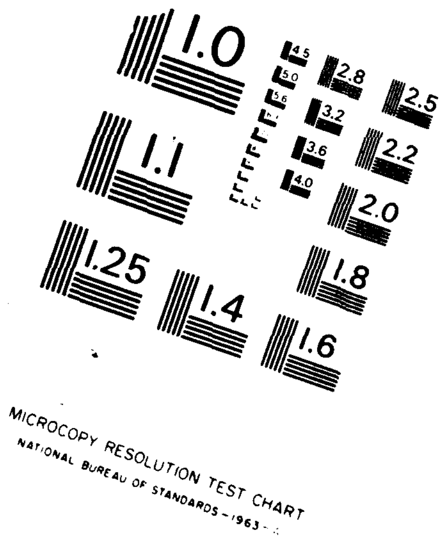
AD-A170 794 THE EVENT-RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING A..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 1/9

UNCLASSIFIED 20 FEB 86 CPL-86-1 F49620-85-C-0041 F/G 5/10 NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

COGNITIVE
PSYCHOPHYSIOLOGY
LABORATORY

Department of Psychology
University of Illinois
Champaign, Illinois 61820

Technical Report No. CPL 86-1
F49620-85-C-0041
January 1986

The Event Related Brain Potential as an Index of Information Processing and Cognitive Activity:

A Program of Basic Research

Annual Progress Report

Prepared for:

The Air Force Office of Scientific Research
Life Sciences Directorate

unclassified
SECURITY CLASSIFICATION OF THIS PAGE

ADA170794

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: unclassified

1b. RESTRICTIVE MARKINGS:

2a. SECURITY CLASSIFICATION AUTHORITY:

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE: -

3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S): CPL 86-1

5. MONITORING ORGANIZATION REPORT NUMBER(S):

6a. NAME OF PERFORMING ORGANIZATION: Department of Psychology, University of Illinois

6b. OFFICE SYMBOL (If applicable): NL

7a. NAME OF MONITORING ORGANIZATION:

6c. ADDRESS (City, State and ZIP Code): 603 East Daniel St., Champaign, IL 61820

7b. ADDRESS (City, State and ZIP Code):

8a. NAME OF FUNDING/SPONSORING ORGANIZATION: Air Force Office of Scientific Research

8b. OFFICE SYMBOL (If applicable):

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F49620-85-C-0041

8c. ADDRESS (City, State and ZIP Code): Bolling Air Force Base, DC 20332-6448

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2313	A1	

11. TITLE (Include Security Classification): The Event-Related Brain Potential as an Index of Information Processing and Cognitive Activity

12. PERSONAL AUTHOR(S): Emanuel Donchin and Arthur Kramer

13a. TYPE OF REPORT: Technical

13b. TIME COVERED: FROM 85Jan01 TO 85Dec31

14. DATE OF REPORT (Yr., Mo., Day): 86Feb20

15. PAGE COUNT: 36pgs + 2 Appendices

16. SUPPLEMENTARY NOTATION:

17. COSATI CODES

FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Event-related brain potential (ERP), P300, Memory, Mental Chronometry, Information Processing, Workload, Automatic vs. Controlled Processing, Resource Reciprocity, Dual Task, Vector Filter

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

We review a program of research designed to understand the event-related brain potential (ERP) so that it can be used as a tool in the study of cognitive function and in the assessment of man-machine systems. We have conducted a series of studies on the functional significance of ERPs and have demonstrated the P300 component is related to memory processes. We have used measures of the same component to evaluate workload, to time mental processes, to study reciprocity of processing resources, and to extend theories of human information processing. We have also made technical advances in the analysis of the distribution of electrical potentials across the scalp.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: NCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION: Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL: Alfred R. Fregly

22b. TELEPHONE NUMBER (Include Area Code): 202/767-5024

22c. OFFICE SYMBOL: AFOSR/NL

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

TABLE OF CONTENTS

1. The Research Strategy

Cognitive Psychophysiology, as its name implies, is a marriage of Cognitive Psychology and Psychophysiology. The basic premise of this union is that the understanding of cognitive processes can be enhanced by augmenting the traditional tools of the cognitive psychologist by adding tools based on the measurement of physiological functions (see 1, 2, 3, 6, 13, 15 -- Appendix A). The psychophysiological data are, of course, useful only to the extent that they complement and expand the view of the mind that can be developed with the use of more traditional techniques. This is the premise that underlies the research described in this report.

1.2 The Event-Related Brain Potential (ERP)

The ERP is a series of voltage oscillations in the brain that can be recorded from the scalp and are time-locked to a discrete event. It is derived by averaging samples (epochs) of the electroencephalogram (EEG) recorded from the human scalp with each sample having the same temporal relationship to a particular event. Note that we can look at activity preceding an event, as well as activity following an event. The voltage oscillations derived in this manner are regarded as manifestations of different "components". Components are defined in terms of their polarity (positive or negative voltage), latency range (temporal relationship to the event), and scalp distribution (variation in voltage with electrode location on the scalp), as well as by their relationship to experimental variables. Components can be quantified using simple magnitude measures or through the application of more advanced techniques such as Principal Component Analysis (PCA) and Vector Analysis (see 10, 11, 13 -- Appendix A). They are labeled by a polarity descriptor (P or N for positive or negative) and a modal latency descriptor (e.g. 300, for 300 msec). Thus, the P300 is a positive ERP component with a modal latency of 300 msec. In some cases, as with Contingent Negative Variation (CNV) and Slow Wave (SW), the numeric descriptors are omitted.

1.3 The Psychophysiological Paradigm

We assume that the voltages we record at the scalp are the result of synchronous activation of neuronal ensembles whose geometry allows their individual fields to summate to a field whose strength can affect scalp electrodes. It is convenient to parse the ERP into a set of components. The component, in our scheme of things, is characterized by a consistent response to experimental manipulations. We further assume that each component is a manifestation at the scalp of an intracranial processing entity. We are not implying that each ERP component corresponds to a specific neuroanatomical entity or that the activity manifested by the component corresponds to a distinct neural process. Rather, we assume that a consistent information processing need, characterized by its eliciting conditions, activates a collection of processes that, for perhaps entirely fortuitous reasons, have the biophysical properties that generate the scalp-recorded activity.

As a working hypothesis we postulate that ERP components are manifestations of functional processing entities that play distinct roles in the algorithmic structure of the information processing system. In other words, we believe that it is possible to describe in detail the transformations that the processing entity applies to the information stream. The goal of Cognitive Psychophysiology, within this framework, is to provide such detailed descriptions. This may be achieved by developing comprehensive descriptions of the conditions governing the elicitation and attributes of the components (the "antecedent" conditions). These descriptions can be used to support theories that attribute certain functions to the subroutine manifested by the component. In turn, the theories should lead to predictions regarding the consequences of the elicitation of the subroutines, predictions that can be tested empirically.

2. Progress Report: The Last Project Year

This section reviews work conducted at the Cognitive Psychophysiology Laboratory (CPL) during the period 01/01/85-12/31/85, with the support of the present project.

In the main, the CPL continued in this period to pursue closely related goals. The primary mission of our research is to develop an understanding of the Event Related Brain Potential (ERP) so that it can be used as a tool in the study of cognitive function and in the assessment of man-machine interactions. To this end, we have conducted studies that fell into four, not altogether distinct, categories, as follows:

A. The elucidation of the functional significance of the ERPs in relation to memory

B. The use of ERPs in studies of cognitive workload

C. The use of ERPs in studies of mental chronometry

D. Methodological studies

Below, we present a systematic review of this research. A list of publications and presentations given during the project period is shown in Appendix A. Appendix B contains a selection of articles, chapters and abstracts.

2.1 Studies of Working Memory

Donchin (1981) proposed that P300 manifests the updating of schemas in working memory ("context updating"). This theory led to the hypothesis that the larger the P300 elicited by a word, the more likely its subsequent recall. In a previous study we tested this hypothesis using a von Restorff paradigm (Karis, Fabiani & Donchin, 1984). We recorded ERPs to words in a series that contained a deviant word (an "isolate"). The isolation was achieved by changing the size of the word. In general, isolated items are better recalled than comparable non-deviant items (the von Restorff effect). We found that subjects who displayed the largest von Restorff effect reported using rote mnemonic strategies, and had a low recall performance. For these subjects, isolates later recalled elicited a larger P300 on initial presentation, than isolates that were not recalled. Subjects displaying the lowest von Restorff effect had the best recall performance and reported using elaborative strategies. For these subjects, P300 amplitude did not predict subsequent recall. Therefore, we concluded that the relationship between P300 amplitude and recall emerges only when it is not overshadowed by subsequent elaborative processing.

2.1.1 P300 and Recall in an Incidental Memory Paradigm

In this experiment Fabiani, Karis, and Donchin (1986) employed an incidental memory paradigm to reduce the use of mnemonic strategies. An "oddball" task consisting of a series of names was presented, and subjects were required to count either the male or the female names. Event-related brain potentials were recorded to the presentation of each name. Following the oddball task, subjects were asked, unexpectedly, to recall as many names as possible. The names that were recalled had elicited, on their initial presentation, larger P300s than names not recalled. Thus, these results confirm our hypothesis: when elaborative strategies are not used, the relationship between P300 and memory emerges more consistently.

2.1.2 Effects of Strategy Manipulation on P300 Amplitude in a von Restorff Paradigm

The interpretation of the data of the Karis et al. (1984) study capitalized on different strategies used by different subjects. In this study Fabiani, Karis, and Donchin (1986) manipulated strategies by instructions to determine whether the relationship between recall and P300 amplitude depends on the subject's mnemonic strategies. Instructions to use "rote" strategies required the subject to repeat each word as it was presented, while "elaborative" instructions required the subject to combine words into images, sentences, or stories.

Nine subjects were run for three sessions in a von Restorff paradigm. The first session was devised to assess the subjects' natural strategies, and their ability to change their strategy according to the instructions. In both the second and third session ERPs were recorded, and subjects were instructed to use one strategy during the first half of the session, and the other during the second half. The order of presentation was counterbalanced across subjects.

Strategy instructions were effective in manipulating the performance of the subjects. When instructed to use rote strategies, subjects recalled significantly fewer words, and displayed a significantly higher von Restorff effect, then when they used elaborative strategies.

ERP analyses on the nine subjects also supported our predictions. When subjects were instructed to use rote strategies, the P300s elicited by words later recalled were significantly larger than those elicited by words later not recalled. However, when subjects were instructed to use elaborative strategies, there was no memory effect on P300 amplitude, but there was a memory effect on a frontal-positive Slow Wave. This component was also observed in the "elaborators" of the Karis et al. (1984) experiment.

Thus, these data also support the theory that P300 manifests the updating of schemas in working memory. The relationship between P300 amplitude and recall is most evident when the subjects base their recall on their first encoding of the word (rote strategies), while it is overshadowed when subjects continue their processing well beyond the P300 time-range (elaborative strategies). There also seems to be a relationship between the amplitude of a later Slow Wave and recall when the subjects use elaborative strategies.

2.1.3 P300 and Memory: A New Study Using A Levels-of-Processing Approach

This experiment will explore the nature of the changes occurring in the memory representation of an event, when a P300 (manifesting the updating process) is elicited. It is hypothesized that P300 manifests the restructuring of the memory representation of an event, and that the amount of restructuring occurring in the representation is indexed by P300 amplitude. Changes occurring in the memory representation are assumed to be beneficial to the subsequent recall of the event, and the data presently available support both these predictions. However, very little is known about the nature of these changes, the reasons why they facilitate recall, and their interactions with stimulus material and subject's strategies in determining subsequent recall performance. This experiment tries to elucidate some of these problems.

In the strategy manipulation experiment outlined above we constrained processing via instructions. An alternative is to devise tasks which require specific types of processing. Tasks can be chosen to require different "levels" of processing, and each level of processing has implications for recall. The level of processing used by the subject should influence the specific aspects of the memory representation that are used for retrieval. For example, if a subject is engaged in semantic processing and encounters a word that is semantically deviant, this may be more useful at retrieval than a word which was orthographically deviant and vice versa. Our hypothesis is that the process manifested by P300 (updating of the memory representation) will always mark (or activate) some attributes of the representation. However, this mark (or enhanced activation) will facilitate recall only when it is congruent with the level of processing, which, in turn, may influence the retrieval search.

A general prediction is that a relationship between P300 and recall will emerge when "shallow" levels of processing are required, but not when "deeper" levels are used. These deep, or semantic levels, will elicit the Slow Wave, and the magnitude of the Slow Wave may be related to the amount of elaboration within a level, or the distinctiveness of an encoding (Craik, 1979). A relationship between P300 amplitude and recall should emerge more strongly after orthographic or phonological processing than after semantic processing. Since orthographic processing should not aid recall, the initial updating of working memory, indexed by P300, will be an important determinant of recall, and the relationship between P300 and recall will be strong. Semantic processing, on the other hand, will increase recall and the relationship between P300 and recall may be overshadowed by this "deeper" processing. Recall improves when elaborative processing increases, and elaborative processing is much more likely to occur during semantic processing than orthographic.

In this experiment subjects will be presented with sequences of words, as in our previous experiments, but a decision will be required after each word. In our basic recall experiment subjects have no overt task to perform while words are being presented. They examine each word and engage in whatever processing they choose in order to prepare for the subsequent free recall task. In this experiment, following the levels of processing approach, each subject will have to process each word in order to make a particular decision with respect to certain attributes of the word. Using a within subject design, the subject will have to make a judgment, in different blocks of trials, based either on a word's orthographic properties (upper case, lower case, letter shapes), or its semantic properties (is it a member of a specific category or related to a particular target word; how pleasant is the word).

2.1.4 Fixed/Varied and Sequential Sternberg Studies

Strayer, Karis, Coles, and Donchin (1984) reported an underadditive relationship between memory set size and response type for P300 latency which was not observed for reaction time. The P300 latency for negative probes was consistently long and did not vary with set size while P300 latency did vary with set size for positive probes. The authors developed a model in which stimuli in memory are either active or inactive. When a probe is presented a serial scan of the active probes is performed and the overt response is based on this outcome. If the probe matches one of the active stimuli, a P300 is elicited at the end of the serial search. Meanwhile, when a probe is presented it activates a relatively constant latency search of long term memory. If the probe does not match any of the active stimuli, then the P300 will be generated when the probe is retrieved from long term memory. The P300 represents a process which maintains or establishes the activation of the probe in working memory. It is important to note that both positive and negative stimuli can be in the active state. Two experiments were designed to test aspects of this model.

Experiment 1 compared two versions of the Sternberg task. The first condition varied the memory set on each trial (hereafter referred to as the varied condition). The second condition used the same memory set for a block of trials (hereafter referred to as the fixed condition). Previous ERP Sternberg studies (e.g., Roth et al., 1975; Gomer et al., 1976; Roth et al. 1977; Adam et al., 1978; Ford et al., 1979; Pfefferbaum et al., 1980; Ford et al., 1982) have failed to observe the underadditive effects of memory set size and response type on P300 latency; however, all of these experiments employed a varied memory set condition, while Strayer et al., (1984) used a fixed memory set condition. Eight subjects received both fixed and varied memory set conditions in counterbalanced order. No systematic differences between the fixed and varied memory sets were observed. Nor were systematic differences observed as a function of the order in which the conditions were presented. However, striking individual differences were observed. Four of the subjects produced underadditive effects, with zero P300 latency slopes for negative stimuli, while the remaining four subjects produced additive effects. Subjects were consistent in their pattern of behavior regardless of the experimental conditions. Examination of the differences between the two groups of subjects revealed no consistent differences in P300 amplitude or wave shape. Thus the data suggest that there are systematic individual differences in the way subjects perform the Sternberg task and that these differences are not due to the fixed and varied memory set conditions as was initially suspected.

Experiment 2 directly tested predictions of the model proposed above. Memory set sizes of 2, 4, and 6 were employed with positive and negative stimuli presented with equal probability. Both negative and positive stimuli could repeat, with the probability that a stimulus repeated =.125. In addition three negative set sizes (2, 4, and 8) were used, where the stimuli in the negative set were the only negative stimuli which could repeat. A fixed memory set condition was employed. We predicted that the pattern of underadditivity would be observed for non-repeating stimuli, but that an additive relationship between memory set size and response type would be produced for repeated stimuli. That is, repeated negative stimuli should also be in an active state and consequently should behave like active positive stimuli. Further, we predicted that this effect would diminish as negative set size increased. This is because as negative set size increases the activation strength of repeated negative stimuli will, on average, diminish thereby increasing the difference in activation strength between positive and negative stimuli.

Ten subjects participated in the experiment. Initial analysis of the P300 latency data support the predictions of the model. An additive relationship between memory set size and response type for repeated stimuli and an underadditive relationship for non-repeated stimuli was obtained. Further, this difference was modulated by the size of the negative set. Analyses of the additional subjects is ongoing. It is useful to note that the model described above shares many similarities with Theios (1974). The pattern of the reaction time data for all subjects is consistent with Theios's model and is inconsistent with Sternberg's (1975) model as well as trace strength theories such as Baddeley and Ecob (1973) and two process theories of familiarity and search such as Atkinson and Juola (1974).

2.1.5 Multiple P300s and Working Memory

This study will examine single and multiple P300 activity to feedback stimuli that induce hypothesis shifts. Levine (1966) presented a paradigm which will be useful to investigate these issues. In a discrimination learning task, subjects indicate which of two stimuli possesses the attribute hypothesized to be "correct." Informative feedback occurs only after every fifth trial. Trials are arranged such that the subject's responses on nonreinforced trials indicate which hypothesis is being used. For example, responses indicating that every stimulus to the left of fixation is correct imply that the subject views "left" to be the correct stimulus attribute. As a result, one can infer which feedback stimuli prompt a change in hypothesis.

In the proposed adaptation of Levine's (1966) paradigm, two letters will be presented sequentially per trial, each of which may be classified along four dimensions: letter (H or B), case (upper or lower), side of presentation (left or right), and value (underlined or not underlined). Consequently, eight hypotheses are possible. ERPs will be recorded to each stimulus and to each of the three possible feedback stimuli (confirming, disconfirming, or noninformative) occurring after each stimulus pair.

Considerable evidence indicates that, as a subject's confidence in his or her classifications increases, P300 amplitude decreases to confirming feedback and increases to disconfirming feedback. This effect should be evident as the subject eliminates hypotheses. Other work indicates that P300s to stimuli possessing the attribute hypothesized to be correct will be larger than P300s to irrelevant-attributable stimuli. This result would allow one to infer the subject's hypothesis without examining overt responses. The extent to which the P300 is indeed related to hypothesis shifts should shed light on the working memory-update model of the functional significance of the P300.

Many feedback paradigms elicit a positive "slow wave" that returns to baseline within 800 msec after the stimulus, following the P300. Johnson and Donchin (1985) contend that the slow wave is composed of overlapping, multiple P300s. An investigation of multiple P300 activity is necessarily largely empirical. Nevertheless, pilot data collected by Gehring and Lorig (1984) indicate that Levine's (1966) paradigm elicits slow wave activity with latencies and scalp distributions that vary with the subject's performance. The proposed adaptation of Levine's (1966) paradigm will allow us to relate multiple P300 activity to each subject's efficiency in performing the task--and thus, to each subject's use of feedback information. Multiple P300 activity may also be related to feedback stimuli prompting different types of hypothesis shifts: to an attribute within the dimension originally hypothesized or to an attribute within a different stimulus dimension. These manipulations should further elucidate the variables affecting multiple P300 activity.

It is anticipated that the study will be conducted during the Summer of 1986 and that analysis of the data set will be completed in the Fall.

2.2 Mental Chronometry

2.2.1 A Psychophysiological Investigation of the Continuous Flow Model of Human Information Processing

The present experiment explored the continuous flow model of information processing by using the P300 component of the event-related brain potential to assess the duration of stimulus evaluation processes, and measures of the electromyogram (EMG) and response force to decompose response processes.

Subjects were required to respond to target letters "H" or "S" by squeezing dynamometers with the left or right hand. Target letters could be surrounded by compatible (e.g., HHHHH) or incompatible (SSHSS) letters. A warning tone preceded the presentation of the letter array by 1000 msec on half the trials. For each trial, latency measures were available for the P300, and correct and incorrect EMG and squeeze activity. The latter measures were also used to classify each trial according to a "degree of error" dimension.

When incorrect squeeze activity was present, execution of the correct response was prolonged, suggesting a process of response competition. This process occurred more often under incompatible conditions, which were also associated with a delayed P300. Thus, the noise/compatibility manipulation influenced both stimulus evaluation and response competition processes. In contrast, the warning tone increased response speed without influencing evaluation time.

These data are consistent with the continuous flow conception and suggest that the latency and accuracy of overt behavioral response are a function of (a) a response activation process continuously controlled by an evaluation process that accumulates evidence gradually, (b) a response activation, or priming, process that is independent of stimulus evaluation, and (c) a response competition process.

A report of this study has been recently published in the Journal of Experimental Psychology: Human Perception and Performance. The expansion of this project includes most of the experiments presented in this report.

2.2.2 Examining Stimulus Evaluation and Response Preparation with Psychophysiological Measures

The purpose of this study was to derive a detailed picture of the activation of response channels over time. The context is a model of information processing that assumes that response are emitted when a hypothetical structure (labelled response channel) is activated at a criterion level. Data from a previous study (Coles, Gratton, Bashore, Eriksen, & Donchin, 1985) indicated that response channels receive input from a mechanism related to stimulus evaluation (a mechanism of "specific activation") and a mechanism independent of the nature of the stimulus (a mechanism of "aspecific activation" that we will label "response priming").

To analyze the contribution of these mechanisms to response channel activity we used two procedures: first, we derived speed/accuracy functions to describe the time course of activation dependent on stimulus evaluation, second, we measured the degree of scalp asymmetry of the Readiness Potential to describe the role and time course of response priming.

At each trial subject were presented tachistoscopically with one of four stimulus arrays (HHHHH, SSHSS, SSSSS, HHSHH). The subject's task was to emit a speeded discriminative response to the central (target) letter of the array, and to ignore the lateral (noise) letters. Note that the noise letters could be the same as the target letter (the compatible noise condition) or of the type calling for the other response (the incompatible noise condition). All the letters were presented in foveal vision.

The set of measures used included analog measures of response force (from which reaction times could be derived), of EMG response, and of ERPs recorded at Fz, Cz, Pz, C3, and C4 referenced to linked mastoids.

The reaction time, error rate, and response classification data (see Coles et al., 1985) replicated our previous findings. The speed/accuracy functions were computed by separating trials with different response latency, and computing accuracy at each response latency. These functions indicated three phases: first, very fast responses (latency less than 200 ms) had a chance level of accuracy for all stimulus arrays; second, responses at intermediate latencies (between 200 and 250 ms) were associated with a higher than chance accuracy for the compatible noise condition, but with a lower than chance accuracy in the incompatible noise condition; third, late response were associated with a high accuracy for both noise conditions.

We interpreted these data as indicating the presence of two phases in the stimulus evaluation process: an early phase where the system does not discriminate between target and noise location, but simply analyzes all of the features presented (we label this phase "feature analysis") and a late phase were the discrimination between target and noise locations takes place (we label this phase "location analysis"). This view is consistent with models of selective attention presented by Treisman and Gelade (1978) and by Eriksen and Yeh (1985). Since feature analysis is sensitive to the noise, we can highlight its time course by computing the difference in accuracy between the compatible and the incompatible noise condition at each moment in time. On the other hand, since location analysis is sensitive to the target location, we can highlight its time course by computing the average between the speed/accuracy functions for compatible and incompatible arrays. These derived functions indicated a partial overlapping between these two phases. However, this overlap may be an artifact due to averaging across trials with different speed of processing (see Meyer et al., 1985 for a discussion of this problem). To partial out the effect of variable processing speed, we estimated the speed of evaluation by measuring P300 latency at each trial. We then computed accuracy as a function of processing time (RT/P300 latency) rather than real time. The functions obtained in this way indicate that the two phases we have previously identified (feature and location analysis) do not overlap, but appear to occur serially (i.e., feature analysis is completed before location analysis is initiated).

As we have said, very fast responses have a chance level of accuracy. These responses can be considered fast guesses. Fast guesses may be due either to a very low response criterion (so that a very small further activation is sufficient to initiate the response - see Grice et al., 1982), or to aspecific (that is, not stimulus related) priming of one of the responses. Aspecific priming implies the activation of response channels (at a sub-threshold level) prior to stimulus presentation. To obtain a direct manifestation of this response preparation in advance to stimulus presentation, we derived measures of the Readiness Potential (Kornhuber & Deecke, 1965). The Readiness Potential is a potential that can be recorded in the period preceding a movement, and is maximum at electrodes placed over the motor cortex contralateral to the movement. We assumed that the Readiness Potential is a manifestation of response preparation, i.e., activation of response channels at a sub-threshold level. Therefore, we reasoned that when the Readiness Potential was larger on the side contralateral to the correct response in advance of stimulus presentation the correct response was prepared, and when the Readiness Potential was larger on the side ipsilateral to the correct response (and contralateral to the incorrect response) the incorrect response was prepared. Thus, we classified trials on the basis of the lateralization of the Readiness Potential prior to stimulus presentation. Indeed, the lateralization of the Readiness Potential predicted the accuracy of fast guesses. We consider these data as evidence that aspecific priming does occur and is, at least in part, responsible for the accuracy of fast guesses.

However, variations in response criteria can still occur. If this is the case, fast responses should be emitted at a lower level of response channel activation than slow responses. To obtain a measure of the activation of response channels at the moment of the response we expanded our study of the lateralization of the Readiness Potential to the whole epoch, rather than limiting it to the period preceding stimulus presentation. Again, we considered the lateralization of the Readiness Potential at a given moment in time as a manifestation of the activation of the response channels. The results indicated that the level of Readiness Potential lateralization at the moment of the response remains fixed, independently of response latency. Thus, the data do not provide support for Grice's variable criterion hypothesis, but rather suggest that the level of activation of the response channels required for the emission of the overt response is fixed, at least for our experimental conditions.

The plan for future expansion of this experiments include the formulation of mathematical models of the processes of response activation and evaluation of their relative merits. Furthermore, we are in the process of writing a report of the study for publication.

2.2.3 A Comparison Between Same/Different and Letter Discrimination Task in a Noise/Compatibility Paradigm

The experiment is based on the comparison between two tasks. In the first task (Letter Discrimination) the subject is asked to respond discriminatively to the central (target) letter of one of four stimulus arrays (i.e., HHHHH, SSHSS, SSSSS, HHSHH). In the other task (Same/Different) the subject is asked to indicate whether the central and lateral letters of the array are the same or different.

