ADDRESSING MISALLOCATION OF VARIANCE IN PRINCIPAL COMPONENTS ANALYSIS OF EVENT-RELATED POTENTIALS

Joseph Dien, Ph.D.

University of California, Davis

This is a post-peer-review, pre-copyedit version of an article published in Brain Topography. The final authenticated version is available online at: https://doi.org/10.1023/A:1022218503558

Abstract

Interpretation of evoked response potentials is complicated by the extensive superposition of multiple electrical events. The most common approach to disentangling these features is principal components analysis (PCA). Critics have demonstrated a number of caveats that complicate interpretation, notably misallocation of variance and latency jitter. This paper describes some further caveats to PCA as well as using simulations to evaluate three potential methods for addressing them: parallel analysis, oblique rotations, and spatial PCA. An improved simulation model is introduced for examining these issues. It is concluded that PCA is an essential statistical tool for event-related potential analysis, but only if applied appropriately.

Descriptors: principal components analysis, event-related potentials

Author's Note

This research was supported in part by NIMH grants MH42129, MH42669, MH18935, MH19554, MH11751, and by a grant from the Pew Memorial Trusts and the James S. McDonnell Foundation to support the Center for the Cognitive Neuroscience of Attention, and by Small Business Innovation Research grant MH51069 to Electrical Geodesics Incorporated, Eugene, Oregon.

This paper is based in part on a doctoral dissertation submitted to the University of Oregon in 1995. Thanks to my dissertation committee (Mike Posner, Libby Schaughency, Don Tucker, and Rick Zinbarg) for their helpful comments. Thanks also to Greg Miller and Kevin Spencer for their helpful comments on an earlier draft.

Addressing Misallocation of Variance in Principal Components Analysis of Evoked Potentials

Joseph Dien

Problems of Superposition

The superposition of activity volume-conducted from multiple regions of brain tissue has become an important obstacle to localizing evoked potentials. While a number of algorithms have been developed to solve this so-called inverse problem, overlapping events can cause significant errors in localization procedures (Zhang & Jewett, 1993). These procedures can also suffer when the number of sources is unknown as is usually the case (Achim, Richer & Saint-Hilaire, 1991). More generally, this problem of superposition is a challenge for all researchers seeking to interpret evoked potentials, whether for localization purposes or as an index of neurocognitive processes.

One method that is sometimes used to disentangle these portions of the waveform is principal components analysis (PCA). PCA is a multivariate technique for uncovering latent factors responsible for patterns of covariance in a set of variables (Gorsuch, 1983). Although localization algorithms (Scherg, 1990) often include PCA in some form, applying it as a separate pre-processing stage can allow it to be more effectively applied and its success independently evaluated. In particular, such a two-step process largely reduces the number of judgments necessary for modeling a dataset to that of how many factors (dipoles) to retain. The utility of this method was recently demonstrated with an auditory attention dataset (Dien, Tucker, Potts & Hartry, 1997). This paper will describe three refinements to this technique. To test these techniques, an improved simulation model will be presented. The results are of potential use to any analysis ERP data.

Temporal PCA

Temporal PCA repeatedly fits a regression line that accounts for the most variance possible, subtracting this variance and then fitting a new line to the residuals. Each such line constitutes a factor and

the correlations of the variables with a given factor are provided by the factor loadings. In the classical PCA of event-related potential (ERP) data (temporal PCA) the variables are the recorded potential at a given time point of the recording epoch (Curry et al., 1983; Donchin & Heffley, 1979; Möcks & Verleger, 1991), as shown in Table 1. Ideally, since the ERP components are the major source of covarying time points, this procedure should yield factors that correspond to each component.

Insert Table 1 about here

For each observation (raw waveform), temporal PCA will generate numbers representing the amplitudes of the latent variables (factor scores). The topography of each factor is encoded by the mean amplitude of its factor scores at each site. One can use this information to reproduce the portion of an observation's waveform represented by a given factor by multiplying the time point factor loadings by the observation's factor score and then multiplying each time point by its standard deviation (Dien et al., 1997). See the appendix for a simple proof.

Limitations of Temporal PCA

Temporal PCA has three major issues: 1) Retention criteria, 2) factor interactions, and 3) latency jitters (although others have been noted, c.f.,. Hunt, 1985). The first issue arises because PCA requires the analyst to decide how many of the factors contain interpretable signal and should therefore be retained. Retaining too many degrades the solution due to retained noise variance and retaining too little warps the solution (Wood, Tataryn & Gorsuch, 1996), producing misallocation.

The second issue, factor interactions, occurs when the factor solution for one component suffers interference from the factor solution of another component. Such interactions arise from rotational indeterminacy. To take a simple case, let us consider a dataset that consists of waveforms with only a P300, of varying amplitude but constant latency. In this case, a single latent waveform would suffice to

model the dataset (suitably weighted for each observation by a factor score) and there is no indeterminacy. The difficulty can be seen for a second dataset in which there is both a P2 and a P300. In this example dataset, one condition produces a P2 alone, a second condition produces a P300 alone, and a third condition produces both. Ideally, the dataset could be accounted for by two factors, one describing the time points involved in the P2 and another describing the time points involved in the P300.

However, it would be equally possible, and mathematically equivalent, to describe the observed waveforms with a latent waveform that describes the extent to which the two features covary for a given observation (consisting of both a P2 and a P300) and a second latent waveform that describes the extent to which the two features differ for a given observation (consisting of a P2 and a P300 of opposite sign). Summed together with the proper weights (factor scores), these two latent waveforms generate the observed waveforms as easily as a P2 and a P300 latent waveform could, as shown by Figure 1. For example, an observation with only a P2 could be modeled by a difference factor weighted such that its P300 segment cancels out that of the other factor, leaving only the P2. One could also generate any number of pairs of latent waveforms intermediate between these two extremes. The problem of rotational indeterminacy is the uncertainty which of these countless possible alternative pairs of statistical waveforms best reflects the real-world electrical waveforms. It is important to keep in mind that this issue is a problem for peak measures as well.

Rotation procedures attempt to address this indeterminacy. The rotation most commonly used, Varimax (Kaiser, 1958), finds the set of equivalent waveforms that maximize the importance (loading) of the time points that are large for the factor and minimize time points that are small (by maximizing the sum of the fourth power of the loadings). This criterion has the value of rotating away from solutions that have only moderate loadings, presumably spread across multiple components. To the extent that ERP components peak at different time points and have relatively focal time courses, this criterion should find factors that approximate the true latent waveforms. This is reasonable because, for the most part, ERP components appear to be monophasic.

Rotation procedures do not necessarily resolve indeterminacy correctly. In an influential paper, it was demonstrated that rotational indeterminacy can result in condition effects being misallocated to an overlapping factor (Wood & McCarthy, 1984). Ironically, although this paper is most frequently cited by critics of this technique, these authors actually advocated its use. Their conclusion was that although misallocation of variance is a problem for PCA, it is also a problem for other measurement techniques as well. Peak amplitude measures, for example, will also confound the effects of overlapping components. Indeed, they stated that, "Other approaches to ERP analysis, measurement of peak amplitudes and latencies for example, are no less subject to the problem of component overlap than PCA; they simply make it easier to ignore by not representing it explicitly," p. 258. If nothing else, PCA can alert the researcher to the presence of overlapping components even when it does not resolve them correctly.

Although Wood and McCarthy (1984) concluded that component overlap is the cause of such misallocation, a reanalysis of their simulation dataset resulted in the conclusion that the misallocation was due to an inadvertent correlation between the two components rather than due to their overlap (Chapman & McCrary, 1995). When the correlation was removed, it was reported that the misallocation was eliminated even though the components were still overlapping. Some simulation data has been presented that seems to suggest that correlated components are only a problem for overlapping components (Hunt, 1985) but it is not clear why this should be. A goal of this paper is to resolve these divergent conclusions.

A point not made by Wood and McCarthy (1984) is that misallocation of variance can result in components being missed entirely. Factors are defined as patterns of correlations (or covariances) between time points without regard to what the source of these correlations are. A single factor can therefore describe multiple ERP components as long as they have identical time courses (Möcks & Verleger, 1991). If two components differ in time course only to a small extent, they may be described by a large factor that encompasses their common time course and a small factor that describes their differences. If this small factor is then obscured by the clutter of random noise, it may appear that there is only one factor rather than two. One might say that one component has been entirely misallocated to another.

