

Evaluation of PCA and ICA of Simulated ERPs: Promax vs. Infomax Rotations

Joseph Dien,^{1*} Wayne Khoe,² and George R. Mangun³

¹Department of Psychology, University of Kansas, Lawrence, Kansas

²Department of Neurosciences, University of California, San Diego, California

³Center for Mind and Brain and Departments of Neurology and Psychology,
University of California, Davis, California

Abstract: Independent components analysis (ICA) and principal components analysis (PCA) are methods used to analyze event-related potential (ERP) and functional imaging (fMRI) data. In the present study, ICA and PCA were directly compared by applying them to simulated ERP datasets. Specifically, PCA was used to generate a subspace of the dataset followed by the application of PCA Promax or ICA Infomax rotations. The simulated datasets were composed of real background EEG activity plus two ERP simulated components. The results suggest that Promax is most effective for temporal analysis, whereas Infomax is most effective for spatial analysis. Failed analyses were examined and used to devise potential diagnostic strategies for both rotations. Finally, the results also showed that decomposition of subject averages yield better results than of grand averages across subjects. *Hum Brain Mapp* 28:742–763, 2007. © 2006 Wiley-Liss, Inc.

Key words: principal components analysis; independent components analysis; event-related potentials

INTRODUCTION

Principal components analysis (PCA) is a multivariate technique that seeks to uncover latent variables responsible for patterns of covariation in numerical datasets [Gorsuch, 1983; Harman, 1976]. It has long been used as a data description and reduction technique to manage the copious quantities of measurements obtained in event-related potential (ERP) studies [Donchin and Heffley, 1979; Möcks et al., 1991]. Although it has been shown to have limita-

tions when applied to ERP data [Wood and McCarthy, 1984] and to be sensitive to parameters like component overlap and correlation [Dien, 1998a], it has nonetheless been utilized with reasonable success in numerous studies when applied in a judicious fashion [Dien, 1999; Dien et al., 1997, 2003a; Spencer et al., 2001; Squires et al., 1975].

Recognition of the limitations of the PCA procedure has given rise to efforts to improve on the process. It has been shown in simulations, for example, that the oblique rotation Promax results in more accurate results with correlated ERP components than the more customary orthogonal rotation Varimax [Dien, 1998a; Dien et al., 2005]. The use of a covariance matrix for the relationship matrix [Kayser and Tenke, 2003] and the inclusion of Kaiser normalization also yield improved results in comparison to using covariance loadings during rotation [Dien et al., 2005].

Recently, a related but quite different procedure called independent components analysis (ICA) has been proposed as an alternative to PCA and some promising results have been reported with both ERPs [Jackson, 1991;

Contract grant sponsor: National Institute of Mental Health (NIMH); Contract grant number: MH11751 (to J.D.); Contract grant numbers: MH55714 and MH02019 (to G.R.M.).

*Correspondence to: Joseph Dien, Department of Psychology, University of Kansas, Fraser Building, 1415 Jayhawk Blvd., Lawrence, Kansas. E-mail: jdien@ku.edu

Received for publication 15 August 2005; Accepted 15 May 2006

DOI: 10.1002/hbm.20304

Published online 28 November 2006 in Wiley InterScience (www.interscience.wiley.com).

Jung et al., 2000; Makeig et al., 1996, 1997, 1999a,b; Vigario, 1997] and hemodynamic measures [Calhoun et al., 2001; Dodel et al., 2000; McKeown et al., 1998; Park et al., 2003]. There has been interest in how the two techniques compare. It is not possible to state which will be more effective for ERP datasets on the basis of the statistical principles alone. Makeig et al. [1997: 10979] noted that ICA "... requires the absence of higher-order as well as second-order correlations between time courses ... [and] is a stronger condition than decorrelation ..."; however, since the default setting in their implementation of ICA is to remove the second-order relationships prior to ICA decomposition via sphering and then to return them afterwards, ICA (used in this fashion) relies on different statistical information from PCA, rather than more statistical information. Indeed, the ability of ICA components to be correlated is a strength of the technique, in contrast to Varimax-rotated PCA solutions [Jung et al., 2000: 1756].

The goal of this exercise is not to determine which is globally better for all purposes, which would be an ill-posed question. Rather, this report will attempt to determine the relative characteristics of the two techniques that will allow investigators to determine which tool to use for a given project. Every statistical technique is based on certain implicit assumptions upon which the model is constructed; the effectiveness of a statistical technique is most often determined by the fit between the statistical assumptions and the characteristics of the datasets. We shall examine the unique aspects of ERP datasets, especially the distinction between using time points or electrodes as variables, and how they relate to the statistical assumptions. This report will first provide a brief review of the algorithms underlying PCA and ICA, followed by a series of tests using simulated and real data.

PRINCIPAL COMPONENTS ANALYSIS

Since comprehensive treatments of PCA are available elsewhere [Gorsuch, 1983; Harman, 1976], this review will focus on highlighting the aspects relevant to the present comparison. Further information on its application to ERP datasets is also available elsewhere [Dien and Frishkoff, 2004; Donchin and Heffley, 1979; Möcks and Verleger, 1991]. PCA has the ultimate purpose of expressing a dataset as a set of linear combinations of variables that are more interpretable, which is to say, relate simply to the latent variables rather than being some sort of complex combination of them. In the case of ERP data, some of these linear combinations would ideally correspond to the ERP components of interest. The linear combinations produced by PCA, as well as ICA, are conventionally termed "components" but in the remainder of this report will be termed "factors" to avoid confusion with ERP "components."

The core procedure of PCA is the decomposition of the so-called relationship matrix. The relationship matrix, typi-

cally a correlation or covariance matrix, summarizes the relationships between each variable and every other variable. In a correlation matrix, the full set of variables is represented by the rows and again by the columns. The entry for each cell of the matrix is the correlation between the two variables represented by the respective row and column. The diagonal of the matrix is the correlation of each variable with itself (unity). A covariance matrix is the same as a correlation matrix except that the variables have not been standardized so that the magnitude of the entries reflects the size of the variable variance as well as the degree of covariation.

The PCA algorithm sequentially fits a linear combination to this matrix that accounts for the greatest possible variance. The matrix is then "residualized," which means that the linear combination is subtracted out, leaving behind the data that has not been accounted for yet, and then the process is repeated with the remaining matrix. In this fashion the dataset is reexpressed as a set of linear combinations (of equal number to the original variables in the absence of collinearity) arranged in order of decreasing size. These factors are uncorrelated with each other, regardless of the nature of the underlying data. The smallest (presumably uninterpretable) factors are then dropped from further analysis.

A rotation procedure is then utilized to increase interpretability of the obtained factors. This step is necessary since the statistically derived factors will usually be linear combinations of the actual latent variables of interest (combinations of different ERP components in the present case). For example, the Varimax rotation [Kaiser, 1958] translates the factors to a mathematically equivalent set of linear combinations, maximizing the variance of the squared factor loadings. This has the effect of generating factors that are as close to zero on some variables as possible, while as large as possible on the others; this may reasonably be expected to yield a solution in which the factors more closely correspond to single ERP components since ERP components are nominally zero on most time points and maximal in a limited set of time points. This process can be graphed as a scatterplot in which each point represents a single variable and the axes represent the two factors. The rotation process rotates the axes of the coordinate system such that the axes pass through the densest groupings of points (which is equivalent to saying that the rotation will arrange for the factor loadings of each variable to be large for one factor and small for the other as much as possible). This process proceeds iteratively for each pairwise combination of the variables until a pass through the full set of pairwise rotations results in rotations that fall below a low criterion point.

The Promax rotation [Hendrickson and White, 1964] utilized in this report performs an initial Varimax rotation and then relaxes the orthogonality restrictions, allowing the factors to become correlated. It does so computationally by rotating individual factors such that they approximate more closely a version of themselves taken to a

higher power (such as a fourth power); in other words, enhancing the large loadings relative to the smaller loadings. Graphically, this is equivalent to saying that each axis is rotated individually without attempting to maintain them at right angles to each other. The higher the power, the greater is this final rotation. If the underlying latent variables, like the ERP components, are in fact correlated, then this can allow for a more accurate solution [Dien, 1998a; Dien et al., 2003b, 2005].

Assumptions and Issues in PCA of ERPs

PCA does not make any strong assumptions about the data. No assumptions are made about the distribution of the variables or of the factor scores [Gorsuch, 1983: 24]. The only assumption is that the variables are linear functions of the factors. Variables do not even need to be linearly related as long as the assumption is met [Gorsuch, 1983: 18]. There is no particular reason to think that this assumption will be violated for ERP datasets.

However, several issues need to be considered for PCA to be successful, the first of which is factor overlap. Factors are defined as being a specific pattern of factor loadings (such as a particular time course for a temporal PCA or a particular scalp topography for a spatial PCA). ERP components that have an identical pattern (such as both peaking at 300 ms for a temporal PCA) cannot by definition be separated into different factors, even if they are separable by some of the variance present in the observations, such as condition variance. The more similar the two components, the more difficult it may be to successfully separate them. This is likely a greater concern for spatial PCA since volume conduction (the property of voltage fields of spreading throughout the conductive medium of the head) ensures that every electrode will be affected by a component and hence every component overlaps substantially with every other component [Dien, 1998a]; conversely, components in the time domain can be completely separate.

A second issue is that of factor correlation. The initial factor decomposition and Varimax rotation are both orthogonal, meaning that the factors are constrained to be uncorrelated even if the actual ERP components are correlated. Such a constraint causes the statistical model to be distorted in order to force the factors to be orthogonal. This issue can be addressed, sometimes quite effectively, by the use of the Promax rotation, which adds a relaxation step [Dien, 1998a; Dien et al., 2003b, 2005]. What is not clear is to what degree this relaxation step can be effective. It is likely that this procedure will only be effective up to some unknown degree of factor correlation. Given the increased amount of overlap found in the spatial dimension, it is expected that this will be a greater concern for temporal PCA insofar as degree of spatial overlap induces factor correlation for temporal PCA, and vice versa, since it determines the extent to which the two components occur in the observations [Dien, 1998a]. Lack of spatial

overlap would induce a negative correlation (observations containing one component would not contain the other component), but it would be diluted by the number of observations containing neither, of which there would be many in most temporal PCAs.

A third issue that can arise is “misretention,” leading to either underextraction or overextraction [Fava and Velicer, 1992; Wood et al., 1996]. This occurs when too few or too many factors are retained for rotation compared to the actual number of substantial latent variables in the dataset. Underextraction can cause ERP components to be combined into a single factor, whereas overextraction can cause minor (perhaps noise) factors being built up at the expense of the major (ERP component) factors and/or factors with only one high loading [Comrey, 1978]. Careful attention to the use of factor retention rules [see Dien, 1998a] and evaluation of factor results are required to address this issue.

Two final issues have been identified for Varimax rotations that could affect the present simulations [Cureton and Mulaik, 1975: 224]. The first occurs when the bulk of the variables load on both factors. One way of describing this issue is by saying that Varimax makes an implicit assumption that there will be large clusters of variables that load only on one or the other factor; if this is not the case, then the rotation will not occur properly. The second occurs when a number of variables have zero loadings on the first unrotated factor. In this case the factor is essentially “pinned” against rotation since Varimax requires that for rotation to occur, the criterion must be increased for each pairwise rotation. In the language of connectionist models, the solution becomes trapped at a local minimum and cannot reach the global minimum. This situation only applies for factor solutions with at least three dimensions. The first unrotated factor typically has loadings on as many of the variables as possible, so it is not clear how often this situation occurs.

A variant of the Varimax rotation, the weighted-Varimax, has been proposed to address these two situations [Cureton and D’Agostino, 1983; Cureton and Mulaik, 1975]. It gives the most weight to factor loadings that are located away from the initial unrotated factor (which is by far the largest), essentially making the assumption that the initial rotation is not aligned with the correct rotation. It is not clear in advance how problematic these two situations might be for spatial and temporal PCAs of ERP data, so it seems worthwhile to evaluate this rotation as well. Since Promax uses Varimax as an initial rotation, we implemented Promax with Weighted-Varimax to supplement the regular Promax with Varimax rotation.

INDEPENDENT COMPONENTS ANALYSIS

Independent components analysis provides an alternative approach to isolating ERP components. Since there are many varieties of ICA, this report will focus on the version

most commonly applied to ERP data, the Infomax rotation [Bell and Sejnowski, 1995], as implemented by the EEGLab toolkit [Delorme and Makeig, 2004]. Since in-depth mathematical treatments already exist [Bell and Sejnowski, 1995; Makeig et al., 1997], this brief review will focus on a more applied description of the algorithm and its implications (as instantiated in the EEGLab software).

A source of confusion for psychologists when discussing ICA is the use of a different terminology grounded in the engineering literature. To reduce reader confusion, for the remainder of this text the equivalent terms from the PCA literature, as summarized in Table I, will be used to refer to both PCA and ICA. Another source of confusion for psychologists is that, unlike PCA, there is no separate extraction step; the rotations can be directly applied to the starting variables. Thus, the term “PCA” applies to the successive steps of extraction and rotation. In contrast, the term “ICA” in effect applies only to a rotation procedure, since no extraction is required (although, as will be discussed below, a PCA extraction may be used as a preprocessing step for ICA).

