Independent Component Analysis of Event-Related Potentials

David M. Groppe[1,*], Scott Makeig[2], Marta Kutas[1]

[1] Department of Cognitive Science,
[2] Swartz Center for Computational Neuroscience
University of California, San Diego, USA

[*] Corresponding author: dgroppe@cogsci.ucsd.edu

Independent component analysis (ICA) is a potentially powerful tool for analyzing event-related potentials (ERPs), one of the most popular measures of brain function in cognitive neuroscience. Based on the statistics of the electroencephalogram (EEG), from which ERPs are derived, ICA may be able to extract multiple, functionally distinct sources of an ERP generated by disparate regions of cerebral cortex. Extracting such sources greatly increases the informativeness of ERPs by providing a cleaner, less ambiguous measure of source activity and by facilitating the identification of this activity across different experimental paradigms. The main purpose of this review article is to explain the logic of ICA, to illustrate how ICA could in principle extract spatiotemporally overlapping ERP sources, and to review evidence that ICA is a well motivated methodology that can extract latent ERP sources in practice. In addition, we close the article by noting potential problems with ICA and by comparing it to three alternative methods for extracting ERP sources/components: spatial principal component analysis, source localization, and temporal principal component analysis.

Due to their impeccable temporal resolution, noninvasiveness, and affordability, event-related potentials (ERPs) have been and continue to be one of the most popular measures of brain activity in cognitive neuroscience. ERP research has provided valuable insight into many domains of brain function (e.g., attention — Martinez et al., 1999, visual object recognition — Johnson & Olshausen, 2003, and language comprehension — Kutas, Van Petten, & Kluender, 2006). However, the informativeness of ERPs (and, more generally, the electro-encephalogram from which ERPs are derived) is limited by the fact that they are most likely generally generated by multiple, functionally distinct neural sources whose scalp potentials overlap temporally and spatially (Dale et al., 2000). This spatiotemporal overlap confounds our ability to measure those sources (e.g., Hagoort, 2003) and makes it difficult to compare ERP phenomena related to different types of stimuli or experimental paradigms (e.g., linguistic and non-linguistic stimuli — Coulson, King, & Kutas, 1998; Osterhout & Hagoort, 1999).

One method for dealing with the problem of spatiotemporally overlapping ERP sources is independent component analysis (ICA). As typically applied to ERP data sets, ICA decomposes the data into a set of independent components (ICs) via a basis of linear spatial

filters (LSFs). To the extent that ICs correspond to latent ERP sources, ICA provides a cleaner, less ambiguous measure of these sources and allows their identification across different experimental paradigms. The main purpose of this review article is to illustrate how ICA could in principle extract spatiotemporally overlapping ERP sources, to briefly explain how ICA learns a basis of LSFs for a particular data set, and to review evidence that ICA is a well-motivated methodology that can extract latent ERP sources in practice. In addition, we close the article by noting potential problems with ICA and by comparing it to three alternative methods for extracting ERP sources/components: spatial principal component analysis, source localization, and temporal principal component analysis.

2. Sources of Scalp Potentials

To understand why LSFs are useful for ERP analysis, one must first understand the relationship between electric potentials recorded at the scalp and their neural and non-neural generators.

2.1 Neural Sources

The electrical activity of the brain that is measurable at the scalp is called the electro-encephalogram (EEG). Event-related potentials (ERPs) are the portion of the EEG whose phase is related to the onset of a class of events and is typically derived by averaging multiple samples of EEG time-locked to the occurrence of such events (e.g., the onset of a visual stimulus or a button press). Much of the brain's electrical activity is practically invisible at the scalp (due to the small magnitude of the activity or interactions between sources of opposite polarity — Kutas & Dale, 1997). Indeed, it is generally believed that the great bulk of the EEG is generated by the synchronous activation of post-synaptic potentials of adjacent cortical pyramidal cells (Baillet, Mosher, & Leahy, 2001; Kutas & Dale, 1997). This activity can be accurately approximated as a finite set of equivalent current dipoles (ECD) or multipoles (ibid.). In this framework, the EEG is the linear combination of the electrical activity of a discrete set of neural electrical sources (ibid.).

To illustrate, consider a situation in which there is only a single neural source[1], y_1, and two scalp electrodes, x_1 and x_2. The voltage recorded at each scalp electrode is simply a scaled version of the source's voltage (Figure 1). The polarity and magnitude of the scaling is determined by the location and orientation of the source relative to the electrode and the conductive properties of the medium between the source and the electrode. This relationship can be represented with a linear function where the source voltage is multiplied by a "gain vector," **g**, to obtain the scalp voltages:

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} .04 \\ -.15 \end{bmatrix} y_1(t) = \mathbf{g} y_1(t) \qquad (1)$$

As long as the position of the source relative to the electrodes and the conductive properties of the head do not change, this linear relationship will remain constant for each time point, *t*. Note that the gain vector defines the scalp topography of a source. With only one source, the topography of the scalp potentials remains constant (Figure 2: B & C).

To make our example slightly more complicated, let us add two more sources. Equation 1 becomes:

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} .04 & -.08 & -.05 \\ -.15 & -.04 & .01 \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{bmatrix} = \mathbf{G}\mathbf{y}(t) \quad (2)$$

where the three gain vectors form a gain matrix, **G**. Now each scalp electrode measures the sum of the

[1] "Neural source" refers to highly correlated patches of cerebral cortex. This could be a single localized patch of cerebral cortex that could be accurately modeled as a single ECD or it could be a distributed set of patches that would require multiple ECDs to be accurately modeled.

scaled voltages of the three sources (Figure 3). As before, the scalp topography of each source is static and defined by its gain vector. However, as there is more than one source and the sources are at least somewhat independent and have distinct topographies, the topographies of the individual sources combine such that the topography of the scalp potentials changes over time (Figure 2: D).

In reality, the neural sources of a segment of EEG are likely to be far more numerous than three. Assuming that a patch of cortex on the order of 25 mm^2 is the smallest neural unit whose electrical activity is practically measurable from the scalp (Baillet et al., 2001) and that the surface area of a hemisphere of human cerebral cortex is approximately 80,000 mm^2 (Van Essen, Drury, Joshi, & Miller, 1998), the number of EEG sources could be as high as 6,400. However, the number of sources contributing to an ERP is probably much smaller than this upper bound. The activity of multiple patches of cortex may be correlated enough that they can be approximated as a single source.

Moreover, the activity of some sources may not be sufficiently phase-dependent on the onset of the event used to derive the ERP to contribute to the ERP or may be cancelled out by sources of opposite polarity. In addition, some sources may be too weak and too far from scalp electrodes to contribute significantly (the strength of a source potential falls off as an inverse function of distance squared — Kutas & Dale, 1997). In fact, many ERP phenomena can be modeled with only a handful of sources (Groppe, 2007). Unfortunately, it is impossible to know with certainty how many sources generated a particular segment of an ERP. For example, multiple ERP sources with similar scalp distributions may be inaccurately modeled by fewer sources (i.e., a few model sources could fit the data very well, even though, in reality, a greater number of sources generated the data). On the other hand, a few ERP sources can be inaccurately modeled by too many sources (Liu, Dale, & Belliveau, 2002).

2.2 EEG Artifacts

Attempts to measure ERPs are complicated by the fact that the brain is not the only source of the electric potentials recorded at the scalp and by technical disturbances (Talsma & Woldorff, 2005). Phenomena such as blinks, eye movements, and head muscle activity generate electrical potentials at amplitudes that can be orders of magnitude larger than ERP amplitudes and can seriously confound ERP analysis. Sensor noise, caused by poor connections between the scalp and electrodes and 60 Hz line noise, also commonly pollute EEG recordings.

The effect of many of these phenomena on recorded scalp potentials can be well approximated by the same framework used to model the EEG: each artifact is modeled as a set of discrete electrical sources with fixed scalp distributions whose activity combines linearly at the scalp with other electrical sources. In terms of Equation 2, each artifact can be approximated by one or more elements in the source vector, **y**, and a corresponding number of columns in the gain matrix, **G**.

For example, blink potentials (Ille, Berg, & Scherg, 2002; Jung, Makeig, Humphries et al., 2000; Jung, Makeig, Westerfield et al., 2000), scalp muscle activity (Jung, Makeig, Humphries et al., 2000), and 60 Hz line noise (Jung, Makeig, Humphries et al., 2000; Tang, Sutherland, & McKinney, 2005) each can be well approximated as a single source. Eye movements can be successfully modeled by two sources, one each for movement in the horizontal and vertical planes (Ille et al., 2002).

3 Linear Spatial Filters

Given that the electric potentials of different EEG and EEG artifact sources linearly combine and can have distinct topographies, linear spatial filters are a potentially powerful tool for teasing

apart these different sources despite their spatiotemporal overlap. A linear spatial filter (LSF) is a weighted sum of the potentials at each electrode. For example, the vector, **w**, below would be an LSF for two electrode data:

$$u(t) = \begin{bmatrix} .06 & -.14 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \mathbf{w}^T \mathbf{x}(t) \qquad (3)$$

The weights to each electrode define the LSF and determine the output, or "activation," of the filter, *u*. While each electrode measures a weighted sum of the electrical activity of neural and EEG artifact sources, the LSF provides a re-weighted sum of those sources and functions like a "virtual electrode" (Spencer, Dien, & Donchin, 2001). Figure 4 (A) plots the activation of the LSF from Equation 3 when input the scalp-recorded data from Figure 3. The activation is very similar to the activity of Source y_1 (r_{cos}=0.89)[2] but smaller in amplitude. In other words, the filter sums the activity of Source y_2 and Source y_3 with small weights and Source y_1 with a relatively large weight.

We can compute the weight that an LSF, **w**, assigns to a source with gain vector, **g**, from the following equation:

$$\cos(\theta)|\mathbf{w}||\mathbf{g}| = \mathbf{w}^T\mathbf{g} \qquad (4)$$

where θ is the angle between the filter and source gain vectors, |**w**| is the length of the filter vector, and |**g**| is the length of the source's gain vector. The first term in this equation, cos(θ), takes a value between -1 and 1. If the filter and source gain vectors are perpendicular (i.e., θ=90 or 270 degrees), then cos(θ) is zero and the filter will not pick up the source's activity at all. As the filter and source gain vectors come closer together, the angle between them shrinks and cos(θ) approaches one. As the angle between the two gain vectors approaches 180 degrees, cos(θ) approaches negative one. The first term interacts with the final term, |**g**|, to make the filter more sensitive to some sources than others. The remaining term in the equation, |**w**|, simply scales the output of the filter and will affect the weights between the filter and all sources equally.

In light of Equation 4, we can now understand why our example filter in Figure 4 works as it does by visualizing the filter and source gain vectors in "electrode space" (Figure 4: B). The filter is nearly perpendicular to the gain vectors of Source y_2 and Source y_3, and nearly parallel to the gain vector of Source y_1. Hence the filter gives greater weight to Source y_1 than to the other sources. The angle between the filter and Source y_1's gain vector is acute, hence the filter picks up Source y_2's activity with the same polarity. Finally, the filter is short, which contributes to the amplitude of the filter output being much smaller than the amplitude of Source y_1.

3.1 LSF Limitations and Difficulties

Note that in this example, the filter cannot perfectly capture one of the sources and completely filter out the other two. That would require the filter to be perpendicular to two of the sources and non-perpendicular to the third. When there are more than *n* sources (where *n* is the number of electrodes) with linearly independent gain vectors[3], it is impossible to perfectly filter out the activity of one source

[2] r_{cos} is the "cosine similarity" between two vectors (i.e., the cosine of the angle between the two vectors) and is analogous to Pearson's linear correlation coefficient, *r*. An r_{cos} of 1 indicates that two vectors are identical save for a possible difference in magnitude (i.e., they point in the same direction). An r_{cos} of — 1 indicates that two vectors point in exactly opposite directions and an r_{cos} of 0 indicates that the vectors are orthogonal. More specifically, the cosine similarity between two vectors **x** and **y** is $\mathbf{x}^T\mathbf{y}/(|\mathbf{x}||\mathbf{y}|)$.

[3] A set of linearly independent gain vectors means that none of the gain vectors equals a weighted sum of the other gain vectors. For example, if there are two sources and two electrodes, the gain vectors are linearly independent if they are not parallel.

from that of all the other sources[4]. Given that the measurements of an electrode will be polluted by a degree of sensor noise that will be somewhat independent of the sensor noise at other electrodes (e.g., Section 4.1), there will always be a set of *n* linearly independent sources in addition to any neural or other artifact sources. Thus, it will be impossible to perfectly isolate any single source. While steps can be taken to minimize this sensor noise (e.g., averaging multiple samples of the EEG responses to the same class of events), there may still be more non-negligible neural and EEG artifact sources than electrodes and these may force a degree of filter error, sometimes called "misallocation of variance" (Wood & McCarthy, 1984). Also notice that if the gain vectors of two of the sources were parallel (i.e., the angle between the vectors was 0 or 180 degrees), that no spatial filter would be able to differentiate between the two of them. Spatial filters require topography differences to separate sources.

While it is important to keep these possible LSF limitations in mind, in practice, LSFs may be able to quite accurately extract sources with large activations or gain vectors especially when numerous electrodes are used to measure the EEG. When a source's activity or gain vector is large, its contribution to the LSF output can effectively overpower that of other sources that the LSF is not able to ignore (Ghahremani, Makeig, Jung, Bell, & Sejnowski, 1996; Kobayashi, James, Nakahori, Akiyama, & Gotman, 1999; Makeig, Jung, Ghahremani, & Sejnowski, 2000). Moreover, with each additional electrode, an LSF gains a free parameter that can improve (or at worst, leave the same) the LSF's ability to extract a particular source. Additional electrodes also reduce the likelihood that multiple sources have indistinguishable topographies[5].

Indeed, the greatest difficulty with using LSFs to extract EEG sources is probably not these fundamental limitations. Rather, it is knowing how to derive an LSF that accurately extracts large EEG sources of interest. This difficulty stems from the fact that it is impossible to know, with certainty, what the sources of a particular set of scalp potentials are. As there are possibly many more scalp potential sources than electrodes, there are an infinite number of source configurations that could have generated a given segment of EEG data (this is known as the "EEG inverse problem" — Kutas & Dale, 1997). Because the exact source configuration is uncertain, it is not clear what the latent sources are to be extracted let alone what LSF, if any, could extract them.

To deal with this problem, it is necessary to make assumptions about the sources of an EEG data set in order to derive LSFs that might extract some of these sources. The following three sections review three of the most popular methods for deriving LSFs from EEG data sets, each based on different source assumptions. The first two of these, independent component analysis (ICA) and principal component analysis (PCA), derive LSFs primarily from statistical assumptions. In contrast, the lattermost class of techniques, source localization, relies primarily on anatomical models to derive LSFs. The emphasis of the review will be on ICA.

[4] For an LSF, **w**, to perfectly filter out one source from *n* other sources whose gain vectors ($\mathbf{g}_i$, $1 \leq i \leq n$) are linearly independent, the following would have to be true: $\mathbf{w}^T\mathbf{g}_i=0$ for all $1 \leq i \leq n$. However, if this were true, the inner product of **w** and any possible gain vector would be zero (i.e., **w** would be the zero vector) since the *n* gain vectors of the unwanted sources span the space of possible vectors (when there are *n* electrodes). Thus **w** would not be able to extract any sources.

[5] If two sources have different topographies, **g** and **g'**, then there must be at least two locations on the scalp such that $g_1/g'_1 \neq g_2/g'_2$. Each additional electrode increases the probability that one records from such a pair of scalp locations.

4. Independent Component Analysis for Deriving LSFs

ICA is a blanket term for a family of algorithms that learns a set of linear filters from the statistics of a data set by trying to make the output of the filters temporally independent[6] or relatively so (Hyvärinen et al., 2001). The logic of ICA, as typically applied to EEG data, is clear when formalized as an attempt to fit a linear generative model to a time series of observations (Roweis & Ghahramani, 1999):

$$\mathbf{x}(t) = \mathbf{A}\mathbf{u}(t) \quad (5)$$

where **u** is an *n*-dimensional vector of independent sources whose activity linearly, instantaneously combines via a full rank "mixing matrix," **A**, to create the *n*-dimensional[7] observations, **x**. When used to model EEG data, **A** is a model of the EEG gain matrix, **G** (i.e., each column of **A** defines the topography of a putative source). The problem for ICA is then to learn **A** or equivalently, the "unmixing matrix," **W**, which is the inverse of **A** and consists of a basis of *n* LSFs that extracts the latent source activity.

$$\mathbf{u}(t) = \mathbf{W}\mathbf{x}(t) = \mathbf{A}^{-1}\mathbf{x}(t) \quad (6)$$

The intuitive reason why it should be possible to learn such "unmixing" LSFs is that the individual dimensions (i.e., scalp electrodes in the context of EEG) of a mixture of independent sources will generally be somewhat dependent, since the activity of a single source contributes to multiple dimensions. Thus, by trying to unmix the data in a way that minimizes these temporal dependencies, ICA may be able to recover the latent sources (e.g., Figure 5). The specifics of how this unmixing is done differ from ICA algorithm to algorithm. Some ICA algorithms find **W** analytically by estimating statistics of **x** (e.g., second-order blind identification — Belouchrani, Abed Meraim, Cardoso, & Moulines, 1997) while others iteratively improve the fit between an estimate of **u** and the assumed generative model (e.g., extended INFOMAX ICA — Lee, Girolami, & Sejnowski, 1999).

To learn **W**, all ICA algorithms have to assume that the sources are somehow non-Gaussian or are somehow non-white (i.e., the activity at each time point is not correlated with the activity at any other time point), because if the sources are white Gaussian, then there are an infinite number of unmixing matrices that will make the unmixed data independent and the problem is under-constrained[8] (Bell & Sejnowski, 1995). The specifics of these assumptions differ across ICA algorithms. For example, extended INFOMAX ICA (Lee et al., 1999), perhaps the most popular ICA algorithm applied to EEG data, assumes independent supergaussian or subgaussian sources but makes no additional assumptions about temporal structure. Supergaussian random variables (e.g., random samples of the potentials

[6] Two random variables, *c* and *d*, are statistically independent if and only if knowing the value of one of the variables provides no information about the value of the other variable. More explicitly, if *c* and *d* are independent, then the conditional probability distribution of *c* given the value of *d* simply equals the probability distribution of *c* (i.e., $P(c|d)=P(c)$) and vice-versa (i.e., $P(d|c)=P(d)$). Independence is related to linear correlation. If two variables are independent then they are uncorrelated. However, two uncorrelated variables are not necessarily independent. A special case is Gaussian random variables, which are independent if they are uncorrelated.

[7] For some ICA algorithms, the dimensionality of the sources and data do not have to be the same, but when ICA is applied to EEG/ERP data it is typically assumed that the number of sources equals the number of electrodes and that **A** is full rank.