The third issue is that of latency jitter, which is variability in the time course of a component across trials. If a given ERP component, such as the P300, occurs at two different latencies in two different conditions, then a single latent waveform will not be able to account for both observed waveforms. Ideally, the dataset could be accounted for by two factors, one wholly describing the P300 at the first latency and one wholly describing the P300 at the second latency. This situation may result in a single latent waveform that captures the time points that are activated by both latencies and a second that captures the difference between the two latencies. Thus, a feature that shifts across the individuals or the conditions can result in an extra principal component resembling the time-derivative of the feature's behavior (Möcks, 1986). Alternatively, it might produce two factors representing the two different latency versions. A given factor may represent latency jitter in a component already modeled by a second factor rather than a distinct ERP component. This is, again, an issue common to peak measurement techniques. A negativity that appears at two different times in two different conditions may in fact be due to two distinct ERP components or it may be due to a single jittered component (c.f., Polich, 1985). The PCA is merely making this issue more apparent.

These difficulties with misallocation and jitter can be addressed via parallel analysis, oblique rotations, and spatial PCA respectively. After the explication of these three techniques, a series of simulations will test their utility.

Parallel Analysis

Up to a point, rotation procedures can address such cases involving difference factors due to overlap or latency jitter. However, the original two components can only be regenerated if the difference factor is retained. PCA will usually produce as many factors as there are variables, mostly representing random noise. For parsimony's sake most of the factors are dropped, retaining only the substantive factors. It is therefore critical that the difference factors representing overlapping components are retained. Retaining too many factors degrades the solution due to retained noise variance and retaining too few warps the solution (Wood et al., 1996), producing misallocation.

Unfortunately, the guidelines for determining the number of factors to retain are imprecise. The most common method, the Kaiser eigenvalue-greater-than-one rule (Kaiser, 1960), has been strongly criticized for providing inferior estimates (Zwick & Velicer, 1986). The second most common decision rule, the Scree test (Cattell, 1966b), is more accurate but contains a substantial subjective component. It is predicated on the fact that a PCA of a random dataset will produce a set of factors with random size. When ordered by size (eigenvalues), they will produce a smoothly descending slope. Factors that are larger than predicted by this steady slope are most likely to have interpretable signal in them. The scree test involves graphing this set of eigenvalues and looking for the "elbow," the point where the factors begin rising above the slope. Simulations indicate that the last point still on the slope, the corner of the elbow, should also be included (Cattell & Jaspers, 1967). The problem is that scree plots of ERP data typically contain multiple elbows, some of which are quite subtle, producing ambiguity as to where the scree starts.

In a comparison of five methods for determining how many factors to retain (Zwick & Velicer, 1986), parallel analysis (Horn, 1965) proved to be the most effective. In this test, PCA is conducted on a random dataset of the same size as the dataset of interest, producing an assortment of random sized factors. When charted in order of size (eigenvalues), this produces a slope. This slope can be used as a baseline against which to compare the factors from the PCA of the experimental dataset. The number of factors that are larger than obtained from the purely random dataset is the number to retain for further analysis. This test therefore uses the same logic as the scree test but removes the guesswork about what constitutes the noise level. When applied to ERP datasets, a complication is that even after averaging the background noise is autocorrelated across the variables. The autocorrelated noise causes factors to appear that are larger than would be expected from a purely random dataset even in the absence of evoked potentials. This point may be addressed by conducting the averaging procedure on the dataset with every other trial inverted, producing what has been termed the +/- reference (Schimmel, 1967; Wong & Bickford, 1980). In this case, the evoked potential should cancel out leaving only the background noise. This inversion dataset should then provide an optimal comparison point for the parallel analysis.

Oblique Rotation

One possible way to address factor interactions is to use oblique rotations. As noted in an important critique of the application of PCA to ERPs (Hunt, 1985), distortions can occur when assumptions of the statistical analysis are violated. An important assumption of PCA and varimax is orthogonality, that the latent variables are uncorrelated. The use of an oblique rotation can address violations of this assumption by allowing factors to be correlated. It has been previously suggested that an oblique rotation might provide better results and was used in passing to demonstrate that two factors were correlated in a simulation dataset (Chapman & McCrary, 1995). The issue was not further pursued or evaluated however.

One of the better oblique rotations (Gorsuch, 1970) is the Promax procedure (Hendrickson & White, 1964), in which a Varimax rotation is relaxed to allow correlation among the factors. As described earlier, in the Varimax procedure the factor vectors are rotated in variable space such that loadings of the variables on the factors are as extreme (either zero or high absolute value) as possible. This results in factors that are restricted to as few variables (time samples in temporal PCA) as possible. This operation is limited by an orthogonality constraint; variance cannot be shared by factors. Factors representing correlated components can be distorted when their shared variance is allocated to other factors.

Promax pursues the Varimax criterion without regard for orthogonality; in effect, allowing multiple factors to share ambiguous variance. Although this approach loses the mathematical simplicity of strict orthogonality, it could allow the individual factors to more closely approximate the underlying components, to the extent that the Varimax criterion is valid. Using the Varimax solution as the starting point, the Promax solution allows factors to become correlated, a condition more suited to brain processes than strict orthogonality. It does so by adjusting each factor in turn to more strongly follow the varimax criterion for simple structure, but this time without regard for maintaining orthogonality to the other factors. Note that the Promax operation is carried out after the retention step so it will not affect the number of factors retained, only their characterization.

Spatial PCA

The high-density montages afforded by advances in current technology have made another approach, spatial PCA, an option for dealing with latency jitter. Although studies using spatial PCA have been published previously (Donchin, Spencer & Dien, 1997; Duffy et al., 1990; Kavanagh, Darcey & Fender, 1976; Skrandies & Lehmann, 1982), these consisted of brief demonstrations. To evaluate the utility of this method, in-depth examination is necessary.

In a spatial PCA, the variables are the microvolts measured at a given channel and the time points serve as observations, as shown in Table 2. The resulting factors consist of topographical patterns with each factor loading describing the weighting of an individual channel. The factor scores indicate the amplitude of these topographical patterns across time. This arrangement will produce a factor solution that differs from a temporal PCA due to emphasis on the spatial variance and due to the rotation procedure.

Insert Table 2 about here

The first reason spatial PCA will differ from temporal PCA lies in the role of four different sources of variance in the data. These four sources are temporal (the waveforms), spatial (topographical patterns), condition (experimental effects), and participant (individual differences). In temporal PCA the dataset consists of waveforms from all the scalp sites from all the conditions from all the participants. In this case, the PCA is conducted on the combination of the temporal variance and the temporal covariances produced by the effects of the site, effects of the conditions, and effects of individual differences. This is appropriate for an ERP dataset since a component should be subject to all three influences. These are also some of the sources of information by which an ERP analyst determines what is a component (Picton & Stuss, 1980). The temporal PCA process helps summarize the information about temporal patterns related to these three sources. Spatial PCA represents the complementary case of analyzing spatial variance and covariance.

It is not correct to say that the difference between spatial and temporal PCA is the same as that between R analysis (in psychometrics, tests as variables and subjects as observations) and Q analysis (subjects as variables and tests as observations) (Cattell, 1966a). Regular PCA operates along two modes (variables and observations). In a simple case where there is only temporal and spatial variance, it does not matter which is used as the variables in this respect. For a temporal PCA, two components with the same time course have the same profile and cannot be distinguished, as noted earlier. Two components could have different time courses but the same spatial distribution and would again not be distinguishable as they would covary absolutely. At whatever site the set of time points affected by one component was large, the time points affected by the other component would be large too. Wherever the former set was small, so would the latter set. The result would be a single bimodal factor. If two components were identical either spatially or temporally, it would not matter whether spatial or temporal PCA was used.