A fundamental difference between the PCA and the ICA procedures concerns the matrix being evaluated. In PCA, during the rotation stage the matrix being evaluated is the loading matrix, which represents the relationship between the factors and the variables; the rotation alters the matrix until the factor loadings meet the criterion (such as the Varimax criterion of maximizing the variance of the squared loadings). In ICA, the procedure evaluates the matrix of factor interrelationships; in other words, the factor scores rather than the factor loadings. The factors are systematically rotated until the relationships between the factors are as close to zero (i.e., independent) as possible.

The ICA algorithm begins by generating factor scores that are initially set equal to the variables (one for each). A relationship matrix is then generated between these factor scores. The factor scoring matrix is then modified such that factors that are different from each other are made even more different. Through a sometimes lengthy training process the factor scoring matrix is modified. New factor scores are generated and used to compute a new relationships matrix; this process is repeated until the changes to the factors drop below a criterion threshold. In this manner the relationships between the factors are gradually reduced as they become increasingly differentiated from each other.

TABLE I. PCA and ICA glossary

PCA	ICA
Factor loading matrix	Mixing matrix
Factor scoring coefficient matrix	Separation matrix
Factor scores	Activations

PCA, principal components analysis; ICA, independent components analysis.

Another difference is the metric by which these relationships are measured. A factor loading matrix, as used in PCA, can be thought of as containing the regression weights needed to predict the variables from the factors. Formally, correlation coefficients are the same as the regression weight needed to predict one variable by the other if the two variables are standardized: $Y = rX$ (where Y is the variable and X is the factor and r is the regression weight). In the ICA relationships matrix, the entries reflect the higher moments as well, such as the third moment: $Y = rX^2$ (keeping in mind that in this case Y is a factor, like X , rather than a variable). These higher-order relations are represented by an exponential sigmoid function that has the form of $y = 1./(1 + \exp(-u))$ and runs from -1 to 1 after some rescaling ($2*y - 1$), where $u =$ the factor score and y is the sigma-transformed factor score. Just like with a correlation, a positive score means a tendency to vary in the same direction and a negative score means a tendency to vary in the opposite direction.

Another issue is that in the relationship matrix the columns are the factor scores and the rows are the sigmoid (sig) transformed versions of the factor scores (fac). This means that the relationship is asymmetric, with the relationship between each factor represented by two numbers (i.e., the product of fac1 and sig[fac2] and the product of sig[fac1] and fac2). In the subsequent rotation step the first value determines the effect of the first factor on the second factor, whereas the second value determines the effect of the second factor on the first.

The ability for one factor to predict another based on these higher moments is related to its Gaussianity. Along the diagonal of the matrix, the entries represent the Gaussianity of the factors. A perfectly Gaussian factor would have a score of zero. The off-diagonals represent the non-Gaussianity of the two factors (i.e., the scores will be maximal when both factors are non-Gaussian and in the same way, which means that the two factors will be related through the higher-order relationships). With each iteration the degree to which a factor will be rotated depends on the relative difference between its diagonal (how much it will stay the same) and the off-diagonals (how much it will change). The more non-Gaussian a factor is, the less it will be rotated. This approach is based on the Central Limit Theorem, which indicates that a mix of two latent variables should be more Gaussian than the pure variables; maximizing non-Gaussianity of the factors should therefore maximize how purely they reflect a single latent variable [Hyvärinen et al., 2001: 9].

The sign of the relationship number controls how the factor scoring coefficients, and hence the factor scores, are changed at each iteration of the process. If the relationship number is positive, a fraction of the second factor’s scoring coefficients are subtracted from those of the first. The more similar (and hence more positive the number), the more is subtracted. The reverse happens if the relationship number is negative (the two are similar in a mirror-like fashion). If the factors started out similar, this process will push them

apart from each other. Ultimately, this process reaches an equilibrium where the changes to the two factors cancel out.

The strongest relationship is likely to be the second-order correlations. For this reason, the default approach is to decorrelate the matrix by “sphering” the data by using matrix division to divide it by the covariance matrix [see also Hyvärinen et al., 2001: 160]. The result is to eliminate the second-order covariances (they now equal zero) but leaving intact the higher-order relations (e.g., the product of the variable with its transformed version is not zero).

The sphering operation also has the important effect of standardizing the data matrix, equalizing the contribution of the different variables to the results, much as PCA normally uses correlational factor loadings at the rotation step. This standardization could, in principle, be performed without also sphering the data; the two do not need to go together. Whereas PCA factor loadings are, by convention, interpreted in correlation form, ICA factor loadings are, by convention, interpreted in covariance form (with microvolt metric). This conversion to microvolt metric automatically occurs when the sphering operation is undone prior to interpretation, and hence simply represents a difference in convention rather than a fundamental difference between PCA and ICA, since PCA factor loadings can also be readily converted to microvolt metric [see Dien et al., 1997].

The sigmoid function also has the purpose of expanding the influence of the most informative part of the data distribution (the center) compared to the outer fringes (the outliers). The outliers are compacted into the floor and ceiling values of -1 and 1 , whereas the central numbers, which may be closely spaced, are spaced further apart in the sigmoid transformed variable. It is this maximization of the information value of the data by this transformation that leads to the name “Infomax” for this ICA algorithm [Bell and Sejnowski, 1995: 1130].

Assumptions and Issues in ICA of ERPs

ICA makes two assumptions about the data. The first is that the data are non-Gaussian in their distribution over different possible values [Hyvärinen et al., 2001: 162], as can be graphed by a histogram, which is to say that they depart from normality. Such non-Gaussian distributions make it possible for the higher-order moments to differentiate the ERP components. It seems likely that most ERP components analyzed in a spatial approach will be highly non-Gaussian, since most of the observations will be zero with just a few time points being nonzero. The actual time course of the components will be relatively unimportant compared to this effect of being temporally circumscribed. It is not as clear what the case will be for a temporal approach.

The second is that the ERP components be independent of each other [Hyvärinen et al., 2001: 152], which means that they should not be only uncorrelated but also unrelated in terms of the higher-order relations, as described in the previous section. One of the prior simulation studies

[Makeig et al., 2000] examined the effect of correlated components on ICA and showed that it can indeed cause distortions in the results. Decorrelating (sphering) the data before the ICA rotation (and recorrelating afterwards) may perhaps address this issue but it remains untested; furthermore, it remains unclear just how independent ERP components tend to be, aside from the second-order correlational moments, or what effect such nonindependence might have on the results. Like with PCA, one would expect that this issue be more serious for the temporal approach. Another situation is when the factor loading matrix is almost singular [Bell and Sejnowski, 1995]. This could happen if two of the factors were too similar, and hence the weights also. This statement is therefore homologous to the issue discussed with regard to PCA that factors that are too similar may be difficult to separate. As with PCA, it is most likely to be an issue for the spatial approach.

The factor loading matrix could also be singular if variables are too similar or if there are more variables than there are latent variables to be modeled by factors. In such a case, the phenomenon of “overfitting” or “overlearning” can occur. One way this phenomenon can be manifest is that factors with isolated temporal bumps that represent portions of a factor’s time series being split between different factors [Hyvärinen et al., 2001; Särelä and Vigário, 2003]. It can also result in effects similar to that described for PCA, such as single ERP components being split into multiple single loading factors.

A final issue more specific to the Infomax algorithm is that it is designed to handle super-Gaussian events that have large amplitudes but limited presence across the observations. When applied as a spatial ICA, typical ERP components meet this description, as they are high amplitude and short duration, which translates to being present in relatively few of the observations (time points). When applied as a temporal ICA, the observations are electrodes and this may no longer be the case. Because ERP components are present in most electrodes due to volume conduction, they should be present in most of the observations (channels) of a temporal ICA. It may therefore be the case that they are better described as having a sub-Gaussian distribution. A variant of the Infomax algorithm, called Extended ICA, has been developed for such cases [Lee et al., 1999]. This variant will also be applied to see if it provides more effective results for temporal ICA.

DIFFERENCES BETWEEN PCA AND ICA

PCA can be applied to ERP datasets using either a temporal [Donchin and Heffley, 1979] or a spatial approach [Dien, 1998a; Kavanagh et al., 1976], a distinction that will play a key part in the present simulation study. In the temporal approach (temporal PCA) the time points are arranged as the variables and the waveforms (combinations of channels, subjects, and conditions) are the observa-

tions. The factors are defined by a specific time course as described by their respective factor loadings. Since the factor loadings are fixed in nature, it is not possible to examine latency changes across conditions or subjects. Scalp topography information, as coded in the factor scores, is free to vary. In the spatial approach (spatial PCA) the channels are the variables and the scalp topographies (combinations of time points, subjects, and conditions) are the observations. With the spatial approach it is possible to examine latency effects but not topography changes. Counterintuitively therefore, the spatial approach is better for studying temporal changes and vice versa.

Aside from the fundamental differences between PCA and ICA, there has also been a history of differences in the application of the techniques. These differences in application need to be addressed as well. At the risk of overgeneralizing, it can be said that PCA studies of ERPs have typically been temporal analyses on subject averages [e.g., Bentin et al., 1985; Chapman et al., 1978; Curry et al., 1983; Dien, 1999; Dien et al., 1997, 2003a; Friedman et al., 1981; Kayser et al., 1998; Kramer and Donchin, 1987; Lutzenberger et al., 1981; Polich, 1985; Rohrbaugh et al., 1978; Ruchkin et al., 1990; Yee et al., 1987], whereas ICA studies have typically been spatial analyses on either single subjects or multisubject grand averages [e.g., Jung et al., 2000; Makeig et al., 1997, 1999a,b; Vigario, 1997].

PCAs have historically been applied on a temporal basis, as ERP researchers typically characterize ERP components in terms of their time course, with scalp topography being a secondary, albeit important, characteristic. Subject averages have been analyzed because they facilitate subsequent analysis of variance (ANOVA) analyses of the factor scores.

The application of ICA to ERP datasets has been motivated, at least to some extent, by a basic difference between the two procedures: PCA is oriented toward “lumping,” while ICA is oriented towards “splitting.” Since the PCA algorithm begins by extracting linear combinations that account for as much variance as possible, the early factors it yields combine as many variables as possible. This results in a maximally parsimonious set of factors that will err toward conflating similar latent variables. ICA, on the other hand, starts the process with a different factor for every variable and, in seeking maximum independence, will tend to make fine distinctions. ICA factors therefore err in the direction of separating activity that should not be separated, sometimes splitting activity into multiple correlated factors that have primary loadings on different variables [Makeig et al., 1999a; McKeown et al., 1998], posing problems of parsimony. These multiple factors may, for example, reflect subtle individual differences. It can even result in background noise combining with a latent factor, splitting it into nearly identical versions at different variables (channels for a spatial analysis). Such splitting would complicate efforts to interpret the ICA results and to compare them with the PCA results.

This parsimony issue is largely avoided with sparse montages when using a spatial analysis. ICA articles that

have successfully applied a spatial approach to subject averages have only examined the data from 14 ERP channel locations [i.e., Matsumoto et al., 2005; Sato et al., 2001]. Such an analysis would yield 14 factors (one for each channel), which would avoid parsimony issues since this is approximately the number of major ERP features (i.e., not leaving any factors to represent subject differences). A study using 29 channels [Pritchard et al., 1999] reported having to analyze each subject and condition separately, because combined analyses resulted in factors corresponding to only a single condition. While the single-subject approach has been successfully used in a case study [Makeig et al., 1997] and in artifact correction [Jung et al., 2000; Vigario, 1997], trying to apply it to multiple subject datasets can be difficult [e.g., Jung et al., 2001]. Conducting a temporal analysis would similarly result in large numbers of factors (a 1-s epoch recorded at 250 Hz yields 250 time points and hence 250 factors) and high levels of splitting; in contrast, the “lumping” bias of PCA means that only a relative handful of these factors explain enough variance to be of interest, minimizing this concern.

One approach to countering this splitting issue is to use the multisubject grand average data to avoid individual difference factors and to reduce complications from background noise [Makeig et al., 1999a]. In principle, this approach could reduce the quality of the results since it loses information about individual difference variance that could be helpful for separating component activity. It also does not solve the fundamental issue of parsimony since a dataset will typically produce as many factors as there are variables, in the absence of collinearity. On the other hand, a multisubject grand average ERP could have the advantage of an improved signal-to-noise ratio with respect to subject averages.

A second approach is to use PCA as a preprocessing step to reduce the dimensionality of the dataset, an option apparently used in only one ERP report thus far [Johnson et al., 2001], but used in a number of fMRI analyses [Calhoun et al., 2001; Dodel et al., 2000; Greicius and Menon, 2004]. Reduction of data dimensionality has been advocated as a strategy for minimizing overfitting [Särelä and Vigário, 2003].

The current report will focus on using the PCA preprocessing approach since it also facilitates comparisons with PCA rotations. One can then conceptualize the contrast as between two different rotations, Promax and Infomax, of the same initial PCA decomposition. Issues about ICA component splitting and factor identification would be comparable to PCA. The multisubject grand average approach will also be evaluated in Simulation 4.

