Behavior Research Methods, Instruments, & Computers
1998, 30 (1), 87-102

DATA ANALYSIS AND INTERPRETATION

Computational and statistical methods for analyzing event-related potential data

GEERT J. M. VAN BOXTEL
Tilburg University, Tilburg, The Netherlands
and University of Amsterdam, Amsterdam, The Netherlands

Some computational and statistical techniques that can be used in the analysis of event-related potential (ERP) data are demonstrated. The techniques are fairly elementary but go one step further than do simple area measurement or peak picking, which are most often used in ERP analysis. Both amplitude and latency measurement techniques are considered. Principal components analysis (PCA) and methods for electromyographic onset determination are presented in detail, and Woody filtering is discussed briefly. The techniques are introduced in a nontechnical, tutorial review style. One and the same existing data set is presented, to which the techniques are applied, and practical guidelines for their use are given. The methods are demonstrated against a background of theoretical notions that are related to the definition of ERP components.

The electroencephalogram (EEG) is the record of the potential differences between electrodes placed on the (human) scalp. EEG records can be classified into two categories: the spontaneous or background EEG and event-related potentials (ERPs). The background EEG consists of periodic voltage fluctuations, which can be classified according to their frequency content—for instance, alpha (8–12 Hz), beta (13–30 Hz), gamma (31–50 Hz), delta (0.5–4 Hz), and theta (5–7 Hz). ERPs are discrete waveforms that are associated with an event. They are hidden within the background EEG and are visualized by averaging discrete EEG epochs that are synchronized to meaningful task events, such as the presentation of a stimulus or the occurrence of a response. Since the background EEG is not assumed to have a temporal relationship to the synchronization event, averaging of discrete EEG epochs will, when a sufficiently large number of epochs (trials) are averaged, result in its attenuation. The activity that does have a temporal relation to the event—the ERP—will then be enhanced in relation to the background EEG and thus become visible. If plotted in time, ERPs appear as a series of positive and negative deflections, which are thought to be the manifestation of underlying *ERP components*.

This paper was prepared while the author was supported by NWO Grant 575-270-082B, awarded to M. W. van der Molen and C. H. M. Brunia. The section on principal components analysis benefited from discussions with Marcel Croon of the Department of Statistics, Tilburg University. The useful suggestions of Martin Eimer, Emanuel Donchin, and two anonymous reviewers are gratefully acknowledged. Correspondence concerning this paper should be addressed to G. J. M. van Boxtel, Department of Psychology (P607), Tilburg University, P. O. Box 90153, 5000 LE Tilburg, The Netherlands (e-mail: g.j.m.vboxtel@kub.nl).

ERP components may be exogenous or endogenous—that is, evoked by events that are extrinsic or intrinsic to the nervous system. The exogenous ERP components occur near the eliciting event, and the time relation to that event is very stable. When a stimulus is presented, for instance, the exogenous ERP components have an early onset latency—that is, they occur within 250 msec from the presentation of the stimulus, and that latency does not vary much from trial to trial. The same can be said about exogenous components that are related to the occurrence of a response, although these relations have been less systematically studied. The characteristics of the exogenous components, such as amplitude and latency, depend only on the physical parameters of the eliciting stimulus—for instance, pitch or loudness, in the case of auditory stimuli. Examples of endogenous ERP components are the P300—a positive peak at about 300 msec after a stimulus—and the slow potentials, which are recorded preceding movement or during other mental activities. In contrast to the exogenous components, endogenous ERP components have a later and more variable onset latency, and their characteristics do not depend, or only partially depend, on stimulus parameters. The endogenous ERP components "are *invoked* by the psychological demands of the situation rather than *evoked* by the presentation of stimuli" (Donchin, Ritter, & McCallum, 1978, p. 350). They are psychophysiological entities in that they are the record of a physiological system from an organism engaged in a psychological task.

The preceding categorizations are an oversimplification that should not be taken too strictly. In practice, the differences are often not very clear-cut. The distinction between background and event-related EEGs, for in-

stance, in the work of Pfurtscheller and colleagues on the event-related desynchronization (ERD; see Pfurtscheller & Klimesch, 1991, for a survey)—in which specific frequencies of the background EEG are related to events—is difficult to make. Likewise, the distinction between exogenous and endogenous ERP components is not always very clear, most notably in the time range between 100 and 300 msec poststimulus. Yet the distinctions have heuristic value and can guide ERP research in practice.

The present paper is concerned with some of the methods that are used to study the latent endogenous components from the observed ERPs measured at the scalp and is focused on their application in experimental psychology. ERPs are now considered by most experimental psychologists to be a useful tool in the study of various psychological research questions (Meyer, Osman, Irwin, & Yantis, 1988). The most important contributions of ERPs to experimental psychology are believed to be in the evaluation of information-processing models, in mental chronometry, and in situations in which continuous measures are required or in which no overt response is available. The popularity of the ERP technique in experimental psychology is pleasing, since it is indicative of a certain degree of maturity in the ERP field. It demonstrates that the field has developed from a stage in which primarily research into the functional significance of the ERP components themselves was done to a stage in which this knowledge is sufficiently large to be applied in various settings. At the same time, care must be taken to ensure that the methodological principles underlying the measurement and identification of ERP components are observed. The application of ERPs in experimental psychology will only be successful if these principles, which are the result of a lot of research in the previous three decades, are carefully pursued. Therefore, this paper reiterates some of these basic methodological principles and assumptions. It is mainly concerned with the issue of how psychophysiological measurements, in particular ERPs, can be reliably decomposed into psychologically meaningful components and with some of the statistical and computational techniques that may be used for that purpose. The paper is by no means complete, and, at appropriate places, the reader is referred to other papers for more detail.

THE DEFINITION OF ERP COMPONENTS

An example of an ERP recorded from six electrode sites is given in Figure 1. The data are taken from an experiment in which 10 subjects were engaged in a choice reaction time task with a constant foreperiod of 4 sec demarcated by two stimuli (van Boxtel, van den Boogaart, & Brunia, 1993). As soon as possible after the second stimulus, the subjects had to squeeze a response button with the thumb and the index finger of the right hand up to 15% of their own maximum force. They had to reach that force criterion either as quickly as possible or about twice as slowly (factor response speed). The information about the required response speed on a particular trial was presented at either the first or the second stimulus (factor instruction stimulus). Figure 1 shows the ERP traces of fast and slow contractions in both conditions, starting from 0.5 sec before the second stimulus, until 2 sec thereafter.[1] Note that, contrary to conventions in other research areas, negativity is plotted upwards and positivity downwards in this figure, as is often done in ERP research. Donchin et al. (1977) estimated that two thirds of the researchers used the *negative up* convention, and a quick survey of recently published articles suggests that this number has not changed dramatically since then.

Visual inspection of Figure 1 reveals two positive deflections. The first peak has a latency of about 400–500 msec after the stimulus, a parietal or centroparietal maximum, and it seems to vary as a function of instruction stimulus, irrespective of response speed. The second peak has a latency of about 700–800 msec, a central maximum, and seems to vary as a function of response speed, irrespective of instruction stimulus. Hence, visual inspection of Figure 1 suggests that there are two independent waves—the first related to the processing of the stimulus and the second related to the execution of the response. The order in which these phenomena occur is in agreement with the predictions of serial stage models of human information processing.

This data set will be used throughout this paper to illustrate the theoretical notions and computational techniques that will be discussed. The first important distinction that needs to be made is the difference between a *component*, on the one hand, and a *peak*, *deflection*, or *waveform*, on the other hand. In Figure 1, for example, the two large positive peaks that can be seen after the respond stimulus are the dependent variables of the experiment, which can be seen to vary as a function of the independent variables—in this case, instruction stimulus and response speed. Peaks such as these are usually defined *observationally* on the basis of obvious characteristics of the deflection, such as sequence, polarity, and latency. For example, one might be interested in the first (sequence) positive (polarity) deflection in the interval between 200 and 600 msec (latency) after the stimulus. One might then take the most positive value in that interval for all electrodes, conditions, and subjects (a peak measure) and submit these values to statistical analysis. Alternatively, one might integrate or average the voltages in a certain time window (an area measure), a method which is less sensitive to noise than is the peak measure, but which may also underestimate differences between electrodes, conditions, and subjects. Fabiani, Gratton, Karis, and Donchin (1987) provided an excellent survey of the observational definitions of ERP components—or operational definitions, as they prefer—and discussed the advantages and disadvantages of several widely used methods in the context of the P300.