[8] Gaussian random variables are independent if they are uncorrelated. For any particular mixture of random variables, one can find an unmixing matrix, **W**, that "spheres" the mixture such that each dimension is uncorrelated and of equal variance. Post-multiplying **W** with an orthonormal matrix **J** will simply rotate the output of **W**. Since a rotated sphere of probability density is indistinguishable from the original sphere, the output of **JW** is also uncorrelated and of equal variance. Since there are an infinite number of orthonormal matrices **J**, there are an infinite number of sphering matrices, **JW**.

generated by occasional bursts of muscle activity or intermittent oscillatory brain activity) have a probability density function (pdf) with a much stronger peak at the mean and heavier tails than a Gaussian distribution with the same variance (Figure 6). A subgaussian random variable (e.g., random samples of a sine wave) has a much flatter pdf around the mean than a Gaussian distribution with the same variance (Figure 6). Alternatively, another ICA algorithm, second-order blind identification (SOBI — Belouchrani et al., 1997), assumes that the sources are uncorrelated at multiple time lags (e.g., the activity of a source at time t is uncorrelated with the activity of another source at time t+100 ms) and have unique normalized spectra, but assumes nothing about the Gaussian nature of the sources.

4.1 An Example of ICA

To illustrate how ICA might be useful for extracting spatiotemporally overlapping ERP sources, consider a hypothetical 12 electrode EEG data set that was generated by 8 sources and 12 mild uncorrelated white Gaussian sensor noise sources (SD=1.25 μV; in comparison the SD of each neural source at the electrode to which it maximally projected was at least 3.42 μV) in two experimental conditions. The average activations of the neural sources (i.e., "source ERPs") in the two conditions are plotted in Figure 7 and the linear correlations between each possible pair of sources are presented in Table 1 (A). The average responses of many sources appear to differ significantly across the two conditions, are generally mildly correlated with one another (median $|r|$=.25) and can be quite correlated. For example, the ERPs of y_5 and y_8 are effectively identical and the ERPs of y_2 are rather correlated to those of y_1 and y_3. However, the strong source ERP correlations do not necessarily reflect the degree of correlation in the single trial activity from which the source ERPs are derived. These correlations are presented in Table 1 (B) and are generally near zero (median $|r|$=.02). Figures 8-9 illustrate how the greater single trial independence is possible. Finally, one pair of sources (y_5 and y_8) continues to exhibit correlated activation even in the single trials because the two ICs are inter-hemisphere homologues of one another.

These sources, plus the 12 sensor noise sources, combine at the scalp via a 20 x 12 gain matrix. Each gain vector is visualized topographically in Figure 10 and the similarity of each possible pair of gain vectors is presented in Table 2. Notice that the gain vectors of the sources generally overlap (median $|r_{cos}|$=.27) and can be quite similar. For example, $g_1$is quite similar to g_2, g_6, and g_7.

The resulting scalp data are plotted in Figure 11. A difference between conditions is most apparent at fronto-central electrodes. Figure 12 visualizes the difference between conditions more clearly by plotting the "difference wave" between the two conditions, the data from Condition Blue minus the data from Condition Red. Based on the statistics of the EEG from both conditions (not shown), extended INFOMAX ICA unmixed the EEG into 12 components. The average activations of these components in the two conditions are plotted in Figure 13. The average activations of the first eight ICs correspond quite closely to those of the eight neural sources (compare to Figure 7; see also Table 3: A—Column 4) although all of the ICs capture a bit of sensor noise. The single trial activations of these ICs are also generally similar to that of the eight neural sources though much less so for ICs 1, 6, and 8 (Table 3: A—Column 5). For the first seven ICs, the less accurate single trial activations stem from the zero-mean sensor noise whose contribution to IC activations is mitigated by averaging (Table 3: B). However, the poor match between the single trial activation of IC 8 and Source 8 reflects the failure of ICA to extract Source 8 because it is highly correlated with its right hemisphere homologue, Source 5 (Table 1: B). The last four ICs correspond to residual sensor noise.

Figure 14 presents the topographies of the 12 ICs. The topographies of the first eight ICs correspond quite closely to those of the eight neural sources

(compare to Figure 10; see also Table 3: A—Column 3). In fact the topographies of ICs and sources are generally more similar than their activations. This illustrates that while it is generally impossible to perfectly extract the activation of a source using an LSF when there are more sources than electrodes, IC topographies can still perfectly match those of EEG sources. The two ICs that most poorly match the topographies of their corresponding sources are ICs 5 and 8. This reflects the fact that the two sources are highly correlated. Indeed, the topography of IC 5 appears to be a blend of the topographies of Sources 5 and 8. The last four ICs correspond to residual sensor noise and have splotchy, non-dipolar topographies (see Footnote 10). Note also that the degree to which an IC's topography matches that of a source does not necessarily correspond to the degree to which that IC's activation matches the activation of that source. For example, the topography of IC 1 is nearly identical to the topography of Source 1 (r_{cos}=.99), which is almost perfect. However, the single trial activation of IC 1 is a relatively mediocre match to that of Source 1 (r_{cos}=.74). One can see how this is possible when one considers that an IC's mixing vector (topography) and unmixing vector (LSF) are somewhat independent. For instance, Figure 15 (B) displays the ideal mixing and unmixing vectors for a hypothetical two source/two electrode data set (Figure 15: A), which perfectly capture each source's activity and topography. Now imagine that Unmixing Vector 2 gets corrupted so that it no longer perfectly extracts Source 2 (Figure 15: C). Counterintuitively, Mixing Vector 2 still perfectly matches the topography of Source 2, but Mixing Vector 1 no longer matches the topography of Source 1 even though Unmixing Vector 1 still perfectly extracts the activity of that source.

To get a sense of what each IC contributes to the scalp data, it is useful to compute the "scalp-projection" (or "back-projection") of a component's activity:

$$\hat{\mathbf{x}}(t) = \mathbf{a}_i u_i(t) \tag{7}$$

where $u_i(t)$ is the activation of the ith component at time t and $\mathbf{a}_i$ is the mixing vector for the ith component. For example, IC 2's contribution to the difference between the ERPs from the two conditions is plotted in Figure 16. IC 2 contributes greatly to the difference at fronto-central midline electrodes, but contributes very little at the posterior and leftmost electrodes.

One can quantify the magnitude of this contribution by measuring the "percentage variance accounted for," ν, by the scalp-projected activity:

$$\nu = 1 - \frac{\sum_j \sum_t (x_j(t) - \hat{x}_j(t))^2}{\sum_j \sum_t x_j(t)^2} \tag{8}$$

where $x_j(t)$ is the ERP voltage at time t and electrode j, $\hat{x}_j(t)$ is the component's scalp-projection at time t and electrode j. ν is one minus the "residual variance," error measure. Residual variance is a unitless measure that is simply the sum-squared difference between the scalp and scalp-projected data normalized by the power of the scalp data. It is 0 if the scalp-projected activity perfectly matches the scalp data and accounts for all of the variance of the data. Residual variance approaches infinity (and ν negative infinity) as the difference between the two grows increasingly large without bound. The percentage variance of the difference wave accounted for by IC 2 is 39%.

Figure 17 plots IC 5's contribution to the difference between the ERPs from the two conditions, which is negligible at all but the lateral posterior electrodes (x_9 and x_{12}). This contribution accounts for 6% of the difference wave's variance. The scalp-projection of ICs 2 and 5 combined accounts for 44% of the difference wave's variance (not shown). Note that the variance accounted for by both components together does not equal the sum of the variance

accounted for by each component alone. This will generally be the case, unless the basis of LSFs is orthogonal (e.g., as in principal component analysis, Section 5), and is a quirk of the percentage variance accounted for measure.

4.2 ICA Ambiguities

4.2.1 ICA Scaling Ambiguity

In this simple example, readers may have noticed two ambiguities in ICA decompositions. The first is that the polarity and magnitude of the LSFs are arbitrary. To illustrate, consider two unmixing matrices **W** and **W′** that are identical except that the first row of **W** is — 2 times the first row of **W′**:

$$\mathbf{w}_1 = -2\mathbf{w}'_1 \qquad (9)$$

The output of the first component of **W** will be twice as large as that of **W′** and of opposite polarity. However, scaling a random variable does not affect its independence relations with other random variables, so **W** and **W′** are equally valid by ICA criteria.

The corresponding mixing matrices, **A** and **A′**, will have the inverse relationship. The first column of **A′** will be -0.5 times the first column of **A**:

$$\mathbf{a}_1 = -\frac{\mathbf{a}'_1}{2} \qquad (10)$$

Consequently, the scalp-projections of the first component of the two decompositions will be exactly the same:

$$\mathbf{a}'_1\mathbf{w}'_1\mathbf{x}(t) = -\frac{\mathbf{a}'_1}{2}(-2\mathbf{w}'_1)\mathbf{x}(t) = \mathbf{a}_1\mathbf{w}_1\mathbf{x}(t) \qquad (11)$$

The scaling ambiguity thus only applies to component activations and scalp topographies, not to their scalp-projections. This means that when comparing the activations or scalp topographies of multiple components, some type of arbitrary normalization must be done (e.g., normalizing activations to unit variance, normalizing scalp maps to unit length). Normalization, however, can be avoided by only comparing component scalp-projections.

4.2.2 ICA Component Order Ambiguity

In addition to IC scaling, the order of ICs is also arbitrary. Take for example two unmixing matrices found by ICA, **W** and **W′**, that are identical except that the first row of **W** equals the second row of **W′** and vice versa. The outputs of the two matrices will be just as independent and equally valid by ICA criteria.

This ambiguity illustrates how it less than straightforward to compare ICs from different ICA decompositions as it is not necessarily obvious which ICs in two or more ICA bases are equivalent. The order ambiguity is commonly resolved by ranking ICs in order of the magnitude of their scalp-projected activity (e.g., the magnitude of their scalp-projected variance). However, such rankings may not be very stable as one IC's scalp-projection may be about as large as several others and the relative amplitudes of latent sources in two data sets may vary due to differences between participants, recording conditions, or task demands. Thus identifying homologous ICs from different decompositions is typically done based on component features (e.g., scalp topographies and activation power spectra — Makeig et al., 2004).

4.3 Physiological Motivation for ICA

While ICA was able to successfully extract almost all of the latent ERP sources in the previous example, this was because the example was constructed to largely meet the assumptions of extended INFOMAX ICA (i.e., all the neural sources were supergaussian, relatively large and few in number, and, save for two sources,

relatively independent of one another). The success of any technique for deriving LSFs to extract EEG sources depends upon how well the technique's assumptions are met by the true data generators and there is good physiological motivation for some of ICA's assumptions. Specifically there are good a priori reasons and some empirical evidence to think that some EEG sources might be independent enough for ICA to extract. A priori, the activity of functionally distinct EEG sources should be somewhat independent if the independent variables that they are sensitive to are manipulated independently by experimenters. For example, in a language comprehension task, if one source is sensitive to word predictability and another is sensitive to the concreteness of a word's meaning, those sources should be at least somewhat independent if predictability and concreteness vary independently in the stimuli. Empirically, there is a great deal of latency variability in ERP components in the raw EEG (Kutas, McCarthy, & Donchin, 1977; Makeig et al., 2004), which should also produce a degree of independence. In addition, there is evidence that the correlation between the activities of nearby cerebral cortical areas (as measured by 20 to 240 second segments of unaveraged local field potentials) can fall off on the order of millimeters in the awake brain (Destexhe, Contreras, & Steriade, 1999).

In practice, ICA's general assumption of non-Gaussian or temporally structured sources is valid to some extent. If all EEG sources were Gaussian, then EEG ICs (i.e., ICs that appear to correspond to EEG sources rather than EEG artifacts) derived by ICA algorithms such as extended INFOMAX ICA (Lee et al., 1999) and Fast ICA (Hyvärinen, 1999) would not be reliable and there are examples of reliable EEG ICs found by these algorithms (Groppe, 2007; Himberg, Hyvärinen, & Esposito, 2004). Likewise, if all EEG sources had identical normalized power spectra, then the EEG ICs derived by SOBI would not be reliable (Belouchrani et al., 1997) and there are examples of reliable EEG ICs from SOBI (Groppe, unpublished; Tang, Pearlmutter, Malaszenko, Phung, & Reeb, 2002). Thus these assumptions are generally met to some extent though there may be some EEG sources that violate the assumptions of particular ICA algorithms and are inextricable by those algorithms. The more specific ICA assumptions about the statistical nature of the sources (i.e., specific source probability density functions or diagonal autocorrelation matrices at specific time lags) are more difficult to justify, but they may not be so problematic as many sources may fit the assumptions of multiple algorithms. Also, different ICA algorithms may be used to cross-check assumptions. For example, if an algorithm that can find Gaussian sources (e.g., SOBI) does not find any such sources in a data set, one is more justified in using an algorithm that assumes non-Gaussian sources (e.g., Extended INFOMAX ICA).

The final potentially problematic assumption of ICA is the assumption of a specific number of sources (typically as many as there are electrodes). As mentioned earlier, it impossible to know how many neural and non-neural electrical sources significantly contribute to scalp potentials. Research on simulated EEG data (Ghahremani et al., 1996; Kobayashi et al., 1999; Makeig et al., 2000) has shown that ICA is capable of accurately extracting sufficiently prominent EEG sources when there are more sources than electrodes.

4.4 Past ICA Findings

The physiological motivation for using ICA to decompose the EEG is corroborated by several successful applications. ICA was first applied to the EEG by Scott Makeig, Tzyy-Ping Jung, and colleagues in 1996 (Makeig, Bell, Jung, & Sejnowski, 1996). Since then ICA has proved useful for removing EEG artifacts such as potentials caused by blinks and 60 Hz line noise (Joyce, Gorodnitsky, & Kutas, 2004; Jung, Makeig, Humphries et al., 2000; Jung, Makeig, Westerfield et al., 2000; Tang et al., 2005), improving the discrimination of event-related

activity in the raw EEG/MEG[9] (Serby, Yom-Tov, & Inbar, 2005; Tang, Pearlmutter, Malaszenko, & Phung, 2002; Tang et al., 2005; Tang, Sutherland, & Wang, 2006; but see Section 4.5.2), and accurately isolating individual early sensory (Tang et al., 2005) and epileptic spike sources (Kobayashi, Merlet, & Gotman, 2001).

It also appears that ICA is useful for extracting ERP sources as multiple studies have found that ICA decomposes the EEG/MEG and ERPs into physiologically plausible, functionally distinct components (Debener, Ullsperger et al., 2005; Makeig et al., 1999; Makeig et al., 2004; Onton, Delorme, & Makeig, 2005; Tang, Pearlmutter, Malaszenko, Phung, & Reeb, 2002). Since the true sources of ERPs are generally unknown it is difficult to evaluate the accuracy of these decompositions but their physiological plausibility is often quite compelling. Probably the most convincing extraction of a source of a "cognitive ERP" to date is a class of ICs of the "error-related negativity" (ERN) ERP component elicited by erroneous experimental participant responses (Debener, Ullsperger et al., 2005). Debener and colleagues found a class of central medial ICs that contributed to the ERN and whose single trial activity predicted future behavior. Moreover, the average topography of the ICs and single trial IC activity was consistent with simultaneously recorded hemodynamic activity in the rostral cingulate zone. Other compelling types of ICs have also been reported. For example, Makeig and colleagues (Delorme, Westerfield, & Makeig, 2007; Makeig et al., 2004; Onton, Westerfield, Townsend, & Makeig, 2006) applied ICA to the P300 ERP component (for reviews see Donchin & Coles, 1988; Nieuwenhuis, Aston-Jones, & Cohen, 2005) elicited in a visual oddball paradigm. They accounted for the P300 with several classes of ICs. These included "P3f" components that were frontally distributed and appeared to be related to decision making/response selection, left and right centrally distributed "mu" components that possibly reflected tactile feedback, "central medial" components whose scalp distributions were similar to the P3a/Novelty P3 component related to attentional orienting (Friedman, Cycowicz, & Gaeta, 2001), and "P3b" components that accounted for a great deal of the P300.

4.5 ICA Caveats

4.5.1 Component Reliability

While ICA is a potentially powerful tool for ERP analysis, there are some caveats that should be kept in mind when interpreting ICA results. The first of these is that any given IC may not be reliable. Since ICs are learned from the statistics of a data set, noisy estimates of those statistics or local minima during training (for iterative ICA algorithms like extended INFOMAX ICA) may lead to spurious ICs. Moreover, there are some sources that are impossible for ICA to reliably decompose (e.g., two white Gaussian sources or two out of phase sine waves — Bell & Sejnowski, 1995; Meinecke, Ziehe, Kawanabe, & Müller, 2002) as multiple decompositions produce equally independent components. Consequently, evidence of ICA reliability is critical for the credibility of ICA results and can be assessed by replicating an analysis on comparable data sets. For example, the reliability of EEG ICs can been established by finding similar ICs across multiple participants (Debener, Ullsperger et al., 2005; Makeig et al., 2004; Onton et al., 2005) or subsets of a data set (Groppe, 2007; Makeig et al., 1999).

4.5.2 Component Accuracy

The second caveat to the interpretation of LSFs derived by ICA (or any other technique — see below) is that because we rarely know what the latent sources of a segment of EEG/ERP are, it is difficult to know how accurately ICs capture

[9] MEG is primarily the linear combination of the magnetic fields generated by the synaptic activity of cortical pyramidal cells (Baillet et al., 2001; Kutas & Dale, 1997). LSFs are just as potentially useful for extracting MEG sources as EEG sources. Some applications of ICA to MEG data are included in this article as evidence of ICA's ability to extract latent sources.

those sources. In the absence of this knowledge, the accuracy of ICs must be assessed indirectly.

One method that has been used to assess IC accuracy is the degree to which the topography of an IC is consistent with a single ECD or a bilaterally symmetric pair of ECDs[10], which is quantified by modeling the source of the IC (Section 6). The motivation for this criterion is the theory that "an independent component should account for synchronous activity within a connected cortical domain, and accordingly its scalp projection should match a single equivalent current dipole (or sometimes two bilaterally symmetric dipoles)" (Debener, Makeig, Delorme, & Engel, 2005, pg. 312). In practice, the ICs that account for the most EEG and ERP variance often meet this criterion (e.g., Figure 18: IC 9, IC 12, & IC 14). The number of such "dipolar" ICs is quite striking, especially when compared to the results of another statistically based method for decomposing the EEG, principal component analysis (e.g., Figures 19-20; Makeig & Delorme, 2004), which is discussed in the next section.

Nevertheless, dipolarity is probably not a necessary condition for IC accuracy. For example, the aforementioned P300, perhaps the single most studied ERP phenomena, is possibly largely produced by the widespread release of norepinephrine in cerebral cortex (Nieuwenhuis et al., 2005). If this is the case, then the P300 is produced by the activity of multiple, distributed patches of cerebral cortex whose activity should be highly correlated and the topography of this contribution to the P300 might be non-dipolar. On the other hand, dipolarity is probably not a sufficient criterion for IC accuracy either, as dipolar ICs may simply reflect the fact that the potentials recorded by nearby electrodes tend to be correlated. Consider, for example, the components of the principal square root of the estimated covariance matrix (i.e., "psPCs") that the EEGLAB toolbox (Delorme & Makeig, 2004) implementation of ICA uses as its first learning step. The set of psPCs decomposes a data set into temporally uncorrelated components and can be thought of as a version of principal component analysis (next section). The topographies of the psPCs of a participant's EEG data from a visual oddball and sentence comprehension experiment (Groppe, 2007) are displayed in Figure 21. Many of the topographies appear to be dipolar. Indeed, in a comparison of multiple statistically-based algorithms for decomposing the EEG, Makeig and Delorme (2004) found that the topographies of the psPCs produced more dipolar components on average than any of the ICA or other PCA algorithms in their analysis. However, while ICA appears to have successfully extracted some readily identifiable EEG artifacts as single components, the psPCs fail to do so. For example, IC 8 (Figure 18) corresponds to horizontal eye movement potentials, but the psPC decomposition splits the artifact into two components, psPC 21 and 36 (Figure 21). Similarly, eye blink potentials are well captured by a single independent component (IC 1 — Figure 18), but are split among several psPCs (1, 2, 3, 5, 7, 8, 9 — Figure 21). This indicates that the psPCs are not accurately extracting large EEG artifacts, which suggests that they are not accurately extracting smaller EEG sources either.