Two simulation comparisons (where it is possible to evaluate accuracy, since the true answer is known) have been made of PCA and ICA of ERP data, both recommending ICA (using the Infomax algorithm) over PCA [Makeig et al., 2000; Richards, 2004]. The present report will seek to extend these studies as follows. First, it will utilize real EEG for the background noise. Second, it will

explicitly examine the distinction between spatial and temporal approaches. Third, it will seek to parameterize the cases in which one or the other technique fails so that users have some basis for choosing which one to use for a given dataset.

This report will also address the Varimax issues noted earlier when there are no large clusters of variables that load on only one factor or if there are a number of variables with zero loadings on the first unrotated factor [Cureton and D'Agostino, 1983 : 224]. A variant of the Varimax rotation, the weighted-Varimax, has been proposed to address these two situations [Cureton and D'Agostino, 1983; Cureton and Mulaik, 1975]. It gives the most weight to factor loadings that are located away from the initial unrotated factor (which is by far the largest), essentially making the assumption that the initial rotation is not aligned with the correct rotation. It is not clear in advance how problematic these two situations might be for spatial and temporal PCAs of ERP data, so it seems worthwhile to evaluate this rotation as well. Since Promax uses Varimax as an initial rotation, we implemented Promax with Weighted-Varimax to supplement the regular Promax with Varimax rotation.

We will also examine the previously described Extended ICA algorithm, which is intended for sub-Gaussian distributions [Lee et al., 1999]. This variant will also be applied to see if it provides more effective results for temporal ICA.

This report consists of five simulations. Simulation 1 examines the reliability of ICA results. Simulation 2 evaluates Infomax and Promax under minimal noise conditions. Simulation 3 examines the effects of different levels of real background EEG noise. Simulation 4 examines the effects of individual differences and of using multisubject grand averages rather than subject averages. Simulation 5 determines if these results still apply when all five simulated components are included in the simulations.

SIMULATION I

Before direct comparisons can be made between ICA and PCA, an essential issue is determining whether ICA solutions are replicable, since there is a random element to the process (the random selection of data subsets) that can cause some variability in the results, as noted on the very helpful website of Makeig and colleagues (<http://www.sccn.ucsd.edu/~scott/tutorial/icafaq.html>).

Methods: Simulation I

A realistic simulation dataset was constructed for testing purposes, as previously described [Dien et al., 2005]. The simulation dataset represents a typical ERP dataset with 20 subjects, two conditions, and 65 channels (using the original montage of the Electrical Geodesics, Eugene, OR, net). Realistic background noise was obtained by using the data obtained from 20 subjects with EEG free from artifacts from a previously published experiment [Dien et al., 2003a]. Trials containing blinks were rejected, resulting in

an average of 55 trials per condition. The data was averaged using the \pm reference [Schimmel, 1967], which eliminates the ERP signal, but preserves the random background noise level, by inverting every other trial. This noise average was filtered using a 30-Hz low-pass filter. The data represents 125 time points, starting 184 ms before baseline, with a sampling rate of 125 Hz. The standard deviation (SD) of the noise ranged from 0.46 to 1.37 (median 1.04) microvolts across the epoch. Each channel of the data were referenced to the average of the data at a given time point, otherwise known as the average reference [Bertrand et al., 1985; Dien, 1998b].

Superimposed on the noise average were two simulated ERP components (Fig. 1). The topography of the ERP components was generated by the Dipole Simulator v. 2.1.0.5

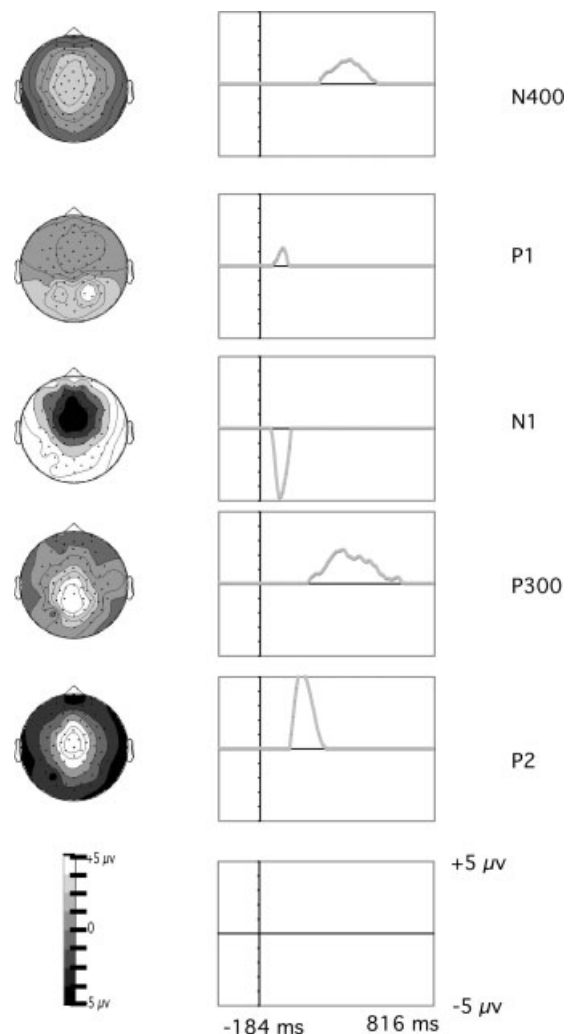


Figure 1.

Simulated ERP components. The scalp topographies represent the voltage map at the peak time point. The time courses represent the voltages at the peak channel.

(written by Patrick Berg and available for download from <http://www.megis.com/udbesa.htm>). One dipole was oriented roughly toward scalp location Cz of the International 10-20 System of electrode placement [Jasper, 1958], while the other dipole was oriented roughly toward Pz. The time course of the two components were generated using a half-sine wave covering 10 and 30 time points each. The peak latencies of the two components are 160 and 256 ms, respectively. The amplitudes of the two components were separately varied from 2–4 microvolts. Subject variance (correlation between the two component amplitudes) was simulated by setting the amplitude of Component 2 to be equal to the Component 1 amplitude plus 2–4 microvolts, divided by 2. The peak amplitude of Component 2 at the focal channel (the channel with the highest amplitude) in the small and large Component 1 cells had a mean (SD in parentheses) of 1.72 μV (0.34) and 1.71 μV (0.34), respectively. A condition effect was introduced by multiplying Component 1 by a factor of 0.9 for the small Component 1 cell and 1.1 for the large Component 1 cell. The peak amplitude of Component 1 at the focal channel had a mean of 2.2 μV (0.30) in the small Component 1 cell and 2.6 μV (0.39) in the large Component 1 cell. This level of effect was intended to yield F-values comparable to published P300 studies, since it has been shown that unrealistically large condition effects can exaggerate the degree of misallocation variance effects [Beauducel and Debener, 2003]. The two components are temporally overlapping and spatially correlated, both of which can be deleterious for PCA solutions [Dien, 1998a].

Aside from the correlation from the simulated subject variance, the substantial temporal overlap induces a correlation between the two ERP components for a spatial analysis since they tend to be present in the same observations [Dien, 1998a], resulting, in this case, in a Pearson's R of 0.10 when calculated across all the observations (the observations are not independent so an inferential test is not warranted). The correlation was calculated for each simulation, Fisher-Z-transformed, averaged, and then back-transformed. A correlation is a reasonable measure of similarity in the context of PCA because even analyses with a covariance matrix actually use correlations (factor loadings) during the rotation step. The choice between covariance and correlation relationship matrices affects the factor retention step, not the rotation step.

In order to evaluate reproducibility of the ICA results, the same dataset was analyzed 100 times using a spatial analysis. EEGLab 4.08 [Delorme and Makeig, 2004] running under Matlab 7.01 (MathWorks, Natick, MA) was used to compute the ICA solutions. Similarity of the analyses was assessed by examining the two factors correlating most highly with the two simulated components. When two factors correlated most highly with the same simulated component, the one that correlated most highly was paired with it.

Because PCA is biased toward combining latent variables together into a single factor, this procedure will tend

to favor ICA; such cases will essentially be tabulated as being an error because the second simulated component will be paired with an unsuitable factor instead of the combined factor. Conversely, ICA has a bias toward splitting components into multiple factors, and since only one of these multiple factors will be chosen, this factor will fit only part of the variance and will score poorly. Implicit in this procedure, therefore, is the conscious judgment that both such situations represent an error on the part of the statistical analysis.

Correlations were assessed with the factor loadings scaled in microvolts, which for ICA takes the form of the pseudoinverse of the product of the sphering matrix and the weight matrix. To examine the effect of using a PCA preprocessing step, the exercise was repeated with six retained factors (as suggested by the Scree test).

Results: Simulation I

While the results across the 100 analysis runs were quite similar, they were not identical. The correlation coefficients for the time course of Component 1, as regenerated from the factor scores, varied from 0.9858–0.9861, while the spatial distribution varied from 0.9979–0.9982. The time course of Component 2 correlated at 0.9993 in all cases, while the scalp distribution varied from 0.9890–0.9893. In general, the parameters were stable up to about the third digit. Examination of the actual factor loadings revealed a similar situation.

PCA preprocessing yielded moderately more stable parameters, with numbers also being generally stable up to the third digit. Component 1 time courses varied from 0.9913–0.9915, while the spatial topography varied from 0.9913–0.9917. The Component 2 time course ranged from 0.9982–0.9982 for temporal patterns and 0.9863–0.9865 for spatial patterns.

Discussion: Simulation I

Although the ICA solutions were largely reliable, variability is observable in the less significant digits. This variability can potentially have a noticeable impact. While the variability was not serious enough in the present study to be an issue, it is unknown whether it may be greater under some conditions such as when the signal-to-noise ratio of the data is lower.

One way to address this issue is to standardize the “random” number generation. Conventional computers cannot generate truly random numbers since their programming is wholly deterministic (<http://computer.howstuffworks.com/question697.htm>); instead, they use a complicated formula with a starting “seed” number to produce unpredictable numbers (the output using the initial seed is utilized as the seed to generate the next “random” number). It may be advisable to modify the pseudorandom number generator so that it uses a known seed that can be replicated as needed. Matlab's pseudorandom number genera-

tor is reset at startup, producing the identical output each time the program is started. The same result can be accomplished by inserting the command “rand(‘state’,0);” at the start of the “runica” code. This approach will be used for the remainder of this report.

SIMULATION 2

With the reliability issues addressed in Simulation 1, a preliminary comparison of ICA and PCA can now be conducted. A simulation dataset will be analyzed using both spatial and temporal approaches with both PCA and ICA. To increase the generalizability to real datasets, as described in the Methods section, five simulated components were constructed from five ERP components from real datasets. For this initial simulation, only minimal noncoherent background noise was added to maximize interpretability of the results (described in Methods). For this reason, the results of Simulation 2 will provide observations about the basic principles that will be largely free from concerns about confounds from the background noise but will not be generalizable to real datasets (maximizing control at the cost of ecological validity). Because it is not possible to systematically vary all aspects of such complicated datasets, Simulation 2 will be approached as representative of a larger universe of possible datasets wherein general principles may be observed for greater understanding of the relevant parameters.

We argue that making an effort to form a fully realistic ERP dataset and offering an evaluation on this basis would be an ill-posed question. Every ERP dataset will have different combinations of ERP components and it would not be possible to cover every eventuality. Instead, the goal of this article is to identify the parameters that result in successful separations of components, such as due to component correlations or the sigma covariance measure that we present later on. We therefore used pairs of components, reasoning that larger sets of simulated components could be difficult to interpret, much as trying to interpret a five-way ANOVA is extremely complicated due to all the possible patterns of interactions. Any time an experimental study of any kind, simulation or otherwise, is conducted one must choose a balance between interpretability and ecological validity; we argue that the current work chose a

balance that allowed us to best meet the goals of this study. We are, however, mindful of the concern that the dataset has a realistic level of dimensionality. We therefore include a high level of coherent background noise in Simulation 3 that provides this dimensionality, while addressing as best we can the potential for interactions between the simulated components and the background noise characteristics by comparing with Simulation 2 results (where there was no coherent background noise). Finally, in Simulation 5 we do include all five simulated components to determine if the prior results generalize to larger numbers of components.

It is not clear how weighted-Varimax will perform with temporal and spatial PCAs. The “pinning” situation (described in the Introduction) only applies to three or more factors, so it will not occur with the present simulation. The off-axis clusters situation (described in the Introduction) might be more likely to occur with spatial PCAs because more overlap (and hence variables with loadings on both factors) should occur in the spatial domain.

Extended ICA will be applied to determine if it provides any benefits. This variant will be most likely to improve temporal ICA since the data is mostly likely to be sub-Gaussian with this approach.

Methods: Simulation 2

Simulation 2 was constructed in the same fashion as Simulation 1 with two modifications. The first modification was that five real ERP components were utilized, as shown in Figure 1 and summarized in Table II. Two visual ERP components were obtained from a previously published study [Dien et al., 2003a]. The left frontal effect (focus near F3, peak at 432 ms), which we term an N400 because it is from an N400 study, although it has a more frontal distribution than usual, was obtained from the multisubject grand average difference wave between the congruent and incongruent ending conditions. A visual P1 was obtained from the same dataset from the congruent ending condition (focus near O2, peak at 120 ms). Three auditory ERP components were obtained from another previously published study [Dien et al., 1997]. An auditory N1 was obtained using the multisubject grand average for all audi-

TABLE II. Statistical properties of the five simulated ERP components

	Peak	Temporal SD	Temporal skew	Temporal kurtosis	Spatial SD	Spatial skew	Spatial kurtosis
N400	432 ms	0.95	-0.04	-0.93	0.63	1.64	1.15
P1	120 ms	0.75	0.77	-0.64	0.30	4.43	19.00
N1	108 ms	2.61	-0.17	-1.21	0.98	-3.64	12.10
P300	400 ms	1.06	0.50	-0.80	0.76	1.13	-0.28
P2	200 ms	2.16	0.84	-0.12	1.32	2.70	5.89

SD, standard deviation, represents the microvolt values at the peak channel.

