp map shows the relative weights or projection strengths (and polarities) of the projection from the component

process to each electrode location. The component activation time series gives the relative amplitude and polarity of the component's activity at each time point. Because we define an EEG source as being spatially stable, a component scalp map remains constant over time. The back-projection of each component process to each scalp channel is the product of the component activation time series with the scalp map weight for that channel. The IC back-projection to all the channels is the portion or component of the scalp data (at all channels) contributed by the component process. The original channel signals are the sums of the back-projected activities of all the independent components. That is, the scalp data are the collection of all the summed back-projections of all the independent components to all the channels.

In simple matrix algebra form, if the scalp data matrix is X and the component activations matrix U, then algebraically WX = U, where W is the unmixing matrix of spatial filters learned by ICA decomposition of the EEG scalp data. This equation simply says that applying spatial filters W to the data X (by simple matrix multiplication) gives the activation time courses of the independent component processes. The converse process that reconstitutes the data from the components is, algebraically, $X = W^{-1}U$ where W^{-1} (the matrix inverse of W) is the component mixing matrix. These same equations can be used to represent any linear decomposition method, though other methods may use different names for the matrices.

Independent components of EEG data

It is important to understand that each scalp EEG recording channel is itself in effect a spatially filtered measure of the varying scalp potential field, recording only the time-varying potential difference between two scalp electrodes, the so-called "active" electrode and one or more "reference" electrodes – at which brain and non-brain signals are potentially just as "active" as the so-called "active" electrode. ICA attempts to replace these scalp channel electrode-difference filters with IC filters using other linear electrode combinations chosen so as to pass individual EEG source signals while rejecting all other sources. The degree of source fidelity ICA can achieve depends on the number of data channels versus the number of active sources – as well as on the length and quality of the data.

Before considering in detail the assumptions underlying ICA and giving heuristic guidelines for how to apply it, let us first show model examples of independent EEG components or ICs. ICs of EEG data can be roughly separated into three types: ICs accounting for brain and non-brain (artifact) processes, respectively, and small ICs whose maps and activities appear noisy and are poorly if at all replicated from session to session. This last category can be considered a 'noisy' part of the EEG signals that ICA is not able to resolve into components dominated by a single source (although not every small IC fits this description). Between these three IC categories, there may be ICs in 'grey areas' whose assignment to one of these three

categories is difficult. Here, let us first consider ICs clearly accounting for activity from particular non-brain (artifact) sources.

Independent non-brain component processes: Noise or signals?

ICA characteristically separates several important classes of non-brain EEG artifact activity from the rest of the EEG signal into separate sources including eye blinks, eye movement potentials, electromyographic (EMG) and electrocardiographic (ECG) signals, line noise, and single-channel noise (Jung et al., 2000b; Jung et al., 2000a). This important benefit of ICA decomposition of EEG data was apparent from the first attempt to apply it (Makeig, 1996). ICA has thus found initial use in many EEG laboratories simply as a method for removing eye blinks and other artifacts from data. For data sets heavily contaminated by eye blinks or other artifacts, for instance data collected from young children, the ability to analyze brain activity in data trials including eye movement artifacts can mean the difference between analyzing and rejecting the subject data altogether.

Unlike regression-based methods for artifact removal, ICA artifact separation allows artifact subtraction (often called artifact 'correction') without requiring a separate ('pure') reference channel for each signal. In practice, regression methods risk eliminating brain signals that also project to the (impure) reference channel (e.g., frontal brain sources also project to an 'electrooculographic (EOG) channel' near the eyes). Figure 3.5 shows scalp maps, spectra, and ERP-image plots (above their trial-average ERPs) for four typical independent artifact source components separated by ICA from the visual selective attention task session considered in earlier figures. The highly distinct activity features separated from the data by ICA make the qualitative implications of temporal independence clear. The recovered component waveforms are the most *temporally distinct* portions of the recorded data. Separation by ICA of non-brain source processes allow detailed analysis of the separated source process time courses. Note, for example, that the subject refrained from blinking for over a second following target stimulus presentations (Fig. 3.5A).

When, as here, the electrode montage includes both head and neck sites, scalp maps of head muscle components exhibit a characteristic polarity reversal at the insertion point of the muscle into the skull, with the direction of the dipole following the direction of the muscle fibers. Note the scalp map associated with an EMG component signal (Fig. 3.5B). Note also the abrupt changes in EMG activity level in the component ERP-image plot near trials 180, 300, and 370 (lower left), a common occurrence for scalp muscle activity recorded during experiments in which the subject is sitting comfortably while attempting to minimize head and eye movements. These marked changes in activity level of this muscle were likely neither willfully controlled nor noted by the subject. By so clearly separating non-brain processes contributing to EEG data, ICA

allows these activities to be analyzed as concurrently recorded biological (or other) signals instead of simply being rejected as non-brain “artifacts.”

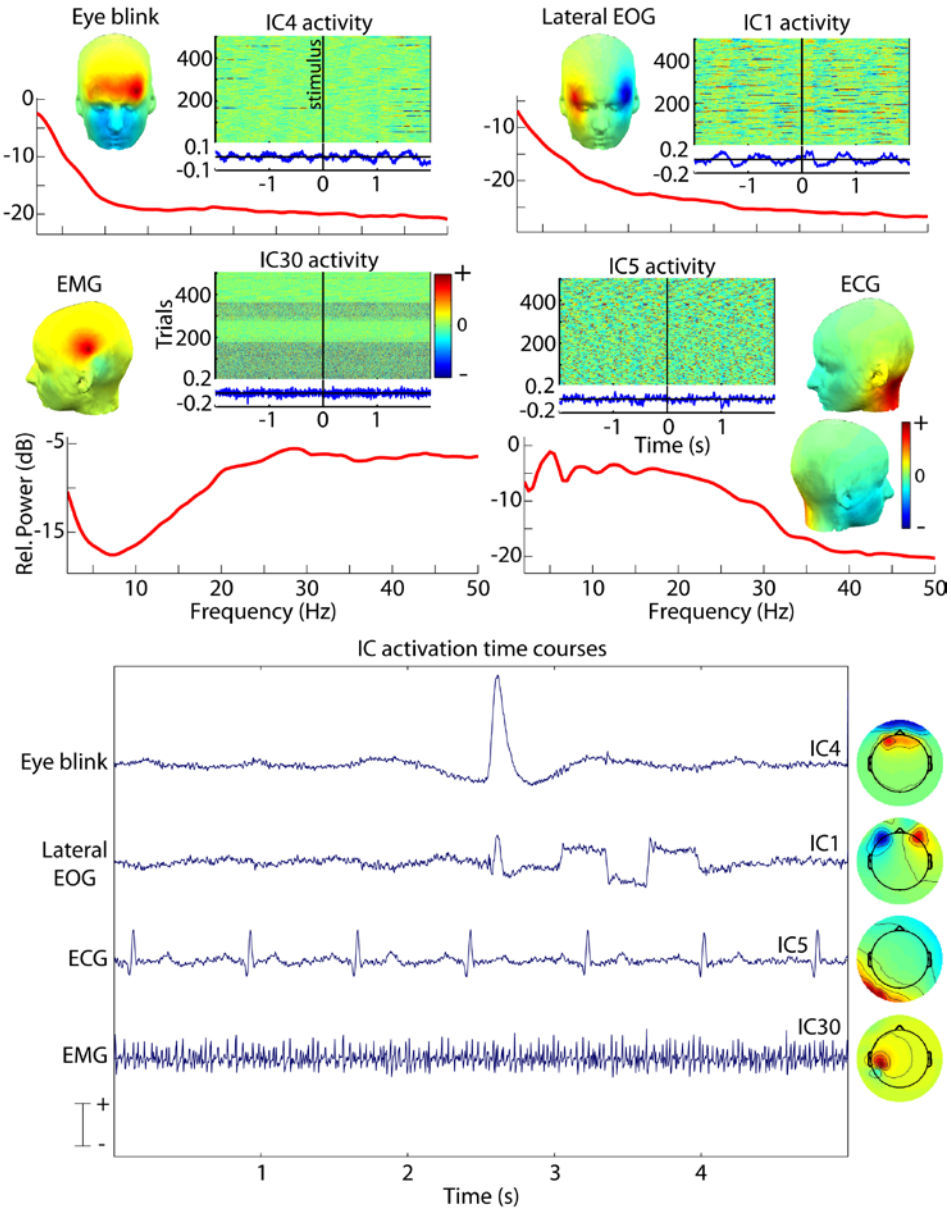

Figure 3.5. *Typical component properties of four non-brain independent component processes accounting respectively for eye blinks, lateral eye movements, left post-auricular electromyographic (EMG) activity, and electrocardiographic (ECG) activity, from the 238-channel EEG recording studied in Figures 3.2 and 3.3. Upper panels show the interpolated component scalp map, activity ERP image, average ERP (below ERP image), and mean power spectrum. Lower panel shows the maximally independent activities of the four processes during a five-second period. Note the characteristic activity elements, also seen in the ERP-image representations.*

Spatially stereotyped versus non-stereotyped artifacts

It is important, however, to understand the distinction between spatially stereotyped non-brain signal sources, such as eye blinks and scalp muscle activities that always project with the same topographic pattern to the scalp channels, and non-stereotyped non-brain signal phenomena that have varying spatial scalp projections. Consider, for example, the case of an unruly subject who vigorously scratches his or her scalp for a second or two during the EEG recording. This quickly produces a series of some hundreds of EEG data points (i.e., EEG scalp maps) whose topographic patterns do not match each other nor appear elsewhere in the data. The one-time-only appearance of each of these scalp maps is in effect temporally independent of all other data sources, possibly hugely increasing the number of 'temporally independent' sources ICA needs to separate into a finite number of component activities. Further, during this period the changes in electrode contact with the skin may alter the spatial pattern with which the other brain and non-brain signal sources project to the electrode array, violating the ICA assumption that these spatial projection patterns are stable throughout the data.

Thus, including a stretch of data dominated by this or other *spatially non-stereotyped (SNS) artifact* in the data given to ICA for decomposition can only limit the success of the decomposition at identifying physiologically distinct EEG source processes. Such SNS periods may be identified by eye while scrolling through the data, by use of simple heuristics (Delorme et al., 2007a), by similar observations of a preliminary ICA decomposition of the whole data, or even automatically during ICA training by computing the probability of each data point fitting the current ICA model and rejecting highly improbable data points from further training.

Cases intermediate between spatially stereotyped and non-stereotyped artifacts are phenomena with spatially stereotyped but non-stationary scalp patterns, for example slow blinks (Onton and Makeig, 2006), ballistocardiographic (BCG) cardiac artifacts recorded within a high-field magnetic resonance scanner (Debener et al., 2008), and slow waves in sleep (Massimini et al., 2004). In such cases, ICA typically finds a small set of *maximally* independent components that each captures one or more time periods of the repeating activity pattern, thereby separating it from other source activities.

Independent brain component processes

Using ICA solely to remove non-brain source processes, while valuable, does not exploit the power of ICA to separate the activities of individual *brain* sources that contribute to the scalp data. Some might ask that since no part of the brain acts *wholly* independently from the rest of the brain, how can ICA decomposition extract physiologically meaningful component signals? The answer to this question is that ICA finds the *maximally* independent components for a data set, even if traces of dependence remain between them. This dependence might be transient – for

example the occasional strong similarity of occasional evoked activity in otherwise maximally independent left and right lateral occipital processes produced by central visual stimulus presentation (Makeig et al., 2002). Or the dependence might be limited, for example only reflected in weakly coherent, low amplitude, high frequency activity.