The nomenclature used for observationally defined ERP waveforms has usually been based on polarity and sequence or on polarity and latency. Under the first convention, the first small peak—which, in Figure 1, can be

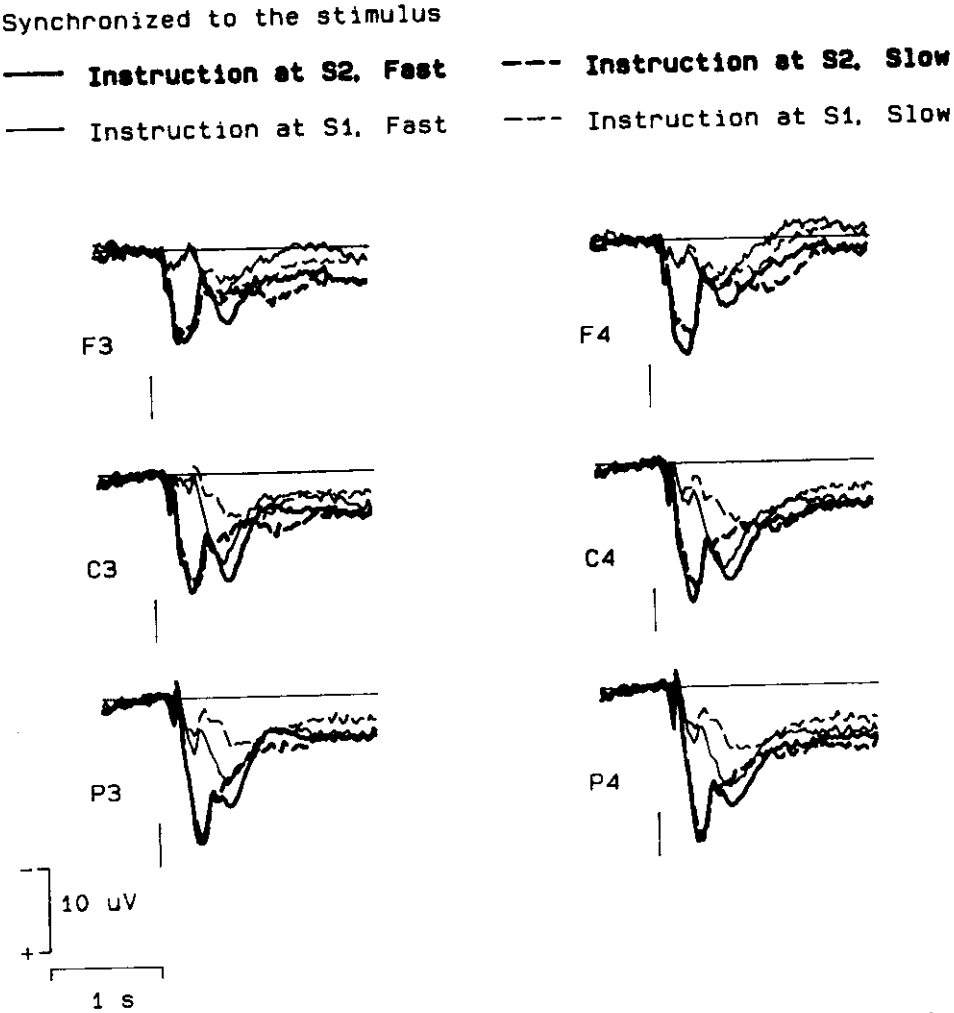

Figure 1. ERP traces from the four instruction stimulus by response speed combinations, averaged over 10 subjects. The synchronization point that was used in averaging was the respond stimulus, of which the occurrence in time is indicated by a vertical line in each set of traces from one electrode position (data are from van Boxtel, van den Boogaart, & Brunia, 1993).

seen to occur just after the stimulus, especially at the central and parietal electrodes—would be designated by P1, which is followed by N1, P2 (the large positive peak), N3, and P3. Under the second convention, these peaks could be described as P200, N300, P500, N650, and P800, for example. Either system has advantages and disadvantages, and the enormous variability in ERP waveforms under different experimental conditions makes it difficult to generally prefer one system over the other. Donchin et al. (1977) recommended the use of the polarity and latency convention, taking the modal or mean latency when describing data from a number of conditions and subjects. Their recommendation is followed by most researchers nowadays, and, in the present paper, the two positivities will also be denoted by P450 and P750, respectively.

Table 1 exhibits the statistical effects of peak and area measures taken from the two positive deflections in Figure 1, P450 and P750, and shows some of the weaknesses and limitations of the peak and area measures. Some of the obvious effects seen in the average ERP traces shown in Figure 1 are not supported by the statistical tests, even though visual inspection of the individual ERP traces of all subjects suggests that these effects are present in all subjects to a certain degree. It is therefore unlikely that the reason for the difference between the visual observation and the statistical results is in a large between-subjects variation that results in small F-ratios. It is more likely that the time window from which the measures were taken had to be large in order to encompass the latency variation from subject to subject, with the side effect that the experimental variation of the P750 was for a large part included in the P450, especially when the area measure was taken.

In contrast to a peak or deflection, the term component should be reserved to denote a *theoretical construct* rather than an observed waveform. This theoretical en-

Table 1
Main Effects and First-Order Interactions of the MANOVA on Area and Peak Measures of the Data Depicted in Figure 1

		F Values			
		Area Measure		Peak Measure	
Effect	*df*	1	2	1	2
Instruction stimulus (S)	1,9	11.22†	2.79	11.05†	1.48
Response speed (R)	1,9	7.76*	1.76	24.12†	7.07*
Electrode position (E)	2,8	2.71	0.58	1.89	0.02
Hemisphere (H)	1,9	7.55*	9.65*	10.55†	4.03
S × R	1,9	2.66	0.49	3.62	0.14
S × E	2,8	6.41*	0.27	3.36	1.03
S × H	1,9	0.00	0.30	0.60	0.19
R × E	2,8	0.37	1.03	1.35	1.58
R × H	1,9	1.04	0.76	3.05	0.20
E × H	2,8	1.42	3.07	0.52	2.06

Note—The analyzed intervals were 200–600 msec after the stimulus for the first peak and 600–1,000 msec for the second. For the peak measure, the most positive value found in each of these intervals was taken; for the area measure, the mean of those intervals was computed. *$p < .05$. †$p < .01$.

tity is believed to represent "some essential physiological, psychological or hypothetical construct whose properties are under study" (Donchin et al., 1977, p. 10). The confusion between the observational and theoretical definitions stems from the fact that the theoretical P300 may observationally appear as P250 or P400. Therefore, Donchin et al. (1977) proposed to make this distinction clearer by using an identifying mark whenever the theoretical rather than the observational definition is intended. Thus, "*P300* would refer to the theoretical component which might observationally appear as P350" (p. 11). This suggestion has never really caught on, however, but it does remain important to distinguish the observational from the theoretical definitions. For the present data set, the related questions would be whether the P450 is a manifestation of the stimulus-related P300 component originally discovered by Sutton, Braren, Zubin, and John (1965) and the P750 a manifestation of the response-related reafferent potential described by Kornhuber and Deecke (1965).

The ERP deflections recorded on the surface of the scalp are the result of volume-conducted electrical activity that is generated in various brain areas during task performance—mainly postsynaptic potentials from large neuronal populations (see, e.g., Caspers, Speckmann, & Lehmenkühler, 1980). According to some authors, the theoretical definition of ERP components is based only on these underlying neuronal generators (see, e.g., Sams, Alho, & Näätänen, 1984). Most authors, however, agree with the notion, most clearly expressed by Donchin et al. (1978), that the theoretical definition of ERP components must also be based on its function—that is, its relation with experimental variables. These authors recognized that the scalp-recorded ERPs are produced by what they called different neuronal aggregates, but they went on to note that "Functionally distinct aggregates need not be anatomically distinct neuronal populations. But it is assumed that neuronal aggregates whose activity will be represented by an ERP component have been *distinctly affected by one or more experimental variables*" (p. 353). Hence, they define components not on the basis of peaks or troughs in the waveform but on the basis of experimental variation, using the adage "*All we can study is that which varies*" (p. 354).

The difficulty with neuronal generators as a defining attribute of ERP components is, of course, that they should be determined every time the definition is applied. For instance, if the neuronal generators were included as a defining attribute for the P300 component, every study using the oddball task would necessarily include some method for defining these generators, such as equivalent dipole modeling, a PET scan, or a similar technique. Otherwise, the criteria for concluding that a P300 was measured would never be fulfilled. It is clear that, even though the neuronal generators of the component may be the fundamental entities ultimately of interest, this is not practical. With the recent availability of recording systems with which the activity of high numbers of electrodes can be sampled at once and the development of sophisticated source localization and imaging techniques, it can nevertheless be expected that the neuronal generators, or at least the scalp distribution, of ERP components will become increasingly important in psychophysiology—not as a defining attribute maybe, but at least as an aid in distinguishing between ERP components and arriving at a more fine-grained level of description. A nice illustration of this approach was recently given by Donchin, Spencer, and Dien (1997), who made a distinction between the classical ("Suttonian") P300 and the Novelty-P300 based on the scalp distribution by spatial principal components analysis.

PRINCIPAL COMPONENTS ANALYSIS

An important yet fairly simple method that can be used as an aid to infer the existence of theoretical components from observed waveforms is principal components analysis (PCA). PCA is one of the techniques usually subsumed under the general label of *factor analysis*.[2] The techniques are intended to describe the complex relations between a large number of variables and to describe these variables in terms of a lesser number of hypothetical, unobserved, latent variables. PCA differs from other factor-analytic techniques in that the factors extracted (termed principal components) are closely related to the original dependent variables, which is not necessarily so in other techniques. In PCA, each principal component is simply a weighted linear combination of all the original dependent variables, and, theoretically, as many principal components may be extracted as there are dependent variables. Furthermore, the principal components are extracted from the data set in a hierarchical fashion: The first component accounts for the largest proportion of the variance in the data, and the successive components must be both orthogonal to the preceding ones and account for the largest portion of the residual variance. For typical ERP data, this percentage drops off rapidly after the first

five or six components, which usually account for 90–95% of the variance in the data.