This experiment has several purposes. First, we had doubt that the P300 in the Coles et al. (1985) experiment was emitted only after a complete categorization of the stimulus (both central and lateral letters). That is, the processing responsible for P300 latency was not solely related to the task of classifying the target letter. Thus, by using two different tasks, we hoped to shed more light on this phenomonon. Second, we wanted to investigate what happens in the same/different task. Two hypotheses for the advantage for the same response are possible. First, subjects may have a response bias for responding same. Second, early in the evaluation of the stimulus the same response is primed more than the different one. In terms of reaction time and accuracy the two hypotheses make similar prediction. However, by using measures of response preparation (lateralization of the Readiness Potential), of stimulus evaluation (latency of P300 and speed/accuracy functions), and response competition (incorrect EMG and squeeze responses as well as interval between EMG and squeeze onset), we believe we will be able to distinguish between the two hypotheses. Explicitely, if the bias hypothesis is true, then the faster the response, the more likely it is for the subject to give the "same" response, independently of the stimulus presented. Furthermore, measures of the lateralization of the Readiness Potential would reveal a priming of the "same" response. If the early evaluation hypothesis is true, then no lateralization should be observed in advance of stimulus presentation. Furthermore, the speed/accuracy functions should show a tri-phasic shape, with very fast responses at a chance level of accuracy, intermediate latency responses with a high accuracy for the "same" response and a low accuracy for the "different" response, and late responses with high accuracy for all stimuli.

At present, one experimental subject has been run for four sessions. We plan to complete the subject running by February or March, and the data analysis by May.

2.2.4 Effect of Noise/Compatibility and Priming

The experiment is designed around the manipulation of response priming in the context of a noise/compatibility paradigm. The imperative stimulus is one of four stimulus arrays (HHHHH, SSHSS, SSSSS, HHSHH). The task of the subject is to respond to the central target letter. The imperative stimulus is preceded by 1,000 msec (long ISI condition) or by 200 msec (short ISI condition) by a priming (warning) stimulus. Four priming conditions (run in different sessions) are employed: a neutral condition (the priming stimulus is a star [*]), a 50/50 condition (the priming stimulus is a letter [H or S] that does not predict the target letter), a 80/20 condition (the priming stimulus is a letter [H or S] that in 80% of the trials is equal to the target letter), and a 20/80 condition (the priming stimulus [H or S] in 20% of the trials is equal to the target letter).

The purpose of this study is to manipulate priming. The prediction is that the lateralization of the Readiness Potential should be most evident in cases where the priming stimulus is highly informative. Furthermore, by manipulating the ISI, we want to investigate (a) the moment at which the priming (indexed by the lateralization of the Readiness Potential) becomes apparent, (b) the time required for the operation of information reversal present in the 20/80 condition. Pilot data show that subjects are able to use the 20/80 information only with the long ISI.

An additional issue is the role of the priming of the incorrect response channel. Pilot data indicate that, when the imperative stimulus does not correspond to the priming stimulus, the effect of the noise disappears. That is, compatible noise trials are as delayed as the incompatible noise trials. Note that, apparently, the response to compatible noise trials is delayed in this condition, but not that to incompatible noise trials. This would suggest that, at the time the incorrect response is primed, even if by the priming rather than the imperative stimulus, it can still produce the usual noise/compatibility effect. This result occurs at both short and long ISIs. Of course, this is only a preliminary finding and needs confirmation.

The experiment is currently underway. We anticipate that we will complete data analysis by July and have a technical report finished by the end of the next project period.

2.2.5 The Consequences of an Error: P300 and Future Action

Donchin (1981) proposed that P300 manifests the operation of maintainance of task-related schemas in working memory. We investigated this hypothesis within the context of a choice reaction time (RT) paradigm. Seven subjects were asked to respond to a series of names according to their gender. The probability of each gender was manipulated in different blocks of trials: either the male names were rare and female names frequent, or male and female names were equiprobable. For half of the blocks the instructions emphasized speed, for the other half accuracy. The speed and probability manipulations were designed so to produce a strong response bias. Indeed, the subjects responded faster and more accurately to frequent than rare stimuli, and were faster and less accurate under speed than under accuracy instructions. In fact, under speed conditions most responses to rare stimuli were fast guesses.

P300 latency was not significantly affected by these manipulations, although it was longer for error than for correct trials. An examination of the speed/accuracy functions revealed that the longer P300 latency for error trials could only be attributed to an active prolongation of the time required for the emission of P300 in these trials. This implies that the subjects processed their responses before emitting the P300. The amplitude of the P300 to error trials predicted the accuracy to the next occurrence of the same stimulus. Thus, in accord with Donchin's hypothesis, P300 manifested the revision of mental schemas related to the subjects' response bias.

The experiment has been completed, and a partial report is in press (Donchin, Gratton, Dupree, & Coles, in press).

2.2.6 ERPs and Information Value

Previous research has found that the task relevance of a stimulus is an important factor in determining the amplitude of the P300 component of the event related brain potential. This finding suggests that P300 amplitude can be used to measure the amount of information extracted from a stimulus. The present study presented subjects with warning stimuli that differed in the amount of information they conveyed about an imperative stimulus. P300 amplitude, reaction time (RT), and accuracy were measured to assess the utility of the information to the subjects. In addition, P300 latency was measured to determine the locus of the effect. Three subjects were presented with sequences of two types of stimuli, informative or imperative. The informative stimuli were the letters H or S, alongside which a dot appeared. The spatial position of the dot indicated the reliability (50% or 80%) with which the informative cues predicted whether the imperative stimulus would be a match (same letter) or a mismatch (the other letter). Subjects responded to the imperative stimuli by pressing one of two buttons. Analyses of reaction time indicate that subjects extracted the information delivered by the informative stimulus. Their reaction times were faster following the more reliable informative stimuli. Further, the subjects exhibited P300s whose amplitudes were related to the information value of the stimuli. To discriminate between expectancy matches and physical matches, the electromyogram (EMG) is being recorded from the forearms of subsequent subjects. This additional information will help determine whether priming is due to perceptual or response facilitation.

2.2.7 Response Channel Activation and Competition in Orthographic and Phonological Matching Tasks

Previous studies (Polich, McCarthy, Wang, & Donchin, 1983; Kramer & Donchin, in press) have used the rhyme/visual paradigm to examine the interaction of phonological and orthographic codes in word processing. These studies have visually presented the subject with pairs of words, and on half the trials asked them to judge whether the words in the pairs look alike, and in the other trials to judge if they rhyme. The stimulus lists were constructed so that subjects were presented with equal numbers of pairs in which the words both rhymed and looked alike, neither rhymed nor looked alike, rhymed but did not look alike, and looked alike but did not rhyme. It was found that, despite instructions to attend to only orthographic or phonological similarity, for those word pairs in which there was a conflict in orthography and phonology, the subject's reaction time was slowed. The phenomenon was strongest in the intrusion of orthographic information into phonological judgements.

In both the Polich et al. and the Kramer and Donchin papers, ERP data were collected in the hopes of elucidating the locus of the conflict. Both of these studies found that the pattern of P300 latencies mimicked that of reaction time. This suggests that the conflict between phonology and orthography had begun to affect processing by the conclusion of the stimulus evaluation phase. The Kramer and Donchin paper looked at the N200 and found that mismatches between the words in a pair were detected even at that earlier latency. Orthographic mismatches elicited an N200 regardless of whether the subject was performing an orthographic or phonological judgement. Phonological mismatches elicited an N200 only in the phonological judgement task. These data not only show that conflicts within the pairs are detected quite soon after stimulus presentation, but also indicate that orthographic analysis occurs in both tasks, while phonological analysis is carried out only when the phonology of the pair is task relevant.

The current rhyme/visual study replicates previous procedures for examining P300 and N200, and in addition, attention will be directed to the response selection and execution stages of processing. The authors of both papers discussed above indicated that the delays seen in the P300 when a conflict exists are not as large as those seen in the reaction time, and have suggested that additional processing delays probably occur at these later stages of processing. We hope to investigate the response selection and execution stages by looking for lateralized CNV and by monitoring EMG. The subject is asked to respond to the phonological or orthographic similarity of the words in a pair by squeezing a static dynamometer with either his left or right hand, where one hand is designated the "yes" hand and the other is the "no" hand. Through the use of lateral electrode placements, we seek to determine if the subject is preparing to respond with one hand by looking for evidence of a CNV present in only one hemisphere, and will attempt to understand the interaction of such preparation with reaction time and accuracy. Monitoring the EMG activity in the subject's forearms will provide more information about the competition of correct and incorrect responses by alerting us to muscle activity that may be too slight to cause a superthreshold response, but may indicate sub-threshold priming of a response. Thus, with this study the development of the delayed reaction time will have been examined from approximately 200 msec post-stimulus, to the execution of the response itself.

The current study was begun in October 1985. At this time, six pilot subjects have been run and the basic phenomenon appears to have been replicated. Actual data collection will not begin until the stimulus lists have been refined to our satisfaction. It is anticipated that the project will be completed by early Fall 1986.

2.3 Multi-Task Processing: The Assessment of Mental Workload

2.3.1 Processing of Stimulus Properties: Evidence for Dual-Task Integrality

The conditions under which dual-task integrality can be fostered were assessed in a study in which we manipulated four factors likely to influence the integrality between tasks: intertask redundancy, the spatial proximity of primary and secondary task displays, the degree to which primary and secondary task displays constitute a single object, and the resource demands of the two tasks. The resource allocation policy is inferred from changes in the amplitude of the P300 component of the event-related brain potential. Twelve subjects participated in three experimental sessions in which they performed both single and dual tasks. The primary task was a pursuit step tracking task. The secondary tasks required subjects to discriminate between different intensities or different spatial positions of a stimulus. Task pairs that required the processing of different properties of the same object resulted in better performance than task pairs that required the processing of different objects. Furthermore, these same object task pairs led to a positive relation between primary task difficulty and the resources allocated to secondary task stimuli. Intertask redundancy and the physical proximity of task displays produced similar effects of reduced magnitude. The results of this study have recently been published (Kramer, Wickens, & Donchin, 1985).

2.3.2 Resource Reciprocity: An Event-Related Brain Potentials Analysis

The amplitude of the P300 component of the Event-Related Potential (ERP) has proved useful in identifying the resource requirements of complex perceptual-motor tasks. In dual-task conditions, increases in primary task difficulty result in decreases in the amplitude of P300s elicited by secondary tasks. Furthermore, P300s elicited by discrete primary task events increase in amplitude with increases in the difficulty of the primary task. The reciprocity in P300 amplitudes has been used to infer the processing tradeoffs that occur during dual-task performance. The present study was designed to investigate further the P300 amplitude reciprocity effect under conditions where primary and secondary task ERPs could be concurrently recorded within the same experimental situation. Forty subjects participated in the study. Measures of P300 amplitude and performance were obtained within the context of a pursuit step tracking task performed alone and with a concurrent auditory discrimination task. Task difficulty was manipulated by varying both the number of dimensions to be tracked (from one to two), and the control dynamics of the system (velocity or acceleration). ERPs were obtained from both secondary task tones and primary task step changes. Average root-mean-square (RMS) error estimates were also obtained for each tracking condition. The data indicated that increased primary task difficulty, reflected in increased RMS error scores, was also associated with decreased secondary task P300 amplitudes and increased primary task P300 amplitudes. Since the increases in primary task P300 amplitudes were complementary to the decrements obtained for the secondary task, the hypothesis of reciprocity between primary and secondary task P300 amplitudes was supported across several different manipulations of primary task difficulty.

Additional research is currently under way to examine a variety of additional issues. If resources are a limited commodity which can be flexibly allocated to the performance of a given task, it should be possible to manipulate policies of resource allocation during dual task performance as a function of instructional set as well as task difficulty. Thus, if a subject is instructed to pay more attention to one task than another, P300 amplitude should reflect an increased allocation of resources to the attentional task. Another area of interest concerns the question of dual-task integrality as opposed to reciprocity. In other words, under what conditions will two tasks be combined to produce a new "single" task (integrality) instead of competing for resources as in the present study (reciprocity).

2.3.3 Keyboard Project: The "Transfer Experiment"

The present project investigates the information processing associated with the use of a newly designed chord keyboard (Gopher, Koenig, & Karis, 1985). In this keyboard letters and numbers are entered by pressing a combination of codes. The keyboard allows entering codes simultaneously with both hands. A further advantage of this keyboard is that fingers can rest continuously over the keys, so that no travelling or visual control is required at any stage of learning.

Previous experiments (described in Gopher, Koenig, & Karis, 1985) showed that the type of relationship between the codes for the right and left hand assigned to each letter had an important effect on the subject's performance in the transcription task. It was clear that when the same keys were used for each letter in the two-hand condition (the SPATIAL code), the performance was superior than when the same fingers were used (the HAND SYMMETRY code). One possible interpretation of this phenomenon is that the difference between the two types of hand relationship is due to the number of codes to be generated for each letter. This explanation is supported by the observation that, even in cases for which the codes are virtually identical to the HAND SYMMETRY condition, when subjects are asked to learn only one set of codes they are at least as fast as in the SPATIAL condition.

These data indicate a role of the storage and retrieval of the memory representation of the key code in this effect. However, one of the main features of the experiments that has been described so far, with the two-hand keyboard, is that subjects were always informed at the beginning of the experiment on the "double" typewriter capabilities of the machine, and required to learn simultaneously the codes for letters on the two hands. Under these conditions, an important question is, whether the clear superiority of spatial patterns as a representation principle reflects the fact that it is the preferred principle under all conditions. Or, it may only reflect the best solution to the problem of representing two simultaneous and competing codes. The fundamental distinction is between the "natural" functioning rules of the human processing system, and the elasticity of this system in adjusting its operation to best fit the constraints imposed on it by the environment (or the task, in the present context).

Therefore we designed an experiment in which the subjects were first asked to type with only one hand (in this case, the right hand) and then, after five sessions of single-hand typing, instructed to "transfer" from a single-hand typing condition to a two-hand typing condition. Again we divided the subjects into two groups (SPATIAL and HAND SYMMETRY), according to the relationship between the codes assigned to the two hands. Note, however, that the two groups were assigned identical codes for the first five sessions, and were only different in the transfer sessions (there were two sessions of two-hand typing). In the studies described earlier we found that the SPATIAL group had faster responses than the HAND SYMMETRY group. The first question in this study is whether this relationship still holds.

Another purpose of this experiment is to investigate the relationship between right and left responses. Previous studies have shown that, when both responses were required, subjects responded first with the left hand, and then with the right hand (Gopher, et al., 1985). However, studies conducted in Israel with the Hebrew language (where words are written from right to left) show that, in that case, subjects responded right, then left. Thus, the left-to-right ordering can be attributed to reading direction and habitual scanning patterns. However, Gopher, et al. (1985) also reported that single left responses were faster than single right (even though all subjects were right handed, and were faster with their right hands on a control task of simple RT). We speculated that only codes for left responses were stored; when a left letter appeared these were accessed directly and executed. When a right letter appeared, the left codes were again accessed, and a transformation carried out to produce a right hand response. This explained the left hand superiority. Left hand codes would be stored because subjects scan left to right, and in learning focus first on the left. An alternative hypothesis is that the codes, being spatial patterns, are stored in the right hemisphere. Once codes are accessed, left hand responses should be faster.

According to the first hypothesis, since subjects start out learning only the right hand in the transfer study, right hand codes should be formed, and when left responses are finally required, there must be a transformation from right codes to left codes. Thus, right responses should be faster than left. If, however, the hemisphere hypotheses is true, then left responses should be equal or faster than right responses.

In the previous experiments, the latency of the P300 component of the ERP mirrored left-right RT differences. That is, P300 latency to single left was shorter than single right. Several studies indicate that the latency of P300 is sensitive to variables affecting stimulus evaluation processes and largely insensitive to variables affecting response execution (Kutas, McCarthy, & Donchin, 1977; Magliero, Bashore, Coles, & Donchin, 1984; McCarthy & Donchin, 1981; Ragot, 1984). Thus, the parallel behavior of P300 and RT is consonant with the hypothesis above that only the single left codes are stored. The assumption is that P300 is produced only after the appropriate code has been selected. Since left codes are stored, left letters are matched with their codes more quickly than right letters. Before right codes are selected a transformation on a left letter code must be performed, and only after this is completed is P300 produced. If only single right codes are now stored, then P300 latency to single right letters should be faster than to single left letters. If P300s to single left letters are still faster, there is no simple interpretation. Since there is a warning dot, the subject should be focusing on the spot where a letter appears. Thus, letters will not be initially projected to one hemisphere. The hemisphere hypothesis depends on codes being stored in one hemisphere (right), thus having more direct access to the left hand.

A large body of evidence indicates that the amplitude of the P300 component of the ERP can be used as an index of resource allocation (Isreal, Wickens, & Donchin, 1980; Kramer, Wickens, & Donchin, 1984; Sirevaag, Kramer, & Donchin, in press; Wickens, Kramer, Vanasse, & Donchin, 1983). In the studies described earlier, we had found that the amplitude of P300 to single letters presented on the left was equal to the amplitude of P300 to single letters presented on the right, while that to dual letter presentations was much larger. Now we would predict that in the first "transfer" session, when the first single left letters appears, the P300 elicited by the left letters should be larger than the P300 elicited by the right letters. This is because the left letters will require more resources. This difference should decrease over time.

The present experiment is being conducted. Data have been collected and are being analyzed. We expect to complete the data analysis in March, and to write a report in the spring or summer of 1986.

3.4 Technical and Methodological Advances

3.4.1 The Interpretation of the Component Structure of Event-Related Brain Potentials: An Ananlysis of Expert Judgments

The analysis of components of the event-related brain potential (ERP) is accomplished through standardized statistical algorithms. However, visual inspection of ERPs is commonly used to guide the selection of the analytic techniques. Thus, the definitional criteria of components employed by the investigators may interact with the choice of analysis procedures. The present study exmaines the criteria which investigators employ to define components of the ERP, the relative importance assigned to different criteria, and the consistency with which investigators use the definitional criteria. ERPs are simulated so as to vary systematically the amplitude, latency, duration, and electrode distribution of the P300 component. Ten experienced ERP researchers rated the similarity of 153 pairs of simulated ERPs. The ratings were analyzed by a multidimensional scaling procedure. Eight unidimensional judgments on each of the 18 ERPs were also obtained. The results suggest that ERP investigators are capable of accurately recovering th underlying dimensions of a set of simulated ERPs. Furthermore, the degree of experience in analyzing and interpreting ERPs affects the weighting structure of the underlying dimensions. These findings support the commonly held, but previously untested, belief that judgments of ERPs based on visual inspection can be both accurate and reliable.

The results of this study have recently been published (Kramer, 1985).

3.4.2 The Reliability of Measurement of the P300 Component of the Event-Related Brain Potential

The purpose of the study was to assess the reliability of several different procedures employed to measure P300 parameters (i.e., amplitude and latency). The study capitalized on the data collected in the course of a previous study on the relationship between P300 parameters and complex psychomotor skills (Karis, et al., 1983). ERP waveforms were collected from 50 subjects in 5 different oddball tasks. Twenty subjects were retested after approximately one month, and ten subjects were tested a third time after approximately three months.

P300 measures were taken both on average waveforms and on single trials. In the latter case, single trial parameters were later averaged. Reliability was estimated both within session (using a split-half random assignment of single trials) and between sessions. P300 amplitude was assessed using base-to-peak, area, covariance, and PCA derived measures. P300 latency was assessed using peak-picking and covariance procedure. Furthermore, some new procedures were also tested. They included Vector filter and a complex procedures that defined P300 peak on the basis of several attributes of the waveform. Another matter of comparison was to assess the reliability of differences between P300 parameters taken on the average waveforms, compared to P300 parameters measured on difference waveforms.

The data indicate that the different procedures used to measure P300 parameters have a high reliability. In most cases the reliability exceeded .60. Measurement on single trials yielded higher reliabilities than measurement on averages, particularly for latency measures. Amplitude measures had higher reliabilities than latency measures. Vector filter yielded slightly higher reliabilities than the traditional procedures of P300 measurement (i.e., peak-picking, area, etc.). The PCA loadings did show some instability that precluded a thorough judgement of the P300 estimates obtained with this procedure. Finally, P300 measures taken on difference waveforms showed a higher reliability than the difference between P300 measures taken on average waveforms.

The results of this study have been reported in a paper dealing with the problem of the definition of P300 (Fabiani, et al., in press). The planned development of this study is to evaluate the reliability of the scalp distribution of the P300, as well as of the waveshape of the ERP taken as an individual characteristic.

3.4.3 A Multivariate Approach to the Analysis of the Scalp Distribution of Event-Related Potentials: Vector Analysis

This paper presents a procedure, Vector Analysis, for the analysis of the scalp distribution of ERPs. The interest in scalp distribution is based on the assumption that the electrical activity of certain brain structures involved in psychological processes is manifested at the scalp by particular components. The scalp distribution of each component reflects anatomical and physiological properties of the structures involved in the generation of the component, as well as conductive characteristics of the interposed media. Most investigators agree that ERP components are characterized by specific scalp distributions. This is probably due to the invariance of the underlying source(s). Hence, the use of information about scalp distribution may be helpful in the identification and analysis of ERP components.

Vector Analysis considers the values observed at several electrode sites as the sum of several components, each characterized by a specific pattern of scalp distribution, and of background noise. Each ERP component is characterized by a specific scalp distribution. This specific scalp distribution may be expressed by a series of weights, one for each electrode. Each component may vary in amplitude as a function of time and experimental manipulation. Our goal is to obtain estimates of the set of weights describing the scalp distribution of the components <u>and</u> of the variations in amplitude as a function of time and experimental manipulations. Within the framework of our model, such estimates would provide a complete description of the ERPs under study. The adoption of a multivariate approach in obtaining these estimates is emphasized.

One possible application of Vector Analysis consists of filtering for a particular scalp distribution (Vector Filter). Filtering for scalp distribution improves the discrimination between signal (component) and noise (background EEG and overlapping components). The power of Vector Filter is illustrated by a simulation study, in which we compared the accuracy of P300 latency estimates obtained with several techniques. Particular attention has been given to the problem of enhancing the discrimination between signal and noise. The best discrimination, and the most accurate latency estimation, are obtained when the characteristics of both the signal and the noise are considered.

To demonstrate the ability of the procedure to discriminate between overlapping components, the results obtained by applying our approach to a study of P300 in aging were illustrated. The data suggest that the scalp distribution of the P300 peak is different for young adults and elderly people over a variety of tasks. The two groups showed both overall differences and task-related differences. The analysis of the ERPs observed in each task revealed that these differences in scalp distribution cannot be entirely attributed to a generally different distribution of P300 in the two groups, to variations in P300 amplitude, or to a combination of these two factors. Rather, they should be attributed, at least in part, to the presence of overlapping components. This is particularly evident for the subjects in the old group.

Plans for the future include a more thorough exploration of the variations in P300 scalp distribution with age. They will be examined in the prospective of the interpretation of the functional significance of P300 obtained from past research and from the studies of relationship between P300 and memory and P300 and workload presented in other parts of this report.

3.4.4 A Simulation Study of Latency Measures of Components of Event-Related Brain Potentials

We compared the accuracy of P300 latency estimation obtained with different procedures under different signal and noise conditions. The procedures included preparatory and signal detection techniques. Preparatory techniques included frequency filters and spatial filters (channel selection and Vector Filter). Signal detection techniques included peak-picking, cross-correlation, and Woody Filter. Different signal-to-noise ratios were simulated by multiplying the signal (P300) by a scaling factor. Two kinds of noise were added: event-related noise (overlapping components), and non-event-related noise (background EEG).

Accuracy in the latency estimation increased exponentially as a function of the signal-to-noise ratio. Cross-correlation provided better estimates than peak-picking. The results with Woody Filter paralleled those obtained with cross-correlation. Vector Filter provided better estimates than channel selection. The use of frequency filtering reduced the advantage of cross-correlation. Large component overlap impaired the accuracy of the estimates obtained with channel selection, but it impaired the accuracy of the estimates obtained with Vector Filter only when the overlapping component had a scalp distribution similar to that of the signal component. The effects of varying noise characteristics, P300 duration and latency, as well as the parameters of Vector Filter were also investigated.

These results indicate an advantage of using all the available information in the estimation of the latency of ERP components. Characteristics of the noise are also relevant to the choice of the appropriate procedure. Some guidelines for this choice are provided in a paper we recently submitted for publication.

Planned expansion of this project will include a computer simulated assessment of the validity of measures of P300 amplitude. This field has received particular interest in the recent past (see Wood & McCarthy, 1984). However, simulated studies comparing the validity of different method of measuring P300 amplitude have not been published yet.

APPENDIX A

Publications and Papers for Project Period

APPENDIX #	ARTICLES AND CHAPTERS
1.	Coles, M. G. H., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. Journal of Experimental Psychology: Human Perception and Performance, 11, 529-553.
2.	Coles, M. G. H., & Gratton, G. (1985). Psychophysiology and contemporary models of human information processing. In D. Papakostopoulos & I. Martin (Eds.), Clinical and Experimental Neuropsychophysiology. Beckenham, England: Croom Helm, Ltd.
3.	Coles, M. G. H., & Gratton, G. (in press). Cognitive psychophysiology and the study of states and processes. In R. Hockey, A. Gaillard, & M. G. H. Coles (Eds.), Energetic and Human Information Processing. Dordrecht, The Netherlands: Nijhoff.
4.	Donchin, E., Gratton, G., Dupree, D., & Coles, M. G. H. (in press). After a rash action: Latency and amplitude of the P300 following fast guesses. In G. Galbraith, M. Klietzman, & E. Donchin (Eds.), Neurophysiology and Psychophysiology: Experimental and Clinical Applications. Hillsdale, NJ: Erlbaum.
5.	Donchin, E., Kramer, A. F., & Wickens, C. (in press). Applications of brain event-related potentials to problems in Engineering Psychology. In M. G. H. Coles, E. Donchin, & S. Porges (Eds.), Psychophysiology: Systems, Processes, and Applications. Middletown, NJ: Till & Till.
6.	Donchin, E., Miller, G. A., & Farwell, L. A. (in press). The endogenous components of the event-related potential -- A diagnostic tool? In E. Fliers (Ed.), Progress in Brain Research. Amsterdam: Elsevier.
7.	Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (in press). The definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), Advances in Psychophysiology. Volume 2. Greenwich, CT: JAI Press, Inc.
8.	Fabiani, M., Karis, D., & Donchin, E. (1986). P300 and recall in an incidental memory paradigm. Submitted for publication.
9.	Fabiani, M., Karis, D., & Donchin, E. (preliminary draft). Effects of mnemonic strategy manipulation in a von Restorff paradigm.