Even in cases in which two ICs have remnant mutual dependence, i.e. when their joint activities could be said to form "a dependent two-dimensional subspace" of the data, ICA should still separate the activity of this subspace from the activities of other, single independent component processes. For example, a moving scalp artifact produced by slow eye blinks might be separated into two or more independent components, each accounting for one phase of the moving blink potential. In this case, the spatiotemporally overlapping component activities would differ from one another, not allowing their parsing as a single IC. Though not completely independent of each other, the time courses of the partially dependent ICs might still be independent of any other IC time course in the data and sufficiently different from each other to require more than one IC (Meinecke et al., 2002).

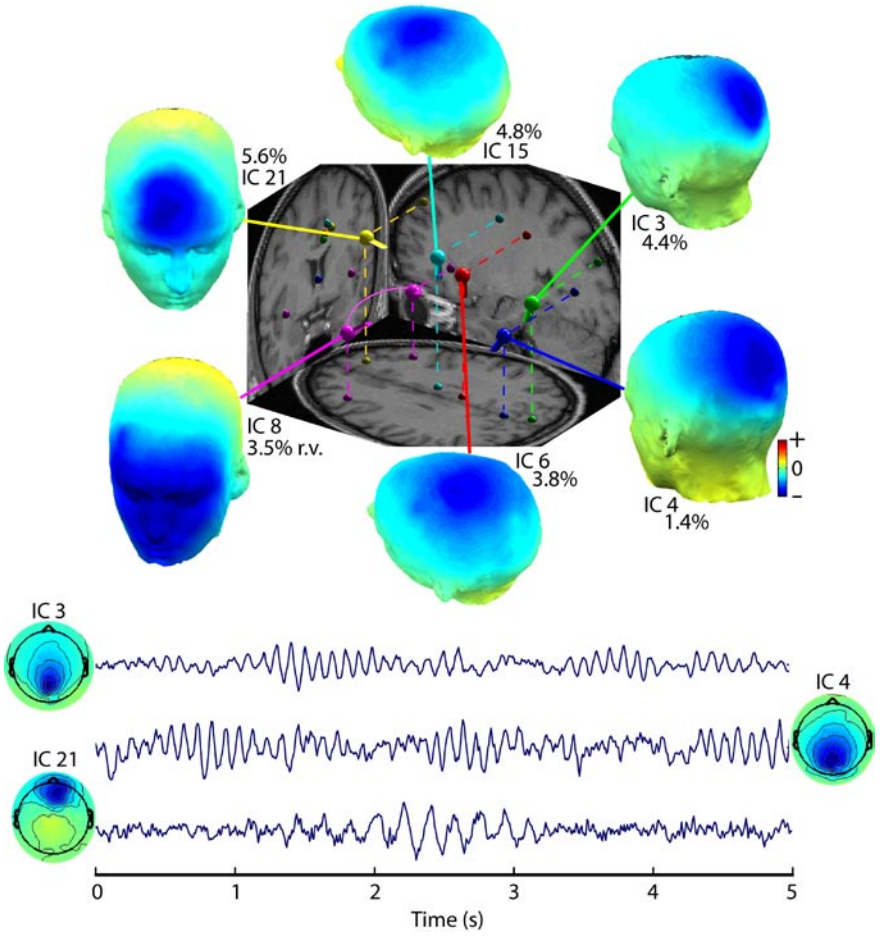

Figure 3.6. *Equivalent dipoles for six maximally independent brain source components. Near each IC index (ranked in order of variance contributed to the scalp data), the residual variance (r.v.) of the equivalent dipole model across the 238-channel component scalp map is indicated, based on fitting the measured 3-D electrode locations to an individualized three-shell boundary element method (BEM) head model. All the residual variances are low (< 6%), indicating that the component maps are compatible with an origin in a single (or, IC8, in dual bilateral cortical patches). The equivalent dipoles (center) are likely situated somewhat closer to the cortical surface than the locations of the equivalent model dipoles. The five-second activation periods shown in the lower panel give representative (not concurrent) examples of bursts of*

frontal midline (IC21) theta (and higher frequency) activity, and posterior (IC3, IC4) alpha source activities for three of the components.

Again, the computed IC single *equivalent dipole* locations cannot represent the spatial distribution of the cortical generator domains. Instead, they represent the computed positions (in the BEM head model) of vanishingly small oriented dipoles whose scalp projection patterns match most closely the actual IC component maps (across all electrodes). In general, an equivalent dipole for a cortical patch source is typically deeper in the brain than the cortical patch itself (Scherg, 1990). Recent advances in distributed inverse source localization methods suggest that it may soon prove possible to estimate, using subject MR images, the patch (or patches) of subject cortex that most likely constitute each IC source domain. Such a goal is likely not reachable by the alternate strategy of first computing an ERP average and then finding inverse source distributions of one or more ERP scalp maps, since the ERP scalp map at any point in time is typically a weighted mixture of contributions from several cortical source areas.

ICA assumptions

As illustrated in Figures 3.5 and 3.6 (above), ICA decomposition has proven to be highly successful for studying EEG data – Why? An important part of the answer must be that there is an approximate fit, at least, between ICA assumptions and the physiological nature of EEG sources themselves (Makeig et al., 2004a; Onton and Makeig, 2006; Onton et al., 2006). Basically, ICA "blindly" separates the scalp data given it into component processes whose spatial and temporal properties are not known in advance, based on the following five assumptions:

(1) That the component source locations (and thereby their topographic projection patterns to the scalp sensors) are fixed throughout the data.
(2) That the projected component source activities are summed linearly at the sensors.
(3) That there are no differential delays involved in projecting the source signals to the different sensors.
(4) That the probability distributions of the individual component source activity values are not precisely Gaussian.
(5) That the component source activity waveforms are (maximally) temporally independent of one another.

The last ('independence') assumption (5) can be translated informally as saying that the component source activity time patterns are *maximally distinct* from one another, an assumption partially supported by intracranial recordings from neighboring areas (Destexhe et al., 1999). More technically, a set of signals are temporally independent, in the sense used for ICA, if knowing the activity (μV) values of any subset of the signals at a given time point gives no clue about the activity values of any subset of remaining sources at the same time point. Thus each

component source signal is, in a particular sense, *an independent source of information* in the data, contributing a temporal pattern not in any way determinable from the values (at the same time point) of the other component source signals.

The spread of information-based signal processing into nearly every signal processing application area in the last decade (Jutten and Karhunen, 2004) derives primarily from the basic interest of investigators in all research areas in identifying the sources of information that contribute to their multi-dimensional data. However, the value of ICA for decomposing any signal is determined by the degree to which the ICA assumptions fit the manner in which the data are actually generated and recorded. For EEG signals, the assumptions of simple summation at the electrodes (2) and lack of differential delay (3) are met precisely. The non-Gaussian distribution assumption (4) is plausible for EEG sources generated by nonlinear cortical dynamics as well as for non-brain artifact sources including cardiac signals, line noise, muscle signals, eye blinks and eye movements, etc. that are not themselves sums of smaller *un*correlated signals.

As mentioned earlier, the ICA spatial source stationarity assumption (1) is consistent with indirect evidence from fMRI and other brain imaging methods, and the independence assumption (5) is consistent with the very sparse long-range cortico-cortical coupling and the predominantly radial thalamocortical connectivity profile. However, both these assumptions have limitations. In particular, optical dye recordings in animals of local field potentials at the millimeter and smaller scale reveal moving wave patterns (Arieli et al., 1995), and comparison of ICA solutions across a group of subjects participating in the same task suggests that the spatially stable EEG sources separated from the data by ICA depend in part on the task the subject is performing (Onton, 2005). Thus, further research is needed on methods of identifying spatial lability in EEG source data (Anemuller et al., 2003) and for identifying changes in the spatial distribution of the sources as subject task, strategy, or preoccupation changes (Lee, 2000). However, given a hypothetical switch between two sites of EEG signal generation as the subject alternately performs two tasks, ICA should in theory return two components each showing the task-related activity only during one performance condition.

Dual-dipole IC processes

From the viewpoint of ICA decomposition, an EEG source is nothing more than an independent time course of information in the data, whatever its scalp projection pattern. The scalp projections (and hence, scalp maps) of ICA components are thus constrained only by the projection patterns of the actual physical sources of the data. Cortical (or other) source signals arising in separate cortical patches may be partially or wholly synchronized if the separate patches are physically linked by dense white matter tracts (such as corpus callosum), or are identically stimulated. In this case, ICA decomposition will (rightly) return a component

summing the scalp projections of the two (physical) source patches. For example, a single IC typically accounts for eye blink artifacts from the two eyes, whose synchronized small upward movements during the blink induce electrical activity accounted for by an equivalent dipole located in each eye.

Similarly, ICA may return one or more brain components whose scalp maps sum the projections of two equivalent dipoles, usually with bilaterally near symmetrical locations and scalp projections, compatible with patches connected by corpus callosum. Theoretically, cortical activities on either end of any dense white matter tract might synchronize and their activities be combined into a single IC, though to us this has not yet been conclusively demonstrated. It makes no sense to say that ICA fails to separate “sources” in this case, unless one for example (re)defines the term “EEG source” to mean activity in a single cortical patch. However, in practice the number of dual-dipolar” ICs is relatively small (except for most ICs accounting for eye movements, thankfully).

ICA ambiguity

Discussions of the polarity and amplitude ambiguity inherent in IC activations in some early ICA papers have been confusing to some readers. In fact, this ambiguity is present only when the IC activations and IC scalp maps are considered separately. We might say that the sign and scaling of the (back-projected) component in the data is split (arbitrarily) between its activation and scalp map. Since $-1 \times -1 = 1$, inverting the signs of both an IC activation and its scalp map will not change their product, or the back-projection of the IC into the original data, which will retain its original polarity. These ambiguities should be kept in mind when examining or comparing IC activations or scalp maps.

However, the μV scaling of the back-projected IC scalp activity is precisely the product of the scalp map values with the activation time series. Thus, ICA decomposition does not lose this information, as is sometimes mistakenly suggested. Also, while IC potentials at the cortical surface are also proportional to the IC activation, accurate source location and electrical head models are needed to determine the actual IC strength on or in the cortex, since this depends on the resistance between scalp and cortex, which in turn varies across heads and source locations.

Note that, ICA does not itself sort the components into any fixed order. Thus, decompositions of similar data, even data from the first and second halves of the same recording session, are not guaranteed to return ICs in the same order. ICs from different data sets need to be compared with each other using one or more measures of their time courses and/or scalp maps, for example their power spectra and equivalent dipole locations.

Number of data channels

How many data channels should be used for ICA filtering to be successful? The most computationally efficient and robust ICA methods, such as infomax ICA, neither increase nor decrease the dimensionality of the data – they find the same number of components as there are data channels and are therefore called "complete" decomposition methods[6]. How many independent sources contribute to EEG data? It is highly likely that there are always more (brain and non-brain) sources with distinct (e.g., near-independent) time courses and unique scalp maps than any possible number of recording channels, since synchronized cortical field activity likely occurs, at least transiently, at more than one spatial scale, and to some extent uncorrelated noise is generated at each of the electrodes. Most such source activities will be small to negligible, but their presence guarantees that the number of degrees of freedom of the recorded data will never be less than the number of data channels.

Data contributions from numbers of sources beyond the number of available component degrees of freedom (i.e., beyond the number of data channels) will be mixed into some or all of the resulting components, thereby adding a kind of 'noise' to the results of the decomposition. The noise inherent to ICA decomposition of EEG data is evidenced by the indeterminate scalp maps of the very smallest ICs in a high-dimensional data decomposition, ICs that may not prove stable under repeated decomposition and whose scalp maps are often far from 'dipolar' (i.e., resembling the projection of a single dipolar source). Because of the need for ICA to 'mix' all of the EEG sources into the available number of components, decomposing data with a larger number of (clean signal) channels may be preferable when there are enough data to decompose them (see following). But decomposing a smaller number of channels will likely prove beneficial as well.

