PRINCIPAL COMPONENT ANALYSIS OF EVENT-RELATED POTENTIALS: SIMULATION STUDIES DEMONSTRATE MISALLOCATION OF VARIANCE ACROSS COMPONENTS [1]

CHARLES C. WOOD and GREGORY McCARTHY

Neuropsychology Laboratory, Veterans Administration Medical Center, West Haven, CT 06516, and Departments of Neurology and Psychology, Yale University, New Haven, CT 06520 (U.S.A.)

Event-related potentials (ERPs) and other time series data pose difficult analytic problems because of inherent statistical dependencies between data values at different time points and because each such value is often the result of a number of overlapping influences. Principal component analysis (PCA), a multivariate statistical procedure closely related to factor analysis, has become a widely used technique for attempting to deal with these problems (for reviews, see Glaser and Ruchkin 1976; Donchin and Heffley 1978; Picton and Stuss 1980).

PCA represents the ERP voltages at successive time points as linear combinations of a new set of variables termed principal components (PCs):

$$V_{ij} = w_{i1}PC_{1j} + w_{i2}PC_{2j}$$
$$+ \dots w_{iT}PC_{Tj} \quad (i = 1, N; j = 1, T) \quad (1)$$

where V_{ij} is the ERP voltage at time j in wave form i; PC_{1j}, PC_{2j}, ... PC_{Tj} are PC 'loadings' representing the contribution of each PC to the voltage at each time point; and w_{i1}, w_{i2}, ... w_{iT} are PC 'scores' representing the contribution of each PC to each of the ERP wave forms. The PCs are orthogonal (and hence statistically independent), so that measures on one PC are guaranteed to provide independent information from measures on another. In addition, each successive PC accounts for a maximal proportion of the original variance uncorrelated with preceding PCs. Thus, when the original

variables are correlated, only a small number of PCs (many fewer than T) can account for a large proportion of the original variance. For ERP data, 4–8 PCs can often account for more than 80% of the total variance.

As typically employed in ERP experiments, PCA is part of a 3-step data analysis strategy. First, PCA is performed on the covariance, cross-product, or correlation matrix of the original ERP voltages. Second, the PCs are rotated using Varimax (Kaiser 1958) or related rotation criteria so that the wave shapes of the derived components correspond more closely to those of hypothesized ERP components. An additive model similar to Eq. 1 also applies following rotation, but with the PCs replaced by terms representing the rotated components. Third, scores on the rotated PCs are used to assess the statistical significance of experimental treatments in analyses of variance (ANOVAs) or other inferential statistics. Thus, PCA is used both as a technique for *component identification* (steps 1 and 2) and as a technique for *component measurement* (step 3). The hope is that the number and wave shapes of the rotated PCs accurately reflect the number and wave shapes of underlying ERP components and that the rotated PC scores accurately reflect variations in magnitude of those components across subjects, electrode locations and experimental conditions.

Although reservations about several aspects of the PCA-Varimax-ANOVA strategy have been expressed (e.g., Donchin and Heffley 1978; Hunt 1979; Wastell 1979, 1981a, b; Wood 1979; McCarthy 1980; Rosler and Manzey 1981), it has

played an increasingly important role in defining and measuring ERP components, particularly in cases of presumed component overlap (e.g., Donchin et al. 1975; Ruchkin et al. 1980; Friedman et al. 1981; Sutton and Ruchkin 1984). Because the true component structure of the ERP data typically analyzed by PCA is largely unknown, the validity of the PCA-Varimax-ANOVA strategy must be evaluated using simulated ERP data in which the component structure is known and can be systematically manipulated. In this paper we report initial simulation studies which demonstrate that PCA can incorrectly allocate variance across components, resulting in large increases in the probability of type I error in tests on PC scores.

Methods

Simulated ERP wave forms were constructed from the three 64-point prototype components shown in Fig. 1 (top). These components were not intended to simulate any particular set of empirical ERP results, but were designed to capture in a general way some of the main features of the ERP components discussed in recent experiments. An initial period of 16 points with no systematic ERP activity was followed by component 1 (12 points in duration), component 2 (26 points in duration), and component 3 (39 points in duration). Component 1 reached its maximum and returned to zero before components 2 and 3 began. The latter components began simultaneously, with component 2 reaching its maximum and returning to zero before component 3 reached its maximum at the end of the epoch. Components 1 and 2 were half-period sine waves and component 3 was a quarter-period sine wave of the durations given above. Each had a maximum amplitude of 1.0 and was tapered at the ends by a cosine function.

In each simulation, the 3 prototype components were combined linearly to form 800 ERP wave forms, corresponding to a $2 \times 2 \times 10$ factorial repeated-measures design with 20 subjects. Such a design might be used, for example, in an experiment comparing ERPs for 2 levels of stimulus intensity, 2 levels of stimulus probability, and 10 electrode locations, in each of 20 subjects. The 800

ERP wave forms were constructed in the following manner. For each component, random weights for each factor in the 3-way factorial design were chosen independently from normal distributions with specified mean and standard deviation. In the present simulations, only main effects of one experimental treatment were investigated; no 2- or 3-way interactions were studied. In simulations with no effect of experimental treatments, the distributions of random weights had means of 100 and standard deviations of 50, yielding composite weight distributions with means and standard deviations which averaged 300 (3×100) and 86.6 ($\sqrt{3 \times 50^2}$), respectively. Main effects were introduced on one of the 2-level treatments by increasing the mean of the random weight distribution for one level from 100 to 200, keeping the standard deviation constant at 50.