However, in real datasets the observations mode is potentially disambiguated by two other sources of variance, individual differences and condition effects. For example, in a temporal PCA, two components with different time courses but identical spatial distributions will not absolutely covary if one is amplified by attention and the other is not. A spatial PCA is no longer simply a transpose of a temporal PCA. Whichever source of variance contributes to the observations mode has the benefit of the two other sources of variance for disambiguating components but is therefore also diluted in its effects on the final factor solution. Thus, spatial PCA will be most effective at separating components with similar time courses although temporal PCA will be most effective at separating components with similar scalp topographies.

Spatial PCA will also produce different results due to the effects of the rotation procedure. As described earlier, Varimax finds the rotation that minimizes the number of variables that loading on the factors. In a temporal PCA, this property will favor factor solutions with temporally delimited waveforms. In a spatial PCA, this property will favor factor solutions with spatially delimited waveforms.

Temporal and spatial PCAs have complementary strengths. A spatial PCA can characterize components that are temporally overlapping or jittered whereas a temporal PCA can identify components

that are spatially overlapping or jittered. For example, two components that occur at nearly the same time but have very different scalp distributions might be characterized by a temporal PCA as a single substantive factor (with the small difference variance lost in the noise factors) while a spatial PCA would characterize them with two different equal-sized interpretable factors. Additionally, a temporal PCA keeps the time course constant but allows the spatial distribution to vary, allowing the distribution of different conditions to be compared. In contrast, a spatial PCA keeps the spatial distribution constant but allows the temporal course to vary, so different conditions may be compared temporally. Thus, each type should detect components that the other misses, as well as providing information about condition changes that the other keeps invariant.

In general, a temporal PCA will allow stronger inferences since all spatial factors necessarily overlap and potentially suffer from misallocation. Balancing this methodological issue is the fact that, to the extent that ERP components are defined as reflecting different neural functions (however, see Picton & Stuss, 1980), components are more likely to have unique spatial signatures than temporal signatures.

While the basic principles elucidated suggest these three techniques should be useful adjuncts to the PCA technique, simulations are necessary to validate the conclusions and to detect unforeseen complications. Simulations were therefore carried out to test their effectiveness. In order to improve upon the simulation dataset first introduced by Wood and McCarthy (1984), a new set of simulations were constructed which permit controlled comparison of the effects of component overlap and component correlation, as well as the individual roles of spatial, condition, and subject variance.

Simulations of Parallel Test

Methods

Test datasets were constructed to represent ten participants in two conditions, each with 65 channels and 65 time points. Two artificial components were constructed from half sine cycles, one with a short period like that of the P2 component(s) and one with a long period like that of the P3 component(s). These two components overlap such that when a correlation is computed between their loadings (treating each pair of loadings as an observation), *r* is approximately zero. To maximize comparability between the temporal and spatial dimensions, the test montage was conceptualized as being a midline montage of 65 electrodes. While the spatial layout of the electrodes is irrelevant to the mathematics of the PCA procedure, this makes it reasonable to display topographies as a linear ordered set in the same manner as the temporal patterns. Figure 2 shows that the half sine cycles can comprise the topography of the components as well, with the short period cycle representing a focal component like that of the P2 and the long period cycle representing a diffuse component like that of the P3. These spatial and temporal weights varied from zero to one. A given data point consisted of the product of the two weights and then arbitrarily multiplied by four microvolts.

Insert Figure 2 about here

Random variability in the dataset is represented by the background EEG of ten subjects from a real dataset (Dien, in press) and was generated by using the +/- reference (Schimmel, 1967). To further minimize time-locked evoked potentials, the 65 time points were taken from the baseline period plus 76 msec. post-stimulus. In order to keep the model simple, no other random variance was included in the model.

Results and Discussion

A base matrix was constructed containing only the two components and no background EEG. When this base data matrix was factored, only two non-zero eigenvalues were obtained, confirming that this artificial dataset has only two components. The background EEG was then added to the base data matrix, producing the final dataset. Figure 3 illustrates how dataset was then factored and the resulting scree compared to that of the background EEG alone. A notable increase is seen in three of the eigenvalues although small increases are seen in the remaining factors as well.

Insert Figure 3 about here

Following the parallel test criteria, four factors were retained (the three plus the elbow) and rotated using the promax algorithm. For comparison's sake, the background EEG was also factored, retaining four factors. Figure 4a shows that the waveforms of the four background EEG factors are quite coherent, verifying that the averaging process has left more than just random noise in the background. Figure 4b of the waveforms for the PCA of the background EEG + base data matrix shows that Components P3 and P2 were indeed mostly recovered along with two noise factors. It appears that Factors 1 and 4 of the background EEG may have been incorporated into the Factors characterizing P2 and P3 due to their similar peak times, distorting the factors somewhat. One can also infer that the reason more than two signal factors were indicated by the parallel test is that some of the variance of the test components have been "misallocated" in this initial extraction. Factor 2 shows a small bump coincident with the P2 peak in b, not evident in a, which suggests it may have incorporated some of the P2 variance. Likewise, Factor 3 show a small dip coincident with the P3 peak in b, but not in a, that suggests is may have incorporated some of the P3 variance. Indeed, there is no reason in principle that misallocation cannot occur in the initial extraction.

On the other hand, four factors turns out to be an appropriate number to retain. Figure 4c demonstrates that if one retains only two factors, the P2 and P3 factors are distorted. This is in line with reports that

retaining too few factors can result in distorted results (Wood et al., 1996). Figure 4d further shows that if one retains six factors the P2 and P3 factors appear even cleaner, adhering closer to zero outside the duration of the component. These findings seem to suggest that the results of the parallel test should be considered merely a higher bound for estimating the dimensionality of the evoked potential. Moreover, it appears that the number of factors retained may need to be an even higher number since features contributed by the background EEG may cause distortions if not characterized by additional factors.

Insert Figure 4 about here

Simulation Tests of Oblique Rotation

The goal of the next simulation test is to examine the effect of the oblique rotation, Promax, in the presence of temporal overlap and component covariation. In this manner, it is also hoped to resolve the dispute between the opposing positions on their role in misallocation of variance (Chapman & McCrary, 1995; Wood & McCarthy, 1984).

Methods

For this next series of simulations, three temporal patterns were used. The temporal patterns of the P2 and P3 patterns from the preceding test were used as well as a third (to be termed the P1 pattern) with the same short duration as the P2 pattern but located earlier so as not to overlap with those of the P3 pattern. The P1 and P2 patterns were used to produce the four possible combinations of having temporal and spatial overlap/non-overlap with the P3 pattern. Overlap in the spatial domain determines overall correlation between the components in a temporal PCA. The P2 pattern results in a zero correlation with the P3 pattern (overlap) while the P1 pattern (non-overlap) results in a -.3 correlation (since the presence of one partially predicts the absence of the other).

The dataset was kept as simple as possible. The basic data consisted of 65 variables (time points) at 65 observations (electrodes). Since there has to be more observations than electrodes to avoid singularity (which interferes with computing the generalized negative inverse used to calculate the factor score coefficients) these observations were doubled (representing two identical conditions, for simplicity's sake). Since simply doubling the data would not prevent the matrix from being singular, a very small random noise term was added. The noise term amounted to only about .5% of the average peak amplitude of the signal (centered on zero) and was added to each data point in the data matrix. One hundred datasets were generated for each simulation. In addition to the noise term, additional randomness was introduced (in the interests of generalizability of the results) by varying each component amplitude within a +/-50% range.

Misallocation of variance was evaluated in terms of defective reconstruction of the components. Effectiveness of component reconstruction was quantified by measuring the correlation between the original time course (or topography) and the corresponding factor using each time point (or electrode site) as the observations. The measure has the additional advantage that since correlations normalize the two variables involved, the waveform amplitudes is controlled for. The correlation was computed between the varimax rotated P1/P2 factor (as the smaller component, it is expected to show misallocation more clearly) and the original time course. The 100 replications were then rank ordered according to the correlations and the median (50th) simulation was selected as the most representative result for each of the simulations.

Results and Discussion

Simulation A is a zero temporal overlap, zero component correlation. The correlations ranged from 1.000 to .9681 with a median of .9994. As Figure 5A presents, there is essentially no misallocation of variance for either time course or topography, with only a subtle dip suggesting some spillover from the P3 component.

