Psychophysiology, 37 (2000), 127–152. Cambridge University Press. Printed in the USA.

COMMITTEE REPORT

Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria

T.W. PICTON,[a] S. BENTIN,[b] P. BERG,[c] E. DONCHIN,[d] S.A. HILLYARD,[e] R. JOHNSON, JR.,[f] G.A. MILLER,[g] W. RITTER,[h] D.S. RUCHKIN,[i] M.D. RUGG,[j] AND M.J. TAYLOR[k]

[a]Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, Canada
[b]Department of Psychology, Hebrew University of Jerusalem, Mount Scopus, Jerusalem, Israel
[c]Department of Psychology, University of Konstanz, Konstanz, Germany
[d]Department of Psychology, University of Illinois, Champaign, USA
[e]Department of Neuroscience, University of California at San Diego, La Jolla, USA
[f]Department of Psychology, Queens College, CUNY, Flushing, New York, USA
[g]Departments of Psychology and Psychiatry, University of Illinois, Champaign, Illinois, USA
[h]Department of Neuroscience, Albert Einstein College of Medicine, Bronx, New York, USA
[i]Department of Physiology, University of Maryland, Baltimore, USA
[j]Institute of Cognitive Neuroscience, University of London, England
[k]Centre de Recherche Cerveau et Cognition, Université Paul Sabatier, Toulouse, France

Abstract

Event-related potentials (ERPs) recorded from the human scalp can provide important information about how the human brain normally processes information and about how this processing may go awry in neurological or psychiatric disorders. Scientists using or studying ERPs must strive to overcome the many technical problems that can occur in the recording and analysis of these potentials. The methods and the results of these ERP studies must be published in a way that allows other scientists to understand exactly what was done so that they can, if necessary, replicate the experiments. The data must then be analyzed and presented in a way that allows different studies to be compared readily. This paper presents guidelines for recording ERPs and criteria for publishing the results.

Descriptors: Event-related potentials, Methods, Artifacts, Measurement, Statistics

Event-related potentials (ERPs) are voltage fluctuations that are associated in time with some physical or mental occurrence. These potentials can be recorded from the human scalp and extracted from the ongoing electroencephalogram (EEG) by means of filtering and signal averaging. Although ERPs can be evaluated in both frequency and time domains, these particular guidelines are concerned with ERPs recorded in the time domain, that is, as waveforms that plot the change in voltage as a function of time. These waveforms contain components that span a continuum between the exogenous potentials (obligatory responses determined by the physical characteristics of the eliciting event in the external world) and the endogenous potentials (manifestations of information processing in the brain that may or may not be invoked by the eliciting event).[1] Because the temporal resolution of these measurements is on the order of milliseconds, ERPs can accurately measure when processing activities take place in the human brain. The spatial resolution of ERP measurements is limited both by theory and by our present technology, but multichannel recordings can allow us to estimate the intracerebral locations of these cerebral processes. The temporal and spatial information provided by ERPs may be used in many different research programs, with goals that range from understanding how the brain implements the mind to making specific diagnoses in medicine or psychology.

Data cannot have scientific value unless they are published for evaluation and replication by other scientists. These ERP guidelines are therefore phrased primarily in terms of publication criteria. The scientific endeavor consists of three main steps, and these map well onto the sections of the published paper. The first step is the most important but the least well understood—the discovery of

Address reprint requests to: Terence W. Picton, Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, Ontario, M6A 2E1, Canada. E-mail: picton@psych.toronto.edu.

[1]In recent years, there has been a tendency to use the term "event-related potentials" to mean the endogenous potentials and to differentiate the event-related potentials from the (exogenous) "evoked potentials." However, this is not what the words mean logically and is certainly not the original meaning of the term "event-related potentials" as "the general class of potentials that display stable time relationships to a definable reference event" (Vaughan, 1969). This paper uses the term "event-related potentials" to include both evoked and emitted potentials. Evoked potentials can be either exogenous or endogenous (or both). Emitted potentials (always endogenous) can be recorded when a cognitive process occurs independently of any specific evoking event (e.g., when a decision is made or a response initiated).

some new way of looking at the world. This step derives from creative processes that are probably similar to those used to solve problems in other domains (Langley, Simon, Bradshaw, & Zytkow, 1987). Unfortunately, this step is often the least documented aspect of a scientific study. Wherever possible, the introduction to a paper should therefore try to describe how the authors arrived at their hypotheses as well as simply stating them. The second step in the scientific process involves the design of an experiment or a set of experiments to test the hypotheses. Setting up the experiments to provide information that convincingly tests the hypotheses and rules out other competing hypotheses requires clarity of thought and elegance of design. The third step involves the careful testing of the hypotheses. Scientific statements are valid as long as they are not falsified when tested (Popper, 1968). The methods and the results of an experimental paper provide the details of how this testing was carried out and what results were obtained. Because the results of an experimental test may be the consequence of a failure in the method or of noise in the measurement, the authors must persuade the reader that the measurements were valid, accurate, and reliable. The discussion section of the paper returns to the creative part of science. The new findings must be related to other published results. Views of the world that have been clearly falsified by the new findings should be summarized. New views justified by the findings must be clearly worked out and formulated for future testing.

The compilation of the present guidelines was initiated by John Cacioppo when he was president of the Society for Psychophysiological Research in 1993. A complementary set of guidelines exists for recording the EEG in research contexts (Pivik et al., 1993). Draft ERP guidelines were then proposed, discussed, and revised by the authors of this report. The paper also benefited from the comments and suggestions of four anonymous reviewers. These ERP guidelines update those deriving from the International Symposium on Cerebral Evoked Potentials in Man held in Brussels in 1974 (Donchin et al., 1977). Since then, several sets of guidelines have been developed for recording exogenous evoked potentials in clinical contexts (American Encephalographic Society, 1994a; Halliday, 1983), but none of these has specifically considered ERPs in relation to normal and abnormal human cognition. Although put together under the aegis of the Society for Psychophysiological Research, these ERP guidelines should apply to papers published anywhere. It is the scientist's responsibility to select a publication venue that can communicate his or her findings to the appropriate audience and to ensure that the rationale, method, results, analysis, and conclusions of the study are presented properly.

The guidelines or recommendations are stated in the titles to each subsection of this paper. The paragraph or paragraphs following these titles explain the committee's reasons for the guidelines and provide advice and suggestions about ERP procedures that can be used to follow them. Although mainly addressed to scientists who are beginning to use ERPs to study cognition, these guidelines should help all who work with ERPs to record their data and communicate their results more effectively. The guidelines use the following codes to indicate committee agreement: "must" indicates that the committee agreed unanimously that the guideline applies in all cases, and "should" indicates that the committee agreed unanimously that the guideline applies in most situations (and that the investigator should be able to justify why the guideline is not followed). Guidelines about specific techniques clearly apply only if this particular technique is used. Some of the guidelines, such as those concerning the rationale for the study and the discussion of the results, are not limited to ERP studies, although they are particularly important in this field.

A. Formulation of the Study

(i) The Rationale for the Study Must Be Presented Clearly

The rationale for an experimental study usually derives from a review of the literature, which either shows important gaps in our knowledge or leads to a reinterpretation of known facts in terms of a new theory. These two situations require further experiment, either to fill in the gaps or to test the new theory. It is essential to communicate the rationale clearly to the readers so that they may see the purpose and significance of the study. It is not sufficient to state that the experiments are intended to clarify something in physiology or psychology without specifying what is to be clarified and why such clarification is important. Because ERP studies relate to both physiology and psychology, terms and concepts specific to one field should be explained (e.g., linguistic categories, chemicals used to evoke olfactory ERPs).

(ii) The Hypotheses of the Experiment(s) Should Be Stated Clearly

Specific hypotheses and predictions about the experimental results must be derived from the rationale. These hypotheses and predictions should be stated in positive terms even though the statistical tests will examine null hypotheses. The first chapter of the *Publication Manual of the American Psychological Association* (American Psychological Association, 1994) provides useful advice for setting out the rationale and hypotheses for an experimental study.

Although true for all areas of research, loosely motivated "shots in the dark" are particularly dangerous in studies in which data are abundant. The overwhelming amount of ERP data along the time and scalp-distribution dimensions can easily lead to incorrect post hoc conclusions based on trial-and-error analyses of multiple time epochs and electrode sites. Huge arrays of data make it easy to obtain "significant" results that are not justified in theory or reliable on replication. Hypotheses should therefore describe particular ERP measurements (e.g., that the experimental manipulation will increase the latency of the P300 wave) rather than nonspecific ERP changes (e.g., that the experimental manipulation will change the ERP in some way).

(iii) As a General Rule, Tasks Should Be Designed Specifically to Elicit the Cognitive Processes Being Studied

If relating attributes of the ERP to cognition is desired, the ERP should be recorded in an experimental paradigm that can be interpreted in terms of the information processing invoked and exercised in the paradigm. To demonstrate ERP concomitants of particular cognitive processes, the ERPs should be recorded when these processes are active (and their activity can be shown through behavioral measurements). It is unlikely (although possible) that an ERP measurement recorded when a subject performs a particular task will turn out to be a specific marker for a cognitive process that does not occur during the task. This result would require that whatever affects the cognitive process independently affects the ERP measurement.

Experimental paradigms that have been well studied, and for which well-developed cognitive models are available, provide a good framework for the study of ERPs. Standard paradigms used by investigators of memory, attention, or decision making will more likely lead to useful mappings of ERP data on cognitive models than new paradigms. However, novel paradigms can yield exciting and useful results, provided the investigators can also present a carefully developed model of the paradigm in information-processing terms.

Historically, the most frequently used ERP paradigm has involved the detection of an improbable target stimulus in a train of standard stimuli. This "oddball" paradigm elicits large ERP components, and provides useful information about how the brain discriminates stimuli and evaluates probability. This paradigm can be adapted to the study of other cognitive processes such as memory and language. However, it is often better to use paradigms more specific to these processes than to force the oddball paradigm to fit the processes. Nevertheless, many other paradigms share characteristics of the oddball task, and it is essential to consider whether the ERPs recorded in these paradigms can be interpreted more parsimoniously in terms of oddball parameters (e.g., probability and discriminability) than in terms of other processes. To prevent confounding the effects of probability with other experimental variables, the investigator should therefore keep the probabilities of stimulus/response categories constant within and across recording conditions.

A final aspect of this recommendation is that the tasks should be adapted to the subjects studied. When studying language in children, for example, researchers must take into consideration the language level of the subjects, and not use vocabulary that would be too advanced for the younger children. When studying subjects with disordered cognition, it is probably worthwhile to adjust the difficulty of the task to their cognitive level. If the subjects cannot perform a task, it is difficult to determine if the absence of particular ERPs are associated with the cause of their cognitive disorder or simply the result of the task not being performed. The tasks need to be of shorter duration for clinical and developmental studies than for ERP studies in normal young adults, because attention span is generally shorter in clinical patients or children. When studying clinical groups, the experimenter can decide to keep the task the same or to adjust the task so that the performance is equivalent between the clinical patients and the normal control subjects (e.g., Holcomb et al., 1995). When the stimuli are the same, the results bear more on differences in sensory processing; when the difficulty is the same, the results are more related to cognitive processes. A related problem concerns whether to compare ERPs only on trials for which performance is correct. Although it is probably best to compare ERPs for both correct and incorrect performance across the subject groupings, this comparison is often impossible unless the task is adjusted so that the accuracy of performance is similar across the groups. These and other issues of how to compare groups with different abilities have been discussed extensively by Chapman and Chapman (1973).

(iv) The Subject's Behavior in the Experimental Paradigm Should Be Assessed

When using ERPs to evaluate the cerebral processes that occur during cognition, the experimenter should usually monitor behavioral responses at the same time as the physiological responses are recorded, provided that this comonitoring can be done without excessive artifactual contamination of the recordings. In many perceptual tasks, a simple motor response to a detected target provides a measure of the speed and accuracy of perceptual performance. In memory tasks, simple yes-no recognition performance measures are helpful not only in monitoring that encoding and retrieval are occurring, but also in averaging ERPs at encoding on the basis of later retrieval. In general, the more behavioral data that are available, the more readily the psychophysiological measures can be evaluated within the context of an information-processing model. The type of behavioral data collected will depend on the type of correlations that may be hypothesized. For example, if the investigators want to consider processing resources they should obtain data for a receiver-operating curve, and if they want to address speed and accuracy, they should have clear behavioral data showing the effects of changing response speed on performance.

In some experiments, ERPs are used as a relatively unobtrusive monitor of cerebral processes without the need for recording overt responses. A classic example is measuring the ERPs to unattended stimuli. This measurement can indicate how these stimuli are processed without the need to ask for overt responses to the unattended stimuli, which could clearly disrupt the focus of attention. In studies of automatic processes, ERPs can be used to assess the brain's responses to stimuli without these stimuli evoking (either perceptually or electrically) controlled cognitive responses. For example, the mismatch negativity is best recorded when the subject is not attending to the auditory stimuli. When the subject attends to the stimuli, the mismatch negativity is difficult to recognize due to the superimposition of other ERP components such as the N2b or P300. When the subject does not attend to the stimuli, a description of what the subject is doing (e.g., reading a book) must be provided, and where possible this activity should be monitored. It is usually better to have the subject perform some task rather than just listen passively. In cases wherein the ERPs are recorded without any attention to the stimuli or behavioral responses, additional studies recording only behavioral responses (or both behavioral and electrical responses) can be helpful in determining the timing and the difficulty of sensory discrimination. For example, the investigator must demonstrate that the stimuli are equally difficult to discriminate before concluding that particular types of deviance elicit mismatch negativities with different latencies or amplitudes (Deoull & Bentin, 1998).

ERP studies of language (Kutas, 1997; Kutas & Van Petten, 1994) provide a clear example where recording behavioral responses at the same time as the ERPs may be counterproductive. Many language processing activities occur without explicit relation to any assigned task, and many studies of semantic processing have been performed in the context of general instructions to "read silently" or "listen" (which do not yield accuracy or reaction time [RT] data). Indeed, many tasks in behavioral psycholinguistics (e.g., lexical decision) are really secondary tasks that do not occur in natural language processing. One clear benefit of the ERP method is that such artificial tasks can be dropped. A detriment to including such tasks is that they elicit decision-related P300s, which may obscure other ERP components such as the N400 wave (see Kutas & Hillyard, 1989; Kutas & Van Petten, 1994). However, even when no overt responses are being made, it is still important to specify as much as possible what the subject is doing during the ERP recordings. Because it is often important to acquire accuracy and RT data in order to compare ERP results with the behavioral literature using tasks such as lexical decision and naming (which is incompatible with ERP recording due to artifacts caused by tongue movements and muscle activity), a useful strategy has been to conduct a behavioral study first, followed by an ERP study with the same stimuli. In other cases, it has been of some interest to compare ERP data obtained under general "read" or "listen" instructions with those obtained with an overt task that forces attention to some aspect of the stimuli. Such comparisons can reveal which aspects of stimuli are processed automatically versus those that are optional. For instance, these comparisons have shown that sentence semantic congruity effects occur independently of the assigned task (Connolly, Stewart, & Phillips, 1990), but that rhyming effects for visually presented word pairs occur only when rhyme monitoring is the assigned task (Rugg & Barrett, 1987).

(v) Subject Strategies Should Be Controlled by Instruction and Experimental Design, and Should Be Evaluated by Debriefing

Perhaps the most difficult variables to bring under experimental control are the cognitive strategies and mental processes underlying the performance of the subject. It is therefore essential to describe in detail how the subjects are instructed about the experimental situation and task. In situations in which subjects are responding actively to the stimuli, the report should clarify whether the subjects have been told to emphasize response speed or accuracy, and which motivating instructions and/or tangible rewards were used. In conditions in which subjects are asked to ignore auditory or somatosensory stimuli, it is generally desirable to give them a task to perform (e.g., read a book, solve a puzzle) in order to have some control over what the subject actually does. Whenever possible it is advisable to use a task with measurable consequences so that the degree to which the subjects actually undertake the assigned task can be assessed. A general description of the task situation such as "passive listening" or "reading" is not adequate in experiments in which state variables could affect the ERPs.

In general, explicit and consistent instructions to subjects can minimize the "subject option" (Sutton, 1969) to react to the situation in an idiosyncratic and uncontrolled fashion. Debriefing the subjects after the experiment can provide information about how they viewed the task and what cognitive strategies they used. Debriefing can be done by simply asking subjects how they performed the task or by using a formal questionnaire that describes the possible strategies that might have been used. Not to ask one's subjects what they were doing in an experiment indicates a faith in one's experimental paradigm that may not be justified. Relations among the ERP measurements, the behavioral data, and these subjective reports can help the investigator interpret what was going on during the task and to test specific hypotheses about how the subjects interpreted the task.

(vi) The Ordering of Experimental Conditions Must Be Controlled and Specified

The way in which the trials for each of the different experimental conditions are put together into blocks must be described clearly. Different experimental conditions can occur in separate blocks or can be combined within blocks. For example, attention can be studied by having subjects attend to stimuli in one block of trials and ignore them in a separate block of trials (block design), or by having subjects attend to some of the stimuli in one block while ignoring others in the same block (mixed design). The amount of time required for each block of trials and the sequence in which the blocks are delivered must be specified. Many aspects of behavior and many components of the ERP change over time, and such changes must not be confounded with the experimental manipulations. It is therefore advisable to balance experimental conditions over time either within each subject or across subjects. Time is but one of many factors that must be controlled. Cognitive behavior is very flexible and heavily influenced by context. Because the general working hypothesis is that different cognitive processes are associated with different ERPs, cognitive electrophysiological studies should exert the same scrupulous control of experimental design as required in experimental psychology when studying cognition.

B. Subjects

(i) Informed Consent Must Be Documented

Informed consent is essential for any research with human subjects (Faden, Beauchamp, & King, 1986). In the case of patients with clinical conditions that might impede informed consent, the experimenter should consider published guidelines for obtaining substitute consent from family or caretakers (e.g., Keyserlingk, Glass, Kogan, & Gauthier, 1995). When the subjects are under 18 years old, the investigator should obtain informed consent from the child's guardians and provide information to the child at a level that the child might understand (Van Eys, 1978). Academic and clinical institutions specify how the rights of human subjects are protected and have committees to approve research protocols and to monitor the research as it proceeds. Investigators must follow the instructions of these committees.

(ii) The Number of Subjects in Each Experiment Must Be Given

The number of subjects in an experiment must be sufficient to allow statistical tests to demonstrate the experimental effects and to support generalization of the results. The number of subjects required to demonstrate a particular size of effect can be estimated using evaluations of statistical power. In addition to being sufficient to demonstrate an experimental effect, the sample size must also be large enough to represent the population over which the results are to be generalized. Because ERP data can vary considerably from one subject to the next, it is often advisable when using small numbers of subjects to sample from a population as homogenous as possible, for example, in terms of age, gender, educational level, and handedness. This method can, of course, limit the generalizability of the results.