More compelling evidence for IC accuracy can come from corroborating findings from other indices of brain function. Although the relationship between other measures of brain function (e.g., fMRI, intracranial EEG) and the EEG is ambiguous (Liu, Belliveau, & Dale, 1998), compelling converging evidence is possible. As mentioned previously a class of ICs of the error-related negativity (Debener, Ullsperger et al., 2005) was consistent with simultaneously recorded hemodynamic activity in

[10] Topographies that are consistent with a single ECD have either a single focus (e.g., Figure 18: IC 12) or two foci of opposite polarity on opposite sides of the head (e.g., Figure 18: IC 14). Topographies that are consistent with a bilaterally symmetric pair of ECDs appear to be the sum of a bilaterally symmetric pair of such topographies (e.g., Figure 18: IC 7). In contrast, "non-dipolar" topographies exhibit a complicated pattern of foci (e.g., Figure 18: IC 44).

the rostral cingulate. Similarly, ICs corresponding to the SI sensory evoked potential elicited by median nerve stimulation have been found to be consistent (within 2 cm on average) with estimates of SI generator locations from previous EEG, MEG, and fMRI studies (Tang et al., 2005).

An IC's credibility is also boosted by evidence that it exhibits a tighter relationship to experimental variables and events than scalp recordings. For example, Tang and colleagues have shown that the activations of ICs corresponding to sensory evoked potentials showed less trial-to-trial variability than the electrode recordings of those sensory evoked potentials (Tang et al., 2005). Moreover, the presence of different types of sensory evoked components was easier to detect in the activations of such ICs than in the scalp data (Serby et al., 2005; Tang, Pearlmutter, Malaszenko, & Phung, 2002; Tang et al., 2006). Such evidence may suggest that the ICs are a more accurate measure of source activity than the scalp electrodes. However, the comparisons by Tang and Serby et al. were all made using scalp data that had not been screened for EEG artifacts. Thus it is not clear if these results indicate anything more than ICA's ability to isolate EEG artifacts.

Finally, the most convincing evidence that a class of ICs is accurate would be evidence that confirmed predictions about the behavior of such ICs. If a class of ICs truly reflects a functionally independent EEG process, then it should be possible to design experiments that selectively manipulate that particular function. For example, as mentioned above, Makeig and colleagues (2004) found a bilateral class of mu ICs in data from a visual oddball paradigm, which generated a post-response potential they suggested might index tactile feedback from pressing a button in the task. This putative class of IC and interpretation could be readily tested with additional experiments. For instance, if the interpretation is correct then mu ICs with different topographies should be obtained in comparable paradigms requiring a foot response (because cortical motor areas that control the hands and feet are spatially quite distinct) and no such post-response mu IC potential should be found when there is no tactile feedback. If predictions like these are confirmed by further experiments, it would increase the credibility of this class of IC.

5. PCA for Deriving LSFs

An alternative method for finding a basis of LSFs to decompose ERPs into components is principal component analysis, (PCA — see Dien & Frishkoff, 2005). In the ERP literature, PCA is often called "spatial PCA" when used to find spatial filters. Like ICA, PCA learns a basis from the statistics of the data such that each component is somewhat independent of other components and is subject to the same ambiguities and caveats as ICA. PCA differs from ICA in that it is only sensitive to second order statistics (i.e., variance/ covariance) and ignores temporal structure (e.g., autocorrelation). In other words, PCA only assumes that the sources of a data set are temporally uncorrelated, which is a weaker statistical assumption than that of ICA algorithms (Footnote 6). Because of the weaker assumption, the results of PCA are fundamentally ambiguous, as for any data set there are an infinite number of bases of LSFs that will unmix the data into uncorrelated components (Footnote 8). The canonical way to resolve this ambiguity is to define the first principal component (PC) as the LSF that accounts for the most variance of any possible LSF. The second PC captures the most variance of any possible LSF that is orthogonal to and uncorrelated with the first PC. The third PC captures the most variance of any possible LSF that is orthogonal to and uncorrelated with both the first and the second PC, and so on. Thus, the canonical PCA solution produces a basis of orthogonal LSFs whose outputs are uncorrelated and the top q components account for the maximum amount data variance.

Often, the canonical PCA solution is modified by ignoring the bottom PCs and rotating the top few PCs to produce a new basis. For example, the varimax rotation produces orthogonal components that tend to have extreme mixing/unmixing vector weights[11] (e.g., strong weights to some electrodes and weak weights to the rest). When applied to EEG/ERP data, this is equivalent to assuming that EEG/ERPs are composed of only a few uncorrelated components, with maximally different (i.e., orthogonal) topographies with a highly variable distribution (i.e., extreme gain vector weights).

PCA can be understood in the same generative model framework as ICA. In this framework, PCA is the optimal unmixing of a data set when the sources, **u**, are white Gaussian random variables, possibly less than the number of electrodes, and embedded in spherical white Gaussian sensor noise, **z** (Roweis & Ghahramani, 1999).

$$\mathbf{x}(t) = \mathbf{A}\mathbf{u}(t) + \mathbf{z}(t) \quad (12)$$

If this were an accurate model of the EEG/ERPs, then PCA would be better than ICA for deriving LSFs. However, this is clearly not the case. The fact that ICA reliably finds the same components in a data set shows that EEG/ERP data sets are not strictly white or Gaussian (Section 4.5.1). Sensor noise is also not uniform across all electrodes. Peri-ocular and temporal electrodes tend to have more noise than other electrodes in EEG/ERP data sets due to blinks/eye movements and muscle activity. Finally, it is unlikely that EEG sources have orthogonal (maximally different) topographies and they may not fit the desideratum of rotations. For example, EEG sources that are deep within the head can have very broad scalp distributions with rather uniform weights at many electrodes. A varimax rotation would not be likely to return such a component.

[11] PCA unmixing and mixing vectors are qualitatively the same. In other words, they are parallel and the polarity and magnitude of one is the inverse of that of the other.

Although the assumptions of ICA are a priori generally more plausible than those of PCA, it is currently impossible to say if ICA or PCA are typically more accurate at extracting ERP sources, since, as mentioned previously, we rarely know what the sources of an ERP are. To our knowledge, the most careful comparison of ICA and PCA derived ERP decompositions (Makeig et al., 1999), found that the 3 main ICs of the P300 correlated more strongly with reaction time and were more reliable than their PCA equivalents. Some researchers have compared ICA and PCA's ability to remove artifacts from EEG recordings. In contrast to ERP sources, artifact components reflecting blinks, eye movements, muscle activity, and line noise are readily identifiable and ICA appears to be much better at capturing these sources as distinct components (Jung, Makeig, Humphries et al., 2000; Tang et al., 2005).

6. Source Localization for Deriving LSFs

A very different alternative for decomposing ERPs into components using LSFs is anatomically-based source localization (Baillet et al., 2001). In contrast to ICA, which learns filters using only the statistics of scalp potentials, source localization builds a model of the EEG gain matrix that can be based on a participant's anatomy, information about probable source locations, the statistics of the scalp potentials, and the biophysics of EEG generation. Once this gain matrix, or "forward model," has been derived, an "inverse model," a set of LSFs, can be computed which estimates the activity of the putative sources. The forward and inverse models are analogous to the ICA mixing and unmixing matrices.

One of the most popular methods for source localization is to approximate the brain, skull, and scalp with a series of concentric, homogeneous spheres and to assume a handful of dipolar generators (Baillet et al., 2001). The locations and orientations of the equivalent current dipoles

(ECDs) are then iteratively fit to the data until a sufficiently small degree of error is reached. The dipoles provide a forward model, $\hat{\mathbf{G}}$, which is an *n* x *m* matrix that maps the activity of *m* sources, $\hat{\mathbf{y}}$, to the potentials at *n* electrodes, **x**, which are embedded in noise, **z**.

$$\mathbf{x}(t) = \hat{\mathbf{G}}\hat{\mathbf{y}}(t) + \mathbf{z}(t) \quad (13)$$

To estimate the source activities from the data, $\hat{\mathbf{y}}$, typically the pseudoinverse of $\hat{\mathbf{G}}$ is used.

$$\hat{\mathbf{y}}(t) = (\hat{\mathbf{G}}^T\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}^T\mathbf{x}(t) \quad (14)$$

This minimizes the sum squared error between the model and the scalp data.

While such an iterative method assumes fewer sources than electrodes, a different approach is to assume more sources than electrodes, which removes the need for iteration. An example of such a method is dynamic statistical parametric mapping (DSPM — Dale, et al., 2000; Liu, et al., 2002). DSPM approximates the cortical surface, inner skull, outer skull, and skin with a boundary element model and a dipolar EEG source is placed about every 1 cm on the cortical surface, oriented perpendicular to it. This approach provides a forward model, $\hat{\mathbf{G}}$, with around three thousand sources per hemisphere. Given the forward model, assumptions about the covariance/variance of the sources and noise, an inverse model can be computed:

$$\hat{\mathbf{y}}(t) = \mathbf{R}\hat{\mathbf{G}}^T(\hat{\mathbf{G}}\mathbf{R}\hat{\mathbf{G}}^T + \mathbf{C})^{-1}\mathbf{x}(t) \quad (15)$$

where **R** and **C** are the covariance matrices of the neural and noise sources respectively. This estimation of source activity is optimal when the sources and noise are Gaussian[12]. To correct for the additive noise, an additional step is taken: the estimated activity of each source is divided by the standard deviation of the noise contributing to that activity.

Relative to ICA, the advantages of source localization for extracting individual EEG/ERP sources are that one can make use of knowledge about the location of these sources and the statistics of the noise when it is available. Furthermore, one does not have to make strong independence assumptions, localization does not suffer from scaling ambiguities, and it is easier to compare results across participants, studies, and other neuroimaging methods (e.g., functional magnetic resonance imaging, intracranial recordings).

The disadvantage of source localization is that inaccuracies in the forward model and statistical assumptions may lead to erroneous results. For example, while the boundary element model used by DSPM is a more accurate approximation of the head than the concentric spheres used in iterative dipole fitting, both ignore the fact that some parts of the brain are anisotropic and that different regions of skull differ in their conductive properties (Baillet et al., 2001). More problematic for source localization are assumptions about the number and locations of sources, which are generally difficult to validate. Iterative dipole fitting may produce accurate results when fitting a very small number of dipoles (e.g., Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2002). However, as the number of dipoles increases, the stability of the solutions rapidly degrades and different initial estimates of source locations and orientation can produce very different results (Baillet et al., 2001). While DSPM does not the have the stability problems of iterative solutions, by potentially greatly

[12] This might be a reasonable assumption for localizing ERPs as they are derived from the mean of a large number of epochs of EEG (usually between 30 and 100). Thus, by the central limit theorem, the sampling distribution of the mean source and noise activity at each time point should be approximately Gaussian if the distribution of their single trial values does not remarkably deviate from a Gaussian distribution (Wilcox, 2002). Gaussian assumptions are less likely to be accurate for localizing single epochs of EEG or ERPs constructed from a limited number of epochs.

over-estimating the number of sources, DSPM will sub-optimally localize activity (Liu et al., 2002).

Source localization and ICA are complementary as they exploit different sources of information and can make compatible assumptions (i.e., source localization depends on strong anatomical assumptions and ICA depends on strong statistical assumptions). Given this, localization and ICA can be used to corroborate one another's results.

Moreover, ICA can function as a useful preprocessing step in iterative source localization. Iterative source localization works best when modeling data with few dipoles. As mentioned previously, the topographies of many ICA components are consistent with a single dipole or dual-symmetric dipolar source (e.g., 28% of non-artifact ICs were consistent with a single dipole in one 64 electrode study — Debener, Makeig et al., 2005). Thus while an EEG/ERP phenomenon may be too spatially complex for iterative localization to find a reliable solution, localizing ICs that contribute to that ERP may simplify the problem enough to make it tractable and thereby improve accuracy. In fact, research on simulated and real data have shown that ICA can improve localization accuracy relative to conventional iterative methods in some situations (Kobayashi, Akiyama, Nakahori, Yoshinaga, & Gotman, 2002; Tang, Pearlmutter, Malaszenko, Phung, & Reeb, 2002) though not all (Kobayashi et al., 2001).

Conversely, source localization may assist ICA (Onton et al., 2005). Localizing ICs will remove ICA's scaling ambiguity, facilitate comparison with results from other imaging methods, and ease the comparison of ICs across participants. The last of these benefits results from the fact that anatomically equivalent sources in two different participants can have remarkably different scalp distributions due to functionally irrelevant differences in cortical folding and head shape. Localizing can correct for these differences and can reduce the dimensionality of an IC's topography (e.g., from 64 electrodes to 3 Cartesian coordinates), which facilitates analysis (e.g., clustering) and visualization.

7. Temporal PCA for Defining ERP Components

In contrast to the previously mentioned methods for extracting EEG/ERP sources with LSFs, some researchers have used linear temporal filters (LTFs) derived by PCA to decompose ERPs into components. This method, called "temporal PCA," is simply PCA (see Section 5) where the individual time samples (e.g., 100, 104, 108 ms post-stimulus) are treated as random variables (Dien & Frishkoff, 2005). As typically applied, the generative model for the data is:

$$\mathbf{x}(e_{pc}) = \mathbf{A}\mathbf{u}(e_{pc}) + \mathbf{z}(e_{pc}) \qquad (16)$$

where each element of the vectors **x** (scalp data), **u** (source), and **z** (spherical white Gaussian noise) represents a different time sample and e_{pc} indexes an electrode, e, on participant p in experimental condition c.

Applied this way, PCA decomposes ERPs into components whose time courses are fixed, orthogonal to one another, and whose topographies can change across experimental conditions and participants. As with spatial PCA, the top few PCs account for the great bulk of data variance and may be rotated to be "simpler," while the lower PCs are ignored.

In contrast to LSFs, temporal PCA is not well motivated physiologically. According to the biophysics of ERP generation, ERPs are a linear combination of temporally varying source activity with fixed topographies. Thus, sets of LSFs can potentially extract some of those sources. Nothing about the physiology of ERP generation suggests that it makes sense to think of ERPs as linear combinations of temporally fixed waveforms with fluid topographies. In fact such components violate

conventional definitions of ERP components as having fixed scalp distributions but varying latencies. Because of this lack of physiological justification, LSFs are more suitable for defining ERP components.

8. Summary

If an EEG phenomenon is generated by distinct patches of cerebral cortex that produce at least somewhat distinct topographies, linear spatial filters (LSFs) can potentially extract those sources. However, because of the ambiguous relationship between scalp electric potentials and their neural and non-neural sources, it is difficult to know how to derive LSFs that successfully do this. By making assumptions about the sources of scalp electric potentials it is possible to derive successful LSFs if the assumptions are valid.

One technique for deriving LSFs from EEG data is independent component analysis (ICA), which learns LSFs by making assumptions about the statistics of the underlying generators. More specifically, ICA assumes that the underlying generators are temporally independent and somehow non-Gaussian or temporally structured. In addition, ICA, as typically applied to EEG data, assumes that the number of sources equals the number of electrodes. There is both a priori and empirical evidence that ICA assumptions about the statistics of the sources are somewhat valid. Moreover, in practice, ICA appears to be able to successfully extract EEG artifacts (e.g., blinks, muscle potentials) and ICA often derives EEG components that are compellingly physiologically plausible. Thus it appears that ICA may be able to accurately extract some EEG sources. However, the results of ICA, or any other technique for deriving LSFs, should be qualified by the fact that they may be erroneous due to inaccurate assumptions and should be validated by converging evidence from other methods and tested predictions. Moreover, since ICA LSFs are learned from the statistics of a data set, it is important to assess the reliability of the derived LSFs.

ICA differs from two other methods for deriving LSFs to extract EEG sources, principal component analysis (PCA) and source localization. Like ICA, PCA derives LSFs by assuming that the sources of an EEG data set are somewhat temporally independent. Specifically, PCA assumes that the sources are simply uncorrelated, which is a weaker form of independence than is assumed by ICA. Because of the weaker assumption, PCA results are ambiguous and users are forced to make additional assumptions (e.g., orthogonal LSFs and highly variable filter weights) to resolve the ambiguity. These additional assumptions are harder to justify than the stronger statistical assumptions of ICA and there is some evidence that ICA is better at extracting latent non-neural and neural sources of EEG data sets. Moreover, principal components are just as subject to the concerns of component validity and reliability as independent components. Thus ICA seems to be a more promising approach for extracting EEG sources though clearly more comparisons of the two methods need to be done.

In contrast to ICA and PCA, source localization relies upon an anatomical model of EEG generation to derive LSFs (in addition to assumptions about source and/or noise statistics). The anatomical model provides some advantages over ICA and PCA in that one does not have to make strong independence assumptions, localization does not suffer from scaling ambiguities, and it is easier to compare results across participants, studies, and other neuroimaging methods (e.g., functional magnetic resonance imaging, intracranial recordings). However, inaccuracies in the anatomical model may lead to erroneous results, and since the locations of the generators of EEG phenomena are generally not known with much certainty, this can be a serious problem for source localization.

Finally, an alternative method to LSFs for parsing ERPs into components is “temporal principal component analysis” (tPCA), which extracts

ostensible ERP components with linear temporal filters (LTFs). As typically applied, PCA decomposes ERPs into components whose time courses are fixed, orthogonal to one another, and whose topographies can change across experimental conditions and participants. In contrast to LSFs, temporal PCA is not well motivated physiologically as nothing about the biophysics of EEG generation suggests that it makes sense to think of ERPs as linear combinations of temporally fixed waveforms with fluid topographies. In fact such components violate conventional definitions of ERP components as having fixed scalp distributions but varying latencies. Because of this lack of physiological justification, LSFs are more suitable for defining ERP components. ■

Acknowledgements

The authors would like to thank Hsin-Hao Yu and two anonymous reviewers for their helpful comments on an earlier version of this article.

References

Baillet, S., Mosher, J. C., & Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, 18(6), 14-30.

Bell, A. J. & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6), 1129-1159.

Belouchrani, A., Abed Meraim, K., Cardoso, J. F., & Moulines, E. (1997). A blind source separation technique using second order statistics. *IEEE Transactions on Signal Processing*, 45, 434-444.

Coulson, S., King, J. W., & Kutas, M. (1998). ERPs and domain specificity: Beating a straw horse. *Language and Cognitive Processes*, 13(62), 653-672.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, 26(1), 55-67.

Debener, S., Makeig, S., Delorme, A., & Engel, A. K. (2005). What is novel in the novelty oddball paradigm? Functional significance of the novelty P3 event-related potential as revealed by independent component analysis. *Cognitive Brain Research*, 22(3), 309-321.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience*, 25(50), 11730-11737.