In simulation B, there is temporal overlap and negative correlation between the components, one or both of which should produce misallocation of variance (Chapman & McCrary, 1995; Wood & McCarthy, 1984). The correlations ranged from .7370 to 1.000 with a median of .9949. Even with this high

correlation, Figure 5B reveals a notable misallocation, especially in the topography. Promax rotation improves the fit slightly to a correlation of .9976.

To disentangle the effects of temporal overlap and component correlation, simulation C contains a negative correlation but no temporal overlap. Correlations range from .8522 to .9979 with a median of .9673, reflecting a notable spillover from the P3 component in both time course and topography as seen in Figure 5C. Promax essentially eliminates the distortion, improving the correlation to .9999.

To determine if the reverse can also cause distortion, simulation D contains temporal overlap but no component correlation. Correlations range from .8877 to 1.000 with a median of .9912, reflecting some misallocation, particularly in the topography, as Figure 5D shows. Promax increased the time course distortion to .9882 but decreased the topography distortion from .9648 to .9945, suggesting this rotation does not reliably address temporal overlap.

Insert Figure 5 about here

These results indicate that both McCarthy and Wood (1984) and Chapman (1995) are correct. At least under some situations, both component overlap and component correlations can produce misallocation of variance. Only when both were absent (Simulation A) was there accurate reconstruction of the components using the customary varimax rotation.

These results also suggest that, as Chapman suggested, an oblique rotation can indeed be helpful. Although it had mixed effects on misallocation due to temporal overlap, Simulation C shows it can essentially eliminate misallocation due to correlated components. Given that it is unlikely that most components will have the exact spatial overlap necessary to produce zero correlation, let alone having uncorrelated sources, such a property seems likely to be of use.

Although this example focused on time course and topography with an eye towards localization efforts, these findings are equally valid for misallocation of experimental effects. When a component has condition effects, it should maintain them when misallocated to different factors as demonstrated by Wood and McCarthy (1984).

Simulation Tests of Spatial PCA

The final simulation test is intended to contrast the characteristics of spatial PCA with the more traditional temporal PCA. From first principles, it is expected that spatial PCA should be able to separate components in some situations that temporal PCA cannot. It is also expected that temporal PCA should be better suited to modeling topography changes and spatial PCA should be better suited for time course changes.

Methods

Simulation datasets were constructed as in the previous series using just the P1 and P3 patterns. Jittered versions were produced by shifting waveforms to the right by five places. Representative replications were chosen arbitrarily by taking the correlation between the Factor 1 (promax rotated) time course and correlating it with the P3 component time course and taking the replication with the median value. For simulations B-D, condition variance was introduced by doubling the size of the P1 component in the 2nd condition. For simulations E-H, jitter is introduced by shifting the appropriate pattern five spaces in the 2nd condition. Representative replications were chosen arbitrarily by choosing the one with the median correlation between Factor One and the non-jittered P1. In simulations E & G, whether the jittered or non-jittered P1 gravitated to Factor One was random so for the median calculation only the replications where Factor One reflects the non-jittered P1 were included so that a replication most representative of one of these groupings was selected rather than an outlier falling in between these two groups.

Results and Discussion

The first set of simulations help delineate one situation where a spatial PCA will dissociate components more effectively than temporal PCA. As already seen, at least with this dataset, a temporal PCA will readily separate the two components when they differ in both temporal and spatial characteristics.

Figure 6A shows (as discussed earlier) that when the two components have the same spatial pattern, temporal PCA cannot dissociate them since they will correlate 100%. Factor 1 accounts for both the P1 and P3 components in this case (>99.9999% of variance) while Factor 2 consists of incoherent noise (<00.0001% variance). It is not clear why the topography of Factor 2 appears to reflect the topography of the P3 component even though the time course appears random and may be a glitch prompted by the nearly non-existent variance involved.

Figure 6B shows that even when two components have the same spatial pattern, temporal PCA can dissociate them as long as there is some differential condition (or subject) variance. In this case, although the PCA was able to identify the dimensionality as being two (96.53% and 3.46% variance respectively), the factors were quite distorted since the components were still highly correlated.

Figure 6C shows that when two components have the same temporal pattern, neither differing spatial topography nor condition variance is sufficient for a temporal PCA to dissociate them. Once again, the first factor accounts for nearly all the variance (>99.9999%) while the second factor represents some residual noise (<00.0001% variance).

Figure 6D demonstrates with this same dataset that a spatial PCA is able to dissociate the components, although there was considerable misallocation of variance. Factor One (96.27% variance) reflects the P3 component and some of the P1 while Factor Two (3.73% variance) characterizes the P1 component and only some of the P3.

Insert Figure 6 about here

The second set of simulations illustrate the performance of temporal and spatial PCA under temporal and spatial jitter. For simplicity's sake, only a single component (the P1) was included in the simulation datasets. As has been demonstrated previously (Möcks, 1986), Figure 7E shows that latency jitter in a component can produce two factors (74.26% and 25.70% variance) in a temporal PCA.

Figure 7F shows that a spatial PCA of this same dataset more meaningfully captures the components with just one factor (99.96% variance). The second factor is inconsequential (<00.0001% variance) although it has some coherence.

Figure 7G shows that spatial jitter can produce a multiple factor effect in a spatial PCA just as temporal jitter does for temporal PCA. The two factors account for 72.92% and 27.05% of the variance respectively.

Figure 7H illustrates that, unlike spatial PCA, a temporal PCA smoothly handles topography changes. The first factor has 99.97% variance while the second is vanishingly small (<00.0001% variance).

Insert Figure 7 about here

These simulations support the predictions made in the introduction that temporal and spatial PCA have complementary strengths and weaknesses. For modeling components with nearly identical time courses or with substantial latency jitter, spatial PCA may produce less misallocation problems. On the other hand, since electrical fields affect the entire head (to varying degrees), unlike in temporal PCA each factor has loadings on every variable (site). This means that misallocation of variance is much more likely to occur since all components overlap.

Conclusion

Whether PCA is successful at parsing components depends on the definition of "component." Unfortunately, there is no simple answer which is why components tend to be defined according to a number of criteria (Picton & Stuss, 1980), forming fuzzy categories. For example, a strict latency definition would fail to categorize a component like the P300 which varies according to stimulus evaluation time. On the other hand, a strict topographical or source definition would fail to group together motor potentials which will be different between fingers and toes, or even between different fingers, and yet are roughly equivalent in function and characteristics. A component, like any other theoretical construct, is essentially an arbitrary category defined by researchers that supports experimentally and theoretically useful generalizations that facilitate interpretation and communication. If PCA produces factors that are useful and generalizeable for the goals of the experiment, then it has been successful. Useful can mean the factor shows coherent condition effects and/or is readily localizeable. Generalizeable means the factor is replicable and has convergent validity when compared to other analysis techniques or other sources of information.

On the basis of these simulations, one can arrive at some recommendations for PCA of ERP data. 1) Careful attention should be focused on the issue of how many factors to retain. The parallel test may provide some assistance but overall does not appear to provide much improvement over a simple scree test. A more practical procedure may be to use the scree test to arrive at an initial estimate and then to determine whether a solution with an additional four factors adds or changes the results in any substantive fashion. 2) Promax appears to provide a substantial improvement over varimax and should be used in all analyses. 3) Temporal PCA is useful when it is expected that the features of interest have spatial jitter or have less temporal overlap than spatial overlap. Spatial PCA is useful when it is expected that there is temporal jitter or less spatial overlap than temporal.

Although improvements on PCA have been proposed (e.g., Bell & Sejnowski, 1995; Maier, Dagnelie, Spekreijse & van Dijk, 1987; Mosher, 1992), in order to be useful they must have their limitations and

strengths mapped out in the same manner as has been done for PCA in this and previous articles. For example, PCA has been generalized to 3-modes or more (Kroonenberg, 1983; Tucker, 1963). An early application used the three modes of subject, conditions, and time points (Donchin, Gerbrandt, Leifer & Tucker, 1972). More recently, it has been proposed using the modes of subjects, channels, and time points (Möcks, 1988). Its very strength of describing components in terms of a fixed time course and topography is also its weakness as this means it has both the weakness of spatial PCA to spatial jitter (as in laterality effects) and temporal PCA to temporal jitter (as in the P300). It is not clear at this point how robust it would be to misallocation of variance issues, although it appears to be promising in cases where both spatial and temporal jitter is lacking (Achim & Marcantoni, 1997). A recent extension to this technique may be able to take latency jitter and stretching into account but remains a work in progress and is not yet publicly available (Achim & Bouchard, 1997). Likewise, independent components analysis (Bell & Sejnowski, 1995; Makeig, Jung, Bell, Ghahremani & Sejnowski, 1997) studies done thus far have been applied in a fashion analagous to spatial PCA and should be evaluated accordingly.