An additive noise term was chosen independently for each of the 64 time points from a normal distribution with zero mean and standard deviation 2. The noise term introduced no covariance between time points and typically accounted for less than 0.5% of the total variance of the simulated ERPs. Finally, the 3 prototype components were multiplied by the corresponding 3 sets of random weights and then summed together with the random noise to yield the composite ERP wave forms. These composite wave forms constituted the raw data for each simulation.

Autocorrelated noise was not used in the present simulations in order to investigate the PCA-Varimax-ANOVA strategy under conditions relatively favorable to its success. Some amount of autocorrelated noise is likely to be present in most ERP data despite signal averaging, and at adverse signal-to-noise ratios such noise can significantly influence the resulting component structure.

Twelve hundred such simulations were performed, each consisting of the generation of an independent set of 800 simulated ERP wave forms as described above, followed by computation of principal components, Varimax rotation and repeated-measures ANOVA. The first set of 400 simulations consisted of 100 with no main effect of any experimental treatment, and 100 each with a single main effect on components 1, 2 and 3, respectively. The second set of 400 investigated

systematic variations in the size of the experimental treatment effect, and the third set investigated systematic variations in the number of subjects employed in the $2 \times 2 \times 10$ design. ERP wave form simulations and data management were performed by special-purpose programs, PCAs and Varimax rotations were performed by programs which included subroutines from the EISPACK system (Smith et al. 1974), and ANOVAs were performed by BMDP2V (Dixon 1981).

Results

In order to yield an accurate reconstruction of the original simulated components, the results of the PCA-Varimax-ANOVA strategy must meet 2 criteria. First, the wave shapes of the rotated PC loadings must correspond closely to those of the original prototype components. Second, the rotated PC scores must be highly correlated with the original random weights used to generate the simulated ERP wave forms. Evidence bearing upon each of these criteria is examined below.

At an even more fundamental level, the number of PCs must correspond to the number of original prototype components (cf., Douglas and Rogers 1983). In the present simulations, only the first 3 PCs were rotated in order to maximize the likelihood of correspondence between the rotated PCs and prototype components. With the relatively simple component structure employed, the commonly employed eigenvalue-equals-one criterion (multiplied by the average variance of the original variables) also yielded 3 components.

Correspondence between rotated PC loadings and prototype components

Fig. 1 (center and bottom) presents the unrotated and rotated PC loadings from a representative example of the 100 simulations involving no experimental treatment effects. Note that the unrotated PCs had non-zero loadings over sizable regions of the sampling epoch and often contributed significantly to latency regions corresponding to more than one of the original prototypes. The rotated PCs, in contrast, corresponded much more closely to the wave shapes of the prototypes, hav-
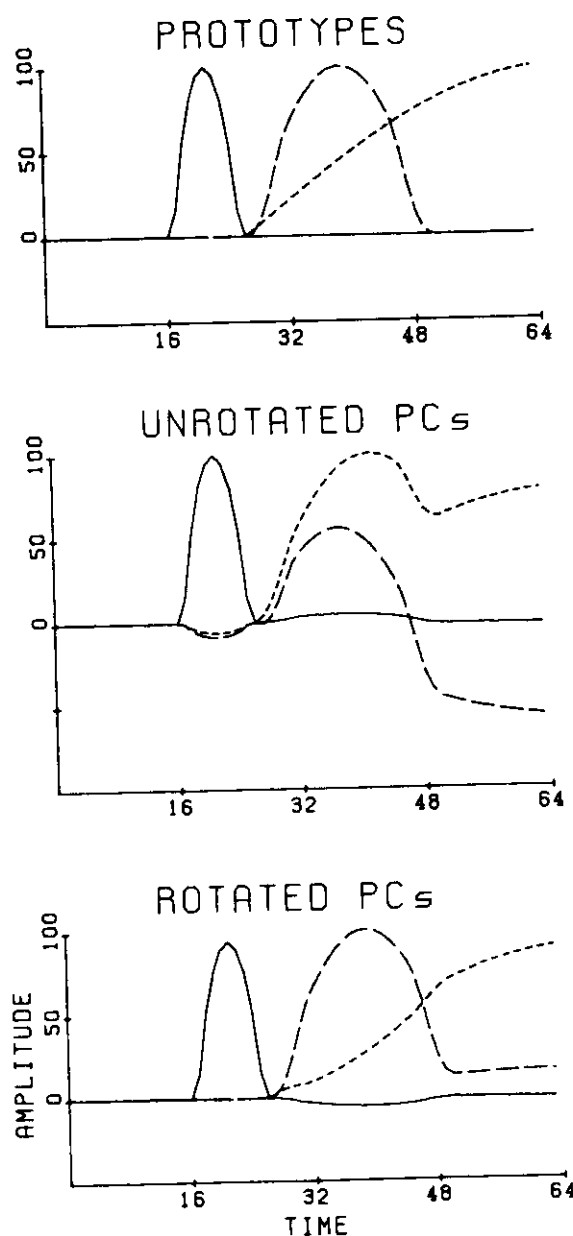

Fig. 1. Plots of prototype components (top), unrotated principal component loadings (center), and rotated principal component loadings (bottom) for a representative simulation having no experimental treatment effect.

ing highly similar onsets, maxima and offsets. The major differences between the rotated PCs and prototypes were: (a) that rotated PC 3 rose more slowly than prototype 3, being initially concave

TABLE I

Proportion of variance accounted for by prototype components and rotated PCs in a representative simulation having no treatment effects.