The total number of subjects recruited and the reasons for not being able to include all of them in the final results (e.g., artifacts, incomplete recordings) should be described. Compared with studies of normal young adults, developmental and clinical studies often have a higher number of subjects who cannot be tested successfully. In these studies, it is particularly important to document the reasons (e.g., lack of cooperation, inability to understand or complete the task), because these reasons may have some bearing on what can be generalized from the results.

(iii) The Age Ranges of Subjects Participating in ERP Experiments Must Be Provided

Because many ERPs change with age, the mean and range of subject ages must be provided. The normal adult age range for most ERP studies can be considered as 18–40 years.[2] When comparing ERPs across groups of subjects, ages should be balanced across the groups (unless, of course, age is one of the variables under study). Subjects older than the age of 40 years should be stratified into decades.

In subjects younger than the age of 18 years, significant ERP changes can occur over short time periods (Friedman, 1991; Stauder, Molenaar, & van der Molen, 1993; Taylor, 1988, 1995). The younger the children, the more marked are these age-related changes. Thus, it is important to use narrower age ranges than for adults. In infants and young children (<24 months) researchers should use 1-month ranges, recording at several points in time (e.g., 6 months, 12 months, and 18 months) rather than averaging across even a few months. In older children, 1-year age groupings are recommended, although 2-year groupings are acceptable over the age of 8 years and 3-year groupings are acceptable among teenagers.

[2] Significant differences can occur even within the age range of 18–40 years. In group studies it is sometimes helpful to use age as a covariate to decrease the noise levels across groups (provided there is no correlation between age and the experimental groups).

(iv) The Gender of the Subjects Must Be Reported

Because gender affects many electrophysiological measurements, the investigator must report how many of the subjects were male and female, and must ensure that any group effects are not confounded by differences in the female/male ratios across groups. When studying normal subjects, the investigator should generally use either a similar number of female and male subjects or subjects of one gender only. It is often worthwhile to include gender as an experimental variable. If the experiment compares normal subjects with subjects with a clinical disorder that is more common in one gender, the male/female ratios should be approximately equivalent across the two groups.

(v) Sensory and Motor Abilities Should Be Described for the Stimuli Being Presented and the Responses Being Recorded

This recommendation is to ensure that subjects can perceive the stimuli normally. For most studies of normal young subjects, it is sufficient to document that all subjects reported normal hearing or vision (with correction). Such self-report is usually correct about normal sensory ability. However, the accuracy of self-report will depend on the type of questions asked. The answer to "Do you have normal hearing?" is much less informative than answers to a set of questions about hearing under different situations (Coren & Hakstian, 1992).

In experiments designed specifically to evaluate perceptual function, particularly in studies of disordered perception, more intensive evaluations should be used to clarify what is normal or to categorize levels of abnormality. For auditory stimuli, subjects should be screened for normal hearing at 20 dB HL at the frequencies tested. For visual stimuli, acuity should be measured (with refractive correction) at a distance appropriate for the stimuli used. Because most visual stimuli are presented at close distances, acuity normally would be checked using Jaeger rather than Snellen charts. If stimulus color is manipulated during the experiment, color vision should be checked (e.g., using one or several Ishihara plates). Unfortunately, there are no widely accepted quantitative screening tests for normal somatic, taste, or smell sensations.

When subjects are making motor responses during the experimental paradigms, the investigator should provide some basic description of the subjects' ability to perform the task. It is usually sufficient to ensure that the subjects report no history of weakness. In all studies using motor responses, the handedness of the subject should be reported and preferably measured using a validated questionnaire.

(vi) The Subjects' Cognitive Abilities Relevant to the Tasks Being Studied Should Be Described

The experimenter should provide some basic assessments of the subjects' ability to perform the tasks being evaluated. In normal subjects, the educational level is a reliable indicator of general cognitive abilities, and descriptions of the subjects such as "undergraduate students" is sufficient. However, this approach is inadequate in the context of clinical patients, children, and the elderly, for whom more specific evaluations should be provided. For example, mental status tests should be used when evaluating the ERPs of demented patients, standardized reading assessments when ERP paradigms that require reading are used in children, and neuropsychological tests of memory when ERPs are used to study memory disorders in the elderly.

(vii) Clinical Subjects Should Be Selected According to Clear Diagnostic Criteria and the Clinical Samples Should Be Made as Homogeneous as Possible

The selection criteria for clinical subjects should be explicitly stated. The *Diagnostic and Statistical Manual of the American Psychiatric Association* (American Psychiatric Association, 1994) provides criteria for most psychiatric disorders. Diagnostic criteria for neurological disorders can be found in the relevant literature. When the clinical disorder is heterogeneous (e.g., schizophrenia, attention deficit disorders), the experimenter should attempt to limit the subjects to one of the various subtypes of the disorder or to stratify the patient sample according to the subtypes. The sample should also be made as homogeneous as possible in terms of both the duration and the severity of the disease process. It is never possible to devise pure patient groupings. Nevertheless, some attempt should be made to limit heterogeneity and any residual sources of heterogeneity should be described. In addition, the sample should be characterized carefully with respect to demographic and psychometric variables. For example, in a study of patients with dementia of Alzheimer type, the investigator should include information about the age and gender of each subject, along with data on current mental status (e.g., Mini-Mental State Examination), premorbid intelligence (e.g., National Adult Reading Test), and memory function (e.g., selected subtests of the Wechsler Memory Scale). For patients with focal brain lesions, such data should also include detailed information about the location and nature of the lesions.

(viii) Medications Used by Subjects Should Be Documented

In ERP studies of normal subjects, the investigator should make sure that the subjects are not taking prescription medications that may affect cognitive processes. It is probably also worthwhile not to use subjects who have taken alcohol or other recreational drugs within the preceding 24 hours. Because clinical patients are commonly treated with medications, it is often difficult to disentangle the effects of the clinical disorder from the effects of the treatment. Wherever possible, some control for medication should be attempted. In some cases unmedicated patients can be studied. If the patients have various dosages of medication, the level of medication should be considered in the statistical analysis, or preferably in the experimental design (e.g., selecting different subgroups of patients with different medication levels). Unfortunately, it is not possible to use an analysis of covariance to remove the effects of different medication levels (or other variables) from other group differences (Chapman & Chapman, 1973, pp. 82–83; Miller, Chapman, & Isaacks, submitted). An analysis of covariance can be used to reduce the variability of measurements in groups that vary randomly on the variable used as a covariate. However, if groups differ on each of two variables, covarying out the effects of one variable will distort any measurement of the effect of the other.

(ix) In Clinical Studies, Control Subjects Should Be Chosen so That They Differ From the Experimental Subjects Only on the Parameters Being Investigated

The selection criteria for the control subjects should be stated clearly, as should the variables on which the control subjects and patients have been matched. In general, the groups should be matched for age, gender, socioeconomic status, and intelligence. The premorbid intelligence of the patient group may be compared with the actual intelligence of the control group using educational level or some more formal psychological assessment such as the National Adult Reading Test. Both control and experimental subjects should be evaluated on standardized behavioral, psycholog-

ical, or neuropsychological tests. These tests should document how the patients are equivalent to the control subjects in some areas but not in others. Exclusion criteria must also be stated explicitly and applied to both clinical and control groups. In many cases, a healthy control group may not be sufficient. Clinical patients with disorders different from those of the patients being studied are often better controls than completely normal subjects. For example, in studies of the effects of a specific focal brain lesion, a helpful control group will consist of patients with lesions of a similar etiology but outside the brain region of interest.

C. Stimuli and Responses

(i) The Stimuli Used in the Experiments Must Be Specified in Sufficient Detail That They Can Be Replicated by Other Scientists

The stimuli must be described accurately in terms of their intensity, duration, and location. The guidelines for clinical evoked potential studies (American Electroencephalographic Society, 1994a) provide clear descriptions of the simple stimuli used in such studies. Where possible, similar descriptions should be provided for the stimuli used in all ERP studies. An extensive description of the different stimuli that have been used in ERP studies and the way in which these stimuli are described and calibrated is given in Regan (1989, pp. 134–155). Investigators using video displays to present visual stimuli can consult Poynton (1996). All stimuli should be calibrated in terms of their intensity and timing using appropriate instrumentation (e.g., a photoreceptor for visual stimuli and a microphone for auditory stimuli). It is important to realize that the presentation of a stimulus in one modality may be associated with stimulation in another modality and the effects of this other stimulus should be masked. For example, airpuff or strobe flash stimuli are often associated with simultaneous acoustic artifacts. If decibels are used to describe intensity, it is essential to provide the reference level because decibels are meaningless without the reference. Common references in the auditory system are sound pressure level (a physical reference), hearing level (relative to normal hearing) and sensation level (relative to the individual's threshold).

(ii) The Timing of the Stimuli Must Be Described.

The minimum temporal parameters that should be described are stimulus duration and the intervals between the stimuli. If the experiment involves trials containing more than one stimulus, the interval between the trials must also be given. The experimenter should clarify whether the intervals are from onset to onset (stimulus onset asynchrony) or from the offset of the preceding stimulus (or trial) to the onset of the next (interstimulus or intertrial interval). If the subjects are expected to execute a motor response or to provide a verbal response, the timing of these responses with respect to the stimuli should be specified. The structure of the stimulus sequences is also an important attribute of the experimental design. Thus, investigators should specify whether trials are initiated by the subject or by the experimenter. They should also specify the rules by which the stimulus sequences are generated (e.g., completely random stimuli according to set probabilities, or random stimuli with the proviso that no two targets occur in succession). Because human subjects are capable (consciously or unconsciously) of picking up regularities and rules of stimulus sequences, subtle changes in these can lead to ERP effects.

Timing is a particular problem when using a video display. The investigator should check the timing of these stimuli using a photoreceptor. An apparently continuous stimulus is actually composed of a series of discrete pulses as the raster process activates the region of the screen beneath the receptor during each screen refresh. The conversion of this stimulus into a sustained visual sensation is described in Busey and Loftus (1994, particularly Appendix D). Because the stimulus is composed of discrete pulses, there are often discrepancies between the programmed onset and duration of the stimulus and the actual stimulus parameters.

(iii) Aspects of the Stimuli Relevant to the Cognitive Processes Being Examined Should Be Described

When words or other complex stimuli are used, they should be selected keeping in mind which of their properties might affect their processing. Because the number of trials necessary to record ERPs is usually larger than the number of trials needed for behavioral measurements,[3] extensive manipulation of stimulus parameters during an ERP paradigm is usually not possible and extra care during stimulus selection is required. Factors such as familiarity, word frequency, and meaning are of paramount importance when studying the ERPs to words. If not manipulated in the experiment, these factors should be controlled rigidly and kept constant across conditions. Whenever possible, the stimuli should be rotated across conditions to prevent any inadvertent confounding of some stimulus parameter with the experimental manipulation. All the relevant stimulus selection criteria and characteristics should be reported (such as the mean and range of the number of letters, phonemes and syllables composing the words, word frequency, and, where relevant, the degree of semantic relatedness of the words). If images or pictures are used, the investigator should specify whether they are drawings or photographs, black and white or color. A figure showing a sample image or images is worth more than many words of description. For auditory stimuli, particularly when words are used, provide the duration (the range, mean, and standard deviation) and the obvious measures such as intensity (root-mean-square [RMS]), frequency, and male or female voice.

(iv) Responses Made by the Subjects Should Be Described

In many ERP paradigms, subjects make overt responses while their ERPs are being recorded. In some paradigms, the ERPs are recorded in reference to these responses instead of or in addition to the sensory stimuli. The investigator must clarify the stimulus-response mapping required during the paradigm (e.g., which button was pressed by which finger in response to which kind of stimulus) and how this response was manipulated. The nature of the response should be described in terms of the limb used to make the response and the type of movement made. When the research focuses on motor-related responses, the force, speed, and extent of the movements should also be measured and reported.

D. Electrodes

(i) The Type of Electrode Should Be Specified

Because electrodes act as filters, they should be chosen so as not to distort the ERP signals being measured. Nonpolarizable Ag/AgCl electrodes can accurately record very slow changes in potential (e.g., Kutas, 1997; Rösler, Heil, & Hennighausen, 1995),

[3]Clear behavioral measurements can be obtained sometimes on single trials (e.g., yes-no decisions about whether a stimulus was perceived) but this is usually not possible with ERPs. In behavioral studies, using more subjects often compensates for the smaller number of trials per subject. This method is not carried out in ERP studies because of the time involved in preparing the subject for the recording.

although precautions must be taken to eliminate drift when ultraslow (less than 0.1 Hz) potentials are recorded (Tassinary, Geen, Cacioppo, & Edelberg, 1990). Such slow drifts in the polarization of the electrodes can be estimated using linear regression techniques and then subtracted from the recordings (Hennighausen, Heil, & Rösler, 1993; Simons, Miller, Weerts, & Lang, 1982). For potentials of higher frequency, a variety of different electrode materials (e.g., gold, tin) may be used. Depending on the electrode material, the surface area of the electrode and the input-impedance of the amplifier, many electrodes will attenuate the low frequencies in the recorded signal (Picton, Lins, & Scherg, 1995). Because many modern EEG amplifiers with high input impedance use very low electrode currents, even these polarizable electrodes can often be used to record slow potentials without distortion. Unfortunately, it is difficult to calibrate the frequency response of the electrode–skin interface and for frequencies less than 0.1 Hz, nonpolarizable electrodes are recommended. The low-frequency response of an electrode can be estimated in situ by observing the signals recorded during sustained eye movements (Polich & Lawson, 1985). The investigator could also estimate the transfer function of the electrodes by measuring the potentials when the eyes follow pendular movements with the same amplitude but different frequencies.

(ii) Interelectrode Impedances Must Be Reported

The recording electrodes are affixed to the surface of the scalp. Subcutaneous needle electrodes should not be used for ERPs because of the risk of infection. The connectivity of the electrode to the scalp is measured by passing very low currents through the electrodes and measuring the impedance to the flow of current. These measurements tell the experimenter four things: how accurately the amplifier will record the potentials, the liability of the electrode to pick up electromagnetic artifacts, the ability of the differential amplifiers to reject common-mode signals, and the intactness of the skin underlying the electrode. For the amplifier to record accurately, the electrode impedance should be less than the input impedance of the amplifier by a factor of at least 100. The higher the impedance of an electrode the greater the effect of electromagnetic fields (e.g., line noise, noise from electric motors, video display systems) on the recording. These effects are caused mainly by currents induced in the electrode circuits. These currents vary with the area surrounded by the circuit (and hence can be reduced by braiding the electrode wires together). Inequalities in the electrode impedance between the two inputs to a differential amplifier will reduce the ability of the amplifier to reject common mode signals (Legatt, 1995). Finally, electrode impedance measures the intactness of the skin and thus its ability to generate skin potentials. Cephalic skin potentials are large, slow potentials that occur when the autonomic nerves and sweat glands in the skin are activated by heat or arousal (Picton & Hillyard, 1972). They are most prominently recorded from the forehead, temples, neck, and mastoid regions.

The interelectrode impedance measured at some frequency within the ERP range (e.g., 10 Hz) should be reduced to less than 10 kΩ by abrading the skin. Electrode–scalp interfaces with higher impedances may yield adequate recordings when amplifiers with high-input impedances are used and when good common mode rejection is available (Taheri, Knight, & Smith, 1994). These systems can be used to record ERPs, but great care must be taken in interpreting slow potentials, because skin potential artifacts can occur easily. To eliminate skin potentials, the impedance at the scalp–electrode junction will need to be reduced (by abrasion or skin puncture) to less than 2 kΩ. Puncturing the skin with a fine sterile disposable needle or lancet is usually less painful than abrasion and leaves visible marks less frequently. The investigator must balance the need for reducing skin potentials with the necessity of preventing any possibility of infection. Impedances of less than 2 kΩ occur only if the skin layer is effectively breached, which clearly increases the risk of infection. Special care must be taken to prevent the transmission of infective agents via the instruments used to reduce the impedance or by the electrodes. Disposable instruments must be used to abrade or puncture the skin, and electrodes must be disinfected properly between subjects. Previously published guidelines for reducing the risk of disease transmission in the psychophysiology laboratory (Putnam, Johnson, & Roth, 1992) must be followed scrupulously.

(iii) The Locations of the Recording Electrodes on the Scalp Must Be Described Clearly

Whenever possible standard electrode positions should be used. The most helpful standard nomenclature is the revision of the original 10-20 system to a 10-10 system as proposed by the American Electroencephalographic Society (1994b). Electrodes should be affixed to the scalp with an accuracy of within 5 mm. Unfortunately there is no standardized placement system for electrode arrays having large numbers of electrodes. The 10-20 system describes 75 electrode locations but does not state which of these should be used in a montage containing a smaller number of channels or how to locate electrodes if more than 75 channels are to be used. In general, we recommend using approximately equal distances between adjacent electrodes, and placing electrodes below as well as above the Fpz-T7-Oz-T8 equator.

The exact locations of the electrodes can be determined relative to some fiducial points (such as the nasion, inion, and preauricular points defined in the 10-20 system) using a three-dimensional digitizer (Echallier, Perrin, & Pernier, 1992). These positions can then be compared with the locations of the 10-20 system by projecting these locations onto a sphere (Lütkenhöner, Pantev, & Hoke, 1990). This projection onto a sphere is necessary for spherical spline interpolations and for source analysis using spherical head models. Various relations between the 10-20 electrode system and the underlying brain have been evaluated (Lagerlund et al., 1993; Towle et al., 1993).

The newly emerging dense-array systems that allow placement of 128 or 256 electrodes present challenges for specifying electrode placement as the number of electrodes clearly exceed the capacity of the 10-20 system. Whatever nomenclature is used, it is important to identify within the dense array landmark electrodes that correspond to the standard sites within the 10-20 system.

(iv) ERPs Should Be Recorded Simultaneously From Multiple Scalp Electrodes

In some cases, simple evoked potentials (e.g., the brainstem auditory-evoked potentials) can be adequately examined for clinical purposes using a single recording channel. However, for most ERPs, simultaneous recording from multiple electrode locations is necessary to disentangle overlapping ERP components on the basis of their topographies, to recognize the contribution of artifactual potentials to the ERP waveform, and to measure different components in the ERP that may be optimally recorded at different scalp sites. As examples, recording from parietal electrodes in addition to frontocentral electrodes can help distinguish between motor and re-afferent somatosensory potentials; time-locked blinks are easily distinguished from the late positive wave by being maximally recorded directly above the eyes; and the mismatch negativity can

usually be distinguished from the N2 wave by its polarity reversal in ear or mastoid electrodes. Many early studies of the endogenous ERPs used midsagittal electrodes (Fz, Cz, Pz) to make some important distinctions among ERP components. However, such locations are not appropriate for the visual-evoked potentials or for any lateralized ERPs. Any developmental studies should use both lateral and midline recording electrodes. Midline electrodes are important (for comparison with both older papers and older subjects), but in developmental studies the largest age-related changes are often seen in lateral electrodes (e.g., Taylor & Smith, 1995).