The parameters are calculated from the base components without the addition of condition or subject variance. Temporal columns represent the figures for the temporal approach and the spatial columns are for the spatial approach.

TABLE III. Relations between pairwise comparisons of simulated components

	C1	C2	Temporal scores	Spatial scores	Temporal loadings	Spatial loadings	Temporal sig cov	Spatial sig cov
1	N400	P1	0.43	0.12	0.13	0.43	1.20	0.17
2	N400	N1	0.81	0.14	0.15	0.82	1.07	0.14
3	N400	P300	0.63	0.91	0.93	0.63	1.32	1.44
4	N400	P2	0.89	0.18	0.19	0.90	1.03	0.23
5	P1	N1	0.75	0.92	0.94	0.76	1.04	2.86
6	P1	P300	0.25	0.16	0.17	0.25	1.61	0.32
7	P1	P2	0.34	0.09	0.09	0.34	1.07	0.11
8	N1	P300	0.22	0.19	0.20	0.22	1.33	0.21
9	N1	P2	0.74	0.10	0.10	0.75	1.37	0.09
10	P300	P2	0.73	0.18	0.19	0.75	1.19	0.33

“Sig Cov” is the sigma covariance measures as described in the text.

C1 and C2 are the two simulated components in the dataset. Temporal scores are the correlation of the true factor scores for the temporal approach (if the factors are reconstructed accurately), including subject, cell, and spatial variance. Spatial scores are the equivalent figure for the spatial approach. The scores are the median results across the ten replicates. Temporal loadings are the correlation between the time course (scaled loadings) of the two components that quantifies the variable overlap for the temporal approach. Spatial loadings are the complementary figures for the spatial approach.

tory conditions (focus near Fz, peak at 108 ms). An auditory P300 was obtained from the difference wave between target and standard conditions (focus between Cz and Pz, peak at 400 ms). Finally, an auditory P2 was obtained from the multisubject grand average of all auditory conditions (focus near Fz, peak at 200 ms).

To ensure that these waveforms are statistically unidimensional in both the spatial and temporal dimensions (without additional dimensions contributed by noise or overlapping components), the components were constructed from the time course at the focus channel matrix (with the periods before and after the component set to zero) multiplied by the maximum normalized scalp distribution at the peak time point (with a reversed N1 topography to ensure correct polarity of reconstructed component). This procedure generated a component with the same time course at all channels and the same scalp distribution at all time points, as should be the case for a single component due to volume conduction of a single source electric field. This procedure was also necessary since it is known that some of these ERP features are in fact composed of multiple components, an issue that is outside the scope of this report [see Näätänen and Picton, 1987; Sutton and Ruchkin, 1984]. This procedure reduced these simulated ERP features from their true unknown dimensionality to a known single dimension.

The second modification is that for this initial simulation the background EEG noise was removed. To prevent the data matrices from becoming singular, a very small amount of noise was added to each data point (-0.01 to $+0.01$ microvolts). The noise had no coherence between data points. The same set of random noise was used for every simulation to keep it constant with regard to the manipulations of interest.

One hundred different datasets were generated and analyzed in each of the four approaches (spatial and temporal, each with PCA and ICA). Test simulations with the 10 dif-

ferent pairwise combinations of the five simulated components were generated. For each of these 10 combinations, 10 simulation datasets were generated for a total of 100 simulation datasets. The component amplitudes were varied as described for Simulation 1 with random variation for each simulated subject average, subject variance, and a condition effect for one of the two components. Table III presents the similarity of the pairs of components in terms of both the factor scores and in terms of the variables.

The PCA Toolbox 1.091 (<http://www.people.ku.edu/~jdien/downloads.html>) was used to compute the PCA solutions. As we have recommended elsewhere on the basis of simulation studies [Dien et al., 2005], the PCAs were carried out using covariance matrices, Kaiser normalization, Promax rotation, and correlation loadings. Promax was conducted with a kappa of 3, which is the parameter that determines how oblique the rotations will be. To convert the factor loadings into microvolt metric for comparison with the original data, the factor pattern matrix was multiplied by the SDs of the variables [Dien et al., 1997]. The ICA was conducted using PCA preprocessing. Only two factors were retained since there is no coherent noise to be accounted for by the PCA solution. The extended-ICA was conducted with the specification that both of the factors would be sub-Gaussian (e.g., “‘extended,’ -2”).

Although we have some reservations about applying inferential statistics to artificial simulation data results, we applied paired t -tests to selected comparisons of interest to further evaluate the results. Note that t -tests compare means, whereas the tables present median statistics. We chose median statistics for the tables because of a judgment that overall consistency is more important than highlighting the effect of dramatic outliers. However, we chose conventional t -tests, which compare means, since they will be the most familiar for readers. For the most part, the resulting t -tests do seem to correspond with the median statistics.

An effort was made to find potential predictors of ICA performance, based on some preliminary examinations of the results. In the Infomax algorithm, independence is operationalized as the product of a factor score with the sigma transformed version of the other factor score. Inspection of the sigma scatterplots (as presented in Figs. 4 and 5) suggested that a useful measure might be the covariance between each pair of independence measures ($u_1 \cdot y_2$ and $u_2 \cdot y_1$) to produce a sigma covariance measure, where u = the factor score and y is the sigma-transformed factor score. Also, the absolute sigma values were generated and the mean of both sets of sigmas were calculated to produce a mean sigma measure. Finally, since the actual rotations are a function of the ratio of the self-sigmas (on the diagonal of the relationships matrix) and the sigmas (on the off-diagonals), the absolute values of all four sigma measures were generated, then the ratio of the off-diagonal and its paired self-sigma were calculated (which jointly determine the rotation of a given factor), and finally, the two ratios were added together to produce a relative sigma measure. These three measures were calculated for each of the 100 simulations and the correlations between these potential diagnostics and the accuracy measures were calculated. The sigma measures were calculated on a prerotation basis (using the known correct results) and on a postrotation basis (using the ICA results). The premeasures assess the utility of the ICA factors given known components and the postmeasures assess their utility when the correct answer is not known. An effort to develop an analogous measure for Promax rotations was not successful.

Results and Discussion: Simulation 2

As can be seen in the median scores of Tables IV and V, spatial ICA provided the most accurate results, followed

closely by temporal PCA. All four analysis types were highly effective for some simulations and highly ineffective for others. Table VII presents the potential rotation diagnostics. Overall, the sigma covariance measure yielded the strongest results.

Comparing PCA using Varimax and Weighted-Varimax presteps, the results were mixed. For temporal analyses, time correlations were borderline significantly better for the conventional Varimax: $t(99) = 1.9, P = 0.055$. However, space correlations were better for weighted Varimax for both temporal PCA ($t(99) = 3.6, P = 0.0005$) and spatial PCA ($t(99) = 5.9, P < 0.0001$). Time correlations for spatial PCA were not significantly different.

Comparing ICA and Extended ICA, the results were much more consistent. For temporal analyses, conventional ICA yielded higher time correlations ($t(99) = 6.0, P < 0.0001$) and space correlations ($t(99) = 8.3, P < 0.0001$). For spatial analyses, conventional ICA also yielded higher time correlations ($t(99) = 52.0, P < 0.0001$) and space correlations ($t(99) = 14.3, P < 0.0001$).

Finally, comparing conventional PCA and conventional ICA, for temporal analyses PCA yielded higher time correlations ($t(99) = 3.2, P = 0.0020$) and space correlations ($t(99) = 2.8, P = 0.0059$). For spatial analyses, conventional ICA also yielded higher time correlations ($t(99) = 6.8, P < 0.0001$) and space correlations ($t(99) = 7.7, P < 0.0001$). Although spatial ICA appeared to be the most effective overall in this simulation, it is important to keep in mind that this simulation did not use realistic background EEG noise and was therefore not representative of real datasets. The goal of the present simulation was to observe whether the analyses perform as expected to variations in the simulation parameters as well as to identify unexpected issues.

TABLE IV. Results of Simulation 2 comparing principal components analysis (PCA) under temporal and spatial approaches

	C1	C2	TPCA time	tPCA space	sPCA time	sPCA space	tPCAw time	tPCAw space	sPCAw time	sPCAw space
1	N400	P1	1.00	1.00	1.00	0.99	0.87	0.96	0.91	0.93
2	N400	N1	0.67	0.69	0.99	0.99	1.00	0.98	0.93	0.99
3	N400	P300	0.97	0.64	0.82	0.95	0.98	0.97	0.93	0.97
4	N400	P2	0.35	0.29	0.84	0.98	0.89	0.99	0.76	0.97
5	P1	N1	0.85	0.99	0.36	0.51	0.37	0.85	0.91	0.88
6	P1	P300	1.00	1.00	0.90	0.59	0.52	0.88	0.69	0.92
7	P1	P2	1.00	1.00	0.99	0.99	0.54	0.90	0.77	0.98
8	N1	P300	1.00	1.00	0.72	0.73	0.99	0.99	0.90	0.87
9	N1	P2	0.96	0.97	0.93	0.93	1.00	1.00	0.99	0.99
10	P300	P2	0.96	0.99	0.78	0.91	0.96	0.99	0.82	0.93
	Median	Totals	0.97	0.99	0.87	0.94	0.93	0.98	0.91	0.95

C1 and C2 are the two simulated components in the dataset. tPCA and tPCAw are the results for the temporal approaches and sPCA and sPCAw are the results for the spatial approaches. PCA means Promax with the regular Varimax prestep, whereas PCAw means Promax with a weighted-Varimax prestep. The “time” columns are the accuracy of the reconstructions of the factor time courses, expressed as the correlation between the scaled factor results and the matching original component. For each analysis the accuracy was calculated for both simulated components and the lowest accuracy of the two factors was recorded. The “space” columns are the accuracy of the reconstructions of the factor spatial topographies. The bottom row is the median score down each column of the table.

TABLE V. Results of Simulation 2 comparing independent components analysis (ICA) under temporal and spatial approaches

	C1	C2	tICA time	tICA space	sICA time	sICA space	tICAe time	tICAe space	sICAe time	sICAe space
1	N400	P1	0.95	0.95	1.00	0.99	0.52	0.58	0.62	0.12
2	N400	N1	0.94	0.70	1.00	1.00	0.65	0.65	0.68	0.54
3	N400	P300	0.40	0.47	0.65	0.91	0.09	0.45	0.45	0.93
4	N400	P2	0.61	0.80	1.00	1.00	0.35	0.28	0.62	0.68
5	P1	N1	0.98	0.84	0.71	0.86	0.91	0.21	0.38	0.99
6	P1	P300	0.40	0.82	0.99	0.95	0.97	0.94	0.67	0.17
7	P1	P2	0.98	0.93	1.00	1.00	0.39	0.74	0.65	0.08
8	N1	P300	0.90	0.90	0.99	1.00	0.85	0.89	0.59	0.65
9	N1	P2	0.82	0.65	1.00	1.00	0.81	0.64	0.68	0.28
10	P300	P2	0.88	0.85	0.99	1.00	0.45	0.45	0.60	0.22
	Median	Totals	0.89	0.83	1.00	1.00	0.58	0.61	0.62	0.41

C1 and C2 are the two simulated components in the dataset. tICA is the results for the temporal approach and sICA is the results for the spatial approach. tICAe and sICAe are for the extended-ICA results. The “time” columns are the accuracy of the reconstructions of the factor time courses, expressed as the correlation between the scaled factor results and the matching original component. For each analysis the accuracy was calculated for both simulated components and the lowest accuracy of the two factors was recorded. The “space” columns are the accuracy of the reconstructions of the factor spatial topographies. The bottom row is the median score down each column of the table.

Figure 2a presents an example simulation, one of the 10 replicates of Case 1 (“Case” being one of the simulated component pairings listed in Table III), where the temporal PCA was highly effective. The first column is a scatterplot of the factor loadings (factor pattern matrix). Each circle represents a single factor loading with the coordinates of the first factor corresponding to the horizontal axis and the coordinates of the second factor corresponding to the vertical axis. Time points not loading on either factor appear in the middle since their factor loadings were essentially zero on both factors. Note how the nonzero factor loadings appear either along the horizontal axis or along the vertical axis, representing time points that load only for one or the other factor. The lines show where the axes should be located if the factor solution is accurate, based on time points that should be zero on one factor and nonzero on the other factor given a successful factor solution. In this case the lines do in fact fall along the two axes of the factor solution, reflecting the success of the solution.