	Component 1	Component 2	Component 3
Prototypes	0.15	0.35	0.49
Rotated PCs	0.15	0.43	0.41

TABLE II

Product-moment correlations between rotated PC scores and original weights in a representative simulation having no treatment effects.

		Rotated PC scores		
		Compo-nent 1	Compo-nent 2	Compo-nent 3
Original weights	Component 1	0.99	0.02	0.02
	Component 2	0.02	0.98	0.20
	Component 3	0.02	0.20	0.98

upward rather than concave downward; and (b) that rotated PC 2 never returned fully to zero in the later portion of the epoch. As shown in Table I, these differences in wave shape were reflected in shifts in the proportion of variance accounted for by each component. Whereas component 2 accounted for 35% of the variance in the original prototypes, it accounted for 43% of the variance in the rotated PC solution. This 8% increase in variance attributed to component 2 by PCA was offset by a corresponding 8% decrease in variance attributed to component 3, which accounted for 49% of the original variance and 41% in the rotated PC solution. Component 1, which did not overlap in time with either of the other components, was virtually identical in shape and accounted for the same proportion of variance in both the original ERP wave forms and the rotated PC solutions.

Although the rotated PC loadings and the proportions of variance shown in Fig. 1 and Table I reflect only one of the 1200 total simulations performed, they are good representatives of the remainder. Only minor variations in wave shape and proportion of variance ($\pm 2\%$ per component) were observed across the other simulations. The data in Fig. 1 and Table I are also representative of the loadings and proportions of variance for simulations in which effects of experimental treatments were introduced, the only differences being the relative increase in variance for the component affected by the experimental treatment.

Correspondence between rotated PC scores and original weights

In order to provide accurate measures of the original components: (a) the rotated PC scores for each component should be highly and selectively correlated with the original weights for the corre-

sponding component; and (b) the effects of experimental treatments should be correctly represented in ANOVAs on the rotated PC scores.

Table II presents correlations between each set of rotated PC scores and each set of original weights for the 3 components. Note first that the correlations between PC scores and weights for corresponding components (i.e., the diagonal elements in Table II) were uniformly high, indicating that the variations in each set of rotated PC scores over subjects and experimental conditions closely mirrored that of the original weights. The off-diagonal elements in Table II indicate low correlations between weights and scores for component 1 and those for each of the other two components. For components 2 and 3, however, there were correlations of 0.20 between rotated PC scores and original weights for the other (i.e., incorrect) component.

Although the misallocation of variance between components 2 and 3 may seem relatively small when expressed as a change in proportion of variance (8-10% of the total variance across the 1200 simulations) or as correlations between the rotated PC scores and the weights for the incorrect original component ($r = 0.20$), it had dramatic consequences on the ANOVAs used to assess the effects of experimental treatments. Table III compares the results of ANOVAs on the original random weights with ANOVAs on the rotated PC scores in the 100 simulations with no treatment effect and the 100 simulations each with treatment effects on components 1, 2 and 3, respectively. Each pair of entries in the table is the number of F ratios out of 100 in which the obtained P value was less than

TABLE III

Number of F ratios with $P < 0.05$ out of 100 ANOVAs on original weights and rotated PC scores.

| | | Component tested | | |
		Component 1 (weights/scores)	Component 2 (weights/scores)	Component 3 (weights/scores)
	No main effect	5/4	3/2	6/7
Locus of main effect	Component 1	100/100	5/4	5/6
	Component 2	3/9	100/100	2/71 *
	Component 3	5/5	4/81 *	100/100

* Inflated type I error in analyses on rotated PC scores.

0.05 for ANOVAs on the original weights and rotated PC scores (left and right values in each column, respectively). For cases in which there was no treatment main effect, the values given in Table II provide estimates of the actual type I error at the 5% level. Thus, when there was no main effect on any component (Table III, top row), the number of F ratios with $P < 0.05$ ranged from 2 to 7 out of 100, closely approximating the 5% type I error rate.

When the treatment effect was introduced on component 1, all 100 ANOVAs on the original weights for component 1 and all 100 ANOVAs on the rotated PC scores for component 1 were significant at the 5% level, and the ANOVAs on the weights and scores for components 2 and 3 approximated the nominal 5% type I error rate (Table III, second row). Thus, for component 1, ANOVAs on the rotated PC scores provided as accurate a test of experimental treatments as ANOVAs on the original weights.

However, when treatment effects were introduced on components 2 and 3, there were large increases in type I error in ANOVAs on rotated PC scores for the other overlapping component. With the treatment effect on component 2, all 100 ANOVAs on PC scores for component 2 were significant at $P < 0.05$, but 71 of the 100 ANOVAs on PC scores for component 3 were also significant at that level. Similarly, with the treatment effect on component 3 all 100 ANOVAs on component 3 scores and 81 ANOVAs on component 2 scores were significant at $P < 0.05$. Thus, PCA's incorrect allocation of variance between components 2 and 3 noted previously in analyses of

proportions of variance and PC score correlations produced type I error rates that were inflated from the nominal 5% level to actual levels of 70–80%.