10. Gratton, G., Coles, M. G. H., & Donchin, E. (1985). A multivariate approach to the analysis of the scalp distribution of event-related potentials: Vector Analysis. Submitted for publication.

11. Gratton, G., Kramer, A. F., Coles, M. G. H., & Donchin, E. (1985). A simulation study of the latency measures of components of event-related brain potentials. Submitted for publication.

12. Heffley, E., Foote, B., Mui, T., & Donchin, E. (1985). PEARL II: Portable laboratory computer system for psychophysiological assessment using event related brain potentials. Neurobehavioral Toxicology and Teratology, 7, 409-414.

13. Kramer, A. F. (1985). The interpretation of the component structure of event-related brain potentials: An analysis of expert judgments. Psychophysiology, 22, 334-344.

14. Kramer, A.F., & Donchin, E. (in press). Brain potentials as indices of orthographic and phonological interaction during word matching. Journal of Experimental Psychology: Learning, Memory, & Cognition.

15. Kramer, A. F., Wickens, C. D., & Donchin, E. (1985). Processing of stimulus properties: Evidence for dual-task integrality. Journal of Experimental Psychology: Human Perception and Performance, 11, 393-408.

16. Sirevaag, E., Kramer, A., Coles, M. G. H., & Donchin, E. (in press). Resource reciprocity: An event-related brain potentials analysis. Human Factors.

ABSTRACTS AND CONFERENCE PRESENTATIONS

17. Fabiani, M., Karis, D., & Donchin, E. (1985). Effects of strategy manipulation on P300 amplitude in a von Restorff paradigm (Abstract). Psychophysiology, 22, 588-589.

18. Farwell, L.A., Chambers, R.D., Miller, G.A., Coles, M.G.H., & Donchin, E. (1985). A specific memory deficit in elderly subjects who lack a P300 (Abstract). Psychophysiology, 22, 589.

19. Gratton, G., Coles, M. G. H., Sirevaag, E., Eriksen, C. W., & Donchin, E. (1985). Examining stimulus evaluation and response preparation with psychophysiological measures (Abstract). Psychophysiology, 22, 592.

20. Kramer, A. F., & Sirevaag, E. (1985). Dual-task processing and visual selective attention: An event-related brain potentials analysis (Abstract). Psychophysiology, 22, 592.

Appendix B1

Journal of Experimental Psychology: Human Perception and Performance

Vol. 11, No. 5 October 1985

A Psychophysiological Investigation of the Continuous Flow Model of Human Information Processing

Michael G. H. Coles, Gabriele Gratton, Theodore R. Bashore, Charles W. Eriksen, and Emanuel Donchin
University of Illinois at Urbana-Champaign

Twelve subjects responded to target letters "H" or "S" by squeezing dynamometers with the left or right hand. Targets could be surrounded by compatible (e.g., HHHHH) or incompatible noise (SSHSS) letters. Measures of the P300 component of the event-related brain potential and of correct and incorrect electromyographic and squeeze activity were used to study stimulus evaluation and response-related processes. When incorrect squeeze activity was present, execution of the correct response was prolonged, indicating a process of response competition. This process occurred more often under incompatible noise conditions, which were also associated with a delayed P300. Thus, the noise/compatibility manipulation influenced both stimulus evaluation and response competition processes. In contrast, a warning tone that preceded array presentation on half the trials, increased response speed without influencing evaluation time. The data suggest that the latency and accuracy of overt behavioral responses are a function of (a) a response activation process controlled by an evaluation process that accumulates evidence gradually, (b) a response priming process that is independent of stimulus evaluation, and (c) a response competition process.

When subjects have to respond to visual displays whose elements call for conflicting responses, their reaction times (RTs) are usually prolonged (e.g., the noise/compatibility effect—Eriksen & Schultz, 1979). The present experiment examined this effect by augmenting the traditional tools of mental chronometry with measures of the latency of the P300 component of the event-related brain potential (ERP) and measures of the electromyogram (EMG). These psychophysiological measures are particularly useful in exploring theories, such as the continuous flow model of Eriksen and Schultz (1979), which attempt to account for the noise/compatibility effect. In particular, the measures can provide information about the interactions between processes associated with stimulus evaluation and processes that are required for the actual execution of responses.

Continuous Flow Models of Human Information Processing

A traditional model of the human information processing system can be traced to

Some of the data reported in this study were presented at the Joint EEG Society/Psychophysiological Society Meeting, Bristol (England), 1983, and at the 7th Evoked Potentials International Conference, Florence (Italy), 1983.

This study was supported in part by a contract from Air Force Office of Scientific Research, Contract #F49620-83-0144, Al Fregly, Project Director; by Public Health Service Research Career Program Award K6-MH-22014 to C. W. Eriksen; and by United States Public Health Service Research Grant MH-01206.

We wish to thank Rich Carlson, Monica Fabiani, Demetrios Karis, Art Kramer, Mick Rugg, Erik Sirevaag, and three anonymous reviewers for their helpful comments on preliminary versions of the manuscript.

Ted Bashore is now at the Medical College of Pennsylvania at Eastern Pennsylvania Psychiatric Institute, Philadelphia, Pennsylvania.

Requests for reprints should be sent to Michael G. H. Coles, University of Illinois, Psychology Department, 603 East Daniel, Champaign, Illinois 61820.

Donders (1969). This model, which has been refined and elaborated by Sternberg (1969), describes a system of elementary processors (i.e., stages) that operate serially. According to this view, a processor is activated upon the *completion* of processing by the preceding element.

An alternative class of models has been proposed in different guises by several investigators (e.g., Eriksen & Schultz, 1979; Grice, Nullmeyer, & Spiker, 1977, 1982; Grossberg, 1982; McClelland, 1979; Turvey, 1973). These models assume that the output of any processor is continuously available to all subsequent, or concurrent, processes. Thus, the partial results of Process A can serve as input to Process B *before* Process A is completed (Eriksen & Schultz, 1979; Grice et al., 1977, 1982; McClelland, 1979).

The continuous flow model of Eriksen and Schultz (1979) is based on the notion that information in the visual modality accumulates gradually over time because of the temporal integrative nature of this sense (Ganz, 1975). According to the model, response activation begins as soon as some visual information is accumulated. Early in the process, the information is consistent with a wide range of responses, and these receive initial activation. As the information continues to accumulate, response activation becomes increasingly focused on responses that remain viable alternatives, given the accumulated data. A given response is actually evoked when the activation of its channel satisfies a criterion. This model assumes, therefore, that during the epoch immediately following the stimulus many responses may be in initial stages of activation. The responses are thus in competition (cf. reciprocal inhibition—Sherrington, 1906). The speed with which a response is executed depends, in part, on the extent of response competition. The greater this competition, the longer the latency of the correct response. A similar model has been proposed by Grice and his colleagues (Grice et al., 1977, 1982).

Consider, for example, the paradigm developed by Eriksen and Eriksen (1974). Subjects are required to move a lever as quickly as possible to the left (right) for the target letter H and to the right (left) for an S. The target letter appears in a clearly defined location, and subjects are instructed to ignore any other letters that occur elsewhere in the visual field. RTs are little affected if the target letter appears flanked by repetitions of itself (compatible noise). However, RTs are appreciably increased if the flanking letters call for the competing response (incompatible noise). Neutral noise letters, that do not call for an experimentally defined response, have an intermediate effect, depending on their feature overlap with the different target letters. If these neutral noise letters share features with the letter that calls for the competing response, they increase RT more than if their features are more congruent with the target letter (Eriksen & Eriksen, 1979; Yeh & Eriksen, 1984).

In accordance with their continuous flow model, Eriksen and his colleagues interpreted the effects of noise/compatibility as evidence that the subject cannot attend solely to the designated target and that both target and noise letters activate their associated responses—that is, they assume a continuous coupling between the processor that analyzes the letter array and the response activation process. The elevated RTs observed when incompatible noise letters appear in the array are due to the activation of both correct and incorrect responses. The responses compete with each other so that the correct response is inhibited and delayed in execution. Furthermore, the effects of feature similarity on RT suggest that the incorrect response can be differentially activated as a function of feature overlap.

On the basis of these studies, Eriksen and his colleagues have argued that the noise/compatibility effect is localized, at least in part, at the response level. To provide further support for this argument, they controlled for the effect of differences in stimulus complexity between compatible and incompatible arrays by assigning each of the two responses to different stimuli (Eriksen & Eriksen, 1979). The subject was instructed to move a lever to the left in response to an H or C and to the right in response to an S or K. In this arrangement, the compatible displays can be as visually complex (e.g., HCH) as the incompatible arrays (e.g., KCK). The data indicated that RT is determined predominantly by the compatibility of the flanking noise and not by the visual heterogeneity of the stimulus array.

Although the continuous flow model appears to provide a satisfactory account of the noise/compatibility effect, the data obtained by Eriksen and his colleagues could also be explained by a strictly serial, discrete stage model, if a number of assumptions are made. Such a model might assume that the stimulus array is "evaluated" in a stimulus evaluation stage and that the results of this evaluation are then passed to a decision stage that identifies the appropriate response. The output of this decision stage is passed to a response execution stage for action. Where would the conflict arise in such a model? Perhaps the conflicting stimuli require a longer evaluation time. Or the full set of information available in the stimulus may be fed to the decision stage so that the choice of response is slowed. It is also possible that a weaker or slower signal is passed to the response execution stage when the preceding stages are subject to conflict.

One of the major differences between continuous flow and serial stage models is the emphasis given to response processes. Serial discrete models (e.g., Sternberg, 1969) typically devote little concern to responses and how they are activated. Their implicit assumption seems to be that on tasks such as choice RT, the end product of the processing stages is a decision or response selection stage whose discrete output is the activation of the appropriate response. As we have seen, continuous flow models (and variable criterion theory, Grice et al., 1977, 1982) do not provide for a separate decision stage responsible for activating or initiating responses. Rather, responses are emitted whenever one of the response channels is activated at a criterion level. This criterion may vary somewhat over trials and conditions, as the subjects adjust their performance to the standards of accuracy expected. Thus, responses can be evoked at different levels of percept development, depending upon the preset criterion, a conception that is consistent with latency operating characteristics or speed–accuracy trade-off functions (Lappin & Disch, 1972a, 1972b).

Another way in which response channels may be activated is through a response priming process that is independent of the nature of the stimulus that is presented and may even precede stimulus presentation. "Aspecific priming" (*aspecific* because the priming is independent of a specific stimulus) may be triggered by such factors as instructions, set, expectancy, pay-off schedules and the like (Eriksen & Schultz, 1979). Note that variations in aspecific priming and variations in response criterion have the *same* influences on response latency and accuracy. Responses that are primed independently of the nature of the stimulus will require less stimulus-related activation for their evocation. Similarly, when subjects lower their criteria for a particular response, less stimulus-related activation is required for an overt response to be given.

As we have seen, the continuous flow model invokes several mechanisms and processes to account for the behavior of overt response systems in the noise/compatibility paradigm. First, there is a process of *stimulus evaluation* that continuously feeds information about the stimulus to associated response activation systems. Second, there is a process of *response competition* by which concurrently activated responses inhibit each other. Third, a process of *aspecific priming* or a mechanism of a *variable response criterion* affects the amount of stimulus-related response activation required for overt response execution. In the next section, we demonstrate how psychophysiological and graded response measures can be used to investigate these processes and mechanisms in the context of the choice RT paradigm used by Eriksen and his colleagues (1974).

Measures

Stimulus-Related Processing

The continuous flow model proposes that responses can be activated throughout the stimulus evaluation process. This view implies that the duration of the evaluation process cannot always be inferred from RT.

One traditional method used to measure the duration of stimulus evaluation has been to derive speed–accuracy trade-off functions (e.g., Pachella, 1974). This method assumes that the accuracy of a response is a function of the evidence accumulated at the time the response is emitted. Thus, by determining the RT associated with a specified level of accuracy, it is possible to infer the duration of stimulus evaluation. However, this method assumes that the duration of stimulus evaluation processes is constant over trials. This assumption may not be valid in all circumstances (e.g., Meyer

& Irwin, 1982). Thus, the speed–accuracy trade-off function may not provide an accurate description of stimulus evaluation processes. For this reason, we need a measure of the duration of stimulus evaluation processes *on each trial*. This measure should be unaffected by those processes associated with response selection and execution.

In the present experiment, we use the latency of the P300 component of the ERP as an estimate of the duration of stimulus evaluation. This use of P300 latency was proposed by Donchin (1979) primarily on the basis of two observations. First, he noted that the P300 component is elicited by the rarer of two events that occur in a Bernoulli sequence (see Duncan-Johnson & Donchin, 1977). It turns out that the rule according to which events are categorized can be quite abstract. Because the "rareness" of an event cannot be established until the event has been properly categorized, it is plausible to suggest that the latency of the P300 depends, at least in part, on categorization, or stimulus evaluation, time. The second observation is that, although both P300 latency and RT are sensitive to categorization time, the two measures can be dissociated (Kutas, McCarthy, & Donchin, 1977). As has been noted by many (for example, see Kutas et al., 1977), the latency of P300 may be shorter than, longer than, or equal to the RT associated with an overt response to the same stimulus. Indeed, the correlation between RT and P300 latency is sometimes high and positive, and sometimes close to zero. It is plausible therefore to propose that P300 latency and RT are determined by two, partially overlapping, sets of processes. The degree to which the two measures are correlated will depend on the extent of the overlap between the two sets of processes.

Several studies have confirmed the view that the set of processes that must be completed before P300 is emitted are related to stimulus evaluation but not to response execution. For example, Kutas et al. (1977) required subjects to categorize each of a series of stimuli into one of two classes and to indicate their decision by making a discriminative button-press response. There were three categorization tasks that were given under both speed and accuracy instructions. The first task required subjects to discriminate the name *Nancy* from the name *David;* in the second task, subjects were presented with a list of first names and had to determine which were male names and which were female; in the third task, subjects were presented with a list of words and had to decide whether a given word was a synonym of the word *prod*. Note that the tasks required increasingly complex levels of categorization for their successful execution. Under accuracy conditions, the latency of P300 increased systematically as the level of categorization increased. In the speed condition, P300 latency was shorter for the David/Nancy task than for the other two tasks. The instructions (speed/accuracy) had a large effect on RT (136 ms) but a small effect (19 ms) on P300 latency. The instructions also had an effect on the *correlation* between RT and P300 latency. The correlation was significantly higher when the subjects were instructed to be accurate. It would appear, then, that when subjects try to be accurate, there is more overlap between the sets of processes that determine RT and P300 latency. These data suggest that P300 latency is (a) sensitive to manipulations of stimulus evaluation time (i.e., complexity of the categorization task), and (b) relatively insensitive to manipulations of response-related processes (i.e., speed vs. accuracy instructions).

A more direct test of the proposed relation between P300 latency and stimulus evaluation time was conducted by McCarthy and Donchin (1981). In this experiment, subjects had to execute a choice response as a function of a target word (LEFT or RIGHT) embedded in a 4 × 6 matrix. On half the trials, the rest of the matrix was filled with (#) signs; on the other trials, randomly selected letters of the alphabet completed the matrix. When the background was made up of letters, it was more difficult to detect the target word, and RT correspondingly increased. Another variable that affected RT was response compatibility. On every trial, a warning stimulus (the word SAME or OPPOSITE) preceded the presentation of the matrix. The words occurred in a random sequence and instructed the subjects to respond with the same hand as that indicated in the matrix or with the opposite hand. Thus, the word LEFT could call for a left- or right-hand response depending on the warning stimulus. The type of matrix—(#s) or letters—had a significant effect on both RT and P300 latency, while the response compatibility manipulation significantly affected RT (91 ms) but not P300 latency (16 ms). This result was replicated and extended by Magliero, Bashore, Coles, and Donchin (1984), who found that the effect of matrix type on P300 latency was evident in counting as well as in RT tasks. These authors also found that graded changes in the confusability of the target word and background characters were associated with graded changes in both P300 latency and RT. As in the studies of McCarthy and Donchin, response compatibility had a large effect on RT and a small effect on P300 latency.

It should be noted that the assertion supported by the data reviewed above is that there are processes that have a significant effect on RT but that do not have an effect on P300 latency. In general, these are processes that appear to have a direct relation to the execution of the response. Strong support for this view is provided in a study by Ragot (1984). In this study subjects were instructed to respond with either crossed or uncrossed hands to stimuli that called for a left- or a right-hand response. The cost of hand crossing in RT was substantial (57 ms). However, crossing the hands had no significant effect on P300 latency (2 ms). It would seem, then, that there is strong evidence to support the claim that P300 latency is largely determined by factors that are independent of the "motor" execution of the response.

There remains some controversy regarding the processes that do affect P300 latency. Ragot (1984) noted that it is possible to detect a small effect (19 ms) of "spatial incompatibility" between stimulus and response. This effect is observed with some regularity even though it tends to be small and often not significant. Coles, Gratton, and Donchin (1984) examined this issue and concluded that such effects of spatial incompatibility can be viewed in terms of strategic changes in the evaluation process. For these reasons, it is possible to use P300 latency as an estimate of the duration of the stimulus evaluation process (cf. Brookhuis, Mulder, Mulder, & Gloerich, 1983; Duncan-Johnson & Kopell, 1981; Ford, Roth, Mohs, Hopkins, & Kopell, 1979; Hoffman, Houck, MacMillan, Simons, & Oatman, 1985).[1]

Response-Related Activity

The concepts of response priming and response competition both imply that the activation of the response systems can occur in a graded fashion, without necessarily achieving the level at which an overt response is actually manifested. Thus, to obtain a detailed description of these processes, we need measures of partial response activation that are more sensitive than measures of the overt manifestation of the response. We use EMG measures and "subthreshold" overt responses to provide such a description.

When electrodes are placed over the muscles involved in the overt response, the difference in electrical potential (EMG) can provide information about both the presence and the timing of response activation. Furthermore, although muscle activation must occur if an overt motor response is to be executed, it is possible for muscle activation to occur without a subsequent overt response if either the activation is weak or if the overt response is aborted. Thus, measures of EMG can be used to assess both the presence of partial response activation as well as the time at which response activation has achieved a particular threshold level for the muscles to be activated.

A second method for assessing partial response activation processes involves the use of an analog response device (such as a dynamometer) rather than a discrete manipulandum (such as a response button). If subjects are required to squeeze a dynamometer with a certain force in order to register a response, then measures of the dynamometer's output can be used to assess both the presence and temporal characteristics of an overt response. As with the EMG, such squeeze responses may not achieve the criterion force level for a "response" to be counted, just as a response button may not be pressed to the point of contact

[1] Note that we are not asserting that P300 is a manifestation of the stimulus evaluation process itself. Rather we propose that P300 is related to a process that is invoked only after stimulus evaluation has been completed (Donchin, 1981; Karis, Fabiani, & Donchin, 1984). In this regard, we should also note a technical consideration. In the present study, P300 latency is assessed on each trial. Because the P300 occurs in a background of EEG activity, special algorithms are required for its detection. In particular, these procedures involve a search for the *peak* of the P300 rather than its onset. Thus, the onset of the P300 process (and the end of stimulus evaluation) can be assumed to have occurred some time before our measure of the latency of the peak (by at least 100 ms). Thus, P300 latency provides a measure of relative, and not absolute, evaluation time.

closure. These partial squeezes may occur if response activation is insufficiently strong or if the response is aborted before complete execution.

These two measures, EMG and dynamometer output, are used in the present experiment to assess the processes of response priming and response competition in the following way. When there is EMG or squeeze activity in a response channel, *but* there is nothing in the stimulus array (target or noise) to call for activation of that response channel, we assume that the process of *aspecific priming* has occurred. If a particular manipulation leads to an increase in the level of aspecific priming of a response channel, less additional activation is required for the threshold for motor response activity to be reached. Therefore, the incidence of EMG and squeeze responses should increase, and they should occur at shorter latencies. *Response competition* is revealed by changes in the temporal aspects of the execution of one response that are associated with the concurrent activation of the other, "competing," response. For example, overt response initiation (as manifested by the EMG) may be delayed, and/or the interval between overt response initiation and completion (as manifested by a squeeze) may be longer, if there is concurrent activation of the other response channel. This concurrent response activation may or may not achieve the thresholds associated with EMG and squeeze activity.

The utility of the EMG measures in the study of response competition is illustrated by the results of a preliminary investigation by Eriksen, Coles, Morris, and O'Hara (in press). These authors measured EMG responses as well as overt motor activity (button presses) in the Eriksen paradigm. Subjects had to respond with the thumbs of the two hands as a function of the target letter. The EMG was recorded from each forearm. Trials were sorted on the basis of the flanking noise (compatible or incompatible) and the presence or absence of EMG activity on the incorrect side. Eriksen et al. (in press) found that incorrect EMG activity occurred more often on incompatible trials and that this incorrect activity tended to appear earlier than the correct EMG activity. Further, on trials when incorrect EMG activity was present, the correct EMG and motor response latencies were delayed. These data provide evidence for a response competition mechanism. The data are also consistent with the continuous flow interpretation of the noise/compatibility effect, because trials on which response competition was evident were more prevalent when the noise was incompatible. However, even when there was no EMG activation on the incorrect side, RTs were still longer for the incompatible arrays. Thus, there was insufficient evidence to attribute all of the noise/compatibility effect on RT to response competition.

Present Experiment

Our psychophysiological exploration of the paradigm described by Eriksen and his colleagues (Eriksen & Schultz, 1979) focuses on the effects of three manipulations. In addition to the noise/compatibility manipulation, we used (a) a manipulation (WARNING) that should affect response-related processes (and EMG and squeeze latency) and (b) a manipulation (BLOCKING) that should affect stimulus evaluation processes (and P300 latency). In this way, we provided different conditions under which RT and P300 latency should be both associated and dissociated.

Noise/Compatibility

We required subjects to make a discriminative response as a function of the central (target) letter in a five-letter array. The flanking noise letters were either the same as the target letter (compatible noise condition) or were those associated with the opposite response (incompatible noise condition). We know from previous research reviewed above that the noise/compatibility manipulation affects RT. In particular, RT is longer in the incompatible noise condition. Eriksen and Schultz (1979) proposed that this effect is due to a greater incidence of response competition. However, this proposal has never been tested directly except in a preliminary study by Eriksen et al. (in press). In the present experiment, we addressed this issue by using measures of partial response activation (EMG and squeeze). We predicted that partial activation of the incorrect response would occur more often in the incompatible noise condition. Furthermore, we looked for direct evidence for the response competition mechanism by evaluating the temporal characteristics of correct response execution when the incorrect response was partially activated. Note that we did not expect that response competition would be absent in the compatible condition. Because incorrect responses can be primed in advance of stimulus presentation, response competition might occur even when the array did not contain information for the incorrect response.

Incompatible noise might also delay processes that occur before response activation. Measures of RT cannot distinguish between this kind of delay and one that is due to response competition. Thus, we obtained measures of the latency of P300 to evaluate the possibility that incompatible noise delays stimulus evaluation. In addition, to understand the elementary processes involved in stimulus evaluation, we examined speed–accuracy trade-off functions. This analysis was designed to study differences in the way information is accumulated in the compatible and incompatible noise conditions.

Warning

On half the trial blocks, a warning tone preceded the presentation of the arrays by 1,000 ms. The tone informed the subject about the timing of array presentation but conveyed no information about the nature of the array. This kind of alerting stimulus should speed RT by facilitating motor preparation rather than stimulus evaluation (cf. Posner, 1978). RT measures cannot easily distinguish between stimulus evaluation effects and motor processes. However, the latency of P300 should be sensitive only to variations in stimulus-related processes. Thus, we predicted that P300 latency would be unaffected by the provision of an alerting stimulus. On the other hand, measures of motor processes (EMG and squeeze) should be affected. In particular, if the level of aspecific priming is higher following the warning, then we would predict that partial response activation should be more evident in warned than in unwarned conditions.

Blocking

Finally, we evaluated the effects of fixing the level of noise within trial blocks. The level of noise was either constant or variable for a series of trials. This manipulation was chosen to study the stimulus evaluation process in detail. In particular, we wanted to create conditions for which complete evaluation was unnecessary for successful task performance (cf. Kutas et al., 1977). By presenting only compatible noise arrays in a trial block, we gave subjects the opportunity to respond correctly without localizing the central target letter, because all the letters in the array were the same. When the noise was always incompatible, the evaluation process could also be facilitated because the central letter was consistently different from the lateral letters. Thus, we predicted that P300 latency (and RT) would be shorter when the level of noise was fixed within a block of trials.

Method

Subjects

Twelve male students at the University of Illinois (between the ages of 18 and 23) served as subjects. They were paid $3.50 per hour, plus a bonus for participating in all sessions.

Design

Subjects were required to make a discriminative response as a function of the target letter in a five-letter stimulus array. They received 12 blocks of 80 trials during each of two sessions. The first 8 blocks of the first session were considered training, and the data obtained from these blocks were not used in the analysis. The remaining 1,280 trials (16 blocks) were divided as follows:

Task. In half (8) of the blocks the subjects were instructed to *respond* with one hand to the target letter H, and with the other to the target letter S. The relation between responding hand and target letter was counterbalanced across subjects. In the other half of the blocks the subjects were instructed to *count* one of the two target letters (counterbalanced over subjects).

Noise. On half the trials, the target letter was surrounded by the same letter (*compatible* noise); on the other half, the surrounding letters were those calling for the opposite response (*incompatible* noise).

Blocking. In half of the blocks, the *fixed* condition, only one type of noise was presented (compatible or incompatible), whereas in the other half, the *random* condition, both types of noise were presented at random. In each case, the probability of each target letter was .5.

Warning. For half the blocks, a warning tone preceded the stimulus. In the other half, no warning was given.

As a result of these manipulations, 80 trials were obtained for each of 16 conditions defined by the factorial combination of two types of task, two types of noise, two types of blocking, and two levels of warning. Note that, with the exception of noise, the level of each variable was always constant for a given block of trials. Trial blocks were randomly ordered with the constraint that no more than two consecutive blocks could have the same level of task, warning, or blocking.

Apparatus and Procedure

On each trial, one of four stimulus arrays, HHHHH, SSSSS, SSHSS, and HHSHH, was back-projected on a translucent screen using a Kodak random access slide projector. Stimulus duration (100 ms) was controlled by a shutter. The interval between two consecutive stimulus presentations varied randomly between 4,500 and 6,500 ms. The subject sat facing the screen at a distance of two meters so that the angle subtended by each letter was 0.5°. Thus, the visual angle subtended by the entire array was 2.5°. A fixation point, placed 0.1° above the location of the central target letter, remained visible throughout the experiment.