Delorme, A. & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9-21.

Delorme, A., Westerfield, M., & Makeig, S. (2007). Medial prefrontal theta bursts precede rapid motor responses during visual selective attention. Journal of *Neuroscience*, 27(44), 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. *Journal of Neuroscience*, 19(11), 4595-4608.

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.

Di Russo, F., Martinez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping*, 15(2), 95-111.

Dien, J. & Frishkoff, G. (2005). Principal components analysis of ERP data. In T. C. Handy (Ed.), *Event-Related Potentials: A Methods Handbook* (1st ed., pp. 189-207). Cambridge, Mass.: MIT Press.

Donchin, E. & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, 112(3), 357-374.

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Biobehavioral Reviews*, 25(4), 355-373.

Ghahremani, D., Makeig, S., Jung, T.-P., Bell, A. J., & Sejnowski, T. J. (1996). Independent component analysis of simulated EEG using a three-shell spherical head model (Technical Report No. 96-01). La Jolla,

California, USA: Institute for Neural Computation, University of California San Diego.

Groppe, D. (2007). Common independent components of the P3b, N400, and P600 ERP components to deviant linguistic events. Unpublished Ph.D. dissertation, University of California, San Diego.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15(6), 883-899.

Himberg, J., Hyvärinen, A., & Esposito, F. (2004). Validating the independent components of neuroimaging time series via clustering and visualization. *NeuroImage*, 22(3), 1214-1222.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3), 626-634.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent Component Analysis*. New York: J. Wiley.

Ille, N., Berg, P., & Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *Journal of Clinical Neurophysiology*, 19(2), 113-124.

Johnson, J. S. & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, 3(7), 499-512.

Joyce, C. A., Gorodnitsky, I. F., & Kutas, M. (2004). Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology*, 41(2), 313-325.

Jung, T.-P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V. J., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(2), 163-178.

Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, 111(10), 1745-1758.

Kobayashi, K., Akiyama, T., Nakahori, T., Yoshinaga, H., & Gotman, J. (2002). Systematic source estimation of spikes by a combination of independent component analysis and RAP-MUSIC. I: Principles and simulation study. *Clinical Neurophysiology*, 113(5), 713-724.

Kobayashi, K., James, C. J., Nakahori, T., Akiyama, T., & Gotman, J. (1999). Isolation of epileptiform discharges from unaveraged EEG by independent component analysis. *Clinical Neurophysiology*, 110(10), 1755-1763.

Kobayashi, K., Merlet, I., & Gotman, J. (2001). Separation of spikes from background by independent component analysis with dipole modeling and comparison to intracranial recording. *Clinical Neurophysiology*, 112(3), 405-413.

Kutas, M. & Dale, A. M. (1997). Electrical and magnetic readings of mental functions. In M. D. Rugg (Ed.), *Cognitive Neuroscience* (first ed., pp. 197-242). Hove East Sussex, UK: Psychology Press.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, 197(4305), 792-795.

Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II (1994-2005). In M. A. Gernsbacher, & M. Traxler (Eds.), *Handbook of Psycholinguistics* (2nd ed., pp. 659-724). New York: Elsevier.

Lee, T. W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and super-gaussian sources. *Neural Computation*, 11(2), 417-441.

Liu, A. K., Belliveau, J. W., & Dale, A. M. (1998). Spatiotemporal imaging of human brain activity using functional MRI constrained magneto-encephalography data: Monte carlo simulations. *Proceedings of the National Academy of Sciences of the United States of America*, 95(15), 8945-8950.

Liu, A. K., Dale, A. M., & Belliveau, J. W. (2002). Monte Carlo simulation studies of EEG and MEG localization accuracy. *Human Brain Mapping*, 16(1), 47-62.

Makeig, S., Bell, A. J., Jung, T.-P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephal-ographic data. In D. Touretzky, M. Mozer & M.

Hasselmo (Eds.), *Advances in Neural Information Processing Systems 8* (pp. 145-151). Cambridge, MA: MIT Press.

Makeig, S. & Delorme, A. (2004). Physiological plausibility and stability of independent component analysis. Cognitive Neuroscience Society Annual Meeting Program, San Francisco.

Makeig, S., Delorme, A., Westerfield, M., Jung, T.-P., Townsend, J., Courchesne, E., et al. (2004). Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biology*, 2(6), e176.

Makeig, S., Jung, T.-P., Ghahremani, D., & Sejnowski, T. J. (2000). Independent component analysis of simulated ERP data. In T. Nakada (Ed.), *Integrated Human Brain Science* (pp. 123-146). New York: Elsevier.

Makeig, S., Westerfield, M., Jung, T.-P., Covington, J., Townsend, J., Sejnowski, T. J., et al. (1999). Functionally independent components of the late positive event-related potential during visual spatial attention. *Journal of Neuroscience*, 19(7), 2665-2680.

Martinez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., et al. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience*, 2(4), 364-369.

Meinecke, F., Ziehe, A., Kawanabe, M., & Müller, K. R. (2002). A resampling approach to estimate the stability of one-dimensional or multidimensional independent components. *IEEE Transactions on Biomedical Engineering*, 49(12), 1514-1525.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, 131(4), 510-532.

Onton, J., Delorme, A., & Makeig, S. (2005). Frontal midline EEG dynamics during working memory. *NeuroImage*, 27(2), 341-356.

Onton, J., Westerfield, M., Townsend, J., & Makeig, S. (2006). Imaging human EEG dynamics using independent component analysis. *Neuroscience and Biobehavioral Reviews*, 30(6), 808-822.

Osterhout, L. & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King, and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes*, 14(1), 1-14.

Roweis, S., & Ghahramani, Z. (1999). A unifying review of linear Gaussian models. *Neural Computation*, 11(2), 305-345.

Serby, H., Yom-Tov, E., & Inbar, G. F. (2005). An improved P300-based brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(1), 89-98.

Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology*, 38(2), 343-358.

Talsma, D. & Woldorff, M. G. (2005). Methods for the estimation and removal of artifacts and overlap in ERP waveforms. In T. C. Handy (Ed.), *Event-Related Potentials: A Methods Handbook* (pp. 115-148). Cambridge, Mass.: MIT Press.

Tang, A. C., Pearlmutter, B. A., Malaszenko, N. A., & Phung, D. B. (2002). Independent components of magnetoencephalography: Single-trial response onset times. *NeuroImage*, 17(4), 1773-1789.

Tang, A. C., Pearlmutter, B. A., Malaszenko, N. A., Phung, D. B., & Reeb, B. C. (2002). Independent components of magnetoencephalography: Localization. *Neural Computation*, 14(8), 1827-1858.

Tang, A. C., Sutherland, M. T., & McKinney, C. J. (2005). Validation of SOBI components from high-density EEG. *NeuroImage*, 25(2), 539-553.

Tang, A. C., Sutherland, M. T., & Wang, Y. (2006). Contrasting single-trial ERPs between experimental manipulations: Improving differentiability by blind source separation. *NeuroImage*, 29(1), 335-346.

Van Essen, D. C., Drury, H. A., Joshi, S., & Miller, M. I. (1998). Functional and structural mapping of human cerebral cortex: Solutions are in the surfaces. *Proceedings of the National Academy of Sciences of the United States of America*, 95(3), 788-795.

Wilcox, R. R. (2002). *Applying Contemporary Statistical Techniques*. London: Academic Press.

Wood, C. C. & McCarthy, G. (1984). Principal components analysis of event-related potentials: Simulation studies demonstrate misallocation of variance across components. *Electroencephalography and Clinical Neurophysiology*, 59, 249-260. ■

Table 1 — (A) Correlation between the hypothetical source ERPs visualized in Figure 7. (B) Correlation between the hypothetical source unaveraged activations from which the source ERPs in Figure 7 are derived.

A: Source ERP Correlation (Pearson's r)

y1	1.00							
y2	.61	1.00						
y3	-.56	-.71	1.00					
y4	.23	.07	.11	1.00				
y5	.14	.25	-.33	-.31	1.00			
y6	.28	.20	-.04	.39	-.42	1.00		
y7	.16	.32	-.14	0	.02	.11	1.00	
y8	.15	.26	-.34	-.32	.99	-.41	.02	1.00
	y1	y2	y3	y4	y5	y6	y7	y8

B: Source EEG Correlation (Pearson's r)

y1	1.00							
y2	-.01	1.00						
y3	-.01	.03	1.00					
y4	.04	-.03	-.01	1.00				
y5	-.02	.01	.03	-.03	1.00			
y6	.01	.03	.02	.04	0	1.00		
y7	.01	.05	0	.03	.01	0	1.00	
y8	-.01	.02	.02	-.02	.83	.01	.01	1.00
	y1	y2	y3	y4	y5	y6	y7	y8

Table 2 — Gain vector similarity of the hypothetical neural ERP sources visualized in Figure 10. Similarity is measured as the cosine of the angle between two gain vectors, r_{cos} (see Footnote 2).

Source Gain Vector Similarity (r_{cos})

	g1	g2	g3	g4	g5	g6	g7	g8
g1	1.00							
g2	.96	1.00						
g3	.37	.56	1.00					
g4	.18	.37	.67	1.00				
g5	.03	.02	-.18	.30	1.00			
g6	.88	.90	.65	.21	-.19	1.00		
g7	-.89	-.84	-.19	-.14	-.29	-.66	1.00	
g8	.02	-.13	-.24	-.41	.15	.04	.04	1.00

Table 3 — Topography and activation similarity of each IC and its analogous source (from the hypothetical data and ICA decomposition presented in Figures 7-14, 16-17). ICs were paired with sources according their topographies. The IC and source with the most similar topographies were paired first. Then, of the remaining ICs and sources, the IC and source with the most similar topographies were paired and so on until each IC was paired with exactly one source. Similarity is defined as the cosine of the angle between feature vectors, r_{cos} (see Footnote 2). The similarities in the second table (B) were computed after removing the 12 sensor noise sources.

A: Similarities Between ICs and Sources

IC	Best Source Match	Topography r_{cos}	ERP r_{cos}	EEG r_{cos}
1	1	.99	.93	.74
2	2	1.00	.92	.88
3	3	.99	.99	.95
4	4	1.00	.98	.98
5	5	.90	.99	.97
6	6	.97	.92	.83
7	7	.97	.91	.86
8	8	.84	.94	.52
9	12	.71	.61	.60
10	16	.69	.77	.74
11	18	.86	.80	.81
12	19	.76	.86	.85

B: Similarities Between ICs and Sources (without Sensor Noise)

IC	Best Source Match	ERP r_{cos}	EEG r_{cos}
1	1	.98	.96
2	2	.94	.95
3	3	.99	.98
4	4	.98	.99
5	5	.99	.98
6	6	.98	.96
7	7	.93	.94
8	8	.98	.65

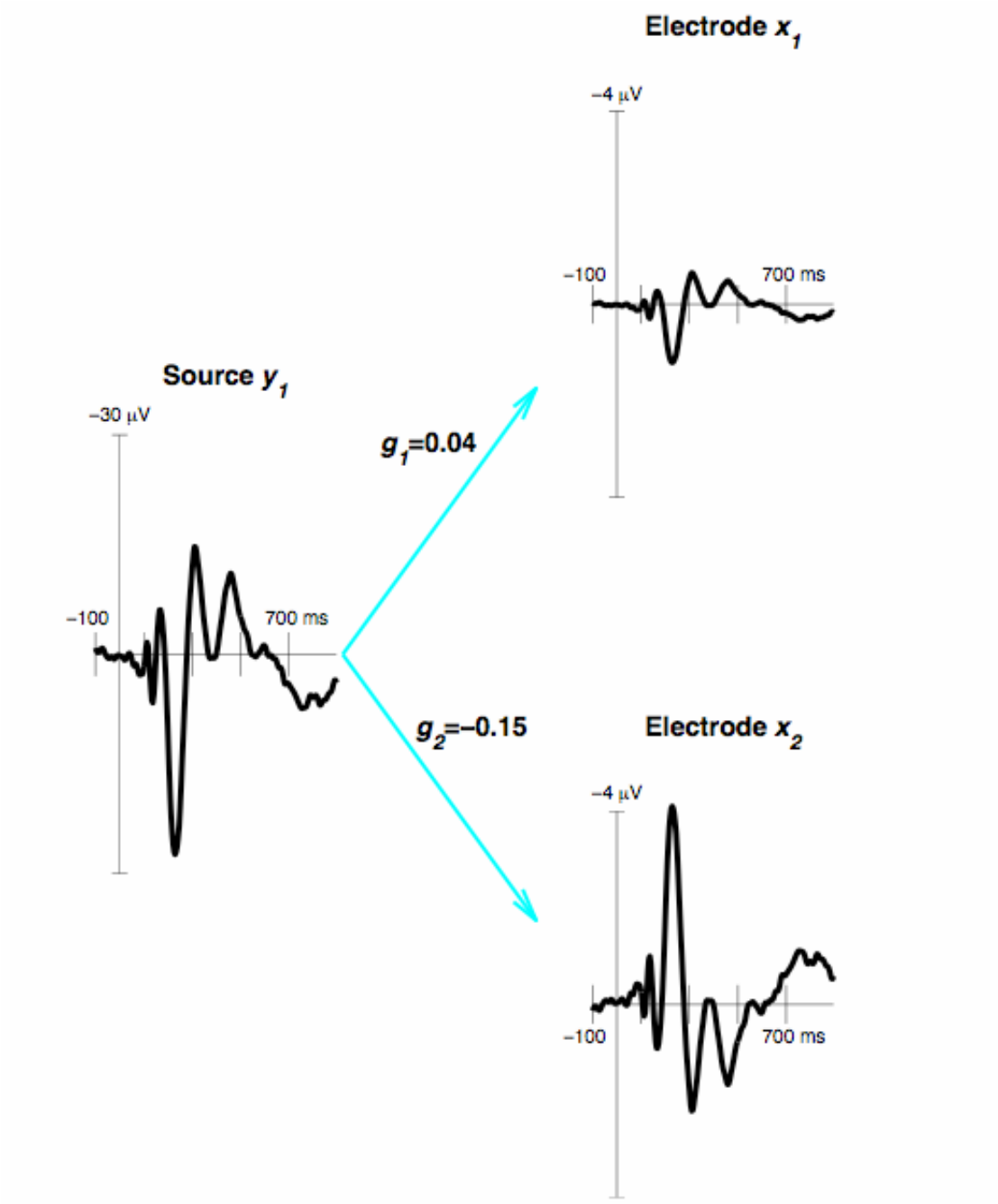

Figure 1 — A single EEG source, y_1, and its projection to two electrodes, x_1 and x_2. The gain values between the source and each electrode are g_1 and g_2. Note the difference in voltage scale between the source and the electrodes.

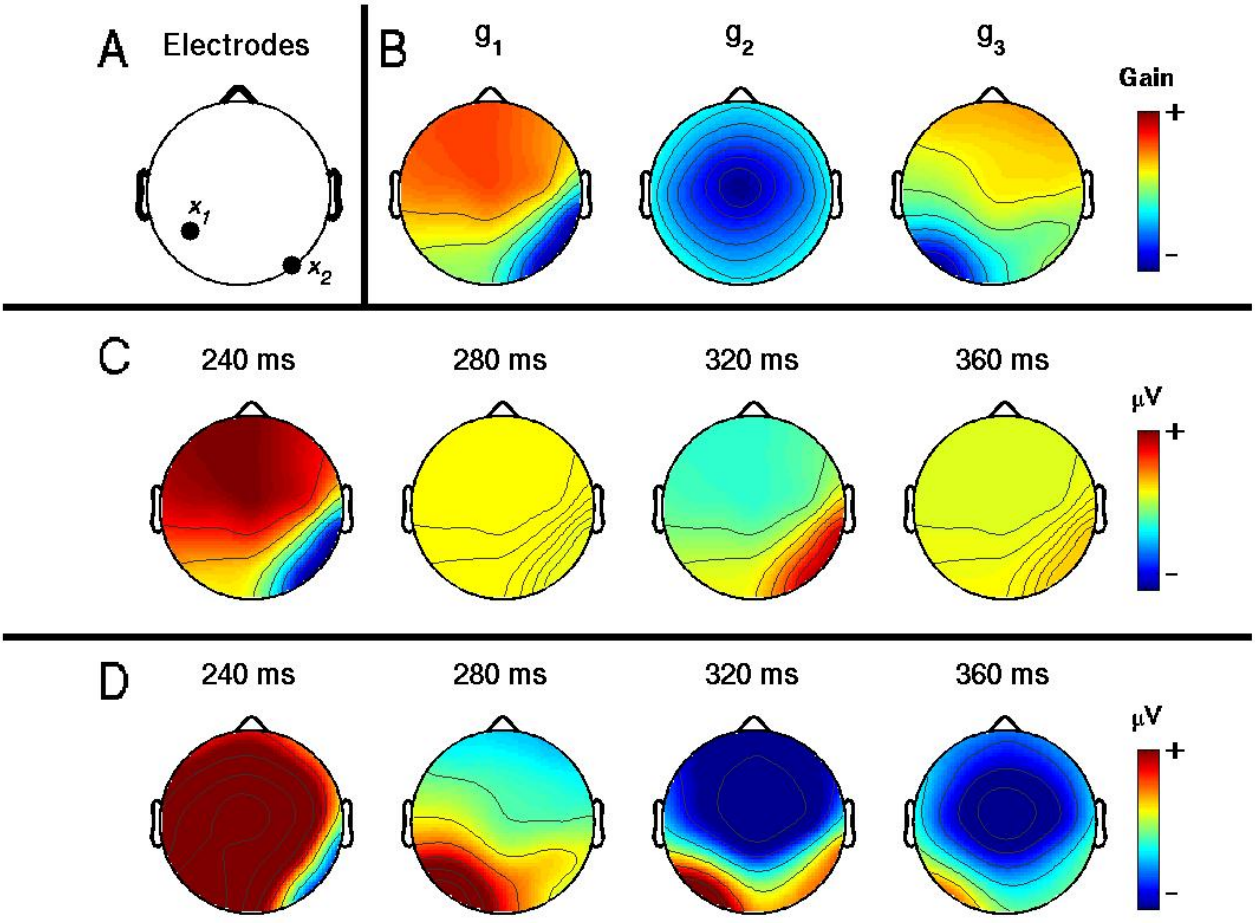

Figure 2 — (**A**) Black dots indicate scalp positions of electrodes in Figures 1 and 3. (**B**) The unitless gain vectors (topographies) of the three EEG sources in Figures 1 and 3. For visualization, each gain vector has been normalized so that its absolute maximal weight is 1. (**C**) Topographies of the scalp potentials from Figure 1 at four points in time. Because there is only one source, the topographies are identical (i.e., the ratio of the potentials at any two points on the scalp is constant). (**D**) Topographies of the scalp potentials from Figure 3 at four points in time. Because there are multiple somewhat independent sources with distinct topographies, the topography of the scalp potentials changes across time.