Note that the scatterplot describes the final loadings that do not entirely correspond to the rotation space. The PCA rotations were conducted using Kaiser normalization, which constrains all the pairs of squared loadings to sum to one, meaning that all the dots actually were placed at the circumference of a round circle. The inactive time points with close to zero loadings were located at essentially random locations on the circle, and hence overall cancelled each other out. For didactic reasons, we find it more helpful to refer to the final loadings as if they were the actual rotation space, which in practical terms is largely equivalent.

The second column presents the unscaled loadings of the two factor waveforms. Note how the waveforms appear box-like, since the corresponding factor loadings are either 0 or 1. If the levels of background noise were

higher, then the factor loadings would look less box-like, as they would represent the ratio between the signal and the background noise. The third column presents the loadings of the two factor waveforms after they have been translated into microvolt scaling by multiplying them by the SDs of the respective time point variables [Dien et al., 1997]. Comparison of the results in Figure 2a with the original time courses in Figure 1 reveals them to be accurate reconstructions. In this case, it seems likely that the results were especially accurate, because these two particular simulated components suffer minimal overlap in both the spatial and temporal dimensions. As we will see, other possible pairings of simulated components can yield less accurate results.

As expected, the first two issues of factor overlap and factor correlation interacted with whether temporal or spatial PCA was utilized. With regard to temporal PCA, factor correlations due to spatial overlap did indeed cause distortions in the factor results, but only to a limited extent due to the use of the Promax rotation rather than the more common Varimax rotation [see Dien, 1998a; Dien et al., 2003b, 2005]. Even the simulations where the factor correlations were as high as about 0.75 showed only moderate distortions. Only the two highest factor correlations (0.81 and 0.89) showed substantial breakdown of the PCA results. The results from the worst case (#4) are presented in Figure 2b. As can be seen, the nonzero factor loadings are clustered together to an extent that the PCA was unable to resolve. The lines show that an accurate solution would require a much more oblique rotation than was provided by the Promax rotation, at least at a kappa of 3. The resulting factor solutions show a classic contrast pattern in which one factor (the gray one) is maximal on all the variables involved while the other factor represents the extent to which the two latent variables are different, being posi-

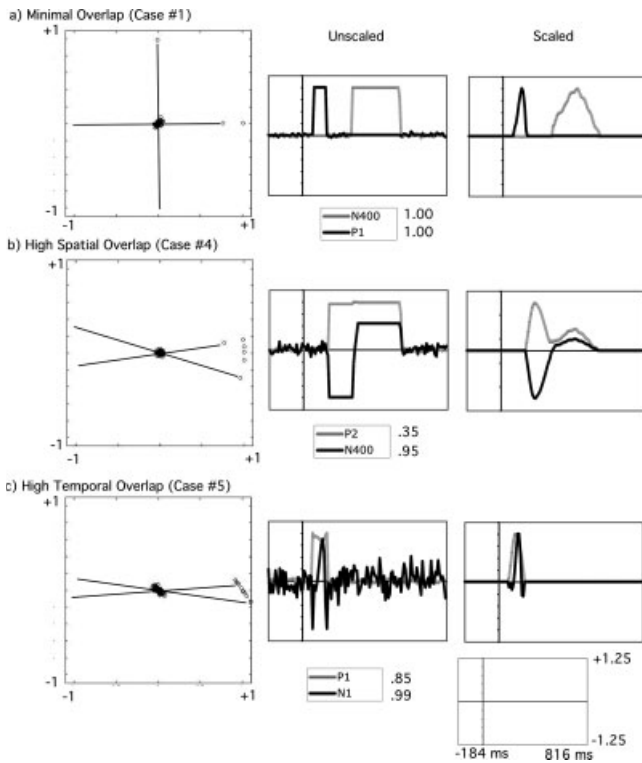


Figure 2.

Simulation 2 temporal PCA results. The scatterplots display the rotated factor loadings. The lines represent the axes of the correct solution. The graphs present the microvolt-scaled factor waveforms normalized to the maximum value (to facilitate comparison of the two factors). Sign of the factor loadings is arbitrary. The first column is for the unscaled factor loadings. The second column is for the microvolt-scaled factor loadings. The numbers next to the legends indicate the accuracy of the scalp topography reconstruction (absolute correlation with the simulated component); the topographies are not shown.

tive for one set and negative for the other. Factor solutions with this kind of pattern are clear indications that the rotation was unsuccessful.

Conversely, temporal overlap had less effect than expected. For example, the overlap between the P1 and the N1 components was nearly total, with the peaks being 108 and 120 ms, respectively. The N1 fully covered all the points corresponding to the P1 but had a time point just before and after the P1 that were without overlap. Nonetheless, the time course reconstruction seen in Figure 2c was quite healthy, with a 0.85 and the spatial reconstruction was nearly perfect with a 0.99. On the other hand, the second highest temporal correlation (between the N400 and the P300) did result in some serious spatial distortion (not shown in figure).

In any case, of the 10 pairwise comparisons, four resulted in notable distortions (less than 0.9 correlation in either time or space). Two corresponded to the highest

temporal factor correlations 2 and 4) and two corresponded to the highest temporal variable correlations (3 and 5). The results are therefore in line with expectations and could be predicted in advance based on the data in Table III, with the results of the simulation providing some guidance as to the point at which distortions can be expected. Simulation 3 was included to determine if the addition of substantial levels of coherent noise changes the threshold points.

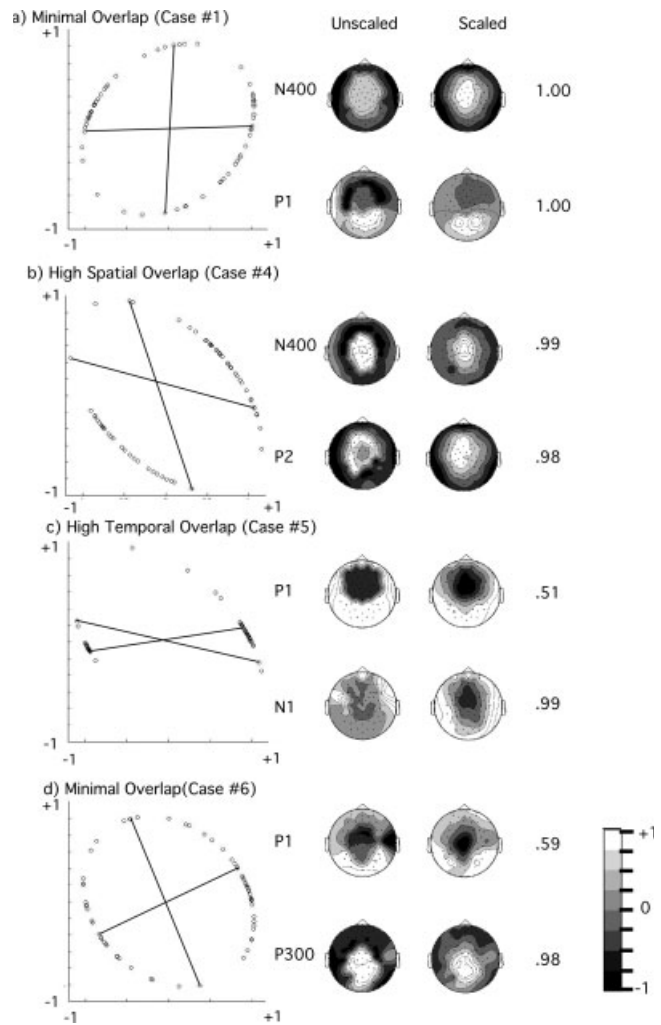


Figure 3.

Simulation 2 spatial PCA results. The scatterplots display the rotated factor loadings. The lines represent the axes of the correct solution. The topography plots indicate the microvolt-scaled factor topographies normalized to the maximum value (to facilitate comparison of the two factors). Sign of the factor loadings is arbitrary. The first column is for the unscaled correlation maps. The second column is for the microvolt-scaled maps. The numbers on the far right side indicate the accuracy of the scalp topography reconstruction (absolute correlation with the simulated component).

Turning to the spatial PCAs, again the expectation is that high factor score correlations and high variable overlap will cause distortions. Figure 3a presents a solution where the spatial PCA was highly successful (Case 1). The first column presents the scatterplot of the factor loadings. Note how the factor rotation has arranged itself so that a maximal number of points fall along the horizontal and vertical axes. This is the physical manifestation of the Varimax criterion, which is further enhanced by the Promax relaxation step. Note also how the increased amount of spatial overlap has manifested as an increased number of points along the intermediate points between the axes, compared to the comparable plots in Figure 2. The second column shows the scalp topography of the unscaled factor loadings and the third column presents the scalp topography of the scaled factor loadings. Keep in mind that the sign of the factor loadings are arbitrary and in this case the factor loadings for the P300 and P1 locations are negative; only the product of the factor loadings and the factor scores correspond to the sign of the original data.

Figure 3b presents the results of the highest factor score correlation (#4) due to high spatial correlation. The lines give a sense of the degree of distortion, which is nonetheless quite moderate (0.98 in the spatial dimension and 0.84 in the time dimension) compared to the results of the temporal PCA shown in Figure 2b. It can therefore be seen how the presence of a high spatial correlation has a much more adverse effect on the temporal PCA than on the spatial PCA. The case with the second-highest spatial correlation (Case 2) suffered only minimal effects (0.99 for space and 0.94 for time).

Figure 3c presents a case of high temporal correlation (Case 5) corresponding to the temporal PCA presented in Figure 2c. Whereas the high temporal correlation had only modest effects on the temporal PCA, the effects were quite severe for the spatial PCA. Both of the cases of high temporal correlation (Case 5, shown, and Case 3, not shown) suffered distortion in the spatial PCA.

In contrast to the temporal PCA results, a number of spatial PCA cases with neither high spatial nor temporal correlation nonetheless were distorted, which appear to reflect a fifth issue of too many variables loading on both factors. One such case is #6, where the factor correlations were only 0.16 and the spatial correlation was only 0.25. Examination of Figure 3d shows that the factor loadings have adopted the classic contrast pattern in which the upper factor represents the electrodes common to both the P1 and the P300, whereas the lower factor is the extent to which the two differ (with positive loadings for the P1 electrodes and negative for the P300 electrodes).

Examination of Figure 3d suggests that the cause of the problem is that most of the variables loaded on both factors. The lines representing the correct solution indicate that in fact the bulk of the points should fall between the two axes. The relaxation step of the Promax rotation cannot provide an improvement in this case, since the factors are nearly uncorrelated with each other (0.16). This prob-

lem is a direct result of the high levels of overlap present in the spatial domain due to volume conduction. There are very few electrodes where one factor is zero and the other factor is nonzero. Volume conduction ensures that all the electrodes will reflect the influence of both components. Thus, the only electrodes that will register zero amplitude for a given component are the ones where the reference choice has designated its level of voltage as corresponding

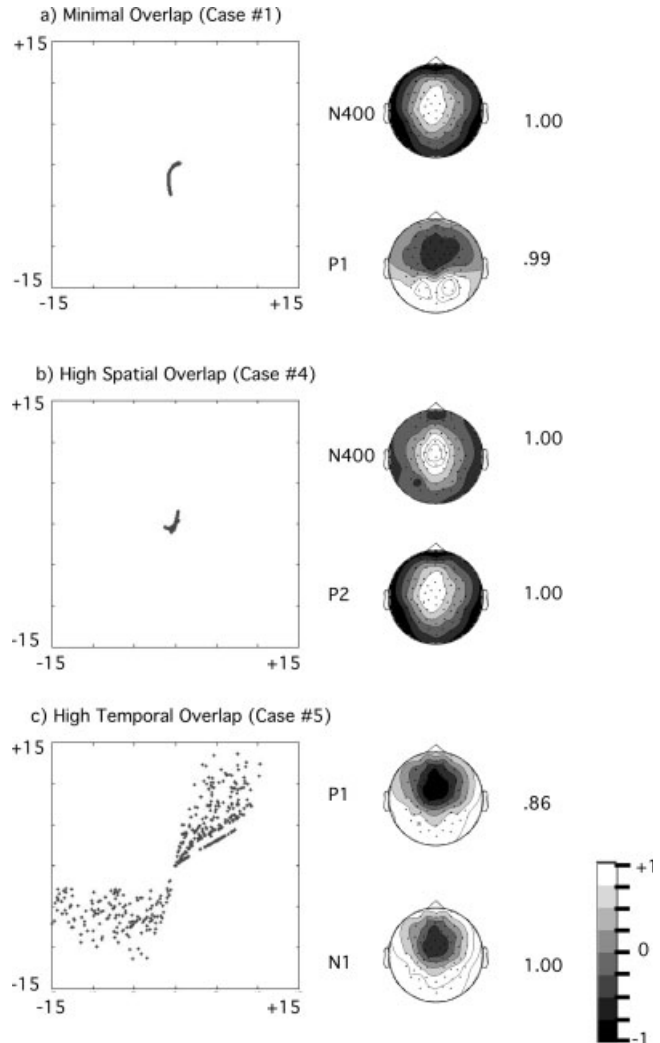


Figure 4.

Simulation 2 spatial ICA results. The scatterplots display the joint sigma values for the two factors using the known true solution. The topography plots indicate the microvolt-scaled factor topographies normalized to the maximum value (to facilitate comparison of the two factors). Sign of the factor loadings is arbitrary. The numbers beside the legends indicate the accuracy of the scalp topography reconstruction (absolute correlation with the simulated component). The original topography of the simulated factors is presented in Figure 1.

to zero voltage, with nearby electrodes being close to zero. For an extended discussion of reference site issues, see Dien [1998b].