A more complete illustration of the increased type I error for ANOVAs on component 3 when the treatment effect was on component 2 is presented in Fig. 2, which shows the complete P distributions for the 100 ANOVAs on the original weights and the rotated PC scores for component 3. In absence of any treatment effect, such distributions should be rectangular with approximately
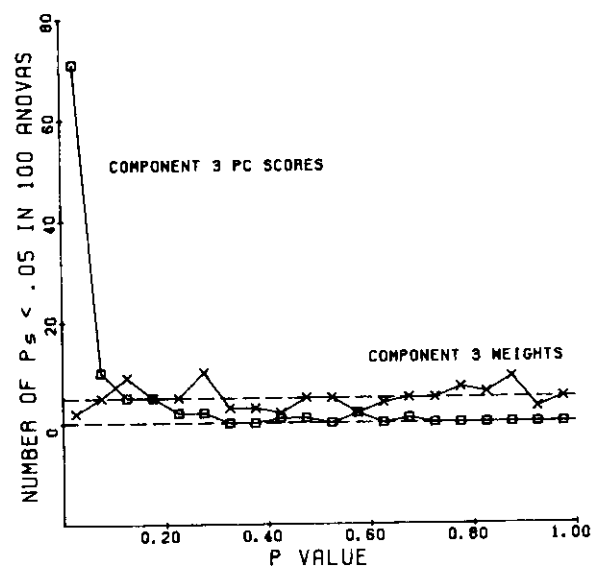

Fig. 2. Distributions of P (bin-width of 0.05) for ANOVAs on component 3 weights and rotated PC scores when the experimental treatment effect was on component 2. The leftmost point for the component 3 PC scores (71 ANOVAs with $P < 0.05$) corresponds to the value shown in Table III.

5% of the ANOVAs occurring at each value of P. The P distribution for component 3 weights closely approximated that pattern, whereas the entire distribution for the component 3 PC scores was shifted leftward toward smaller P values. Thus, the misallocation of variance influenced the entire P distribution in subsequent ANOVAs.

As an initial exploration of the dependence of the inflated type I error rates upon the size of the simulated experimental treatment effects and upon the number of subjects employed, 8 additional sets of 100 simulations with treatment effects on component 2 were performed, 4 of which manipulated size of the treatment effect and 4 of which manipulated number of subjects. Treatment effect size was reduced from 100 to 75, 50, 25 and 10, and N was reduced from 20 to 16, 12, 8 and 4, respectively.

The results of the treatment effect manipulations are shown in Fig. 3, in which the obtained numbers of F ratios with $P < 0.05$ are shown for ANOVAs on original weights and rotated PC scores for components 2 and 3. Stated in terms of the strength of association measure ω^2 (e.g., Hays

1963), these effect sizes ranged from less than 0.01 to a maximum of 0.23. Each point in the figure is based on 100 sets of simulated data analyzed using the PCA-Varimax-ANOVA strategy in the same manner described above. Since the treatment effect was always introduced on component 2, the values for component 2 in the figure represent the probability of correctly rejecting the null hypothesis at the 0.05 level for effects of different sizes. The values for component 3 represent the probability of type I error in component 3 ANOVAs as a function of effect size on component 2.

In analyses on the original weights (dotted lines), the number of correct rejections in component 2 ANOVAs increased sharply as effect size increased from zero, reaching 95% at an effect size of one-third to one-half a standard deviation. ANOVAs on component 3 weights approximated the 5% nominal type I error rate regardless of effect size on component 2. In analyses on the rotated PC scores (solid lines), the number of correct rejections in component 2 ANOVAs closely paralleled those for ANOVAs on the original weights, indicating that the rotated PC scores for component 2 were as sensitive to treatment effects on that com-
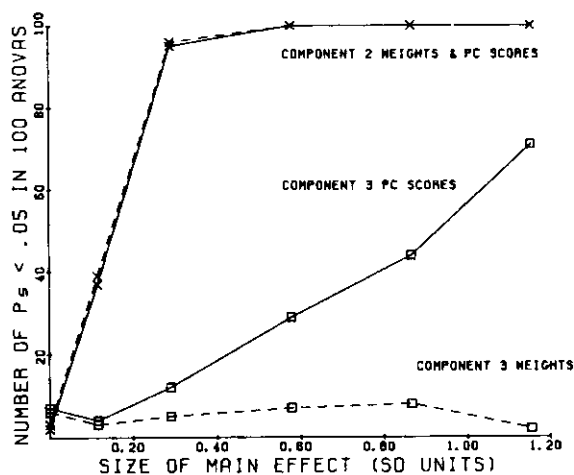

Fig. 3. Effects of manipulating the size of the treatment effect introduced for component 2 on correct rejection rates for ANOVAs on component 2 and for type 1 error rates for ANOVAs on component 3. Number of Ps < 0.05 are plotted against the size of the treatment effect in standard deviation units of the composite random weight distribution for component 2. The largest and smallest treatment effects (1.15 and 0.0 standard deviation units, respectively) correspond to the treatment sizes of 100 and 0 shown in Table III.
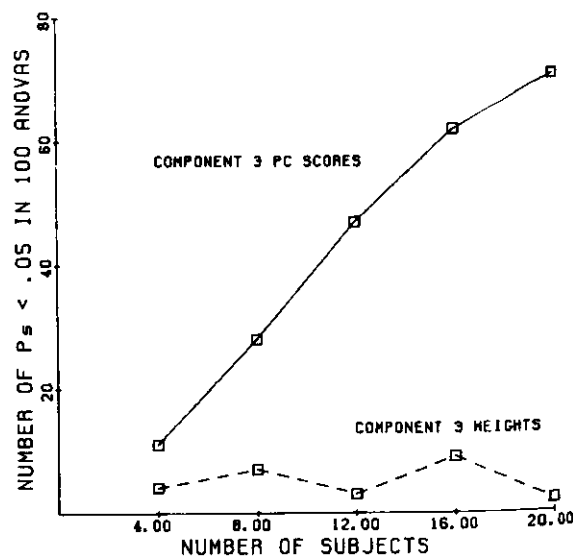

Fig. 4. Effects of manipulating N on type 1 error rates for ANOVAs on component 3 when the treatment effect was introduced for component 2. The points for N = 20 correspond to the data shown in Fig. 2 and Table III.

ponent as were the original weights. The type I errors for ANOVAs on rotated PC scores for component 3 began to increase above the nominal 5% level at effect sizes of approximately one-third a standard deviation, reaching 71% at the largest effect size studied.