Besides reducing the often huge amount of data that is usually collected in typical ERP measurements, the promise of PCA is that it gives insight into the unobserved, theoretical components from the observed measurements recorded at the scalp. The use of PCA as a tool for the study of ERPs was advocated by Donchin (1966). Tutorial reviews of the use of PCA for ERP data are given by Chapman and McCrary (1995), Donchin and Heffley (1978), and Glaser and Ruchkin (1976), among others. The way in which PCA is most frequently used is to arrange the recorded ERPs so that the successive time points of the single ERPs are treated as *variables* in the PCA program (often a statistical package such as SAS, SPSS, or BMDP). The different ERPs, recorded from the single electrode positions, experimental conditions, and subjects, are treated as *cases* (observations). For instance, in the data shown in Figure 1, there were 2 sec after the stimulus, and the sampling rate was 200 Hz, so there would be $2 * 200 = 400$ variables. Data were recorded from six electrode sites (F3, F4, C3, C4, P3, and P4), during two contraction speeds (fast, slow), under two experimental conditions (instruction at S1 or instruction at S2), and from 10 subjects. Hence the number of cases would be $6 * 2 * 2 * 10 = 240$.

This procedure may be referred to as a *temporal PCA*, in contrast to a *spatial PCA*, which will be briefly discussed later. The basic model for the temporal PCA—for simplicity, without electrodes, experimental conditions, subjects, and error—is given by:

$$x(t) = \sum_{k=1}^{K} a_k c_k(t), \qquad (1)$$

where t are time points $(1, \ldots T)$, and k are components $(1, \ldots K)$. The equation is given here to show that the observed ERP waveform, the $x(t)$—which, as indicated by the (t), is time-variant—is conceived of as a linear combination of time-invariant and time-variant variables (the a_k and $c_k(t)$, respectively). The result of the PCA is a set of K time-variant *component loadings*, which represent the contribution of each component to the voltage at each time point, and a set of time-invariant *component scores*, which represent the contribution of each component to each of the ERP waveforms. That is, the component loadings are the *basic waveforms*, indicating the instants in the ERP in which amplitude variability exists, and the scores indicate the nature of that variability that can be analyzed by the usual statistical techniques, such as analysis of variance.

The PCA technique will be demonstrated with the use of the data set presented in Figure 1. A number of important choices that have to be made in the course of the application of the PCA will be discussed, and some useful suggestions for optimizing the results will also be given. However, before doing so, one important aspect of PCA should be stressed: PCA is a *correlation* technique. This implies that researchers have to be careful in giving a causal interpretation to the results obtained. The extracted components cannot directly be interpreted as anatomical, physiological, or even psychological entities. PCA just analyzes and describes variance in the observed data. For this and for other reasons, some of which will be discussed below, the use of PCA as a tool for analyzing ERPs has also been criticized. Most, if not all, of these criticisms have been sufficiently repelled to justify its use, but PCA should in no way be viewed as a kind of magic tool. It should rather be viewed as just another way of analyzing multichannel data, like simple peak picking, albeit a technique with some notable advantages over some other methods that are often used.

Number of Variables

The time interval to be analyzed may be constrained by practical limitations, the most important of which is the number of variables (time points). Some PCA programs have an upper limit on the number of variables they can handle, often related to the particular computer software or hardware at hand (particularly the availability of memory). A maximum of about 100 variables is not uncommon for typical personal computer configurations. In addition, the number of variables should be less than the number of cases (observations). Although most PCA programs nowadays allow the number of variables to be larger than the number of cases, the reliability of the PCA increases with the number of cases. In the present example, there were 400 variables and 240 cases, so either the time interval should be shortened, or the number of variables should be reduced by down sampling (i.e., picking an equally spaced number of samples from the data set). Since it is undesirable to limit the time interval before the first PCA has even been calculated, the second method is preferred.

The number of variables in the present data was reduced from 400 to 50 by calculating means of 8 successive time points. This method was used, instead of simply picking every eighth point, in order to avoid aliasing. Aliasing occurs if the cutoff frequency of the low-pass filter used in data collection is more than half of the sampling frequency (see, e.g., Srinivasan, Tucker, & Murias, 1998). Condensing by a factor of eight means down sampling from 200 to 25 Hz, and the present data were low-pass filtered at 30 Hz. Hence, a low-pass filter of a maximum of 12.5 Hz should be applied to the data before selecting every eighth point (see Ruchkin & Glaser, 1978, for simple low-pass filters in software). Taking the mean of 8 consecutive time points is just another way of low-pass filtering the data, and is often computationally faster. Downsampling by a factor of eight is justified because the components of interest are all expected to have a frequency content well under 10 Hz, and it makes the dimensions of the data matrix (50 time points by 240 cases) reasonable.

Baseline Level

Because the EEG does not have a natural zero level, at least in behaving subjects, an artificial baseline is calcu-

lated. Often the mean of an interval of up to 1 sec before the stimulus is taken and subtracted from all the successive time points to be analyzed. Wastell (1981a) argued that this common procedure leads to the extraction by PCA of a first or second component exhibiting high loadings toward the end of the analyzed time interval. Such a component, which is the result of the autocorrelated nature of the data, is shown in Figure 3 (Component 1). Its component scores usually do not show any meaningful statistical effects. The probability that such a "spurious" component will be extracted is reduced by DC removal, either by correcting the total time interval of each individual ERP (case) to zero microvolts or by setting the mean of all cases (electrodes and conditions) of one single subject to zero. The latter two procedures generally produce similar results. If the zero baseline level is not situated at the beginning of the analyzed time interval, the first component extracted will usually reflect the existing baseline differences. Wastell noted that even in such cases PCA tends to extract components in the order of their frequency content—that is, first the components that vary slowly over time, followed by the faster components (slowly varying components usually operate over a longer epoch and hence explain more variability). This implies that one should be careful to attach significance to the amount of explained variance of a component, which decreases with each component that is extracted. The fourth or fifth component may not explain as much variance as does the first, but this does not imply that it is less meaningful. The variance in the component scores is more important, because it is related to the experimental manipulations.

Association Matrix

The first step in the temporal PCA consists of computing the association between all individual time points (variables), which results in a symmetrical association matrix with a size of the number of time points. The general idea is, of course, that associated time points belong to the same underlying component. The question is which association measure to take, and the choice is usually restricted to cross products, covariances, or correlations. The cross products matrix is produced by summing the results of the multiplication of all possible pairs of variables across cases. The covariance matrix is computed in the same way, except that the mean of each variable is subtracted from each case before the cross products are computed. In addition, for the calculation of the correlation matrix, the values of the individual cases for each variable are also divided by the standard deviation across all cases of that variable, with the result that all variables have equal variance before the cross products are computed.

The choice of the association measure has great consequences for the interpretation of the PCA results. As Donchin and Heffley (1978) pointed out, PCA of the cross products matrix will result in components that are related to large peaks in the original waveforms, even when such peaks do not show any experimental effects. Furthermore, the loadings of the first component usually represent the grand mean waveform of the original data. An advantage of analyzing the cross products matrix is that the resulting component loadings and scores can be interpreted in terms of the original data—that is, the measurement units are the same (usually microvolts). This is of particular importance in interpreting the polarity of a component. Only when the cross products matrix is analyzed, a positive component score is surely indicative of positivity in the original data. However, the undesirable property that the cross products PCA may result in components without experimental variation causes this matrix to be an unsuitable choice for PCA.

Analysis of the covariance matrix will lead to the extraction of principal components that correspond to the variance around the grand mean ERP. The extent to which the individual ERPs differ from the grand mean, rather than their absolute amplitude, determines which principal components are extracted. This is a desirable feature, because it implies that components are only extracted if there is variation across electrodes, conditions, and subjects, and that is exactly what researchers are looking for. The resulting component loadings and scores should be interpreted with respect to the grand mean ERP data—that is, a positive component score indicates that the ERP in that situation is more positive than is the grand mean. PCA on the correlation matrix results in components that can be interpreted in roughly the same way. The use of the covariance instead of the correlation matrix is justified by the fact that the values of all variables represent a voltage. Hence, the variables have the same scale, and there is no need to scale the data by the standard deviations. However, there is no a priori reason why the standard deviations of all variables are the same, even when identical measurement units are used. The difference between the PCA of the covariance matrix and that of the correlation matrix is usually insignificant for typical ERP data (Chapman & McCrary, 1995). Practical considerations may guide the choice between the covariance and the correlation matrix, such as the analysis package available. BMDP allows all three association matrices to be factored; SAS allows a choice between the covariance and correlation matrix; and in SPSS only the correlation matrix is analyzed. But even in the latter case, it is easy to rescale the obtained component loadings to the original measurement unit by multiplying the value of each time point by the standard deviation of that time point. For the present data set, the covariance matrix was analyzed, using SAS PROC FACTOR.