The optimal number of recording channels is not yet known. This number will depend on the spatial frequencies that are present in the scalp recordings (Srinivasan, Tucker, & Murias, 1998), provided that such frequencies are determined by the geometry of the intracerebral generators and not by errors in positioning the electrodes or modeling the impedances of the head. The proper use of high-density electrode arrays requires techniques for accurately measuring the location of the electrodes and for handling the loss of one or several recording channels through poor contact. Variance in the placement of the electrodes (or the measurement of such placements) acts as noise in any analysis of topographies or intracerebral sources.

(v) The Way in Which the Electrodes Are Affixed to the Scalp Should Be Described

The hair presents the major problem in keeping electrodes in good contact with the scalp. Ordinary metal electrodes can be affixed with adhesive paste, which serves both to hold the electrode in place and to connect it electrically to the scalp, or with collodion (either directly or in gauze). Collodion can be removed with acetone or (preferably) ethyl alcohol. In nonhairy regions of the head, the electrodes can be affixed using sticky tape or two-sided adhesive collars. Ag/AgCl electrodes in plastic housings do not work well with either adhesive paste or collodion. They can be affixed to the scalp by using collodion (alone or in gauze) to mat the hair around the site and then using two-sided adhesive decals. When using large numbers of electrodes, an elastic cap (Blom & Anneveldt, 1982) or net (Tucker, 1993) is helpful to hold electrodes in position. Care must be taken to ensure that the cap or net fits well and that the electrodes are located properly. A range of cap sizes to cover the different head sizes is clearly necessary. In children, an electrode cap is definitely preferable to applying electrodes individually. Although electrodes can be placed individually, the intersubject variability would be greater in children due to placing the electrodes on moving targets. Infants and young children do not always like having a cap on, but they often do not care for electrodes either, and at least when the electrode cap is placed successfully there is greater chance that the electrodes will be in the correct locations.

(vi) The Way in Which Artifact-Contaminated Single Channels Are Treated Should Be Described

In high-density multichannel recordings, one or more channels frequently contain large artifacts due to a poor contact between the electrode and the scalp or some amplifier malfunction. The number of such channels should be reported. The number of bad channels in any one recording should not exceed 5% of the total. Even if the number of channels is small, however, it is difficult to decide what should be done to integrate these data with other data from the same or other subjects. When generating averages, it makes little sense to include the bad channel in any rejection protocol (because all epochs might be rejected), but inclusion of the bad data would add unnecessary noise to grand averages. On the other hand, if the channel is omitted, data averaged across conditions (or across subjects) would then be available only for those channels that were recorded in all conditions (or all subjects).

One useful solution to this problem is to estimate the missing data, either using linear or spherical spline (Perrin, Pernier, Bertrand, & Echallier, 1989) interpolation. Although linear interpolation is mathematically simpler, it has the disadvantages that (a) electrodes at the edge of the array cannot be estimated, and (b) only a few adjacent electrodes are used to estimate the interpolation. Using spherical splines, an estimate of the signal at one missing electrode location is made from the signals at all the other electrodes, leading to less sensitivity to noise at individual electrodes. Missing data at the edge of the electrode array may also be estimated, because the splines assume continuity over the whole (spherical) head.

The method of spherical splines has other useful applications, apart from mapping, for which it was originally intended. Using the same interpolation method, a set of data recorded at digitized locations can be "normalized" to generate data at a set of standard 10-10 or 10-20 locations. Grand averages can then be generated from the normalized data. Another possible application is the automatic detection of bad electrodes. Data from each electrode are compared with the estimate computed from the other electrodes. A bad electrode/signal is detected when the differences between the real and estimated data become larger than a given threshold.

(vii) Referential Recordings Should Be Used and the Reference Should Be Specified

Almost all ERP recordings are made using differential amplifiers so that electrical noise in phase at the two inputs can be canceled. These differential recordings can be made using either referential montages (wherein the second input to all channels is a common reference) or bipolar montages (that link electrodes in chains with the second input to one channel becoming the first input to the next channel). By providing the slope of the potential field, bipolar recordings help localize a maximum or minimum at the point at which the recording inverts in polarity. However, they are often very difficult to interpret in ERP studies. Because bipolar montages can always be recalculated from referential montages but not vice versa, referential recordings are recommended for ERP studies.

The experimenter must specify the reference. A variety of reference electrodes can be used depending on the type of ERP and the recording system. Offline calculations can allow the subsequent rereferencing to any site or set of sites desired (Dien, 1998a; Picton et al., 1995). The physical linking of electrodes together to form a reference is not recommended because the shunting of currents between electrode sites may distort the distribution of the scalp voltages (Miller, Lutzenberger, & Elbert, 1991). Most recording systems will allow such a linked-electrode reference to be recalculated later if each electrode in the reference is recorded separately. If the recordings are obtained using a single reference, an average reference calculated as the sum of the activity in all recorded channels divided by the number of channels plus one (i.e., the number of electrodes) is perhaps the least biased of the possible references (Dien, 1998a). This approach allows activity to be displayed at the original reference site (equivalent to zero minus the value of the average reference). If the activity at the original reference site is not to be evaluated, the calculation of the average reference is determined by dividing by the number of recording channels. This calculation might be done, for example, if data to be used in source analysis were recorded using a linked-ear reference

(because the location of such a reference cannot be specified accurately). Average-reference recordings are particularly appropriate for topographic comparisons because they are not biased by a single reference site, for source analyses that usually convert the data to average-reference format prior to modeling and for correlation-based analyses, because the correlations are not inflated by the activity at a single reference site. The interpretation of the average reference has been the subject of some controversy and many of the assumptions underlying the use of average reference are not satisfied in actual recordings. However, if recordings are obtained from a reasonable sample of head locations (i.e., including electrodes below the Fpz-T7-Oz-T8 equator), the signals relative to an average reference will approximate the true voltages over the head, which must average to zero (Bertrand, Perrin, & Pernier, 1985; Dien, 1998a). When comparing waveforms and maps to those in the literature, it is essential to consider differences in the reference. For example, the classic adult P300 or P3b wave is usually recorded at Fpz as a negative deflection when using an average reference but as a positive deflection when using an ear or mastoid reference. It is often helpful when comparing waveforms with those in the literature that use another reference to plot the waveforms using both references or, if one is using the average reference, to include the waveform for the other reference electrode in the figure.

E. Amplification and Analog-to-Digital (A/D) Conversion

(i) The Gain or Resolution of the Recording System Must Be Specified

The recording system consists of the amplifiers that bring the microvolt signals into some range where they can be digitized accurately and the converters that change these signals from analog to digital form. The amplifier gain is the ratio of the output signal to the input signal. The resolution of the A/D converter is the number of levels that are discriminated over a particular range, usually expressed as a power of 2 (bits). For most ERP purposes an A/D converter using 12 bits (4,096 values) is sufficient, provided that the incoming signal typically ranges over at least 8 bits of this converter range and does not lead to blocking. Converters with greater precision are necessary if large DC shifts are being monitored without baseline compensation so that the resolution is sufficient even when the signal covers only a portion of the range.

The gain of the recording system can be specified in terms of resolution, that is, as the number of microvolts per least significant bit (smallest level discriminated by the A/D converter) or, inversely, as the number of bits per microvolt. This calculation combines both the amplifier gain and the resolution of the A/D converter. For example, if the amplifier increases the recorded EEG by a factor of 20,000× and the 12 bit A/D converter blocks at ±5 V, the range of the A/D conversion in terms of the input signal is ±250 μV, and the system resolution is 0.122 μV/bit (calculated as 10/[20,000 × 4,096]).

Amplifiers should have a sufficient common-mode rejection ratio (at least 100 dB) so that noise signals occurring equally at each of the electrodes can be eliminated. Subjects should be grounded to prevent charge accumulation and the ground should be protected from leakage currents. Under certain clinical circumstances, full electrical isolation of the inputs (e.g., using optical transmission) may be needed. These and other considerations of electrical safety are reviewed more fully elsewhere (e.g., Cadwell & Villarreal, 1999; Tyner, Knott, & Mayer, 1983).

The most common technique for calibrating the amplifiers uses a square wave lasting between one fifth and one half the recording sweep and having an amplitude typical of the largest ERP measurements to be made. Optimally the amplifier is calibrated in series with the A/D converter and averaging computer so that the whole recording system is evaluated. Another technique uses sine-wave signals at an amplitude and frequency typical of the EEG (or ERP) to calibrate the amplifier and A/D converter. With multichannel recording systems it is essential to measure separate gains for each channel (and to use these channel-specific gains in the amplitude measurements). These gains should be within 10% of the mean gain.

(ii) The Filtering Characteristics of the Recording System Must Be Specified

Analog filtering is usually performed at the same time as amplification. The bandpass of the amplifier must be provided in terms of the low and high cut-off frequencies (−3 dB points). We recommend describing the cut-offs in terms of frequencies rather than time constants, although the measurements are theoretically equivalent. In cases for which the filter cut-offs are close to the frequencies in the ERPs being measured, the slope of the filters (in dB/octave) should also be described, because analog filters with steep slopes can distort the ERP waveform significantly.

Analog filtering should be limited at the high end to what is necessary to prevent aliasing in the A/D converter (i.e., less than one half the frequency of A/D conversion) and at the low end to what is necessary to prevent blocking the converter by slow changes in baseline. Aliasing occurs when signals at frequencies greater than twice A/D conversion rate are reflected back into the sampled data at frequencies equal to subharmonics of the original frequencies (and at other frequencies that depend on the relation between the original signal and the A/D rate). Rough rules of thumb are to set the high cut-off to approximately one quarter of the A/D rate and the low cut-off to approximately the reciprocal of four times the sweep duration (Picton & Hink, 1974). When recording 1-s sweeps using a 200 Hz A/D conversion rate, these rules of thumb would lead to bandpass of 0.25–50 Hz. Further filtering can be done offline using digital filtering techniques. Filters do not completely remove frequencies beyond the cut-off frequency. For example, if a simple (6dB/octave) high-pass filter with a cut-off (−3 dB point) at one quarter of the digitization frequency is used, the attenuation of a signal at half the digitization frequency is only 9 dB (i.e., the amplitude is 35.5% of what it was before filtering), and strong signals well above the filter frequency may still lead to aliasing. The high-frequency noise from a video display may be a particular problem because the noise is locked to the stimulus. For example, a 90-Hz video refresh rate may alias into the ERP at a frequency of 5.625 Hz. Notch-filters to exclude the line frequency range (50−60 Hz) may significantly distort the recording and are therefore not recommended.

(iii) The Rate of A/D Conversion Must Be Specified

A/D conversion should be carried out at a rate that is sufficiently rapid to allow the adequate registration of those frequencies in the signal that determines the measurements. The minimum rate is twice the highest frequency in the signal to be measured. Frequencies in the recording higher than one half the A/D rate must be attenuated by analog filtering to prevent aliasing.

The multiplexing of the different recording channels to the A/D converters should be set up so that the delay interval between the measurements of different channels does not significantly dis-

tort any between-channel latency measurements (Miller, 1990). The most usual form of multiplexing switches among the channels using a rapid rate that is independent of the interval to switch to the next sample time. Provided this multiplexing rate is much faster than the A/D rate used for ERP studies, there will not be significant latency distortion. Optimal sampling would use a separate A/D converter for every channel, so that all channels could be sampled simultaneously. Alternatively, a single, multiplexed A/D converter could be preceded by separate sample-and-hold circuits for each channel. The simplest way to check that the multiplexing is not causing signal distortion is to record calibration sine-wave signals simultaneously in all channels and to ensure that the phase of the digital signal is equivalent in each channel. This method will also check for between-channel differences in the analog filters.

F. Signal Analysis

(i) Averaging Must Be Sufficient to Make the Measurements Distinguishable From Noise

The number of responses that need to be averaged will depend on the measurements being taken and the level of background noise present in single-trial recordings. The noise should be assessed in the frequency band in which the component is measured. Thus, it often takes fewer trials to record a recognizable contingent negative variation than a recognizable N100 of similar amplitude in an eyes-closed condition in which the EEG noise near 10 Hz is high. Many different techniques can assess the noise levels of averaged recordings (reviewed in Picton et al., 1995). Most of these measure the variance of individual trials or subaverages of the response. A simple way to demonstrate the noise level in a recording is to superimpose replicate tracings of subaverages of the response. Unfortunately, in recent years the incidence of such replicate ERP figures has declined.

The first question that might be asked is whether or not an ERP is present. This question is important when using the ERPs to estimate the threshold for detecting a stimulus or discriminating a difference between stimuli. The answer to this question will need some demonstration that the averaged ERP is or is not significantly different from the level of activity that would be present if the averaging had been performed on the recorded EEG without any ERP being present. This assessment must, of course, take into account the number of tests being performed. If every one of 200 points in an ERP waveform is tested automatically to determine whether it is significantly different from noise, approximately 10 of these tests will be significant at $p < .05$ by chance alone. Techniques are available to determine how many such "significant" results are necessary to indicate a truly significant difference (Blair & Karniski, 1993; Guthrie & Buchwald, 1991). Several other techniques are available to demonstrate whether a recorded waveform is significantly different from what might be expected by chance (e.g., Achim, 1995; Ponton, Don, Eggermont, & Kwong, 1997).

A second question is whether ERPs recorded under different conditions are significantly different. In general, if one wishes to demonstrate significant differences between ERPs, the noise level for each averaged ERP waveform should be reduced below the level of the expected difference. Differences between pairs of ERPs recorded under different conditions can be evaluated and depicted by computing the difference between the two ERPs. The variance of this difference waveform will equal the sum of the variances of the individual ERPs (provided that the noises of the two ERPs are not correlated). For example, if the variances of the two ERPs are roughly the same, then the standard deviation of the difference ERP will be larger than the standard deviations of the original ERPs by a factor of 1.41.

Source analysis is particularly susceptible to residual background noise because the analysis procedures will attempt to model both the noise and the signal. For source analysis, the noise variance (assessed independently of the source analysis) should be less than 5% of the signal variance. If the analysis is highly constrained, the signal-to-noise requirements for source analysis can be less stringent. This might occur, for example, if one bases the analyses of individual ERP waveforms on the analysis of the grand mean data by maintaining the source locations and just allowing the sources to change their orientations.

(ii) The Way in Which ERPs are Time Locked to the Stimuli or the Responses Should Be Described

The averaging process is locked to some triggering mechanism that ensures that the ERPs are reliably time locked to the events to which they are related. For ERPs evoked by external stimuli, this is usually done by recording a trigger at the same time as the stimulus. There are two sources of variability in this timing. The first concerns the relationship between the trigger and the stimulus. If the stimulus is presented on a video display, there may be some lag between the trigger and the occurrence of the stimulus when the raster scanning reaches the location of the screen where the stimulus is located. If the trigger is locked to the screen refresh rate, this lag will be a constant fraction of the refresh rate. The second source of variability derives from the way in which the triggers are registered in the recording device. The accuracy of this registration often depends on the speed of A/D conversion.

When the ERPs are locked to responses, it is essential to describe what response measurement is used (Deecke, Grözinger, & Kornhuber, 1976; Shibasaki, Barrett, Halliday, & Halliday, 1980). Two main trigger signals are possible: a mechanical signal such as button press or some measurement of the electromyogram (EMG). EMG measurements require recordings from electrodes placed over the main muscle used to make the response. The recorded signal is rectified and a threshold level is selected for initiating the trigger. The locations of the electrodes and the triggering level should be described clearly. Even when triggering on a mechanical response, it is helpful to record the rectified EMG. This recording will allow some estimate of the time between the EMG and the mechanical signal and the variability of this time.

(iii) When Latency-Compensation Procedures Are Used, They Should Be Defined Clearly and the Amount of Compensation Should Be Specified

One of the assumptions of averaging is that the ERP is time locked to the eliciting event. This statement means that the latency of each ERP component should remain constant across the trials that are used to compute the signal average. Any "latency jitter" that occurs when the timing of a component varies across trials can substantially reduce the peak amplitude of the average ERP. Latency jitter is particularly common when the ERP component of interest is a manifestation of a processing activity that is invoked at variable times following the external stimulus. In such cases, using the external stimulus to define the zero time for averaging can create substantial latency jitter in the data and the results can be misleading. The investigator must be particularly careful when comparing ERP amplitudes across conditions that vary in latency jitter. A reduction in the amplitude of an averaged ERP may be caused by greater latency jitter rather than a change in amplitude of the individual ERPs.

Various techniques can be used to adjust for latency jitter (Möcks, Köhler, Gasser, & Pham, 1988; Picton et al., 1995; Ruchkin, 1988; Woody, 1967). Most of these require that the ERP be relatively simple in shape and recognizable in single trials. The basic Woody technique cross-correlates the single trial waveform with the average waveform (template), shifts the latency of each single trial waveform to the latency of its maximum correlation with the template, recomputes the average using the shifted single-trial waveforms, and then iterates until some criterion is reached. The procedure can be facilitated by filtering the single-trial data (Ruchkin, 1988). In conditions that encourage latency jitter, some attempt to compensate for latency jitter is mandatory. Without such adjustments, amplitude comparisons should be avoided. Area measures (or mean amplitudes over a specified time window) may be helpful if the jittered waveform is mainly monophasic. When using latency compensation, the amount of compensation must be specified in terms of the average amount of latency shifting over trials, as well as the maximum and minimum of these shifts. The filter settings used to preprocess the single-trial waveforms should also be specified.

It is important to demonstrate that the outcome of the latency jitter adjustment is not merely the result of the technique lining up background noise (Donchin & Heffley, 1978). One way to check the solutions provided by the iterative Woody filter procedure is to compare the shape of the temporal distribution of the identified peaks with the shape of the raw average. If there is not an inordinate degree of amplitude variability across the trials in the average and if the component that is jittered is mainly monophasic, then the shape of the raw average should approximate the distribution of the single-trial peaks. Thus, one can have confidence in the Woody solution to the extent that the histogram of when the single-trial peaks were identified is a rough approximation of the raw average. Additional comparisons between the shapes of this plot and the RT distribution provides another converging measure in conditions wherein ERP latencies and RT are correlated. Another approach is only to use ERP trials in which the correlation between the ERP and the template exceeds the correlation between recordings where there is no ERP (e.g., over the prestimulus baseline) and the template.