The results of manipulating N are shown in Fig. 4, which compares type I error rates for ANOVAs on the original random weights and the rotated PC scores for component 3. For ANOVAs on the random weights, type I error closely approximated the 5% level and was essentially independent of N over the range studied. In contrast, for ANOVAs on rotated PC scores, type I error was proportional to N, ranging from a minimum value of 11% with N = 4 to 71% with N = 20.

To investigate whether the misallocation problem is due exclusively to the differences between the rotated PC loadings and the original prototypes or due in part to the computation of rotated PC scores, we calculated for each of the simulations described above another set of component scores derived from the original prototypes instead ˙he rotated PC loadings. We shall refer to these 'prototype scores' to distinguish them both from the original random weights and the rotated PC scores. Unlike the random weights (which were stipulated by the simulations), the prototype scores were computed from the composite ERP wave forms in a manner similar to the rotated PC scores. The only difference was that the point-by-point standard deviations of the weighted prototypes were used as estimates of the component wave shapes instead of the rotated PC loadings. In other words, the weighted prototype wave forms were substituted for the rotated PC loadings, and the computations of component scores and ANOVAs were subsequently performed in an identical manner to the original simulations. Thus, this procedure asks whether the original prototype components can be accurately measured in the simulated ERP wave forms if the estimates of the component wave shapes are more accurate than those provided by the rotated PC loadings.

Fig. 5 (top) compares the wave shapes of the rotated PC loadings from the condition with no treatment effect (Fig. 1, top) with those of the weighted prototypes as defined in the preceding
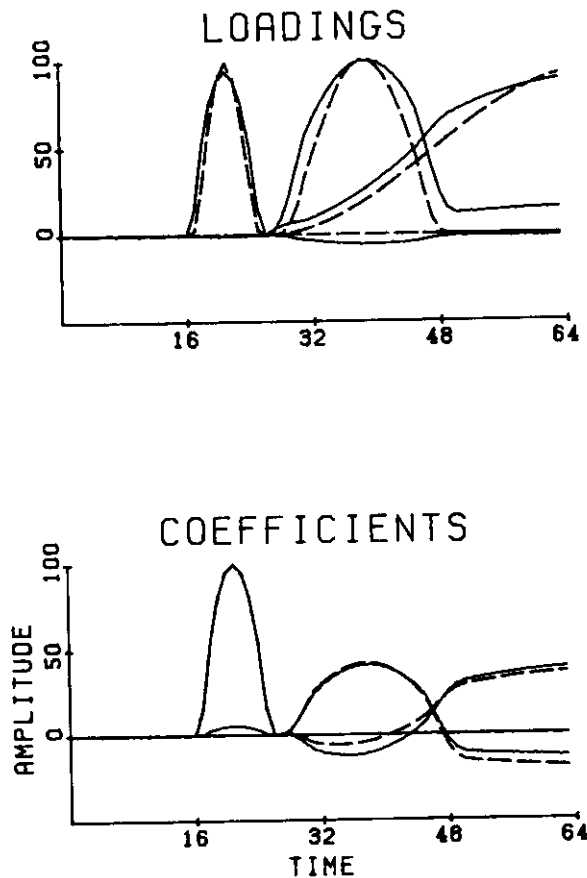

Fig. 5. Top: comparison of rotated PC loadings (solid) with weighted prototypes (dashed) calculated from the standard deviations of the voltages at each time point. Bottom: comparison of component score coefficients derived from the rotated PC scores (solid) and the weighted prototypes (dashed) shown in the top of the figure.

paragraph. Two main differences should be noted: (a) the weighted prototype for component 2 is narrower and returns to zero before the end of the epoch; and (b) the shape of the rising portion of rotated component 3 is distorted relative to the corresponding weighted prototype. Fig. 5 (bottom) plots the score coefficients corresponding to the weighted prototypes and rotated PC scores in Fig. 5 (top). These coefficients are multiplied, point-by-point, by the ERP wave forms and summed to yield the component scores for each set of components. Thus, they constitute filters through which the ERP data are passed in order to obtain mea-

surements of each component. Note that the major differences between coefficients for the prototype scores and rotated PCs again lie in portions of the sampling epoch where the other component is maximal.

The use of the prototype scores instead of the rotated PC scores resulted in a marked improvement in accuracy of estimating the original random weights. In correlations between the prototype scores and random weights analogous to those for the rotated PC scores shown in Table II, the on-diagonal correlations between corresponding components were all in excess of 0.9999, and the off-diagonal (i.e., erroneous) correlations between components 2 and 3 decreased to less than 0.04. The ANOVAs on the prototype scores showed no increase in type I error between components 2 and 3 when there were treatment effects on the other component. When plotted as in Fig. 3, the number of significant ANOVAs on the prototype scores were identical, point-by-point, to the ANOVAs on the original weights. Thus, it is the inaccurate estimation of component wave shape by the rotated PC loadings, not any inherent difficulty in the calculation of component scores, which is responsible for the misallocation of variance and for the consequent increase in type I errors in ANOVAs on rotated PC scores.