Rotation Criterion and Number of Components

The purpose of rotation is to obtain a simpler interpretation of the components. The solution of a PCA with more than one component is not unique but just one of the infinite number of possible solutions. By rotation, one tries to find the solution that possesses the property of *simple structure*, according to which a single variable has a high

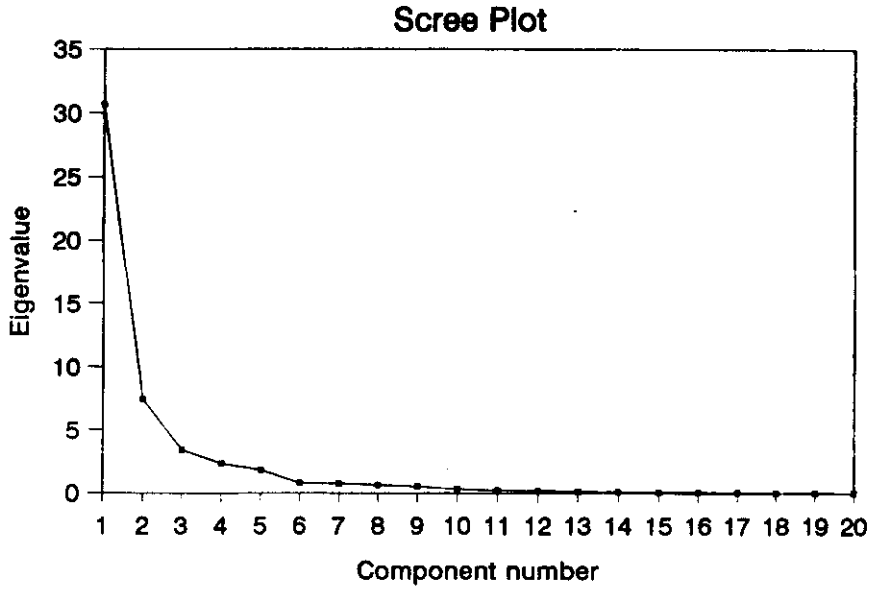

Figure 2. Scree test of the PCA of the correlation matrix, for which the data displayed in Figure 1 were used. Note that the original number of components was 50 (equal to the number of variables), but only the first 20 components are displayed to enhance resolution.

loading on only one component and a zero or negligible loading on all other components. In terms of the present temporal PCAs on ERP data, this means that the temporal overlap of the principal components is minimized. Often a Varimax rotation is done, which tries to find components which are either large or small, not intermediate, while maintaining orthogonality—that is, independence between components. Oblique rotation methods—for instance, Direct Quartimin—may achieve a greater degree of simple structure but do not have the desirable property of independent components. There has been some discussion about presenting rotated or unrotated component loadings (see, e.g., Rösler & Manzey, 1981; Wastell, 1981b), but the consensus is that component loadings should be rotated and that the Varimax criterion is appropriate for most practical applications.

Irrespective of the rotation method, it should be determined how many components should be rotated. The total number of components in a PCA—K in Equation 1—is theoretically equal to the number of variables that the PCA was based on—that is, 50 in the present case. For typical ERP data, this number is much less, usually below 10, although this depends on the experimental conditions of the particular data. Since PCA extracts components in the order of the percentage of explained variance, an obvious method is to find a criterion based on the amount of explained variance, which is indicated by the eigenvalue of the association matrix. Such a criterion ensures that the nonretained components explain little variance in the original data and can be treated as noise. The best-known criterion is the *eigenvalue equals one* rule, by which only those components are retained that have eigenvalues greater than or equal to one. This is identical to stating that a component should explain the variance of at least one variable (time point). It should be noted that the rule only makes sense if the correlation matrix is factored, so that, if the covariance matrix is to be analyzed, one could do a separate PCA on the correlation matrix just to determine the number of components and then extract that number from the covariance matrix subsequently.

The eigenvalue equals one rule often overestimates the number of relevant components somewhat. Another method, which generally leads to a slightly smaller number of components, is the scree test, in which the components' eigenvalues are plotted against the ordinal component number (Figure 2). Then, looking backward from right to left along the x-axis, one tries to determine the component number at which the eigenvalue starts to increase. For the present data, this suggests that five components should be retained and rotated, which in this case agrees with the number of components that are based on the eigenvalue equals one rule. Although the scree test allows for a certain degree of subjectivity in determining the number of components, simulation studies have shown that it results in a more accurate number of components than does the eigenvalue equals one rule. Chapman and McCrary (1995) recommended the performance of rotation on both numbers of components and, if they are different, the selection of the most appropriate solution on the basis of the theoretical interpretation of the resulting components.

Statistical Tests on Component Scores

The scores of each component that results from the PCA may be subjected to the same statistical tests as all other ERP measures, such as peak or area measures.

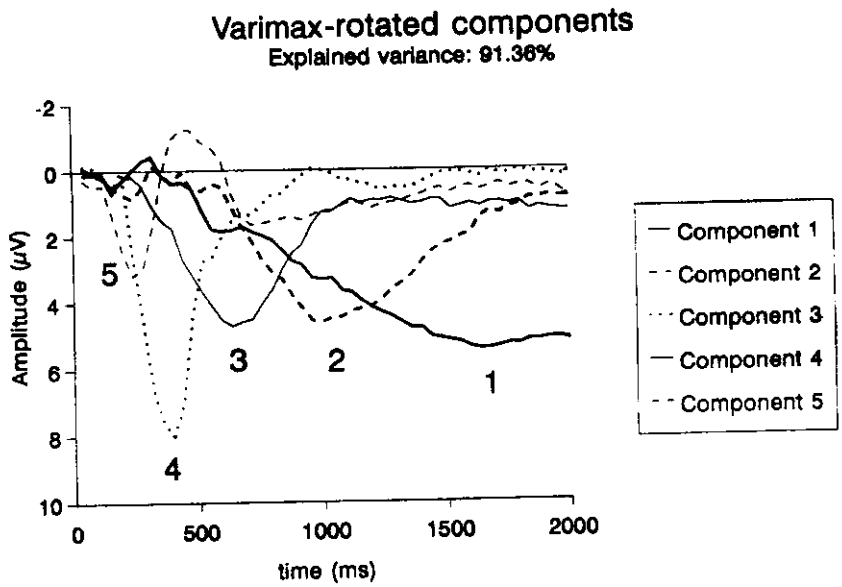

Figure 3. Component loadings obtained after factoring the covariance matrix, and rotating five components to a Varimax criterion. The data are the same as in Figure 1, but condensed to 50 time points by taking means of 8 successive points. The statistical effects on the associated component scores are shown in Table 2.

Often, univariate analysis of variance (ANOVA) is done separately for each component. If the experimental conditions are administered within subjects, as in the present as well as in most psychophysiological experiments, a repeated measures ANOVA should be performed. For this test, it is assumed that the population variance is homogeneous across treatments (in this case electrodes and conditions), and violation of this assumption leads to an increase in Type I error. Because ERP measurements usually consist of multiple electrodes, and the values recorded at nearby electrodes are usually more correlated than are those of distant electrodes, this problem is very likely to occur in ERP analysis (Vasey & Thayer, 1987). Therefore, one should either perform univariate ANOVAs, in which the degrees of freedom involved in the tests are corrected, for instance by the Greenhouse-Geisser ε, or perform a multivariate analysis of variance (MANOVA), which has another way of dealing with this problem (see O'Brien & Kaiser, 1985, for an excellent introduction to MANOVA).

Results of Present PCA

With the present data set, a PCA was computed on the covariance matrix using 50 variables and 240 cases. The scree test presented in Figure 2 and the eigenvalue equals one rule converged on extracting five components, which were rotated to simple structure using a Varimax criterion. The resulting component loadings are exhibited in Figure 3, and the most important results of MANOVAs on the corresponding component scores are given in Table 2. In the component loadings plot, a large positive peak at about 450 msec after the stimulus can be observed (Component 4). Note, again, that the positivity of this peak indicates that the values are positive with respect to the grand mean, given the fact that the covariance matrix was factored. The component scores reveal a large effect of instruction stimulus and no effect of response speed, suggesting that this component is stimulus-related. Component 3 peaks around 750 msec poststimulus, and its scores show no main effect of instruction stimulus but a large effect of response speed. Thus, this component is interpreted to be response-related. Figure 4 displays the scores of Components 3 and 4, and it is immediately clear from that figure that Component 3 has an effect of response speed but not of instruction stimulus, whereas the reverse is true for Component 4. Note that the PCA

Table 2
Main Effects and First-Order Interactions of the MANOVA on the Component Scores of 5 Varimax-Rotated Components Obtained by PCA of the Covariance Matrix

		Component Number				
Effect	*df*	1	2	3	4	5
Instruction stimulus (S)	1,9	14.76†	0.61	1.10	77.29†	0.35
Response speed (R)	1,9	0.05	3.02	14.92†	0.67	10.24*
Electrode position (E)	2,8	6.48*	0.27	9.54†	12.25†	6.35*
Hemisphere (H)	1,9	2.43	79.42†	1.71	13.26†	2.90
S × R	1,9	0.83	0.16	1.21	0.25	11.19†
S × E	2,8	5.28*	4.65*	12.53†	3.72	17.21†
S × H	1,9	0.06	10.63†	5.76*	0.63	1.50
R × E	2,8	0.18	1.70	8.94†	4.89*	3.18
R × H	1,9	0.91	9.65*	1.57	3.26	2.10
E × H	2,8	6.36*	2.76	0.39	13.58†	0.65

Note—The loadings of these components are shown in Figure 3.
*$p < .05$. †$p < .01$.