In the *respond* conditions, the task of the subject was to respond to the central target letter (H or S) by squeezing one of two zero-displacement dynamometers (Daytronic Linear Velocity Force Transducers, Model 152A, with Conditioner Amplifiers, Model 830A; see Kutas & Donchin, 1977). The force applied to the dynamometer was transformed into a voltage by the transducer. This voltage was digitized at 100 Hz for 1,000 ms following array presentation. The output of the transducer was processed by a circuit to determine when the force exceeded a prescribed criterion value. This value defined the occurrence of an overt response and was used to determine RTs. Before the practice trials, the value of each subject's maximum squeeze force was determined for each hand separately. Then, criterion values corresponding to 25% of maximum force were established. During the practice trials, a click was presented to the subject over a loudspeaker whenever the force exerted on the transducer crossed the criterion.

In the *count* condition, subjects were required to count the number of trials on which a designated central target letter was presented. For half the subjects, the counted letter was H, while for the others it was S.

On half the blocks, a warning tone (1000 Hz, 50-ms duration, 65 dB re 20 $\mu N/m^2$) preceded the presentation of the array by 1,000 ms. These blocks constituted the *warned* condition. Note that the interstimulus interval (time between arrays) was the same for both warned and unwarned blocks.

For half the blocks, the level of noise (compatible or incompatible) was *fixed* within a block; for the other half it was *random*. Thus, in the fixed condition only two of

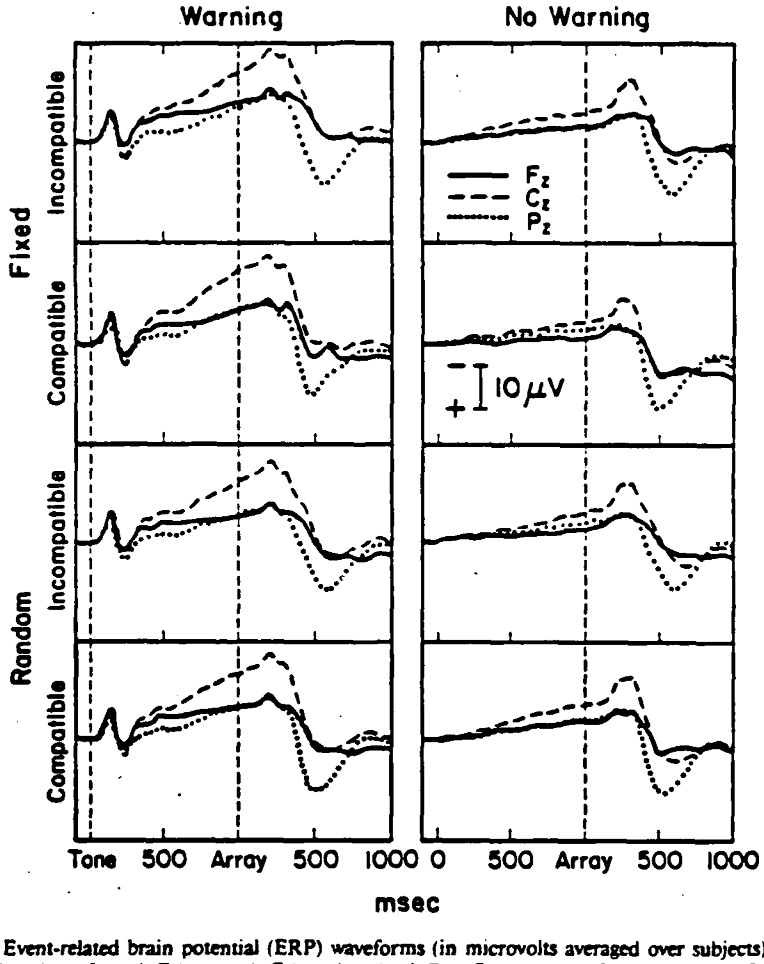


Figure 1. Event-related brain potential (ERP) waveforms (in microvolts averaged over subjects) for three electrode locations: frontal (Fz), central (Cz), and parietal (Pz). (Separate waveforms are shown for the eight different experimental conditions of the respond task.)

the four arrays were presented, while in the random condition any one of the four arrays could occur on any trial.

Psychophysiological Recording

The electroencephalogram (EEG) was recorded from Fz, Cz, and Pz (according to the 10/20 international system, Jasper, 1958) referenced to linked mastoids using Burden Ag/AgCl electrodes affixed with collodion. Vertical electrooculographic activity (EOG) was recorded from Burden electrodes placed above and below the right eye. The EMG was recorded by attaching pairs of Beckman Ag/AgCl electrodes on both the right and the left forearm using standard forearm flexor placements (Lippold, 1967). For EEG and EOG electrodes the impedance was less than 5 Kohm; for EMG, impedance was below 15 Kohm.

The EEG and EOG signals were amplified by Grass amplifiers (model 7P122), and filtered on-line using a high-frequency cut-off point at 35 Hz and a time constant equal to 8 s. The EMG signals were conditioned using a Grass Model 7P3B preamplifier and integrator combination. The preamplifier had a ½ amplitude low-frequency cut-off at 0.3 Hz, while the output of the integrator (full-wave rectification) was passed through a filter with time constant of 0.05 s.

In each case, the derived Voltage × Time functions were digitized at 100 Hz, for an epoch of 2,100 ms starting 1,100 ms before array presentation. For the warned condition, this provided a 100-ms sample before the presentation of the warning tone.

Data Reduction

Overt responses. As we noted above, the subjects were required to squeeze the dynamometers to a criterion of at least 25% of maximum force to register a "response." Thus, an overt response was deemed to have occurred if this criterion was achieved, and RT was defined as the interval between array onset and the point at which the criterion was crossed. By evaluating the outputs of both force transducers, we were able to establish both the accuracy and the latency of these overt responses on every trial.

The squeeze response requirement was used to provide additional information about the dynamics of overt response execution. Thus, the output of the force transducer could be used not only to assess when the force exerted by the subject crossed the criterion but also to determine when an overt response was initiated. In particular, we established the minimum value of output of the force transducer that was discriminable from noise. This value became the criterion for overt response *initiation*, and the time at which this occurred was used to define the latency of squeeze *onset*.

In this way, for each squeeze of either dynamometer to criterion, two latency measures were available: the latency of squeeze onset and the RT. Because the outputs of both dynamometers were evaluated on each trial, these two measures were available for both correct and incorrect responses. Furthermore, on some trials overt responses were initiated but not completed—that is, the force exerted did not exceed the 25% criterion. Thus, for these trials we were able to determine both the presence and latency of "partial" squeezes. When they occurred, these partial squeezes were generally made by the incorrect hand and were accompanied by complete overt response execution by the correct hand.

Psychophysiological data. For every trial, the variance of the EOG activity was computed. When this exceeded a preset criterion, the data from that trial were discarded. In fact, this occurred for less than 10% of the trials. To provide a sense of the ERP waveforms recorded under the conditions of the experiment, we show in Figure 1 the grand average ERPs for the eight conditions of the RT experiment. Note that negative going potentials are represented by an *upward* deflection of the curve.

For the warned condition, we note a response to the warning stimulus followed by a slow increase in negativity (particularly at Cz) that may correspond to the contingent negative variation (CNV, Walter, Cooper, Aldridge, McCallum, & Winter, 1964). The stimulus array elicits a "classic" P300 characterized by maximal positivity at the Pz electrode. In the unwarned condition, we also see the classic P300 following presentation. The ERP data for the count conditions will not be considered in detail. These conditions were included to confirm that any effects of the independent variables on ERP measures in the RT task could not be attributed to the motor response requirement.

The single-trial data from the three scalp electrodes (Fz, Cz, and Pz) were smoothed using a low-pass digital filter (high-frequency cut-off point at 3.14 Hz, two iterations). The three waveforms were then combined to yield a composite waveform by differentially weighting the three electrodes (vector filter, Gratton, Coles, & Donchin, 1983). The weights were chosen to reflect the scalp distribution usually observed for P300 (Pz > Cz > Fz). This procedure has proved to be both reliable and valid (Gratton, Kramer, & Coles, 1984; Fabiani, Gratton, Karis, & Donchin, in press). P300 latency was then estimated by finding the latency of the maximum value of the composite waveform in a time window between 300 ms and 1,000 ms after array presentation. In this way, for each individual trial, except those where excessive eye movements occurred, a value for P300 latency was obtained.[2]

For the respond task only, the integrated EMG activity from both arms was evaluated on each trial. The integrated EMG traces typically exhibited small, unsystematic variation prior to array presentation. Following the array, a response was observed in one or both traces. To determine the latency of the onset of an EMG response and to evaluate whether an EMG response was present, a criterion value was established. This was accomplished using a procedure similar to that described above for the onset of squeeze activity. Thus, we determined (for each subject) the minimum value of the integrated EMG output sufficient to discriminate a change from random variations in background EMG. When the integrated EMG exceeded this criterion, an EMG response was deemed to have been initiated, and the latency of this activity was noted. As with

[2] We should note that we also used a more traditional method, peak-picking at Pz, to determine the latency of P300 on single trials. There was a close correspondence between the data obtained using the traditional procedure and those from vector filter. However, analyses of latency measures derived from the vector procedure yielded consistently higher F values than those based on the peak-picking procedure.

the squeeze responses, EMG responses in both arms could be observed on the same trial.

Results and Discussion

This section is organized in the following way. First, we present the results of an analysis of the RT and error data. This will show that we have replicated the effects of noise/compatibility reported by Eriksen and his colleagues and that both warning and blocking have effects on these measures. Second, to provide evidence that partial response activation occurs in this paradigm, we present analyses of graded responses. Then, we consider how partial activation is related to measures of the latency of the psychophysiological and squeeze responses. Next, we review the data relating to the effects of the three manipulations—noise, warning, and blocking—on partial activation and stimulus evaluation. Finally, we present speed–accuracy trade-off functions for the different conditions of the experiment as well as for different latencies of the P300 responses.

Reaction Time and Error Rate

The RT data replicated the results reported by Eriksen and Eriksen (1974). Subjects responded faster to compatible noise arrays (397 ms) than to incompatible noise arrays (444 ms). Furthermore, both warning and blocking manipulations affected RT. When a warning tone preceded the presentation of the stimulus array, RTs were shorter (410 ms) than when no warning was given (430 ms). When level of noise was fixed within a block of trials, RTs were shorter (413 ms) than when both compatible and incompatible arrays could occur (428 ms). However, the advantage for the fixed condition was more pronounced for compatible arrays (19 ms) than for incompatible arrays (11 ms).[3] These effects can be seen in Figure 2. They were supported by an analysis of variance (ANOVA) on mean correct RTs for each subject and each of the eight conditions (defined by the three manipulations), which revealed significant main effects of noise, $F(1, 11) = 129.59$, $p < .001$; warning, $F(1, 11) = 44.39$, $p < .001$; and blocking, $F(1, 11) = 15.60$, $p < .01$; and a significant interaction between blocking and noise, $F(1, 11) = 5.14$, $p < .05$. Note that, for this analysis, RT was defined as the latency at which the squeeze response crossed the criterion (25% of maximum force).

Errors (defined as squeezes above the 25% force criterion with the incorrect hand) were analyzed using a similar ANOVA. Mean error rates for the different conditions are shown in Figure 2. Subjects made more errors in response to incompatible noise arrays than on compatible noise trials, $F(1, 11) = 30.97$, $p < .001$. However, the effects of noise and blocking interacted, $F(1, 11) = 34.53$, $p < .001$. In fact, fixing the level of noise for a block of trials reduced the error rate for the incompatible noise condition but increased the error rate for the compatible noise condition.

When these data are considered together with those for RT, the following picture emerges. For compatible noise, error rate is larger and RT shorter for the fixed than for the random condition. This suggests that subjects adopt a less conservative strategy in the fixed condition. In contrast, for incompatible noise, error rate is smaller and RT shorter for the fixed than for the random condition. This pattern of data cannot be readily explained in terms of a difference in the conservatism of the response criterion. Rather, it appears that the processing of the incompatible array is facilitated in fixed versus random conditions. As we shall discuss later, we believe that this processing advantage is actually present for both compatible and incompatible conditions. However, it is not apparent in the compatible condition because of a concurrent change in strategy. The problem of interpretation introduced by variations in response strategy may be resolved by the P300 data, which we consider later.

Graded Response Analysis

One major aim of this experiment is to explore the role of response competition and aspecific priming in the noise/compatibility paradigm. In this section, then, we consider evidence for the presence of partial response activation. Next, we review the results of analyses of the effects of the three experimental manipulations on both the frequency and the

[3] When a significant interaction was obtained, an analysis of simple main effects was performed to interpret the interaction. In all cases, the alpha level was set at .05.

latency of partial activation of response channels.

The EMG and squeeze measures serve as the basis for identifying four levels of response activation for each of the two response channels (left/right or correct/incorrect hand). These levels are zero activation, EMG activation, partial squeeze activation, and criterion squeeze activation (a squeeze with at least 25% maximum force). In principle, then, we could have identified many different configurations of response activation in our data set. However, the number of configurations is limited for both practical and theoretical reasons. First, the levels of activation within a given channel are not independent. Thus, if a criterion squeeze is evident in a channel, EMG activation and partial squeeze activation *must* have occurred in that channel. This restricts the number of possible configurations to 16. Second, trials on which neither channel achieves a criterion squeeze level are uninteresting because, in traditional terms, no response occurred. Third, some configurations occur so infrequently that reliable estimates of their characteristics are not possible. For example, subjects seldom exhibit criterion squeeze responses in both channels. These considerations led us to consider only four response configurations for the purposes of classifying the

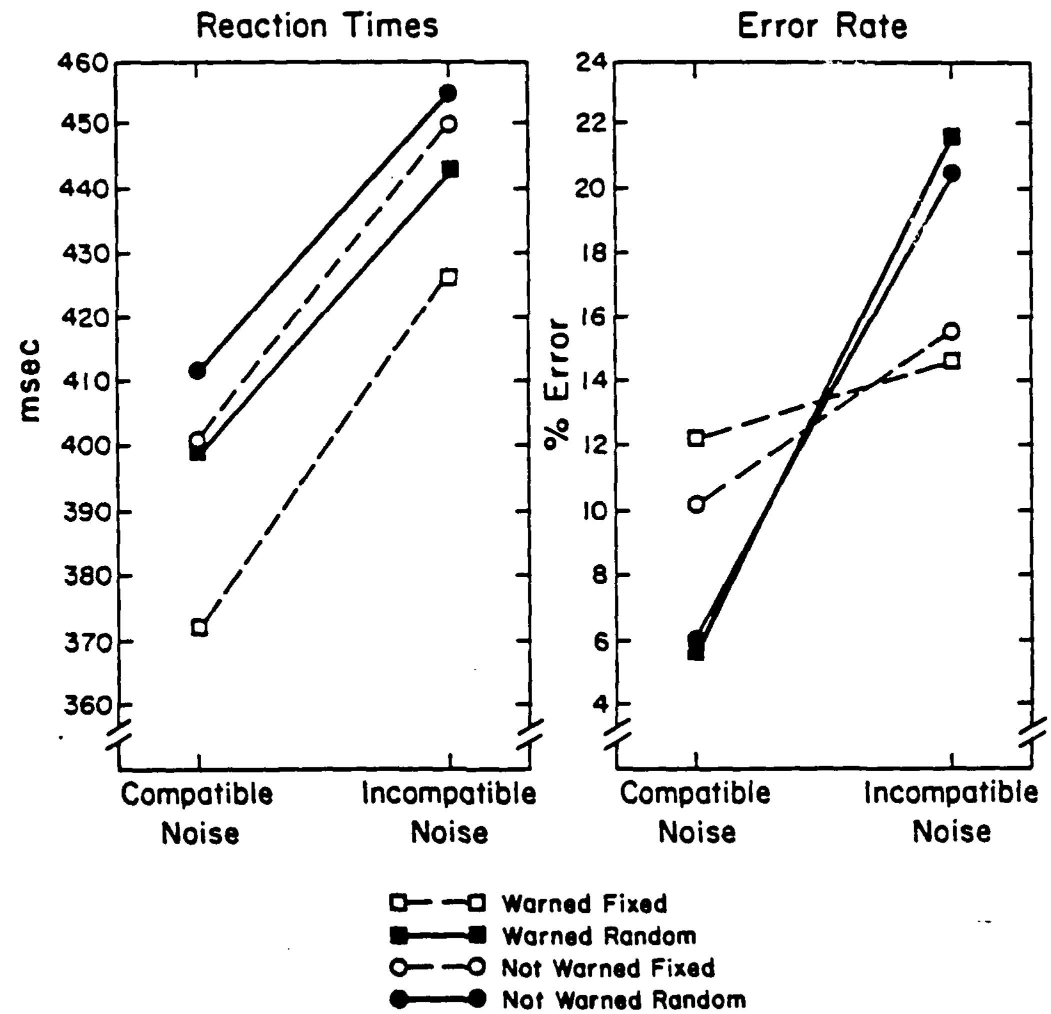


Figure 2. Reaction times (in milliseconds) and error rates as a function of noise, warning, and blocking conditions.

trials. These four configurations have the property of including (a) completely correct trials, where there is no evidence of partial activation of the incorrect response channel; (b) correct trials for which there is partial activation of the incorrect response channel at the level of the EMG; (c) trials with squeeze activity on both sides, which, depending on whether or when a criterion squeeze occurs, may be correct or incorrect; and (d) completely incorrect trials, which may or may not include partial EMG activation of the correct channel. In fact, 99.4% of all trials could be classified into one of these four categories.

The formal definitions of the four configurations are as follows:

- N — Activity only on the correct side in EMG and squeeze channels. (*N*o activity on the incorrect side)
- E — Activity on the correct side for EMG and squeeze channels; activity also present for EMG on the incorrect side. (*E*MG activity on the incorrect side)
- S — Activity on the correct side for EMG and squeeze channels; activity also present for both EMG and squeeze channels on the incorrect side. The incorrect squeeze may or may not reach the 25% of maximum force criterion. (*S*queeze activity on the incorrect side)
- Error — Activity on the incorrect side for EMG and squeeze channels; EMG activity on the correct side may or may not be present. However, no correct squeeze activity is present.

Note that in terms of a conventional error analysis, trials classified as either N or E would be considered "correct" trials. On the other hand, trials classified as Error would be considered "incorrect" trials. The S trials might be considered either correct or incorrect, depending on the magnitude and timing of the two squeeze responses. However, on most trials, the incorrect response (partial or complete) preceded the correct response (see below).

For each subject and each of the eight conditions, we determined the number of trials falling into each of the four categories described above (N, E, S, and Error) and then expressed the frequency of trials in each category as a percentage of the total number of trials for that condition. The mean percentages over subjects and conditions were N = 47%, E = 31%, S = 16%, Error = 6%. Thus, on 47% of the trials (E and S), partial activation of the incorrect response channel occurred even though the correct response was also activated. Note that half the S trials were counted as incorrect responses in the traditional error analysis described earlier.

In spite of our efforts to assure that each response category was associated with a sufficient number of trials, for 1 subject for some conditions no trials were classified in the N category. This subject's data were not considered in any of the subsequent analyses. For 7 other subjects, the Error category was sometimes empty. The data for these subjects were retained for most of the analyses. The frequencies with which trials were classified in each category as a function of condition are shown Figure 3.

Latency analysis. We now consider the relation between our response classification system and measures of the latencies of EMG and squeeze onset for the correct side, EMG and squeeze onset for the incorrect side, and P300. The effects of the experimental manipulations on these latency measures are also analyzed.

Figure 4 shows mean latency values for the different conditions of the experiment for each of the five latency measures. The data are segregated for the four response categories. To highlight the effects of the noise/compatibility and warning manipulations, we present the latency data for these manipulations in Figures 5 and 6, respectively. The latter two figures also provide information about the frequency of the different response categories for the two manipulations.

a. *Response classification.* The analyses to be reported in this section are designed to address three questions: (a) Does our response classification system represent a "degree of error dimension"? (b) Does response competition occur when two response channels are activated concurrently? (c) Is the degree of error related to the time required to evaluate the stimulus?

Inspection of Figures 4, 5, and 6 suggests that the pattern of latencies varies with response category. These variations are consistent with the view that the response categories

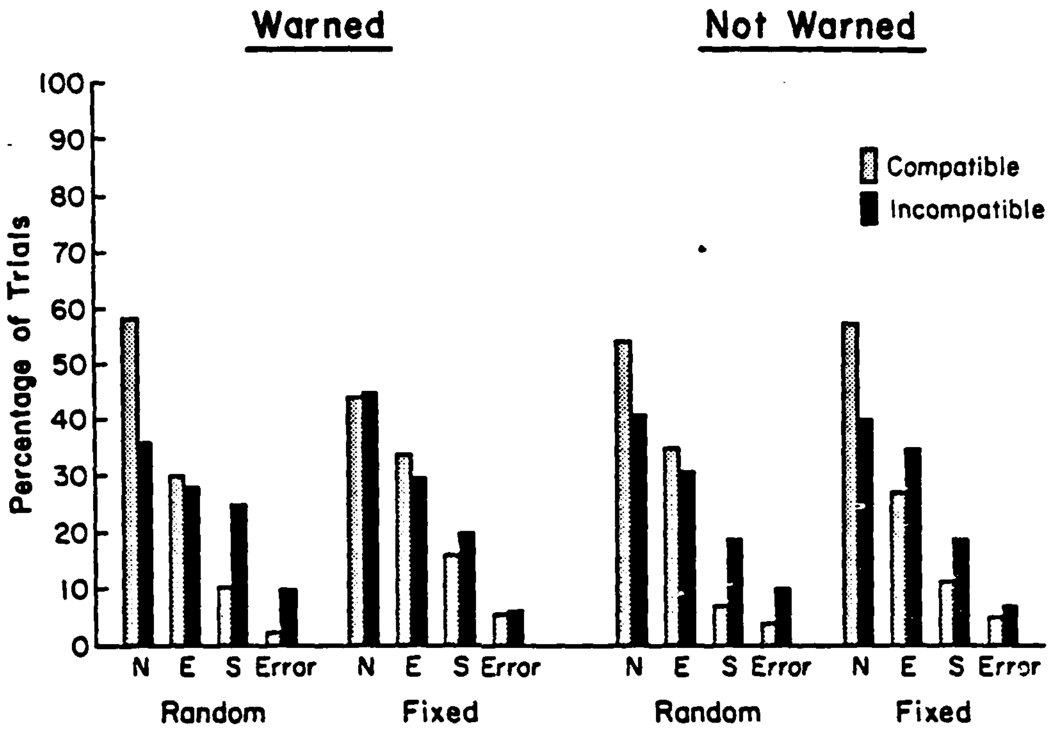


Figure 3. Frequency distributions of trials as a function of the four response categories. (*N*, *E*, and *S* are correct response trials associated with either no [N], electromyographic [E], or squeeze [S], activity on the incorrect side. *Error* trials are associated with an incorrect squeeze and no correct squeeze activity [see text]. Separate distributions are shown for the eight different experimental conditions.)

can be considered as ordered levels of a degree of error dimension. In fact, the onset latency of correct motor activity (both EMG and squeeze) increases monotonically from the N to E to S categories. Similarly, the latency of the incorrect motor activity decreases monotonically from the E to S to Error categories. These conclusions are confirmed by ANOVAs whose results are reported in Table 1.[4] Thus, for both correct and incorrect response channels, the latencies of EMG and squeeze onset are longer when activity is present on the other side. Furthermore, there is a larger increase when the contralateral activity includes a squeeze than when it includes only EMG activity. Because responses are delayed to the degree that activation of the competing response channel occurs, these data satisfy our criterion for the existence of a response competition mechanism.

Further support for the response competition mechanism comes from an analysis of the interval between the initiation of the correct response (as shown by the onset of EMG activity) and its execution (as shown by the onset of squeeze activity). This interval was longer for the S (80 ms) than for the E (53 ms) and N (57 ms) categories, $F(2, 20) = 32.30$, $p < .001$. These results indicate that as the amount of motor activity on the contralateral side increases (from N and E to S), the execution of the correct response is disrupted.

---

[4] Whenever a significant main effect was obtained for a factor with more than two levels, Tukey's HSD test (Tukey, 1953) was used to determine which levels were significantly different from each other (alpha level = .05). For the latency of correct activity (both EMG and squeeze onset), the S category was longer than N or E, which did not differ significantly from each other. For the onset latency of incorrect EMG activity, the Error category was shorter than the E category. Note that, whenever the Error category was included in an analysis, the ANOVA was based on the data from 4 rather than 11 subjects. However, the picture that emerges from the analysis of 4 subjects replicates that provided by the whole sample of 11 subjects as far as the differences among N, E, and S are concerned.

Thus, not only is the onset of a response delayed when there is squeeze activity in the contralateral side but also the actual execution of the response is prolonged. These data confirm the existence of a response competition mechanism.