Data requirements

Successful ICA decomposition requires an adequate amount of data. We may say, metaphorically, that the independence of many source signals cannot be "expressed" in brief mixtures of them. To "express" their independence (or less metaphorically, for an ICA algorithm to recognize it), a considerable amount of data is typically required. Thus, successful ICA decomposition typically profits from being applied to a large amount of data, typically *the entire collection of continuous data or extracted and then concatenated single trials from an event-related EEG/ERP task session*. The most frequent mistake researchers make in attempting to apply ICA to their data is to attempt to apply ICA decomposition to too few data points. For data with large numbers of channels (64 or more), we suggest it is optimal to decompose a number of time points at least 20 or more times the number of channels *squared*. This is only a heuristic standard, *not* a strict minimum, and using this much data does not in itself guarantee an optimal decomposition. For very dense scalp arrays, this standard could require an unreasonable amount

of data. For example, to decompose 256-channel data 20×256^2 time points at 256 points/sec would require over 80 minutes of recording time and occupy nearly 1.5 GB, though by this same standard for 64-channel data a 22-minute recording occupying about 0.35 GB would suffice.

We are not as sure about the influence of sampling rate on ICA decomposition. Doubling the sampling rate during a recording period shortened by half might not produce as effective a decomposition, since the higher frequencies captured in the data acquired with a higher sampling rate would be small, relative to lower frequency activity, and might have lower source-signal-to-noise ratio. See Onton & Makeig (Onton and Makeig, 2006) for further discussion.

Optimally the data should be from a period in which the subject is predominantly in the same state (for example, awake and attentive), and performing the same type of task or tasks. Although standard ICA methods are theoretically able to separate data into sources that are principally active at different periods in the data set, a promising newer mode of ICA decomposition allows learning multiple sets of independent components wherein each time point is associated with only one decomposition (Palmer, 2008).

Since most ICA decompositions do not use relationships among time points to perform the source separation (infomax ICA actually re-shuffles the order of the time points in each training step), it makes no difference whether the data are from contiguous time periods or from separate data epochs. For example, Makeig et al. (2002) reported a 31-channel decomposition of data only from the N1 response-peak period following presentations of visual stimuli. This was possible because of the large number of such stimuli (2,500) viewed by each subject, and the relatively small number of channels recorded (31).

Finally, it should be noted that ICA is reference free, since any re-referencing of the data that preserves its dimensionality does not change its information content or its sources. After re-referencing, the IC scalp maps will change but IC activation dynamics and equivalent model dipole locations should not change except as a result of normal statistical variability, which is typically small for ICs with highly ‘dipolar’ scalp maps.

ICA versus PCA

Another well-known method of linear decomposition of multi-channel data, *principal component analysis* (PCA), transforms multi-channel data into a sum of uncorrelated principal components so named because they each, in sequence, account for the most possible (or ‘principal’) variance in the remaining uncorrelated (or orthogonal) portion of the signal data not accounted for by the preceding principal components. By contrast, independent components (ICs) produced by ICA have no natural order – though it is common to sort them by descending variance of the (back-projected) scalp data they each account for. Again, for either PCA or ICA the whole scalp data

are the sum of the individual component contributions. The simple system of Fig. 3.4 thus applies to PCA as well, though for PCA, W^{-1} is called the eigenvector matrix and W is its inverse. Also, in PCA both the eigenvector matrix and the activations or factor weights matrices are normalized, and an intervening diagonal matrix E, the *eigenvalues* matrix, is used to hold the relative scaling of the components (i.e., $X = W^{-1}EU$), while the columns of the mixing matrix as well as the rows of the activations matrices are each normalized (to have unity root-mean-square values).

The important difference between ICA and PCA is in their quite different goals or objectives. We may say that PCA attempts to *lump* together maximum signal variance from however many sources into as few principal components as possible, whereas ICA attempts to *split* the signal into its separate information sources, regardless of their variance. This makes PCA useful for compressing the number of dimensions in the data while preserving as much as possible of the data variance. However, elimination of low-variance PCs for the purpose of dimension reduction most likely deletes portions of nearly *all* the source activities, not just the smaller ones. When data length is not long enough to successfully decompose all available channels, another possibility is to perform ICA decomposition of data from some channel subset. The relative value of these two approaches (principal subspace versus channel subspace) is difficult to evaluate in advance.

The ‘maximum successive variance’ objective of PCA also forces both the principal component activities and the scalp projections (scalp maps) to be mutually uncorrelated (orthogonal). Since the scalp projections of brain (and non-brain) sources are rarely themselves orthogonal, this property forces all but the first very few principal component scalp maps to resemble checkerboards that cannot reasonably represent the activity of single EEG sources. In general, principal component maps do not resemble the projection of a single EEG source unless one source (often, eye blinks) or two sources with near-orthogonal maps (for example, lateral and vertical eye movements) dominate the signal variance.

For this reason, some ERP researchers advocate the use of post-PCA component rotation methods developed for earlier factor analysis approaches, such as Varimax or Promax (Dien et al., 2005). These may help focus the scalp maps of the very first components to emphasize a few large source activities (such as eye blink artifacts and lateral eye movements), but both simulations and actual decompositions show their power to accomplish this for many brain and non-brain sources pales in comparison to ICA methods when properly applied to sufficient data (Makeig et al., 1999b; Makeig et al., 2002).

Independence among source waveforms, however, is a much stronger assumption than the simple absence of correlations between source pair signals. Substituting the stronger assumption of independence between component activities instead of requiring them only to be

uncorrelated allows ICA to return independent components (ICs) having any (non-identical) scalp maps. Every IC scalp map is then free to represent the projection of a single brain or non-brain signal source, whereas PCA component maps are constrained to be uncorrelated and therefore most have a “checkerboard” appearance not compatible with a single cortical (or other) source projection.

Theoretically, exact independence is such a strict requirement that it can never be established for EEG signals with finite length. ICA algorithms, therefore, may at best produce components with *maximal* independence by ensuring that components continually *approach* independence as the ICA algorithm iteratively applied to the data. The degree of IC independence achieved may differ for different data sets and also for different ICA algorithms applied to the same dataset. Our discussion of the nature of brain EEG sources (Section I) implies that the more independent the recovered IC activities, the more dipolar (or occasionally, bilateral dual-dipolar) the IC scalp maps of brain components, a trend supported by recent tests (Delorme, unpublished data).

Independent component contributions to single trials and ERPs

By definition and design, independent component processes contribute nearly independent temporal variability to sets of single-trial epochs. Each IC represents an independent EEG process whose continuous activity variations in the single trials are available for inspection and analysis. In particular, brain-based ICs with near-dipolar scalp maps may each be presumed to index the near-synchronous field activity arising in a single patch of cortical neuropile (or occasionally, simultaneously in two bilaterally symmetric and likely tightly-coupled cortical patches). Examining the trial-to-trial variability in the IC activities relative to a set of time-locking events may allow a more detailed understanding of event-related brain dynamics than examination of raw scalp channel data themselves, since the effects of source (and artifact) mixing by volume conduction have been removed or strongly reduced by ICA.

As an example of this, Figure 3.7 shows ERP-image plots of the activities of six independent components in the same response-locked single-trial data as earlier figures. In each ERP image, IC activity is scaled (in μV) as in it projects to the near-vertex (Cz) electrode. Locations of the equivalent IC dipoles are shown in the central panel. Trials are ordered exactly as in Fig. 3.3, by the amplitude of the LPC peak 120 ms after the button press (0 ms) at the vertex. Note the quite high degree of overlap of the ‘dipolar’ scalp maps of these midline components that contribute temporally *independent* contributions to the recorded EEG signals.

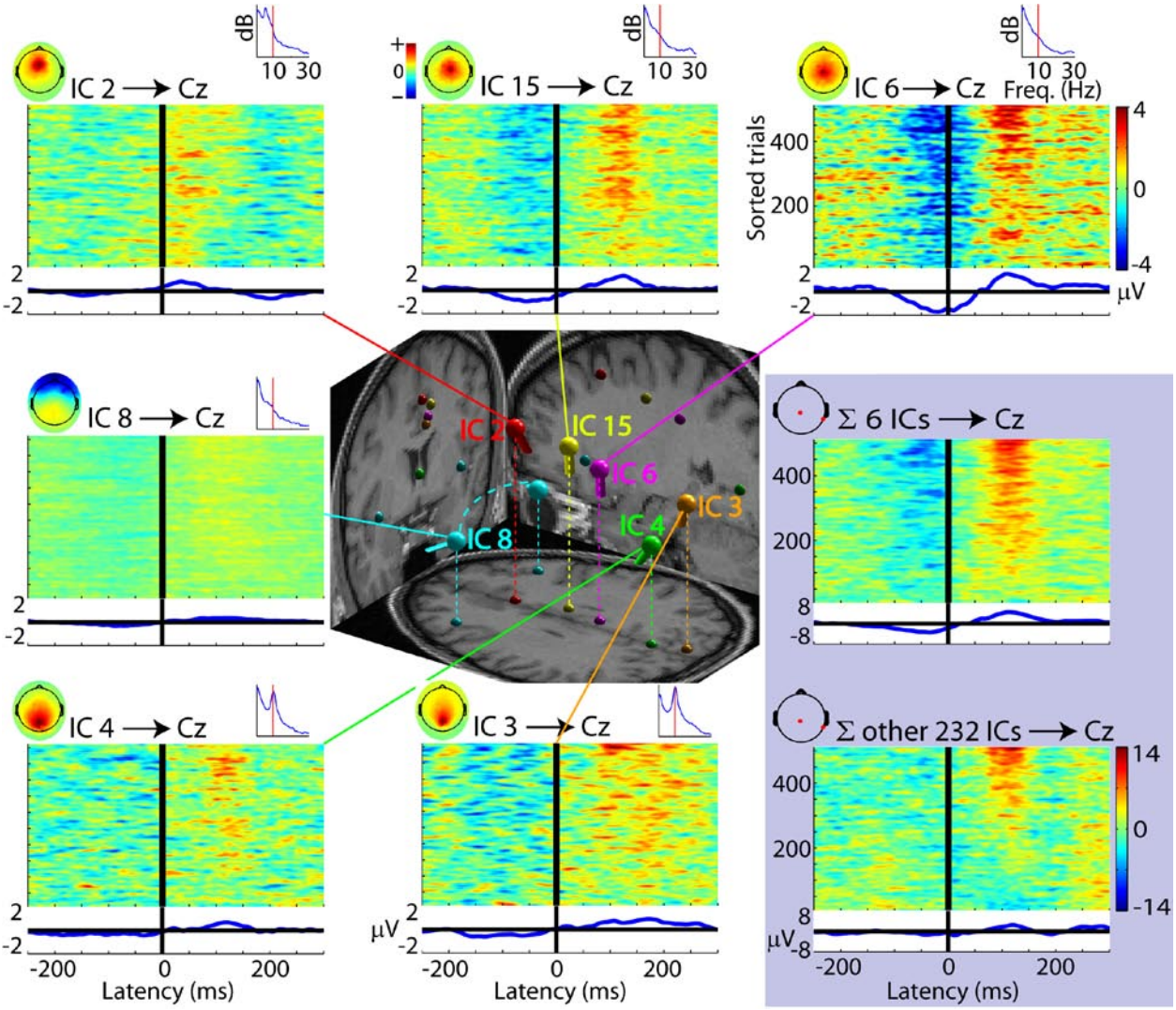

Figure 3.7. *ERP images for six midline and one bilateral brain independent components (ICs) in the same session as earlier figures (compare Fig. 3.6), with trials sorted in the same order as Fig. 3.3, and scaled as they contribute to the near-vertex channel signal imaged in Fig. 3.3. There, the trials were sorted by the amplitude of the LPC peak centered 120 ms after the button press (at latency 0 ms) at the near-vertex channel shown on the head cartoon (middle right panel). The sum of the signals projected by these six component processes is shown in the same panel (note difference in color scale). The summed contributions of the other 232 non-artifact components to the same channel are shown in the lower right panel. The channel ERP-image plot (Fig. 3.3A) is thus the sum of the two ERP-image panels with grey backgrounds at the right of this figure. Separately, the contributions to the ERP of each of the other components summed in the lower right panel are smaller than those of the six components whose contributions are shown in the other panels.*

The sum of the signals projected by these six component processes is shown in the middle right (grey) panel. Note that the trial order used here, which sorts the trials by ERP amplitude at the selected channel (as in Fig. 3.3B), only partially sorts the post-response amplitude of each IC activation. This is shown by the uneven gradations of the post-motor response positivity in the component ERP-image panels and in the sum of their contributions at the same near-vertex channel (middle right panel). This implies that contributions to the activity fitting the mean ERP template in Fig. 3.3B sum *varying spatial combinations* of these and other brain source processes in the different trials.