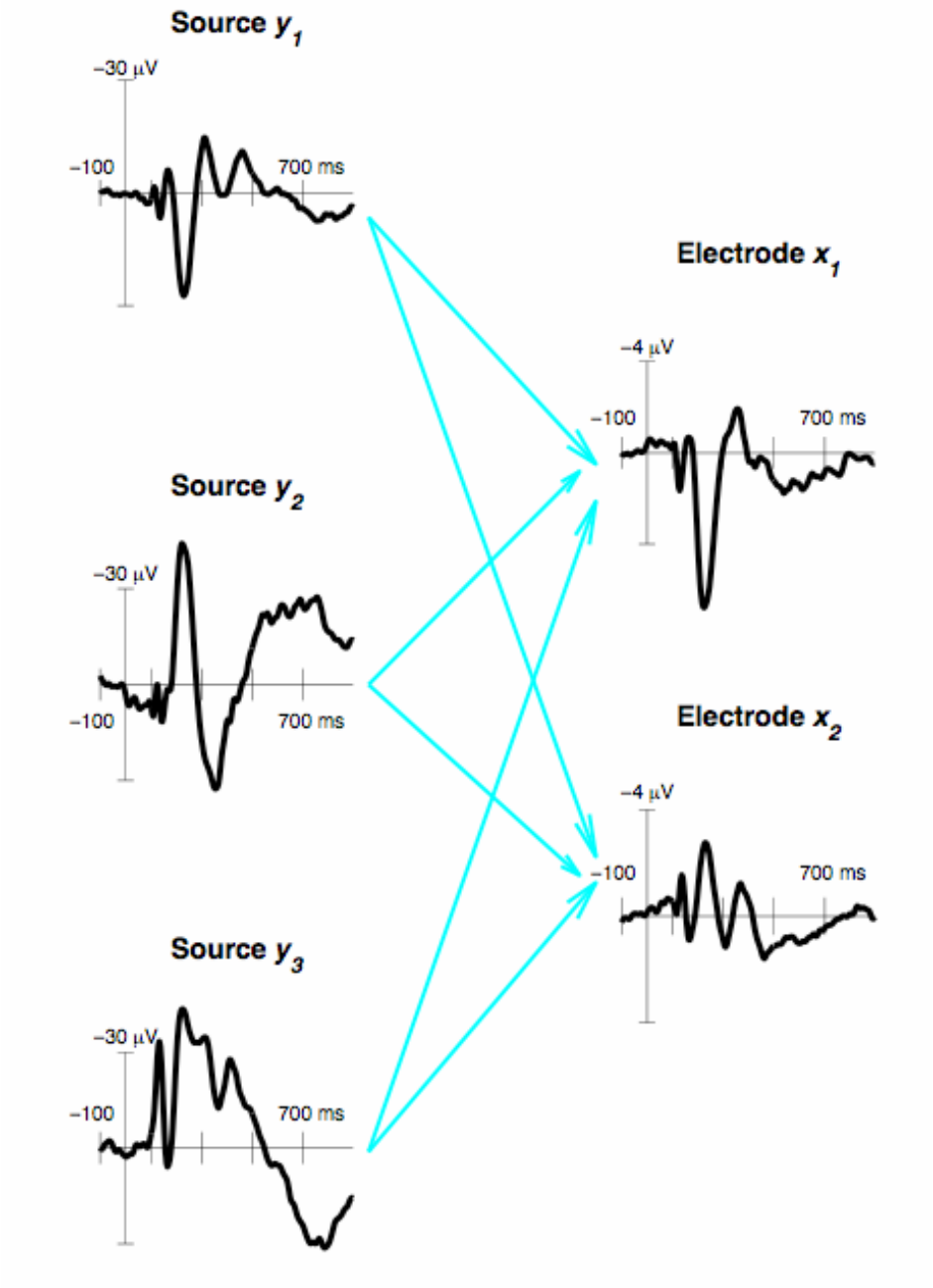

Figure 3 — Three EEG sources (y_1 ,y_2 , and y_3), and their combined projections to two electrodes, x_1 and x_2. Note the difference in voltage scale between the sources and the electrodes.

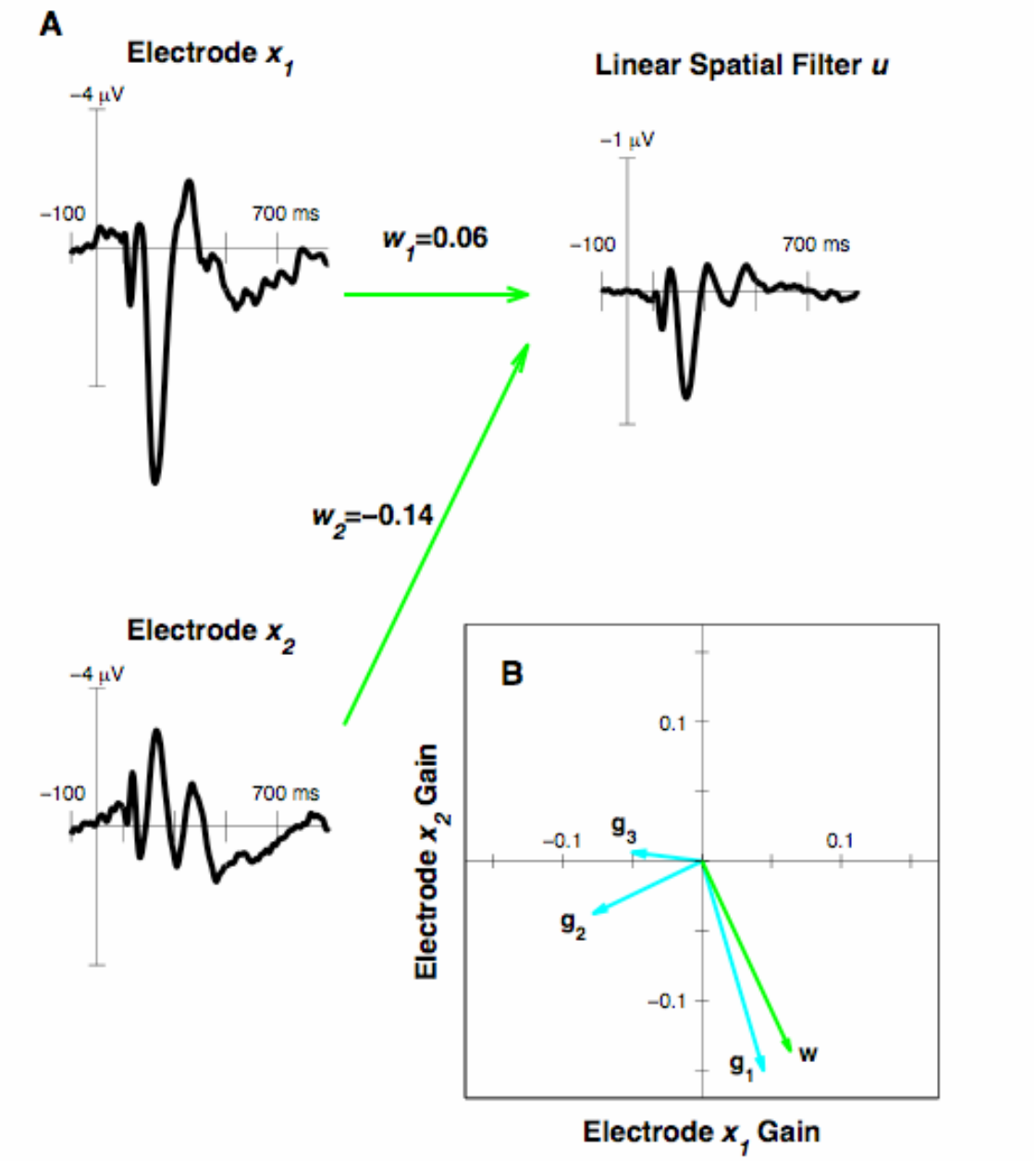

Figure 4 — (**A**) The output, *u*, of a linear spatial filter, **w**, applied to the potentials at two electrodes from Figure 3. Note the difference in voltage scale between the electrodes and the linear spatial filter (LSF). (**B**) The gain vectors ($\mathbf{g}_1$, $\mathbf{g}_2$, $\mathbf{g}_3$) of the three sources in Figure 3 and the gain vector of the LSF, **w**. Since **w** is nearly perpendicular to $\mathbf{g}_2$ and $\mathbf{g}_3$, and nearly parallel to $\mathbf{g}_1$, the filter output closely correlates with that of Source y_1 (shown in Figure 3).

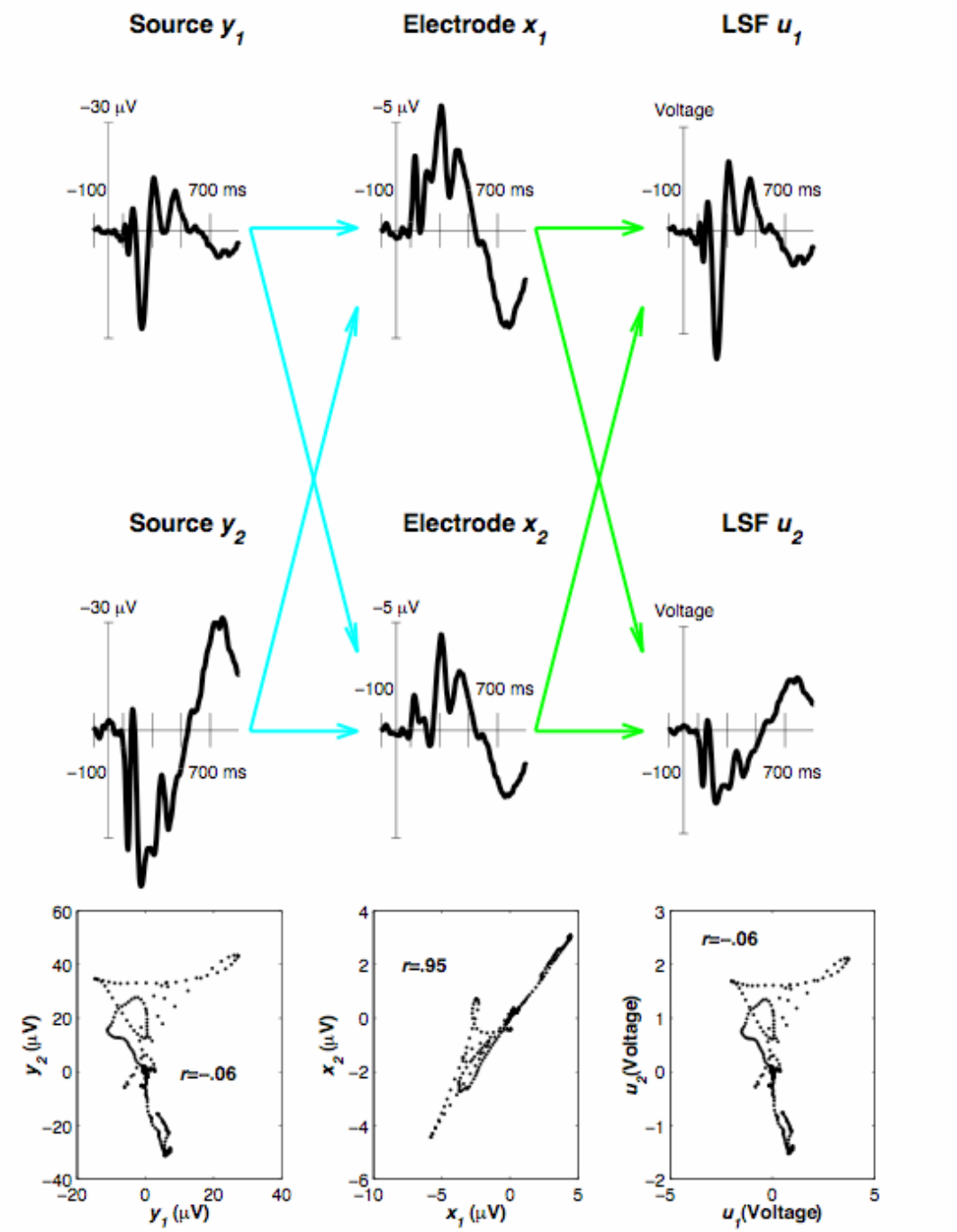

Figure 5 — (**Left Column**) Two hypothetical EEG sources. A scatter plot of the source activity (**Left Column-Bottom**) illustrates a degree of independence between the sources. (**Middle Column**) The combined projection of the two sources at two scalp electrodes. The potentials at the two electrodes are highly correlated. (**Right Column**) The output of two linear spatial filters perfectly match the original sources and are thus somewhat independent. Blue arrows represent gain weights. Green arrows represent filter weights. Scale of LSF output is arbitrary (Section 4.2.1).

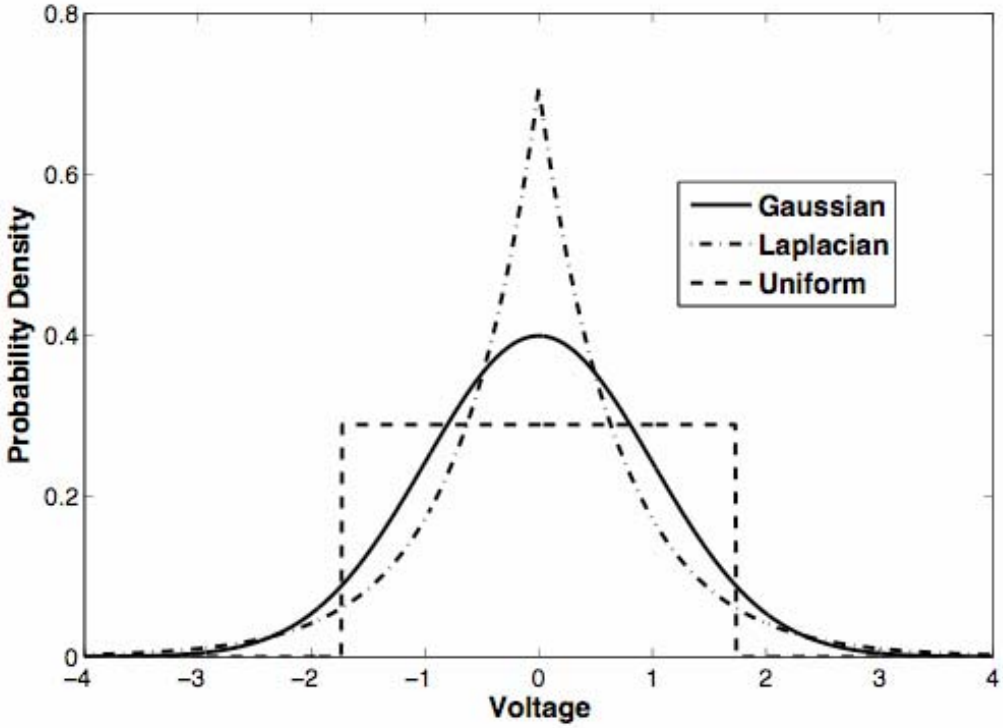

Figure 6 — Examples of Gaussian, supergaussian, and subgaussian distributions. All distributions are zero mean and unit variance. Supergaussian random variables (in this case one with a Laplacian pdf) are more likely to take values close to the mean or extreme values than Gaussian variables. Random samples of the electric potentials generated by occasional bursts of muscle activity or intermittent oscillatory brain activity have supergaussian distributions because most of the time their activation is near zero, but occasionally they produce very strong potentials. Subgaussian distributions (in this case one with a uniform pdf) are more uniformly distributed around the mean than Gaussian variables. Random samples from a sine wave have a subgaussian distribution.

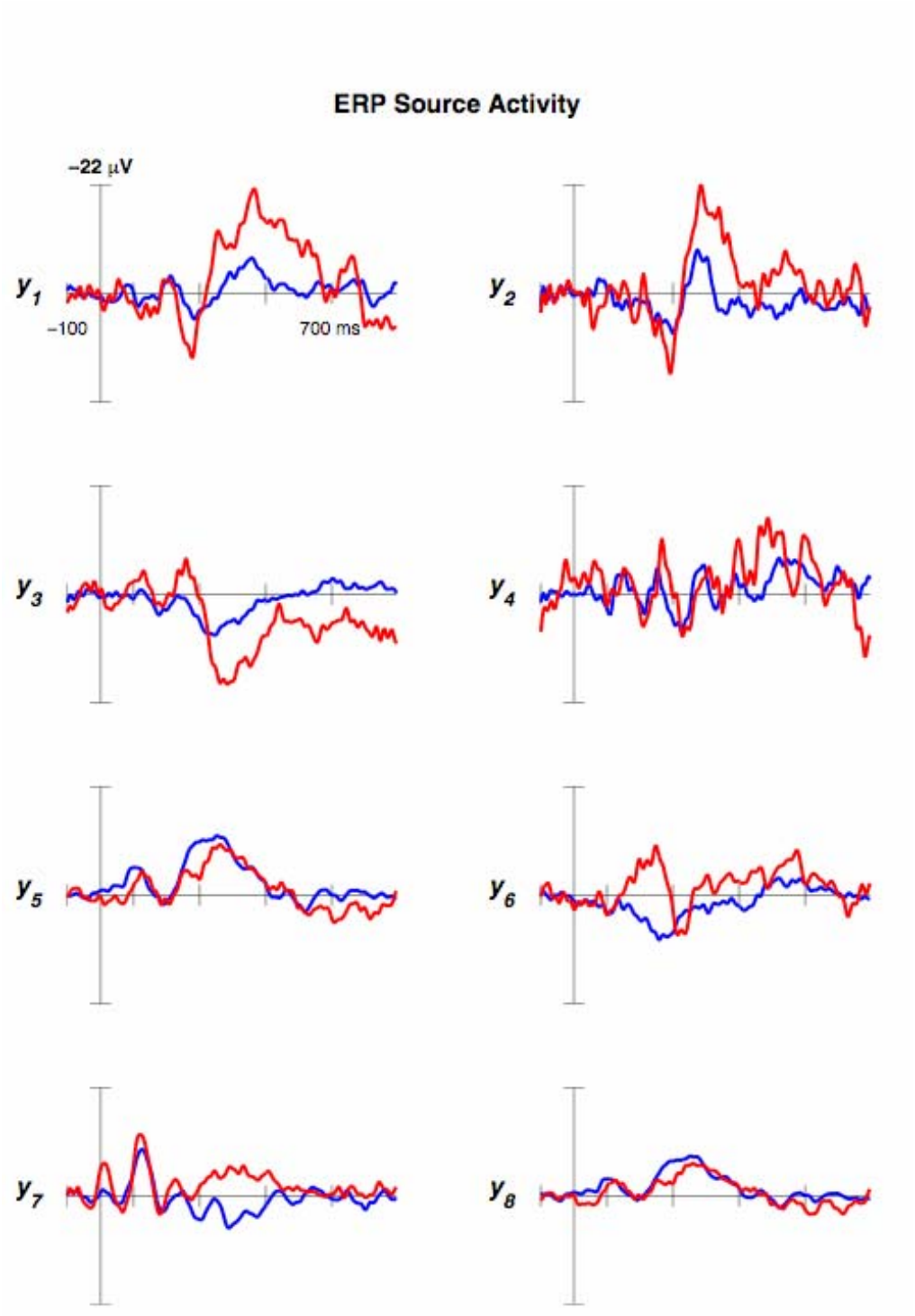

Figure 7 — Average activity of eight hypothetical sources in two experimental conditions (Red and Blue). The activity of 12 additional noise sources are not shown.