In conclusion, both PCA methods provide useful information about the component structure of the ERP. It should also be apparent that the richness of the evoked potential, as revealed by PCA, cannot be appreciated by a simple windowing procedure. While concerns about misallocation of variance are valid, to avoid therefore the use of PCA is to kill the messenger carrying the bad news that coping with superposition is a challenge for ERP analyses in general. Ideally, one should use experimental manipulations to isolate components as best possible and then characterize the component structure of the resulting effects with the appropriate PCA. ERP studies not doing so should have to justify any conclusions they make concerning componentry, particularly latency measures and neural generators. Such a PCA may also prove helpful with subsequent localization efforts. With sufficient observations it could be feasible to apply these techniques to functional MRI as well, separately or in conjunction with ERP data.

References

Table Legends

TABLE 1. Example of dataset for temporal PCA. Variables consist of the voltage measured at each of *t* time points. The observations consist of *m* waveforms (the waveforms measured at all the combinations of *p* participants, *c* conditions, and *n* channels).

TABLE 2. Example of dataset for spatial PCA. Variables consist of the voltage measured at each of *n* channels. The observations consist of *m* topographies (the scalp patterns measured at all the combinations of *p* participants, *c* conditions, and *t* time points). Note that the spatial PCA dataset is not simply the transpose of the temporal PCA dataset but rather the resultant of the separate transposition of each of the *p* x *c* blocks of *t* time point by *n* channels.

Tables

Table 1.

	Time 1	**Time 2**	**Time *t***
Waveform 1 (participant 1, condition 1, channel 1)	0 μv	1 μv	1 μv
Waveform 2 (participant 1, condition 1, channel 2)	0 μv	2 μv	2 μv
Waveform *m* (participant *p*, condition *c*, channel *n*)	3 μv	3 μv	0 μv

Table 2.

	Chan 1	**Chan 2**	**Chan *n***
Topography 1 (participant 1, condition 1, time 1)	0 μv	0 μv	2 μv
Topography 2 (participant 1, condition 1, time 2)	1 μv	2 μv	4 μv
Topography m (participant *p*, condition *c*, time *t*)	1 μv	2 μv	0 μv

Figure Captions

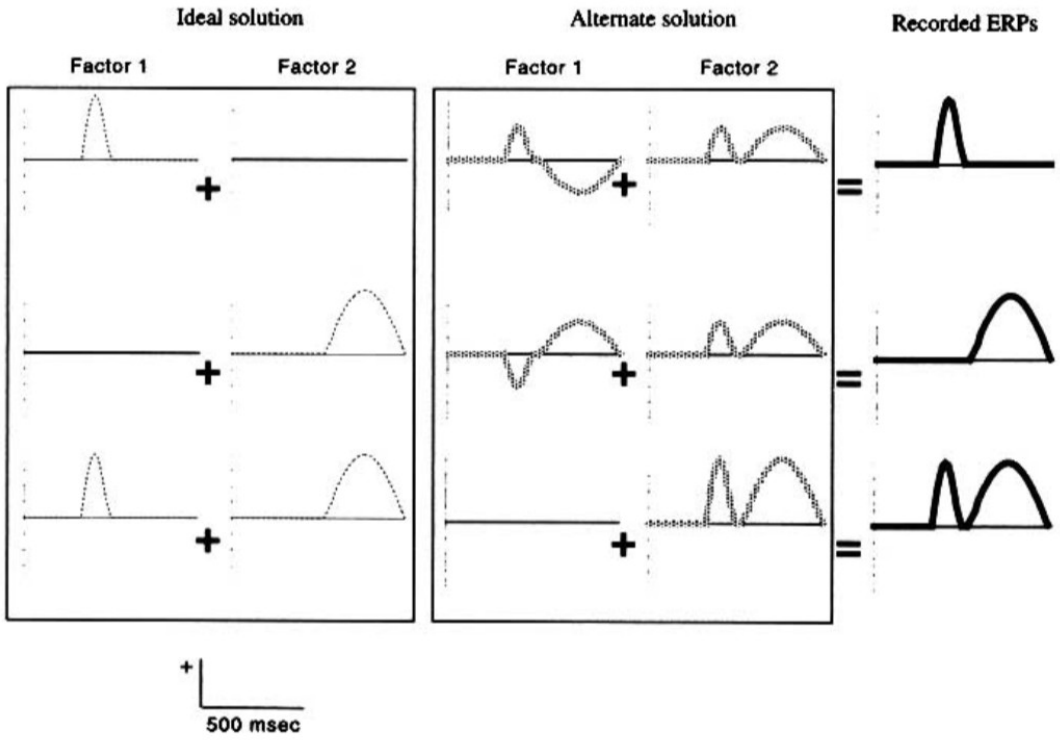


FIGURE 1. Example of alternative factor solution that can account for the P2 and P3. The ideal solution consists of separate factors for each component that together can account for waveforms (rightmost column) consisting of the P2 alone, the P3 alone, and both together. Because of rotational indeterminacy, another pair of factors can also be generated (alternate solution) consisting of a contrast between the two components and the combination of the two components that together can also account for the same three recorded waveforms. While mathematically equivalent, the electrophysiological literature clearly indicates that the first solution is more plausible.

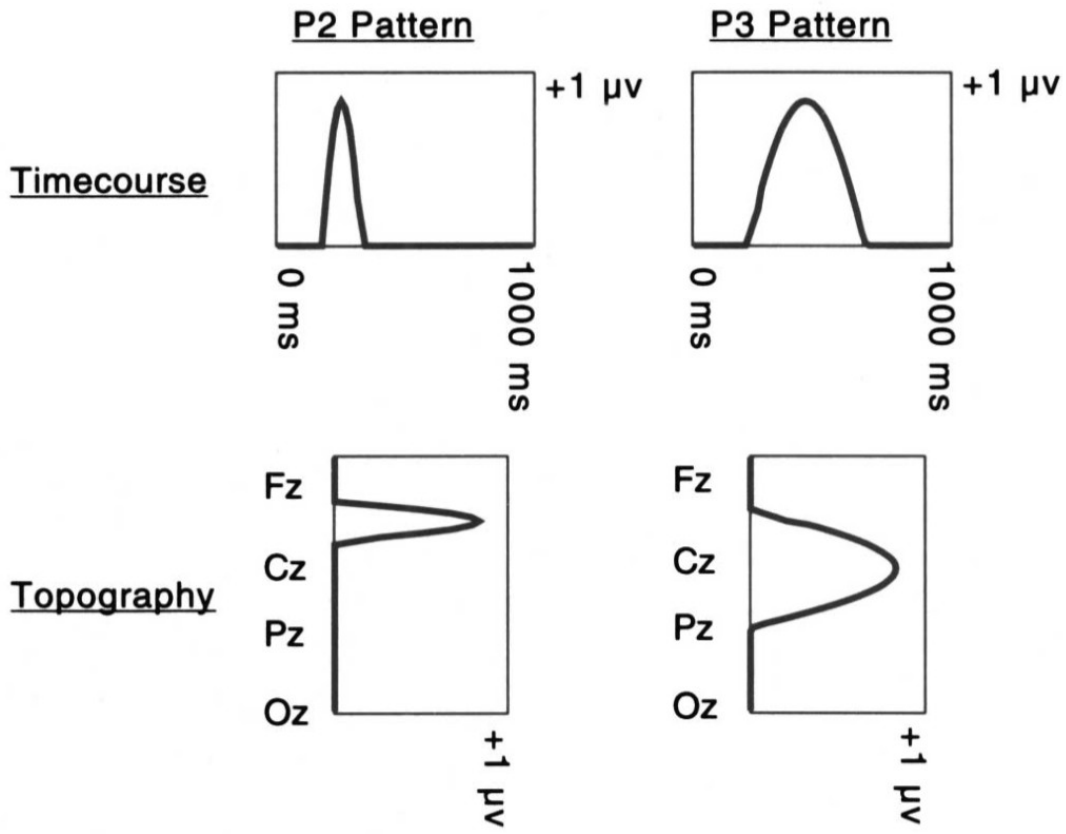


FIGURE 2. Artificial components for simulation studies. The figure illustrates how the "P1" and "P3" patterns can be used to form both the time course and the topography of a component, if the topography is defined as amplitudes measured along a midline array down the center of the scalp.