Note also in Fig. 3.7 the *different peak latencies* of the LPC peak for ICs 2, 6, and 15 (top row). The trial ordering selected in Fig. 3.3 based on the amplitude of the average ERP ignores these single trial and component process differences. It sorts trials according to the amplitude of the *average summed contributions* of these and other independent sources, rather than on the varying amplitudes and latencies of the individual source processes.

The summed and similarly smoothed (smaller) contributions of all the other 232 non-artifact components to the same channel in the single trials are shown in the lower right (grey) panel. The channel ERP image in Fig. 3.3A is thus the sum of the two right (grey) panels, as well as the sum of the two ERP-image panels Fig. 3.3 B and C. Neither of the 'remainder' ERP images (Fig. 3.3C or Fig. 3.7 lower right) suggest a satisfactory modeling of the LPC as summing just two factors – i.e. neither the 'six-ICs' model imaged in Fig. 3.7 (lower right) nor the 'invariant-ERP' model in Fig. 3.3 (B and C). Using the ICA model, however, we may examine, for example, whether the particular trial-to-trial differences in the LPC window of each identified IC may indicate that its response varies with some dimension of the varying trial context (e.g., each trial's particular cognitive and behavioral demands and demand history).

Figure 3.8 shows that a portion of the IC trial-to-trial variability highlighted in Fig. 3.7 is indeed linked in orderly ways to behavioral trial differences. It shows ERP-image plots for the same six independent components (ICs) as in Fig. 3.7, again scaled as they contribute to the central scalp channel (center right) but here sorted by subject reaction time and then smoothed with a wider (50-trial) moving window to more clearly visualize trends. Note that the wider averaging window reduces the overall amplitude of the imaged data through phase cancellation of trial-to-trial IC variability in neighboring trials (compare the color μV scale limits here with those in Fig. 3.7).

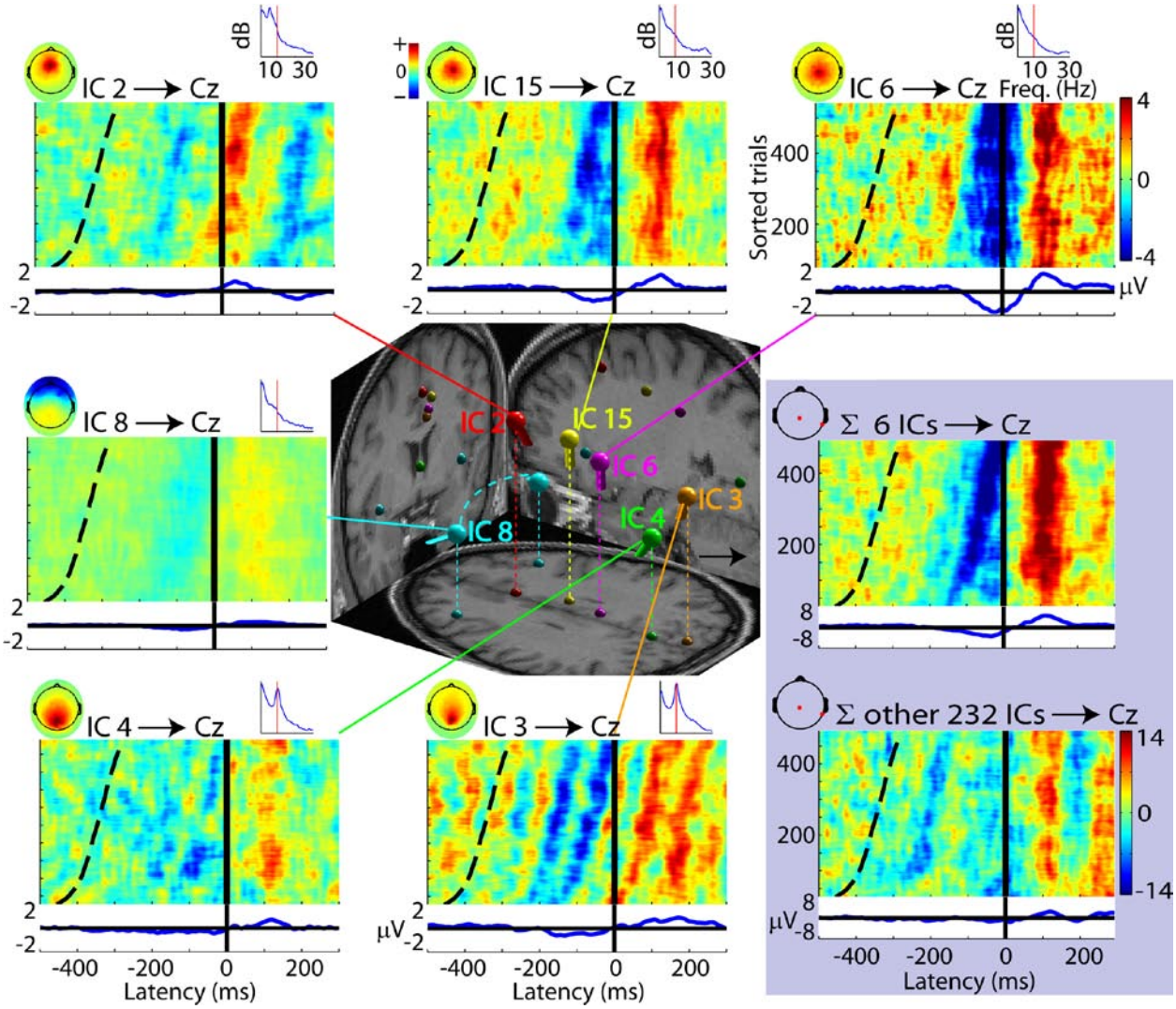

Figure 3.8. *ERP images for the same six independent components (ICs) as in Fig. 3.7, again scaled as they contribute to the central scalp channel (center right cartoon head) but here sorted by subject reaction time and then smoothed with a broader (50-trial) moving window (note color scales). The middle right panel shows the summed contribution of the six ICs to the whole channel signal, with again (lower right) the difference between the whole signals and the sum of these six source contributions. This view reveals that a portion of the trial variability evidenced in Fig. 3.7 is tightly linked to differences in subject reaction time. For anterior sources ICs 2 and 8, the negativity preceding the button press is time-locked to stimulus onsets, while the subsequent LPC is mainly time-locked to the subject button press. For central midline sources IC15 and IC6, the negativity onset and offset are time-locked to the stimulus and button press respectively. Posterior sources IC3 and IC4 appear to exhibit partial phase resetting of their alpha activities by stimulus presentations. The single scalp channel signal sums all these (and other) event-related source dynamics.*

The middle right panel shows the summed contribution of the six ICs to the whole channel signal, and the lower right panel, again the remainder of the whole channel signals they do not account for. This view reveals that a portion of the trial variability evidenced in Fig. 3.7 is tightly linked to differences in subject reaction time. For anterior sources ICs 2 and 8, the negativity preceding the button press is time-locked to stimulus onsets, while the subsequent LPC is mainly time-locked to the subject button press. For central midline sources IC15 and IC6, the negativity onset and offset are time-locked to the stimulus and button press respectively. Posterior sources IC3 and IC4 appear to exhibit partial phase resetting of their alpha activities following stimulus presentations (see Section IV). The single scalp channel data and ERP sum all these (and doubtless other) event-related source process contributions.

It may be worth the reader's effort to examine again carefully the trial variability in different dimensions visualized for scalp channel data in Figs. 3.2 and 3.3, and for IC data in Figs. 3.7 and 3.8.

Independent component clustering

To compare, group, or further average ERPs across subjects and/or sessions, channel data are typically identified by the labeled (Cz, Pz, etc.) or measured (x,y,z) channel positions on the scalp. Though equating of equivalent scalp locations across sessions and subjects is adequate for many purposes, it ignores the variety of individual cortical configuration differences, particularly in the positions and orientations of cortical sulci, that may orient anatomically equivalent EEG source projections toward different scalp areas in different subjects. In this case, functionally equivalent sources may have quite different scalp maps, and therefore electrodes at analogous locations will record different weighted mixtures of source activities. Thus, for example, signals from 'my Cz' and 'your Cz' may not be equivalent, even if our brains have equivalent cortical areas that function identically. This produces unavoidable and rarely considered variability in scalp recordings that are compared or averaged across subjects.

Since under favorable circumstances ICA can separate scalp-recorded signals into the volume-conducted activities of maximally independent brain sources, it may be more accurate to group, compare, and characterize *functionally* equivalent clusters of ICs across subjects and/or sessions. Finding these IC equivalence classes is the challenge of *IC clustering* across subjects and/or sessions. IC clusters may be selected on the basis of their equivalent dipole locations, ERPs, and/or other measures.[7]

Figure 3.9 shows a sample application of IC clustering to a grand mean ERP averaging data from 12 subjects who participated in a visual attention-shift experiment. Throughout the experiment, subjects made speeded manual choice responses to indicate in which dimension (shape or color) the lateral target stimulus (presented at 0 ms) differed from a simultaneously

presented neutral background stimulus. In the 12 subjects' ICA-decomposed data, we identified 22 clusters of similarly located and similarly reacting ICs by comparing equivalent dipole locations, mean power spectra and event-related spectral perturbations (ERSPs, see Section IV) in three stimulus conditions. Figure 3.9 focuses on a grand-mean ERP time locked to stimulus presentation (at latency 0 s) in one condition. The central panel shows IC equivalent dipole locations for four of 22 identified IC clusters.