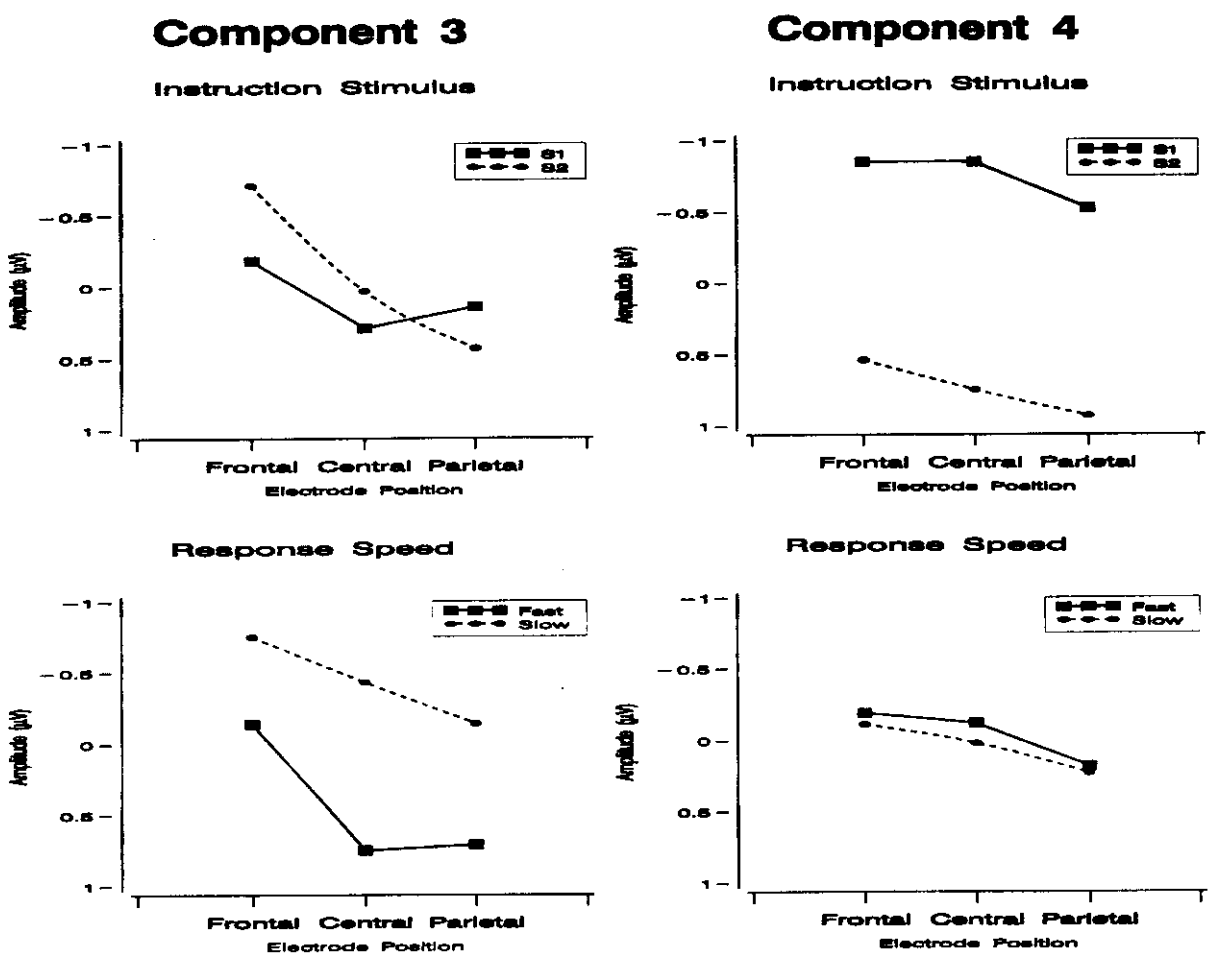

Figure 4. Effects of the factors instruction stimulus and response speed on the scores of Components 3 and 4, the loadings of which are displayed in Figure 3, as a function of electrode position.

resulted in separate stimulus- and response-related components, unlike the peak and area measures, in which main effects of the within-subjects factor response speed were also observed during the P450 interval (Table 1).

The other extracted components reveal a few other aspects of the data. As already noted, the first component is an example of a spurious component resulting from the autocorrelation in the data, when the baseline level is pinned at the beginning of the recording interval. It attracts some variance that is due to instruction stimulus and the scalp distribution of the data. The second component, peaking at about 1 sec poststimulus, attracts most of the variance that is due to hemisphere differences, whereas Component 5 reflects the interaction between stimulus- and response-related activity and can be seen to temporally overlap Components 3 and 4.

Taken together, the results of this PCA support the existence and independence of two components in the recorded data, the first of which is stimulus-related and probably a manifestation of the P300. The second is response-related, presumably an instance of the reafferent potential. It is important to note that this conclusion does not directly follow from the PCA. The PCA just looks for instants at which there is systematic variation because of experimental treatments, electrodes, subjects, and so forth. A peak at about 450 msec in the loadings of Component 3 just indicates that there is systematic variation at that time, the source of which can be analyzed by the statistical tests of the component scores. Because the existence of systematic variation is one of the key prerequisites for defining theoretical components, as was argued in the previous section, the PCA results yield important arguments for the theoretical interpretation of the observed waveforms.

Possible Criticisms

The use of temporal PCA for analyzing ERP data has been criticized mainly on two grounds. First, PCA is said to have difficulties extracting components with temporal overlap. Wood and McCarthy (1984) reported that under certain circumstances, especially with overlapping components, the PCA–Varimax rotation—ANOVA of factor

scores strategy may result in "misallocation of variance" across components. A detailed discussion of this problem is beyond the scope of this paper, and it might have been somewhat overemphasized by critics of the PCA method. Even a quick glance at Tables 1 and 2 strongly suggests that this problem is even more serious in the peak and area measures, inasmuch as almost the entire variance associated with response speed was contained in the interval of the P450. It is quite clear that, when a peak or area measure is taken and there is overlap of the underlying components, the resulting measure is necessarily the sum of a combination of these components. This kind of misallocation is not made explicit when peak or area measures are taken, but, when a PCA is done, the misallocation issue often comes up, even though the PCA handles component overlap much better than do peak or area measures.

A second criticism relates to PCA's difficulty in extracting components with temporal variability (latency jitter). It is true that PCA extracts components on the basis of variability in amplitude and not in latency; thus, it treats latency variability as amplitude variability. Donchin and Heffley (1978) have shown that a PCA of a data matrix that contains waveforms of fixed amplitude but variable latency results in a biphasic component representing the latency shift. One may criticize PCA for extracting such a component and hence not yielding the "true" component structure of the data, but, as already noted, PCA is not a magic tool which can be applied blindly. A careful analysis of the component structure in relation to the observed data will quickly provide insights into such effects, and this is a more constructive approach than just dismissing PCA altogether is. In addition, as already suggested by Donchin and Heffley and by various others, such problems can fairly easily be circumvented by special techniques that deal with the latency variability and that can be applied prior to the PCA, as will be shown in the section on Woody filtering.

Spatial PCA

Until now, only the temporal PCA has been discussed, in which time points are treated as variables and the ERPs obtained from the different electrode positions, experimental conditions, and subjects as cases (observations). However, this is only one of the possible uses of PCA. Another method is to treat the electrode positions as variables and the time points obtained in a particular experimental condition as cases. The resulting component loadings will now reveal at which electrode, or cluster of electrodes, activity is present that can be separated from the activity at other electrodes. The component scores will then provide an indication of the time points at which that activity is present. In other words, the spatial PCA decomposes the potential distribution of the data, which is justified because, as discussed previously, one of the key characteristics of a theoretically defined ERP component is its scalp distribution, being an indication of the underlying neurophysiological generators. Since modern equipment allows the activity from large arrays of electrodes to be fairly easily recorded, it can be expected that the spatial PCA will become increasingly important. Strangely enough, the spatial PCA is not yet included as an option in most modern source localization software packages, but it can easily be computed with standard statistical software by using the same guidelines as those indicated above for the temporal PCA.

Spatial PCA is most useful in combination with temporal PCA. This can be done in one of two ways. First, a temporal PCA as described above can be done to identify the epochs within the ERP where variability exists. Subsequently a spatial PCA follows the temporal PCA to decompose the scalp distribution in that epoch. Alternatively, the spatial PCA may be computed immediately, not for just one particular epoch, but over the whole recording interval. All time points are then entered as cases, which yields very large numbers of cases (usually over 10,000) and hence reliable results. Each resulting component loading can then be mapped by using interpolation techniques, such as those that one would apply to the standard observed ERPs. Subsequently, a temporal PCA is computed on the component scores, which are split into a time-variant and time-invariant part. An illustration of the latter method is provided by Donchin et al. (1997).

Another interesting recent development in the spatial domain is Möcks's (1988) trilinear decomposition of ERPs. He suggested a PCA model in which an extra parameter that represents electrode coefficients was added. Instead of decomposition into two parts—the loadings and the scores, the latter including the variation due to electrodes—Möcks proposed to decompose the observed ERP into three parts: loadings, electrode coefficients, and scores. In this case, the scores represent the variation that is due to experimental treatments and subjects only. This model has some interesting properties, most notably that it has unique solutions for a fixed number of components, so that rotation is no longer needed. However, it is computationally more complicated and not available in standard statistical packages. This might be an interesting area for future developments.