As designed, the weighted-Varimax provided some improvement for this example (Case 6) and the others with similar characteristics, but yielded inferior results for still others. Overall, the weighted-Varimax produced better results for spatial PCA, consistent with our discussion of Figure 3d, but worse for temporal PCA. Even for spatial PCA the weighted-Varimax rotation improved half of the cases but degraded the other half. It appears that, in general, weighted-Varimax may have some utility for spatial PCA but not for temporal PCA. Overall, there did not appear to be compelling reasons to switch to weighted-Varimax, especially given the stronger performance of spatial ICA.

This simulation did not address the other two issues. There were no issues of misretention (described in the Introduction) since there were only two sources of coherent variance in the datasets, the two simulated components. There were likewise no pinning issues, since this only occurs when there are at least three factors being rotated.

Next, Table V presents the results of the ICA for both spatial and temporal approaches. For the temporal approach, ICA yielded clearly inferior results compared to both types of PCA. In contrast, the results for the spatial approach were quite impressive, yielding the best results of any of the procedures thus far evaluated (Figure 4).

There are two possible reasons for why the spatial ICA yielded better results than the temporal ICA. The first possibility is that, as seen in Table I, overall the components fit the super-Gaussian distribution (with kurtoses over 1, except for P300), whereas the components in the temporal arrangement are all sub-Gaussian. The two spatial ICA cases that yielded weak results (Cases 3 and 5) had component time courses that were highly similar, as seen in the Spatial Scores column of Table III (both over 0.9, whereas the others are below 0.2). Some evidence against this interpretation is provided by extended ICA, which is designed to work with sub-Gaussian distributions. As seen

in Table VI, this algorithm did not change the pattern of temporal analyses, overall performing more poorly than spatial analyses (and overall yielding inferior results).

The other possibility is that the spatial scores are simply more easily separated than the temporal scores (in other words, these results are a property of the relationship between the two scores rather than of the Gaussianity of the scores taken individually). Evidence for this supposition is found in the sigma covariance measure, which is overall higher for the temporal scores. Furthermore, the sigma covariance measures are especially high for the two spatial cases, which did not separate successfully.

As for the temporal ICA cases, while the results are generally less successful than for the spatial ICA approach, there are still successful solutions. Examples of both successful and unsuccessful solutions are displayed in Figure 5. There was a nearly significant correlation between the pre-analysis diagnostic measures and the accuracy of the factor loadings for the 10 temporal ICA cases ($r[8] = -0.60$, $P = 0.066$) as seen in Table VI in the row marked “covariance,” which is impressive given the low n . (It may not be appropriate to use the full 100 cases as the n since they consist of 10 replicates of the 10 cases. Of course, whether to use all 100 cases as the sample size depends on what larger population is to be generalized to; since this is an artificial dataset, it is unclear what statistical significance means, let alone which is the appropriate sample. Nonetheless, these tests do provide some sense of how to interpret the results).

SIMULATION 3

While Simulation 2 provided results that provided some insights into the boundary conditions for PCA and ICA, and how they are affected by the differences between the temporal and the spatial approaches, the comparisons cannot be generalized to real datasets until realistic noise levels have been added to the simulations. In the following simulations the background EEG noise used in Simulation 1 are added to the simulated ERP components.

TABLE VI. Independent components analysis (ICA) rotation diagnostics for Simulation 2

	tICA time Pre	tICA space Pre	sICA time Pre	sICA space Pre	tICA time Post	tICA space Post	sICA time Post	sICA space Post
Covariance	0.60	0.22	0.88	0.96	0.52	0.58	0.65	0.83
Mean	0.34	0.14	0.67	0.49	0.65	0.65	0.12	0.22
Relative	0.06	0.47	0.26	0.16	0.09	0.45	0.59	0.79

tICA is the results for the temporal approach and sICA is the results for the spatial approach. The numbers are the correlations between the putative ICA diagnostic measure and the accuracy measures. The “time” columns are for the reconstructions of the factor time courses and the “space” columns are for the reconstructions of the factor spatial topographies. The nature of the three putative diagnostic measures are described in the Methods section. The accuracy measures were calculated as the correlation between the scaled factor results and the matching original component. For each analysis the accuracy was calculated for both simulated components and the lowest accuracy of the two factors was used. The “pre” measures were based on the correct solution and the “post” measures were based on the actual ICA results.

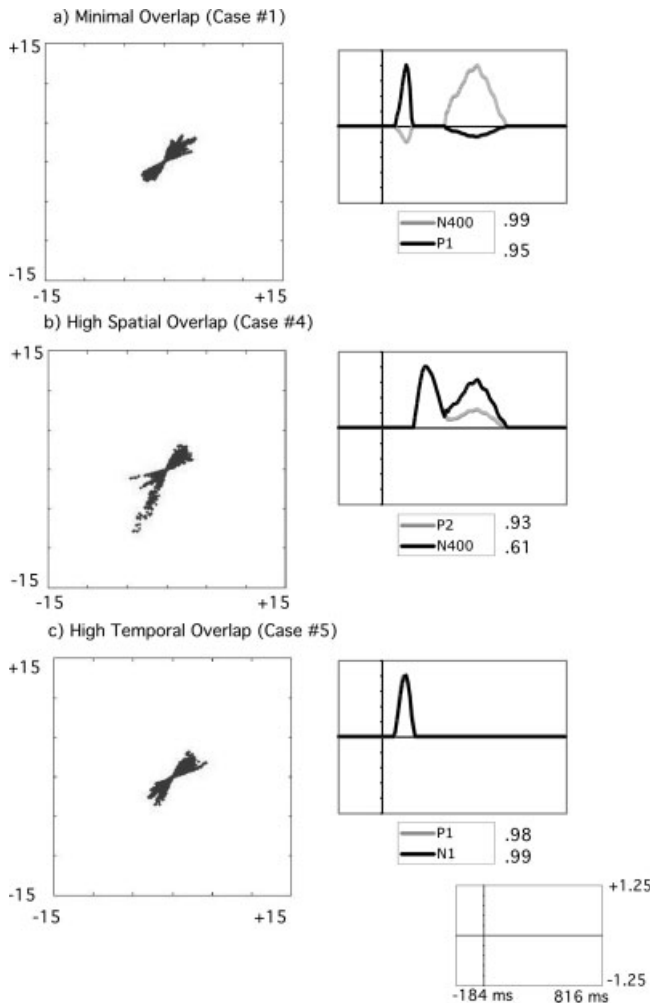


Figure 5.

Simulation 2 temporal ICA results. The scatterplots display the joint sigma values for the two factors using the known true solution. The waveforms indicate the microvolt-scaled factor waveforms normalized to the maximum value (to facilitate comparison of the two factors). Sign of the factor loadings is arbitrary. The numbers beside the legends indicate the accuracy of the time course reconstruction (absolute correlation with the simulated component). The original time course of the simulated factors are presented in Figure 1.

Methods: Simulation 3

Simulation 3 was constructed in the same fashion as Simulation 2, with the one modification of using the real background EEG noise from Simulation 1. This noise corresponds to roughly the noise from 55 averaged trials. A low-noise version was also constructed with the noise amplitudes reduced by half and a high noise version was constructed with double noise amplitudes. Insofar as the relationship between noise levels and numbers of trials is a square root function [Regan, 1989: 47], this corresponds

to 110 and 28 trials, respectively. Scree plots [Cattell, 1966; Cattell and Jaspers, 1967] suggested that retaining seven factors was appropriate for the temporal analyses and six factors was appropriate for the spatial analyses. The Scree test was chosen as it is a well-established procedure for estimating dimensionality and because it should not be biased toward either rotation. Other methods of estimating dimensionality do exist [e.g., Hansen et al., 2001].

Results and Discussion: Simulation 3

As can be seen in the median scores of Tables VII–IX, noise levels had a marked effect on accuracy. For low noise, the time correlations were not different for either temporal or spatial analyses. For space correlations, the temporal PCA was higher than the temporal ICA: $t(99) = 2.9$, $P = 0.0045$. Conversely, the spatial ICA was higher than the spatial PCA: $t(99) = 6.8$, $P < 0.0001$. For medium noise, the only significant difference was higher spatial correlations for spatial ICA than for spatial PCA: $t(99) = 5.4$, $P < 0.0001$. For high noise, there were no significant differences. Overall, the conclusions of Simulation 2 hold fairly well: PCA is most effective for temporal analyses and ICA is most effective for spatial analyses at all three noise levels, although the differences became statistically insignificant for higher noise levels.

SIMULATION 4

Simulation 4 seeks to further improve the realism of the simulation analyses by adding a further complication. In the simulations thus far the time course and the spatial topography of the simulated components are the same for all the simulated subjects. In real data there is considerable variability, which could affect the relative performances of PCA and ICA. The present simulation therefore adds such individual variability. This addition also makes it possible to evaluate the alternative approaches of using the subject averages (favored by PCA studies) and the grand averages (favored by ICA studies).

Methods: Simulation 4

Simulation 4 was constructed in the same manner as Simulation 3's medium noise condition. The main difference was that instead of using the same scalp topography and time course for each simulated ERP component (derived from the grand average) for each simulated subject, an individual scalp topography and time course was used for each simulated subject. These individual topographies and time courses were derived from the real individual subject averages, thus adding a realistic level of individual variation into the simulated dataset. Although this means this simulation is a step closer to using real ERPs, this dataset is still simulated in that the dimensionality of the ERP features was reduced to a single dimension, and so the correct answer is still known. The time windows

TABLE VII. Results of Simulation 3 ICA and PCA under low noise conditions

	C1	C2	tPCA time	tPCA space	sPCA time	sPCA space	tICA time	tICA space	sICA time	sICA space
1	N400	P1	1.00	1.00	0.99	0.90	0.94	0.95	1.00	1.00
2	N400	N1	0.78	0.90	0.95	0.77	0.93	0.55	1.00	1.00
3	N400	P300	0.53	0.72	0.97	0.95	0.96	0.77	0.66	0.91
4	N400	P2	0.41	0.43	0.64	0.93	0.62	0.81	0.99	1.00
5	P1	N1	0.88	0.78	0.86	0.69	0.97	0.88	0.24	0.52
6	P1	P300	1.00	1.00	0.92	0.73	0.51	0.83	0.99	0.97
7	P1	P2	1.00	1.00	0.99	0.99	0.95	0.92	1.00	1.00
8	N1	P300	1.00	1.00	0.72	0.78	0.94	0.92	0.99	1.00
9	N1	P2	0.94	0.93	0.90	0.89	0.84	0.69	1.00	1.00
10	P300	P2	0.90	0.97	0.80	0.93	0.86	0.87	0.99	1.00
	Median	Totals	0.92	0.95	0.91	0.89	0.93	0.85	0.99	1.00

PCA, principal components analysis; ICA, independent components analysis. C1 and C2 are the two simulated components in the dataset. tPCA is the results for the temporal PCA approach and sPCA is the results for the spatial approach. tICA and sICA are for the ICA results. The “time” columns are the accuracy of the reconstructions of the factor time courses, expressed as the correlation between the scaled factor results and the matching original component. For each analysis the accuracy was calculated for both simulated components and the lowest accuracy of the two factors was recorded. The “space” columns are the accuracy of the reconstructions of the factor spatial topographies. The bottom row is the median score down each column of the table.

were all the same, corresponding to those used in the prior simulations (such that if the subject averages were averaged, it would result in the grand average waveform). Likewise, the time point used to form the scalp topographies were the same for all the subjects, corresponding to those used for the prior simulations. Because the auditory components were obtained from a dataset with only 16 subjects, it was necessary to have only 16 simulated subjects in this simulation. Scree charts suggested that seven factors be retained for the temporal analyses and six factors be retained for the spatial analyses.

Although in principle it would be desirable to also include temporal jitter, this step was not taken in the interests of interpretability. Once temporal jitter was

added, it would be difficult to evaluate the accuracy of the analyses. For example, in the presence of temporal jitter (where each waveform has a different time course), what is the definition of an accurate solution? Would accuracy be having a different factor corresponding to each of the different subject averages? Would accuracy be having a single factor corresponding to the central tendency? The former would favor Infomax and the latter would favor Promax. And how would one define central tendency? We therefore elected to keep the time course constant to provide a basis for making an unambiguous evaluation of analysis accuracy. Furthermore, we suggest that the issue of subject variability is already addressed by the observations made for the spatial domain and it is

TABLE VIII. Results of Simulation 3 ICA and PCA under medium noise conditions

	C1	C2	tPCA time	tPCA space	sPCA time	sPCA space	tICA time	tICA space	sICA time	sICA space
1	N400	P1	1.00	1.00	0.63	0.25	0.87	0.94	0.99	0.99
2	N400	N1	0.84	0.95	0.92	0.59	0.91	0.76	0.99	0.97
3	N400	P300	0.44	0.62	0.99	0.89	0.96	0.83	0.48	0.36
4	N400	P2	0.37	0.43	0.46	0.70	0.65	0.81	0.99	0.98
5	P1	N1	0.84	0.74	0.94	0.58	0.95	0.91	0.16	0.55
6	P1	P300	1.00	1.00	0.90	0.82	0.49	0.84	0.99	0.96
7	P1	P2	0.99	1.00	0.92	0.96	0.93	0.92	0.99	0.95
8	N1	P300	1.00	1.00	0.75	0.81	0.95	0.92	0.99	1.00
9	N1	P2	0.96	0.96	0.82	0.67	0.85	0.72	1.00	1.00
10	P300	P2	0.93	0.99	0.62	0.85	0.87	0.88	0.99	0.99
	Median	Totals	0.95	0.97	0.86	0.75	0.89	0.86	0.99	0.98

PCA, principal components analysis; ICA, independent components analysis. C1 and C2 are the two simulated components in the dataset. tPCA is the results for the temporal PCA approach and sPCA is the results for the spatial approach. tICA and sICA are for the ICA results. The “time” columns are the accuracy of the reconstructions of the factor time courses, expressed as the correlation between the scaled factor results and the matching original component. For each analysis the accuracy was calculated for both simulated components and the lowest accuracy of the two factors was recorded. The “space” columns are the accuracy of the reconstructions of the factor spatial topographies. The bottom row is the median score down each column of the table.