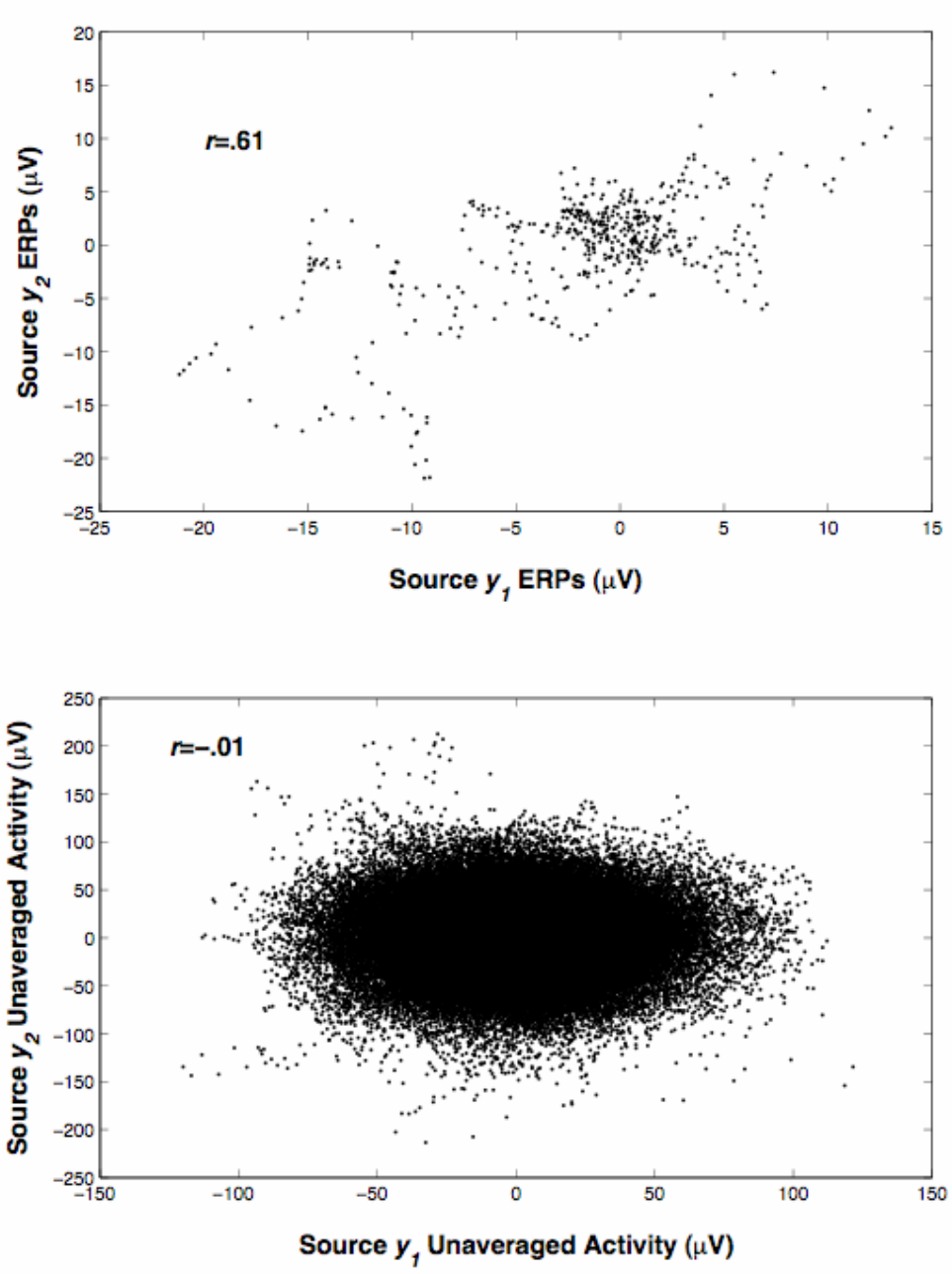

Figure 8 — (**Top**) A scatter plot of the mean activity of Sources 1 and 2 in the two experimental conditions shows that the ERPs of the two sources are correlated. (**Bottom**) A scatter plot of the unaveraged activity of the same sources reveals that the each source's contribution to the raw EEG is somewhat independent of the other. *r* is Pearson's linear correlation coefficient.

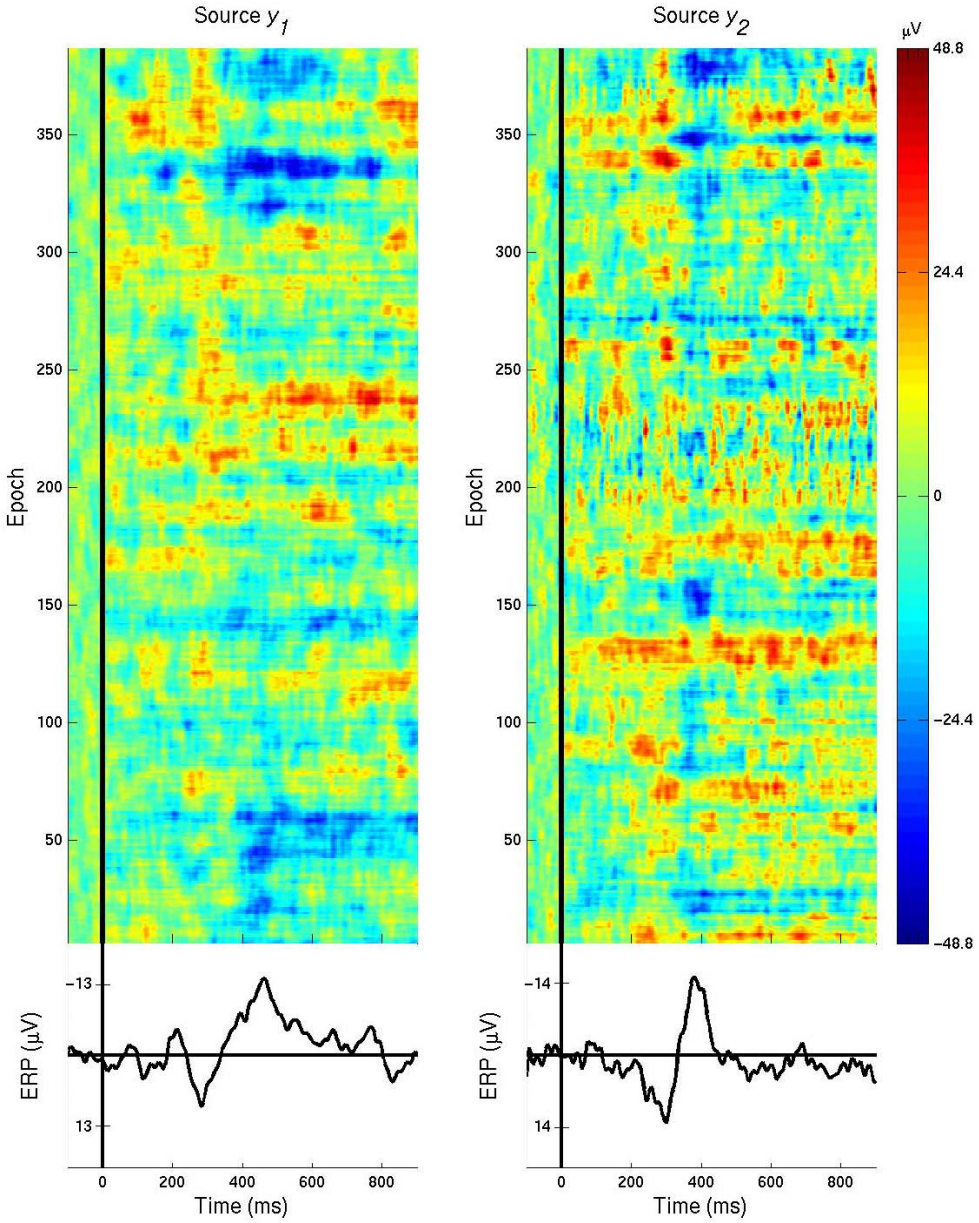

Figure 9 — **(Top)** Single trial activity of Source 1 (**Left**) and Source 2 (**Right**) in both experimental conditions. Each horizontal line represents source activity from a single experimental trial or "epoch." Voltage is represented with color. Single trials are smoothed using a 10 trial vertical moving average. **(Bottom)** Mean single trial activity (ERP) of Source 1 (**Left**) and Source 2 (**Right)**. Although the average activations of the two sources are somewhat similar, the single trial activations are uncorrelated due to the large amount of single trial variation.

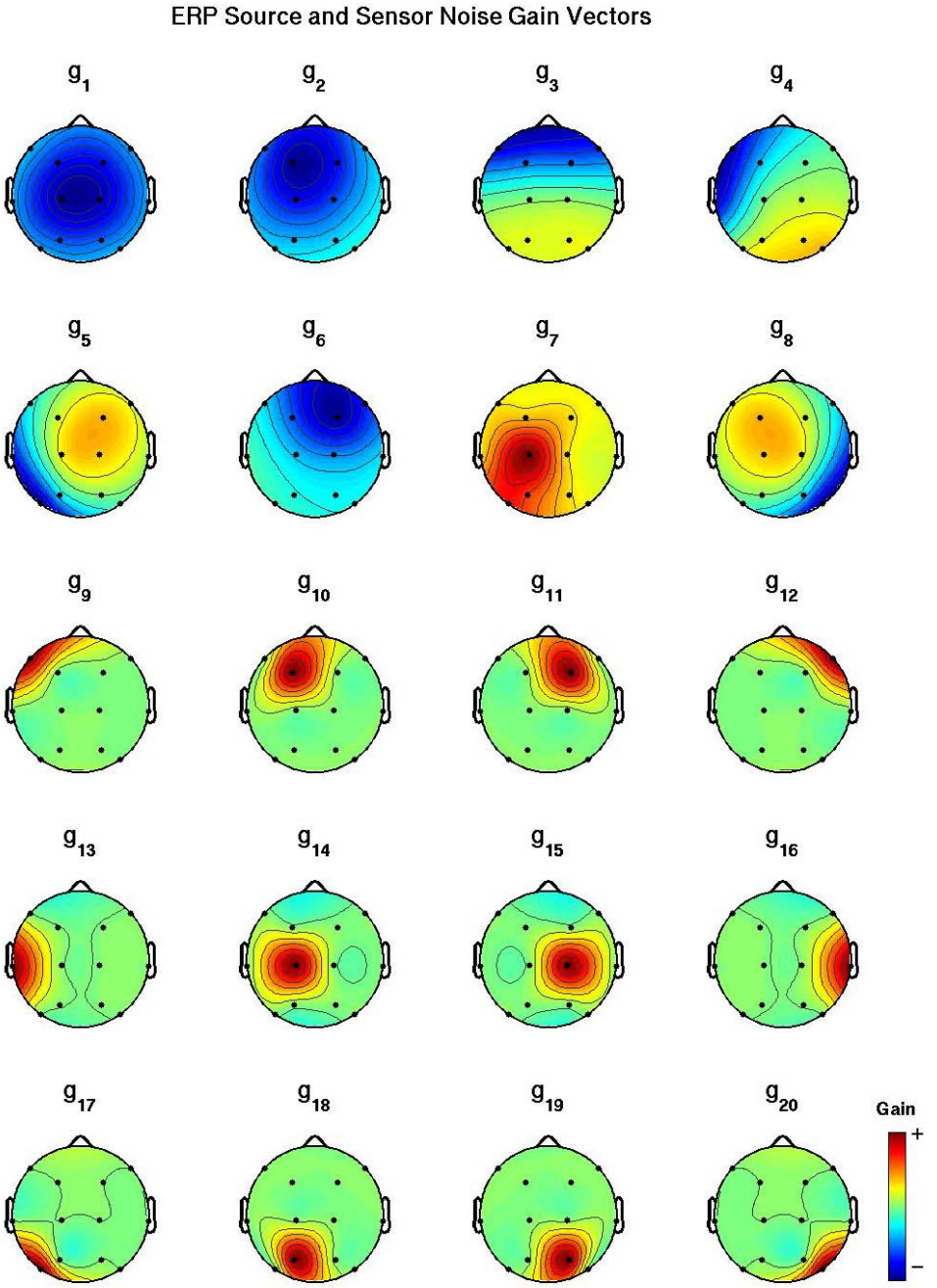

Figure 10 — The first eight topographies correspond to the ERP sources in Figure 7 and the last 12 correspond to sensor noise. Black dots indicate the locations of the 12 electrodes. For visualization, each gain vector has been normalized so that its absolute maximal weight is 1. Note that although the sensor noise is confined to single electrodes, their gain weights appear to extend beyond each electrode due to the interpolation used for visualization.

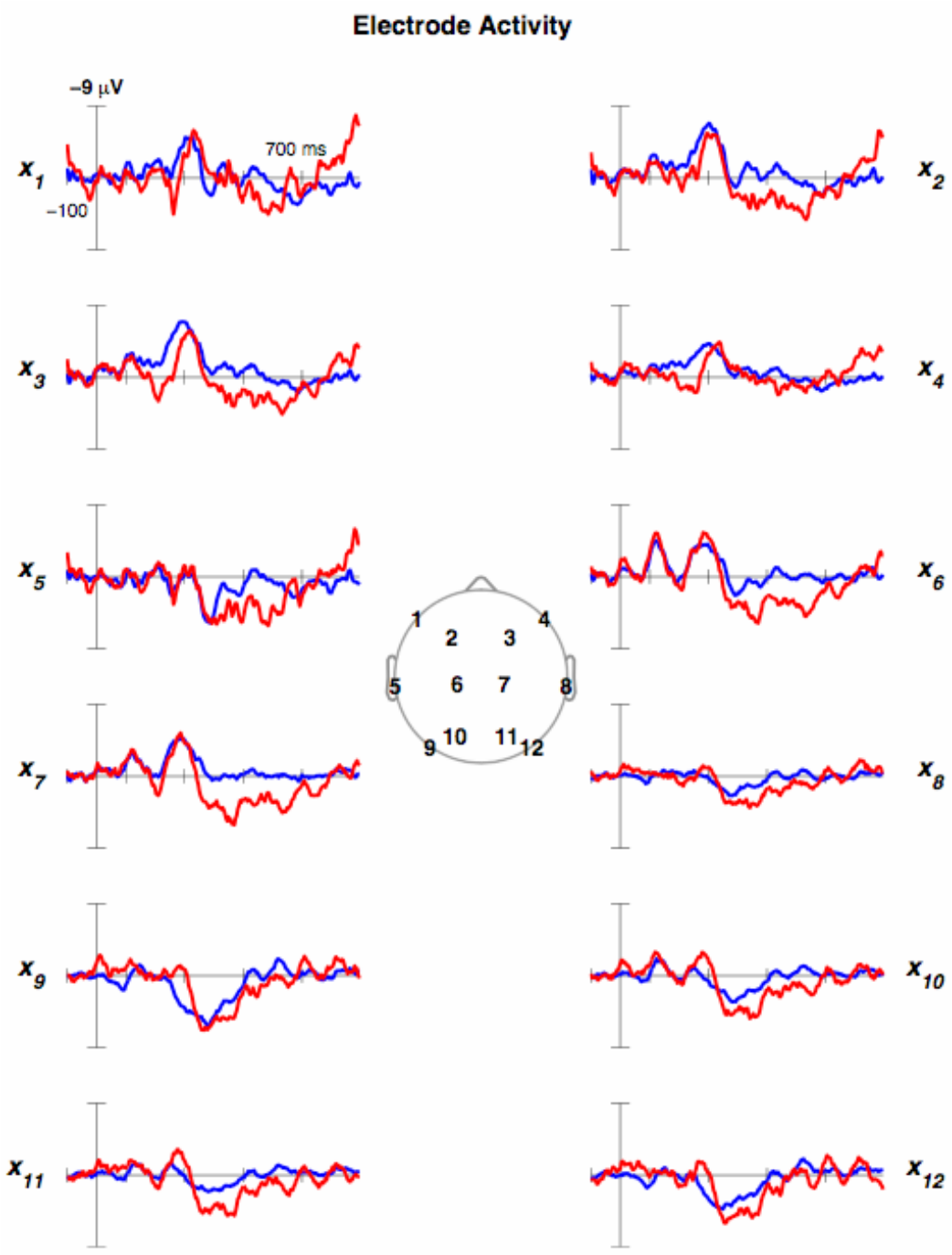

Figure 11 — Mean electrode activity in two experimental conditions (Red and Blue) generated by the sources whose mean activations and topographies are illustrated in Figures 7 and 10, respectively. Cartoon head shows electrode locations.

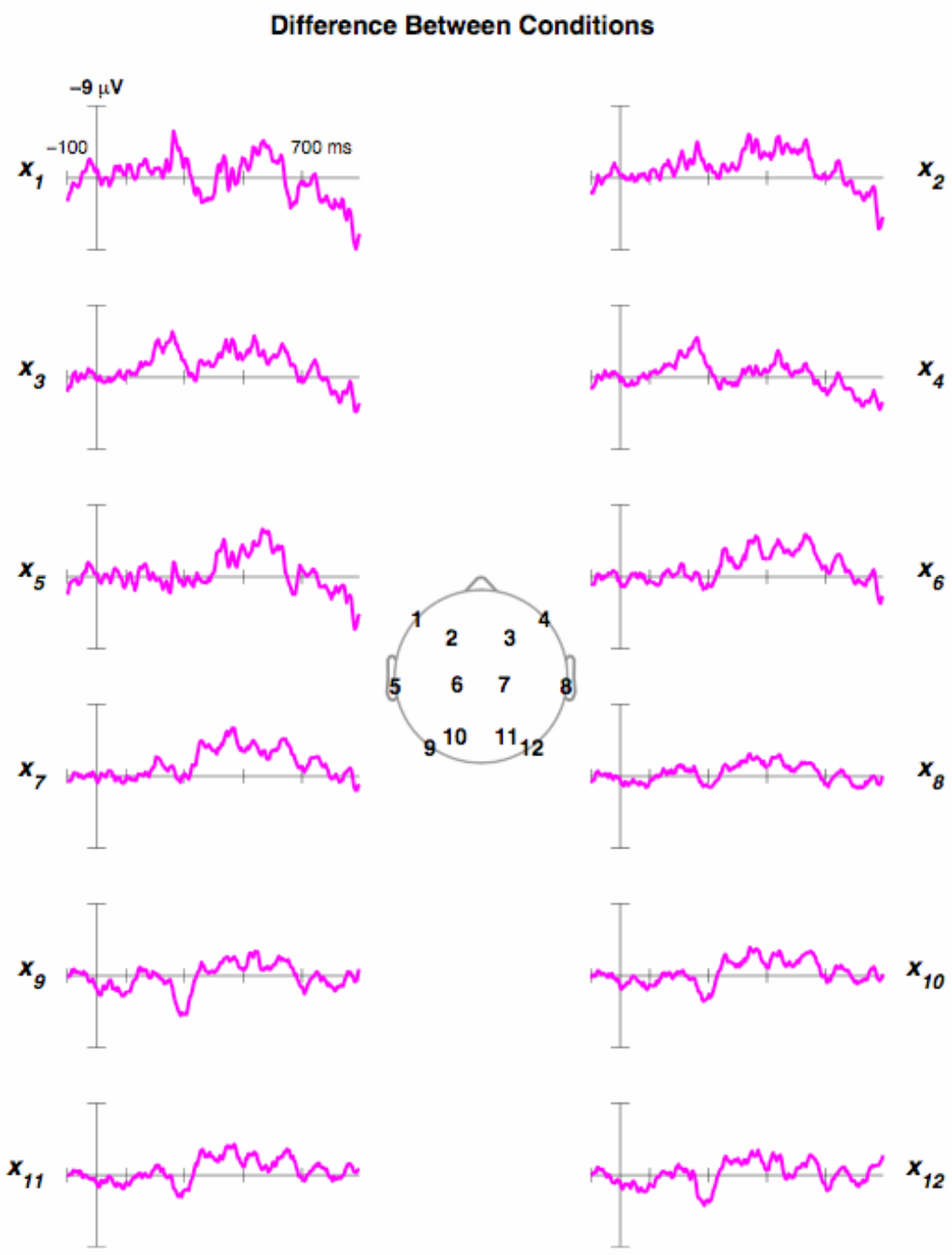

Figure 12 — "Difference waves" obtained by subtracting the ERPs from Condition Red from those from Condition Blue (the two sets of ERPs shown in Figure 11). Cartoon head shows electrode locations.