DEALING WITH TEMPORAL VARIABILITY OF ERP COMPONENTS

In the above section, it was noted that PCA has some difficulties in dealing with temporal variability in waveforms. Indeed, it should not be forgotten that most ERP analyses are done on data that are averaged over a sometimes substantial number of trials that belong to the same experimental manipulation. For instance, for the data of Figure 1, the average number of trials per condition was about 60 for each subject. Of course, the first positive peak after the response, labeled P450, will not occur at 450 msec on each single trial. Changes in the state of the subject's attention and other sources of uncontrolled

variation will lead to temporal variability in the peak of the deflection (also referred to as latency jitter). The effect of trial-to-trial variability of peak latencies on the resulting average is that peaks will be smaller and broader than in the "true" average—that is, the average without latency variability. In the case of the present data, it can be expected that latency jitter affects the P750 more than it does the P450, because the total variability of the P750 is the sum of the variabilities that are due to stimulus processing, central processing, and response execution. The P450, on the other hand, is supposed to reflect stimulus processing only, so it is likely to show less variability (unless there is no variability in central processing and response execution, of course). Judged from Figure 1, the P750 does indeed seem smaller and broader than does the P450.

In order to estimate the response-related P750 better, one could produce another average that includes the same data trials but in which the instant of the overt response is aligned, rather than the instant of the stimulus, as in Figure 1. This will eliminate the temporal variability of the response and, as a consequence, introduce temporal variability in the stimulus. Since the P450 is thought to be related to the stimulus and the P750 to the response, the new average is likely to have a smaller and broader P450 peak and a greater and narrower P750 peak, as compared with the average synchronized to the stimulus. To eliminate response-related variability even further, one could take the onset electromyographic (EMG) activity as the synchronizing point—inasmuch as that instant is assumed to be closer in time to the events taking place in the brain—and, hence, the variability between EMG onset and the overt button press can also be eliminated.

In the present experiment, there were only two external events—the stimulus and the response. The P450 peak is thought to be related to the stimulus, so its temporal variability is lowest in the average synchronized to the stimulus. However, that average still has inherent temporal variability of the P450 peak, which may even be greater in one condition than in another. In order to eliminate that variability, an average synchronized to the internal response to the stimulus should be calculated, but that is exactly the instant that the P450 peak is thought to represent. One could try then to synchronize the average on the P450 peak itself; the most frequently used procedure for doing so is called Woody filtering. Woody filtering is very briefly discussed below, after which a more elaborate analysis of the detection of EMG onset is given.

For producing latency-corrected ERP averages, an estimation of the latency of the signal of interest (the P300 or the EMG, in this case) on the single trial is needed. Estimating the overall latency from averaged ERP data is also possible but has its own set of caveats, some of which are discussed by Smulders, Kenemans, and Kok (1996). The most important pitfall for onset latency measurement from averaged data is that the onset will be determined by the trials in which it occurs relatively early. On the basis of simulations, Smulders et al. suggested that this pitfall could be avoided by taking the instant at which the component of interest exceeds 50% of its maximum amplitude (or some other threshold, depending on the signal). Often only differences in latency between experimental conditions are needed, not the absolute values of the latency themselves. In this case, the method recently introduced by Miller, Patterson, and Ulrich (1998) was shown to be very accurate and computationally fast and simple. Although the latter two methods have the enormous advantage that calculations can be done on the single subjects average, however, they cannot be used to produce latency-corrected averages, for which calculations need to be done on the single trials.

Woody Filtering

Woody (1967) proposed an adaptive filtering method for dealing with the temporal variability of the signal, by which a latency-corrected average is produced. The method uses a template, which is usually the initial average synchronized to the stimulus, although sometimes a low-frequency positive half-cycle sine wave is used. The covariance between the template and the observed single trial ERP is then computed across a range of time shifts— that is, the template is shifted from left to right across the single trial ERP. When the covariance reaches a maximum, the signal is said to be detected. This procedure is repeated for every single trial, and each trial is then shifted by a number of points corresponding to the maximum of the covariance function for that trial. Subsequently, a new average is computed, which provides a better view of the latency-corrected signal. The new average then replaces the old template, and the entire procedure is repeated until, on some index, no further improvement is seen.

A major danger of Woody's (1967) procedure and of all methods dealing with latency alignment, for that matter, is that the template may be erroneously aligned to background noise instead of to the single trial peak. Wastell (1977) argued that this is often the case when more than one iteration is done, which also corresponds to the findings of Möcks, Köhler, Gasser, and Pham (1988) using simulated data. The latter authors also presented a new method for obtaining estimates of latency shifts in the frequency domain, which performed better than Woody's algorithm and which also allows a statistical test on the presence of latency jitter. Their method, which is based on a maximum likelihood estimation, is too complicated to describe here in nonmathematical terms, but those interested are referred to their Appendix, in which the full details of the method are given, with a view to implementing it in computer software.

Producing latency-corrected averages should be viewed as an essential step in the data analysis procedure, especially when the initial (often stimulus-locked) average suggests that latency jitter may influence the interpretation of the results. If one records ERPs in different conditions and the ERP in one condition is smaller but broader

than it is in another condition, one can be sure to have a latency jitter problem, and adjustment is an absolute necessity for a useful interpretation of the results. Woody's (1967) algorithm is very simple and hardly computationally intensive, given today's availability of computing power. Yet it has proved to be very useful in a number of studies too large to enumerate here.

The Detection of EMG Onset

Methods for detecting EMG onset are described in detail here because not much attention is given to them in the literature. It seems as if each laboratory has its own method of EMG onset determination, which is trusted so much that it is not considered necessary to provide information about its accuracy; at least, this is seldom reported in articles. Yet, automated methods for EMG onset determination differ much in their ability to detect the onset accurately, although these differences can be minimized substantially by applying a few additional rules. The detection of EMG onset is not only important for producing averages synchronized to the response. In some cases, ERP components are defined with respect to EMG onset, such as the motor potential of the readiness potential (Kornhuber & Deecke, 1965). In addition, in modern mental chronometry, the reaction time interval is divided up into an increasing number of epochs, each of which is thought to be associated with a certain psychological process. Psychophysiology provides a number of methods for delineating those epochs, and the EMG onset is one of them.

EMG onset can only be reliably scored if the EMG signal is rectified—that is, if the absolute value at each time point is taken. The correct recording procedure is as follows. First, the EMG signal is high-pass filtered in order to eliminate slow movement artifacts. Then the signal is (full wave) rectified, an operation by which all negative values are flipped to positive ones. This will approximately double the frequency content of the signal; hence, it should be done before low-pass anti-aliasing filtering. The cutoff frequency of the high-pass filter will determine the steepness of the initial EMG burst; the higher that frequency, the steeper will be the resulting burst. However, the cutoff frequency will also be limited by the cutoff frequency of the low-pass filter and the desired sampling frequency. For instance, if the data are to be sampled at a rate of 200 Hz, the maximum low-pass cutoff frequency is 100 Hz (preferably lower, depending on the filter's roll-off), and the cutoff frequency of the high-pass filter is necessarily lower than is that of the low-pass filter. In practice, good values for the high-pass filter are 20 or 30 Hz, allowing a low-pass filter cutoff of 50 Hz and a sampling rate of at least 100 Hz to avoid aliasing. The resulting bandpass of 20 or 30 Hz to 50 Hz seems to be very narrow, but it is a practical range if one is interested in automatically scoring the EMG onset for typical ERP applications.

Barrett, Shibasaki, and Neshige (1985) presented a method of visual EMG onset determination. The point at which the EMG trace exceeds a threshold voltage preset by the experimenter is determined by a computer, which then displays the rectified EMG on the screen and draws a vertical line through that point. The vertical line can be moved by cursors, and the precise onset can be determined manually by the experimenter. This method allows for precise EMG onset determination but is very laborious if large numbers of trials and subjects are involved; moreover, it allows for a certain degree of subjectivity in onset determination.

Van Boxtel, Geraats, van den Berg-Lenssen, and Brunia (1993) compared the accuracy of some automated methods for EMG onset determination. They asked five experienced researchers to visually determine the onset of minimally filtered EMG traces from a large number of trials to the nearest millisecond. They then filtered and down sampled the same trials to make them comparable to EMG recordings obtained in real experimental applications. Subsequently, the various onset detection methods were applied, and the results of those methods were compared to the visually determined onset. Only those methods were studied that were presented in articles whose main purpose was to describe an EMG onset detection method. This does not imply that other methods may not be just as accurate, or even more accurate, but as explained above, there are probably just as many methods as there are psychophysiological laboratories.

In Figure 5, an example of a single-trial EMG record is displayed, annotated to show the guiding principle behind the methods that were investigated. All methods initially rely on a threshold voltage comparison—that is, they locate the time point in which the EMG burst exceeds a certain fixed voltage. The time point found is then regarded as a kind of *candidate onset*, and additional criteria have to be met in order for this candidate point to qualify as the real EMG onset. If the additional criteria are not met, a next candidate onset is searched (usually the next sample). The additional criteria vary from method to method and usually fall into one of three classes. In the first class, an analysis of the temporal properties of the EMG burst is done: How long does the EMG remain above threshold level? In the second class, an amplitude analysis is done: Is there an increase in amplitude after the candidate onset? Finally, the third class uses an area analysis: How large is the area under the EMG burst curve? The following methods are variants of these general principles. It should be noted that intuitively obvious methods for determining the EMG onset do not work well. For instance, determining a high threshold and interpolating back to the zero level, calculating the first or second derivative of the EMG trace, and Woody filtering are all methods which significantly overestimate the onset.