(iv) Any Digital Filtering Algorithms Used in the Analysis Must Be Specified

Digital filtering of the ERP waveform can help to increase the signal-to-noise ratio by eliminating those frequencies in the recording that are irrelevant to the measurements (Cook & Miller, 1992; Glaser & Ruchkin, 1976; Nitschke, Miller, & Cook, 1998). Digital filtering has clear advantages over analog filtering. First, the original data can be maintained for evaluation using other filter settings. Second, digital filters can be set up so as not to alter the phase of frequencies in the waveform. Such zero-phase digital filtering does not distort the morphology of the ERP waveform as much as analog filters with similar bandpass. Third, digital filtering can more easily adapt its settings than filtering that depends on hardware components. It is therefore probably best to restrict analog filtering to what is required to prevent aliasing or blocking of the A/D converter, and to use digital filtering for signal analysis.

G. Noncerebral Artifacts

(i) Possible Noncerebral Artifacts Should Be Monitored

Unfortunately (from the point of view of recording ERPs), the brain is not the only source of electrical activity recorded from the human scalp. The scalp muscles, tongue, eyes, and skin can all contribute to these electrical recordings. The activity from scalp muscles and scalp skin potentials can usually be monitored adequately on the same channels as are used to record the ERPs. This is not true, however, for ocular and tongue potentials.

It is essential to monitor ocular artifacts using electrodes near the eyes when recording most ERPs.[4] If all recording electrodes (including the reference) are located on a single plane, a single electrooculogram (EOG) monitor channel can be used (with electrodes located on the same plane or one parallel). For example, a string of midsagittal electrodes (Fz, Cz, Pz, Oz) can be combined with a single supraorbital or infraorbital electrode. However, if electrodes are located over the entire scalp, at least two separate (and roughly orthogonal) channels should be used for monitoring the EOG artifacts. A single diagonal channel can be used to reject trials that are contaminated by blinks or eye movements (provided the movement is not orthogonal to the electrode derivation). However, this approach is not adequate if the purpose of the monitoring is to subtract the electroocular artifacts from the recordings.

In tasks involving overt speech or tasks wherein subvocalization might occur, electrodes should monitor the effects of tongue and jaw movement. Investigators can monitor these artifacts with electrodes over the cheeks and below the jaw (realizing that these electrodes will pick up ocular as well as glossokinetic artifacts). Several authors have suggested that these potentials are so large and variable that it is impossible to record cerebral ERPs associated with speech production (Brooker & Donald, 1980; Szirtes & Vaughan, 1977).

The average ERPs should then include simultaneously averaged waveforms from the monitoring channels. Like the cerebral ERPs, potentials deriving from noncerebral generators may appear on the averaged waveform without being recognizable in single-trial waveforms. For example, looking toward the responding hand can create an artifact that mimics the lateralized readiness potential recorded over the contralateral scalp preceding the response. If compensation procedures for EOG artifact are not used, any demonstration of the readiness potential should therefore include a simultaneously recorded horizontal EOG.

(ii) Subjects Should Be Informed About the Problem of Artifacts and Encouraged to Reduce Them

It is far more efficient to reduce artifacts before the recording than to remove them later by increased averaging or by compensation. Instructions to blink only during the intervals between trials can help, provided this request does not impose too heavy an attentional burden. It is not good for subjects to expend all their cognitive resources on timing their blinks and have none left for the experimental task. Young children pose particular problems for controlling artifacts. When testing children, instructions similar to those given to adults usually suffice for children of 5 years and older. When studying cognitive processes in younger subjects, the experimenter needs access to a pause button, so that ERPs are only recorded when the child is alert, quiet, and looking at the screen. For visual ERPs this is obviously necessary, but for auditory ERP studies such fixation is also important simply to reduce eye-

[4] If the ERPs being measured have latencies of less than 50 ms (e.g., auditory brainstem responses), it is unnecessary to monitor ocular artifacts, because the latency of time-locked artifacts is longer than 50 ms and because the frequency spectrum of the potentials associated with blinks and saccades is lower than the frequency spectrum of the potentials being recorded.

movement artifact. An interesting screensaver on a computer screen is extremely useful. If a young infant has a pacifier and can suck on it gently, the child may be calmer and more attentive. Children from 2 years through to at least 12 years (and older if clinical populations are included) will usually perform a task more attentively and produce fewer artifacts if an experimenter sits beside them and offers (at random intervals) words of encouragement (e.g., "That's great!" or "You're doing well!").

(iii) Criteria for Rejecting Artifact-Contaminated Trials Must Be Specified

Potentials generated by noncerebral sources often occur randomly with respect to the events eliciting ERPs. If so, they merely serve to increase the background noise and can be removed by averaging. However, because the potentials may be much larger than the ongoing EEG background, the extra averaging required to remove such potentials can be exorbitant. When the artifacts are intermittent and infrequent, the investigator should remove contaminated trials from the averaging process. Any trials showing electrical activity greater than a criterion level (e.g., ± 200 μV) in any recording channel should be rejected from averaging. The criterion would obviously vary with different recording situations. A ± 200-μV criterion would not be appropriate to recordings taken during sleep when the background EEG could be much larger, or to recordings taken with direct-coupled electrodes where there could be large baseline fluctuations. Rejection protocols do not obviate the need to average the recordings that monitor artifacts. It is always possible that small artifacts can escape rejection and still contribute significantly to the ERP.

Eye movements and blinks are particularly difficult to remove by simple averaging because they are frequently time locked to the stimuli. Rejection protocols may use criteria similar to those described above to eliminate from the averaging any trials contaminated by eye blinks or large eye movements. If rejection occurs when the activity recorded from supraorbital electrodes (referred to a distant reference or to an electrode below the eye) exceeds ± 100 μV, trials containing blinks will be eliminated. Other rejection procedures may use a more relative measure such as eliminating any trials in which the RMS value on eye monitoring channels exceeded a value that is, for example, two standard deviations larger than the mean RMS value for that channel.

The investigator should describe the percentage of trials rejected from analysis, and the range of this percentage across the different subjects and experimental conditions. Rejection protocols decrease the number of trials available for averaging. Young children require at least double, preferably triple, the number of trials used in adults due to the higher rejection rates due to ocular and muscle artifact, and behavioral errors (misses, false alarms). The rejection rate increases with decreasing age, and in infants rejection rates of 40% or more are routine. This problem is balanced somewhat by the larger ERPs that can often be recorded in younger children.

If the number of rejected trials is very high (more than a third in adults), the data may become difficult to interpret. Given a set amount of time or number of stimuli presented, the ERPs will show increased background noise because fewer trials will be accepted for averaging; given a set number of accepted trials, cognitive processes may habituate because of the longer time required to reach this number. As well, the trials may not be representative of the cognitive processes occurring: trials with EOG artifact may differ systematically from those without (Simons, Russo, & Hoffman, 1988). In these conditions, compensation protocols are preferable to rejection procedures. One way to assess whether trials with artifact are similar to those without is to compare the means and standard deviations of some behavioral measurement, such as the reaction time, before and after artifact rejection.

(iv) Artifact Compensation Procedures Must Be Documented Clearly

Although rejection procedures can be used to eliminate artifacts in many normal subjects, these protocols will not be satisfactory if the artifacts are very frequent. Rejecting artifact-contaminated trials from the averaging process may then leave too few trials to obtain an interpretable recording. In such conditions, compensation procedures can be used to remove the effect of the artifacts on the ERP recordings. Compensation procedures for ocular artifacts are well developed, and it is generally more efficient to compensate for these artifacts than to reject artifact-contaminated trials from analysis. Compensation will only attenuate the electrical effects of the artifacts, and other reasons may still exist for rejecting trials contaminated by ocular artifact. For example, the experimenter may not wish to average the responses to visual stimuli if these were presented when the subject blinked (and did not perceive the stimulus).

The most widely used methods to remove ocular artifacts from the EEG recordings subtract part of the monitored EOG signal from each EEG signal (for a comparison among several such algorithms see Brunia et al., 1989). This approach assumes that the EEG recorded at the scalp consists of the true EEG signal plus some fraction of the EOG. This fraction (or propagation factor) represents how much of the EOG signal spreads to the recording electrode. When using both vertical and horizontal EOG monitors to calculate the factors, it is essential to consider both channels of information in a simultaneous multiple regression (Croft & Barry, in press). The assumption that the contamination by ocular potentials is a linear function of the EOG amplitudes is reasonable for eye blinks, and for saccadic eye movements when the movements are within ± 15 degrees of visual angle. This general approach also assumes that the monitored EOG signal contains only EOG, with no contribution from the EEG, an assumption that is clearly not correct and that can lead to problems in estimating the true EEG signal, particularly in scalp regions near the eyes.

For effective artifact correction, two problems must be solved. The first is to compute the propagation factors for each electrode site. The second is to perform the correction. To compute the propagation factors accurately it is important to have enough variance in the eye activity. Blinks produce consistently large potentials and are usually frequent enough to compute propagation factors using the recorded data. Because the scalp distribution of an eye blink artifact is distinctly different from the scalp distribution of the artifact related to a vertical saccade, separate propagation factors should be calculated for eye movements and for blinks.[5] Although eye movements in the recorded data may be small but

[5]The potentials associated with blinks and saccades are generated by distinctly different processes. The eyeball is polarized with the cornea being positive with respect to the retina. Saccade potentials are caused by rotation of this corneoretinal dipole. Blink potentials are caused by the eyelid sliding down over the positively charged cornea, permitting current to flow up toward the forehead region (Lins, Picton, Berg, & Scherg, 1993a; Matsuo, Peters, & Reilly, 1975). Contrary to widespread beliefs, the eyeball does not roll upward during normal blinks (Collewijn, Van Der Steen, & Steinman, 1985). The different mechanisms for a vertical saccade and a blink account for the distinct scalp topographies of the potentials associated with them.

consistent enough to affect the EEG averages, they may nevertheless be too small to allow an accurate estimation of propagation factors. We therefore recommend that these propagation factors be measured using separate calibration recordings in which consistent saccades of the order of ±15 degrees are generated in left, right, up, and down directions. Blink factors can be calculated either from blinks recorded during the ERP trials or from blinks recorded during this calibration recording.

A proper correction procedure must somehow distinguish the different types of electroocular activity. Horizontal eye movements are well identified by the horizontal EOG, consisting of a bipolar recording of electrodes placed adjacent to the outer canthi of the left and right eyes (or with separate referential recordings from each electrode). Vertical eye movements and blinks are both recorded by the vertical EOG recorded from or between supra- and infraorbital electrodes. Blinks can be distinguished from vertical eye movements on the basis of their time course (Gratton, 1998; Gratton, Coles, & Donchin, 1983), although this method cannot cope with overlap, as in blink-like rider artifacts at the beginning of saccades (Lins et al., 1993a). Vertical eye movements and blinks can be distinguished on the basis of their relative magnitudes above and below the eye when a remote reference is utilized. For blinks, above the eye there is a large positive deflection, whereas below the eye there is a much smaller negative deflection, of the order of 1/10th the magnitude of the deflection above the eye. For vertical movements, the above/below eye deflections are also of opposite polarity, but the magnitudes of the above/below deflections are of the same order of magnitude. An alternative approach is to record an additional EOG channel that contains a different combination of vertical eye movements and blinks. By subtracting the appropriate combination of the two EOG channels, the two types of eye activity can be eliminated, even when they overlap. A useful additional EOG channel is the "radial EOG" (Elbert, Lutzenberger, Rockstroh, & Birbaumer, 1985), which can be computed by taking the average of the channels around the eyes, referred to a combination of channels further back on the head (e.g., linked ears). Using multiple regression to compute propagation factors between horizontal, vertical, and radial EOG and each EEG channel, any overlap of different types of eye activity can be corrected in the EEG data (Berg & Scherg, 1994; Elbert et al., 1985). When saccades are infrequent, it is possible to compensate for blink artifacts and to eliminate epochs containing other types of eye movement on the basis of visual inspection of the recorded data.

The use of propagation factors to compensate for the EOG artifacts in EEG recordings is not perfect. There may be changes in propagation factors over time due, for instance, to changes in the subject's posture and therefore direction of gaze, or to changes in the electrode–skin interface especially around the eyes. The use of one EOG channel for each type of eye movement is an approximation. EOG electrodes record EEG from the frontal regions of the brain as well as eye activity. This recording causes two problems. First, it can distort the regression equation used to calculate the EOG propagation factors. This distortion can be decreased by subtracting any stimulus-synchronized contribution (e.g., Gratton et al., 1983), by low-pass filtering the recording or by averaging the recordings using the onset of the eye-movement for synchronization (Lins, Picton, Berg, & Scherg, 1993b). Second, multiplying the EOG recording by the propagation factors and then subtracting this scaled waveform from the scalp EEG recording will remove a portion of the frontal EEG signal as well as the EOG.

A new approach to eliminating eye artifacts in multiple electrode data uses a source component analysis (Berg & Scherg, 1991, 1994; Ille, Berg, & Scherg, 1997; Lins et al., 1993b) to estimate the eye activity independent of the frontal EEG. Instead of considering propagation factors between EOG and EEG, source components or "characteristic topographies" are computed for each type of eye activity. These source components are combined with a dipole model (Berg & Scherg, 1994; Lins et al., 1993b) or principal components analysis (PCA)-based topographic description (Ille et al., 1997) of the brain activity to produce an operator that is applied to the data matrix to generate waveforms that are estimates of the overlapping eye and brain activity. The estimated eye activity is then subtracted from all EEG (and EOG) channels using the propagation factors defined by the source components. This technique has several advantages. First, it generates a better estimate of eye activity than is provided by EOG channels. Second, it allows the EOG channels to be used for their EEG information. Third, if separate source components are generated for each type of eye activity, their associated waveforms provide an estimate and a display of the overlapping eye movements: for example, the blink rider artifact overlapping a saccade is separated into a blink waveform and a saccade waveform. The quality of separation of eye and brain activity depends on the quality of the model of brain activity, but even a relatively simple dipole model provides a better estimate of eye activity than the EOG. Using this technique, the exact placement of the EOG electrodes is not important, although multiple electrodes near the eyes are required to estimate the eye activity. Six or more periocular electrodes are recommended for monitoring the EOG to obtain adequate source components for compensation. Because of this requirement, the technique is mainly appropriate to recordings with large numbers (32 or more) of electrodes.

H. Presentation of Data

(i) ERP Waveforms Must Be Shown

The presentation of averaged ERP waveforms that illustrate the principal phenomena being reported is mandatory. It is not sufficient to present only schematic versions of the waveforms or line or bar graphs representing selected waveform measures. There are several reasons why ERP waveforms are required. First, given the ambiguities inherent in current methods for ERP quantification, the nature of an experimental effect can often be understood most effectively by visual inspection of the appropriate waveforms. Second, visual criteria of waveform similarity are useful for comparing results across different laboratories. Third, inspection of the actual waveforms can reveal the size of the experimental effect in relation to the background noise remaining in the waveforms. Fourth, without a display of the waveforms the reader has no way of evaluating the validity of the measurement procedures used in data analysis.

Grand-mean ERPs (across all the subjects) are appropriate in cases in which individual responses display approximately the same waveshape. If there is substantial interindividual variability, however, representative waveforms from individual subjects should be presented. In all cases, some clear indication of intersubject variability should be given—this may take the form of graphical or tabular presentation of the latency and amplitude variability of the principal measurements. When the main findings concern a correlation between ERP measurements and a continuous variable, grand-mean waveforms can be presented for different ranges of the variable. For example, one could provide the waveforms representing each decade of age, or each quartile of a measurement of disease severity.

It is often helpful to overlay waveforms from different conditions to allow the reader to see the pattern of the ERP differences. Due regard must be paid to how easily these waveforms can be discriminated when they are reduced for publication. Clearly different lines must be used, and, in general, no more than three waveforms should be superimposed in a single graph.

(ii) Both Temporal and Spatial Aspects of the ERP Data Should Be Shown

ERPs are voltages that are recorded over both time and space. There are two main ways to display these data. The first is as a change in voltage over time—the ERP waveform. The second is as a change in voltage over space—the ERP topography (scalp-distribution map). Both time and space can be represented by using either multiple maps or multiple waveforms. For example, the scalp distribution of an ERP waveform can be shown by plotting all of the time waveforms on a diagrammatic scalp. Multiple maps from different points in time can show the time course of the scalp distribution (providing a movie of the brain's activity). When scalp distributions are compared statistically, it is more helpful to graph the results or some subset thereof than to provide the data in tabular form.

In most cases, it is useful to present ERPs from multiple electrode sites that span the scalp areas where the effects of interest are occurring. Changes in ERP waveforms across the scalp provide important evidence about the number and the topographies of the underlying components and are crucial for comparing experimental effects across subjects and laboratories. Topographic information is also invaluable for distinguishing ERPs from extracranial artifacts arising from eye movement or time-locked muscle activity. Finally, examining the fine structure of the waveform at different sites is very useful for interpreting topographic voltage contour maps, for example, for determining whether more than one component is contributing to a voltage measure at a particular time point. A map at a given point in time often cannot adequately substitute for waveforms at multiple electrodes when determining the component structure.

(iii) ERP Waveforms Must Include Both Voltage and Time Calibrations

Ideally the figure layout should be such that readers can easily measure amplitudes and latencies for themselves. The voltage calibration line must show the size of a simple number of microvolts (i.e., $+5\ \mu$V rather than $+4.8\ \mu$V). We recommend that the time calibration span the whole duration of the sweep. We further recommend hash marks on the time calibration to indicate subdivisions consisting of a simple number of milliseconds (e.g., 100 ms rather than 75 ms). This temporal calibration line must also clearly show the timing of the sensory stimuli and motor responses.

(iv) The Polarity Convention of the ERP Waveform Must Be Indicated Clearly

ERP waveforms can be plotted with upward deflections indicating positive or negative potentials at the active electrode relative to the reference. Both conventions are used in the literature and no general consensus exists as to which is preferable. Whatever polarity convention is used must be represented in the figure and not just in the figure legend. The preferred way is to indicate the calibration voltage with a sign ("+" or "-") at the upper end of the voltage calibration. This can often be done together with the voltage measurement (e.g., $+10\ \mu$V). Another approach is to place the "+" and "-" signs at the ends of the voltage calibration and the magnitude in the middle.

(v) The Locations of the Electrodes Should Be Given With the ERP Waveforms

These locations can be given either by giving the name of the electrode adjacent to the waveform or by suggesting their location by the position of the waveform in the figure. The reference must be clearly specified in the figure or figure legend.

(vi) If Subtractions Are Used, the Original ERPs From Which the Difference Waveforms Were Derived Should Be Presented Together With the Difference Waveforms

One design for psychophysiological experiments is to compare physiological measurements recorded under two conditions that were presumably chosen so that one or more psychological processes differ between the conditions without any differences in other variables that might affect the physiological recording. Given this design, a simple way to examine ERP differences between the conditions is to subtract the recorded waveform in one condition from that recorded in the other. The resultant "difference waveform" is assumed to represent physiological processes that are different between the two conditions.