Discussion

The simulations reported here demonstrate that the PCA-Varimax-ANOVA strategy can, under conditions similar to those of empirical ERP experiments, incorrectly allocate variance across components, resulting in serious misinterpretation of the effects of experimental treatments. This problem requires careful consideration in evaluating experimental reports involving the PCA-Varimax-ANOVA strategy.

The need for systematic simulation studies of the PCA-Varimax-ANOVA strategy has been recognized for some time, and limited simulation studies have been reported by us and by others (e.g., Wood 1979; McCarthy 1980; Rosler and Manzey 1981). Although the scope of the present simulations far exceeds that of any available in the

literature, it is nevertheless limited. Only one set of simulated components, one experimental design, one type of association matrix (variance-covariance) and significant effects on one experimental treatment were studied. We therefore believe that it would be unwise to overgeneralize these results or to view them as invalidating any particular set of empirical conclusions based on the PCA-Varimax-ANOVA strategy. Additional simulation studies are needed to evaluate the magnitude of the variance misallocation problem across systematic variations in the number, shape, overlap and variance of simulated components, and in the experimental design, simulated treatment effects and specific form of PCA employed.

Although detailed conclusions about the nature of the variance misallocation problem must await the results of such studies, a number of tentative suggestions can be made at this point:

(1) The components for the present simulations were designed so that there was considerable temporal overlap between components 2 and 3 and no overlap between component 1 and either of the other two. The obtained variance misallocation was limited to components 2 and 3, and component 1 was well reconstructed both in terms of the wave shape of rotated PC loadings and the correlations and ANOVAs on rotated PC scores. Thus, the amount and form of the overlap between components may be expected to play an important role in determining the magnitude of variance misallocation.

(2) The present simulations were based on PCAs of variance-covariance matrices because they are probably the most common type of analysis used for ERP data. However, PCAs performed on correlation and cross-product matrices in a few selected examples from those presented above demonstrated variance misallocation of roughly the same magnitude. None of the 3 main forms of association matrix used for PCAs appears to be immune from the variance misallocation problem.

(3) The relationships between type I error and the manipulations of treatment effect size and N shown in Figs. 3 and 4 indicate that the magnitude of the type I error covaries with the sensitivity (i.e., power) of the ANOVAs for detecting true treatment effects in the data. That is, larger treatment

effects and larger N's produced both an increased likelihood of detecting true treatment effects and an increased likelihood of type I errors for tests on components whose variance has been misallocated by PCA. Decreasing the alpha level would decrease both the likelihood of type I error and the likelihood of correct rejections.

(4) The type I error rate rose more slowly with increasing treatment effect size than did the correct rejection rate (Fig. 3). Hence, at least for this set of simulations, there was a range of treatment effect sizes for which sensitivity to true effects was relatively high and type I error rate was relatively low (less than twice the nominal alpha level). Whether such regions of acceptable sensitivity to true effects and relative insensitivity to type I error can be generalized to other components, designs and patterns of treatment effect remains to be determined.

(5) The increases in type I error demonstrated by the present simulations occurred in ANOVAs on components with no treatment effect when there was a significant effect present on an overlapping component. Hunt (1979) and Wastell (1981a) have suggested that ANOVAs on PC scores are biased toward exaggerating the statistical significance of true treatment effects because variance due to experimental treatments is used to define the dimensions of measurement (i.e., the component axes). Although the present simulations were not designed to address this issue, the manipulations of treatment effect size in Fig. 3 show little evidence of such exaggeration. The correct rejection rates for ANOVAs on component 2 PC scores were comparable to those for component 2 weights for small treatment effects.

(6) The direction of the spurious mean differences in ANOVAs on rotated PC scores varied as a function of the locus of the true treatment effect. When a treatment effect was introduced on component 2, then the spuriously significant mean differences for component 3 were in the opposite direction to those on component 2. However, when a treatment effect was introduced on component 3, then the mean differences between treatment levels were in the same direction for components 2 and 3. Thus, no generalizations should be drawn concerning the relative directions of correctly detected

and spurious treatment effects. The sizes of the spuriously significant treatment effects tended to be smaller than those of the correctly detected effects, with the mean differences between treatment levels for the latter roughly 20–25% those of the former.

(7) Perhaps the most disturbing aspect of the present result is the fact that type I error rates as high as 81% were obtained with misallocation of as little as 8–10% of the original variance, and with rotated PCs whose wave shapes corresponded as closely to those of the prototypes as do those in Fig. 1. In preparation for these simulations, we investigated a number of other sets of components, some of which yielded considerably poorer solutions than those in Fig. 1 in terms of number and wave shape of derived components. Although this paper has focused upon the existence and consequences of variance misallocation in a case in which the PCA solution appears intuitively good, readers should not overlook the likelihood of considerably poorer solutions with other sets of components.

The fact that the principal (i.e., unrotated) components of a given data set need not, of mathematical necessity, accurately reconstruct the component structure of the data has been emphasized in introductory treatments of PCA (e.g., Harris 1975, pp. 162–163) as well as in the specific context of PCAs of electrophysiological data (Van Rotterdam 1970; Freeman 1979). Although PCs have the important properties of being mathematically unique, orthogonal and accounting for a maximal proportion of the original variance, they need not correspond to the components which generate the data in question. This point is clearly illustrated by the unrotated PC loadings in the present simulations (Fig. 1, center), which account for over 99% of the original variance, but which contribute significantly to segments of the sampling epoch associated with more than one prototype. Such a result is common for unrotated PCs of ERPs and other time series data and argues against the recommendation of Rosler and Manzey (1981) that unrotated PCs should be preferred to rotated PCs in ERP analyses.