Threshold voltage comparison. Most researchers determined EMG onset by simply comparing the rectified

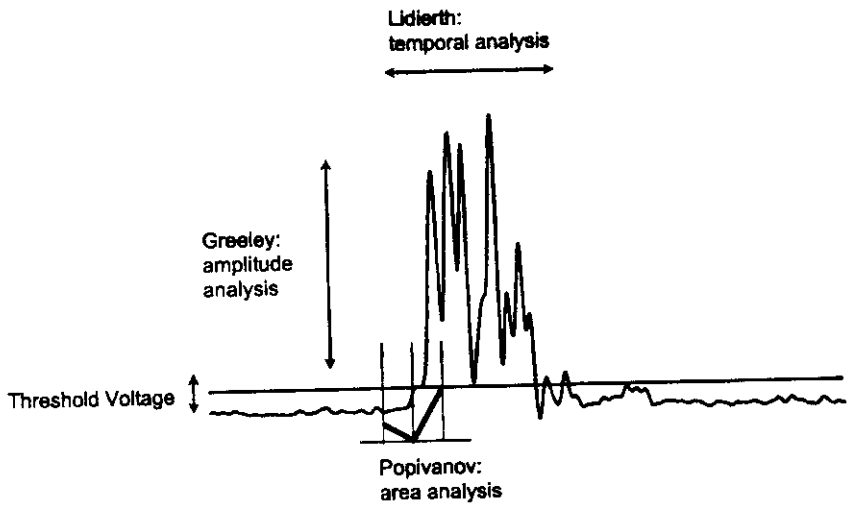

Figure 5. Example of a single-trial EMG recorded bipolarly from the palmar aspect of the right forearm, high-pass filtered at 20 Hz, full-wave rectified, and then low-pass filtered at 50 Hz. The annotations are intended to clarify the nature of the various EMG onset scoring methods.

signal against a preset threshold voltage. The advantages of this method are that it is extremely simple and may be performed both on line on the analog signal and off line on the digitized trace. It also has some drawbacks. The preset threshold level must be determined for each subject separately. If it is set too low, the method will be extremely sensitive to small preliminary bursts of EMG activity and, hence, will underestimate EMG onset. If it is set too high, it will overestimate the onset. Moreover, this method is susceptible to trial-to-trial differences in EMG rise time. Instead of this simple fixed threshold, a variable threshold can be used, which can be used for all subjects. This is done by calculating a running confidence interval of a fixed number of samples (N), starting at the stimulus, until a sample is detected that is larger than the upper confidence limit. The upper confidence limit can be adapted by a factor (F), so that the onset is said to be at sample I if, for the voltage V_I the following equation is satisfied:

$$V_I \geq M + F \cdot t_{(\alpha, N-1)} \cdot \frac{S}{\sqrt{(N-1)}}, \qquad (2)$$

where M is the mean, S is the standard deviation of the previous N samples, and t is the value from Student's t distribution at probability level α and $N-1$ degrees of freedom. This method was found to be always better than the fixed threshold voltage comparison, and the values for the parameters, assuming a sampling frequency of 200 Hz, are suggested to be set at $N = 15$, $\alpha = .01$, $t_{(.01, 14)} = 2.624$ and $F = 10.5$.

Greeley: Threshold voltage and amplitude analysis. Greeley (1984) proposed an on-line technique, implemented in hardware, for estimating the EMG onset. This technique is easily implemented in software and works as follows. If several successive points of rectified EMG within a fixed interval have larger amplitudes than does the preset threshold level, the onset is said to be detected. The software implementation can be slightly adapted in order to optimize the technique by specifying that n_1 samples out of n_2 equidistantly spaced samples in an interval of T msec should have larger amplitudes than does the threshold level, without the assumption of successivity. The threshold level should be determined by Equation 2 with $F = 11$. Appropriate values for n_1, n_2, and T are 3, 6, and 64, respectively (based on a sampling frequency of 200 Hz). Hence, for each candidate onset sample number k, $n_2 = 6$ equidistantly spaced samples are taken from the subsequent interval of $T = 64$ msec (12 samples), resulting in the series $k+1, k+3, \ldots k+11$. Of these series, $n_1 = 3$ samples must be above threshold level.

Lidierth: Threshold voltage and temporal analysis. The algorithm proposed by Lidierth (1986) relies on a threshold voltage comparison that is followed by an analysis of the EMG burst duration. If the rectified EMG has a larger voltage than does the threshold level, a check is made to determine that the burst has a minimum length of t_1 msec and a maximum length of t_2 msec. However, transient decreases below the threshold are ignored if they are not longer than t_3 msec. The threshold voltage can be calculated from Equation 2 with $F = 6$, and the optimal values for t_1 and t_2 are 90 and 1,000 msec, respectively. For fast EMG bursts, as they are mostly recorded, t_3 should be set to zero; for slow bursts, it can be set to a slightly higher value, for instance 10, but this will very much depend on the characteristics of the burst.

Popivanov: Threshold voltage and area analysis. Popivanov (1986) proposed a slightly more complicated procedure: If the rectified EMG exceeds a preset threshold value at sample I, cumulative sums are computed forward and backward for a preset interval C. This procedure is iterated for the next samples I with voltages V_I until

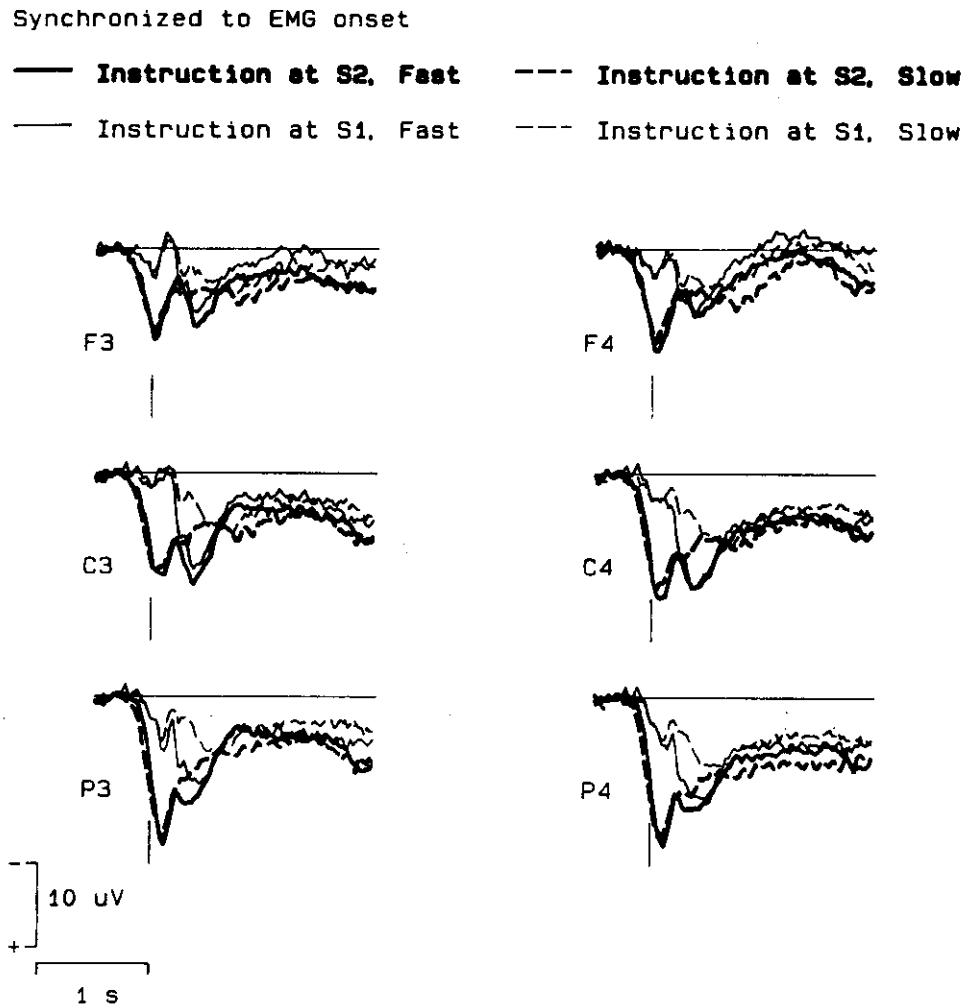

Figure 6. The same ERP traces as displayed in Figure 1, but averaged with the EMG onsets as the synchronization point. The EMG onsets were determined in the single trials using the method of Lidierth (1986), complemented by backward calculation to the preceding local minimum. The vertical lines below each trace indicate the instant of the average EMG onset.

the ratio of the cumulative sums becomes equal to or higher than a preset number A:

$$\sum_{i=I}^{I+C} V_i \geq A \cdot \sum_{i=I-C}^{I} V_i. \qquad (3)$$

The parameters for this method are very hard to optimize, but the best values were found to be $F = 7$ for the threshold voltage comparison, $C = 10$ msec, and $A = 2.5$.