The weakness of this approach, however, is that physiological processes are usually not additive, that is, do not occur such that the physiological processes in one condition equal those processes in the other conditions plus or minus one (or more) other processes. Consequently, the interpretation of the difference waveform is not straightforward. The difference waveform represents activity caused by the physiological processes that are present in one condition but not in the other. The difference waveform by itself does not show which of the original waveforms contained the additional components. Indeed, the difference waveform may represent the superimposed effects of processes that were specific to both the minuend and subtrahend ERP waveforms. These issues concerning "cognitive subtractions" are not unique to ERP research and arise with other techniques, particularly studies of cerebral blood flow (Friston et al., 1996).

When using difference waveforms, authors should bear in mind various factors that might affect the subtraction by differentially affecting the two recordings from which the difference is calculated. Cognitive factors include changes in the state of the subject and changes in the manner of processing information between the two recordings. More physiological factors include changes in latency of one or more components in the unsubtracted ERPs. When a particular subtraction has been commonly used in a well-known paradigm, these considerations need not be discussed in the paper. However, any new or uncommon subtractions warrant some discussion of these issues.

Whenever difference waveforms are used it is essential to describe exactly how the subtraction was carried out and to delineate the polarity of the resulting difference waveform. A lateralized readiness potential can be demonstrated by subtracting the ERP recorded over the frontocentral scalp region ipsilateral to the responding hand from the ERP recorded contralaterally. This difference waveform can then be averaged across left- and right-hand responses to obtain a waveform indicating the time course of response activation independent of hand activated (Coles, 1989). However, because the subtractions may be performed and combined in other ways (e.g., De Jong, Wierda, Mulder, & Mulder, 1988), the investigator should be very clear about what was done to calculate the resultant difference waveform (Eimer, 1998).

(vii) Maps Should Identify Clearly What Is Represented and Should Be Plotted Using Smooth Interpolations and a Resolution Appropriate to the Number of Electrodes

It is essential to tell the reader what the map represents. Generally this explanation requires that the map be characterized by the type of measurement (e.g., voltage, current source density), latency, reference (for voltage maps, current source density maps are reference free), and mode of interpolation. It is important to realize that most data points in a scalp-distribution map are interpolated from recorded data rather than recorded directly. Smooth interpolation routines such as those using spherical splines (Perrinet al., 1989) are preferable to nearest-neighbor routines that often show spurious edge effects. Contours in the map (or different colors) should follow a resolution that is appropriate to the values recorded. For most ERP maps, a resolution of 10 levels is sufficient to show the topographical features. Multiple maps can be scaled in two ways: a magnitude scale plots the actual voltage or voltage slope (for current source density) and a relative scale plots values from the minimum to the maximum for each map. A magnitude scale highlights the differences in size of the recorded activity across maps, whereas a relative scale highlights differences in topography across maps. The figure legend should indicate the type of scale and the same scale should be used for all maps within one figure.

(viii) The Viewpoint for Scalp-Distribution Maps Must Be Indicated Clearly

The scalp distribution of the recorded voltages or current source densities can be viewed from above, from the side, from the front, or from the back. Other viewpoints are not recommended since it is difficult to document the view and without such documentation the map loses meaning. The viewpoint can be indicated diagrammatically by using landmarks such as the ears, eyes, and nose, provided these landmarks are easily visible and not ambiguous. Unless there are compelling reasons otherwise, maps viewed from above should be plotted with the front of the head at the top and the left of the head at the left. Because radiological imaging often uses an opposite convention, left and right should be clearly indicated on the figure. Similarly, lateral and anteroposterior views should indicate front-back and left-right.

(ix) Color Should Not Distort the Information in a Map

Color scales can sometimes help clarify the contours of a map, but these scales are not linear. In a scale based on the visual spectrum, the changes from orange to yellow and from yellow to green are much more distinct than the changes from red to orange or from green to blue. Some of this nonlinearity derives from the confounding of color and luminance: the yellow color in the middle of the scale is generally brighter than the colors at either end. Wherever possible, color scales should be chosen so that there is reasonable correspondence between changes in color and changes in luminance (the Xerox criterion). Because red-green color blindness is not uncommon, we recommend that scales using both these colors not be used. This allows two main color scales: the heat scale (purple-red-orange-yellow-white) and the sea scale (purple-blue-green-yellow-white). In general, gradations of a parameter are better shown by changes in color saturation of a single hue, whereas changes from one parameter to another can be displayed by a change in hue. These suggestions imply that whereas negative and positive polarities in a voltage map can be represented by two different colors (e.g., blue and red), gradations of positivity and negativity may be shown by modulating the saturation of these colors.

I. Measurement of ERP Waveforms

(i) Measured Waves Must Be Defined Clearly

Once the ERPs have been recorded they must be measured. Measurement requires that the components of a waveform be defined in some way.[6] The simplest approach is to consider the ERP waveform as a set of waves, to pick the peaks (and troughs) of these waves, and to measure the amplitude and latency at these deflections. This traditional approach has worked surprisingly well for many purposes, despite the fact that there is no a priori reason to believe that interesting aspects of cerebral processing would be reflected in these positive and negative maxima. More complex analyses (e.g., principal component analyses) are often performed in an attempt obtain some better index of the psychophysiological processes. Nevertheless, the results of these analyses are often presented as waveforms over time and measured in terms of peaks and troughs.

Several ERP labeling systems are currently in use, each with both advantages and drawbacks. The two most common approaches are to designate the observed peaks and troughs in the waveform in terms of polarity and order of occurrence in the waveform (N1, P2, etc.) or in terms of polarity and typical peak latency (N125, P200, etc.). A variant of the latter system can be used to describe a mean deflection over a specified time window (e.g., P20-50, N300-500). Negative latencies may used to label movement-related potentials that precede response onset (Shibasaki et al., 1980). For example, N-90 indicates a negative deflection that peaks 90 ms prior to the response as measured by initial peak of the rectified EMG. There are inherent problems with both the latency and the ordinal systems, because a waveform feature representing a particular psychophysiological process may vary in its timing or order of appearance depending upon experimental circumstances, age or clinical status. To minimize such ambiguities, authors must be absolutely clear about how their labels are applied. For both the ordinal and the latency convention, the observed latency range and mean value for each peak should be specified, and variations as a function of scalp site and experimental variables noted. To emphasize variations among components at different scalp areas, the recording site may at times be usefully incorporated in the label (e.g., N175/Oz).

An important distinction needs to be made between observational terminology, which refers to the waveform features measured in a given data set, and theoretical terminology, which designates ERP components that represent particular psychophysiological processes or constructs (Donchin, Ritter, & McCallum, 1978). For some ERPs, theoretical labels have been assigned that identify the hypothesized functional roles of the components, such as “mismatch negativity,” “processing negativity,” or “readiness potential.” In other cases, polarity-latency labels such as P300 or

[6]The word “component” is used in the ERP literature in several ways (Picton & Stuss, 1980). The word indicates the parts or constituent elements that make up a whole. In its general sense, the word therefore describes the parts of an ERP waveform analyzed according to some concept of its structure. This structure should be defined, either directly or by context. Three structures are often used. First, the ERP can be considered as a simple waveform composed of waves or deflections. Second, the ERP can be considered in terms of how it has been manipulated experimentally. Within this concept one can analyze the waveform into parts using subtractions or using a statistical analysis of principal components. Third, the ERP can be considered in terms of how it is generated by sources within the brain. Ultimately, the goal is to understand the ERP waveforms in terms of both intracerebral sources and experimental manipulations. A component would then be a temporal pattern of activity in a particular region of the brain that relates in a specific way to how the brain processes information.

N400 have been used in a theoretical sense, referring not to a waveform feature but to a psychophysiological entity with specific functional properties. One useful suggestion for keeping observational and theoretical nomenclature separate is to identify the latter with a line over the name (e.g., $\overline{\text{P300}}$). The proliferation of cognitive ERP studies in recent years has resulted in such a menagerie of components that it is often difficult to know whether the theoretical entities identified in one study are in fact equivalent to those of another study. Sorting out this situation will be made easier by keeping observational and theoretical terminology distinct.

Peak amplitude measurements are typically made relative to either a prestimulus baseline (baseline-to-peak) or with respect to an adjacent peak (or trough) in the waveform (peak-to-peak). The baseline period should be long enough to average out noise fluctuations in the average waveforms. Baseline periods shorter than 100 ms may increase the noise of the measurements by adding the residual noise in the baseline to the residual noise in the peak measurement. In general, baseline-to-peak measurements are preferable to peak-to-peak measurements, given that successive peaks may well reflect different physiological and/or functional processes that would be confounded in a peak-to-peak measure. However, in cases in which the peaks of interest are superimposed on a slower wave or a sloping baseline shift, the peak-to-peak measure may be a more veridical index of temporally localized activity. Peak-to-peak measures are also appropriate in cases in which an adjacent peak-trough ensemble is considered to reflect the same functional process or in which one member of such an ensemble remains constant under the experimental manipulations.

The choice of a baseline is particularly problematic when studying response-locked potentials. When measuring potentials that occur before a response, the baseline period should be chosen at a latency sufficiently early to demonstrate slow preparatory processes. Although the potentials specifically related to a motor act occur some tens of milliseconds prior to the act, readiness potentials begin several hundreds of milliseconds or even seconds earlier. It is often necessary to use more than one baseline period to measure different parts of response-locked potentials. Examples would be an early preresponse baseline for measuring the preparatory and motor potentials and an immediately preresponse period for measuring the postresponse potentials. When both stimulus- and response-locked potentials overlap (for example, in potentials related to making an incorrect response), the baseline should be chosen prior to the occurrence of any of the stimuli so as to be unaffected by latency-jittered remnants of the stimulus-evoked potential. Another approach to this problem (or its inverse) would be to estimate the latency jittered stimulus-evoked potential and to subtract this away from the response-locked average (Woldorff, 1993).

Although peaks are usually picked at the point of maximum (or minimum) voltage, this selection may be problematic if the data are noisy or if the waveforms are not symmetrical about the peak. An alternative method of determining peak latency and amplitude uses a midlatency procedure (Tukey, 1978). In this procedure, the maximum amplitude in a time window at a specified electrode is found and then the leading and lagging edges of the peak are searched to find the latencies where the amplitudes are some specified fraction (e.g., 70%) of the maximum value. These two latencies are then averaged to yield a measure of the peak latency. The procedure is most appropriate when there is a broad, flat peak.

An important pitfall must be kept in mind when comparing peak measurements if the peaks are being defined as the maximum deflections (either positive of negative) within a specified time window. In this case, it is only appropriate to compare the measured amplitudes of averaged waveforms that are based on a similar number of trials (stimulus presentations). The fewer trials included in the average, the more residual noise is superimposed on the peak, and the more the maximal peak (or trough) in the interval will be determined by the residual noise in the average rather than by the peak of interest. For this reason, averaged ERPs based on fewer trials will tend to have larger amplitudes (and more variable latencies) when measured by a peak-within-a-window algorithm.

This type of artifact may be mitigated by measuring peak amplitudes at a fixed latency, by low-pass filtering the data to remove some of the unaveraged noise, or by measuring mean amplitudes over a specified time window (essentially the same as low-pass filtering). The mean amplitude is more stable than the amplitude at a fixed latency. Furthermore, the time windows for a mean measurements may be adjusted to encompass those parts of the waveform where effects of interest are expected to occur, whether or not they contain any clear peaks. The choice of the time window, however, is not simple, and tends to be influenced by post hoc considerations. It is also difficult to apply when experimental groups have different peak latencies and/or more or less dispersed waveforms. It is desirable, therefore, either to determine time epochs of interest a priori, on the basis of previous studies, or to determine the window limits using an objective algorithm for finding the onsets and offsets of components.

Quantifying the onset and offset of an ERP wave might better capture the time course of cerebral processes than measuring its peak latency. A component's onset may be used to measure the beginning of a particular stage of processing, and a component's duration may index the duration of that processing stage. However, defining the onset and offset of a component is difficult, since these measurements are very susceptible to any residual noise in the ERP waveform. A possible approach is to use point-by-point statistics and define the onset as the first latency (within a predefined time range) at which the difference between the waveforms elicited in the two conditions of interest, or between the waveform and its baseline, starts being significant and does not return to insignificant values before the offset of the component. In a similar way one might define and measure offsets of components. Another approach (Scheffers, Johnson, & Ruchkin, 1991) is to measure onset and offset latencies by using suitably defined points on the leading and trailing slopes of a component. For example, even when onset/offset latency are not observable, latencies can be measured at amplitudes that are a specified fraction of peak amplitude (e.g., half-amplitude). Although such "fractional" latencies do not provide absolute measures of onset/offset latency, they do provide relative measures, in the sense that fractional latencies covary with onset/offset latencies. In addition, the resulting measurements are independent of any amplitude differences across experimental manipulations or subjects. The measurement of onset is particularly important when studying the lateralized readiness potential, because the onset is closely related to the decision processes that initiates selective response activation (Coles, 1989; Eimer, 1998). Two methods have been proposed specifically to measure this onset latency (Miller, Patterson, & Ulrich, 1998; Schwarzenau, Falkenstein, Hoormann, & Hohnsbein, 1998).

(ii) Measurements of a Peak at Different Electrodes in a Single Subject and Experimental Condition Should Be Taken at the Same Latency

If the scalp topography of a peak is to be considered, measurements should not be taken at different latencies for different electrodes. To do so would confound any rational definition of a peak and would be extremely susceptible to noise. Unfortunately, soft-

ware to measure the maximum peak within a latency range often does this calculation independently for each electrode location. If a peak inverts in polarity, these methods will attenuate (and sometimes eliminate) the inversion by measuring noise peaks of uninverted polarity. The topography should therefore be measured at one selected latency. The latency of a peak may be difficult to identify if it varies across different electrodes. If the peak is clearly maximal at one electrode location, its latency at this location should normally be used. For widely distributed peaks, the average latency at a set of electrodes may be used, or peaks may be identified in a measurement of global field power (Lehmann, 1987; Lehmann & Skrandies, 1980). Sometimes, it may be worthwhile to measure the waveforms at peak latencies determined at different electrodes, for example, auditory N1a, N1b, and N1c waves from frontal, vertex, and temporal electrodes, respectively (McCallum & Curry, 1980).

When comparing (or combining) topographies across subjects and/or conditions, the investigator should use the latency determined for each subject and/or condition. It is inappropriate to represent differences between ERPs recorded in different conditions as the difference between two maps recorded at the same latency. Because of latency shifts in the ERP across conditions, the two original maps may represent two different phases of the same ERP component. If so, the difference map does not reflect a change in the component across conditions but rather the difference between early and late phases of a latency-varying component.

(iii) Mean Amplitude Measurements Over a Period of Time Should Not Span Clearly Different ERP Components

One of the ways to handle problems of peak identification and the latency variance between subjects is to take a mean amplitude measurement of the waveform over a defined period of time. This period may derive from measurements of peak latency in grand-mean waveforms or may be arbitrarily defined. Although this mean measurement may be converted to an area measurement by multiplying by the time period, we recommend using the simple mean amplitude. When measuring slow or sustained potentials the latency range can span several hundred milliseconds. However, if the scalp distribution of the ERP changes significantly during the measurement period, the resultant measurements may become impossible to interpret.

(iv) Area Measurements Should Be Described Clearly and Used With Caution

An "area" measurement calculates the mean amplitude of a waveform between two defined time points and multiplies this mean by the difference in time. If the time points are defined arbitrarily, simply calculating the mean amplitude is preferable because amplitude units are easier to understand than amplitude-time units. If the experimenter wishes to measure the combined duration and amplitude of an ERP-wave, the time points for the area measurement would be defined on the basis of the waveform (e.g., the onset and offset of a wave of a particular polarity). In this case, the experimenter should be careful because slight changes in the level of residual noise or the estimation of baselines can cause large changes in these latencies.

J. Principal Component Analysis (PCA)

(i) The Type of Association Matrix on Which the PCA Is Based Must Be Described

Multiple different brain processes can generate measurable electrical fields at a distance from where they are generated. These fields linearly superimpose to produce the ERP waveforms observed on the scalp. A voltage measured at particular time point and a particular scalp location may therefore represent the activity of multiple ERP components. Each of these "components" of the ERP has a specific topography, occurs over a particular period of time, and is related in a characteristic way to the experimental manipulations. ERP components are defined in terms of how they are distributed across the scalp and how they are affected by experimental manipulations. Donchin et al. (1978) thus proposed that an ERP component was a "source of controlled, observable variability," and suggested that the ERP can be decomposed into a linear combination of components, each of which can be independently affected by the experimental manipulations. Such a model fits easily with the procedures of PCA (Donchin, 1966; Donchin & Heffley, 1978; Glaser & Ruchkin, 1976, pp. 233–290; Möcks & Verleger, 1991; Ruchkin, Villegas, & John, 1964; Van Boxtel, 1998), which is a method for linearly decomposing a multivariate data matrix. When applied to a set of ERPs, PCA produces a set of components. Associated with each component is an array of component "coefficients" or "scores" (one for each ERP in the original set). The product of a component and its coefficient for a given ERP specifies the contribution of the component to the ERP. In the most common way that the PCA has been used to study the ERPs, the variables examined in the analysis are the time points of the ERP waveform and the resultant components are therefore waveforms. The coefficients then represent the amplitudes of the different components in the recorded ERPs. In the terminology of factor analysis, the components are often referred to as "factor loadings" and the coefficients as "factor scores".

For PCA to be effective, there must be an ample and systematic variation within the set of ERPs being analyzed. Hence, ERPs are usually obtained from a variety of scalp sites, from more than one experimental condition, and from a set of subjects. Diversity in the ERPs, as a function of scalp location and/or experimental condition, is essential for decomposing the recordings into the underlying constituent processes. A study is only as good as the degree to which the investigator has induced systematic variance into the measurements and gained control over that variance by the experimental manipulations.

In ERP research, two different types of PCA formulations have been used. One type is temporal PCA, in which the data are conceptualized as waveforms and the data matrix is laid out with the time variable nested innermost. The second type is spatial PCA, in which the data are conceptualized as topographies and the data matrix is laid out with the electrode location variable nested innermost (for details see Dien, 1998b; Spencer, Dien, & Donchin, 1999).[7] The formulation and nesting arrangement of the data must

[7]The use of PCA in ERP studies can be compared with the (original) use of the PCA in psychometrics, in which the data consist of measurements on many "variables" obtained for a number of "cases." In psychometrics the variables are usually scores on some test and the cases are individual subjects. In a "temporal" analysis of an ERP data set, the cases can be the specific ERPs recorded from a particular electrode and associated with a specific event. The variables in this case are the voltages measured at each time point. The association matrix is then computed between the variables (time points) across all cases (electrode by condition). A time point can also be treated as a case, and the electrodes as variables. This view is used when performing a "spatial" PCA, in which the association is computed between the electrodes as variables across the time points that are then the cases. For temporal PCA, the manipulations could be electrode, experimental condition, and subject; for spatial PCA, the manipulations could be time, experimental condition, and subject. The structure of these manipulations is not addressed directly by the PCA. Techniques have been developed for multimodal decomposition of such data structures but have not yet been applied widely in ERP studies.

be specified explicitly. Although the temporal PCA has been the version used most frequently in the ERP literature, spatial PCA approaches have been applied in the correction of ocular artifacts (Berg & Scherg, 1994) and the derivation of sources (Mosher, Lewis, & Leahy, 1992).