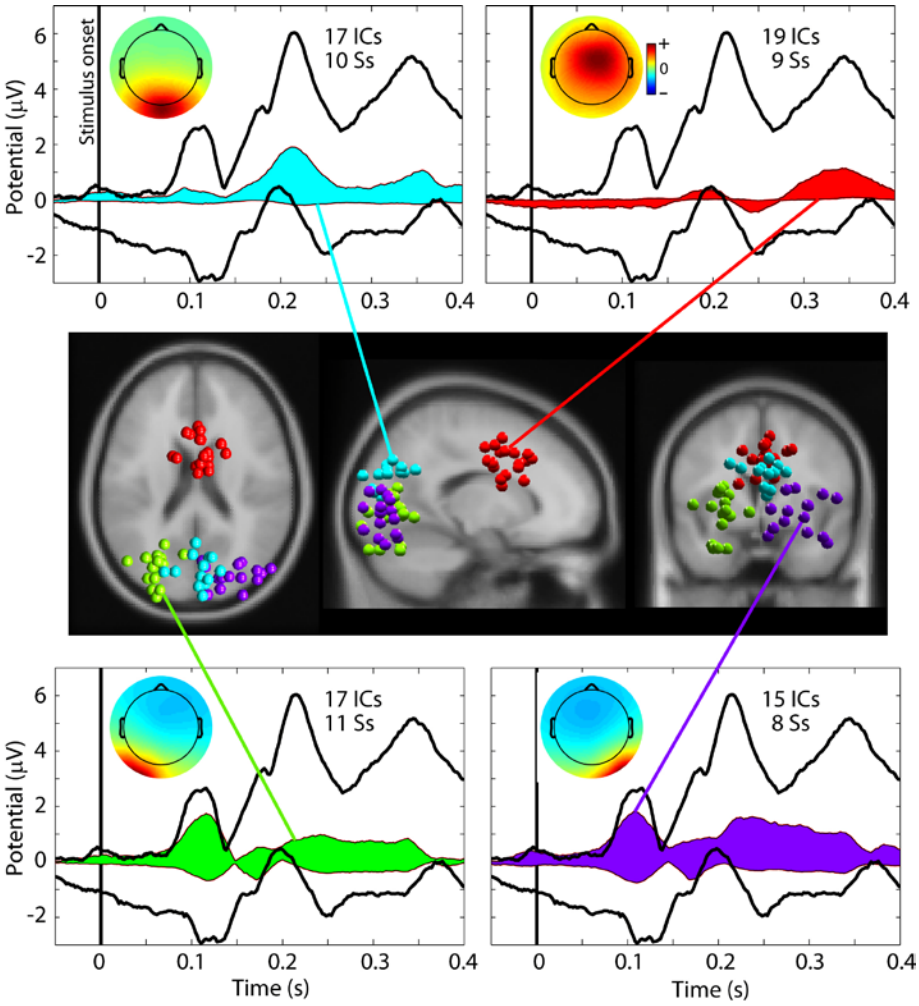

Figure 3.9. *Equivalent model dipole locations, mean scalp maps, and cluster projection envelopes of four (of 22) independent component cluster contributions to a visual stimulus ERP (grand mean over 12 subjects) in a visual attention-shift experiment. ERP envelopes show only the most positive and most negative channel values at each response latency. Here, the envelope (see text) of the grand-average back-projection of the indicated IC cluster is color-filled. The outer black traces are the envelope of the whole grand-mean ERP after removing contributions of component clusters and outlier components accounting for eye, muscle, and other non-brain artifacts. The bottom two panels show clusters accounting for most of the P1 peak in the grand-mean ERP. The upper two panels indicate the portions of the grand-mean ERP accounted for by a central posterior cluster (blue, with maximal contribution to the peak labeled P2) and a midline cluster (red, with maximal contribution to the later peak labeled P3).*

The black traces in the four top and bottom plot panels show the *envelope* of the grand mean ERP (i.e., its maximum and minimum channel values at each latency). The four top and bottom panels show the cluster-mean scalp maps and the boundaries of the colored regions, the *envelopes* of those portions of the grand-mean ERP accounted for by each of the four clusters.[8] Envelope plotting allows the ERP contributions of one or more ICs or IC clusters to be visually compared with the envelope of the whole scalp ERP.

The lower panels show two lateral occipital IC clusters (see the green and purple IC dipoles) that accounted for nearly all the bilateral positive peak near 110 ms in the ERP, plus a later sustained "ridge-like" feature. The upper two panels show the portions of the grand mean ERP accounted for by a central posterior cluster (blue) whose maximum ERP contribution was to the positive peak near 220 ms, and a midline cluster (red) that contributed maximally to a later positive peak near 350 ms.

Note that although the model dipoles are represented, for visual convenience, as small balls, the actual uncertainty in their individual locations is rather larger, as are the distributions of cortical territory across which synchronized local field activity (in our model) produce the far-field potentials recorded by the scalp electrodes. Sources of dipole location error in Fig. 3.9 include possible differences in recorded electrode positions relative to each other and the scalp, errors in co-registering the electrodes to the head model, and differences in head shape, and possible differences in head tissue conductivity parameters. Although the equivalent model dipole locations shown in the middle panel are relatively tightly grouped, their spread may also reflect differences in the locations of functionally equivalent cortical areas across subjects, since similarities between activity measures were also considered in assigning components to clusters.

IC clustering is required to compare ICA decompositions from more than one subject or session. It can be used to understand the locations and dynamics of independent component processes contributing to average ERPs as well as to the unaveraged single trials. IC clustering provides an involved but under favorable circumstances, we believe a more adequate answer to the inverse problem of estimating the distributed sources of ERP scalp maps and the relationship of the source dynamics to experimental events and conditions. In particular, IC clustering gives a more adequate solution than simply attempting to model the distributed cortical sources of ERP scalp maps themselves. IC clustering also allows testing for differences within and/or between subject groups reflected in the presence or absence of ICs in one or more clusters and/or on details of the clustered IC locations or activities.

IV. Time/frequency analysis of event-related EEG data

Time-locked but not phase-locked: Event-related spectral perturbations (ERSPs)

To understand the relationship of ERP features to the event-related dynamics of the entire EEG signals from which they are derived, it is convenient to use time/frequency analysis that models the single-trial data as summing an ever-changing collection of sinusoidal bursts across a wide frequency range. Note that producing this representation of the data does not mean that the EEG is necessarily composed of such bursts, or that the burst shape or *window* employed in the analysis is necessarily a physiologically accurate template. Rather, as Joseph Fourier first showed for heat flows along a copper tube, frequency analysis, and later non-stationary time/frequency analysis, can be used to represent *any* temporal activity pattern, not limited to those portions of the recorded signals that do indeed resemble single time/frequency basis elements, e.g. symmetric and smoothly tapered bursts at a single frequency. However, the frequent appearance of periodicities at multiple frequencies is a clear and remarkable feature of EEG records and this property of the signals show quite clear and spatially distinct changes accompany changes in arousal and attention, making time/frequency analysis clearly useful for EEG analysis.

Rather than averaging the recorded ('time-domain') event-related data epochs directly, one may average their time/frequency transforms (see also Chapter 2, this volume). Averaging time/frequency power or log power values in a regular grid of time/frequency windows gives an event-related spectrogram that is nearly always dominated by relatively large low-frequency activities. Normalizing the result, therefore, by subtracting the mean log power spectrum within some defined 'baseline' period (pre-stimulus or otherwise, as relevant to the analysis) allows a color-coded time/frequency image of mean log spectral differences we call the *event-related spectral perturbation* (ERSP) image (Makeig, 1993). Basing the ERSP on changes in log power implicitly assumes a multiplicative model by which EEG spectral changes represent the multiplication or division of the baseline power at each frequency in each latency window relative to the time-locking events.

Determining either the amplitude or the phase of activity at a particular time/frequency point involves matching the data in a window surrounding the given time point to the oscillatory basis element (typically a tapered sinusoidal burst or 'wavelet'). To measure low frequencies, this window must be relatively long, limiting the frequency range considered for short data epochs. Also, event-related changes in spectral power may last longer than significant features in the ERP. For these reasons, our own typical time/frequency analyses use epochs including at least 1 s before the time-locking event and continuing to 2 s or more following it, allowing a frequency decomposition based on a three-cycle tapered sinusoidal wavelet down to 3 Hz.

The mean ERSP of a set of event-related data epochs can index event-related dynamics that leave no trace at all in the ERP average of the same epochs, as first shown for alpha band

activity by Pfurtscheller and Aranibar (Pfurtscheller and Aranibar, 1977). Thus the ERSP transform of the average ERP for a set of data epochs, while of possible interest to compute, may bear little or no resemblance to the average ERSP for the same collection of epochs. For one, significant ERSP features may long outlast the reliable ERP features. For example, Figure 3.10 (top panel) shows a mean event-related spectral perturbation (ERSP) time/frequency image for a left-frontal independent component (IC2) time-locked to button presses following target stimuli (from the same session as Fig. 3.1-3.8). Regions of non-significant difference from baseline (here $p < .001$, uncorrected for multiple comparison) are masked with light green. The ERSP image reveals that mean alpha band power just below 10 Hz increases weakly following the button press, while mean low-beta activity (15-20 Hz) in two frequency ranges increases most markedly after 400 ms. Activity at the 6-Hz baseline spectral peak (see top left side-facing blue baseline spectrum) does not change, though activity below 5 Hz increases weakly around the button press, and then decreases beginning 200 ms after the button press.

Spectral power in the average ERP is often referred to as the spectrum of activity *evoked* by events, while changes in spectral power appearing in the ERSP are dubbed changes *induced* by events. However, this terminological distinction should not suggest that the two are necessarily physiologically distinct. To see this, we need to consider changes in phase statistics associated with experimental events.

Phase-locking across trials: inter-trial coherence (ITC)

The ERSP disregards completely the consistency or inconsistency of the *phase* of the activity at each frequency and latency in a set of event-related epochs. *Inter-trial coherence* (ITC), or more precisely, inter-trial phase coherence, introduced as ‘phase-locking factor’ by Tallon-Baudry et al. (Tallon-Baudry et al., 1996), measures the degree of consistency, across trials, of the phase of the best-fitting time/frequency basis element at each latency/frequency point. Phase consistency is measured on a scale from 0 (no consistency, phase across trials is random and uniform around the phase circle) to 1 (phase perfectly consistent across trials). The ITC for any *finite* set of randomly-selected data epochs will typically not be 0. Therefore, it is important to compute a baseline threshold for the appearance of significantly non-random phase coherence. An ITC reliability threshold for a set of trial data can be found using either parametric or non-parametric statistical methods (Mardia, 1972; Delorme and Makeig, 2004).

It is important to note that the ITC and *ERSP* images for a given set of event-locked data epochs may have few or even *no* common features. For example, in Fig. 3.10 the post-motor response increases in alpha and then in beta-band power in the frontal midline IC spectrum are not mirrored by significant changes in ITC at the same latencies and frequencies.

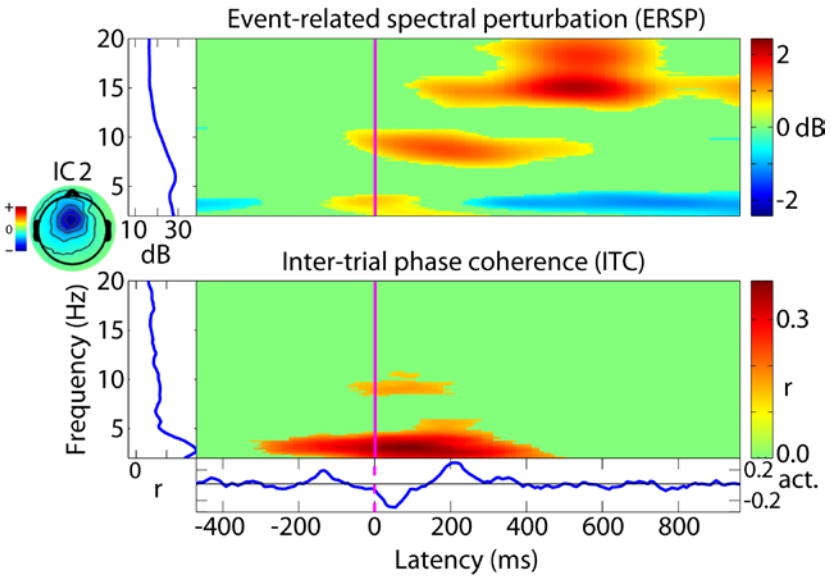

Figure 3.10. *Event-related time/frequency analysis of the set of independent component trials shown in Fig. 3.7 and 3.8 (upper left). Mean event-related spectral perturbation (ERSP, top) and inter-trial coherence (ITC, bottom) time/frequency images for a leftward-pointing mid-frontal independent component (IC2) time-locked to button presses following target stimuli. Regions of non-significant difference from baseline (p< .001, uncorrected) are masked with light green. The top (ERSP) image reveals that mean alpha band power at 10 Hz increases weakly following the button press. On average, low-beta activity (15-20 Hz) activities first decrease slightly, then increase after 400 ms. Activity at the baseline spectral peak (6 Hz, see top left baseline spectrum plot) does not change, though activity below 5 Hz is maximal at the button press. The bottom (ITC) image shows that 4-Hz activity becomes partially but significantly phase-locked around the button press, meaning the portion of the component ERP (lower trace) near 4 Hz is statistically significant (compare the ERP trace below), as are its weak 10-Hz "scalloping" between -50 and 200 ms. Component activation units ('act.') are proportional to scalp μV. The statistically significant changes in mean spectral power in the beta band, shown in the upper panel, are not associated with significant ERP features and therefore represent changes in component activity time-aligned but not phase-aligned to the button presses.*

However, there is an intimate relationship between the ITC and the *ERP*. In particular, the occurrence of a significant ERP peak or other feature requires significant ITC (see Section II). In this sense, a significant ERP value at any time point reflects and requires significant ITC values at one or more frequencies at that time point (except in odd, improbable cases). Note also that the ITC for any frequency may be significant even at latencies at which the mean potential ERP value is 0. For example, if the 0 value in the ERP occurs during the zero crossing of an alpha oscillation in each trial, then the ITC at that alpha frequency might be highly significant, although the ERP at that time point might have a value of 0. The ITC may also be significant, at a particular latency, at more than one frequency. If so, this will be reflected in the shape of the ERP waveform surrounding the latency in question.