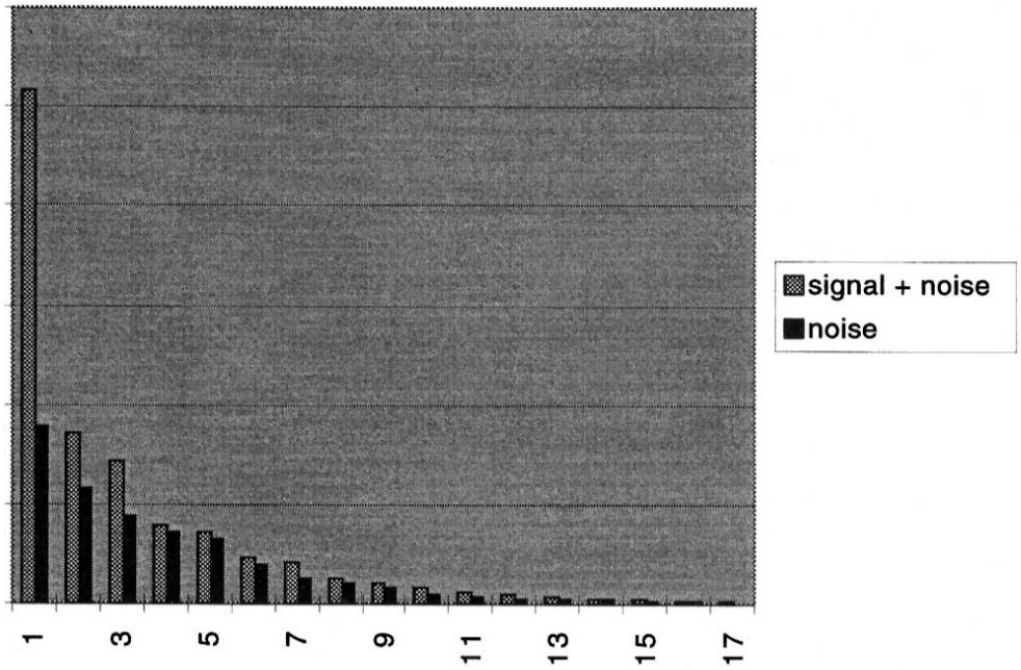
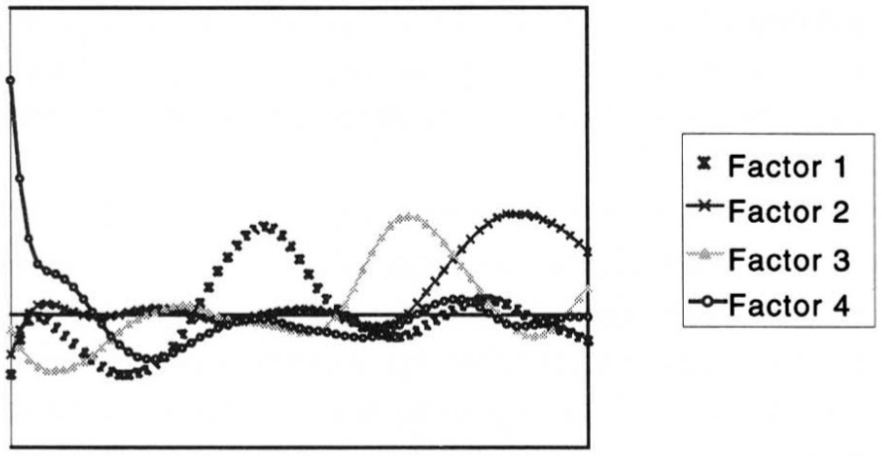
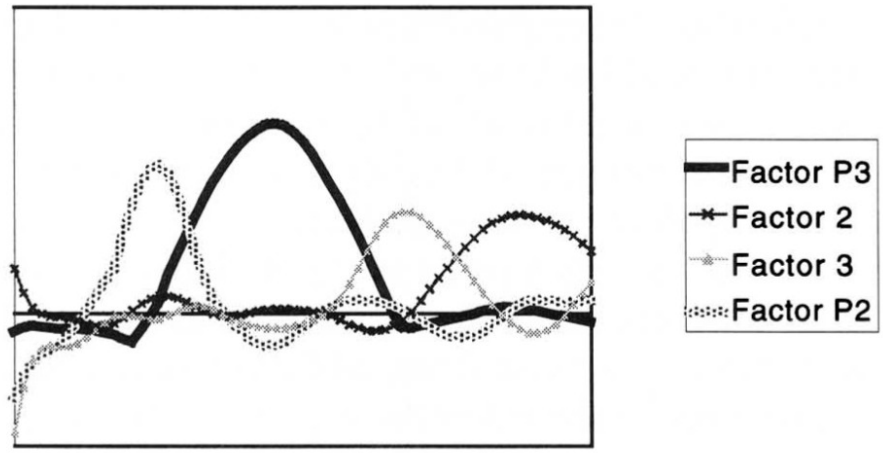


FIGURE 3. Parallel Test simulation. Chart shows eigenvalues of Factors 1 through 17 for both background EEG and base data matrix + background EEG. Although addition of signal has increased the eigenvalues for all the factors, it has done so most for the first three. Application of the scree criterion indicates retention of the first four factors (the three factors plus one more).

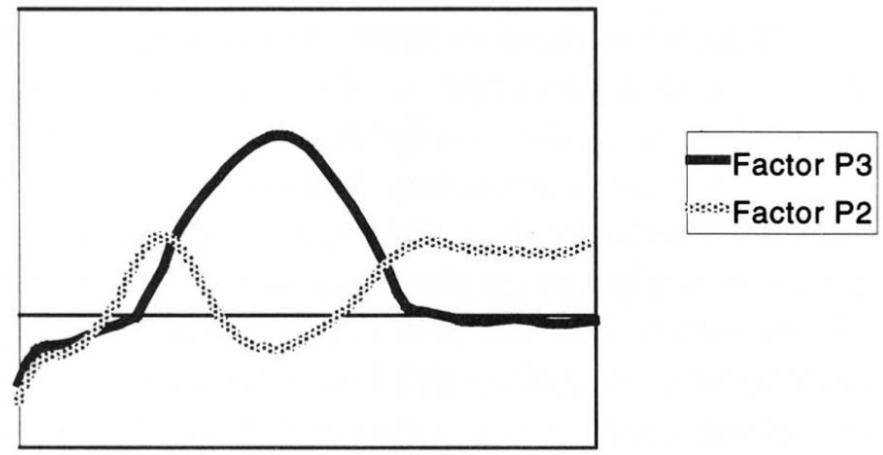
a)


Factor 1
Factor 2
Factor 3
Factor 4

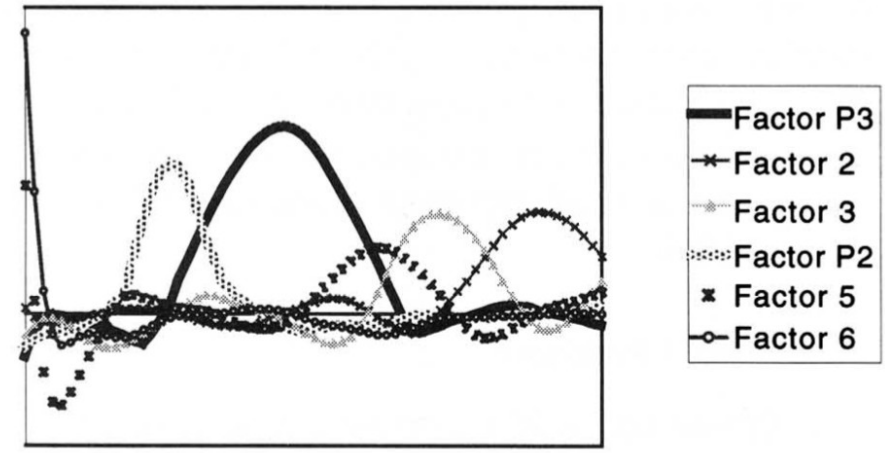
b)