All of these methods are able to determine the onset in more than 99% of all trials. In the plain application of the algorithms, the method of Greeley (1984) produced onsets that were closest to the visually determined onsets in the raw traces. However, one important point should be noted. These methods all rely on an initial threshold level comparison that, by its nature, detects the onset too late. Therefore, an obvious procedure to be applied, after the plain application of these methods, is to calculate backward from the onset produced by the methods to the preceding local minimum. This procedure greatly improves the results for all methods, and the algorithms of Greeley (1984) and Lidierth (1986) produce the most accurate results after application of this additional procedure. The average difference from the visually determined onset is then below 5 msec for both methods, which equals one sample point at a sampling frequency of 200 Hz. Lidierth's method produces a slightly lower average standard deviation around the mean, and its parameters are easier to adjust than is the algorithm suggested by Greeley. Hence, Lidierth's method can be expected to be ap-

plicable to a larger variety of different muscles and to be easier to adapt when special requirements must be met, as, for instance, in the case of partial responses.

The data displayed in Figure 1 were averaged again, taking as the point of synchronization the EMG onsets determined by the method of Lidierth (1986), complemented by backward calculation to the preceding local minimum. The resulting waveforms are displayed in Figure 6. These new averages can now form the basis of a new set of tests that include area measurement, peak picking, and PCA. When comparing the results to the stimulus-locked averages (Figure 1), it can be observed that the response-related peak previously denoted as P750 has become larger, and the stimulus-related P450 has become smaller. At the C3 electrode, which is over the motor area of the hemisphere contralateral to the response, the P750 is even larger than is the P450 in the new average. These results, which are supported by statistical tests not presented here, support the interpretation that the first peak after the stimulus is related to the processing of that stimulus, which is followed by a second peak related to the execution of the response. This in turn supports the presence of the existence of independent stimulus- and response-related ERP components in these data.

There is probably no single best algorithm for finding EMG onset automatically. The accuracy of a particular method may vary with the quality of the EMG recording, which in turn depends on various aspects of the muscle from which the activity is recorded and on data collection procedures. This survey is, therefore, not given to convince researchers to use this or that particular algorithm. However, it is felt that researchers should look into the accuracy of their method and report these accuracy findings in their papers. This survey and the study of van Boxtel, Geraats, et al. (1993) merely provide some useful suggestions to this end.

REFERENCES

BARRETT, G., SHIBASAKI, H., & NESHIGE, R. (1985). A computer-assisted method for averaging movement-related cortical potentials with respect to EMG onset. *Electroencephalography & Clinical Neurophysiology*, **60**, 276-281.

CASPERS, H., SPECKMANN, E.-J., & LEHMENKÜHLER, A. (1980). Electrogenesis of cortical DC potentials. In H. H. Kornhuber & L. Deecke (Eds.), *Progress in brain research: Vol. 54. Motivation, motor and sensory processes of the brain: Electrical potentials, behavior and clinical use* (pp. 3-15). Amsterdam: Elsevier.

CHAPMAN, R. M., & MCCRARY, J. W. (1995). EP component identification and measurement by principal components analysis. *Brain & Cognition*, **27**, 288-310.

DONCHIN, E. (1966). A multivariate approach to the analysis of average evoked potentials. *IEEE Transactions of Biomedical Engineering*, **13**, 131-139.

DONCHIN, E., CALLAWAY, E., COOPER, R., DESMEDT, J. E., GOFF, W. R., HILLYARD, S. A., & SUTTON, S. (1977). Publication criteria for studies of evoked potentials (EP) in man. In J. E. Desmedt (Ed.), *Progress in clinical neurophysiology: Vol. 1. Attention, voluntary contraction and event-related cerebral potentials* (pp. 1-11). Basel: Karger.

DONCHIN, E., & HEFFLEY, E. F. (1978). Multivariate analysis of event-related potential data: A tutorial review. In D. A. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potential research* (pp. 555-572). Washington, DC: U.S. Government Printing Office.

DONCHIN, E., RITTER, W., & MCCALLUM, W. C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. H. Koslow (Eds.), *Event-related brain potentials in man* (pp. 349-411). New York: Academic Press.

DONCHIN, E., SPENCER, K. M., & DIEN, J. (1997). The varieties of deviant experience: ERP manifestations of deviance processors. In G. J. M. van Boxtel & K. B. E. Böcker (Eds.), *Brain and behavior: Past, present, and future* (pp. 67-91). Tilburg: Tilburg University Press.

FABIANI, M., GRATTON, G., KARIS, D., & DONCHIN, E. (1987). Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in psychophysiology* (Vol. 2, pp. 1-78). Hillsdale, NJ: JAI Press.

GLASER, E. M., & RUCHKIN, D. S. (1976). *Principles of neurobiological signal analysis*. New York: Academic Press.

GREELEY, H. P. (1984). Real-time detection of precise EMG onset for synchronisation of an EEG averager. *Clinical Physics & Physiological Measurement*, **4**, 13-19.

KORNHUBER, H. H., & DEECKE, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflügers Archiv*, **284**, 1-17.

LIDIERTH, M. (1986). A computer based method for automated measurement of the periods of muscular activity from an EMG and its application to locomotor EMGs. *Electroencephalography & Clinical Neurophysiology*, **64**, 378-380.

MEYER, D. E., OSMAN, A. M., IRWIN, D. E., & YANTIS, S. (1988). Modern mental chronometry. *Biological Psychology*, **26**, 3-67.

MILLER, J., PATTERSON, T., & ULRICH, R. (1998). A jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, **35**, 99-115.

MÖCKS, J. (1988). Decomposing event-related potentials: A new topographic components model. *Biological Psychology*, **26**, 199-215.

MÖCKS, J., KÖHLER, W., GASSER, T., & PHAM, D. T. (1988). Novel approaches to the problem of latency jitter. *Psychophysiology*, **25**, 217-226.

O'BRIEN, R. G., & KAISER, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin*, **97**, 316-333.

PFURTSCHELLER, G., & KLIMESCH, W. (1991). Event-related desynchronization during motor behavior and visual information processing. In C. H. M. Brunia, G. Mulder, & M. N. Verbaten (Eds.), *Event-related brain research* (EEG Supplement 42, pp. 58-65). Amsterdam: Elsevier.

POPIVANOV, D. (1986). Computer detection of EMG edges for synchronization of movement-related brain potentials. *Electroencephalography & Clinical Neurophysiology*, **64**, 171-176.

RÖSLER, F., & MANZEY, D. (1981). Principal components and VARIMAX-rotated components in event-related potential research: Some remarks on their interpretation. *Biological Psychology*, **13**, 3-26.

RUCHKIN, D. S., & GLASER, E. M. (1978). Simple digital filters for examining CNV and P300 on a single trial basis. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potentials* (pp. 579-581). Washington, DC: U.S. Government Printing Office.

SAMS, M., ALHO, K., & NÄÄTÄNEN, R. (1984). Short-term habituation and dishabituation of the mismatch negativity of the ERP. *Psychophysiology*, **21**, 434-441.

SMULDERS, F. T. Y., KENEMANS, J. L., & KOK, A. (1996). Effects of task variables on measures of the mean onset latency of LRP depend on the scoring method. *Psychophysiology*, **33**, 194-205.

SRINIVASAN, R., TUCKER, D. M., & MURIAS, M. (1998). Estimating the spatial Nyquist of the human EEG. *Behavior Research Methods, Instruments, & Computers*, **30**, 8-19.

SUTTON, S., BRAREN, M., ZUBIN, J., & JOHN, E. R. (1965). Evoked potential correlates of stimulus uncertainty. *Science*, **150**, 1187-1188.

VAN BOXTEL, G. J. M., GERAATS, L. H. D., VAN DEN BERG-LENSSEN, M. M. C., & BRUNIA, C. H. M. (1993). Detection of EMG onset in ERP research. *Psychophysiology*, **30**, 405-412.

VAN BOXTEL, G. J. M., VAN DEN BOOGAART, B., & BRUNIA, C. H. M. (1993). The contingent negative variation in a choice reaction time task. *Journal of Psychophysiology*, **7**, 11-23.

VASEY, M. W., & THAYER, J. F. (1987). The continuing problem of false positives in repeated measures ANOVA in psychophysiology: A multivariate solution. *Psychophysiology*, **24**, 479-486.

WASTELL, D. G. (1977). Statistical detection of individual evoked responses: An evaluation of Woody's adaptive filter. *Electroencephalography & Clinical Neurophysiology*, **42**, 835-839.

WASTELL, D. G. (1981a). On the correlated nature of evoked brain activity: Biophysical and statistical considerations. *Biological Psychology*, **13**, 51-69.

WASTELL, D. G. (1981b). PCA and VARIMAX rotation: Some comments on Rösler and Manzey. *Biological Psychology*, **13**, 27-29.

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

WOODY, C. D. (1967). Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. *Medical & Biological Engineering*, **5**, 539-553.

NOTES

1. In the original data, the prestimulus CNV was about 3 μV more negative when the instruction was transmitted at the second stimulus, as opposed to the first. For all the present analyses, this difference was removed by calculating a new baseline over the 500-msec interval immediately preceding the stimulus. This inflated the difference in the factor instruction stimulus on the first positivity after the stimulus.

2. Actually, PCA is the more general technique, of which factor analysis is a variant.