A further interesting finding comes from a comparison of the latency of the onset of the squeeze response on the correct and incorrect sides. This comparison can be performed only when squeeze activity is present on both sides—that is, for the S category. In this case,

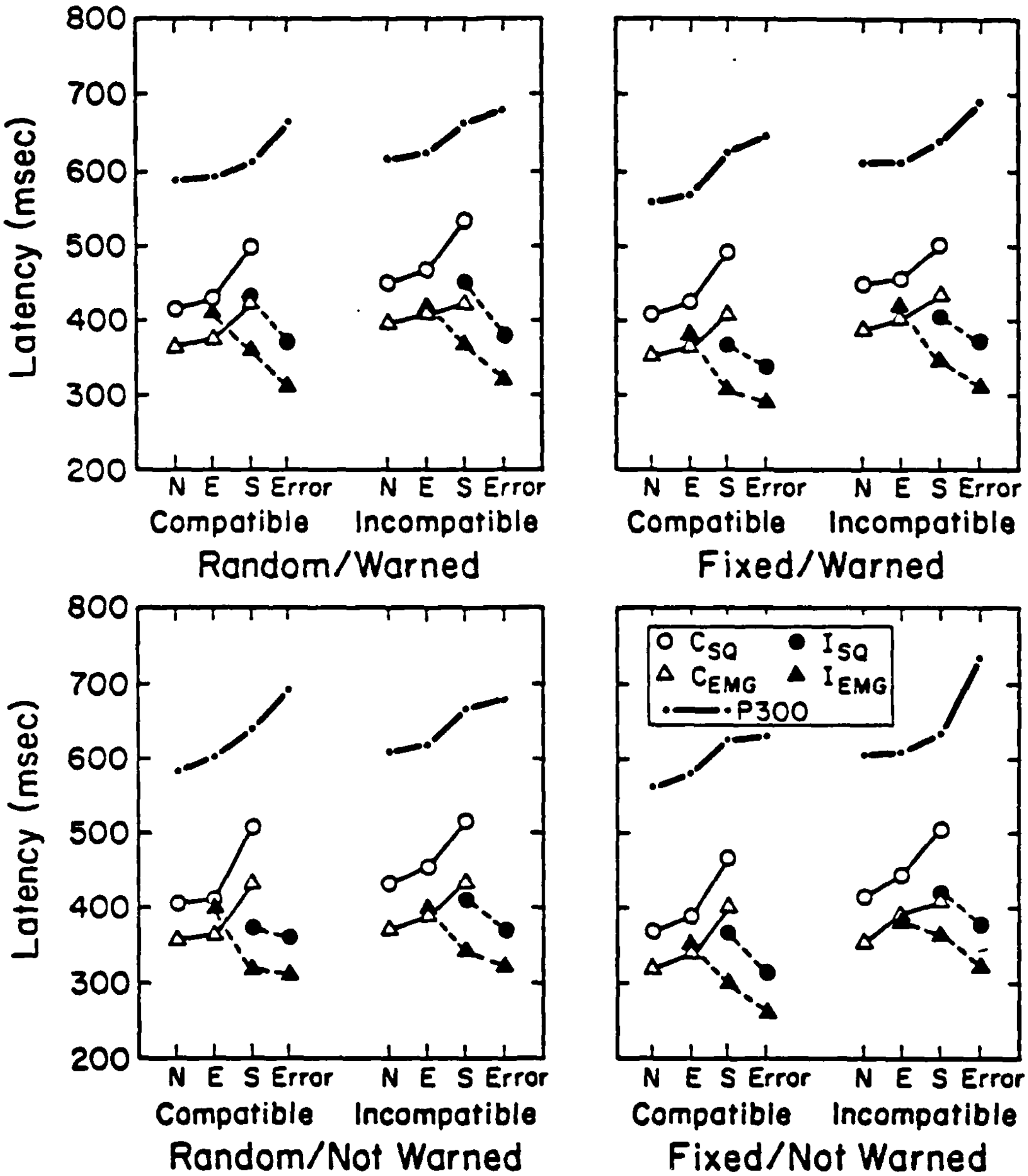


Figure 4. Values (in milliseconds) for the five latency measures as a function of response category and the eight conditions of the experiment. (For N, E, and S categories, the latency data are based on 11 subjects. For the Error category, the data are based on 4 subjects [see text]. P300 = latency of the P300; Csq = latency of onset of the correct squeeze response; Cemg = latency of onset of the correct electromyogram [EMG] response; Isq = latency of onset of the incorrect squeeze response; Iemg = latency of onset of the incorrect EMG response.)

the onset latency of the incorrect squeeze (396 ms) was shorter than that of the correct squeeze (501 ms), $F(1, 10) = 79.57$, $p < .001$. This result indicates that even though both squeeze responses are executed, they are not executed simultaneously—the incorrect response occurs first.

Together, these data suggest the following picture: (a) Both response channels may be activated on the same trial; (b) if this activation reaches the level of a squeeze, the two response channels inhibit each other (response competition); (c) response activation is not an all-or-none phenomenon—rather, several levels of activation are possible; (d) the activation of the correct response to the threshold for squeeze emission may occur after the emission of an incorrect squeeze, but the converse is not true.

The latency of the P300 component of the ERP also increases monotonically from N, to E, to S, to Error categories. The results of the relevant ANOVAs are shown in Table 1.[5] Because we interpret the latency of the P300 as a measure of the duration of evaluation processes, we infer that there is an association between the duration of evaluative processes and the likelihood of incorrect activity (at least at the squeeze level). Although this is only a correlational finding, it may suggest that a slowing of the stimulus evaluation process enhances the probability of the appearance of incorrect motor activity. We shall return to this point later.

b. *Noise/compatibility effect.* Inspection of the distribution of trials according to response category (see Figures 3 and 5) reveals that more trials were classified as N and fewer as S and Error when the noise was compatible. This was confirmed by an ANOVA on transformed (arcsine) percentage values for the E, S, and Error categories, which gave a significant main effect

[5] Tukey HSD tests revealed that the differences between E and S were significant. The N and Error categories were not statistically distinguishable from the E and S categories, respectively.

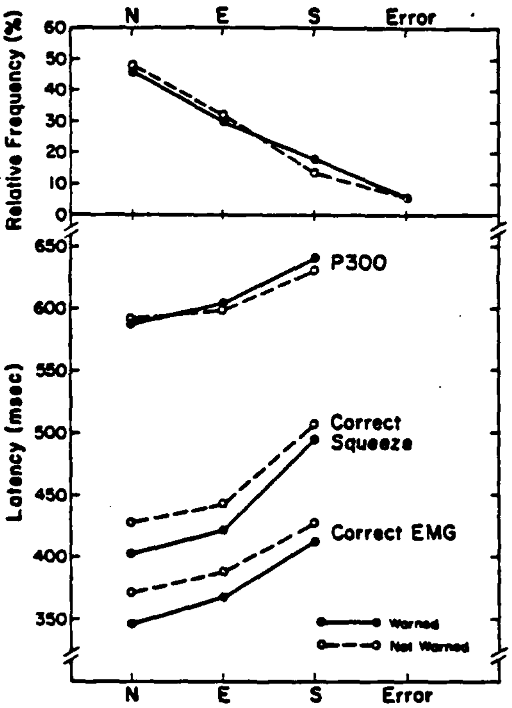


Figure 6. Latency of onset in milliseconds of correct electromyogram (EMG) and squeeze activity and of P300 as a function of response category for warned and not warned trials. (The relative frequencies of each response category for warned and not warned trials are shown in the upper panel.)

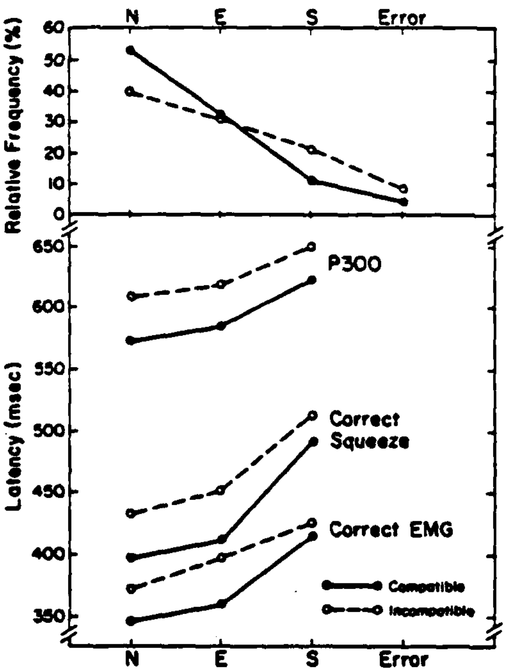


Figure 5. Latency of onset in milliseconds of correct electromyogram (EMG) and squeeze activity and of P300 as a function of response category for compatible and incompatible arrays. (The relative frequencies of each response category for compatible and incompatible arrays are shown in the upper panel.)

Table 1
Results of Analyses of Variance on Latency Measures

Main effect & response side	EMG onset latency		Squeeze onset latency		P300 latency	
	df	*F*	*df*	*F*	*df*	*F*
Noise[a]					1, 10	26.44**
Incorrect	1, 10	8.87*	1, 10	17.94**		
Correct	1, 10	17.13**	1, 10	67.67**		
Warning[a]					1, 10	1.00
Incorrect	1, 10	5.78*	1, 10	9.32**		
Correct	1, 10	8.81*	1, 10	16.44**		
Blocking[a]					1, 10	12.19**
Incorrect	1, 10	4.75	1, 10	2.90		
Correct	1, 10	6.32*	1, 10	4.98		
Response category[a]					2, 20	17.13**
Incorrect	1, 10	26.60**				
Correct	2, 20	51.80**	2, 20	109.24**		
Response category[b]					3, 9	10.61**
Incorrect	2, 6	9.10*	1, 3	23.47*		
Correct						

Note. EMG = electromyogram.
[a] Analysis based on 11 subjects.
[b] Analysis based on 4 subjects.
* $p < .05$. ** $p < .01$.

of noise, $F(1, 10) = 22.13$, $p < .001$, and a significant Noise × Response Category interaction, $F(2, 20) = 8.52$, $p < .01$. Note that these data are consistent with the "traditional" error rate analysis described earlier. However, they provide the important additional information that trials with both squeeze responses were more common when the array was incompatible.

These results confirm the previous findings of Eriksen et al. (in press) and are consistent with the continuous flow model. Evidence for the incorrect response is present in the incompatible array, and this evidence appears to lead to the activation of the incorrect response even though a correct response may be given ultimately.

As we noted above, there were more S trials and fewer N trials when the array contained incompatible noise. Furthermore, in the previous section we saw that response competition occurs on S trials. This is suggested by the delay in both the initiation and execution of the correct response on these trials. Thus, one way in which incompatible noise delays the average RT is by increasing the number of trials on which response competition occurs. If one computes RT without regard to response category (as we did in our initial RT analysis and as would be done in a traditional analysis), the cost of incompatible noise is 47 ms. The larger frequency of S trials for incompatible noise arrays (23%) than for compatible noise arrays (11%) accounts for an effect of 10 ms. This value is derived by weighting mean squeeze latency values for N, E, and S categories by the proportion of trials that were classified in each category. This leaves a 37-ms effect of noise/compatibility that is not yet explained.

Now, even when the level of incorrect response activation is controlled (that is, response category is a factor in the ANOVA), the interval between EMG and squeeze onset in the correct channel is still longer for incompatible noise arrays (67 ms) than for compatible noise arrays (59 ms), $F(1, 10) = 5.31$, $p < .05$. That is, *within* N, E, and S categories the interval between correct EMG and squeeze onset is, on the average, 8 ms longer for incompatible noise arrays. If this value is recomputed on the basis of appropriately weighted means (see above), then the value is 12 ms. Thus, we find an effect of noise/compatibility on the temporal aspects of correct response execution, even when the presence of incorrect activity is controlled. If it is assumed that a prolongation of the interval between EMG and squeeze onset is a sign of response competition, then response competition must have an effect on correct response execution that is not associated with the peripheral activation of the incorrect response channel (i.e., muscle and squeeze activity). This implies that response competition can occur when the activation of the incorrect response channel is below the threshold required for EMG or squeeze activity. Thus, of the 47-ms weighted mean effect of noise/compatibility on correct squeeze response latency, 10 ms can be attributed to a form of response competition that is associated with the emission of an incorrect EMG and squeeze response, and a further 12 ms to a form of response competition that is associated with subthreshold incorrect response activation. This leaves 25 ms to be explained.

The previous analyses indicated that the interval between EMG and squeeze onsets is affected by noise/compatibility. However, the *onset latency* of the correct EMG activity is also affected by noise/compatibility. In fact, the EMG onset latency is 28 ms longer for incompatible noise arrays than for compatible noise arrays even when response category is considered as a factor in the ANOVA. (See Table 1 for the results of ANOVAs and Figures 4 and 5 for the means.) Can this effect also be explained in terms of response competition? To answer this question, we need to examine the P300 data to determine whether noise compatibility affects stimulus evaluation. These data reveal that indeed, stimulus evaluation is longer for incompatible arrays, because the latency of the P300 is delayed. (See Table 1 for results of the relevant ANOVAs.) In fact, the delay in P300 associated with incompatible noise is 32 ms for an unweighted means analysis while the corresponding weighted mean value is 27 ms.[6] The latter value is very close to the 25-ms effect of noise/compatibility that remained after the effects of response competition had been removed.

This series of analyses reveals that the prolongation in the overt response latency for incompatible noise trials (47 ms) is due both to a slowing down of the evaluation process (27 ms) and to an increase in response competition (22 ms). The discrepancy of 2 ms is within the limits of rounding errors. A continuous flow model accounts for this dual effect in terms of the same cause: Incompatible noise produces conflict in stimulus evaluation, which slows the evaluation process *and* activates both response channels, which in turn results in response competition.

This is not the whole picture, however. Subjects also make incorrect responses and exhibit activity on the incorrect side on *compatible* trials, when there is nothing in the stimulus array to activate the incorrect side. This observation suggests the operation of another response-driving process that is independent of the stimulus. This is the process we have labeled *aspecific priming*.

c. *Warning effect.* We expected the process of aspecific priming to be more evident under warned conditions, because of the hypothesized increase in indiscriminant response activation resulting from the warning tone. Indeed, there was a tendency for fewer trials to be classified as N, and more as S, when the warning tone was presented, although the Warning × Category interaction was not significant.

The presence of an uninformative warning tone results in faster motor responses (as shown by EMG and squeeze onset latencies), both for the correct and the incorrect side. However, the latency of P300 is not affected by the warning manipulation (see Table 1 for the results of the corresponding ANOVAs). Furthermore, the interval between the onset of correct EMG activity and the peak of the P300 is longer in the warned condition, $F(1, 10) = 10.22$, $p < .01$. Together, these findings indicate that the warning facilitates motor responses without influencing the speed of evaluation processes. Recall that the presence of the warning tone also affects the number of trials with incorrect squeeze activity (although not significantly). Thus, the presence of the warning tone induces the subjects to respond faster but at a slightly higher error rate.[7] This effect of warning may

---

[6] A similar analysis of P300 latency for the count task, when no overt motor response was required, also revealed a significant main effect of noise, $F(1, 11) = 11.90$, $p < .01$. P300 latency was 16 ms longer for incompatible arrays.

[7] We have argued that the presence of a warning tone does not affect the evaluation process. Rather it leads subjects to become less conservative—they respond faster and make more errors. One apparently troubling aspect of the data is the lack of a significant effect of warning on error rate. Analysis of speed-accuracy trade-off functions such as those presented in Figure 9, Panel c, indicates that a 20-ms decrease in RT (the mean effect of warning) should be associated with an increase in error rate of approximately 3%. This was, in fact, the increase in error rate

be attributed to a greater aspecific priming or to lower response criteria.

Note that the warned condition was characterized by the presence of a negative-going potential (CNV) in the interval between the warning tone and the stimulus array (see Figure 1). Several investigators have related similar scalp negativities to motor preparation (see Deecke, Bashore, Brunia, Grunewald-Zuberbier, Grunewald, & Kristeva, 1984, for a review). Furthermore, some researchers (e.g., Gaillard, 1977; Kok, 1978; Rohrbaugh & Gaillard, 1983; Rohrbaugh, Syndulko, & Lindsley, 1976) have argued that later aspects of the CNV are related to motor preparation. In this sense, then, the late CNV may be a manifestation of aspecific priming.

d. *Effect of blocking.* When the level of noise was fixed rather than random within a block of trials, onset latencies of both EMG and squeeze responses on the correct side and of P300 were significantly shorter (by 17 ms, 15 ms, and 14 ms, respectively). (See Table 1 for the results of ANOVAs and Figure 4 for the means.) These data suggest that stimulus evaluation processes are faster when the level of noise is fixed. For *both* noise/compatibility conditions, it is apparently easier for subjects to perform the task when they know in advance what kind of noise will be presented.

However, there is more to the blocking manipulation than a simple main effect on stimulus evaluation. When we consider the distribution of trials across the different response categories, we find that the effect of fixing the level of noise was different for the different noise/compatibility conditions, $F(2, 20) = 3.84$, $p < .05$. Subsequent analyses revealed that for the fixed compatible noise condition, fewer trials were classified as N and more as S than for the random compatible condition. On the other hand, for incompatible conditions, fixing the level of noise did *not* lead to a larger frequency of S trials. These data confirm our previous conclusion that subjects adopt a less conservative strategy when they are confronted with the fixed compatible condition. Thus, the effect of fixing the level of noise is to speed evaluation processes for both noise/compatibility conditions and to change response strategy when the noise is compatible.

Speed–Accuracy Trade-Off Functions

Up to this point, we have considered the effects of the manipulations on the average duration of the stimulus evaluation process. In this section, we examine speed–accuracy trade-off functions for the various conditions of the experiment. We will show (a) that the noise/compatibility manipulation affects the time course of evidence accumulation, (b) that the warning does not affect the evaluation process, and (c) that fast responses are mainly controlled by the letters flanking the target.

The speed–accuracy functions are obtained by plotting response accuracy as a function of response latency. They are intended to provide a representation of the manner in which stimulus evaluation processes proceed over time that is uncontaminated by response bias factors (e.g., Pachella, 1974). However, as we have noted, this interpretation is predicated on the assumption that the speed of stimulus evaluation processes is constant for a given condition. This assumption may not be valid (see Meyer & Irwin, 1982). Thus, in the analysis reviewed here, we compute *separate* speed–accuracy trade-off functions for trials with *different* durations of stimulus evaluation. We do this by using P300 latency as a parameter. That is, trials are first sorted according to the latency of the P300. Then, for each P300 latency bin, we plot response accuracy against response latency.

We obtained our functions in the following way. First, for each of the 12 subjects, and for each of the eight conditions, the latency of the onset of first EMG response, the correctness of that response, and the P300 latency for each trial were tabulated. Second, we defined each trial as a fast or slow P300 trial if P300 latency on that trial was longer or shorter than the median P300 latency for that subject and condition. We also classified the trials into four quartiles on the basis of EMG onset latency

---

when computed using the definition of an error described in this section. Because error rate was computed on a relatively small number of trials, our estimate was not sufficiently reliable to permit a 3% difference to be significant in an ANOVA. If more reliable estimates were obtained, we could determine whether the difference is "real" or whether, in fact, the subjects are able to respond faster, but at the same accuracy level, when a warning is present. If this is the case, then the effect of the warning might be to change the slope of the response activation function, that is, to speed motor processes.

for that subject and condition. In this way, trials were sorted into eight groups on the basis of P300 latency (fast/slow) and quartile. For each of these groups, accuracy was computed by dividing the number of correct trials by the total number of trials for that group.

We should note that we use EMG onset latency, rather than squeeze latency, as our measure of response speed in these analyses because the activity in the EMG channel occurs first and is a more sensitive sign of response activation.

Figures 7 and 8 display the speed–accuracy functions for each condition of the experiment for fast and slow P300 latency trials separately. The standard errors for each mean are also shown. Figure 9 displays a summary of the speed–accuracy trade-off functions for different P300 latencies and for the two noise and the two warning conditions. Figures 7, 8, and 9 (Panel a) reveal two important points. First, accuracy increases as EMG latency increases, regardless of the latency of the P300 (i.e., the duration of stimulus evaluation); that is, the slower the response, the more likely it is to be correct, $F(3, 33) = 101.55$, $p < .01$. Second, accuracy is lower for all response speeds when P300 latency is long, $F(1, 11) = 39.99$, $p < .01$. Furthermore, *similar* levels of accuracy are achieved either by the conjunction of a slower EMG response and a slow P300 or by a faster EMG response and a fast P300. In other words, P300 latency, and by implication stimulus evaluation time, appears to determine the relative position of the speed–accuracy trade-off function. Together, these data suggest that the accuracy of a response depends on its timing relative to the evaluation process. When

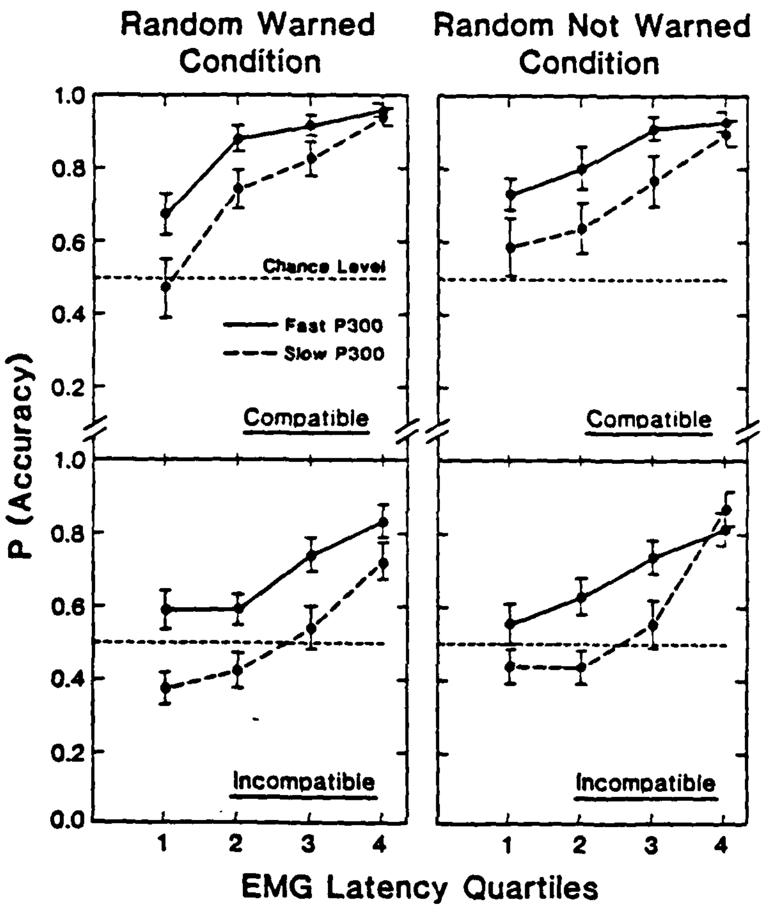


Figure 7. Speed–accuracy trade-off curves as a function of P300 latency for compatible and incompatible noise trials, when noise was randomized within trial blocks, for the two warning conditions separately.

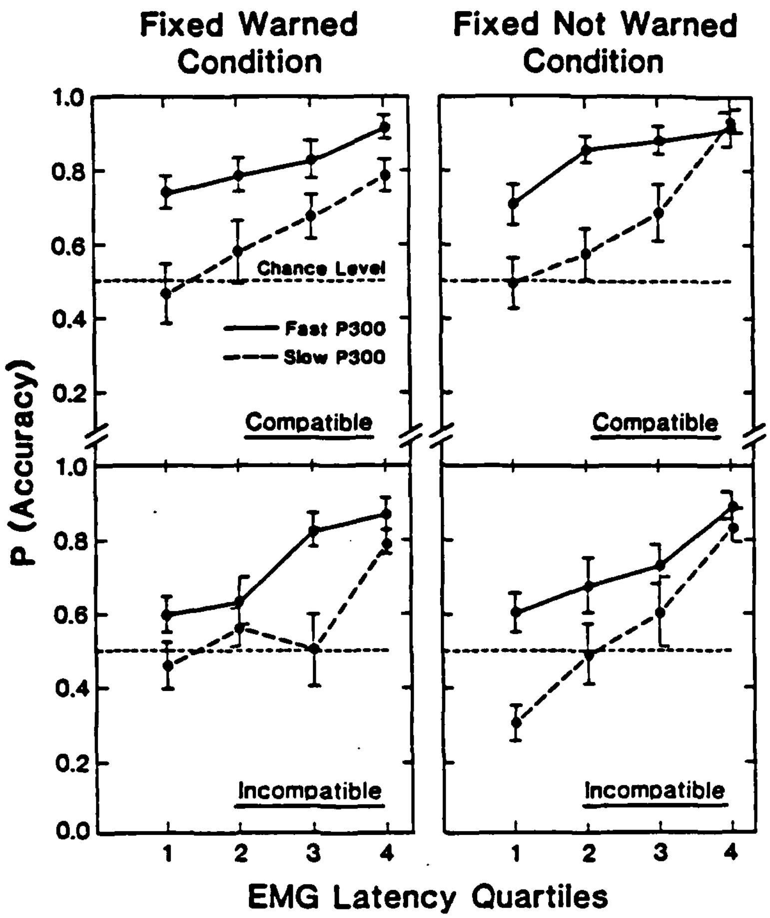


Figure 8. Speed-accuracy trade-off curves as a function of P300 latency for compatible and incompatible noise trials, when noise was fixed within trial blocks, for the two warning conditions separately.

evaluation proceeds quickly, a high level of accuracy is achieved even when responses are fast; conversely, when evaluation proceeds slowly, a high level of accuracy is achieved only when RTs are long.[8] These data illustrate how measures of the P300 can be used to overcome the difficulties raised by the assumption that the duration of the evaluation process is constant on every trial.

Figures 7, 8, and 9 (Panel b) show that speed-accuracy functions for compatible and incompatible noise arrays are different. For each quartile, accuracy is lower for the incompatible arrays, $F(1, 11) = 56.98$, $p < .01$. This confirms that the evaluation process is slower, or at least different, for these arrays.

Figures 7, 8, and 9 (Panel c) show that the functions for warned and unwarned trials are

[8] We have interpreted the interaction among P300, EMG onset latency, and accuracy in terms of an effect on accuracy of the relative time during the evaluation process at which a response is emitted. When subjects respond quickly and evaluation is slow, they are likely to make errors. Note that we are inferring that accuracy is a function of P300 and EMG onset latency, although our data are correlational in nature. An alternative interpretation is that the P300 is delayed when the subject makes an error. In fact, we have evidence from another experiment (Gratton, Dupree, Coles, & Donchin, 1985) that P300 can be actively delayed by a process of error recognition. The conditions under which this result was obtained involved a choice RT task under speed instructions. The instructions led the subjects to respond very quickly and at a low-accuracy level. As we have outlined elsewhere (Coles, Gratton, & Donchin, 1984), these two interpretations can be distinguished on the basis of the accuracy level for trials on which responses are fast and P300 latency is long. In particular, accuracy should be close to zero for these kinds of trials if the error recognition interpretation is valid. Such a finding

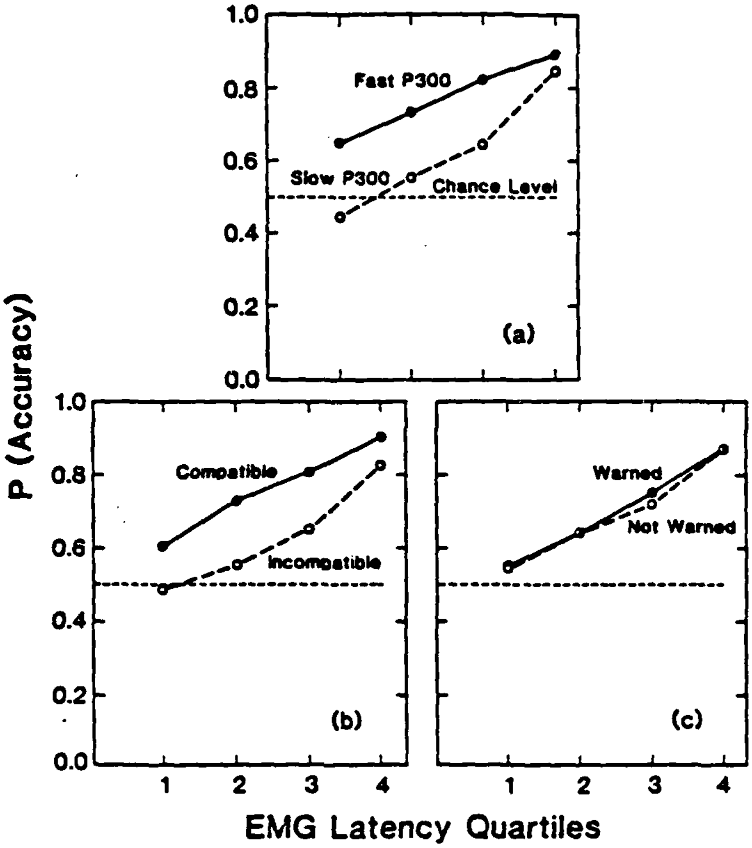


Figure 9. Speed–accuracy trade-off curves as a function of P300 latency (Panel a), noise (Panel b), and warning (Panel c).

essentially identical. For the main effect of warning, $F(1, 11) = 0.19$, $p = 0.67$; for the Warning × Quartile interaction, $F(3, 33) = 0.56$, $p = 0.65$. These observations confirm the conclusion we drew earlier that the presence of a warning stimulus does not affect the evaluation process. Rather, the difference between these two conditions in mean response latencies and error rates reflects a difference in the average point on the speed–accuracy trade-off function at which the subject is operating. As we argued above, the greater aspecific priming (or a lower criterion) on warned trials leads to a less conservative response (i.e., responses are released on the basis of less information).