The first step in a PCA is to compute an association matrix[8] from the data. It is crucial that the type and means of computation of the association matrix are specified. The matrix can consist of cross-products, covariances, or Pearson product-moment correlation coefficients. The resulting PCA will differ as a function of the type of association matrix and way in which the data are entered into this matrix. For temporal PCAs, the associations are most commonly calculated between different time points in the ERP waveforms. The association matrix is then dimensioned by the number of time points, and the resultant components are temporal waveforms. For spatial PCAs, the associations are usually calculated between different electrodes, and the matrix is dimensioned by the number of recording channels. The derived components for a spatial PCA are topographies, or variations in amplitude across electrodes.

(ii) The Criterion for Determining the Number of Components Must Be Given

A PCA examines the multivariate space defined by the original variables measured in the study. The PCA fits a new set of coordinates in which the data can be described, in which each of the new dimensions is a linear combination of the original variables. The new dimensions are defined so that the first component accounts for the largest percentage of the variance in the data, the next component accounts for the largest percentage of the residual data and is orthogonal to the first, and so on. The data are thus described in a space of new, orthogonal, "principal" components. The number of extracted components needed to account for the variance is usually smaller than the number of the original variables. Because the PCA defines the data in terms of components that explain successively smaller proportions of the variance, the first set of components usually accounts for the signal, or at least the most important parts of the signal. The remaining components account for the noise, and for constituents of the signal that cannot be distinguished from the noise. The second step in a PCA is therefore to determine how many components to retain. The number of meaningful components can be determined by using various criteria for deciding where to place the cut-off between signal and noise (Gorsuch, 1983, Chapter 8).

(iii) The Type of Rotation Used (if any) Must Be Described

The mathematics of PCA constrains both the set of components and the set of coefficients to be orthogonal. A second (optional) step in the analysis relaxes one (but not both) of these constraints via a varimax rotation of either the components or the coefficients. In ERP applications, the varimax rotation has usually been applied such that the resulting coefficients are orthonormal and the components are nonorthogonal and tend to be temporally compact for temporal PCAs and spatially compact for spatial PCAs. This compactness derives from an interaction between the rotational criterion and the structure of the data. It is also possible to apply the varimax rotation such that the components are orthonormal and the coefficients are nonorthogonal and concentrated over a limited set of electrodes and conditions.[9] Other rotations are possible (e.g., Dien, 1998b) but have not been used widely in ERP studies.

(iv) The Components Must Be Presented Graphically

Each component must be plotted. For a temporal PCA, the components must be plotted as waveforms, using a time scale similar to that used for the ERP waveforms. Each component may be scaled directly in voltages, or plotted so that its amplitude varies over time as a function of the amount of variance associated specifically with that component. For a spatial PCA, the components should be plotted topographically as maps.

(v) The Nature of the Components Should Be Described in Terms of the Experimental Manipulations

The nature of a component is best described in terms of what part of the experimental variance it represents. This can be demonstrated by presenting the coefficients or scores of the components in graphs, plotted as functions of electrode and experimental conditions, or in maps with one map for each component. The component scores measure the amount of a component within a given ERP and can be evaluated in statistical tests in the same way as amplitude measurements. For example, the scores can show the topography, that is, variation across electrodes, of the component waveforms obtained from a temporal PCA, or the waveform, that is, variation over time, of a component topography obtained from a spatial PCA. These analyses of variance (ANOVAs) should be used to demonstrate the nature of the components rather than to demonstrate significant experimental effects, since one can criticize the ANOVAs as being susceptible to Type 1 error. The logic is that a significant component of the variance the data exists and is related to particular experimental variations.

PCA is essentially a method to parse the experimentally induced variance into a small number of independent components.

[8] The matrix consists of a square matrix of association indices with a size equal to the number of points in the waveform (or the number of electrodes in the topography if a spatial PCA is being carried out). Calculating these indices by simple multiplication yields a cross-products matrix. Subtracting the mean waveform from each individual waveform before multiplication will give a covariance matrix. Standardizing each point so that all points have the same variance before calculating the indices gives a correlation matrix. PCA uses the variance between the points across experimental manipulations to extract the components. A cross-products matrix contains the total variance of the data. A covariance matrix contains the variance related to the experimental manipulations. A correlation matrix contains the experimental variance for standardized measurements. Because interest is generally in components that are affected by the experimental manipulations, a PCA of the covariance matrix is the most commonly used method for analyzing ERPs. A cross-products matrix represents all the energy in the measurements, but emphasizes large measurements independently of the experimental effects. A correlation matrix tends to accentuate small differences at some points. Because the ERP values all use the same units (voltage), there is no real need to scale the measurements by the standard deviations.

[9] A matrix of components (or arrays of coefficients) is orthogonal if each component (or array of coefficients) is uncorrelated with all other components (or coefficient arrays) in the matrix. A matrix is orthonormal if, in addition to being orthogonal, the mean square amplitudes of each component (or array of coefficients) are the same for all components (or coefficient arrays) in the matrix. A PCA of a set of ERPs consists of a set of coefficients and set of components. One of these sets will be orthonormal, with a dimensionless scale, and the other will be orthogonal, being scaled in microvolts. When an orthogonal rotation is applied to the results of the PCA, the orthonormal set will remain orthonormal, but the orthogonal set becomes nonorthogonal, while still being scaled in microvolts. Temporal PCAs are typically implemented such that the coefficients are orthonormal and the components are scaled in microvolts, and hence, after an orthogonal rotation, the components are nonorthogonal. However, it is always possible to rescale a PCA such that either the components or the coefficients are orthonormal, so that it is possible for either one or the other to remain orthonormal after rotation.

As is true of all analysis techniques, the use of PCA requires art and experience, and the interpretation of the components requires caution. Components of the ERP that contribute only small amounts of variance during the experimental manipulations may not show up clearly in the analysis. The orthogonality constraint is likely to result in an imperfect mapping between the "actual" physiological components and the components produced by PCA (with or without a subsequent rotation). Noise in the data may add to this problem of "misallocation of variance" (Wood & McCarthy, 1984; see also Achim & Marcantoni, 1997; Dien, 1998b).[10]

Like other ERP measures, the temporal PCA is susceptible to the effects of latency jitter. If the ERPs in a set of similar conditions contain an ERP component that has different latencies in different conditions or in different subjects, the PCA will correctly identify this latency variability as a source of variance and may identify multiple components where only one physiological component exists (Donchin & Heffley, 1978). Hence, PCA should be applied only after the investigators have examined the latency distributions of their data. A corresponding problem exists if there are variations in topography ("spatial jitter").

At its most basic and most powerful level, PCA is a method of simplifying complex, multichannel ERP data sets by reducing their temporal and/or spatial dimensionality. At a higher level, PCA can provide some insight into how ERPs are affected by the experimental manipulations.

K. Source Analysis

(i) The Type of Source Analysis and the Procedures Followed Must Be Specified

Source analysis is the name given to a variety of routines that attempt to model the scalp-recorded fields on the basis of generators within the brain. There are several approaches. One distinction is between moving and stationary sources. The moving source approach models each point in time with the best possible source or set of sources that can explain the potentials recorded at that point in time at the different scalp locations (Fender, 1987; Gulrajani, Roberge, & Savard, 1984). The stationary source approach (Miltner, Braun, Johnson, Simpson, & Ruchkin, 1994; de Munck 1990; Scherg, 1990; Scherg & Picton, 1991) postulates a set of sources that remain constant in location and orientation during the recording. This type of analysis then models how the contribution of these stationary sources to the ERP waveform varies over time. This analysis provides the time course of activity at each of the sources.

Another distinction is between discrete and distributed sources. Discrete source analyses consider the scalp-recorded activity to be generated by a small number of distinct dipolar sources that differ in location and/or orientation. Distributed source analyses interpret the scalp-recorded fields in terms of currents at a large number of locations within the brain. This distinction between discrete and distributed analyses can also be considered as referring to models that assume that the number of sources is less or more than the number of electrodes. For models with fewer sources than electrodes, the locations and orientations of sources are usually fit to the data using nonlinear search algorithms that try to minimize a cost function such as the residual variance between the modeled waveforms and the actual waveforms recorded at the scalp. For models with more sources than electrodes, a fixed set of sources distributed through the brain or over the cortical surface is assumed. To obtain a meaningful solution, constraints are applied on the currents generated at these sources. A minimum norm analysis gives intracerebral currents with the minimum total current (Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993; Hämäläinen & Ilmoniemi, 1984). Low-resolution electromagnetic tomography (LORETA) gives intracerebral currents that show the greatest smoothness, that is, that change least from one location to the next (Pascual-Marqui, Michel, & Lehmann, 1994).

Source models can be used in different ways. At one extreme, they may describe in a tractable way the topography of the data because a fitted dipole source indicates the center of gravity of a field distribution. At the other extreme, they can attempt to explain the underlying brain generators and their overlapping activity over time. Depending on the researcher's goals and the quality of the data, dipole models may be applied anywhere between these two extremes.

In view of the continuing developments in the field and the variations among the methods, it is difficult to make recommendations that apply to all methods. Although some of the following points apply to all methods, they concentrate on the methods that assume fewer sources than electrodes (both moving source and spatiotemporal methods), because these methods are still the most frequently used.

(ii) The Constraints and Assumptions Used in the Source Analysis Should Be Described

Because of the low spatial resolution of the EEG, and because of the infinite number of possible generator combinations that can give rise to the surface potentials, it is necessary to make a number of assumptions before using source analysis to identify generators. Such assumptions can include (1) a limited number of sources, (2) hemispheric symmetry of the sources (Scherg & Berg, 1991), (3) minimum energy of sources (Hämäläinen & Ilmoniemi, 1984), and (4) sources constrained to the cortical surface (Dale & Sereno, 1993). Other assumptions are incorporated into the analysis in the head model that describes the conductivity and dimensions of the scalp, skull, brain, and cerebrospinal fluid.

Spatiotemporal models are often developed in interaction with software using heuristic strategies that involve the input of certain assumptions or hypotheses by the human user, and the output of feedback in terms of goodness of fit and source waveforms (cf. Scherg, 1990). Such interactions allow the method to be applied in many different ways, depending on the hypotheses being tested, prior knowledge about the generators, and the nature of the data. The development of models is analogous or equivalent to the development of theories in any area of science: models are evaluated with respect to how well they fit the data; specific models can be tested, compared, and rejected; and models derived from one set of data can be tested with other measurements. In all cases, the constraints, assumptions, and strategies should be specified in such a way that other researchers can test and replicate the results. The decision processes whereby one model was preferred over another should be described clearly.

Methods using fewer sources than electrodes describe the sources in terms of equivalent dipoles. Even assuming an accurate model of the dimensions and conductivities of the head, the location of an equivalent dipole may not necessarily correspond to the real loca-

[10] It should be noted that under the conditions in which PCA may misallocate variance, measurements based on windowed peak measurements will also misallocate variance. The misallocation of variance is a problem general to any analysis of ERP data in which components overlap.

tion of the source when the dipole is modeling the activity of an extended sheet of cortex or several synchronously active sources. Even so, the location and orientation of an equivalent source can still provide useful information, and the time course of source activity can track the overlap of different processes.

Care is required when interpreting differences in source analyses between clinical and control populations, because it is possible that the patients' pathology may have altered generator geometry or conductivity. For example, scalp potential fields can be distorted by skull defects following neurosurgery, which can produce localized paths of low resistance between brain and scalp. Distortions may also occur when the skull is intact, because in large atrophic lesions brain tissue is replaced by CSF, which has a higher conductivity than brain. These issues are of particular importance for source localization techniques that assume a standard head model.

(iii) Source Analysis Should Be Applied Only to Data That Contain Low Levels of Noise

Noise affecting source analysis can occur either traditionally in the form of residual background activity in the average ERP waveforms or topographically in the sense of inaccurate electrode locations. Some effort should be made to illustrate how the topography of the signal has been recorded and has not been distorted by noise or artifacts. Presenting the signal-to-noise ratio is one possibility. A useful check is to present replications, obtained from repeated or split-half measurements, showing that the topography is similar in a pair of measurements.

The topography of the recorded activity (the relative signal amplitude at each electrode) and the change in this topography over time are critical for source analysis. Any manipulation of the data that alters the topography can have a critical effect on the results of the analysis. Baseline correction is one such manipulation, because it incorporates the assumption that the time range over which the baseline is computed contains no source activity. High-pass digital filtering can interact with baseline correction to distort topography. High-pass filtering applied to epoched data can, depending on the algorithm, significantly distort the potentials at the start and end of the epoch. Baseline correction, as typically computed over a period at beginning of the epoch, introduces these distortions into the whole time range of the epoch. High-pass filtering should therefore be applied to the continuous data before conversion to epochs or (failing that) after baseline correction of epoched data. The topography can also be distorted by eye artifacts or by attempts to remove these from the recording using propagation factors. Source analysis profits from a widespread head coverage, and the inclusion of additional electrodes below the standard 10-20 positions (e.g., F9, P9, Iz) is recommended in order to be able to pick up activity from sources in the base of the brain. Using exact electrode positions recorded with a 3D-digitizer, rather than the positions desired during their placement, can alleviate the distortion of topography that results from spatial noise.

(iv) The Goodness of Fit of a Source Model Must Be Determined

How well a model fits the recorded data can be measured in several ways. One technique is to measure the residual variance, which is the percentage of the variance in the data not explained by the model. It is essentially the mean square error between the model and the data expressed as a percentage of the data variance. An equivalent measure is the "goodness of fit," which is the percentage of the data variance explained by the model. When displaying results, the goodness of fit (or residual variance) should be presented over the time range of interest. These measurements depend on the overall strength of the signal at any time point, because the residual variance is expressed as a percentage of the data variance. If there is little recorded activity at a particular latency, the residual variance may be high (and the goodness of fit low) even though the absolute value of the residual variance remains constant. Some measure of data variance such as the global field power (Lehmann & Skrandies, 1980) should therefore be presented in parallel to the goodness of fit.

(v) The Investigator Should Provide Some Assessment of the Reliability of the Sources

Source analysis is often performed on grand-mean data, because such data are relatively noise free. Just as it is incumbent upon the investigator to show the variability of the ERP waveforms, it is similarly necessary to show the variability of the sources from one subject to the next. This can be done by analyzing the sources in individual subjects and describing or plotting the confidence limits for the solutions, or by using the solution for the grand-mean data and plotting the source waveforms obtained in the individual subjects using this solution. Another aspect of the source variability is how different source locations or orientations can explain the data almost as well as the source configuration finally accepted. If the final solution was accepted because it minimized the residual variance, the investigator should describe the range of source locations and orientations that could explain the data with only a small increase in this variance.

L. Statistical Analysis

(i) The Experimenter Must Use Statistical Analyses That Are Appropriate to Both the Nature of the Data and the Goal of the Study

In designing statistical analyses for their data, investigators should not feel bound by one specific or commonly used statistical method. Although parametric statistics have advantages that have rightly given them pride of place, there are many other approaches to statistical inference. In many situations, techniques such as nonparametric statistics, permutational statistics (e.g., Blair & Karniski, 1993), and bootstrapping (e.g., Wasserman & Bockenholt, 1989) may be more appropriate, because they make no assumptions about the distribution of the data. These techniques may be particularly helpful in the analysis of multichannel scalp distributions (Fabiani, Gratton, Corballis, Cheng, & Friedman, 1998; Karniski, Blair, & Snider, 1994). As Tukey (1978) pointed out, statistical analysis can be used as a tool for either decision making or data exploration (heuristics). Hence, investigators should not view statistical analysis as a ritual designed to obtain the blessing of a "level of significance" but as a way to interact with the data.

(ii) Analyses Using Repeated Measures Must Use Appropriate Corrections

Experimental designs with repeated measures are used often in ERP research. In general, univariate ANOVAs are performed on these data. Such ANOVAs assume that the data are normally distributed with homogeneous variance among groups. With repeated-measures data, univariate ANOVAs assume sphericity, or equal covariance among all pairs of levels of the repeated measures. This

assumption is usually violated by psychophysiological data (Jennings, 1987). To compensate for such violations the degrees of freedom can be reduced by calculating epsilon as described by Greenhouse and Geisser (1959) or Huynh and Feldt (1976). Epsilon (ϵ) is a measure (between 1 and 0) of the homogeneity of the variances and covariances. As these become inhomogeneous, the value of ϵ becomes smaller and the degrees of freedom should be reduced before assessing the probability. If this technique is used, the results of a univariate ANOVA with repeated measures and more than two degrees of freedom can be provided using a format which gives the uncorrected degrees of freedom, the corrected p value, and epsilon: $F(29,522) = 2.89$, $p < .05$, $\epsilon = 0.099$ (Jennings & Wood, 1976). Most such cases can be more evaluated precisely using a multivariate analysis of variance (MANOVA) (Vasey & Thayer, 1987), which does not assume sphericity. Not widely appreciated is the fact that MANOVA can be used for analyses involving a single dependent measure. Other approaches that might be used to obtain valid assessments of repeated measurements have been recently reviewed by Keselman (1998).

(iii) Analyses of Scalp Distribution Using Electrode by Condition Designs Should Consider Removing Condition Effects

Topographic profile analyses can be used to determine whether amplitude measurements, obtained at different latencies or in different experimental conditions, reflect the activity of more than one combination of neural generators. It is assumed that ERP activity recorded on the scalp is due to a combination of neural sources located in various brain regions and/or with different orientations. If, in different experimental conditions or different time intervals, the combination of brain source activities is the same, then the corresponding shapes of scalp topographies will be the same. Conversely, if the shapes of the scalp topographies are different in different experimental conditions or at different times within the same condition, then the underlying combination of activities at the brain sources must also be different. The difference can occur if different sources are involved or if the same sources are involved but with different relative strengths (Alain, Achim, & Woods, 1999).