For example, in the ITC image in Figure 3.10 (lower panel), activity near 4 Hz becomes partially but significantly phase-locked around the button press event (0 ms), meaning the portion of the component ERP (bottom trace) at 4 Hz is statistically significant. As well, ITC becomes (barely) significant at 10 Hz, a fact reflected in the weak 10-Hz 'scalloping' in the ERP waveform between -200 and +300 ms. No other ITC frequencies, and therefore no other ERP frequencies, are significantly different from chance. The statistically significant ERSP changes in mean spectral power in the beta and low-gamma bands, shown in the upper panel, are not associated with significant ITC features and therefore represent component activity that is phase inconsistent, i.e. not phase-aligned (or phase-coherent) across trials, and so does not contribute significantly to the trial-average ERP.

ERPs and partial phase resetting

The relatively low peak ITC values in Fig. 3.10 (~0.4) are not unusual for longer-latency ERP features. An alternate model, first proposed for selected event-related data as early as 1974 by Sayers (Sayers et al., 1974), is known as the phase-resetting or partial phase-resetting model. Phase resetting refers to a phenomenon seen both in mathematical models and in biological systems in which the phase of an ongoing periodicity (e.g., the cardiac or circadian cycle) is reset to a fixed value relative to the delivered perturbing stimulus. For example, brief exposure to strong light delivered to a dark-adapted rat (or human) at almost any phase of the wake-sleep cycle, will tend to reset the cycle to a fixed phase value (Winfree, 1980; Czeisler et al., 1986; Honma et al., 1987; Tass, 1999). At the frequency of the ongoing, spontaneous rhythm, an ITC measure time-locked to comparable events delivered at random time points throughout the session will become significant as the phase of the rhythm in some or most of the trials is reset to a fixed value. If the phase of the rhythmic activity then tends to continue to advance in a regular manner from its initial reset value, the ITC time-locked to the events of interest will remain significant for some number of cycles until, across trials, natural variability randomly separates the advancing phase values.

The term 'phase resetting' has been applied to EEG dynamics in a less formal sense, since in most cases there is no constant, ongoing rhythm for experimental events to perturb. Rather, in many cases the signal contains only intermittent bursts of alpha or other frequency spindles of various lengths. The term 'phase resetting', therefore, can be formally applied in some statistical sense to mean that the phase statistics (as measured, for example, by the ITC) are transiently perturbed following events of interest. If, whenever rhythmic activity at a given frequency is present, its phase distribution following the time-locking events becomes non-uniform, ITC will increase and may tend to remain significant for as long as the rhythmic activity is present.

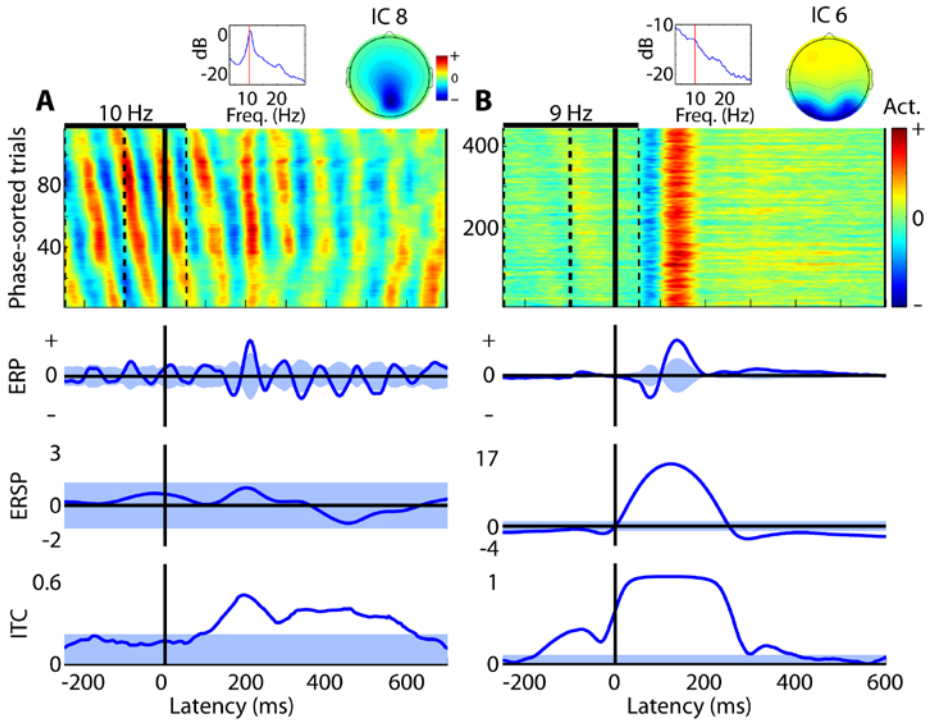

Figure 3.11. *ERPs and partial phase resetting. The right panel shows an ERP-image plot visualizing the responses in over 400 single trials of a bilateral occipital independent component (IC) process following presentation of a letter at the central fixation point of a subject participating in a working memory task. The bilateral component (IC6, sixth by variance expressed in the data) produces a response largely resembling a one-cycle sinusoid at 9 Hz. The mean ERP trace (below the ERP image) plots its mean time course, time locked to stimulus onset. The ERSP trace (below that) shows that the mean level of 9-Hz energy in the data, during this period, is 15 dB or more higher than in the pre-stimulus (or ensuing) period. The ITC trace (below that) confirms that during this ERP feature the phase of the entire 9-Hz activity in the trials is highly consistent (ITC approaching 1). The blue backgrounds show $p < .01$ probability limits, demonstrating that all three measures are highly significantly different from baseline in this period. The panel on the left shows a quite different set of over 100 data trials for a medial (or medial bilateral) occipital IC process in the five-box task of Fig. 3.1 in which stimuli were presented above and left of a central fixation cross, while the subject retained fixation. The large amount of alpha band activity produced by this IC process under these circumstances likely reflects 'alpha flooding' of relevant visual cortex when visual attention is forced by the task to remain elsewhere in the visual field (Worden et al., 2000).*

Figure 3.11 shows ERP-image plots for two sets of visual-stimulus locked trial data from two ICs captured in different subjects under rather different task conditions. In panel 3.11A (left), the stimulus is a briefly-flashed disk presented at a central, visually unattended target square located above a central fixation cross during a visual selective attention task. The IC shown here has a bilateral equivalent dipole model in or near primary visual cortex, and produces abundant alpha-band activity (see power spectrum), likely reflecting the 'alpha flooding' of visual cortical areas sensitive to the foveal fixation region when the subject places his or her

visual attention elsewhere in the visual field (Worden et al., 2000). This alpha activity appears to be 'partially phase-reset' (ITC ~ 0.4) for nearly 500 ms (5 alpha cycles) following stimulus presentation. In the ERP image, trials are sorted by alpha phase in a three-cycle window ending 50 ms after stimulus onset. The possibility of 'partial phase resetting' is suggested by the bending and then near-vertical alignment of the positive and negative wave fronts beginning near 100 ms in the ERP image, when the ITC becomes significant.

Note that the visual evidence presented by this ERP-image, including the finding of a significant ITC (lower ITC trace), are *not* in themselves sufficient evidence to prove that these data truly fit a 'phase resetting' model (for more discussion, see Chapter 2, this volume). Nor do they necessarily rule out a 'true-ERP' model for the data, e.g. a model in which the same ERP (upper ERP trace) resembling an alpha burst is simply *added* to ongoing alpha and other EEG activity in every trial (as in Fig. 3.3), and that the ongoing alpha activity is in turn reduced in amplitude just enough to make total mean alpha power at each latency constant, as observed here (middle ERSP trace). However, keep in mind that these data are the result of spatial filtering by ICA of a single independent source, very likely focused on a single source area (or closely spaced medial bilateral areas), given the highly 'dipolar' from of the IC scalp map. It thus seems to us physiologically implausible that, following these visual events, this same cortical source area produces ongoing random-phase alpha activity *plus* a fixed but wholly *unrelated* alpha-burst ERP. Several groups have recently proposed measures to further test phase resetting models on data such as these (Mazaheri and Jensen, 2006; Hanslmayr et al., 2007; Martinez-Montes et al., 2008). Ultimately, the issue will likely be settled by fitting concurrent scalp and intracranial EEG recordings to generative models of cortical field dynamics, a process begun by groups studying human brain responses during cortical recording (Wang et al., 2005).

In panel 3.11B (right), on the other hand, the time-locking stimulus is a letter presented *at fixation* in a letter working memory task. The spectrum of the bilateral lateral-occipital IC (inset) has only a weak alpha band peak, and no sign of prolonged alpha-band phase resetting following the highly stereotyped (ITC > 0.8) component IC stimulus-evoked response (which contributes strongly to the P1-N1-P2 features of the full scalp ERP, not shown). At the frequency best fitting the ERP complex (9 Hz), mean single-trial amplitude during the ERP is nearly 6 times (over 15 dB) higher than the mean amplitude of activity at the same frequency in the pre-stimulus baseline. Phase-sorting the single trials at 9 Hz in a window ending 50 ms after stimulus onset (as in A) shows that the phase of the weak alpha activity present in single trials during the baseline period has no obvious effect on the latencies of the subsequent evoked-response activity in the same trials. For this IC stimulus response, therefore, a 'partial phase-resetting model' seems unnatural and a 'true ERP' model adequate. However, even here one may ask whether, for example, the frequency peak of the ERP (9 Hz) may not also be a peak of the spontaneous (baseline) spectrum of this cortical area.

Many authors have attempted to draw a hard distinction between evoked and induced event-related activities, defining *evoked* activity as being activity completely time-locked and phase-locked to the stimulus (ITC = 1) and thereby composing the ERP, while the remainder of the single-trial activity, having no phase locking to the time-locking events (ITC = 0) is defined as *induced* (Galambos, 1992). While this distinction may be useful for some purposes, drawing this terminological distinction does not mean this decomposition of the EEG signal into evoked (ERP) activity plus induced (other EEG) activity has any natural physiological basis. Think of a stack of five pennies – Again, does this stack 'really' sum two groups of two and three, or of groups of four and one? In fact, the stack of pennies retains no trace of how it was constructed and thus cannot be said to be any more 'really' 3+2 than 4+1, no matter how it was originally constructed. The same applies to the model of event-related EEG data illustrated in Fig. 3.3: *EEG data = ERP + Other*, a model that, as ICA decomposition and Figs. 3.7 and 3.8 suggest, disregards the varying single-trial contributions of spatially separable data information sources, some clearly linked to trial-by-trial behavioral differences.