A further interesting aspect of the functions shown in Figures 7 and 8 concerns the accuracy for fast EMG responses and slow P300s. In the compatible noise conditions, accuracy is approximately 50%. We infer from this that when subjects respond quickly on trials where the duration of stimulus evaluation is long (P300 latency is long), they are essentially guessing. However, on incompatible trials, the combination of fast EMG responses (the first quartile) and slow P300s (across warning and blocking conditions) is associated with an accuracy value that is *below* chance, $t(11) = 3.83$, $p < .01$.

One explanation for this excessive error rate is that early in the evaluation of an incompatible noise array, there is more evidence for the incorrect response. It should be recalled that

was obtained in the Gratton et al. (1985) study. However, in the present experiment, the accuracy level for fast response/slow P300 trials is close to 50%. We do find that accuracy level falls below 50% in the incompatible noise condition, but this finding is most readily explained in terms of the potency of the flanking noise in driving the incorrect response.

an incompatible array contains one letter associated with the correct response and four letters associated with the incorrect response. Thus, when the subject responds quickly and evaluation is proceeding slowly, the evidence available at the time of response favors the incorrect response. Note that this excessive error rate is not seen in the data for compatible arrays. Our data suggest, then, that early in the evaluation process, the subject performs an analysis of the features of *all* the letters in the array without selecting the information provided by the target letter in the central location. We refer to this process as *feature*, or *letter*, analysis. Selection for the features of the center letter (*location* analysis) appears to occur later. These two aspects of stimulus evaluation, feature, or letter, analysis and location analysis, can both activate the response channels directly. The two processes may occur in sequence or in parallel. However, in the latter case, feature analysis should be faster than location analysis. Thus, fast responses, based mainly on the feature analysis, are likely to be incorrect for an incompatible noise trial, but correct for a compatible noise trial. The process of aspecific priming, discussed earlier, also controls activation of response channels. If one or other of the responses is heavily primed (for example, because of guessing), then that response may be released without being influenced by either feature or location analyses.

Conclusions

The results of this experiment clearly indicate that both the correct and incorrect response channels can be activated concurrently. The activation of the response channels occurs in a graded fashion, so that partial response activation of one response channel may accompany complete response activation of the other channel. When both response channels are activated, *response competition* occurs, and the temporal characteristics of correct response execution are affected. Response activation itself appears to be controlled by two processes: stimulus evaluation *and* aspecific priming. The influence of the first process increases over time after array presentation, because slower responses are more accurate. Furthermore, when the array contains information calling for the incorrect response, this response is more likely to be activated. In fact, when subjects respond early, the incorrect information dominates to such an extent that error rates are greater than chance. The second process, aspecific priming, results in an activation of response channels that is independent of the stimulus. This is evident from the fact that activation of the incorrect response is observed when there is no corresponding information in the stimulus array.

This picture is consistent with the continuous flow model proposed by Eriksen and Schultz (1979). Although it was not the purpose of this study to address the question of the viability of serial stage models, our data are not easily accommodated by a strictly serial stage model (e.g., Sternberg, 1969). For example, to account for our observation of concurrent activation of both response channels, a serial stage model would have to assume that a decision stage emits an output to each of the response channels that is proportional to the evidence accumulated at the moment of the decision.[9] However, this would be inconsistent with the observed temporal relations between the correct and the incorrect responses when both occur on the same trial (the S category). In fact, the incorrect response occurs *before* the correct response on S trials. To explain this finding, one would have to assume several decision stages. Thus, although it is possible to increase the complexity of a serial stage model to account for our data, it is clear that the continuous flow model (and other parallel models) provides a more parsimonious explanation.

The analysis of the EMG and subthreshold squeeze data have important implications for the concept of response competition. First, we find that when incorrect squeeze activity is present, initiation of correct activity is delayed. Second, we find that the temporal characteristics of correct response execution are affected by the degree to which incorrect activity is present. When an incorrect squeeze response is produced (the S category), the interval between correct EMG onset and correct squeeze onset is increased. Finally, when there is evidence in the array for both responses (incompatible condition), this interval is also prolonged, although there may be no peripheral manifestation of activation of the incorrect response (as in the N category). Together, these

[9] This model was suggested by an anonymous reviewer.

findings are most readily explained in terms of the operation of a response competition mechanism. Furthermore, the fact that the temporal characteristics of response execution can be modified and that responses can be initiated without being executed, suggests that response execution is best conceived of as a continuous process. This view contrasts with that of McClelland (1979), for whom response execution is the only discrete process in the human information processing system.

The manipulations we used in our experiment have different effects on the information processing system. One effect of introducing incompatible *noise* to the stimulus array is to increase the number of trials on which incorrect activity occurs. In general, the presence of incorrect activity is associated with an increase in the time taken to execute a correct response. Thus, the mean RT difference between compatible and incompatible noise is due, at least in part, to response competition. However, the effect of incompatible noise is also to slow down the evaluation process, as indexed by P300 latency. Thus, the noise/compatibility effect on mean RT appears to be due both to an effect on the incidence of response competition *and* to an effect on the stimulus evaluation process.

In contrast to the noise manipulation, the *warning* conditions provided a clear dissociation between P300 latency and the latency of motor response measures (correct and incorrect squeeze and EMG onset latencies). The latter were in fact shortened by the warning, whereas the presence of a warning had no effect on P300 latency. This result suggests that the warning did not influence stimulus evaluation processes, but it was clearly effective in increasing the aspecific priming of the two response channels. These data contrast in an interesting manner with the results of Duncan-Johnson and Donchin (1982). These investigators presented imperative stimuli that either matched or failed to match an antecedent warning stimulus. When the stimuli mismatched, the P300 latency to the imperative stimulus increased. Thus, there are conditions in which the information carried by a warning stimulus can affect the duration of stimulus evaluation processes for a subsequent event, suggesting the operation of *perceptual* priming. However, in the present study, the warning stimulus (a tone) did not match the imperative stimuli (letters). Under these circumstances, there is apparently no opportunity for an effect of perceptual priming on the evaluation process.

By *fixing* the level of noise within a block of trials, correct responses were speeded and P300 latency was shortened. This indicates that fixing the level of noise facilitates the stimulus evaluation process. However, this manipulation also leads to a modification in the response criterion or to a greater aspecific priming in the compatible noise condition, so that subjects respond faster but less accurately.

Insights into the nature of the stimulus evaluation process were provided by the speed–accuracy trade-off functions with stimulus evaluation time controlled. These functions suggest that in our experiment the stimulus evaluation process consists of at least two subprocesses, feature or letter analysis and location analysis. Note that our conception of the process of stimulus evaluation is similar to that discussed by Treisman and her colleagues (Treisman & Gelade, 1980; Treisman, Sykes, & Gelade, 1977). They argue that an early, parallel process of feature analysis precedes the detection of the feature location. Our data suggest that the output of the feature analysis should be available before that of the location analysis, although these two subprocesses may occur in sequence or in parallel. Both feature (letter) and location analyses appear to activate the response channels directly. In fact, the speed–accuracy functions for incompatible arrays reveal that early responses are driven more by the lateral letters than by the central target letter. This short cut of the information processing flow is inconsistent with the assumptions of a strictly serial *and* a strictly cascade model (e.g., McClelland, 1979). Both these models assume that the flow of information proceeds through an ordered sequence of processing elements. On the other hand, these kinds of short cuts are not inconsistent with the assumptions of the continuous flow model (Eriksen & Schultz, 1979).

An interesting integration of serial and parallel models has been proposed recently by Miller (1982, 1983). His model can be described as a hybrid *parallel–discrete* model. He suggests that information is *not* transferred continuously between processing elements. Rather, the transfer occurs only when an element has completely processed a "grain" of information. Thus, information represented by a grain is transferred discretely. However, when there is more than one grain, different processing elements can be engaged in parallel. Note that, when all the relevant information is contained in one grain, his model is formally equivalent to a serial model. When the relevant information can be partitioned into an infinite number of grains, his model is formally equivalent to a cascade model. In terms of Miller's model, our data suggest that the information is partitioned into more than one grain, because responses are activated on the basis of partial information about the stimulus array. Furthermore, at the level of feature (or letter) analysis, several grains must be handled in parallel. On the other hand, at the level of location analysis information may be transferred in only one grain.

In summary, the results of our experiment are consistent with the continuous flow model (Eriksen & Schultz, 1979), although they are not inconsistent with other parallel models, such as those proposed by Miller (1982) or Grice and his colleagues (Grice et al., 1977, 1982). We have provided evidence for two relatively independent sources of response activation: an aspecific, stimulus-independent process, and a specific, stimulus-dependent process. As evidence accumulates in the stimulus evaluation system, specific activation of the associated response systems occurs. Activation of the incorrect channel is determined both by the amount of aspecific priming and by the evaluation process, when there is evidence in the stimulus for the incorrect response. Activation of the incorrect response channel can interfere with correct response execution through a response competition process.

References

Brookhuis, M. A., Mulder, G., Mulder, L. J. M., & Gloerich, A. B. M. (1983). The P3 complex as an index of information processing: The effects of response probability. *Biological Psychology, 17*, 277–296.

Coles, M. G. H., Gratton, G., & Donchin, E. (1984, September). *Flies in the ointment.* Paper presented at the IIIrd International Conference on Cognitive Neuroscience, Bristol, England.

Deecke, L., Bashore, T., Brunia, C. H. M., Grunewald-Zuberbier, E., Grunewald, G., & Kristeva, R. (1984). Movement-associated potentials and motor control. In R. Karrer, J. Cohen, & P. Tueting (Eds.), *Brain and information: Event-related potentials* (pp. 398–428). New York: New York Academy of Sciences.

Donchin, E. (1979). Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), *Evoked potentials and behavior* (pp. 13–75). New York: Plenum Press.

Donchin, E. (1981). Surprise! . . . Surprise? *Psychophysiology, 18*, 493–513.

Donders, F. C. (1969). On the speed of mental processes. In W. G. Koster (Ed. and trans.), *Attention and performance II* (pp. 412–431). Amsterdam: North-Holland.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology, 14*, 456–467.

Duncan-Johnson, C. C., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. *Biological Psychology, 14*, 1–52.

Duncan-Johnson, C. C., & Kopell, B. S. (1981). The Stroop effect: Brain potentials localize the source of the interference. *Science, 214*, 938–940.

Eriksen, C. W., Coles, M. G. H., Morris, L. R., & O'Hara, W. P. (in press). An electromyographic examination of response competition. *Bulletin of the Psychonomic Society.*

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*, 143–149.

Eriksen, C. W., & Eriksen, B. A. (1979). Target redundancy in visual search: Do repetitions of the target within the display impair processing? *Perception & Psychophysics, 26*, 195–205.

Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics, 25*, 249–263.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (in press). The definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in psychophysiology: Vol. 2.* Guilford, CT: JAI Press.

Ford, J. M., Roth, W. T., Mohs, R. C., Hopkins, W. F., & Kopell, B. S. (1979). Event-related potentials recorded from young and old adults during a memory retrieval task. *Electroencephalography and Clinical Neurophysiology, 47*, 450–459.

Gaillard, A. W. K. (1977). The late CNV wave: Preparation versus expectancy. *Psychophysiology, 14*, 563–568.

Ganz, L. (1975). Temporal factors in visual perception. In E. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 5, pp. 169–231). New York: Academic Press.

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). Filtering for scalp distribution: A new approach (vector filter). *Psychophysiology, 20*, 443–444. (Abstract)

Gratton, G., Dupree, D., Coles, M. G. H., & Donchin, E. (1985). *P300 and error processing.* Manuscript in preparation.

Gratton, G., Kramer, A. F., & Coles, M. G. H. (1984). A comparative study of measures of the latency of event-related brain components. *Psychophysiology, 21*, 578–579. (Abstract)

Grice, G. R., Nullmeyer, R., & Spiker, V. A. (1977). Application of variable criterion theory to choice reaction time. *Perception & Psychophysics, 22*, 431–449.

Grice, G. R., Nullmeyer, R., & Spiker, V. A. (1982). Human reaction times: Toward a general theory. *Journal of Experimental Psychology: General, 111,* 135–153.

Grossberg, S. (1982). *Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control.* Boston: Reidel Press.

Hoffman, J. E., Houck, M. R., MacMillan, F. W., Simons, R. F., & Oatman, L. C. (1985). Event-related potentials elicited by automatic targets: A dual-task analysis. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 50–61.

Jasper, H. H. (1958). The ten–twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10,* 371–375.

Karis, D., Fabiani, M., & Donchin, E. (1984). P300 and memory: Individual differences in the von Restorff effect. *Cognitive Psychology, 16,* 177–216.

Kok, A. (1978). The effect of warning stimulus novelty on the P300 and components of the contingent negative variation. *Biological Psychology, 6,* 219–233.

Kutas, M., & Donchin, E. (1977). The effect of handedness, of responding hand, and of response force on the contralateral dominance of the readiness potential. In J. Desmedt (Ed.), *Attention, voluntary contraction and event-related cerebral potentials* (pp. 189–210). Basel: Karger.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science, 197,* 792–795.

Lappin, J. S., & Disch, K. (1972a). The latency operating characteristics: I. Effects of stimulus probability on choice reaction time. *Journal of Experimental Psychology, 92,* 419–427.

Lappin, J. S., & Disch, K. (1972b). The latency operating characteristics: II. Effects of visual stimulus intensity on choice reaction time. *Journal of Experimental Psychology, 93,* 367–372.

Lippold, O. C. J. (1967). Electromyography. In P. H. Venables & I. Martin (Eds.), *A manual of psychophysiological methods* (pp. 245–297). Amsterdam: North-Holland.

Magliero, A., Bashore, T. R., Coles, M. G. H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology, 21,* 171–186.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science, 211,* 77–80.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–330.

Meyer, D. E., & Irwin, D. E. (1982). *On the time course of rapid information processing* (Tech. Rep. No. 43). Ann Arbor: University of Michigan, Cognitive Science Program.

Miller, J. (1982). Discrete versus continuous stage models of human information processing: In search of partial output. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 273–296.

Miller, J. (1983). Can response preparation begin before stimulus recognition finishes? *Journal of Experimental Psychology: Human Perception and Performance, 9,* 161–192.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41–82). Hillsdale, NJ: Erlbaum.

Posner, M. I. (1978). *Chronometric explorations of mind.* Hillsdale, NJ: Erlbaum.

Ragot, R. (1984). Perceptual and motor space representation: An event-related potential study. *Psychophysiology, 21,* 159–170.

Rohrbaugh, J. W., & Gaillard, A. W. K. (1983). Sensory and motor aspects of the contingent negative variation. In A. W. K. Gaillard & W. Ritter (Eds.), *Tutorials in event-related potential research: Endogenous components* (pp. 269–310). Amsterdam: North-Holland.

Rohrbaugh, J. W., Syndulko, K., & Lindsley, D. B. (1976). Brain components of the contingent negative variation in humans. *Science, 191,* 1055–1057.

Sherrington, C. S. (1906). *Integrative action of the nervous system.* New York: Scribner.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II* (pp. 276–315). Amsterdam: North-Holland.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Treisman, A., Sykes, M., & Gelade, G. (1977). Selective attention and stimulus integration. In S. Dornic (Ed.), *Attention and performance VI* (pp. 333–361). Hillsdale, NJ: Erlbaum.

Tukey, J. W. (1953). *The problem of multiple comparisons.* Princeton, NJ: Princeton University Press.

Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review, 80,* 1–52.

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: An electrical sign of sensorimotor association and expectancy in the human brain. *Nature, 203,* 380–384.

Yeh, Y.-Y., & Eriksen, C. W. (1984). Name codes and features in the discrimination of letter forms. *Perception & Psychophysics, 36,* 225–233.

Received December 3, 1984
Revision received April 16, 1985 ■

Appendix B2

In: D. Papakostopoulos, S. Butler, & I. Martin (Eds.), Clinical and Experimental Neuropsychophysiology. Beckenham, England: Croom Helm, Ltd., 1985.

PSYCHOPHYSIOLOGY AND CONTEMPORARY MODELS OF HUMAN INFORMATION PROCESSING

Michael G.H. Coles & Gabriele Gratton, Cognitive Psychophysiology Laboratory, University of Illinois, Champaign, Illinois 61820, USA.

Cognitive psychologists are interested in how a particular input (stimulus) to the human information processing system is translated into a particular output (response). They propose that different processing structures perform different transformations on the input such that, ultimately, an output is produced. The number of structures, and their function, varies among theories; however, in general, they include perceptual, central, and response structures (e.g. Wickens, 1980).

Traditionally, it has been proposed that the processes associated with the different structures occur sequentially - that is, the process associated with one structure must be completed before another process begins. In recent years, these discrete models have been challenged by those who propose that information can be transmitted from one structure to another before the process performed by the first is completed. Thus, continuous models imply that several processes can occur simultaneously (or in parallel) and that a given process can operate on the partial information provided by another process. (See Miller, 1982, for a discussion of the difference between discrete and continuous models).

The measurements taken by cognitive psychologists (reaction time (RT), percent correct, etc.) are seriously limited in terms of their ability to test continuous theories, principally because they represent a single output measure which

is determined by many intervening processes, and their interactions. Thus, a particular experimental manipulation may affect not only the duration of a mental process - but also the rate of transmission between processing structures, and/or the criterion level at which the output is emitted. Changes in RT that result from a manipulation might be the result of any of these factors. Thus, cognitive psychologists are faced with a serious problem if they wish to use RT measures to make inferences about the nature of the effect of a particular experimental manipulation. It will be our thesis that psychophysiological measures might prove to be extremely useful in dealing with this problem, because they can provide information about the processes that intervene between input and behavioural output. In this sense, they can be considered as adjuncts to traditional measures of reaction time and accuracy.

In this chapter, we will provide an overview of cognitive models with special emphasis on contemporary continuous models. We will then indicate how psychophysiological measures may be used to assist in understanding cognitive processes.

MODELS OF HUMAN COGNITIVE FUNCTION

SERIAL MODELS

In this section, we present a brief review of models proposed by Donders and Sternberg. For a more complete treatment of this topic, see Pachella (1974).

Donders' model. Serial, or stage, models of human information processing derive from Donders' pioneering work in the 1860s. Donders (1868) distinguished among three classes of reaction time task: a,b and c. In the a-reaction, a single stimulus requires a single response: in the b-reaction, one of two stimuli are presented and the subject must produce one response to one stimulus and another response to the other: in the c-reaction, again one of two stimuli is presented but the subject must only respond to one of them.

Donders argued that the three types of reaction time tasks differed in terms of the number of stages or processes involved in their successful execution. Thus, the c-reaction time task adds a process of discrimination to the a-reaction time task, while the b-reaction time task adds an additional process of response selection. If RT is measured in each task, then by subtracting the RTs for various types of task it should be possible, according to Donders, to identify the time taken for discrimination and response selection processes respectively. Note that the subtractive method advocated by Donders presupposes that (a) the various stages of human information processing are arranged serially, (b) the duration of each stage is causally independent of the duration of the other stages, and (c) it is possible to use an experimental manipulation to add a stage (or process) without affecting other stages. The latter assumption has been referred to as the "postulate of pure insertion" (Ashby and Townsend, 1980). Given these assumptions, we can see that RT represents the sum of the duration of several component processes, and that it is possible to determine the duration of a stage "inserted" by a manipulation.

Sternberg's model. Sternberg (1969a) proposed a similar model ("stage model") to that advocated by Donders. Like Donders, Sternberg assumed that the human information processing system consists of a number of serially arranged stages and that RT represents the sum of the durations of each stage. However, the two theorists differ in the interpretation of the effect of an experimental manipulation. While Donders believed that a manipulation results in the insertion of a stage, Sternberg argues that it affects the duration of a particular stage, without affecting the duration of other stages. This is referred to as the postulate of "selective influence" (see Pieters, 1983).

As was mentioned above, Donders' model relied on the "subtraction method" to yield information about the duration of mental processes. Likewise, the Sternberg model is closely tied to a particular methodology - namely, the "additive factors method" based on the analysis of variance. This procedure involves the manipulation of two or more experimental factors and the observation of their effects on RTs. In particular, analysis of variance is used to determine the significance of the effects of these factors, both as main effects and in

interaction. Factors that affect RTs independently (that is, their interaction is not significant) are held to influence different stages of processing. On the other hand, when two factors have an interactive effect on RT, they are held to influence the same stage.

Sternberg's model is most closely identified with a memory task (Sternberg, 1969a, 1969b). In one instance of the task, the subject was given a set of items to remember, and then was required to indicate whether a probe item was ("yes" response), or was not ("no" response), a member of the memory set. Experimental manipulations included stimulus quality, memory set size, response type (yes or no), and response frequency. All four factors had significant main effects on RT. None of the two-way interactions tested by Sternberg was significant (the quality x frequency interaction was not tested). From these data, he argued for the existence of four discrete stages (encoding, comparison, decision and response organisation).

Pachella (1974) notes that "the additive factor method itself does not supply either the description of each stage, as given above, or the order in which the stages are effective. Those conjectures arise from a consideration of the nature of each factor and a logical argument concerning the dependency of each subsequent transformation on the processes preceding it". (p. 55)

Discussion of serial models. Serial models have a basic appeal because of their simplicity and elegance, and this is probably responsible in part for their dominance in the field for more than a century.

The postulates of pure insertion and selective influence, if supported, imply that the cognitive psychologist has the relatively simple task of using a variety of experimental manipulations to identify the number and function of the information processing stages. This task is accomplished using straightforward statistical procedures such as "t"-test or analysis of variance. Unfortunately, the real world is not so simple.

First, even a cursory knowledge of biological systems suggests that the nervous system does not function in a serial manner. "For," as Woodworth (1938) said, "there is nothing to prevent two cerebral processes from occurring simultaneously. Two responses, one perceptual and one motor, may take their start simultaneously from the same stimulus. If the motor response is not made to depend on the perception, it can start at once, as in the simple reaction. The brain is not a one track road" (p. 305).

Second, it has been argued (Pachella, 1974, p.57) that the additive factors method relies on the acceptance of the null hypothesis for the inference of "independence". That is, if the interaction between two experimental factors is not significant, then it is claimed that the two factors influence different stages. This is clearly a risky procedure, unless great care is taken to insure sufficient statistical power.

Third, it has been argued that the postulates of pure insertion and selective influence have no basis in experimental data (Pachella, 1974). These postulates assume that the information processing sequence and the activities of the different stages remain unaffected by an experimental manipulation that either adds, or influences, a particular stage. The postulates can be questioned by a single counter-demonstration (see Pachella, 1974). A related problem is that it is not always easy to see whether the effect of an experimental manipulation is to add a stage of processing (pure insertion) or to modify an existing process (selective influence).

Fourth, analysis of the relationship between speed and accuracy of responses reveals that the output of the processing system varies as a function of time. In fact, the quality of performance (accuracy) varies continuously with reaction time. This suggests that the outputs of the processes are available at any moment in time rather than only after a given delay (the duration of the process). Proponents of serial models counter this argument by only accepting data from a certain point on the speed-accuracy trade-off function - namely, the point at which error rates are low while reaction times are fast. It is clear that such a narrow view of "acceptable data" is problematic not only because of the difficulty inherent in determining whether the subject is operating at the required level but

also because it denies the relevance of all other points on the speed-accuracy trade-off function. (See Pachella, 1974, for a discussion of this issue).

PARALLEL MODELS

In response to the problems raised above in connection with serial models, several investigators have recently argued for a different approach to the description of the human information processing system. This approach is characterized by two fundamental propositions. First, several processes can occur at the same time; second, the output of the process associated with each structure can feed continuously into other structures. As a result, this new approach has been described as "parallel" and "continuous" to distinguish it from the "serial" and "discrete" approach discussed in the previous section.

In this section we shall review the proposals of three of the principal adherents of the new approach, McClelland, Grice and Eriksen. Rather than describe each of their views in detail, we shall focus on the special aspects of their proposals. We shall also consider the views of Miller who advances a theoretical framework that encompasses both serial and parallel approaches.

McClelland's cascade model. As its name implies, McClelland's model proposes that the human information processing system consists of a collection of processes arranged "in cascade" (McClelland, 1979). Although the processes are ordered, each process is continuously active. It operates on the basis of its input, and, at any particular time, provides an output that corresponds as closely as possible to an ideal transformation of the input. Since the transformation takes a finite time, each process introduces a delay into the system. An important assumption of the model is that any piece of information that enters the system must proceed in an orderly manner through all the processes. No side trips, back-tracking, or short-cuts are allowed. Furthermore, an important exception to the continuous activation of the processes is the response activation system. This system is assumed to operate in a discrete manner. When the input to this processing level exceeds a prescribed level, a response is emitted (see Grice, Nullmeyer and Spiker, 1982; and see below).

McClelland provides a mathematical formulation to describe the behaviour of each process. For present purposes, it is sufficient to note that the output of each process is assumed to be an exponential function of its input.

Grice's general theory. This theory was developed with the express intention of accounting for Donders' three types of reactions. Like McClelland, Grice believes that different aspects of the human information processing system can be active at the same time (Grice et al., 1982). Thus, when a stimulus is presented there is a gradual priming of responses as the strength of the representation of the stimulus increases. This is referred to as the development of associative strength. Also important for the theory is the notion of response channel. There is one response channel for each possible response. When the level of activation in a response channel exceeds some pre-specified criterion (responses bias), that response is emitted. Thus, the choice reaction time task is seen as a "race" among the response channels. Grice differs from McClelland in that he allows for some by-passing of processes. He proposes a mechanism - excitation - whereby response processes are activated independently of the nature of the stimulus. In other words, the mere presentation of a stimulus, any stimulus, is sufficient to activate responses without stimulus evaluation.