Factor P3
Factor 2
Factor 3
Factor P2

c)


Factor P3
Factor P2

d)


Factor P3
Factor 2
Factor 3
Factor P2
Factor 5
Factor 6

A 0

FIGURE 4. Factor waveforms from parallel test simulation. Factor loadings were multiplied by the time point standard deviations to rescale them and then overplotted. The corresponding factors in the four different analyses were given the same line type to facilitate comparisons. a) Factors from PCA of background noise alone (4 factors retained to facilitate comparison with b). b) Factors from PCA of data matrix, retaining 4 factors as indicated by the parallel test. c) Retaining only two factors results in distorted characterization of P2 and P3. d) Retaining more factors than indicated by parallel test (6 factors) moderately improves reconstruction even more.

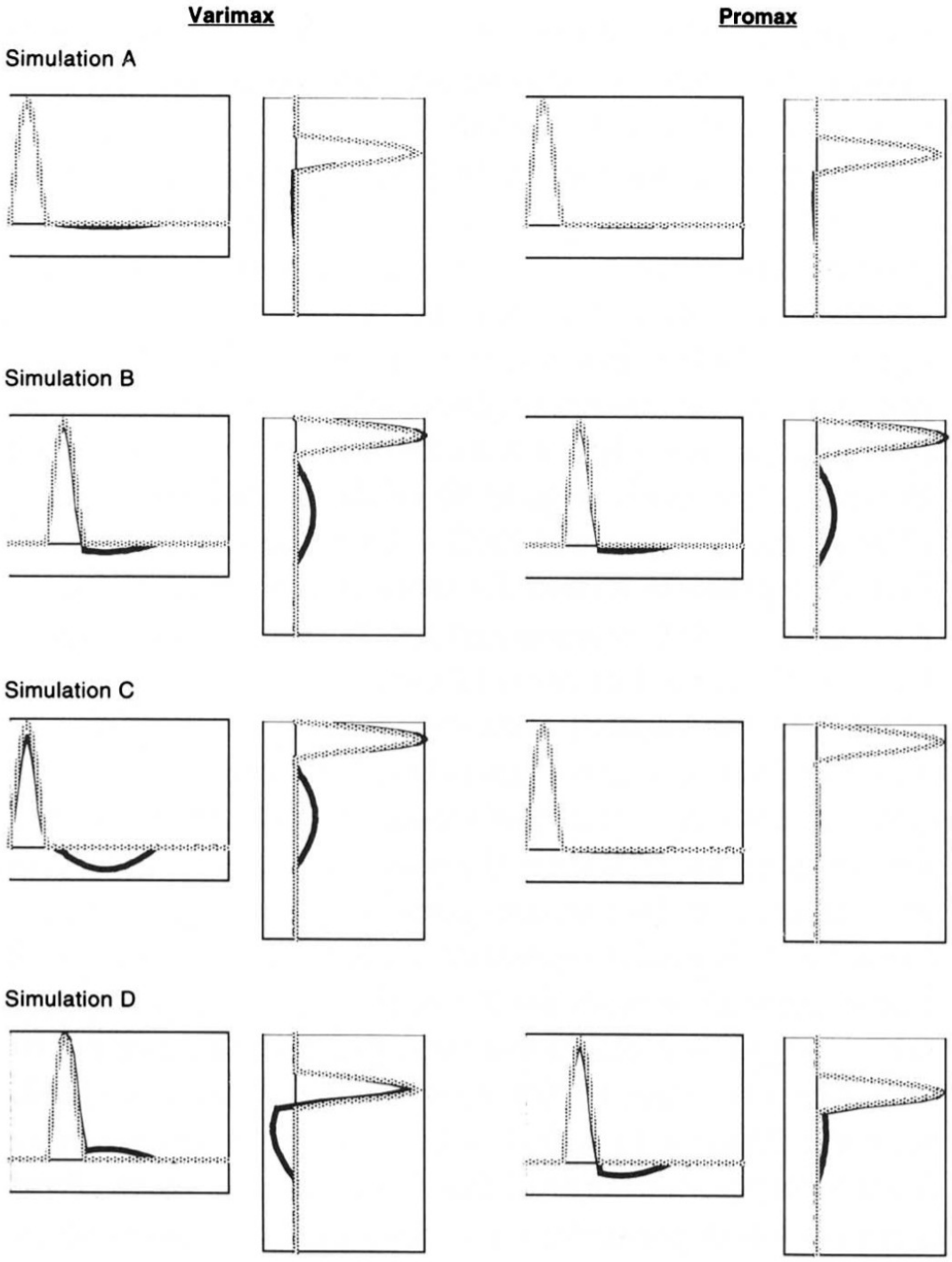


FIGURE 5. Results of promax simulation tests. The results for the P1/P2 factor only are illustrated. The horizontal figures represent the time course and the vertical figures represent the topography. Results are shown for both the varimax and promax rotations. The original pattern is shown is gray and the factor reconstruction is drawn in black, such that only deviations from the original pattern can be seen. A) No

temporal overlap or component correlation between the two components (P1 and P3) results in a good reconstruction of the P1. B) Both temporal overlap and component correlation between the two components results in distortions . C) Component correlation but no overlap results in distortion only for the varimax rotation. D) Temporal overlap but no component correlation results in distortion for both rotations.