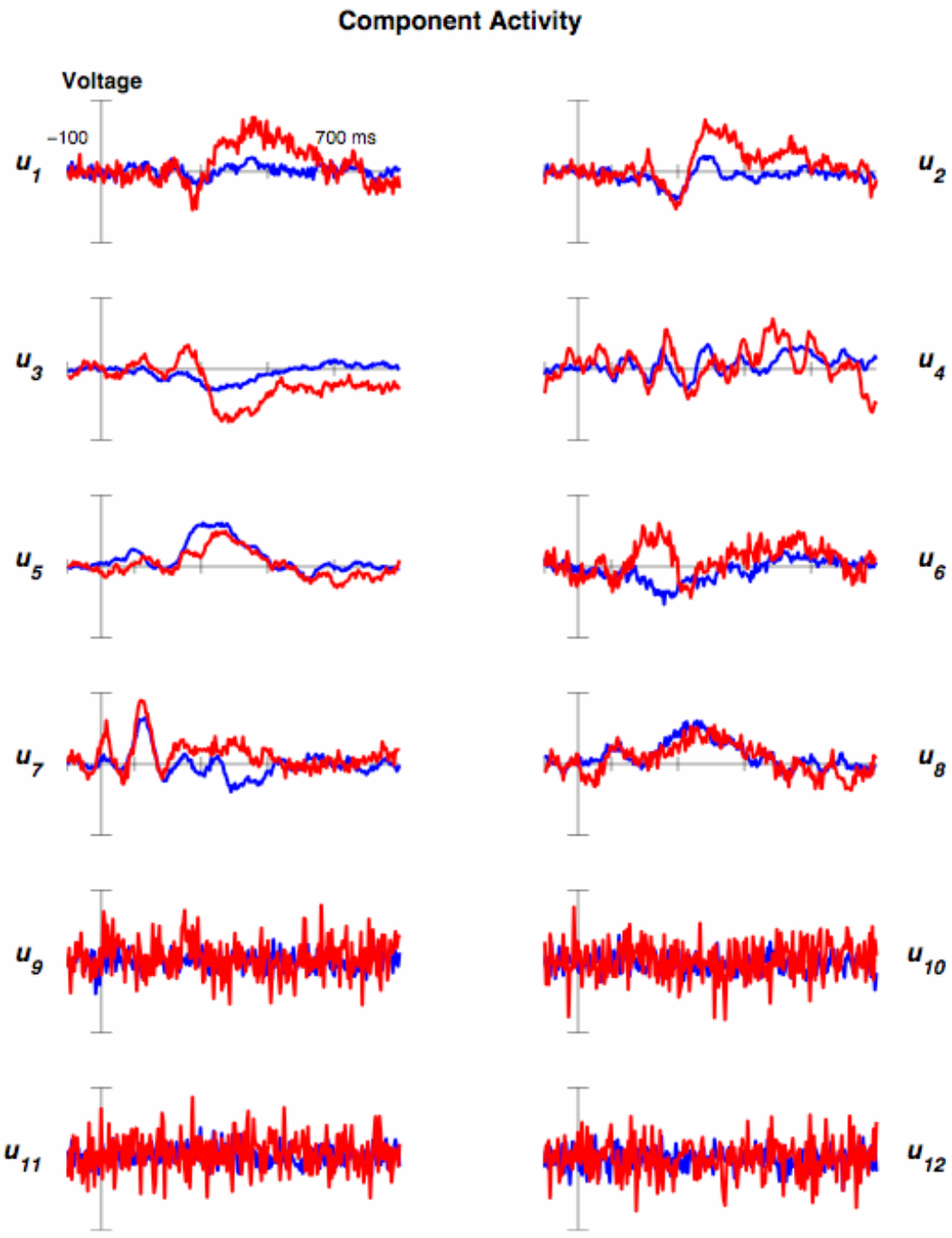

Figure 13 — Independent component activity derived from the simulated EEG from which the ERPs in Figure 11 were derived. For visualization, the average activation of each IC has been normalized to unit variance. The activations of the first eight ICs are similar to those of the eight underlying neural sources (see Figure 7).

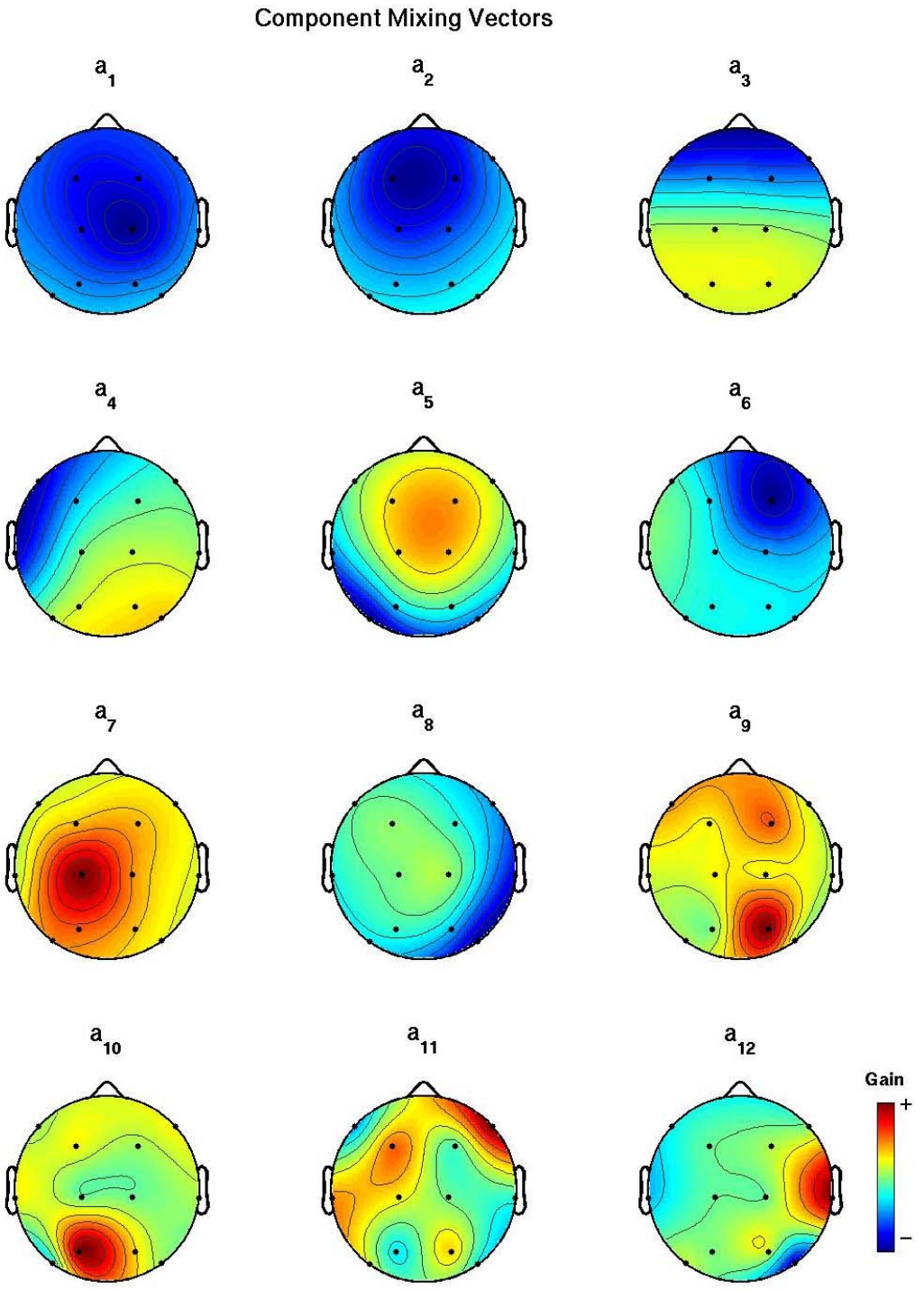

Figure 14 — Mixing vectors (topographies) of the 12 independent components whose activations are shown in Figure 13. Black dots indicate the locations of the 12 electrodes. For visualization, each vector has been normalized so that its absolute maximal weight is 1. The topographies of the first eight ICs are similar to those of the eight underlying neural sources (see Figure 10).

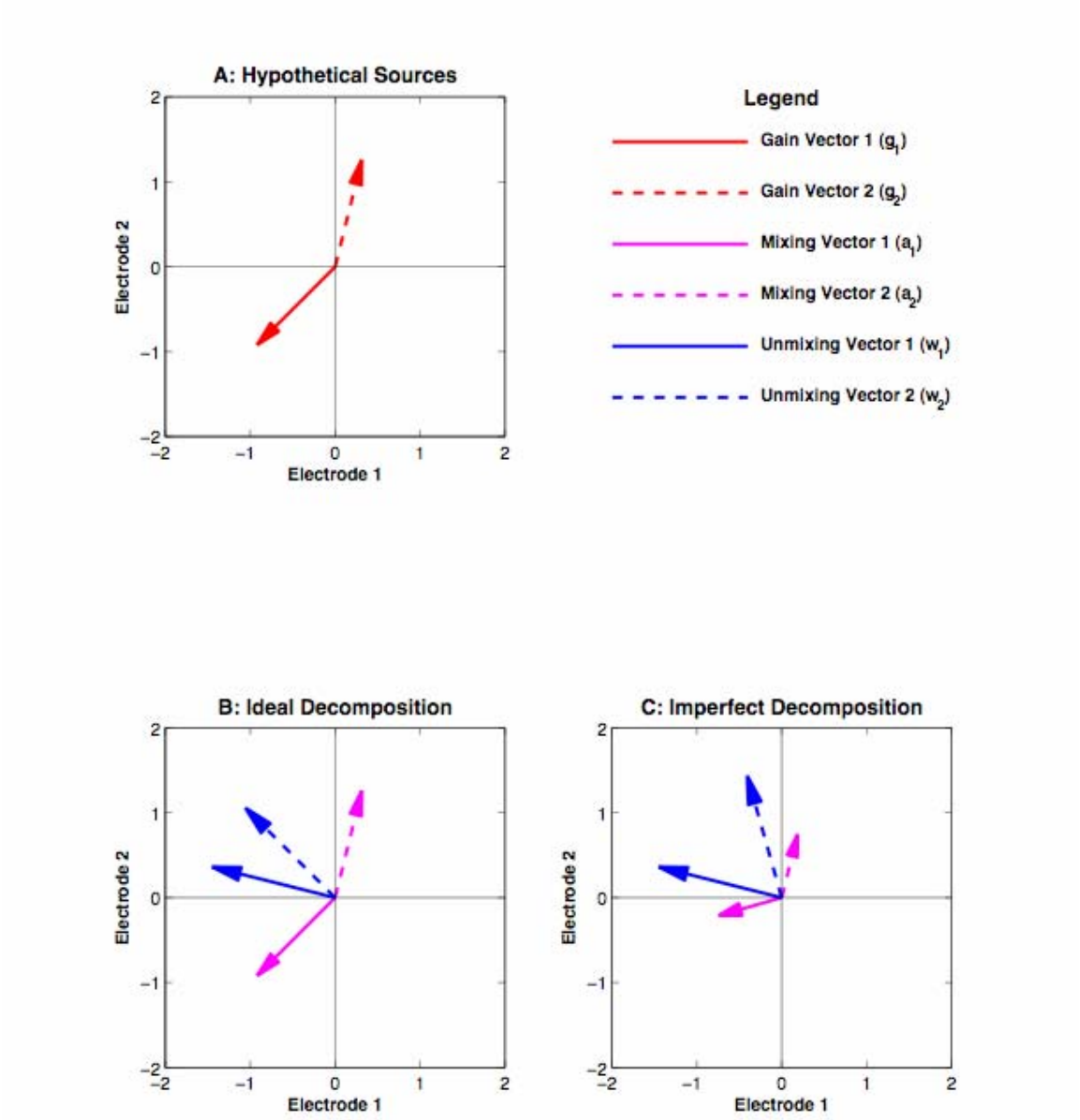

Figure 15 — (**A**) The gain vectors of a hypothetical source configuration. (**B**) An ideal ICA decomposition of the hypothetical source configuration. The mixing vectors are identical to the source gain vectors and each unmixing vector perfectly extracts one source's activity from the other. (**C**) An imperfect ICA decomposition derived from the ideal ICA decomposition by corrupting Unmixing Vector 2. Unmixing Vector 1 still perfectly extracts the activity of Source 1, but Mixing Vector 1 no longer matches the gain vector for Source 1. Conversely, Mixing Vector 2 still matches the direction of the Source 2's gain vector, but Unmixing Vector 2 will no longer perfectly extract the activity of Source 2.

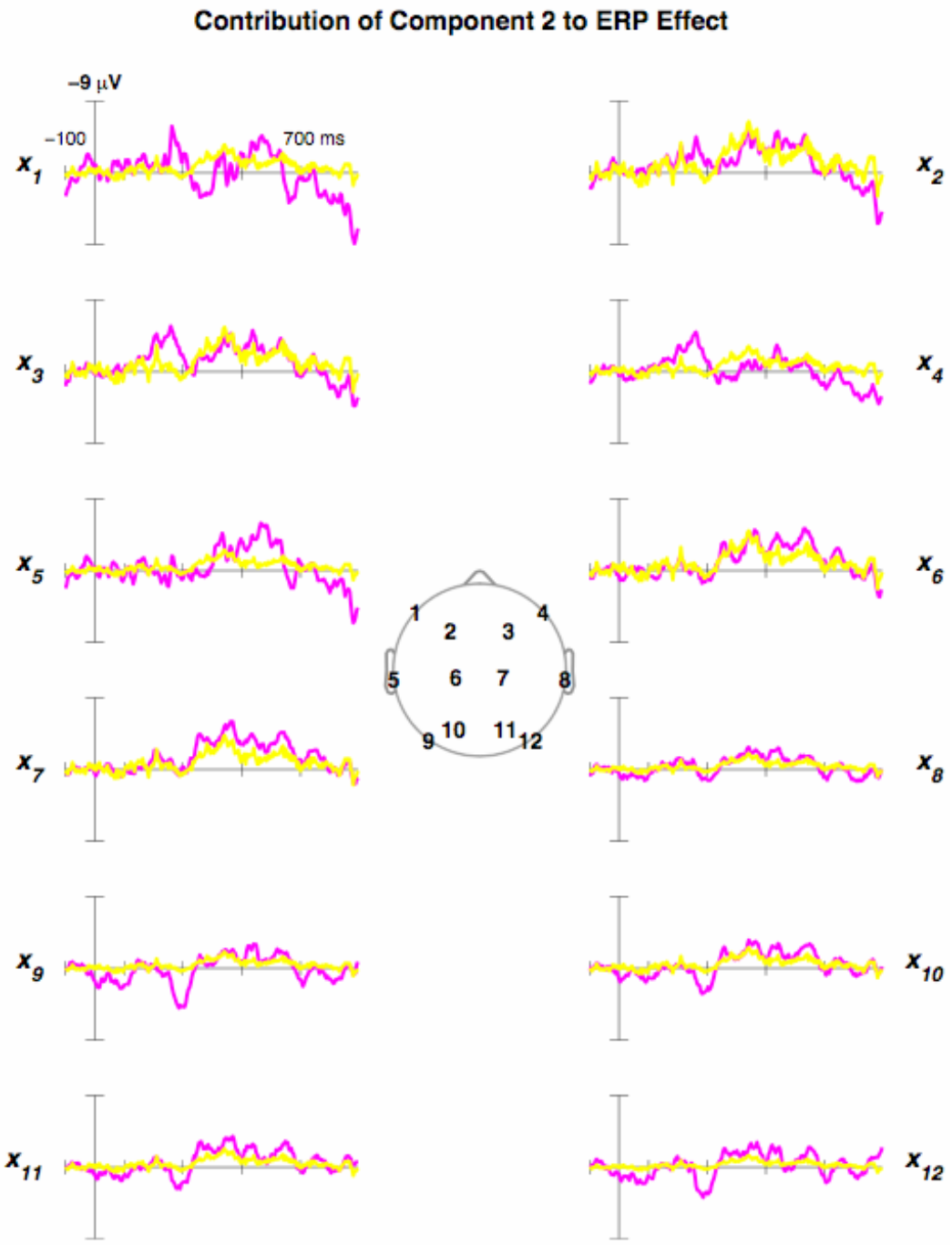

Figure 16 — The contribution of IC 2 (yellow) to the difference waves from Figure 12 (pink). Cartoon head shows electrode locations.