Figs. 3.7 and 3.8 suggest that scalp ERPs sum channel activity arising from different mixtures of spatial source processes in different trials. But how should we think of the average response of a single IC? Assuming that an IC activation does index locally-synchronous or near-synchronous field activity of a single patch of cortex, can the IC activity producing the IC *"ERP"* activity (strictly time-locked to the set of evoking events) be physiologically distinct from o*ther* (non phase-locked) EEG activity originating at the same moments in presumably *the same cortical patch*?

Linear summation in cortex, even of direct sensory input and ongoing cortical dynamics, appears physiologically implausible without strong nonlinear interactions. Fiser and colleagues (Fiser et al., 2004) have noted that even at prototypical sensory cortex – the input layer of primary visual cortex (in ferrets), only a few percent of the synapses deliver information directly from the eyes via the lateral geniculate nucleus (LGN). In accord with this fact, they report that "at all ages including the mature animal, correlations in spontaneous neural firing [during natural vision] were only slightly modified by visual stimulation, irrespective of the sensory input. These results suggest that in both the developing and mature visual cortex, sensory evoked neural activity represents the modulation and triggering of ongoing circuit dynamics by input signals, rather than directly reflecting the structure of the input signal itself" (Fiser et al., 2004). If this is the case even for V1, it should not be less so for cortical areas that are not primary sensory areas. Clearly, deeper understanding of the EEG dynamic changes associated with sensory and other events will require more detailed observation and modeling of brain dynamics at multiple spatial scales. In terms of EEG research, more detailed observations and modeling are needed of trial-by-trial differences in oscillatory activity and its relationship to its transformation by experimental events.

Event-related coherence

Cognitive events – moments at which we apperceive the *significance* of some sensory event and mentally *'grasp'* its immediate consequences for our attention and behavioral planning – must involve and/or produce complex and distributed changes in EEG dynamics. Furthermore, some mechanism of information transfer between brain regions must exist that is dynamically dependent both on the nature of the stimulus and its relation to subject expectations and intentions. A possible mechanism for this transfer may be indexed by transient temporal coupling between pairs of sources relative to experimental events. One measure of this coupling is event-related coherence (ERC) (Delorme and Makeig, 2003).

The preponderance of coherence of all sorts observed between pairs of scalp channel signals is accounted for by ICA as deriving from common IC projections to both scalp channels. A change in amplitude of a single IC, relative to other ICs that project to the same channel pair, may produce a change in their measured (zero-lag) scalp channel coherence *without* any actual coherence changes occurring at the cortical source level. By maximally reducing the effects of volume conduction on the data, ICA decomposition allows a more principled study of transient or intermittent coherence between IC source activities.

Recently, we used ICA decomposition of target response data from the same five-box task used here for illustrative purposes to show that brief and weakly spatially coherent theta wave complexes arise in frontal midline, somatomotor, and parietal cortex in many subjects following significant events (Makeig et al., 2004b), often beginning in frontal polar cortex (Delorme et al., 2007b). But how can the activities of 'independent' components be (occasionally) phase coherent? As described previously, ICA decomposition actually derives *maximally* independent components – this allows the discovered IC activity patterns to exhibit occasional transient dependence – for example, in the five-box data at one frequency in at most a fifth of the trials. In this and related cases, we found the partial coherences to have non-zero phase lags, and to remain when each component ERP was (artificially) regressed out of each single trial activity and coherence was computed only on the remainder. Event-related coherence is another measure that cannot be deduced from ERP waveforms alone, and cannot be confidently interpreted when computed for pairs of scalp-channel signals. ICA preserves only those coherences that represent transient coupling of the frequency-domain activities of two EEG sources with a fixed latency difference.

A 'close up' example of similar 'phase reorganization' in human brain was recently provided by the study by Wang and colleagues of event-related local field activity in multi-channel 'thumbtack' electrodes pushed through a small piece of intact cortex in anterior cingulate before its clinically required removal in a brain operation (Wang et al., 2005). They reported that theta band activity was generated in superficial layers of anterior cingulate cortex

(ACC) both before and after presentations of a variety of task-relevant stimuli, while after presentations phase-locking between ACC and other brain areas increased transiently.

V. Meeting the Challenge of the Moment

For us to survive and thrive, at each moment our brain must integrate its awareness of its present situation and environment, including existing plans for action and/or inaction, with its emerging sensory experience and mnemonic associations. It must optimally engage or revise its attentional distribution, action plans, and physiological body state in a way adequate t*o meet the challenge of the moment.*

This volume summarizes the results of nearly fifty years of scientific experience in studying the shapes and sizes of average event-related potential (ERP) responses of scalp EEG signals to sensory or other events, responses that depend in large part on the *significance* of the events to the subject and on the *context* in which they occur. EEG is the oldest and most non-invasive functional brain imaging modality; it is also the least expensive and most highly portable. The continuing promise of EEG brain imaging is that the highly labile dynamics of EEG scalp fields, signaling changes in local field synchrony within and between cortical areas, can provide detailed indices of changes in human attentional, intentional, and affective state, both *post hoc* and even, to an increasing extent, online, with potentially important applications to basic scientific research, to clinical and workplace monitoring, and to other fields of human interest and endeavor.

In this chapter, we have discussed the origins in *local cortical synchrony* of both EEG signals and ERP waveforms derived from them. We have defined the concept of an EEG source, based on both EEG analysis and physiological evidence, and have demonstrated the utility of independent component analysis (ICA) for separating multi-channel EEG recordings into a set of temporally and functionally independent brain and non-brain source processes. Finally, we have shown a simple example of using ICA decomposition to study the sources that contribute to (as well as those that contaminate) ERPs, and their activities in the single-trial EEG data. We have given examples of using ERP-image plotting to visualize the dependence of EEG responses in single trials on behavioral, EEG, or other parameters, have introduced time/frequency analysis in the form of inter-trial coherence (ITC) to show that the activity captured in average ERPs reflects trial-to-trial phase consistency, and have introduced the concept that *some* ERP features may reflect reorganization (or *perturbation*) of the exact timing or phase statistics of *ongoing* activity in the same cortical areas, as long suggested by investigators familiar with dynamic modeling methods used in engineering.

We believe the increasingly urgent challenge for the field of ERP and more general EEG research is to discover the brain source dynamics that produce the characteristic features of

evoked responses and to model the trial-by-trial (and condition-by-condition) differences in EEG (and ERP) dynamics associated with the large variety of events that unfold continually in our daily lives, within an ever-evolving situational context – events that pose a wide variety of challenges to which our brains respond effectively.

We believe this to be an exciting time to study human electrophysiology, an era in which non-invasive EEG recording is moving toward fulfilling its promise of becoming a true functional brain imaging modality. Current knowledge and understanding of EEG dynamics is likely to advance steadily as new analysis tools developed for this purpose become more widely applied. One result should be a deeper and fuller understanding of the nature and significance of ERP features[9].

Endnotes

[1]For example, at 20 Hz and traveling at 1 m/s, a radiating 'pond ripple' would reach the edge of a 1-cm source domain in 5 ms, one tenth of a 20-Hz cycle. Therefore, there would only be a $2\pi/10$ = 36 degree phase lag between the center and the edge of the patch, and the spatiotemporal pattern of potentials at scalp electrodes would be highly correlated with the pattern produced by completely synchronous 20-Hz activity across the same 1-cm domain.

[2]Though less commonly appreciated, intracranial electrodes also record volume-conducted signals from distant sources along with local field activity produced just under the electrode (Lee, 2005)**.**

[3] These EEG data were collected synchronously from 250 scalp plus four infra-ocular and two electrocardiographic (ECG) electrodes with an active reference (Biosemi, Amsterdam) at a sampling rate of 256 Hz and 24-bit A/D resolution. Onsets and offsets of target discs, as well subject button presses, were recorded in a simultaneously acquired event channel. The recording montage covered most of the skull, forehead, and lateral face surface, omitting chin and fleshy cheek areas. Locations of the electrodes relative to skull landmarks for each subject were recorded (Polhemus, Inc.). Electrodes with grossly abnormal activity patterns were removed from the data, leaving 238 channels. After re-referencing to digitally linked mastoids, the data were digitally filtered to emphasize frequencies above 1 Hz. Data periods containing broadly distributed, high-amplitude muscle noise and other irregular artifacts were identified by tests for high kurtosis or low-probability activity and removed from analysis. Occurrence of eye blink, other eye movement, or isolated muscle noise artifact was not a criterion for rejection. Remaining data time points were then concatenated and submitted to decomposition by extended infomax ICA using the binica function available in the EEGLAB toolbox (http://sccn.ucsd.edu/eeglab). Decompositions used extended-mode infomax ICA (Makeig et al., 1997) with default training parameters. Extended infomax was used to allow recovery of any components with sub-gaussian activity distributions, including 60-Hz line noise contamination. ICA components clearly and predominantly accounting for eye movement, muscle, cardiac, single-channel, or other artifactual activity were removed from the ERP data. Both the target stimulus-locked and motor response-locked epochs analyzed in the figures were referred to a mean baseline in a 500-ms period before target stimulus onsets.

[4]Data figures in this chapter were produced using software tools from the freely available EEGLAB Matlab software environment (sccn.ucsd.edu/eeglab/). The single-subject 256-channel data set from which we derived most of the figures was recorded and first studied by Delorme et al. (Delorme et al., 2007b) and is available for download in raw and in EEGLAB formats from the EEGLAB web site (above).

[5]In particular, the phase of a digitally recorded signal cannot be defined above its Nyquist frequency (half of its sampling rate) and is ambiguous at its Nyquist frequency.

[6]Methods that find more components are available, but require narrower source assumptions and more computation time.

[7]EEGLAB includes Matlab-based tools for applying, evaluating, and exploring component clustering.

[8]Data used for the IC cluster figure were collected by Klaus Gramann at the University of Munich from 12 subjects performing a visual feature discrimination task. The electroencephalogram (EEG) was recorded continuously at a sampling rate of 500 Hz using 64 Ag/AgCl electrodes mounted on an elastic cap. EEG signals were amplified using a 0.1–100-Hz bandpass filter and filtered off-line using a 1–40-Hz bandpass. All electrodes were recorded referenced to Cz and then re-referenced off-line to linked mastoids. Average ERPs in an 800-ms epoch were computed relative to a 200-ms pre-stimulus baseline. ICA decomposition used extended infomax. ICs were clustered across subjects using EEGLAB clustering functions based on their respective dynamics under three target stimulus-difference conditions (whether or not the target had a different color, different shape, or both than the accompanying standard stimuli). Only ICs whose equivalent dipole projection to the scalp had a residual variance from the IC scalp map below 15% and an equivalent dipole location within the brain volume were considered for clustering. These ICs were separated into 22 clusters based on their equivalent dipole locations, event-related spectral perturbations (ERSPs) and inter-trial coherences (ITCs) in the 500 ms following target onsets.

[9]Thanks to Arnaud Delorme and David Groppe for helpful discussions, and to our many colleagues from the Swartz Center for Computational Neurosciences, UCSD and Computational Neurobiology Laboratory, Salk Institute for their invaluable support, insight, and companionship on our long and continuing quest to better understand human event-related brain dynamics.

References

Akalin-Acar Z, Gencer NG (2004) An advanced boundary element method (BEM) implementation for the forward problem of electromagnetic source imaging. Phys Med Biol 49:5011-5028.

Anemuller J, Sejnowski TJ, Makeig S (2003) Complex independent component analysis of frequency-domain electroencephalographic data. Neural Netw 16:1311-1323.

Arieli A, Shoham D, Hildesheim R, Grinvald A (1995) Coherent spatiotemporal patterns of ongoing activity revealed by real-time optical imaging coupled with single-unit recording in the cat visual cortex. J Neurophysiol 73:2072-2093.