To determine quantitatively whether topographic shapes are different, it is necessary to remove amplitude differences prior to the comparison of shapes. Failure to do so can result in amplitude differences being confounded with shape differences (McCarthy & Wood, 1985). For example, such a confound can occur when using a significant ANOVA interaction between electrode and experimental manipulation to indicate different topographic shapes. One strategy to eliminate this confound is to normalize the data across different conditions by finding the maximum and minimum values in each condition, subtracting the minimum from each data point and dividing the result by the difference between maximum and minimum (McCarthy & Wood, 1985). Unfortunately, this approach may sometimes obscure true differences in topography (Haig, Gordon, & Hook, 1997). Vector scaling, the second strategy described by McCarthy and Wood, however, provides a reliable approach to detecting differences in topography (Ruchkin, Johnson, & Friedman, 1999). In this method the data are scaled so that the RMS values of the across-subject averages from the different conditions (or times) are the same. Within each condition, RMS amplitude is obtained by computing the square root of the over-electrodes mean of squared across-subjects averaged amplitudes. The data within each condition are divided by the RMS amplitude specific for each condition.[11] After the data have been scaled, epsilon-corrected ANOVA or MANOVA can be used to assess the significance of topographic profile interactions with the experimental manipulations. The removal of amplitude differences when analyzing ANOVA electrode by experimental manipulation interactions is only required when the issue is whether topographic shapes are different. In other cases, scaling is not necessary. Furthermore, the interpretation of a detected topographic difference should consider both the unscaled and the scaled data, because points of maximum difference in the original data may become attenuated in the scaled data. When making between-group comparisons with ANOVA, the assumption of equal covariance matrices that underlies their use may be invalidated by the scaling procedure. This problem does not occur with within-group, repeated-measures designs.

(iv) Responses That Are Not Significantly Different Should Not Be Interpreted as Though They Were the Same

One recurrent mistake is to assume that the absence of a statistically significant difference means that the responses are the same. Unfortunately, few statistical tests can prove significant similarities. This mistake usually comes in the following guise. An ERP in condition A is significantly different from the ERP in condition B for group I but not for group II. These findings do not mean that group I is different from group II unless there is a significant group by condition interaction or a significant difference in the A-B differences between the two groups.

(v) When Making Comparisons Between Groups, the Investigator Should Demonstrate Some Homology Between the Components Being Measured

ERPs can differ between groups in many ways. Changes between groups may occur in amplitudes, in latencies, and in scalp topography, and interactions can occur between group effects and experimental manipulations. If one group shows no evidence of a particular ERP component, comparisons are relatively easy. However, other changes in the ERP waveform may be difficult to interpret, because one is never sure that one is comparing the same thing in the two groups. Component identification in patient studies is more complex than usual, because alterations in latency, amplitude, and topography can occur in one or more components of the ERP (Johnson, 1992). Thus, it is important for the experimenter to evaluate whether the patients' ERP components have been correctly identified. For example, arbitrary comparisons (of amplitude or scalp topography) at set latencies will always run into difficulties if there is any reason to believe that the speed of processing differs between the groups. The problem can be illustrated with an example in which the stimuli elicit a large positive peak with a latency of 400 ms in the control subjects and a smaller positive peak with a latency around 560 ms in the patients. The question that must be addressed is whether the peaks at 400 and 560 ms represent activity arising from the same or different generators. Component identification is based on the two most important properties of any ERP component: (1) response to experimental

[11] Other approaches to scaling might also be possible. For example, if differences across subjects is not a concern, the data might be scaled in each condition for each subject by the RMS value for that subject-condition. However, these techniques have not been validated yet, and they may lead possibly to unforeseen problems in multicondition factorial designs. For the present, only the approach described in the text is recommended.

variables and (2) scalp distribution. If the potentials at 400 ms in controls and 560 ms in patients respond to these experimental variables in the same manner and have similar scalp distributions, then, by the definition of components offered by Donchin et al. (1978) these potentials probably represent the same ERP component, and presumably the same brain processes. This conclusion assumes that the latency shift is immaterial to the component's definition, that is, that the same component can appear at different latencies. The experimenter can then reasonably interpret the patients' potential as a delayed version of the control subjects' potential.

(vi) Comparisons Between Groups Should Consider Differences in Variability Between the Groups

Investigators must ensure that clinical data are presented in a form that allows the quality and the variability of the ERP data to be assessed. Even in studies of young healthy subjects, merely to present grand-average waveforms may omit much that is important. When studying clinical cases the use of grand averages is even more bothersome. Because clinical groups are often small and heterogeneous, grand averages, and other measures of central tendency, can give a misleading impression. Almost all patient groups will show smaller amplitudes than normal controls because of increased latency variability in the clinical group. Therefore, averaging ERP data across patients should be avoided or calculated with extreme care. If data are averaged, the presentation of grand averages should be supplemented with representative waveforms from single subjects, and all summary statistics should include measures of variability. A simple way of demonstrating the variability of simple ERP measurements such as latency or amplitude within patient groups and within the normal subjects is to present all the individual data points in a scatter graph or histogram. The reader can then see clearly the extent of overlap between the groups (e.g., Johnson, 1992).

Any investigation of clinical cases has an inherent problem of generalization. Patients always differ in the extent and exact location of the lesion to their brain, and/or in the specific manifestation of their pathology or cognitive dysfunction. Moreover, these differences may be superimposed on different premorbid neuroanatomical variations, different cognitive abilities, and different disease etiologies. The presentation of a single "representative" case is therefore as insufficient as the presentation of the grand mean. If the research goal is to generalize the findings, presenting data from several representative subjects (both those showing the general effects and those not) or from all the individual subjects (if the numbers make this feasible) is essential.

Signal-to-noise ratios will often be lower in the clinical group than in controls because of more lost trials, greater levels of muscle and movement artifact, and lower ERP amplitudes. Thus, the failure to find a significant experimental effect in patients does not necessarily mean that no such effect exists—merely that the statistical power of the contrast was lower than in the control group.

(vii) Single-Case Studies Must Use Properly Matched Control Subjects and Must Demonstrate the Reliability of the Single-Case Data

As in other areas of neuropsychology, single-case ERP studies are of great value but present additional methodological challenges. First, sufficient well-matched controls are required to establish the normal limits for the ERP effect under investigation. Second, the reliability and reproducibility of the data from the patient must be demonstrable. At the least, this verification requires that multiple sets of data be collected and presented in a way that allows them to be compared. Ideally, techniques should be used that permit the presence (or the absence) of an experimental effect to be demonstrated at an appropriate level of statistical significance. Bootstrapping techniques (Wasserman & Bockenholt, 1989) can be helpful in demonstrating differences between a single case and a group of normal subjects.

(viii) In Comparisons Between Groups, Appropriate Statistics Should Be Used to Assess Both Groups and Individuals Within the Group

Studies comparing two groups of subjects can be used in two distinct ways. First, differences can show that psychophysiological processing differs between the groups. Second, differences might show whether a particular individual belongs to one or the other group. In comparisons between clinical subjects and normal controls, this distinction translates into statistically significant differences, which may be used to describe and understand the disorder, and clinically significant differences, which can be used to diagnose the disorder in a particular individual (Oken, 1997). Determining whether a difference is clinically significant requires attention to the standard deviation of the measurements in addition to the standard error of the mean. The best way to demonstrate how a measurement can be used as a diagnostic test is to provide a scatter graph of the measurement in both normal subjects and subjects with the clinical disorder.

The possible diagnostic accuracy of ERP measurements is assessed by evaluating the probabilities of true- and false-positive and true- and false-negative outcomes for the measurement (Sackett, Haynes, Guyatt, & Tugwell, 1991; Swets, 1988). Clinical tests require setting some criterion level that divides the results into positive and negative. A good clinical test is one that much more probably indicates the presence of disease than not when the result is positive ("sensitivity") and much more probably indicates the absence of disease than not when the result is negative ("specificity"). Ultimately, a clinical test is best evaluated in a population that is similar to the subjects who will be assessed. For example, schizophrenic subjects could be compared with other patients presenting with the possible diagnosis of schizophrenia rather than with completely normal subjects.

(ix) Comparisons Between Groups Should Not Be Limited to One Measurement

It is much more powerful to show that one measurement changes whereas another does not than to demonstrate a change in a single measurement alone. Such dissociation can be used to infer meaningful distinctions between lesions in different brain areas (Shallice, 1988) or to differentiate different types or subtypes of psychopathology (Chapman & Chapman, 1973). Alterations in ERP amplitudes and/or latencies are a frequent finding in clinical studies. However, the interpretation of such results depends on whether earlier components also show similar alterations. If all earlier components have normal latencies, one can conclude that the deficit occurs after a normal initial analysis of sensory information. In contrast, if earlier components are also delayed, one would have to show that the later delays are longer to demonstrate that these stages are specifically deficient and not just affected by receiving a delayed input. Another factor vital to the interpretation of clinical data concerns the response of ERP components to experimental variables. In the presence of amplitude and/or latency differences between patients and controls, it is useful to determine whether these measures varied in response to the experimental variables in the same way in both groups. For example, if the patient group

showed significantly reduced or delayed P300s in an oddball paradigm, it is important to determine whether the amplitude was nevertheless inversely related to stimulus probability and whether target stimuli elicited larger P300s than nontargets for both groups (Duncan-Johnson, Roth, & Kopell, 1984). Alterations in the scalp topography of a measured wave are as helpful in determining what is going wrong in a damaged brain as alterations in the wave's amplitude or latency (e.g., Johnson, 1992, 1995).

An especially powerful method evaluates two different tasks in two different patients or two patient groups. A double dissociation occurs if one patient is impaired on one task but not the other and the reverse occurs in the other patient. This dissociation strongly supports the hypothesis that the two tasks require distinct cerebral processes (only one of which is damaged in each of the patients). The same logic can be applied to ERP components that may be affected differentially by different clinical disorders. When possible, more than two levels of the chosen variable should be administered to ensure that the double dissociation is not an artifact of floor or ceiling effects. (see Shallice, 1988, for a review of the difficulties in demonstrating double dissociation).

M. Discussion of the Results

(i) New Findings Should Be Related to Those Already Published

If the experiments are successful, tests of the hypotheses will yield results that were not known before. The final task of the paper is thus to place these new results in the context of what was known before—what was described in the introduction as leading to the present study. It is essential to relate the experimental results to those obtained by others. Similarities should be summarized. Differences should be explained logically by differences in the experimental methods or the types of analyses. If the new data contradict those previously published, it is essential to describe why. New ways of understanding often shine through such discrepancies.

(ii) The Generalizability of the Results Should Be Described

It is important to consider the extent to which the experimental results can be generalized from the actual recording situation and the particular subjects used in the experiments. This generalization can be evaluated by considering the nature of the subject sample and the similarity of results to those recorded by others.

(iii) Unexpected Findings That Were Not Predicted in the Hypotheses Should Be Described When Relevant

Often, the results may contain findings that were not considered in the planning of the experiment but that are relevant to the processes being studied. Although these findings do not have the same scientific weight as those predicted in the hypotheses, they remain important as data from which new hypotheses can be formulated.

(iii) The Implications of the Results Should Be Described

The meaning of the experimental findings must be delineated within the domain described in the experimental rationale and according to the hypotheses formulated in the introduction to the paper. As well, authors should consider their results in relation to adjacent fields of knowledge. If the hypotheses were mainly physiological, what are the implications of the ERP findings for our understanding of human cognition? If the hypotheses were mainly psychological, are there any physiological implications? What are the possibilities of clinical applications? Thus, the discussion begins to prepare the rationale for further experiments and the process of science continues.

N. Conclusions

Science depends on data that are recorded reliably, analyzed properly, and interpreted creatively. A scientist must pay attention to the details and ensure that they are documented sufficiently so that others can replicate published results. The experiments must be designed so that the measurements will test one explanation and rule out others. The data must be measured accurately and analyzed with care to distinguish meaningful effects from noise. A combination of competence, caution, and creativity can lead to powerful interpretations of the world and predictions for the future. The guidelines and recommendations of this paper have attempted to bring these general principles of science into the specific arena of the ERPs.

REFERENCES

Achim, A. (1995). Signal detection in averaged evoked potentials: Monte Carlo comparison of the sensitivity of different methods. *Electroencephalography and Clinical Neurophysiology*, *96*, 574–584.

Achim, A., & Marcantoni, W. (1997). Principal component analysis of event-related potentials: Misallocation of variance revisited. *Psychophysiology*, *34*, 597–606.

Alain, C., Achim, A., & Woods, D. L. (1999). Separate memory-related processing for auditory frequencies and patterns. *Psychophysiology*, *36*, 737–744.

American Electroencephalographic Society. (1994a). Guidelines on evoked potentials. *Journal of Clinical Neurophysiology*, *11*, 40–73.

American Electroencephalographic Society. (1994b). Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *11*, 111–113.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.

Berg, P., & Scherg, M. (1991). Dipole models of eye movements and blinks. *Electroencephalography and Clinical Neurophysiology*, *79*, 36–44.

Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology*, *90*, 229–241.

Bertrand, O., Perrin, F., & Pernier, J. (1985). A theoretical justification of the average reference in topographic evoked potential studies. *Electroencephalography and Clinical Neurophysiology*, *62*, 462–464.

Blair, R. C., & Karniski, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology*, *30*, 518–524.

Blom, J. L., & Anneveldt, M. (1982). An electrode cap tested. *Electroencephalography and Clinical Neurophysiology*, *54*, 591–594.

Brooker, B. H., & Donald, M. W. (1980). Contribution of the speech musculature to apparent human EEG asymmetries prior to vocalization. *Brain and Language*, *9*, 226–245.

Brunia, C. H. M., Möcks, J., van den Berg-Lenssen, M., Coelho, M., Coles, M. G. H., Elbert, T., Gasser, T., Gratton, G., Ifeachor, E. C., Jervis, B. W., Lutzenberger, W., Sroka, L., van Blokland-Vogelesang, A. W., van Driel, G., Woestenburg, J. C., Berg, P., McCallum, W. C., Tuan, P. H. D., Pocock, P. V., & Roth, W. T. (1989). Correcting ocular artifacts—A comparison of several methods. *Journal of Psychophysiology*, *3*, 1–50.

Busey, T. A., & Loftus, G. R. (1994). Sensory and cognitive components of visual information acquisition. *Psychological Review*, *101*, 446–469.

Cadwell, J. A., & Villarreal, R. A. (1999). Electrophysiologic equipment and electrical safety. In M. J. Aminoff (Ed.), *Electrodiagnosis in clinical neurology* (4th ed., pp. 15–33). New York: Churchill Livingstone.

Chapman, L. J., & Chapman, J. P. (1973). *Disordered thought in schizophrenia*. New York: Appleton-Century-Crofts.

Coles, M. G. H. (1989). Modern mind-brain reading: Psychophysiology, physiology and cognition. *Psychophysiology*, *26*, 251–269.

Collewijn, H., Van Der Steen, J., & Steinman, R. M. (1985). Human eye movements associated with blinks and prolonged eyelid closure. *Journal of Neurophysiology*, *54*, 11–27.

Cook, E. W., & Miller, G. A. (1992). Digital filtering: Background and tutorial for psychophysiologists. *Psychophysiology*, *29*, 350–367.

Connolly, J. F., Stewart, S. H., & Phillips, N. A. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language*, *39*, 302–318.

Coren, S., & Hakstian, R. A. (1992). The development and cross validation of the self-report inventory to assess pure-tone threshold hearing sensitivity. *Journal of Speech and Hearing Research*, *35*, 921–928.

Croft, R. J., & Barry, R. J. (in press). EOG correction: Which regression should we use? *Psychophysiology*, *37*, 123–125.

Dale, A. M., & Sereno, M. I. (1993). Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *Journal of Cognitive Neuroscience*, *5*, 162–176.

Deecke, L., Grözinger, B., & Kornhuber, H. H. (1976). Voluntary finger movement in man: Cerebral potentials and theory. *Biological Cybernetics*, *23*, 99–119.

De Jong, R., Wierda, M., Mulder, G., & Mulder, L. J. (1988). Use of partial stimulus information in response processing. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 682–692.

Deoull, L. Y., & Bentin, S. (1998). Variable cerebral responses to equally distinct deviance in four auditory dimensions: A mismatch negativity study. *Psychophysiology*, *35*, 745–754.

Dien, J. (1998a). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments and Computers*, *30*, 34–43.

Dien, J. (1998b). Addressing misallocation of variance in principal components analysis of event-related potentials. *Brain Topography*, *11*, 43–55.

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: Methodology and publication criteria. In J. E. Desmedt (Ed.), *Progress in clinical neurophysiology: Vol. 1. Attention, voluntary contraction and event-related cerebral potentials* (pp. 1–11). Basel, Switzerland: Karger.

Donchin, E., & Heffley, E. F. (1978). Multivariate analysis of event-related potential data: A tutorial review. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potentials research* (pp. 555–572). Washington, DC: U.S. Environmental Protection Agency.

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. 349–441). New York: Academic Press.

Duncan-Johnson, C. C., Roth, W. T., & Kopell, B. S. (1984). Effects of stimulus sequence on P300 and reaction time in schizophrenics. *Annals of the New York Academy of Sciences*, *425*, 570–577.

Echallier, J. F., Perrin, F., & Pernier, J. (1992). Computer-assisted placement of electrodes on the human head. *Electroencephalography and Clinical Neurophysiology*, *82*, 160–163.

Eimer, M. (1998). The lateralized readiness potentials as an on-line measure of central response activation processes. *Behavior Research Methods, Instruments and Computers*, *30*, 146–156.

Elbert, T., Lutzenberger, W., Rockstroh, B., & Birbaumer, N. (1985). Removal of ocular artifacts from the EEG—A biophysical approach to the EEG. *Electroencephalography and Clinical Neurophysiology*, *60*, 455–463.

Fabiani, M., Gratton, G., Corballis, P. M., Cheng, J., & Friedman, D. (1998). Bootstrap assessment of the reliability of maxima in surface maps of brain activity of individual subjects derived with electrophysiological and optical methods. *Behavior Research Methods, Instruments, and Computers*, *30*, 78–86.

Faden, R. R., Beauchamp, T. L., & King, N. N. (1986). *A history and theory of informed consent*. Oxford, UK: Oxford University Press.

Fender, D. H. (1987). Source localization of brain electrical activity. In A. S. Gevins & A. Rémond (Eds.), *Handbook of electroencephalography and clinical neurophysiology: Revised series, Vol. 1. Analysis of electrical and magnetic signals* (pp. 355–403). Amsterdam: Elsevier.

Friedman, D. (1991). The endogenous scalp-recorded brain potentials and their relationship to cognitive development. In J. R. Jennings & M. G. H. Coles (Eds.), *Handbook of cognitive psychophysiology: Central and autonomic nervous system approaches* (pp. 621–656). New York: Wiley.

Friston, K., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S. J., & Dolan, R. J. (1996). The trouble with cognitive subtraction. *NeuroImage*, *4*, 97–104.