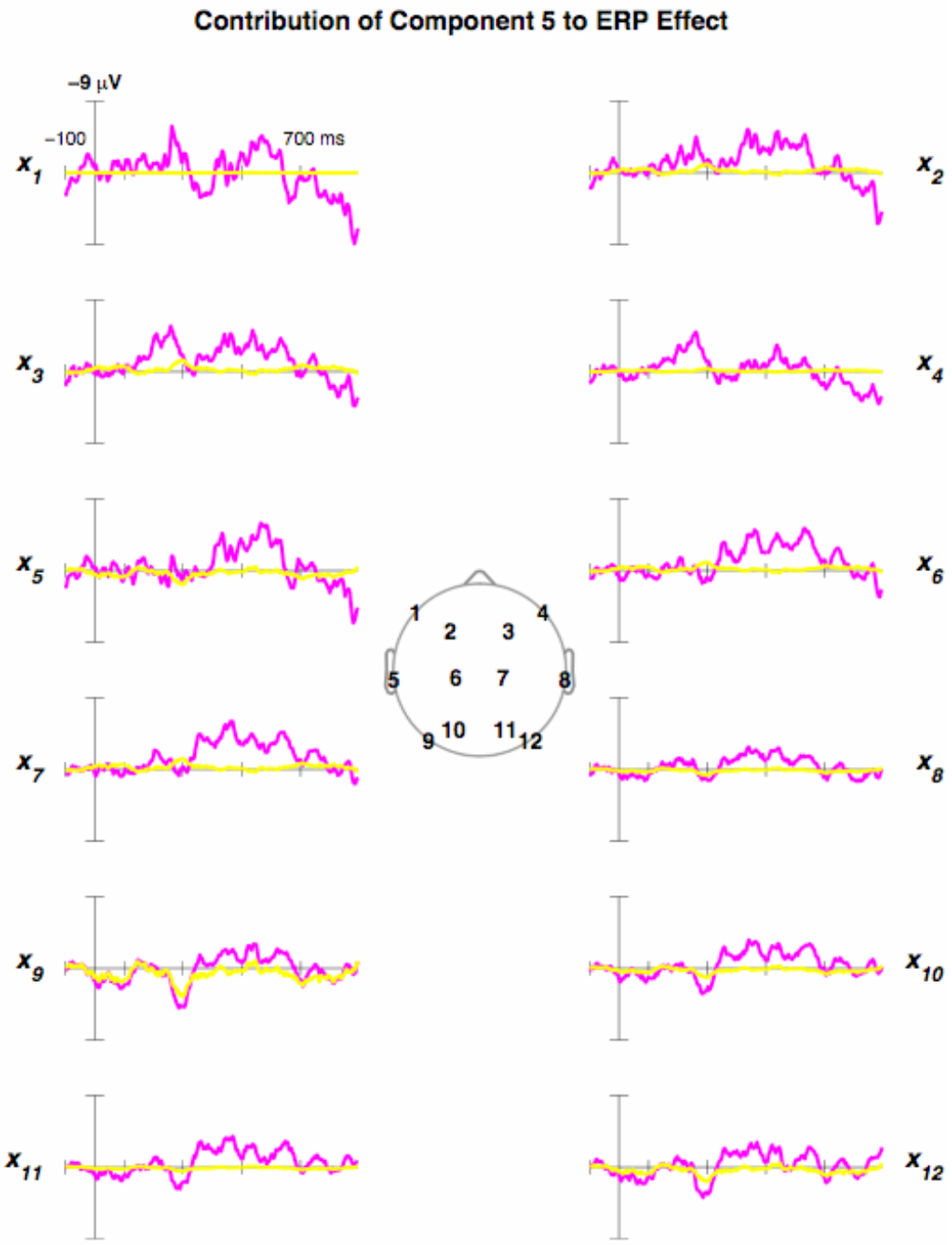

Figure 17 — The contribution of IC 5 (yellow) to the difference waves from Figure 12 (pink). Cartoon head shows electrode locations.

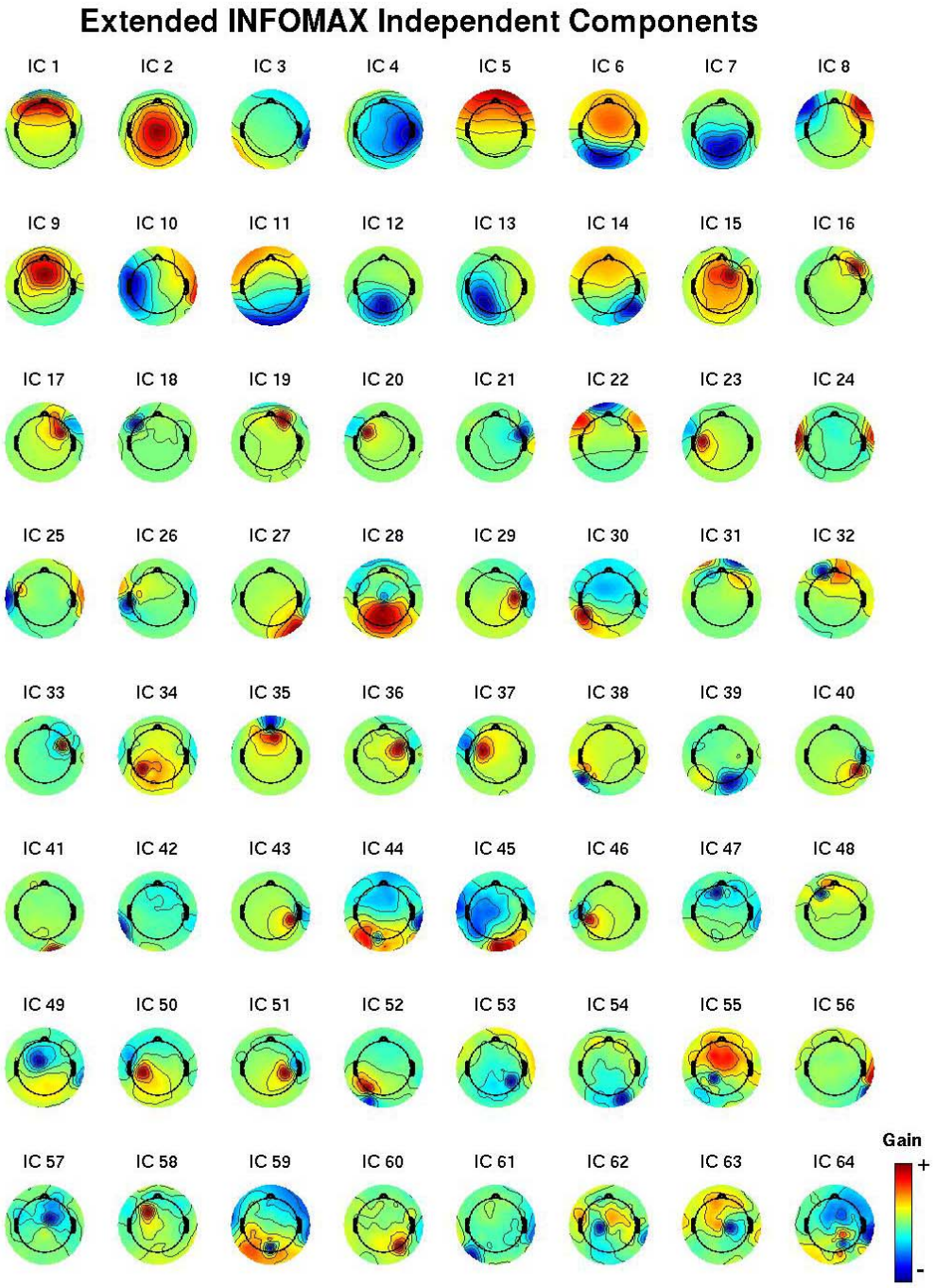

Figure 18 — Mixing vectors (topographies) of the 64 independent components derived from the EEG of a person performing visual oddball and sentence comprehension tasks (Groppe, 2007). For visualization, each vector has been normalized so that its absolute maximal weight is one and the 64 electrodes are not shown. Vector weights below the head's equator are plotted progressively beyond the radius of the head. ICs are ranked in decreasing order of scalp-projected variance (summed across all electrodes).

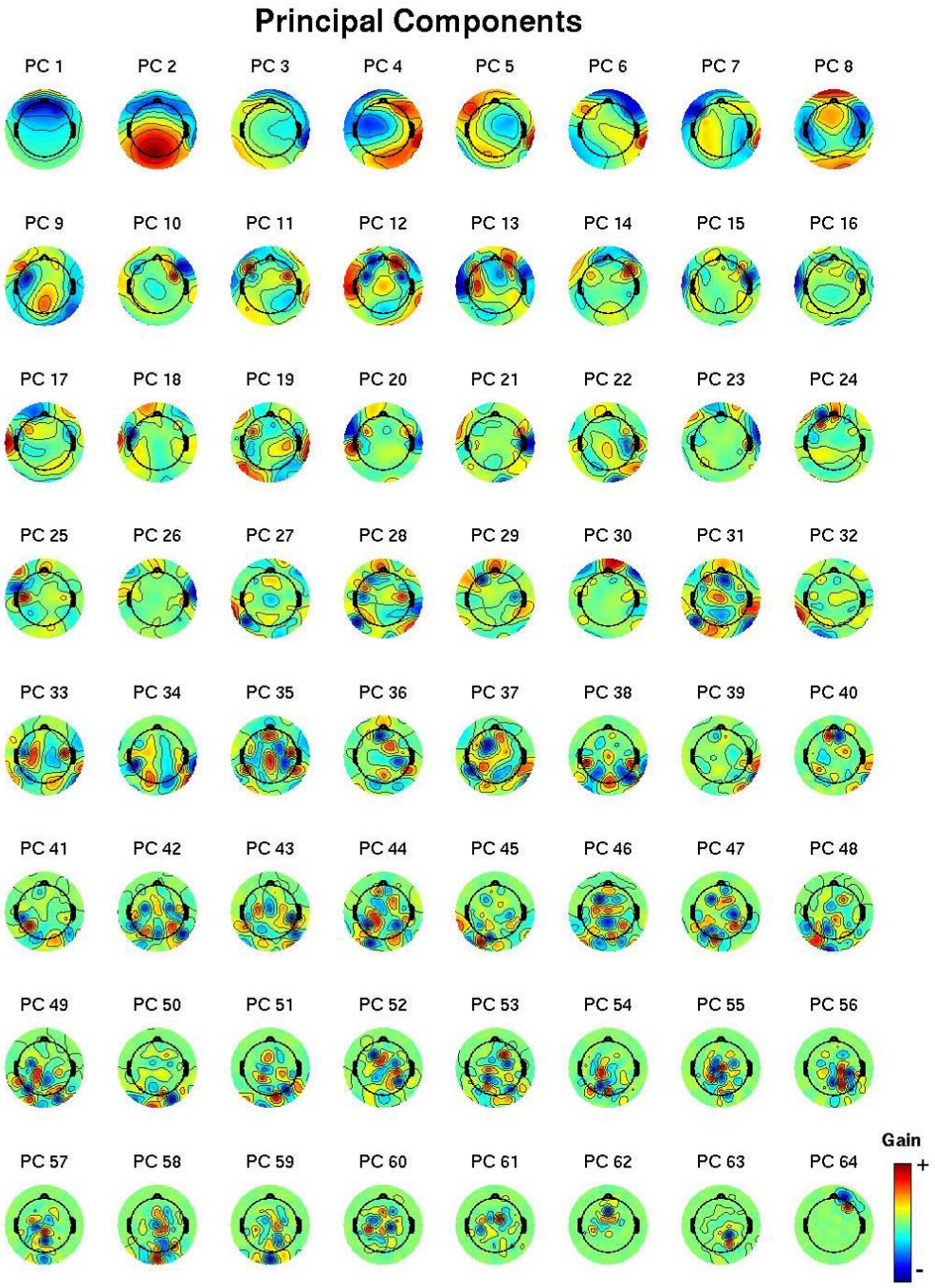

Figure 19 — Mixing vectors (topographies) of the 64 principal components derived from the same data from which the ICs in Figure 18 were derived. These PCs are the "canonical" PCs (Section 5). For visualization, each gain vector has been normalized so that its absolute maximal weight is one and the 64 electrodes are not shown. Vector weights below the head's equator are plotted progressively beyond the radius of the head.

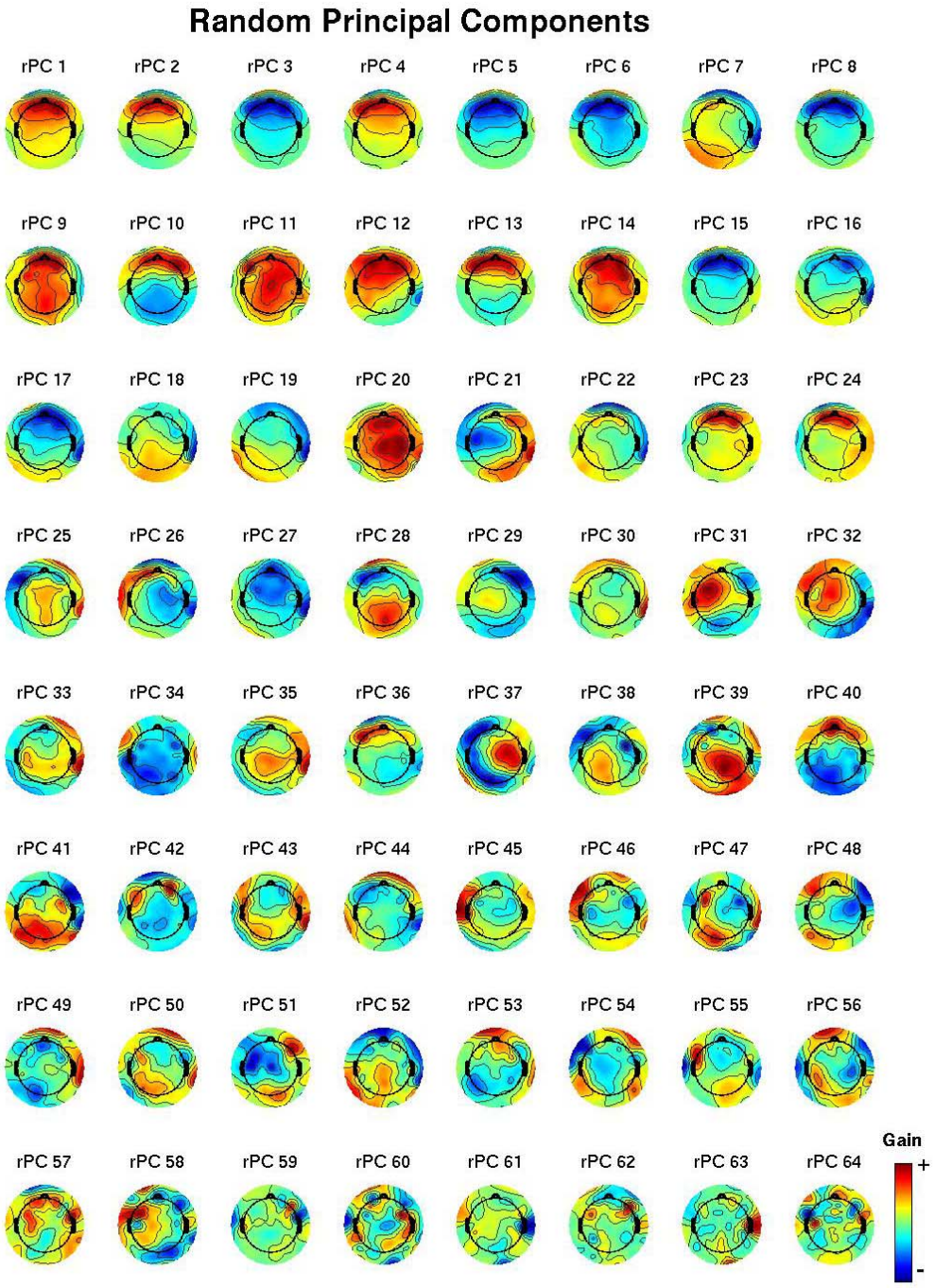

Figure 20 — Mixing vectors (topographies) of a random rotation (Devroye, 1986, pg. 607) of the principal components displayed in Figure 19. Like the original PCs, this decomposition unmixes the data from which they were derived into uncorrelated components. For visualization, each gain vector has been normalized so that its absolute maximal weight is one and the 64 electrodes are not shown. Vector weights below the head's equator are plotted progressively beyond the radius of the head. Random PCs are ranked in decreasing order of scalp-projected variance (summed across all electrodes).

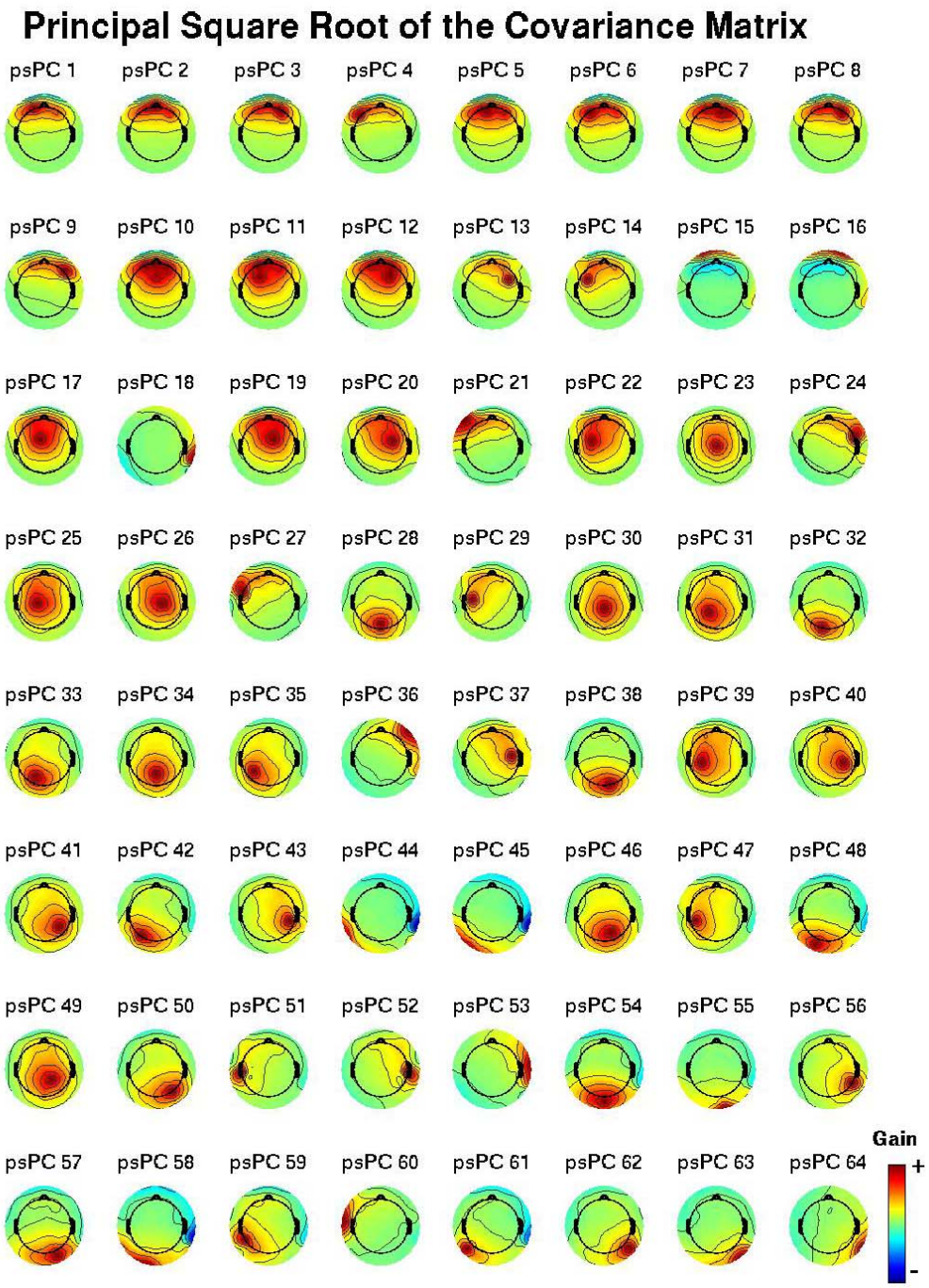

Figure 21 — Mixing vectors (topographies) of the 64 components of the principal square root of the covariance matrix of the same data used to derive the decompositions in Figures 18-20. Like PCs, this decomposition unmixes the data from which they were derived into uncorrelated components. For visualization, each gain vector has been normalized so that its absolute maximal weight is one and the 64 electrodes are not shown. Vector weights below the head's equator are plotted progressively beyond the radius of the head. Components are ranked in decreasing order of scalp-projected variance (summed across all electrodes).