The use of Varimax and other rotation criteria may be viewed as an attempt to trade the mathe-

matical uniqueness and orthogonality of the initial PC solution for an increased likelihood of approximating the wave shapes of the assumed underlying components. Since the component structure of ERPs typically analyzed by the PC-Varimax-ANOVA strategy is unknown, investigators have relied upon intuitive criteria, primarily the similarity of rotated PC loadings to the wave shapes of the presumed underlying ERP components as evidence for the validity of the strategy. Although rotated PC loadings appear to correspond more closely to the hypothesized wave shapes of many ERP components, there is again no mathematical necessity that the rotated PCs accurately correspond to the true component structure of the data.

We have thus far considered the problem of misallocation of variance exclusively from the perspective of PCA. However, it should be emphasized that unless the ERP voltage at a given time point is due entirely to a single component (a situation which, although possible, appears increasingly unlikely for many ERP data of interest, see McCallum and Curry 1979; Picton and Stuss 1980; Wood et al. 1984), then the problem of correctly allocating ERP variance to overlapping components must be faced by any technique for component identification and measurement. 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. Misallocation of variance and misinterpretation of experimental effects are just as possible using such techniques as they are with PCA.

Although PCA is capable of making misallocation errors of the sort demonstrated here, it does have two advantages not shared by most other ERP analysis techniques. First, PCA explicitly acknowledges the possibility of component overlap and provides a quantitative means for representing it. Second, PCA's representation of ERP data as linear combinations of a set of underlying components is consistent both with the principle of superposition which governs potential fields in volume conductors and with most investigators' pretheoretical ideas about the composition of ERP

data. These observations suggest that improvements over PCA might be achieved by retaining the fundamental linear model inherent in PCA but substituting other estimates of the underlying component wave forms, based either on theory (cf., Freeman 1975, 1979) or alternative analytic techniques, for the unrotated or rotated loadings that PCA provides by solving the eigenvalue problem.

In conclusion, we believe that the present results demonstrate: (a) that the use of rotated or unrotated PC loadings derived from PCA as estimates of underlying ERP components can result in misallocation of variance across components; (b) that the use of PC scores to assess the effects of experimental treatments can result in serious misinterpretation of treatment effects; and (c) that such a possibility demands caution in relying heavily or exclusively upon the PCA-Varimax-ANOVA strategy in analyzing ERP data. Comprehensive simulation studies are needed in which the variance misallocation problem is investigated across systematic variations in the simulated components, experimental designs, treatment effects and forms of PCA employed. Such simulations will not only allow the extent of the misallocation problem to be assessed across the range of data likely to be encountered in ERP experiments, but will also provide test cases for assessing alternative techniques for component identification and measurement.

Summary

Simulated event-related potential (ERP) components were used to investigate the ability of principal component analysis (PCA), Varimax rotation and univariate analysis of variance (ANOVA) to reconstruct component wave shapes, to allocate variance correctly across components, and to identify the correct locus of simulated experimental treatments. The simulated ERPs consisted of 800 randomly weighted combinations of three 64-point components, corresponding to a $2 \times 2 \times 10$ repeated-measures design with 20 subjects. Covariance PCAs, Varimax rotations and univariate ANOVAs were performed on each of 400 such simulations, 100 with no effect of any

experimental treatment and 100 each with main effects on each of the 3 components. Eight hundred additional simulations were performed to investigate the effects of systematic variations in the size of the experimental treatments and the number of subjects per experiment.

The wave shapes of the simulated components were reconstructed reasonably well, although not completely, by the rotated principal component (PC) loadings. However, comparison of rotated PC scores with the random weights used to generate the simulated ERPs indicated that PCA incorrectly allocated variance across overlapping components, producing dramatic increases in type I error (the largest in excess of 80%) for ANOVAs on one component when the true treatment effect was on another. Although these results should not be overgeneralized, they clearly demonstrate that the PCA-Varimax-ANOVA strategy can incorrectly distribute variance across components, resulting in serious misinterpretation of treatment effects. Additional simulation studies are needed to determine the generality of the variance misallocation problem; pending the outcome of such studies, results obtained with the PCA-Varimax-ANOVA strategy should be interpreted cautiously.

Résumé

Analyse en composantes principales des potentiels liés à l'événement; des études par simulation révèlent que des erreurs peuvent être commises dans l'évaluation de la variance entre composantes

Les composantes simulées de potentiels liés à l'événement (PLE) ont été utilisées pour étudier comment l'analyse en composantes principales (ACP), la rotation Varimax et l'analyse univariée de variance (ANOVA), permettent de reconstruire les formes des composantes, d'attribuer correctement la variance entre composantes et d'identifier le niveau exact des traitements expérimentaux simulés. Les PLE simulés ont consisté en 800 combinaisons aléatoirement pondérées de 3 composantes de 64 points, correspondants à un schéma de $2 \times 2 \times 10$ mesures répétées sur 20 sujets. Des ACP de covariance, des rotations Varimax et des ANOVAs univariées ont été effectuées chaque fois sur 400 simulations de ce type, 100 sans effet d'aucun traitement expérimental et 100 avec pour chacune des effets importants sur chacune des 3 composantes. 800 simulations supplémentaires ont été pratiquées pour rechercher les effets de variations systématiques de la dimension des traitements expérimentaux et du nombre de sujets par expérience.