Eriksen's continuous flow model. Like Grice, Eriksen believes in the separation among response channels. Because the stimulus evaluation processes feed directly and continuously into the response channels, different responses can be simultaneously activated to the degree that the information contained in the stimulus calls for different responses (Eriksen & Schultz, 1979). At the perceptual end of the model, great stress is given to the inability of the human information processing system to completely filter out irrelevant information. Much of Eriksen's work is concerned with visual displays where a target stimulus is surrounded by noise stimuli. The noise may or may not contain cues that are associated with responses

that are incompatible with the required response. Eriksen finds that, when the noise is incompatible reactions times are longer than for compatible noise. This difference is attributed to the "involuntary" processing of the extraneous noise cues and to a process of response competition. When the incorrect response is activated by the incompatible noise, this response competes with the correct response and reaction time is prolonged. This process of response competition is a special feature of Eriksen's model.

Miller's "grain" hypothesis. The basis for Miller's hypothesis is the view that the distinction between discrete and continuous models of human information processing is an oversimplification (Miller, 1982). This distinction can be described in terms of the way in which information is transmitted. A critical concept for Miller is that of "grain". He uses it to refer to the units in which the information is transferred. For discrete models, information is transmitted in large "whole" grains - that is, all the information about the stimulus is passed on at one time. For continuous models, information is transmitted in infinitely small grains - that is, information is passed on continuously. Extending this logic, Miller argues that there may be circumstances where the grain is neither "whole" and large nor infinitely small. In this case, partial information may be transferred but the number of units is finite and may be small. He argues that these units may correspond to mental codes such as those for letters and numbers.

In a series of ingenious experiments in which partial information about the stimulus is given in advance of stimulus presentation, Miller has demonstrated that grain size varies with experimental conditions.

CONCLUSIONS

In this section, we have reviewed two classes of information processing models. While the propriety of the two classes may vary with the experimental situation (Miller, 1982), several investigators believe that continuous models provide the more veridical description of the nature of human information processing in most situations.

Although the application of serial models to psychological phenomena has been accomplished successfully with simple behavioural measures (reaction time and accuracy), it seems unlikely that parallel models will prove as tractable. This is because these models permit many more degrees of freedom in terms of the behaviour of the underlying mechanisms. Thus, a particular behavioural output such as a reaction time response is determined not by the sum of the duration of the processes intervening between stimulus and response but by a complex relationship among these processes. For example, for each information processing structure we must know the function describing the relationship between its output and its input (McClelland, 1979; Eriksen & Schultz, 1979). We must also consider the possibility that a given structure may receive inputs from several structures (Grice et al., 1982). Thus, in order to explore continuous models, we must have information about the behaviour of the intervening processes. This information is difficult to gain from the final output of the system (i.e., reaction time, accuracy). It is in this regard that psychophysiological measures may be useful in the analysis of contemporary models of human information processing.

THE ROLE OF PSYCHOPHYSIOLOGY

As we argued above, psychophysiological measures are useful to the extent that they provide information about the human information processing system that is not readily available using traditional behavioural measures. In what sense can they be useful? First, they provide information about which process occurs, and, second, they provide information about when that process occurs. With this knowledge, it should be possible to test and modify models of information flow. In a schematic representation of the human information processing system, we can conceptualize the psychophysiological responses as outputs emitted at different levels of the information processing flow. Such a representation also illustrates the value of psychophysiological responses as "mapping devices". Variables which affect the flow of information above the psychophysiological response output will also affect that response. Those variables which affect the flow of information below the

psychophysiological response output will not affect that response. In the same way, the time course of the processes may be reflected by parallel variations in the time course of psychophysiological responses. Furthermore, we can derive hypotheses about those processes which affect psychophysiological responses, but which do not affect current overt behavioural responses.

Given these considerations, several criteria are applicable to the selection of suitable measures of psychophysiological activity. First, the measures should have temporal properties (latency, etc.) which are similar to those of the processes of interest. This criterion is particularly appropriate for studies of mental chronometry, where the interest is in the timing of mental processes. In the case of preparatory processes, we would expect there to be a similarity between the time courses of both the psychophysiological changes and the processes of which the changes are manifestations. Second, psychophysiological and traditional "behavioural" measures should be dissociable, at least under some circumstances (Donchin, 1982). If psychophysiological and behavioural measures are perfectly correlated, then the former will merely serve as "substitutes" for the latter. Such redundancy would trivialize the value of psychophysiological measures, since measures of reaction time are easier and cheaper to obtain. Third, the measure should be a manifestation of the activity of a structure involved in human information processing. In some cases, there may be a match between the process manifested by the psychophysiological measure and a process proposed by some cognitive theory. If this is the case, then the measure can be used to evaluate some aspect of the theory. If such a match does not exist, the psychophysiological measure should be seen as providing just as valid data about human information processing as a behavioural measure. In this case, some revision of cognitive theories is indicated so that they account for both behavioural and physiological data. In fact, it has been argued that psychophysiological data suggest that "notions of information processing involving additive, exclusive serial processes must be oversimple. We must also admit the operation of parallel processes and feedback mechanisms, which are well-known features of neural systems, even of reflex organization" (Vaughan and Ritter, 1973: p.133).

CANDIDATE MEASURES

In this section, we consider several psychophysiological measures that are, on a priori grounds, good candidates for use in the study of human information processing. This review is not intended to be exhaustive. Rather, we have selected measures that appear to be particularly useful in evaluating some aspects of Eriksen's continuous flow model.

P300. This is a positive going component of the event-related brain potential (ERP). It occurs at least 300 msec after an event and appears to be related to the process of context-updating or the revision of working memory (Donchin, 1981). For our purposes, the most important characteristic of the P300 is that it occurs after stimulus evaluation processes have been completed, and is relatively independent of response selection and execution (McCarthy & Donchin, 1981; Magliero, Bashore, Coles & Donchin, 1984). Thus, the occurrence of P300 can be used as a marker for the end of the stimulus evaluation processes, and P300 latency can be used as a measure of their duration.

CNV. The contingent negative variation (CNV) is a negative going component of the ERP. It is observed in the interval between two stimuli when some contingency has been established. Although there is some controversy concerning the functional significance of the component, it is generally believed that it is, at least in part, related to motor preparation (see Donchin, Coles & Gratton, 1984). Thus, we propose that the CNV can be used as a marker for the presence of motor preparation - or the activation of response-related processes.

EMG. The electromyogram (EMG) is generated by the electrical activity of the muscles - and is therefore a manifestation of muscle activity. Our interest in this measure is focussed on the observation that (a) muscle activity precedes movement, (b) EMG is sensitive to "subliminal" muscle activity that does not result in an "overt" response (movement).

An example - H and S

We will now review an experiment in which we use the psychophysiological approach to evaluate Eriksen's continuous flow model (Coles, Gratton, Bashore, Eriksen and Donchin, in preparation).

The particular setting we chose for a test of the approach was an apparently simple one. Twelve male subjects were required to make a discriminative response as a function of the center letter (target) in a five letter stimulus array. There were four arrays: HHHHH, SSSSS, HHSHH and SSHSS. The responses we required of the subject were slightly unusual - a squeeze with the left or right hand of zero-displacement dynamometers at 25% of maximum force.

This particular task is a favorite of Eriksen's. As was noted above, in a variety of studies, his laboratory has found a consistent difference in reaction time between compatible and incompatible arrays. In accordance with his continuous flow model, his explanation for the differences goes s follows.

Incompatible arrays contain information about the incorrect response. As stimulus evaluation proceeds and before it is completed, this incorrect information is passed to the response activation system. Thus, the incorrect response is activated. Although the correct response may ultimately be given, the activation of the incorrect response will interfere with the execution of the correct response (through the process of response competition), thereby prolonging reaction time for incompatible arrays.

Thus, our first mission in this experiment was to determine whether the reaction time difference between compatible and incompatible arrays is indeed due to response competition effects. As we note below, we believe that we can assess the response competition process by evaluating EMG and squeeze responses on the incorrect side. An alternative hypothesis would be that it takes subjects longer to evaluate incompatible arrays, thus prolonging reaction time. This hypothesis can be tested by using P300 latency as a measure of stimulus evaluation time.

A second question in this experiment concerned the effects of a non-informative tone stimulus that preceded array presentation. We know from previous research that provision of such a warning stimulus speeds reaction time. We were interested in confirming Posner's claim that the alerting effect of a non-informative warning stimulus in this type of situation is due to a change in response-related processes (Posner, 1978). We propose that this claim can be investigated by using P300, EMG and squeeze measures, as well as measures of preparatory processes (late CNV).

A critical aspect of the continuous flow model is that it proposes that the activation of the incorrect response interferes with the execution of the correct response thereby postponing reaction time. To analyze this process of response competition, we used measures of EMG activity and squeeze activity on the incorrect side to classify trials in terms of their degree of error.

N - Activity only on the correct side in EMG and squeeze channels

E - Activity on the correct side for EMG and squeeze channels: activity also present for EMG on the incorrect side

S - Activity on the correct side for EMG and squeeze channels: activity also present for both EMG and squeeze channels on the incorrect side.

Error - Activity on the incorrect side for EMG and squeeze channels. EMG activity on the correct side may or may not be present.

We will first review evidence concerning the nature of the compatibility effect (see Figure 1).

The upper part of Figure 1 illustrates that S and Error trials (where the wrong squeeze response was produced) occurred more often when the arrays were incompatible. The lower part of Figure 1 illustrates two main points. First, the latency of activity on the correct side increases as the degree of activity on the incorrect side increases - that is, correct responses were longer when there was

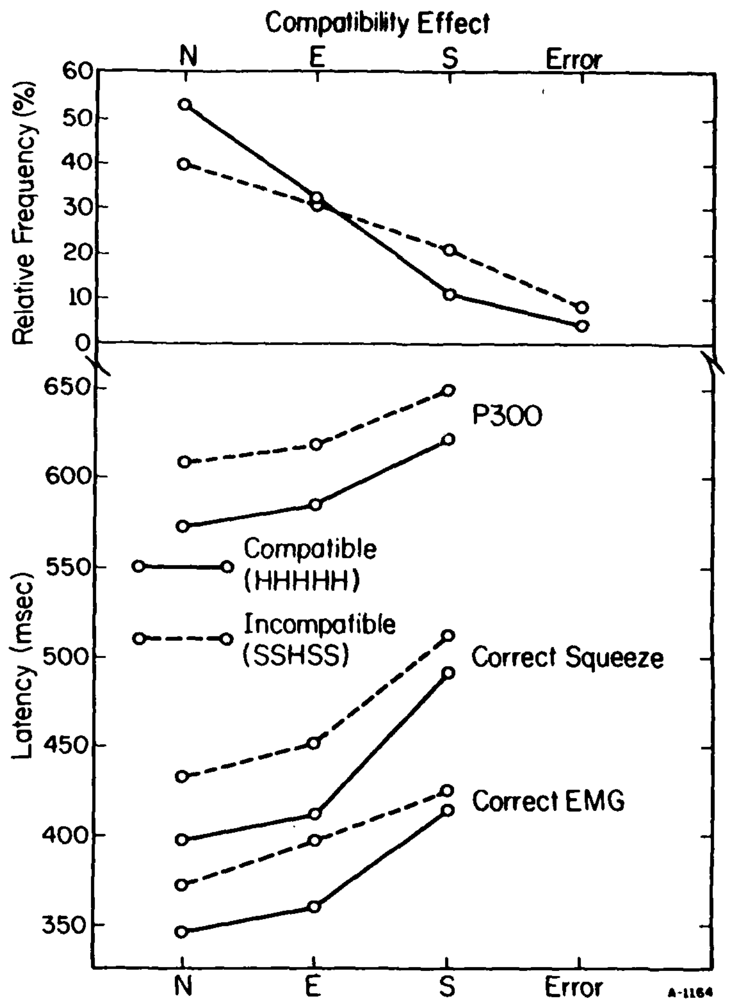


FIGURE 1 Latency of correct EMG and squeeze activity and of P300 as a function of the degree of incorrect activity for compatible and incompatible arrays. The relative frequency of each response class or compatible and incompatible arrays is also shown.

activity on the incorrect side (E and S categories). Second, the time between the manifestation of response initiation (by the EMG) and response completion (the squeeze) increases with the degree of error. This can be seen most clearly by comparing (within compatible or incompatible conditions) the difference between EMG and squeeze onset for S trials with the comparable differences for N or E trials. The difference is significantly greater for S trials. Thus, with incompatible arrays there are more trials in which the subject squeezed the incorrect dynamometer (S), and on these trials the interval between EMG onset and squeeze onset for the correct dynamometer is longer than for the other conditions (N,E). There is evidence of response competition when EMG activity occurs on the incorrect side (E and S), and when force is also applied to the incorrect dynamometer (S) response competition is greatest. Taken together, these findings support the hypothesis that at least one reason for the longer incompatible reaction times is the fact that response competition (especially in the S category) occurs more often when the array is incompatible, and reaction time is prolonged by response competition.

This is not the whole story, however. When we look at the latency data for compatible and incompatible arrays within the different error categories we note that, in all cases, both P300 latency and reaction time are prolonged. (P300 latency is presented in the middle of Figure 1.) Thus, the effect of incompatibility on reaction time appears to be due both to a greater incidence of response competition for incompatible trials and to the generally longer stimulus evaluation time. If we combine all correct trials (N,E,S), then responses to incompatible trials are approximately 40 msec. longer than responses to compatible trials. We can explain about 70% of this difference (28 msec) by the change in stimulus evaluation (indexed by P300 latency), while the other 30% (12 msec) stems from response competition. There is also a relationship between P300 latency and the degree of error: the longer the stimulus evaluation time, the more likely it is that the subject will activate the incorrect response.

Our second question concerned the effect of the non-informative warning tone (non-informative in

terms of response choice). Why does the warning stimulus speed up responses?

The relevant data are shown in Figure 2. First, note that for all response classes, both EMG and squeeze latencies are shorter for the warned condition. Second, note that P300 latency (that is, stimulus evaluation) is not affected by the warning. Third, note that warned trials are associated with a slightly higher incidence of incorrect activity. This is most evident for the S category.

Thus, the effect of the warning stimulus is to decrease response latency by about 30 msec., and increase the incidence of incorrect activity (S) by about 3%. At the same time, the warning stimulus does not affect the latency of P300 (stimulus evaluation time). These findings are most parsimoniously interpreted in terms of the speed-accuracy trade-off function. The warning tone leads the subject to adopt a less conservative strategy.

We believe that this strategy is best considered in terms of an "aspecific activation" process, that is, activation of the response *channels* can occur independent of the specific nature of the stimulus. In the case of the warning, this activation occurs during the foreperiod and may be manifested by the large CNV that is present.

Variations in the level of aspecific activation are also responsible for the presence of incorrect activity on compatible trials, that is, errors occur even when there is nothing in the stimulus to trigger an incorrect response.

Our data, then, suggest that the compatibility effect on reaction time is due to both greater incidence of response competition and slower stimulus evaluation for incompatible arrays. The warning effect is due to greater aspecific activation on warned trials. This manifests itself in the form of a change in response bias towards a less conservative strategy. Since activation of both correct and incorrect responses can occur on the same trials, some sort of continuous flow conception (as proposed by Eriksen) is warranted. However, since incorrect activity can occur even when there is nothing in the stimulus array to call

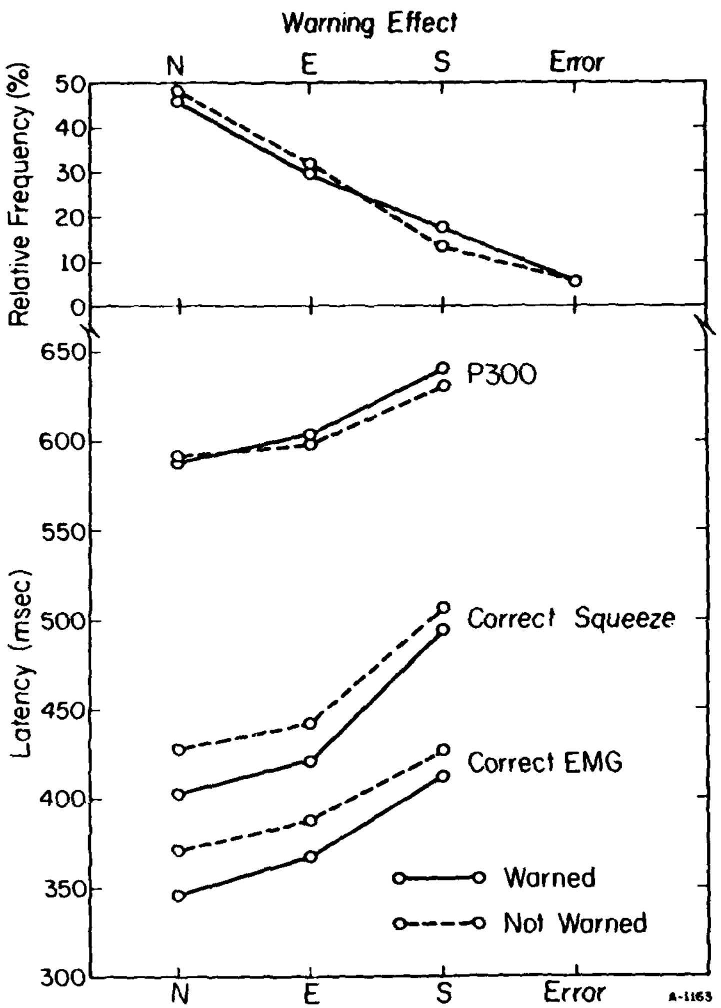


FIGURE 2. Latency of correct EMG and squeeze activity and of P300 peak as a *function of* the degree of incorrect activity for warned and not warned trials. The relative frequency of each response class for warned and not warned trials is also shown.

for the incorrect response, it seems likely that we must also infer that some form of aspecific activation can occur (as proposed by Grice et al., 1982).

CONCLUSIONS

What understanding of cognitive processes have we gained by adopting the psychophysiological approach to the H and S experiment?

The combined use of P300 and EMG response measures has allowed us to make inferences about the cognitive locus of the Eriksen effect. Similarly, information provided by the P300, and to some extent by the CNV, enabled us to say something about the locus of the warning effect. Beyond this, we have several indications of how, with some modifications, the psychophysiological approach might be even more useful. First, if the hypothesis that the CNV is a manifestation of aspecific activation is correct, then a more detailed analysis of the CNV will help us understand this process in more detail. For example, the late CNV is lateralised as a function of the response channel (left or right) which is activated prior to overt response (Rohrbaugh, Syndulko and Lindsley, 1976). Thus, lateral CNV measures derived on individual trials will provide a description of processes that occur during the foreperiod at a time when no overt behaviour is available. Second, we can obtain a precise description of the stimulus evaluation process by looking at speed-accuracy trade-off functions for trials with different P300 latencies. This will enable us to describe in detail the differences in evaluation between compatible and incompatible displays. Together, these two psychophysiological approaches should provide us with a detailed description of two processes that are determinants of the final overt response - stimulus evaluation and aspecific activation - and their inter-relationship.

ACKNOWLEDGEMENTS

We wish to thank Charles Eriksen, Monica Fabiani, Demetrios Karis, Art Kramer, Greg Miller and Walter Schneider for their helpful comments on preliminary versions of the manuscript. Preparation of the chapter was supported in part from AFOSR (F49620-83-0144), Dr. Al Fregly, Project Director.

FOOTNOTE

In this paper, we propose that psychophysiological data can be used to test and refine models of cognitive function. A complementary view is that psychophysiological data can be used to build models of brain function. An example of this view can be found in Papakostopoulos, D. (1968).

REFERENCES

Ashby, F.G. & Townsend, J.T. (1980). Decomposing reaction time distributions. Pure insertion and selective influence revisited. Journal of Mathematical Psychology. 21, 93-123.

Coles, M.G.H., Gratton, G., Bashore, T.R., Eriksen, C.W. and Donchin, E. (in preparation) The psychophysiological investigation of the continuous flow model of human information processing.

Donchin, E. (1981). Surprise!.....Surprise? Psychophysiology. 18, 493-515.

Donchin, E. (1982). The relevance of dissociations and the irrelevance of Dissociationism: a reply to Schwartz and Pritchard. Psychophysiology. 19, 457-463.

Donchin, E., Coles, M.G.H. and Gratton, G. (1984). Cognitive psychophysiology and preparatory processes: A case study. In: S. Kornblum and J. Requin (Eds.) Preparatory States and Processes. Hillsdale, NJ: Erlbaum. (pp. 155-178)

Donders, F.C. (1868). On the speed of mental processes. In: W.G. Koster (Ed. and trans.) Attention and Performance II Amsterdam: North-Holland (1969). (pp. 412-431)

Eriksen, C.W. and Schultz, D.W. (1979). Information processing in visual search: A continuous flow conception and experimental results. Perception and Psychophysics, 25, 249-263.

Grice, G.R., Nullmeyer, R. and Spiker, V.A. (1982). Human reaction times: Toward a general theory. Journal of Experimental Psychology: General, III, 135-153.

Magliero, A., Bashore, T.R., Coles, M.G.H. and Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. Psychophysiology. 21, 171-186.

McCarthy, G. and Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. Science. 211, 77-80.

McClelland, J.L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. Psychological Review, 86: 287-330.

Miller, J. (1982). Discrete versus continuous stage models of human information processing: In search of partial output. Journal of Experimental Psychology: Human Perception and Performance. 8, 273-296.

Pachella, R.G. (1974). The interpretation of reaction time in information processing research. In: B.H. Kantowitz (Ed.) Human Information Processing: Tutorials in Performance and Cognition Hillsdale, N.J.: Erlbaum, pp. 41-82.

Papakostopoulos, D. (1978). Integrative models: Macropotentials as a Source for Brain Models. In: D. Otto (Ed.) Multidisciplinary Perspectives in Event-Related Brain Potential Research. U.S. Environmental Protection Agency, Washington, D.C., 635-639.

Pieters, J.P.M. (1983). Sternberg's additive factor method and underlying psychological processes: Some theoretical considerations. Psychological Bulletin. 93, 411-426.

Posner, M.I. (1978). Chronometric Explorations of Mind. Hillsdale, N.J.: Erlbaum.

Rohrbaugh, J.W., Syndulko, K. and Lindsley, D.B. (1976). Brain components of the Contingent Negative Variation in humans. Science. 191, 1055-1057.

Sternberg, S. (1969a). The discovery of processing stages: Extensions of Donders method. In:W.G. Koster (Ed.) Attention and Performance II Amsterdam: North-Holland. (pp. 276-315).

Sternerg, S. (1969b). Memory scanning: Mental processes revealed by reaction time experiments. American Scientist. 57, 421-457.

Vaughan, H.G. Jr. and Ritter, W. (1973). Physiologic approaches to the analysis of attention and performance. In: S. Kornblum (Ed.) Attention and Performance IV New York: Academic Press. (pp. 129-154).

Wickens, C.D. (1980). The structure of attentional resources. In: R. Nickerson (Ed.). Attention and Performance VIII Hillsdale, N.J.: Erlbaum. (pp. 239-257).

Woodworth, R.S. (1938). Experimental Psychology. New York: Holt.

A NEUROPHYSIOLOGICAL INVESTIGATION OF MINOR HEAD INJURY

R.J. McClelland, The Queens University of Belfast, Whitla Medical Building, 97 Lisburn Road, Belfast, BT9 7BL.

Head injury, particularly head injury from road traffic accidents, consitutes a major health problem particularly in young adults. Probably the most reliable guide to the incidence of head injury in the community is attendance at Accident and Emergency Departments and one survey of Scottish hospitals attendance rate indicated 18 in 1,000 of the population (Strang et al., 1978). Male attendances were over twice as common as females and the rate for children and young adults was approximately twice the overall attendance rate. Two major causes, road accident and assault, accounted for 31% of all attendances and for over half of attendances in the age range 15-24.

Head injury embraces all degrees of severity, from a symptomless blow on the head to irreversible traumatic brain damage. Therefore to understand better the natural history of head injury and the contributions made by the different therapeutic investigations, measures of severity are critical (Jennett, 1976). Clinical measures which have been used to assess the severity of head injury are duration and depth of consciousness, duration of post-traumatic amnesia and depth of injury. Implementation of the Glasgow Coma Scale for the assessment of unconsciousness has considerably reduced the amount of semantic confusion and subjectivity that surrounds the assessment of this clinical parameter (Teasdale and Jennett, 1974).

However, problems still exist in making a

To appear in: R. Hockey, A. Gaillard, & M. G. H. Coles (Eds.), Energetic and Human Information Processing. Dordrecht, The Netherlands: Nijhoff.

COGNITIVE PSYCHOPHYSIOLOGY AND THE STUDY OF STATES AND PROCESSES

Michael G. H. Coles and Gabriele Gratton

Cognitive Psychophysiology Laboratory

Department of Psychology

University of Illinois

Champaign, Illinois 61820, USA

ABSTRACT

The complexities introduced by recent theorizing suggest that discrete final output measures may be insufficient to provide a satisfactory description of the human information processing system. We propose that this problem may be resolved by adopting the psychophysiological approach, which is based on the assumption that different psychophysiological measures are sensitive to different aspects of information processing. We illustrate the utility of the approach by describing an experimental program that utilizes measures of the electromyogram and the event-related brain potential to provide direct assessments of stimulus evaluation and response-related processes. In particular, we are able to describe the relative influence of response priming, and of different aspects of stimulus evaluation, on the accuracy of overt responses emitted at different latencies. We are also able to obtain a continuous measure of the activation of response

channels. We consider these findings in the context of questions about the relationship between states and processes, and conclude that our psychophysiological measures may be used to describe how two process parameters, initial level and gain, can be related to state variations.

INTRODUCTION

The premise of cognitive psychophysiology is that understanding of human cognitive processes can be enhanced through the use of psychophysiological measures in conjunction with measures of overt behavior (e.g., see Coles & Gratton, in press; Donchin, 1979; Donchin, Coles, & Gratton, 1984). In this chapter, we shall illustrate the approach of cognitive psychophysiology by describing some recent research in our laboratory (Coles, Gratton, Bashore, Eriksen, & Donchin, 1985; Gratton, Coles, Sirevaag, Eriksen, & Donchin, in preparation). We shall also discuss how this research may inform the questions with which this book is concerned, namely the relationship between energetical states and psychological processes.