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

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Gratton, G. (1998). Dealing with artifacts: The EOG contamination of the event-related brain potential. *Behavior Research Methods, Instruments and Computers*, *30*, 44–53.

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–484.

Greenhouse, W. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95–112.

Gulrajani, R. M., Roberge, F. A., & Savard, P. (1984). Moving dipole inverse ECG and EEG solutions. *IEEE Transactions on Biomedical Engineering*, *31*, 903–910.

Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology*, *28*, 240–244.

Haig, A. R., Gordon, E., & Hook, S. (1997). To scale or not to scale: McCarthy and Wood revisited. *Electroencephalography and Clinical Neurophysiology*, *103*, 323–325.

Halliday, A. M. (1983). Standards of clinical practice for the recording of evoked potentials (EPs). In International Federation of Societies for Electroencephalography and Clinical Neurophysiology (Eds.), *Recommendations for the practice of clinical neurophysiology* (pp. 69–80). Amsterdam: Elsevier.

Hämäläinen, M. S., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography—Theory, instrumentation, and applications to non-invasive studies of the working human brain. *Reviews of Modern Physics*, *65*, 413–497.

Hämäläinen, M. S., & Ilmoniemi, R. S. (1984). *Interpreting measured magnetic fields of the brain: Estimates of current distributions*. Report TKK-F-A559. Espoo, Finland: Helsinki University of Technology.

Hennighausen, E., Heil, M., & Rösler, F. (1993). A correction method for DC drift artifacts. *Electroencephalography and Clinical Neurophysiology*, *86*, 199–204.

Holcomb, H. H., Ritzl, E. K., Medoff, D. R., Nevitt, J., Gordon, B., & Tamminga, C. A. (1995). Tone discrimination performance in schizophrenic patients and normal volunteers: Impact of stimulus presentation levels and frequency differences. *Psychiatry Research*, *57*, 75–82.

Hoormann, J., Falkenstein, M., Schwarzenau, P., & Hohnsbein, J. (1998). Methods for the quantification and statistical testing of ERP differences across conditions. *Behavior Research Methods, Instruments and Computers*, *30*, 103–109.

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics*, *1*, 69–82.

Ille, N., Berg, P., & Scherg, M. (1997). A spatial components method for continuous artifact correction in EEG and MEG. *Biomedical Techniques and Biomedical Engineering*, *42*(Suppl. 1), 80–83.

Jennings, J. R. (1987). Editorial policy on analyses of variance with repeated measures. *Psychophysiology*, *24*, 474–475.

Jennings, J. R., & Wood, C. C. (1976). The ϵ-adjustment procedure for repeated measures analyses of variance. *Psychophysiology*, *13*, 277–278.

Johnson, R., Jr. (1992). Event-related potentials. In Litvan, I., & Agid, Y. (Eds.), *Progressive supranuclear palsy: Clinical and research approaches* (pp. 122–154). New York: Oxford University Press.

Johnson, R., Jr. (1995). Event-related potential insights into altered sensory and cognitive processing in dementia. In F. Boller & J. Grafman (Series Eds.), & R. Johnson, Jr. (Section Ed.), *Handbook of neuropsychology: Vol. 10, section 14. Event-related brain potentials and cognition* (pp. 241–267). Amsterdam: Elsevier.

Karniski, W., Blair, R. C., & Snider, A. D. (1994). An exact statistical method for comparing topographic maps, with any number of subjects and electrodes. *Brain Topography*, *6*, 203–210.

Keselman, H. J. (1998). Testing treatment effects in repeated measures

designs: An update for psychophysiological researchers. *Psychophysiology*, *35*, 470–478.
Keyserlingk, E. W., Glass, K., Kogan, S., & Gauthier, S. (1995). Proposed guidelines for participation of persons with dementia as research subjects. *Perspectives in Biology and Medicine*, *38*, 319–362.
Kutas, M. (1997). Views on how the electrical activity that the brain generates reflects the functions of different language structures. *Psychophysiology*, *34*, 383–398.
Kutas, M., & Hillyard, S. A. (1989). An electrophysiological probe of incidental semantic association. *Journal of Cognitive Neuroscience*, *1*, 38–49.
Kutas, M., & Van Petten, C. K. (1994). Psycholinguistics electrified: Event-related potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83–143). San Diego, CA: Academic Press.
Lagerlund, T. D., Sharbrough, F. W., Jack, C. R., Jr., Erickson, B. J., Strelow, D. C., Cicora, K. M., & Busacker, N. E. (1993). Determination of 10-20 system electrode locations using magnetic resonance image scanning with markers. *Electroencephalography and Clinical Neurophysiology*, *86*, 7–14.
Langley, P., Simon, H. A., Bradshaw, G. L., & Zytkow, J. M. (1987). *Scientific discovery: Computational explorations of the creative process*. Cambridge, MA: MIT Press.
Legatt, A. D. (1995). Impairment of common mode rejection by mismatched electrode impedances: Quantitative analysis. *American Journal of EEG Technology*, *35*, 296–302.
Lehmann, D. (1987). Principles of spatial analysis. In A. S. Gevins & A. Rémond (Eds.), *Handbook of electroencephalography and clinical neurophysiology: Revised series, Vol. 1. Analysis of electrical and magnetic signals* (pp. 309–354). Amsterdam: Elsevier.
Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology*, *48*, 609–621.
Lins, O. G., Picton, T. W., Berg, P., & Scherg, M. (1993a). Ocular artifacts in EEG and event-related potentials. I. Scalp topography. *Brain Topography*, *6*, 51–63.
Lins, O. G., Picton, T. W., Berg, P., & Scherg, M. (1993b). Ocular artifacts in recording EEGs and event-related potentials. II. Source dipoles and source components. *Brain Topography*, *6*, 65–78.
Lütkenhöner, B., Pantev, C., & Hoke, M. (1990). Comparison between different methods to approximate an area of the human head by a sphere. In F. Grandori, M. Hoke, & G. L. Romani (Eds.), *Advances in audiology: Vol. 6. Auditory evoked magnetic fields and electric potentials* (pp. 165–193). Basel, Switzerland: Karger.
Matsuo, F., Peters, J. F., & Reilly, E. L. (1975). Electrical phenomena associated with movements of the eyelid. *Electroencephalography and Clinical Neurophysiology*, *38*, 507–511.
McCallum, W. C., & Curry, S. H. (1980). The form and distribution of auditory evoked potentials and CNVs when stimuli and responses are lateralized. In H. H. Kornhuber & L. Deecke (Eds.), *Progress in brain research: Vol. 54. Motivation, motor and sensory processes of the brain: Electrical potentials, behaviour and clinical use* (pp. 767–775). Amsterdam: Elsevier.
McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, *62*, 203–208.
Miller, G. A. (1990). DMA-mode timing question for A/D converters. *Psychophysiology*, *27*, 358–359.
Miller, G. A., Chapman, J. P., & Isaacks, B. G. (submitted). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*.
Miller, G. A., Lutzenberger, W., & Elbert, T. (1991). The linked-reference issue in EEG and ERP recording. *Journal of Psychophysiology*, *5*, 273–276.
Miller, J., Patterson, T., & Ulrich, R. (1998). A jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, *35*, 99–115.
Miltner, W., Braun, C., Johnson, R., Simpson, G. V., & Ruchkin, D. S. (1994). A test of brain electrical source analysis (BESA): A simulation study. *Electroencephalography and Clinical Neurophysiology*, *91*, 295–310.
Möcks, J., Köhler, W., Gasser, T., & Pham, D. T. (1988). Novel approaches to the problem of latency jitter. *Psychophysiology*, *25*, 217–226.
Möcks, J., & Verleger, R. (1991). Multivariate methods in biosignal analysis: Application of principal component analysis to event-related potentials. In R. Weitkunat (Ed.), *Digital biosignal processing: Vol. 5. Techniques in the behavioral and neural sciences* (pp. 399–458). Amsterdam: Elsevier.
Mosher, J. C., Lewis, P. S., & Leahy, R. (1992). Multiple dipole modelling and localization from spatio-temporal MEG data. *IEEE Transactions Biomedical Engineering*, *39*, 551–557.
de Munck, J. C. (1990). The estimation of time varying dipoles on the basis of evoked potentials. *Electroencephalography and Clinical Neurophysiology*, *77*, 156–160.
Nitschke, J. B., Miller, G. A., & Cook, E. W., III. (1998). Digital filtering in EEG/ERP analysis: Some technical and empirical comparisons. *Behavior Research Methods, Instruments and Computers*, *30*, 54–67.
Oken, B. S. (1997). Statistics for evoked potentials. In K. H. Chiappa (Ed.), *Evoked potentials in clinical medicine* (3rd ed., pp. 565–577). Philadelphia: Lippincott–Raven.
Pascual-Marqui, R. D., Michel, C. M., & Lehmann, D. (1994). Low-resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *International Journal of Psychophysiology*, *18*, 49–65.
Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, *72*, 184–187. (Note: Corrigendum [1990] *Electroencephalography and Clinical Neurophysiology*, *76*, 565.)
Picton, T. W., & Hillyard, S. A. (1972). Cephalic skin potentials in electroencephalography. *Electroencephalography and Clinical Neurophysiology*, *33*, 419–424.
Picton, T. W., & Hink, R. F. (1974). Evoked potentials: How? What? And Why? *American Journal of EEG Technology*, *14*, 9–44.
Picton, T. W., Lins, O., & Scherg, M. (1995). The recording and analysis of event-related potentials. In F. Boller & J. Grafman (Series Eds.), & R. Johnson, Jr. (Section Ed.), *Handbook of neuropsychology: Vol. 10, section 14. Event-related brain potentials and cognition* (pp. 3–73). Amsterdam: Elsevier.
Picton, T. W., & Stuss, D. T. (1980). The component structure of the human event-related potentials. In H. H. Kornhuber & L. Deecke (Eds.), *Progress in brain research: Vol. 54. Motivation, motor and sensory processes of the brain: Electric potentials, behaviour and clinical use* (pp. 17–49). Amsterdam: Elsevier.
Pivik, R. T., Broughton, R.J ., Coppola, R., Davidson, R. J., Fox, N., & Nuwer, M. R. (1993). Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts. *Psychophysiology*, *30*, 547–558.
Polich, J., & Lawson, D. (1985). Event-related potential paradigms using tin electrodes. *American Journal of EEG Technology*, *26*, 187–92.
Ponton, C. W., Don, M., Eggermont, J. J., & Kwong, B. (1997). Integrated mismatch negativity (MMNi): A noise-free representation of evoked responses allowing single-point distribution-free statistical tests. *Electroencephalography and Clinical Neurophysiology*, *104*, 143–150.
Popper, K. R. (1968). *The logic of scientific discovery*. New York: Harper & Row.
Poynton, C. A. (1996). *A technical introduction to digital video*. New York: Wiley.
Putnam, L. E., Johnson, R., Jr., & Roth, W. T. (1992). Guidelines for reducing the risk of disease transmission in the psychophysiology laboratory. *Psychophysiology*, *29*, 127–141.
Regan, D. (1989). *Human brain electrophysiology: Evoked potentials and evoked magnetic fields in science and medicine*. Amsterdam: Elsevier.
Rösler, F., Heil, M., & Hennighausen, E. (1995). Distinct cortical activation patterns during long-term memory retrieval of verbal, spatial and color information. *Journal of Cognitive Neuroscience*, *7*, 57–65.
Ruchkin, D. S. (1988). Measurement of event-related potentials: Signal extraction. In T. W. Picton (Ed.), *Handbook of electroencephalography and clinical neurophysiology: Revised series, Vol. 3. Human event-related potentials* (pp. 7–43). Amsterdam: Elsevier.
Ruchkin, D. S., Johnson, R., Jr., & Friedman, D. (1999). Scaling is necessary when making comparisons between shape of event-related potential topographies: A reply to Haig et al. *Psychophysiology*, *36*, 832–834.
Ruchkin, D. S., Villegas, J., & John, E. R. (1964). An analysis of average evoked potentials making use of least mean square techniques. *Annals of the New York Academy of Sciences*, *115*, 799–826.
Rugg, M. D., & Barrett, S. E. (1987). Event-related potentials and the interaction between orthographic and phonological information in a rhyme-judgement task. *Brain and Language*, *32*, 336–361.
Sackett, D. L., Haynes, R. B., Guyatt, G. H., & Tugwell, P. (1991). The interpretation of diagnostic data. In D. L. Sackett, R. B. Haynes, G. H.

Guyatt, & P. Tugwell (Eds.), *Clinical epidemiology: A basic science for clinical medicine* (2nd ed., pp. 69–152). Boston: Little Brown.

Scheffers, M., Johnson, R., Jr., & Ruchkin, D. S. (1991). P300 in patients with unilateral temporal lobectomies: The effects of reduced stimulus quality. *Psychophysiology*, *28*, 274–284.

Scherg, M. (1990). Fundamentals of dipole source potential analysis. In F. Grandori, M. Hoke, & G. L. Romani (Eds.), *Auditory evoked magnetic fields and electric potentials. Advances in audiology* (Vol. 5, pp. 40–69). Basel, Switzerland: Karger.

Scherg, M., & Berg, P. (1991). Use of prior knowledge in brain electromagnetic source analysis. *Brain Topography*, *4*, 143–150.

Scherg, M., & Picton, T. W. (1991). Separation and identification of event-related potential components by brain electric source analysis. In C. H. M. Brunia, G. Mulder, & M. N. Verbaten (Eds.), *Event-related brain research. Electroencephalography and Clinical Neurophysiology: Supplement 42* (pp. 24–37). Amsterdam: Elsevier.

Schwarzenau, P., Falkenstein, M., Hoormann, J., & Hohnsbein, J. (1998). A new method for the estimation of the onset of the lateralized readiness potential. *Behavior Research Methods, Instruments and Computers*, *30*, 110–117.

Shallice, T. (1988). On method: Single-case studies. In T. Shallice (Ed.), *From neuropsychology to mental structure* (pp. 217–244). Cambridge, UK: Cambridge University Press.

Shibasaki, H., Barrett, G., Halliday, E., & Halliday, A. M. (1980). Components of the movement-related potential and their scalp topography. *Electroencephalography and Clinical Neurophysiology*, *49*, 213–226.

Simons, R. F., Miller, G. A., Weerts, T. C., & Lang, P. J. (1982). Correcting baseline drift artifact in slow potential recording. *Psychophysiology*, *19*, 691–700.

Simons, R. F., Russo, K. R., & Hoffman, J. E. (1988). Event-related potential and eye-movement relationships during psychophysical judgments: The biasing effects of rejected trials. *Journal of Psychophysiology*, *2*, 27–37.

Spencer, K. M., Dien, J., & Donchin, E. (1999). A componential analysis of the ERP elicited by novel events using a dense electrode array. *Psychophysiology*, *36*, 409–414.

Srinivasan, R., Tucker, D. M., & Murias, M. (1998). Estimating the spatial Nyquist of the human EEG. *Behavior Research Methods, Instruments and Computers*, *30*, 8–19.

Stauder, J. E. A., Molenaar, P. C. M., & van der Molen, M. W. (1993). Scalp topography of event-related brain potentials and cognitive transition during childhood. *Child Development*, *64*, 769–788.

Sutton, S. The specification of psychological variables in an average evoked potential experiment. In Donchin, E., & Lindsley, D. B. (Eds.), *Average evoked potentials: Methods, results and evaluations* (pp. 237–297). Washington, DC: National Aeronautics and Space Administration (SP-191).

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*, 1285–1293.

Szirtes, J., & Vaughan, H. G., Jr. (1977). Characteristics of cranial and facial potentials associated with speech production. *Electroencephalography and Clinical Neurophysiology, 43*, 386–396.

Taheri, B. A., Knight, R. T., & Smith, R. L. (1994). A dry electrode for EEG recording. *Electroencephalography and Clinical Neurophysiology*, *90*, 376–383.

Tassinary, L. G., Geen, T. H., Cacioppo, J. T., & Edelberg, R. (1990). Issues in biometrics: Offset potentials and the electrical stability of Ag/AgCl electrodes. *Psychophysiology*, *27*, 236–242.

Taylor, M. J. (1988). Developmental changes in ERPs to visual language stimuli. *Biological Psychology*, *26*, 321–338.

Taylor, M. J. (1995). The role of event-related potentials in the study of normal and abnormal cognitive development. In F. Boller & J. Grafman (Series Eds.), & R. Johnson, Jr. (Section Ed.), *Handbook of neuropsychology: Vol. 10, section 14. Event-related brain potentials and cognition* (pp. 187–211). Amsterdam: Elsevier.

Taylor, M. J., & Smith, M. L. (1995). Age-related ERP changes to verbal and nonverbal memory tasks. *Journal of Psychophysiology*, *9*, 283–297.

Towle, V. L., Bolanos, J., Suarez, D., Tan, K., Grzeszczuk, R., Levin, D. N., Cakmur, R., Frank, S. A., & Spire, J. P. (1993). The spatial location of EEG electrodes: Locating the best-fitting sphere relative to cortical anatomy. *Electroencephalography and Clinical Neurophysiology*, *86*, 1–6.

Tucker, D. M. (1993). Spatial sampling of head electrical fields: The geodesic sensor net. *Electroencephalography and Clinical Neurophysiology*, *87*, 154–163.

Tukey, J. W. (1978). Measurement of event-related potentials. Commentary, a data analyst's comments on a variety of points and issues. In E. Callaway, P. Tueting, & S. H. Koslow (Eds.), *Event-related brain potentials in man* (pp. 139–151). New York: Academic Press.

Tyner, F. S., Knott, J. R., & Mayer, W. B., Jr. (1983). Electrical safety. In F. S. Tyner, J. R. Knott, & W. B. Mayer, Jr. (Eds.), *Fundamentals of EEG technology: Vol. 1. Basic concepts and methods* (Chapter 6, pp. 70–82). New York: Raven Press.

Van Boxtel, G. T. M. (1998). Computational and statistical methods for analyzing event-related potential data. *Behavior Research Methods, Instruments and Computers*, *30*, 8–19.

Van Eys, J. (Ed.) (1978). *Research on children. Medical imperatives, ethical quandaries, and legal constraints*. Baltimore, MD: University Park Press.

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.

Vaughan, H. G., Jr. (1969). The relationship of brain activity to scalp recordings of event-related potentials. In E. Donchin & D. B. Lindsley (Eds.), *Average evoked potentials. Methods, results and evaluations* (pp. 45–75). Washington, DC: National Aeronautics and Space Administration.

Wasserman, S., & Bockenholt, U. (1989). Bootstrapping: Applications to psychophysiology. *Psychophysiology*, *26*, 208–221.

Woldorff, M. G. (1993). Distortion of ERP averages due to overlap from adjacent ERPs: Analysis and correction. *Psychophysiology*, *30*, 98–119.

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

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

(Received March 9, 1999; Accepted May 24, 1999)