TABLE IX. Results of Simulation 3 ICA and PCA under high noise conditions

	C1	C2	tPCA time	tPCA space	sPCA time	sPCA space	tICA time	tICA space	sICA time	sICA space
1	N400	P1	0.98	0.99	0.65	0.80	0.79	0.88	0.89	0.89
2	N400	N1	0.96	0.99	0.83	0.88	0.85	0.79	0.95	0.74
3	N400	P300	0.35	0.58	0.99	0.84	0.91	0.88	0.27	0.33
4	N400	P2	0.70	0.12	0.39	0.76	0.73	0.83	0.97	0.75
5	P1	N1	0.60	0.59	0.91	0.30	0.89	0.84	0.08	0.55
6	P1	P300	0.98	0.99	0.89	0.87	0.54	0.86	0.98	0.89
7	P1	P2	0.97	0.99	0.46	0.82	0.74	0.93	0.95	0.76
8	N1	P300	0.99	0.99	0.74	0.74	0.93	0.95	0.99	0.98
9	N1	P2	0.98	0.98	0.83	0.65	0.84	0.63	0.99	0.99
10	P300	P2	0.99	0.99	0.53	0.79	0.90	0.89	0.97	0.89
	Median	Totals	0.98	0.99	0.79	0.79	0.84	0.87	0.96	0.83

PCA, principal components analysis; ICA, independent components analysis. C1 and C2 are the two simulated components in the dataset. tPCA is the results for the temporal PCA approach and sPCA is the results for the spatial approach. tICA and sICA are for the ICA results. The “time” columns are the accuracy of the reconstructions of the factor time courses, expressed as the correlation between the scaled factor results and the matching original component. For each analysis the accuracy was calculated for both simulated components and the lowest accuracy of the two factors was recorded. The “space” columns are the accuracy of the reconstructions of the factor spatial topographies. The bottom row is the median score down each column of the table.

therefore not necessary to repeat this examination in the temporal domain.

A follow-up comparison between using subject and grand averages was also conducted. For the grand average analysis, the analyses were applied to the grand average and then the resulting factor scoring coefficients were applied to the subject average data. The factor loadings and factor scores were then evaluated in the usual manner.

Results and Discussion: Simulation 4

The initial subject average results can be observed in Table X. The time correlations of the temporal analyses were not significantly different. The space correlations of

the temporal PCA was higher than the temporal ICA: $t(99) = 2.1, P = 0.036$. The time correlations of the spatial ICA were higher than the spatial PCA: $t(99) = 5.9, P < 0.0001$. The space correlations of the spatial ICA were higher than the spatial PCA: $t(99) = 18.2, P < 0.0001$.

The grand average results are presented in Table XI. The time correlations of the temporal analyses were not significantly different. The space correlations of the temporal PCA was higher than the temporal ICA: $t(99) = 7.7, P < 0.0001$. The time correlations of the spatial ICA were higher than the spatial PCA: $t(99) = 4.0, P = 0.0001$. The space correlations of the spatial ICA were higher than the spatial PCA: $t(99) = 8.2, P < 0.0001$.

TABLE X. Results of Simulation 4 comparing PCA and ICA under temporal and spatial approaches

	C1	C2	tPCA time	tPCA space	sPCA time	sPCA space	tICA time	tICA space	sICA time	sICA space
1	N400	P1	0.97	0.99	0.82	0.78	0.95	0.93	0.97	0.94
2	N400	N1	0.97	0.99	0.91	0.57	0.88	0.95	0.96	0.93
3	N400	P300	0.31	0.49	0.93	0.56	0.88	0.21	0.74	0.76
4	N400	P2	0.97	0.98	0.64	0.72	0.90	0.95	0.97	0.94
5	P1	N1	0.44	0.63	0.96	0.70	0.96	0.94	0.96	0.87
6	P1	P300	0.95	0.93	0.88	0.72	0.88	0.96	0.89	0.90
7	P1	P2	0.97	0.99	0.65	0.80	0.97	0.98	0.96	0.93
8	N1	P300	0.91	0.94	0.41	0.75	0.51	0.81	0.77	0.87
9	N1	P2	0.98	1.00	0.89	0.75	0.90	0.91	0.99	0.98
10	P300	P2	0.56	0.67	0.24	0.66	0.71	0.88	0.28	0.75
	Median	Totals	0.96	0.96	0.85	0.72	0.89	0.94	0.96	0.91

PCA, principal components analysis; ICA, independent components analysis. C1 and C2 are the two simulated components in the dataset. tPCA and tICA are the results for the temporal approaches and sPCA and sICA are the results for the spatial approaches. The “time” columns are the accuracy of the reconstructions of the factor time courses, expressed as the correlation between the scaled factor results and the matching original component. For each analysis the accuracy was calculated for both simulated components and the lowest accuracy of the two factors was recorded. The “space” columns are the accuracy of the reconstructions of the factor spatial topographies. The bottom row is the median score down each column of the table.

TABLE XI. Results of Simulation 4 comparing PCA and ICA under temporal and spatial approaches using the grand average

	C1	C2	tPCA time	tPCA space	sPCA time	sPCA space	tICA time	tICA space	sICA time	sICA space
1	N400	P1	0.98	0.98	0.91	0.89	0.95	0.90	0.94	0.99
2	N400	N1	0.72	0.98	0.92	0.61	0.86	0.52	0.93	0.99
3	N400	P300	0.34	0.52	0.55	0.64	0.35	0.57	0.47	0.70
4	N400	P2	0.46	0.51	0.86	0.96	0.87	0.63	0.87	0.98
5	P1	N1	0.98	0.90	0.96	0.80	0.54	0.84	0.91	0.80
6	P1	P300	0.90	0.81	0.56	0.50	0.80	0.04	0.88	0.89
7	P1	P2	0.95	0.94	0.69	0.86	0.96	0.89	0.96	0.99
8	N1	P300	0.85	0.39	0.79	0.93	0.86	0.43	0.81	0.83
9	N1	P2	0.79	0.97	0.93	0.99	0.64	0.27	0.91	1.00
10	P300	P2	0.77	0.67	0.52	0.42	0.78	0.37	0.72	0.83
	Median	Totals	0.82	0.86	0.82	0.83	0.83	0.55	0.90	0.94

PCA, principal components analysis; ICA, independent components analysis. C1 and C2 are the two simulated components in the dataset. tPCA and tICA are the results for the temporal approaches and sPCA and sICA are the results for the spatial approaches. The “time” columns are the accuracy of the reconstructions of the factor time courses, expressed as the correlation between the scaled factor results and the matching original component. For each analysis the accuracy was calculated for both simulated components and the lowest accuracy of the two factors was recorded. The “space” columns are the accuracy of the reconstructions of the factor spatial topographies. The bottom row is the median score down each column of the table.

Efforts to directly compare the results of Simulation 4 to the prior simulations must be tempered by the understanding that Simulation 4 had fewer simulated subjects (16 rather than 20) since real data was available from only 16 subjects for the auditory components. Nonetheless, the data in Table X bears a striking resemblance to that in the comparable medium noise Table VIII of Simulation 3. The median totals were essentially the same, except for that of spatial ICA, which were noticeably reduced. This difference was somewhat balanced by an improvement for the temporal ICA (space) scores. Overall, PCA yields the best results for the temporal approach and ICA yields the best results for the spatial approach.

Turning to the grand average results presented in Table XI, it appears that it yields degraded results compared to that obtained with subjects averages. This makes sense in that a factor scoring coefficient obtained from one dataset may not fully apply to a dataset with different variance components; even going from a grand average to subject averages could potentially change the composition of the dataset variance.

SIMULATION 5

A final question that we shall examine is whether these observations continue to apply when all five simulated components are included in the simulations.

Methods: Simulation 5

Simulation 5 was constructed in the same manner as Simulation 4 except that all five simulated components were included in each of the 100 simulated datasets. One other difference is that for this dataset no condition effect

or individual difference variance was included. For each simulation the factor most closely matching each original simulated component was identified with the constraint that each factor could only be matched to one simulated component; if a factor was the best match for two simulated components, the simulated component with the lesser fit was instead matched to the next best factor. Overall fit of each simulation was computed as the average of the absolute correlations of each of the five matched factors; the median value across the 100 simulations was utilized as the summary statistic. Ten factors were retained for spatial analyses and nine for temporal analyses.

Results and Discussion: Simulation 5

The overall results are summarized in Table XII. For the temporal analyses, the time correlations of the PCA were higher than for the ICA: $t(99) = 19.5, P < 0.0001$. Likewise, the space correlations of the PCA were higher than for the ICA: $t(99) = 21.8, P < 0.0001$. Conversely, for the spatial analyses, the time correlations of the ICA were higher than

TABLE XII. Results of Simulation 5 comparing results with all five simulated components

	Temporal time	Temporal space	Spatial time	Spatial space
PCA	0.74	0.75	0.61	0.60
ICA	0.70	0.53	0.84	0.67

PCA, principal components analysis; ICA, independent components analysis. Temporal, temporal approach; Spatial, spatial approach; Time, time course; Space, scalp topography.

for the PCA: $t(99) = 6.7$, $P < 0.0001$. Likewise, the space correlations of the ICA were higher than for the PCA: $t(99) = 77.0$, $P < 0.0001$.

DISCUSSION

In this report we sought to compare the efficacy of Promax and Infomax rotations for decomposing averaged ERPs. Since published analyses of these two techniques have often differed in terms of the approach (temporal vs. spatial) and the averaging procedure (subject averages and grand averages), these two parameters were examined as well. In addition, related issues such as variability in ICA solutions and solutions diagnostics were evaluated. Overall, it was found that Promax is most effective for the temporal approach and Infomax was the most effective for the spatial approach.

The first simulation demonstrated the presence of variability in ICA results, as previously noted by Makeig and colleagues in their web page (see above), suggesting it may be prudent to reset the pseudorandom number generator before each ICA run. The second simulation suggested that PCA may be most effective for the temporal approach and ICA may be most effective for the spatial approach. The third simulation suggested that these observations apply even in the presence of real noise. The fourth simulation suggested that this conclusion is applicable even with the addition of subject variability in component time course and topography. It also suggested that analyses based on the subject averages are preferable to those based on the grand average. The fifth simulation suggested that these results generalize to the case of all five simulated components.

The generality of these conclusions are circumscribed in a number of ways. The five simulated components represent five common ERP components but do not by any means constitute an exhaustive sampling of ERP components. The simulated datasets also differ from real datasets in that no more than five components were present. This choice was made in order to facilitate evaluation of the results. We argue that the addition of real EEG background noise increased the dimensionality to something approaching that of real datasets. Furthermore, we suggest that the use of PCA dimensionality reduction rendered such concerns largely moot since all but the largest sources of variance would be eliminated. The simulation datasets also did not include condition time course and topography variability. Given the minimal effects of introducing subject variability, we do not think such an addition would have notable effects. Finally, it should be kept in mind that there are many varieties of both PCA and ICA rotations, and so the present results are restricted to the ones that were evaluated.

Overall, we consider the general question of whether PCA or ICA is more effective to be poorly posed. The results of these simulations suggest that when utilizing a temporal approach to ERPs the PCA rotation Promax can be more effective, whereas when utilizing a spatial ap-

proach the ICA rotation Infomax can be more effective. This result can be described as occurring because for ERP data components are most cleanly separated in the temporal domain. Promax performs better for temporal analyses because it puts the temporal variance in the factor loadings, which is what Promax operates on when rotating. Infomax performs better for spatial analyses because it puts the temporal variance in the factor scores, which is what Infomax operates on when rotating. However, as discussed earlier, these conclusions should be considered to be general guidelines and not necessarily applicable to every dataset; for a given dataset, the characteristics of the features should be evaluated with respect to the parameters identified in this article.