Les formes d'ondes des composantes simulées ont été correctement reconstituées, bien qu'incomplètement, par pondération de la composante principale (CP) après rotation. Toutefois, la comparaison des performances de la CP après rotation avec des pondérations aléatoires utilisées pour la genèse des PLE simulés, indique que l'ACP attribue incorrectement la variance entre composantes qui se chevauchent, et entraîne ainsi d'importantes augmentations d'erreurs de type I (la plus grande de 80% par excès) dans des ANOVAs sur une composante, alors que le véritable effet du traitement était d'affecter une autre composante. Bien que ces résultats ne doivent pas être généralisés exagérément, ils démontrent clairement que la stratégie ACP-Varimax-ANOVA peut distribuer incorrectement la variance entre les composantes, ceci résultant en une interprétation gravement faussée de l'effet des traitements. Des études complémentaires avec simulation seront nécessaires pour déterminer la généralité de ce problème d'attributions erronées de la variance; en fonction des résultats de telles études, les résultats obtenus avec la stratégie CP-Varimax-ANOVA devront être interprétés avec précaution.

References

Dixon, W.J. (Ed.) BMDP Statistical Software 1981. University of California Press, Berkeley, CA, 1981.

Donchin, E. and Heffley, E.F. Multivariate analysis of event-related potential data: a tutorial review. In: D.A. Otto (Ed.), Multidisciplinary Perspectives in Event-Related Brain Potential Research. U.S. Environmental Protection Agency, Washington, DC, 1978: 555–572.

Donchin, E., Tueting, P., Ritter, W., Kutas, M. and Heffley, E.F. On the independence of CNV and the P300 components of the human averaged evoked potential. Electroenceph. clin. Neurophysiol., 1975, 38: 449–461.

Douglas, R.R. and Rogers, L.J. Comparative factor analysis models for an empirical study of EEG data. Int. J. Neurosci., 1983, 18: 211–226.

Freeman, W.J. Mass Action in the Nervous System. Academic Press, New York, 1975.

Freeman, W.J. Measurement of cortical evoked potentials by decomposition of their wave forms. J. Cybernet. Inf. Sci., 1979, 2: 44–56.

Friedman, D., Vaughan, H.G. and Erlenmeyer-Kimling, L. Multiple late positive potentials in two visual discrimination tasks. Psychophysiology, 1981, 18: 635–649.

Glaser, E.M. and Ruchkin, D.S. Principles of Neurobiological Signal Analysis. Academic Press, New York, 1976.

Harris, R.J. A Primer of Multivariate Statistics. Academic Press, New York, 1975.

Hays, W.L. Statistics for Psychologists. Holt, Rinehart and Winston, New York, 1963.

Hunt, E. The validity of assumptions underlying the use of PCA and ANOVA in ERP research. Paper presented to the Workshop on Analysis of Event-Related Potentials. Vancouver, B.C., 1979.

Kaiser, H.F. The varimax criterion for analytic rotation in factor analysis. Psychometrika, 1958, 23: 187–200.

McCallum, W.C. and Curry, S.H. Hemisphere differences in event related potentials and CNVs associated with monaural stimuli and lateralized motor responses. In: D. Lehmann and E. Callaway (Eds.), Human Evoked Potentials: Applications and Problems. Plenum, New York, 1979: 235–250.

McCarthy, G. Simulation studies of PCA and ANOVA in the analysis of ERP data. Paper presented to the Workshop on Analysis of Event-Related Potentials, Washington, DC, 1980.

Picton, T.W. and Stuss, D.T. The component structure of human event-related potentials. In: H.H. Kornhuber and L.

Deecke (Eds.), Motivation, Motor and Sensory Processes of the Brain: Electrical Potentials. Behaviour and Clinical Use. Progress in Brain Research. Vol. 54. Elsevier/North-Holland, Amsterdam, 1980: pp. 17–49.

Rosler, F. and Manzey, D. Principal components and varimax rotated components in event-related potential research: some remarks on their interpretation. Biol. Psychol., 1981, 13: 3–26.

Ruchkin, D.S., Sutton, S. and Stega, M. Emitted P300 and slow wave event-related potentials in guessing and detection tasks. Electroenceph. clin. Neurophysiol., 1980, 49: 1–14.

Smith, B.T., Boyle, J.M., Dongarra, J.J., Garbow, B.S., Ikebe, Y., Klema, V.C. and Moler, C.B. Matrix Eigensystem Routines — EISPACK Guide; Lecture Notes in Computer Science, Vol. 6. Springer, New York, 1974.

Sutton, S. and Ruchkin, D.S. The late positive complex — advances and new problems. Ann. N.Y. Acad. Sci., 1984, in press.

Van Rotterdam, A. Limitations and difficulties in signal processing by means of the principal-components analysis. IEEE Trans. biomed. Engng, 1970. BME-17: 268–269.

Wastell, D.G. On the independence of P300 and the CNV: a short critique of the principal components analysis of Donchin et al. (1975). Biol. Psychol., 1979, 9: 171–176.

Wastell, D.G. PCA and varimax rotation: some comments on Rosler and Manzey. Biol. Psychol., 1981a, 13: 27–29.

Wastell, D.G. On the correlated nature of evoked brain activity: biophysical and statistical considerations. Biol. Psychol., 1981b, 13: 51–69.

Wood, C.C. Interpretation of principal components. Paper presented to the Workshop on Analysis of Event-Related Potentials, Vancouver, B.C., 1979.

Wood, C.C., McCarthy, G., Squires, N.K., Vaughan, H.G., Woods, D.L. and McCallum, W.C. Anatomical and physiological substrates of event-related potentials: two case studies. Ann. N.Y. Acad. Sci., 1984, in press.