Beggs JM, Plenz D (2003) Neuronal avalanches in neocortical circuits. J Neurosci 23:11167-11177.

Bell AJ, Sejnowski TJ (1995) An information-maximization approach to blind separation and blind deconvolution. Neural Comput 7:1129-1159.

Bollimunta A, Chen Y, Schroeder CE, Ding M (2008) Neuronal mechanisms of cortical alpha oscillations in awake-behaving macaques. J Neurosci 28:9976-9988.

Bullock TH (1983) Electrical signs of activity in assemblies of neurons: compound field potentials as objects of study in their own right. Acta Morphol Hung 31:39-62.

Comon P (1994) Independent Component Analysis, a New Concept. Signal Processing 36:287-314.

Czeisler CA, Allan JS, Strogatz SH, Ronda JM, Sanchez R, Rios CD, Freitag WO, Richardson GS, Kronauer RE (1986) Bright light resets the human circadian pacemaker independent of the timing of the sleep-wake cycle. Science 233:667-671.

Dan Y, Poo MM (2004) Spike timing-dependent plasticity of neural circuits. Neuron 44:23-30.

Debener S, Mullinger KJ, Niazy RK, Bowtell RW (2008) Properties of the ballistocardiogram artefact as revealed by EEG recordings at 1.5, 3 and 7 T static magnetic field strength. Int J Psychophysiol 67:189-199.

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 including independent component analysis. J Neurosci Methods 134:9-21.

Delorme A, Sejnowski T, Makeig S (2007a) Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. Neuroimage 34:1443-1449.

Delorme A, Westerfield M, Makeig S (2007b) Medial prefrontal theta bursts precede rapid motor responses during visual selective attention. J Neurosci 27:11949-11959.

Destexhe A, Contreras D, Steriade M (1999) Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states. J Neurosci 19:4595-4608.

Dien J, Beal DJ, Berg P (2005) Optimizing principal components analysis of event-related potentials: matrix type, factor loading weighting, extraction, and rotations. Clin Neurophysiol 116:1808-1825.

Fischl B, Salat DH, van der Kouwe AJ, Makris N, Segonne F, Quinn BT, Dale AM (2004) Sequence-independent segmentation of magnetic resonance images. Neuroimage 23 Suppl 1:S69-84.

Fiser J, Chiu C, Weliky M (2004) Small modulation of ongoing cortical dynamics by sensory input during natural vision. Nature 431:573-578.

Foxe JJ, Schroeder CE (2005) The case for feedforward multisensory convergence during early cortical processing. Neuroreport 16:419-423.

Francis JT, Gluckman BJ, Schiff SJ (2003) Sensitivity of neurons to weak electric fields. J Neurosci 23:7255-7261.

Freeman WJ (2000) Mesoscopic neurodynamics: from neuron to brain. J Physiol Paris 94:303-322.

Freeman WJ, Barrie JM (2000) Analysis of spatial patterns of phase in neocortical gamma EEGs in rabbit. J Neurophysiol 84:1266-1278.

Fries P, Nikolic D, Singer W (2007) The gamma cycle. Trends Neurosci 30:309-316.

Frost DO, Caviness VS, Jr. (1980) Radial organization of thalamic projections to the neocortex in the mouse. J Comp Neurol 194:369-393.

Galambos R (1992) A comparison of certain gamma band (40-Hz) brain rhythms in cat and man. In: Induced Rhythms in the Brain (Basar E. BTH, ed), pp 201-216. Boston: Birkhauser.

Gencer NG, Akalin-Acar Z (2005) Use of the isolated problem approach for multi-compartment BEM models of electro-magnetic source imaging. Phys Med Biol 50:3007-3022.

Grave de Peralta-Menendez R, Gonzalez-Andino SL (1998) A critical analysis of linear inverse solutions to the neuroelectromagnetic inverse problem. IEEE Trans Biomed Eng 45:440-448.

Grinvald A, Lieke EE, Frostig RD, Hildesheim R (1994) Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. J Neurosci 14:2545-2568.

Halgren E, Marinkovic K, Chauvel P (1998) Generators of the late cognitive potentials in auditory and visual oddball tasks. Electroencephalogr Clin Neurophysiol 106:156-164.

Hanslmayr S, Klimesch W, Sauseng P, Gruber W, Doppelmayr M, Freunberger R, Pecherstorfer T, Birbaumer N (2007) Alpha phase reset contributes to the generation of ERPs. Cereb Cortex 17:1-8.

Honma K, Honma S, Wada T (1987) Phase-dependent shift of free-running human circadian rhythms in response to a single bright light pulse. Experientia 43:1205-1207.

Iyer VK, Ploysongsang Y, Ramamoorthy PA (1990) Adaptive filtering in biological signal processing. Crit Rev Biomed Eng 17:531-584.

Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ (2000a) Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. Clin Neurophysiol 111:1745-1758.

Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ (2001) Analysis and visualization of single-trial event-related potentials. Hum Brain Mapp 14:166-185.

Jung TP, Makeig S, Humphries C, Lee TW, McKeown MJ, Iragui V, Sejnowski TJ (2000b) Removing electroencephalographic artifacts by blind source separation. Psychophysiology 37:163-178.

Jutten C, Herault J (1991) Blind Separation of Sources .1. An Adaptive Algorithm Based on Neuromimetic Architecture. Signal Processing 24:1-10.

Jutten C, Karhunen J (2004) Advances in blind source separation (BSS) and independent component analysis (ICA) for nonlinear mixtures. Int J Neural Syst 14:267-292.

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.

Lee I, Worrell, G, Makeig, S (2005) Relationships between concurrently recorded scalp and intracranial electrical signals in humans. Human Brain Mapping Abstracts.

Lee TW, Lewicki, M.S. (2000) The generalized Gaussian mixture model using ICA. International Workshop on Independent Component Analysis.

Linkenkaer-Hansen K, Nikouline VV, Palva JM, Ilmoniemi RJ (2001) Long-range temporal correlations and scaling behavior in human brain oscillations. J Neurosci 21:1370-1377.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature 412:150-157.

Luck SJ (2005) An introduction to event-related potentials and their neural origins. In: An Introduction to the Event-Related Potential Technique, pp 1-50: The MIT Press.

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, Debener S, Onton J, Delorme A (2004a) Mining event-related brain dynamics. Trends Cogn Sci 8:204-210.

Makeig S, Jung TP, Bell AJ, Ghahremani D, Sejnowski TJ (1997) Blind separation of auditory event-related brain responses into independent components. Proc Natl Acad Sci U S A 94:10979-10984.

Makeig S, Westerfield M, Townsend J, Jung TP, Courchesne E, Sejnowski TJ (1999a) Functionally independent components of early event-related potentials in a visual spatial attention task. Philos Trans R Soc Lond B Biol Sci 354:1135-1144.

Makeig S, Westerfield M, Jung TP, Covington J, Townsend J, Sejnowski TJ, Courchesne E (1999b) Functionally independent components of the late positive event-related potential during visual spatial attention. J Neurosci 19:2665-2680.

Makeig S, Westerfield M, Jung TP, Enghoff S, Townsend J, Courchesne E, Sejnowski TJ (2002) Dynamic brain sources of visual evoked responses. Science 295:690-694.

Makeig S, Delorme A, Westerfield M, Jung TP, Townsend J, Courchesne E, Sejnowski TJ (2004b) Electroencephalographic brain dynamics following manually responded visual targets. PLoS Biol 2:e176.

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.

Mardia KV (1972) Statistics of directional data. New York, NY: Academic Press.

Martinez-Montes E, Cuspineda-Bravo ER, El-Deredy W, Sanchez-Bornot JM, Lage-Castellanos A, Valdes-Sosa PA (2008) Exploring event-related brain dynamics with tests on complex valued time-frequency representations. Stat Med 27:2922-2947.

Massimini M, Huber R, Ferrarelli F, Hill S, Tononi G (2004) The sleep slow oscillation as a traveling wave. J Neurosci 24:6862-6870.

Mazaheri A, Jensen O (2006) Posterior alpha activity is not phase-reset by visual stimuli. Proc Natl Acad Sci U S A 103:2948-2952.

Meinecke F, Ziehe A, Kawanabe M, Muller KR (2002) A resampling approach to estimate the stability of one-dimensional or multidimensional independent components. IEEE Trans Biomed Eng 49:1514-1525.

Murre JM, Sturdy DP (1995) The connectivity of the brain: multi-level quantitative analysis. Biol Cybern 73:529-545.

Nunez P (1977) The dipole layer as a model for scalp potentials. TIT J Life Sci 7:65-72.

Nunez P, Srinivasan, R. (2005) Electric fields of the brain; The neurophysics of EEG Oxford University Press.

Onton J, Makeig S (2006) Information-based modeling of event-related brain dynamics. Prog Brain Res 159:99-120.

Onton J, Westerfield M, Townsend J, Makeig S (2006) Imaging human EEG dynamics using independent component analysis. Neurosci Biobehav Rev 30:808-822.

Onton J, Makeig,S. (2005) Independent Component Analysis (ICA) source locations vary according to task demands. Organization for Human Brain Mapping Abstracts.

Oostenveld R, Oostendorp TF (2002) Validating the boundary element method for forward and inverse EEG computations in the presence of a hole in the skull. Hum Brain Mapp 17:179-192.

Palmer JA, Makeig, S., Kreutz-Delgado, K. Rao, B.D. (2008) Newton Method for the ICA Mixture Model. ICASSP:1805-1808.

Pfurtscheller G, Aranibar A (1977) Event-related cortical desynchronization detected by power measurements of scalp EEG. Electroencephalogr Clin Neurophysiol 42:817-826.

Radman T, Parra L, Bikson M (2006) Amplification of small electric fields by neurons; implications for spike timing. Conf Proc IEEE Eng Med Biol Soc 1:4949-4952.

Ranganath C, Rainer G (2003) Neural mechanisms for detecting and remembering novel events. Nat Rev Neurosci 4:193-202.

Rulkov NF, Timofeev I, Bazhenov M (2004) Oscillations in large-scale cortical networks: map-based model. J Comput Neurosci 17:203-223.

Sayers BM, Beagley HA, Henshall WR (1974) The mechansim of auditory evoked EEG responses. Nature 247:481-483.

Scherg M (1990) Fundamentals of dipole source potential analysis. Advances in Audiology 6:40-69.

Schroeder CE, Mehta AD, Givre SJ (1998) A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. Cereb Cortex 8:575-592.

Stettler DD, Das A, Bennett J, Gilbert CD (2002) Lateral connectivity and contextual interactions in macaque primary visual cortex. Neuron 36:739-750.

Swadlow HA, Gusev AG (2000) The influence of single VB thalamocortical impulses on barrel columns of rabbit somatosensory cortex. J Neurophysiol 83:2802-2813.

Tallon-Baudry C, Bertrand O, Delpuech C, Pernier J (1996) Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. J Neurosci 16:4240-4249.

Tass P (1999) Phase Resetting in Medicine and Biology: Stochastic Modelling and Data Analysis Springer.

Voronin LL, Volgushev M, Sokolov M, Kasyanov A, Chistiakova M, Reymann KG (1999) Evidence for an ephaptic feedback in cortical synapses: postsynaptic hyperpolarization alters the number of response failures and quantal content. Neuroscience 92:399-405.

Wang C, Ulbert I, Schomer DL, Marinkovic K, Halgren E (2005) Responses of human anterior cingulate cortex microdomains to error detection, conflict monitoring, stimulus-response mapping, familiarity, and orienting. J Neurosci 25:604-613.

Winfree AT (1980) The geometry of biological time. Biomathematics 8.

Worden MS, Foxe JJ, Wang N, Simpson GV (2000) Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. J Neurosci 20:RC63.