Whether to use a temporal Promax or a spatial Infomax, on the other hand, cannot be stated categorically. As these simulations show, sometimes one or the other approach is more effective, depending on the characteristics of the data. Furthermore, as discussed elsewhere [Dien, 1998a], the spatial approach is more appropriate for analyzing temporal changes, whereas the temporal approach is more appropriate for analyzing spatial changes (such as laterality shifts). Finally, the case has been made elsewhere that for many datasets a two-step PCA [Dien et al., 2003b; Spencer et al., 1999] approach (spatio-temporal or temporo-spatial) is appropriate. For such an analysis, the present results suggest that the optimal procedure would use both PCA Promax and ICA Infomax for the temporal and spatial steps, respectively; however, since the first step collapses one of the dimensions (spatial or temporal), it is not clear if these observations would still apply for the second step. A future report is in the planning stages to examine the issues involved in two-step PCA in more detail.

The present simulations also provide some potential diagnostics for determining whether results may be problematic for both PCA and ICA. PCA results can be diagnosed by plotting the factor loadings in 2D space. The ICA results can be diagnosed by computing the sigma covariance measure. Further studies will be required for the latter measure to confirm its utility. In general, it is possible to diagnose problems by evaluating whether the factor time course has the accustomed unipolar pattern (although bipolar ERP effects are known). Problems with spatial topography are more difficult to diagnose but using degree of fit to dipole models may have utility. Mismatches between the topography of condition effects and the overall factor topography can also be indicative of problems in the analysis [see Dien et al., 1997, 2003a].

The pace of methodological improvements has quickened in recent years with the increased availability of computing resources. We may expect continuing improvements in both PCA and ICA methodologies to parallel this computing power. Currently, the main challenge is translating the methodological innovations of engineers and statisticians into more general use. It is sincerely hoped that this report helps this process and that it serves as a roadmap for physiologists and cognitive neuroscientists utilizing PCA and ICA to analyze their data.

ACKNOWLEDGMENT

We thank Todd Little for helpful comments on an earlier draft of the article.

REFERENCES

- Beauducel A, Debener S (2003): Misallocation of variance in event-related potentials: simulation studies on the effects of test power, topography, and baseline-to-peak versus principal component quantifications. *J Neurosci Methods* 124:103–112.
- Bell AJ, Sejnowski TJ (1995): An information-maximisation approach to blind separation and blind deconvolution. *Neural Comput* 7:1129–1159.
- Bentin S, McCarthy G, Wood CC (1985): Event-related potentials, lexical decision and semantic priming. *Electroencephalogr Clin Neurophysiol* 60:343–355.
- Bertrand O, Perrin F, Pernier J (1985): A theoretical justification of the average reference in topographic evoked potential studies. *Electroencephalogr Clin Neurophysiol* 62:462–464.
- Calhoun VD, Adali T, Pearlson GD, Pekar JJ (2001): Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. *Hum Brain Imag* 13:43–53.
- Cattell RB (1966): The scree test for the number of factors. *Multivariate Behav Res* 1:245–276.
- Cattell RB, Jaspers J (1967): A general plasmode (No. 3010-5-2) for factor analytic exercises and research. *Multivariate Behav Res Monogr* 67–63:1–212.
- Chapman RM, McCrary JW, Chapman JA, Bragdon HR (1978): Brain responses related to semantic meaning. *Brain Lang* 5:195–205.
- Comrey AL (1978): Common methodological problems in factor analytic studies. *J Consult Clin Psychol* 46:648–659.
- Cureton EE, D'Agostino RB (1983): Factor analysis: an applied approach. Hillsdale, NJ: Lawrence Erlbaum.
- Cureton EE, Mulaik SA (1975): The weighted varimax solution and the Promax rotation. *Psychometrika* 40:183–195.
- Curry SH, Cooper R, McCallum WC, Pocock PV, Papakostopoulos D, Skidmore S, et al. (1983): The principal components of auditory target detection. In: Gaillard AWK, Ritter W, editors. *Tutorials in ERP research: endogenous components*. Amsterdam: North-Holland. p 79–117.
- Delorme A, Makeig S (2004): EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 134:9–21.
- Dien J (1998a): Addressing misallocation of variance in principal components analysis of event-related potentials. *Brain Topogr* 11:43–55.
- Dien J (1998b): Issues in the application of the average reference: review, critiques, and recommendations. *Behav Res Methods Instrum Comput* 30:34–43.
- Dien J (1999): Differential lateralization of trait anxiety and trait fearfulness: evoked potential correlates. *Personal Individ Differ* 26:333–356.
- Dien J, Tucker DM, Potts G, Hartry A (1997): Localization of auditory evoked potentials related to selective intermodal attention. *J Cogn Neurosci* 9:799–823.
- Dien J, Frishkoff GA, Carbone A, Tucker DM (2003a): Parametric analysis of event-related potentials in semantic comprehension: evidence for parallel brain mechanisms. *Cogn Brain Res* 15:137–153.
- Dien J, Spencer KM, Donchin E (2003b): Localization of the event-related potential novelty response as defined by principal components analysis. *Cogn Brain Res* 17:637–650.
- Dien J, Frishkoff GA (2004): Principal components analysis of event-related potential datasets. In: Handy T, editor. *Event-related potentials: a methods handbook*. Cambridge, MA: MIT Press.
- Dien J, Beal DJ, Berg P (2005): Optimizing principal components analysis of event-related potential analysis: matrix type, factor loading weighting, extraction, and rotations. *Clin Neurophysiol* 116:1808–1825.
- Dodel S, Herrmann JM, Geisel T (2000): Localization of brain activity—blind separation for fMRI data. *Neurocomputing* 32–33: 701–708.
- Donchin E, Heffley E (1979): Multivariate analysis of event-related potential data: a tutorial review. In: Otto D, editor. *Multidisciplinary perspectives in event-related potential research* (EPA 600/9-77-043). Washington, DC: U.S. Government Printing Office. p 555–572.
- Fava JL, Velicer WF (1992): The effects of overextraction on factor and component analysis. *Multivar Behav Res* 27:387–415.
- Friedman D, Vaughan HG, Erlenmeyer-Kimling L (1981): Multiple late positive potentials in two visual discrimination tasks. *Psychophysiology* 18:635–649.
- Gorsuch RL (1983): Factor analysis, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.
- Greicius MD, Menon V (2004): Default-mode activity during a passive sensory task: uncoupled from deactivation but impacting activation. *J Cogn Neurosci* 16:1484–1492.
- Hansen LK, Larsen J, Kolenda T (2001): Blind detection of independent dynamic components. *Proc IEEE Int Conf Acoust Speech Signal Process* 5:3197–3200.
- Harman HH (1976): *Modern factor analysis*, 3rd ed. Chicago: University of Chicago Press.
- Hendrickson AE, White PO (1964): Promax: a quick method for rotation to oblique simple structure. *Br J Stat Psychol* 17:65–70.
- Hyvärinen A, Karhunen J, Oja E (2001): *Independent component analysis*. New York: John Wiley & Sons.
- Jackson JE (1991): *A user's guide to principal components*. New York: John Wiley & Sons.
- Jasper H (1958): Report on the committee on methods of clinical examination in electroencephalography. *Electroencephalogr Clin Neurophysiol* 10:370–375.
- Johnson MH, de Haan M, Oliver A, Smith W, Hatzakis H, Tucker LA, et al. (2001): Recording and analyzing high-density event-related potentials with infants using the geodesic sensor net. *Dev Neuropsychol* 19:295–323.
- Jung T-P, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ (2000): Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin Neurophysiol* 111:1745–1758.
- Jung T-P, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski T (2001): Analysis and visualization of single-trial event-related potentials. *Hum Brain Imag* 14:166–185.
- Kaiser HF (1958): The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 23:187–200.
- Kavanagh RN, Darcey TM, Fender DH (1976): The dimensionality of the human visual evoked scalp potential. *Electroencephalogr Clin Neurophysiol* 40:633–644.
- Kayser J, Tenke CE (2003): Optimizing PCA methodology for ERP component identification and measurement: theoretical rationale and empirical evaluation. *Clin Neurophysiol* 114:2307–2325.
- Kayser J, Tenke CE, Bruder GE (1998): Dissociation of brain ERP topographies for tonal and phonetic oddball tasks. *Psychophysiology* 35:576–590.
- Kramer AF, Donchin E (1987): Brain potentials as indices of orthographic and phonological interaction during word matching. *J Exp Psychol Learn Mem Cogn* 13:76–86.

- Lee T-W, Girolami M, Sejnowski T (1999): Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources. *Neural Comput* 11:609–633.
- Lutzenberger W, Elbert T, Rockstroh B, Birbaumer N (1981): Principal component analysis of slow brain potentials during six second anticipation intervals. *Biol Psychol* 13:271–279.
- Makeig S, Bell AJ, Jung T, Sejnowski TJ (1996): Independent component analysis of electroencephalographic data. *Adv Neural Inform Process Syst* 8:145–151.
- Makeig S, Jung T-P, Bell AJ, Ghahremani D, Sejnowski TJ (1997): Blind separation of auditory event-related brain responses into independent components. *Proc Natl Acad Sci U S A* 94:10979–10984.
- Makeig S, Westerfield M, Jung T-P, Covington J, Townsend J, Sejnowski TJ, et al. (1999a): Functionally independent components of the late positive event-related potential during visual spatial attention. *J Neurosci* 19:2665–2680.
- Makeig S, Westerfield M, Townsend J, Jung T-P, Courchesne E, Sejnowski TJ (1999b): Functionally independent components of early event-related potentials in a visual spatial attention task. *Philos Trans R Soc Lond B Biol Sci* 354:1135–1144.
- Makeig S, Jung T-P, Ghahremani DG, Sejnowski TJ (2000): Independent component analysis of simulated ERP data. In: Nakada T, editor. *Integrated human brain science: theory, method, application (music)*. Amsterdam: Elsevier. p 123–146.
- Matsumoto A, Iidaka T, Haneda K, Okada T, Sadato N (2005): Linking semantic priming effect in functional MRI and event-related potentials. *Neuroimage* 24:624–634.
- McKeown MJ, Makeig S, Brown GG, Jung T-P, Kindermann SS, Bell AJ, et al. (1998): Analysis of fMRI data by blind separation into independent spatial components. *Hum Brain Imag* 6:160–188.
- Möcks J, Verleger R (1991): Multivariate methods in biosignal analysis: application of principal component analysis to event-related potentials. In: Weitekamp R, editor. *Digital biosignal processing*. Amsterdam: Elsevier. p 399–458.
- Näätänen R, Picton T (1987): The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24:375–425.
- Park HJ, Kim JJ, Youn T, Lee DS, Lee MC, Kwon JS (2003): Independent component model for cognitive functions of multiple subjects using ^{15}O PET images. *Hum Brain Imag* 18:284–295.
- Polich J (1985): Semantic categorization and event-related potentials. *Brain Cogn* 26:304–321.
- Pritchard WS, Houlihan ME, Robinson JH (1999): P300 and response selection: a new look using independent-components analysis. *Brain Topogr* 12:31–37.
- Regan D (1989): *Human brain electrophysiology: evoked potentials and evoked magnetic fields in science and medicine*. Amsterdam: Elsevier.
- Richards JE (2004): Recovering dipole sources from scalp-recorded event-related-potentials using component analysis: principal component analysis and independent component analysis. *Int J Psychophysiol* 54:201–220.
- Rohrbaugh JW, Syndulko K, Lindsley DB (1978): Cortical slow negative waves following non-paired stimuli: effects of task factors. *Electroencephalogr Clin Neurophysiol* 45:551–567.
- Ruchkin DS, Johnson R Jr, Canoune HL, Ritter W, Hammer M (1990): Multiple sources of P3b associated with different types of information. *Psychophysiology* 27:157–176.
- Särelä J, Vigário R (2003): Overlearning in marginal distribution-based ICA: analysis and solutions. *J Machine Learn Res* 4:1447–1469.
- Sato W, Kochiyama T, Yoshikawa S, Matsumura M (2001): Emotional expression boosts early visual processing of the face: ERP recording and its decomposition by independent component analysis. *Neuroreport* 12:709–714.
- Schimmel H (1967): The (+/-) reference: accuracy of estimated mean components in average response studies. *Science* 157:92–94.
- Spencer KM, Dien J, Donchin E (1999): A componential analysis of the ERP elicited by novel events using a dense electrode array. *Psychophysiology* 36:409–414.
- Spencer KM, Dien J, Donchin E (2001): Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology* 38:343–358.
- Squires NK, Squires KC, Hillyard SA (1975): Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalogr Clin Neurophysiol* 38:387–401.
- Sutton S, Ruchkin DS (1984): The late positive complex: advances and new problems. *Ann N Y Acad Sci* 425:1–23.
- Vigario RN (1997): Extraction of ocular artefacts from EEG using independent component analysis. *Electroencephalogr Clin Neurophysiol* 103:395–404.
- Wood CC, McCarthy G (1984): Principal component analysis of event-related potentials: simulation studies demonstrate misallocation of variance across components. *Electroencephalogr Clin Neurophysiol* 59:249–260.
- Wood JM, Tataryn DJ, Gorsuch RL (1996): Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychol Methods* 1:354–365.
- Yee CM, Miller GA (1987): Affective valence and information processing. In: Johnson R, Rohrbaugh JW, Parasuraman R, editors. *Current trends in event-related potential research*. *Electroencephalogr Clin Neurophysiol Suppl* 40. Amsterdam: Elsevier. p 300–307.