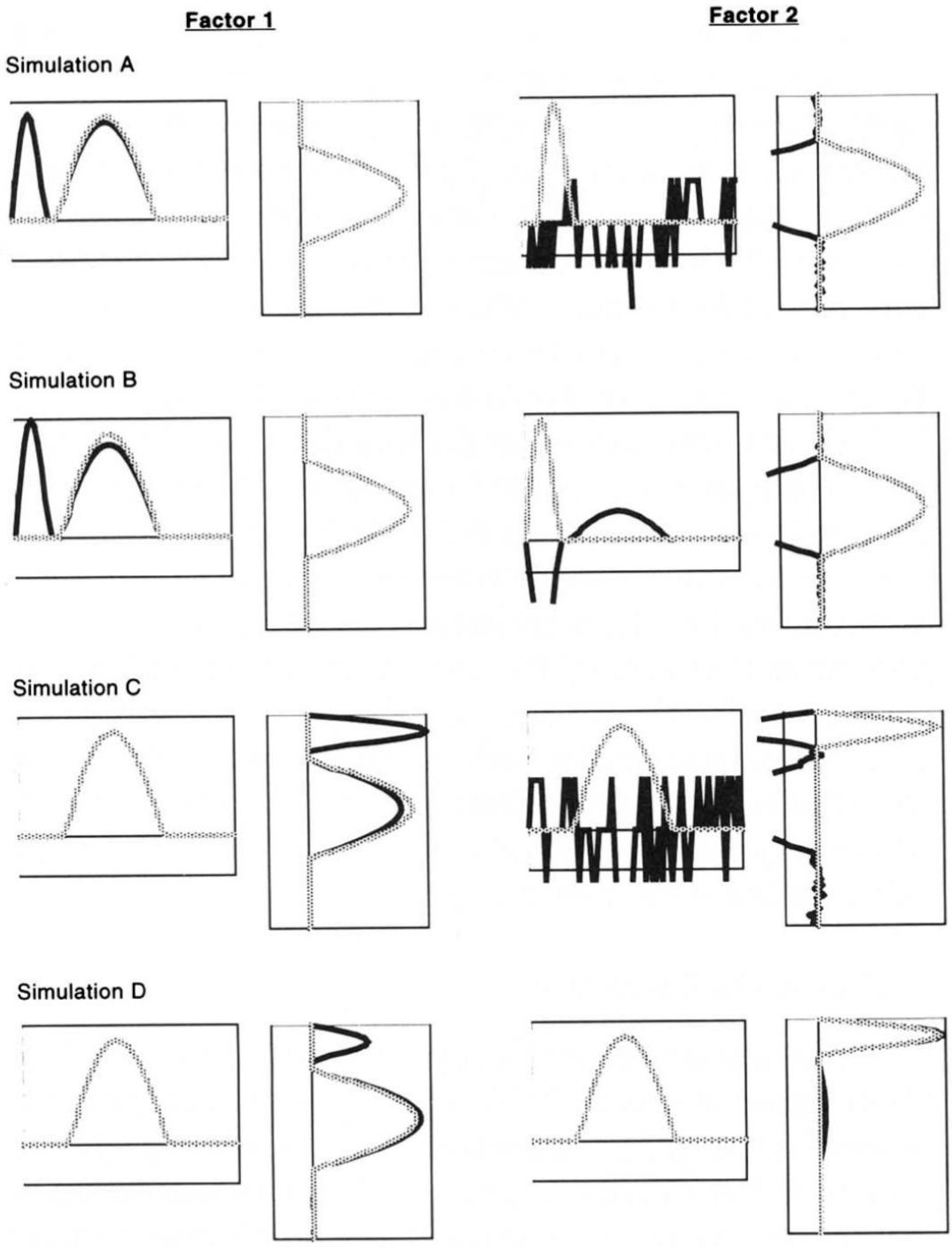


FIGURE 6. Results of spatial PCA simulation tests. The results for the promax rotation only are illustrated. The horizontal figures represent the time course and the vertical figures represent the topography. Results are shown for both Factor One and Factor Two. The original pattern is shown is gray and the factor reconstruction is drawn in black, such that only deviations from the original pattern can be

seen. A) Components have same spatial pattern. B) Components have same spatial pattern but differential condition effect. C) Components have same temporal pattern and differential condition effect. D) Spatial PCA in case of components with the same temporal pattern and differential condition effect.

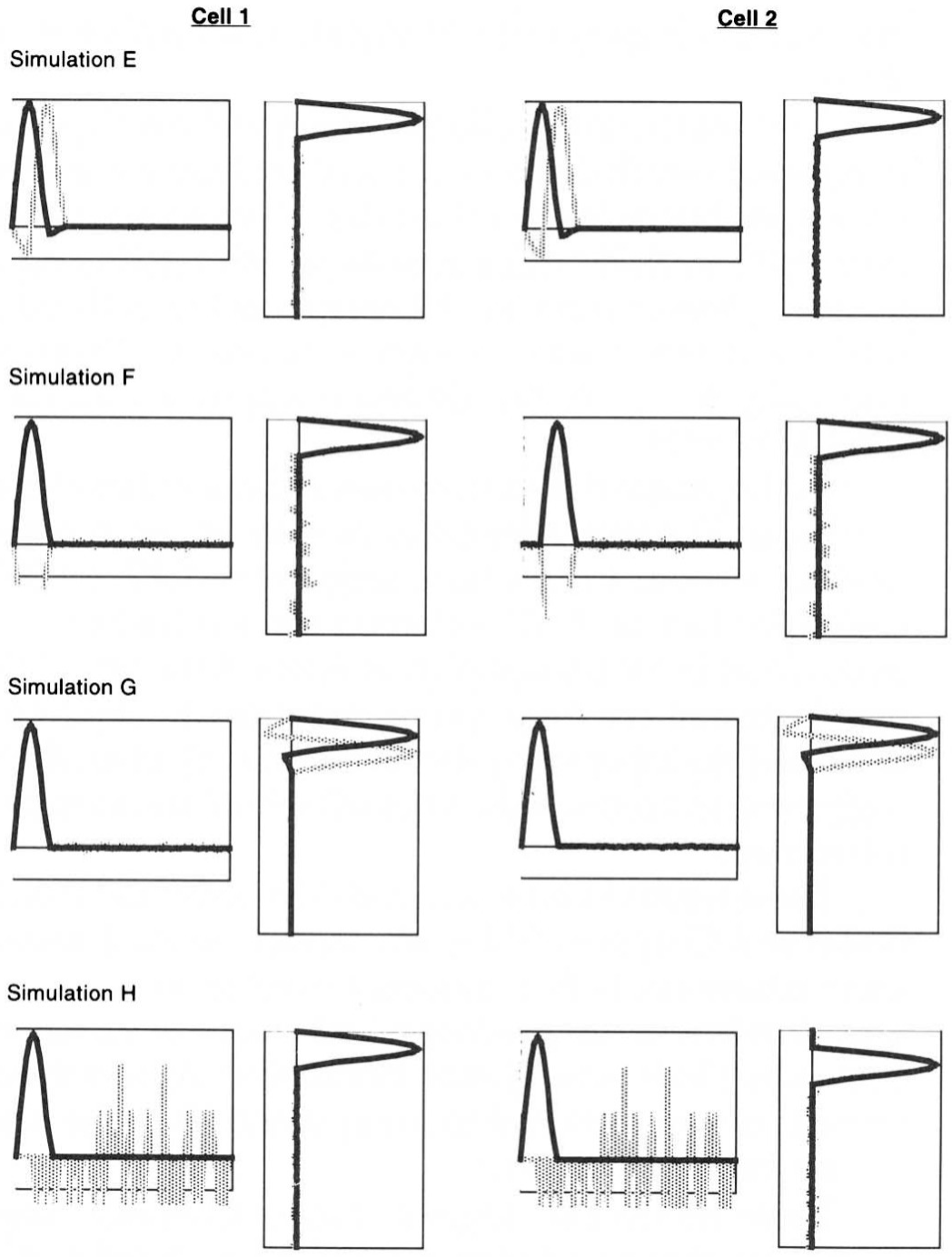


FIGURE 7. Results of jitter simulation tests. The results for the promax rotation only are illustrated. The horizontal figures represent the time course and the vertical figures represent the topography. Results

are shown for both Cell One and Cell Two. Factor One is shown is black and Factor Two is in gray. Waveforms indicate only the shape of the factors, not the amplitude. A) Temporal PCA of component with temporal jitter (different latencies in Cell 1 and Cell 2). Factor 1 describes the component in Cell 1 and Factor 2 describes the component in Cell 2. B) Spatial PCA of component with temporal jitter. Factor 1 describes the component in both cells while Factor 2 is essentially non-existent. C) Spatial PCA of component with spatial jitter (different topographies in Cell 1 and Cell 2). Factor 1 describes the component in Cell 1 and Factor 2 describes the component in Cell 2. D) Spatial PCA of component with temporal jitter. Factor 1 describes the component in both cells while Factor 2 reflects essentially only noise.

Appendix

A simple proof, that the portion of the recorded waveform accounted for by a factor can be computed by multiplying the factor loading by the factor score by the standard deviations of the variables, starts with the fact that the relation between two variables can be expressed as:

(1) $Z_Y = rZ_X$

where r is the correlation coefficient and Z_Y and Z_X are variables that have been standardized (p. 116, Glass & Hopkins, 1984).

Keeping in mind that a factor loading is the correlation between a variable and the factor (p. 599, Tabachnick & Fidell, 1989), this means that:

(2) $Z_t = rZ_S$

where Z_t is a given time point variable (standardized), r is the factor loading, and Z_S is the factor score (standardized). Since:

(3) $Z_t = (T - \mu) / \sigma$

where σ is the standard deviation of T and μ is the mean of T. Rearranged, this is:

(4) $(T - \mu) = \sigma Z_t$

substituting in (2) produces:

(5) $T - \mu = \sigma r Z_S$

or

(6) $T = \sigma r Z_S + \mu$

If one carried out this operation for each factor, added together the resulting latent waveforms, and then added μ to the result, one would reconstitute the original raw waveform (p. 15, Jackson, 1991). PCA is therefore literally the process of decomposing the observed waveforms into inferred latent waveforms in that the factor scores, factor loadings, variable standard deviations, and variable means together contain the full information necessary to regenerate the raw data. Note that the variable means (μ) represent variance that could not be associated with any factors and should therefore not be used when regenerating individual factor waveforms.