Contemporary theories of human information processing have made some fairly radical proposals concerning the manner in which the elements of the system operate. Rather than conceiving of the system as a series of elementary processors arranged in series, with the activation of one processor possible only when the previous element has finished its operation, these new theories propose that elements communicate continuously (c.f. Eriksen &

Schultz, 1979; Grice, Nullmeyer, & Spiker, 1982; McClelland, 1979). This view implies that elements of the information processing system can be simultaneously active and has led to a considerable increase in the complexity of the procedures required to understand how a particular input to the processing system is "translated" into a particular output. For example, a given experimental manipulation may affect the duration of an individual process (c.f. the serial view) -- but it could also affect the rate at which one element transmits information to other processing elements, and/or the criterion level at which such information transfer occurs. Furthermore, the level of initial activation of a process (priming) may be influenced either by experimental manipulations or by variation in internal states.

Given this potential complexity, how can we understand the way in which the system deals with an input so as to produce a particular output? Cognitive psychologists, whose interest is in information processing, have proposed some ingenious solutions to this problem that rely both on new experimental procedures and on new methods of data analysis (e.g., Grice, Nullmeyer, & Spiker, 1982; McClelland, 1979; Meyer, Yantis, Osman, & Smith, 1985; Miller, 1982). However, the new methods are still applied to measures of overt behavioral output (reaction time and accuracy) when the overarching interest is in the <u>intervening</u> processes that give rise to the overt behavior. The task of describing the information processing system would be considerably easier if we had measures of the activity of particular elements of the system, as well as measures of its output. This is where psychophysiology may help. If psychophysiological measures are sensitive to particular information processing activities, then we should be able to use them to understand how these processes interact to produce the behavioral output.

After describing the experimental setting in which we apply the psychophysiological approach, we shall review both psychophysiological and behavioral data that illustrate how the approach can provide useful insights into the nature of human information processing.

THE EXPERIMENTAL SETTING

The setting for our experiments (Coles, Gratton, Bashore, Eriksen, & Donchin, 1985; Gratton, Coles, Sirevaag, Eriksen, & Donchin, in preparation) involves a visual search task. Subjects are asked to respond with their left or right hands to a target letter (H or S) that appears at the center of a five-letter stimulus array. Hand-letter assignment is counter-balanced across subjects.

The target letters are surrounded by either the same letter (compatible noise condition -- e.g., HHHHH or SSSSS) or by the letter calling for the opposite response (incompatible noise condition -- SSHSS or HHSHH). In all cases, the center letter is presented at fixation and the array (target and noise letters) subtends 2.5 degrees of visual angle.

In most conditions, the four arrays occur in a random order such that the global probability of each array is .25 and the probability of each of the two target letters (and of each response) is .5. Usually, a warning tone precedes array presentation on each trial by one second. Note that the warning tone only provides information about the timing of array presentation: it does not tell the subjects which array is to be presented.

For reasons that will be apparent later, we require a somewhat unconventional response -- a squeeze with the right or left hand of zero displacement dynamometers. To register a "response", squeeze force has to exceed 25% of maximum force. Subjects are given extensive practice with feedback to learn what it feels like to squeeze at, or above, this force level.

Eriksen and his colleagues (e.g. see Eriksen & Schultz, 1979, for review) have shown that reaction times are slower, and error rates are higher, for incompatible arrays. We refer to this as the "noise compatibility effect". Our overt behavioral data are entirely consistent with these previous findings.

RESPONSE CHANNELS AND THEIR ACTIVATION

Most contemporary models of information processing propose that a response is emitted when a hypothetical element of the system, a response "channel", is activated at a criterion level (e.g. Eriksen & Schultz, 1979; Grice et al., 1982; McClelland, 1979).

Given the variety of responses available to our subjects, there are, in principle, a large number of response channels that could be activated in our experimental situation. However, for present purposes, we shall focus only on the two response channels that are critical for successful task performance -- one associated with the left hand and the other with the right. Since our interest is in response accuracy, and since target letter/response hand assignment is varied between subjects, we shall refer to these as "correct" and "incorrect" response channels.

We propose that for each channel, there is an activation process, such that, at any time, response channel activation can vary between minimum and maximum levels (see Figure 1). Overt behavior (in the form of a squeeze of the dynamometer) occurs when a criterion, or threshold level of activation is achieved. Muscle activity also occurs at a threshold level -- however, since muscle activity can occur in the absence of overt behavior, muscle activity must have a lower criterion level of response channel activation than the squeeze. These proposals are illustrated in Figure 1.

We propose that activation in the two response channels can be influenced by three factors: stimulus related information processing, response competition, and response priming or bias.

Stimulus-Related Information Processing

According to Eriksen's continuous flow model (Eriksen & Schultz, 1979), as information about a stimulus is accumulated in a stimulus evaluation system, the response channel associated with that stimulus is activated. Thus, if correct information is accumulated, the correct response channel will be activated; if incorrect information is accumulated, the incorrect response channel will be activated. This is shown diagrammatically in Figure 1. Since, in our experiment, the stimulus array sometimes contains information for both responses (the incompatible noise condition), both response channels may be simultaneously activated by the same stimulus array. Furthermore, incompatible noise may also introduce some perceptual conflict that can interfere with the evaluation process such that transfer of information from stimulus evaluation to response activation systems is delayed.

Insert Figure 1 About Here

Response Competition

When the two response channels are activated simultaneously (there is more than zero activation in both channels), the responses will tend to interfere with each other such that their activation levels are reduced and/or the slope of their activation functions decreases. This competition mechanism may be due to either a "hard-wired" mechanism by which movement of one hand inhibits movement of the other or to a competition set up by the experimental instructions when the subject is told that only one response is correct on any trial. Eriksen & Schultz (1979) propose that this response competition mechanism is responsible for the noise compatibility effect. As we noted above, because incompatible arrays contain information about both correct and incorrect responses, these arrays will tend to be associated with activation of both response channels. Because of response competition, execution of the correct response will be delayed by the concurrent activation of the incorrect response (see Figure 1).

Priming and Bias

In a situation where subjects may have to respond with either hand on any trial, they may prime one or other of their response channels in advance of stimulus presentation. While this "guessing" behavior may be guided by antecedent events (such as the stimulus on the previous trial), it is, by definition, independent of the stimulus on the current trial. Thus, correct or incorrect response channel activation can occur which is independent of the nature of the current stimulus (see Figure 1).

PSYCHOPHYSIOLOGICAL AND BEHAVIORAL DATA

The principle concern in our experiments was to understand the factors that control the accuracy and speed of subjects' responses. We argued in the previous section that three factors may be responsible for the activation of correct and incorrect response channels, and, by implication, these factors should control both the accuracy and latency of overt responses. We expected to find that incompatible noise arrays would be associated with slower reaction times and more errors than compatible noise arrays. In fact, our overt behavioral measures (the latency at which the squeeze response crossed the 25% force threshold and the accuracy of that response) were entirely consistent with the data obtained previously by Eriksen and his colleagues (Eriksen & Schultz, 1979). In our case, the effect of noise incompatibility was to increase reaction time by 47 msec and error rate by about 9%.

Stimulus-Related Processing

To explore the influence of stimulus-related processing on overt response accuracy and latency, we measured the P300 component of the event-related brain potential (ERP). From a variety of studies, evidence has been accumulated to suggest that the latency of the P300 is related to the duration of stimulus evaluation processes and is relatively unaffected by processes associated with response execution (see Coles et al., 1985, and Coles, Gratton, & Donchin, 1984, for a review of these studies). Thus, to address the question of a noise compatibility effect on stimulus processing, we began by determining whether the latency of the P300 was longer for incompatible arrays. While the effect of noise compatibility on squeeze latency was 47 msec, P300 latency was 27 msec longer for incompatible arrays. Thus, part of the noise-compatibility effect could be attributed to differences in the duration of stimulus evaluation. As we will see, the remaining effect of noise-compatibility can be attributed to overt and latent response competition.

While this result was interesting, there is clearly more to the stimulus evaluation process than its duration. Thus, our next set of analyses was designed to determine whether the evaluation process itself was different in the two noise conditions. We began by employing a technique commonly used in experimental psychology -- the speed-accuracy trade-off curve (note 1) -- in which trials are sorted into different response latency bins, and the accuracy for each bin is computed. Figure 2a shows separate speed-accuracy trade-off functions for the two noise conditions. If evaluation processes were merely delayed in the incompatible condition, we would expect that the only difference between the functions for compatible and incompatible noise arrays would lie in their offset -- that is, the incompatible function would be shifted along the abscissa relative to the compatible function. However, the incompatible array appears to have a different shape, with a decrease in response accuracy below the 50% chance level for responses with latencies of between 200 and 300 msec.

Before we consider the significance of this shape difference, we need to review an objection that has been raised by Meyer and his colleagues (Meyer & Irwin, 1982) concerning the use of speed-accuracy trade-off functions to make inferences about the nature and time course of evaluation processes. Meyer has argued that these functions may give one the false idea that there is a gradual increase over time following stimulus presentation in the amount of information available for performing the task. In fact, Meyer has shown that one can obtain gradually increasing functions even if evaluation is an all-or-none process -- that is, there is a discrete transition between the state where "no information about the stimulus is available" and the state where "all information about the stimulus is available". This will happen if there is variability over trials in the time at which such a discrete transition occurs -- that is, variability in the duration of the evaluation process -- and if data are averaged over trials (as they always are, when speed-accuracy trade-off functions are computed).

Insert Figure 2 About Here

To deal with this problem, what we need is an index of the duration of the evaluation process. If the process is indeed discrete, then if we adjust each trial for the timing of the response relative to the duration of the evaluation process, the speed-accuracy trade-off function should approximate a discrete, step function.

As we stated earlier, P300 appears to occur after the end of evaluation. Thus, to control for the duration of evaluation processes we divided each subject's reaction time (that is, latency of EMG onset) by the latency of P300 on every trial. Conceptually, this gives us a ratio measure that describes the proportion of the evaluation process that has transpired at the moment the reaction time response is released (or relative reaction time). Figure 2b shows speed-accuracy trade-off functions with relative, rather than absolute, reaction time on the abscissa. Several aspects of these functions are notable. They do not suggest that the evaluation is a discrete, all-or-none process. However, they do suggest the presence of several phases in the evaluation process. If we examine the course of both compatible and incompatible functions as relative reaction time increases, three phases become apparent. Initially, response accuracy for both noise conditions is at 50% (or chance) level. When subjects respond this quickly, then, they appear to be guessing in the sense that their responses are independent of any information conveyed by the array. As responses become relatively slower, accuracy increases for compatible noise arrays, but decreases below chance level for incompatible arrays. This difference between the two arrays is presumably due to the difference in the peripheral noise letters. Thus, at this intermediate phase of evaluation, subjects appear to be getting information about the features of the letters in the array, but they appear to be confused about the location of the features. We refer to this phase of evaluation as the "feature analysis" phase. As relative response times increase further, accuracy increases for each type of array. Since the arrays share the property of having the same target letter, we infer that at this phase of evaluation, subjects are able to locate the central target letter -- that is, they are able to perform "location analysis".

To explicate the relative contribution over time of these two types of analysis to response accuracy, we obtained two derived functions. Since feature analysis is manifested by the difference in the adjusted speed-accuracy functions for the two types of array, the derived feature analysis function is obtained by subtracting the functions for compatible and incompatible arrays. On the other hand, location analysis is manifested by the similarity in the speed-accuracy functions for compatible and incompatible arrays. Therefore, the location analysis function is obtained by taking the mean of the speed-accuracy functions for the two noise conditions. Both derived functions are shown in Figure 3. They have been adjusted by a constant so as to represent the contribution of the two types of analysis to response accuracy at each relative response latency. Note that the original functions shown in Figure 2b can be derived from the derived functions by adding (or subtracting) the values shown in Figure 3 to the chance level (50%).

What information have we gained about stimulus evaluation processes in our paradigm? The P300 latency data suggest that stimulus evaluation is slower when the noise is incompatible. We also have evidence from speed-accuracy trade-off functions adjusted for P300 latency that the evaluation process has three distinct phases -- a phase when no information has been accumulated, a phase when information about the features is available but there is no information about their location, and a phase when information about the features of the letter located in the central target information is available. The proposal that the evaluation of the array consists of two phases, feature and location analysis, is consistent with conclusions reached by Treisman and Gelade (1980). In contrast to their approach, we have relied on psychophysiological measures to understand how evaluation proceeds. In particular, we have avoided the problems with speed-accuracy functions described by Meyer. In this sense, then, a psychophysiological measure (P300 latency) has enhanced our understanding of the nature and time course of evaluation processes.

Insert Figure 3 About Here

Response Competition

As is illustrated in Figure 1, response competition should be evident in a decrease in the slope of the activation function of a particular response channel. This decrease in slope is, in turn, associated with an increase in the interval between muscle and squeeze activity. Thus, to study the process of response competition we measured the electromyogram (EMG) from the muscles associated with the squeeze response in both hands. The interval between the onset of EMG activity and the onset of squeeze activity was assessed on each trial. If activation of the incorrect response channel interferes with the execution of the correct response, this interval should be longer when incorrect response activity is evident. We used measures of EMG and squeeze activity associated with the incorrect response to classify trials into different categories according to the degree of incorrect activity present (Coles, Gratton, Bashore, Eriksen, & Donchin, 1985). We assumed that this classification reflected different levels of activation in the incorrect response channel. In fact, we found that the interval between the onset of EMG and squeeze activity on the correct side varied with the level of activation of the incorrect response channel. In particular, the interval was longer when the activation achieved the level of a squeeze response than when it did not reach this level. These data provide clear evidence for the existence of a response competition process.

As we have noted, Eriksen's theory proposes that incompatible noise arrays are associated with the activation of the incorrect response and that it is this incorrect response activation that delays correct response execution. We have seen that, in general, correct response execution is delayed when the incorrect squeeze response is activated. However, our data also indicated that the incompatible noise condition was associated with a higher frequency of trials on which incorrect squeeze responses occurred. In fact, this effect of incompatible noise accounts for 10 msec of the 47 msec overall difference in reaction time between compatible and incompatible arrays.

Now, even when there was no overt manifestation of activation of the incorrect response channel, the interval between correct muscle and correct squeeze activity was longer for the incompatible condition. This finding suggests that response competition processes may operate even when activation of the incorrect response channel does not reach the criterion for peripheral manifestation (in the form of the muscle or squeeze activity). This latent form of response competition accounts for an additional 12 msec of the overall noise-compatibility effect.

These data indicate that the process of correct response execution can vary as a function of the activity in other response channels (response competition). It is clear that insights into the response execution process, as well as the activation of response channels, can be gained by measuring muscle activity and by using a continuous rather than discrete response manipulandum.

Response Priming and Bias

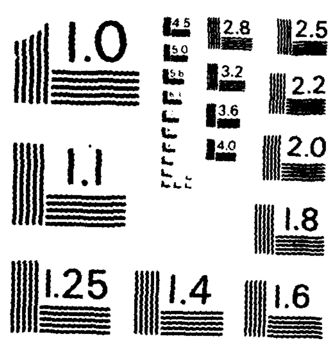
We have seen that very fast responses (either in absolute terms or relative to P300 latency) are associated with chance levels of accuracy. Furthermore, we argued earlier that, irrespective of the type of array, subjects may prime their responses in advance of, or at least independently of, array presentation. The question we now address is whether the accuracy of these guesses can be predicted by a psychophysiological measure of priming -- that is, can we use psychophysiological measures, obtained <u>before</u> array presentation, to determine whether subjects will emit correct or incorrect responses <u>after</u> array presentation?

To assess this response priming process, we turn to measures of the ERP obtained during the time-period just prior to array presentation. We know from previous research that, during the foreperiods of warned reaction time tasks, a slow negative wave (the contingent negative variation - CNV) is observed. Furthermore, it has been suggested that the part of this wave that occurs just before the stimulus is related to motor preparation (see, for example, Rohrbaugh & Gaillard, 1983; Rohrbaugh, Syndulko, & Lindsley, 1976). In addition, this research and other studies of voluntary movements (e.g., Kutas & Donchin, 1977) have indicated that this negative wave is lateralized (larger on one side of the scalp) when subjects anticipate responding with a particular hand. For right-handed subjects, left-hand movements are associated with larger negativity on the right side of the scalp and vice versa for right-hand movements. In terms of the notion of response channels, these data suggest that activation of these channels may be manifested by lateralized negativity at the scalp. In particular a scalp electrode placed over the motor cortex contralateral to the intended response will exhibit larger negativity than an ipsilateral electrode. Therefore, in our experiment, we use measures of the lateralized negativity recorded just prior to array presentation to infer which response channel has been primed on that particular trial.

By evaluating the pattern of negativities in the 100 msec before array presentation on each trial, we classified the trials according to the following scheme. <u>Contralateral</u> trials were those for which the negativity was larger at the scalp site that was on the opposite side of the body from that of the correct response on that trial. Thus, these trials may be thought of as those for which the correct response was primed. <u>Ipsilateral</u> trials were those for which negativity was larger at the scalp site on the same side of the body as the correct response. These trials may be thought of as those for which the incorrect response

AD-A170 794 THE EVENT-RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING A..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 2/8

UNCLASSIFIED 20 FEB 86 CPL-86-1 F49620-85-C-0041 F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

was primed. No difference trials were those for which there was a small difference between the two lateral scalp sites. These trials may be thought of as those for which neither response was preferentially primed. (Note that our measure does not differentiate between the situation when neither response is primed and that when both responses are primed.)

Insert Figure 4 About Here

Having classified our trials in this fashion, we looked at the relationship between priming and speed-accuracy trade-off functions. As can be seen in Figure 4 for both compatible and incompatible conditions, trial classification is related to response accuracy. Accuracy is higher for contralateral (correctly primed) trials than for ipsilateral (incorrectly primed) trials, at least for relatively fast responses. Accuracy for "no difference" trials is intermediate.

To obtain a representation of the time course of this effect, we computed a derived function by taking the difference between the functions for contralateral and ipsilateral trials. This derived function is shown in Figure 3. As with the other functions shown in this figure, it has been adjusted by a constant so that the original speed-accuracy functions can be derived from it. In this case, if the trial is a correctly primed trial (Contralateral), then the value of the priming function at any relative response latency is added to the chance level (50%). If it is an incorrectly primed trial (Ipsilateral), the value of the priming function is subtracted.

What the derived function (Figure 3) shows is that the accuracy of relatively fast responses is more influenced by priming processes than that of slower responses. The accuracy of these slower responses, as we have seen, is more under the control of stimulus evaluation processes.

Our data indicate that lateralization of the negativity at the scalp provides an index of response priming. While priming can be investigated in other ways (for example, by the probe method), the psychophysiological measure has the advantage of being totally unobtrusive.

Response Channel Activation

The results of the analyses described in previous sections support the proposal that the activation of response channels is controlled by priming, evaluation, and response competition processes. The <u>timing</u> of responses is controlled to some extent by response competition. Figure 3 provides a graphic representation of the relative importance of evaluation and priming processes in determining response <u>accuracy</u> as a function of the time at which a response is released relative to the duration of evaluative processes. If we were able to obtain a continuous measure of response activation, we should be able to see the influence of these processes at different points in time. Such a continuous measure should also enable us to determine whether the criterion for overt response activation is constant for responses of different latencies (see Figure 1), or whether variability in response latency is associated with variability in criterion levels of activation. If the latter is the case, then we might expect to find that fast responses are released at a lower level of response channel activation than are slow responses.

We have begun to study response activation directly using a measure of the ERP. In fact, it is the same measure we described earlier in connection with response priming. We record the voltage at lateral scalp sites (above the motor cortices) and examine the difference between the voltage at these two sites as it changes over time. This difference, derived for the period beginning before the warning tone and extending beyond array presentation, is our measure of relative response activation. Conceptually, the measure may be thought of as representing the difference in the level of activation between the two response channels. Activation of the correct response is greater to the degree that the voltage at the scalp site contralateral to the correct response is more negative. Greater ipsilateral negativity represents greater incorrect response activation.

Trials were sorted on the basis of the type of array (compatible or incompatible), the accuracy of the subject's response, and the response latency. Then, average values of the difference in voltage at the lateral scalp sites were obtained. The data are shown in Figure 5. (There were insufficient trials to obtain stable averages for the compatible correct condition.)

Insert Figure 5 About Here

Several aspects of the data shown in Figure 5 are notable. First, for fast correct trials, be they compatible or incompatible, the correct response channel is activated <u>in advance of</u> array presentation. Conversely, for fast incorrect, incompatible, trials the incorrect response channel is activated prematurely. These data are consistent with the priming results described earlier which showed that response accuracy was related to priming for relatively fast responses. A second interesting aspect of the data concerns the behavior of the activation functions <u>after</u> array presentation. In each panel of Figure 5, the activation functions "take-off" in the direction of the response that was actually given. The four vertical lines in each panel represent the beginning of each reaction time bin, and the point at which each vertical line crosses its associated activation function has been indicated by a check mark. This point represents the degree of asymmetry at which the EMG response is released -- and, conceptually, corresponds to the relative level of activation at which a response is released. The interesting aspect of these data is that, regardless of reaction time and condition, there appears to be a fixed degree of asymmetry (or a criterion activation level) which, when exceeded, is associated with an overt response. In this sense, our data are consistent with a fixed criterion model of response execution.

Further inspection of the data for incompatible, correct trials provides an indication of an influence of evaluation processes on response channel activation. After stimulus presentation, most of the waveforms are characterized by a small dip towards incorrect response activation followed by a sharp deflection toward the correct response activation. The latter sharp deflection presumably reflects the influence of information concerning the target letter on response activation -- that is, it reflects the outcome of location analysis. Can we also see evidence for feature analysis in the activation functions? A speculative but interesting hypothesis is that the small early dip toward incorrect response activation represents the contribution of feature analysis. Early in the evaluation of incompatible arrays, evidence for the incorrect response activates the incorrect response channel. This leads to a change in relative response activation in the direction of the incorrect response. Note that the time at which this "incorrect dip" occurs corresponds quite closely to the time at which feature analysis is dominant, as is evident in Figure 2a. Note also that this dip is not present in the waveforms for the compatible, correct trials. In this case, both feature and location analysis drive response channel activation in the direction of the correct side.

These preliminary data concerning response activation suggest that we may fruitfully use measures of lateral scalp negativity to derive a continuous measure of relative response activation. The measure appears to provide information about activation _as it_ occurs before overt responses are evident either at the muscle or at the level of movement.

Of course, there are limitations to the measure. First, in principle it is possible for the two response channels to exhibit relatively independent activation functions. However, our measure is, by definition, only sensitive to the relative difference between activation functions. Since electrical fields are volume conducted by the brain, measures of the electrical activity at a particular scalp site do not necessarily represent the electrical activity of a source directly below that site. In fact, it is possible for an electrode placed over the left motor cortex to be sensitive to the electrical activity generated in the right motc cortex (as well as in the left motor cortex). The problems raised by volume conduction are serious -- especially if we want to obtain a measure of left activation that is uncontaminated by right activation (or vice versa). We are currently working on this problem.

A second limitation of the data presented in Figure 5 is that it does not take account of the fact that, on some trials, both response channels are activated to the level of muscle, and, in some cases, squeeze responses. Evidence for this was discussed earlier when we considered response competition. The problem with studying response activation processes under different degrees of response competition is that the latency of, say, muscle activity on the correct side is relatively independent of the latency of muscle activity on the incorrect side. Thus, to obtain stable average measures of scalp activity that are time-locked to both correct <u>and</u> incorrect muscle activity would require an extraordinarily large number of trials. To obtain the averages shown in Figure 5, our subjects performed 3600 trials each! This number will have to be increased by a large factor if we are to be able to look directly at the effects of response competition on activation processes.

These technical problems aside, the data discussed in this section and presented in Figure 5 suggest that (a) response activation processes begin in advance of array presentation, (b) evaluation processes feed directly into activation processes, and (c) responses are released when a fixed level of activation is attained.

Summary

What understanding of human information processing have we gained through the use of psychophysiological measures? Measures of P300 latency have helped to describe the time course of two aspects of stimulus evaluation -- feature and location analysis. Measures of the EMG (and of squeeze responses) have provided evidence for a response competition process and implicated this process as being partially responsible for the noise-compatibility effect. Measures of lateral scalp negativity have enabled us to look at priming processes and to trace the course of response activation during the foreperiod and after array presentation. Such insights would have been difficult, if not impossible, to obtain from measures of the overt behavior alone.

IMPLICATIONS FOR THE STUDY OF THE STATE/PROCESS RELATIONSHIP

The psychophysiological approach exemplified in preceding sections, and the data themselves, have implications for the study of the state/process relationship. We will base our discussion on the distinction between states and processes described in the Introduction, and we will also refer to the position presented by Wickens (this volume).

We propose that the elements of the system can be analyzed in terms of the processes they perform and of the states affecting the parameters of their operation. If we could describe the processing activity of an element in terms of an "activation x time" function, then this function could vary in its initial level (i.e., the baseline) and in its rate of change (i.e., the gain). Thus, the parameters affected by state variables include the baseline level of activation, as well as the gain of an element (see Wickens, this volume).

To illustrate these properties of processing elements, we will first consider the process of response channel activation. As we have shown, the initial level, or <u>state</u>, of activation of the response channels can vary on a trial to trial basis (see Figure 5). These modulations in the initial level of response channel activation lead to variations in the amount of further activation (based on evidence accumulation) required for the emission of the overt response. We labelled these variations in the initial state, response "priming" or "bias." We could conceive of situations involving stress, sleep deprivation, noise, etc., in which this initial level would be modulated in a more generalized and enduring fashion. For example, the effects of noise may be to increment initial levels of response activation in both response channels. While generalized, long-term state changes were not considered in our experiments, its seems possible that these changes share a common *modus operandi* with the short-term, specific changes in initial levels that we observed.

Scrutiny of Figure 5 reveals that the rate of change of activation that occurs prior to a motor response is apparently constant for different levels of priming and for different response latencies. This implies that the gain of response activation systems is not correlated with the initial level of activation, at least in our studies. In this sense, there is no evidence for gain modulation in the response activation system. However, gain modulation is apparent in the stimulus evaluation system. P300 latency varies from trial to trial, and we use this measure as an index of the speed of evaluation processes. By computing the RT/P300 latency ratio on each trial, we are, in essence, using P300 latency to adjust the overt response latency on that trial for the rate of evidence accumulation (gain). Our analysis further implies that both feature and location analysis share a common gain, and that trial to trial fluctuations in this gain can be indexed by P300 latency.

We did not manipulate state variables which might affect gain in an enduring, long-term fashion. Nevertheless, it seems reasonable to propose that the operation of such variables on the gain of the evaluation process can be detected by using measures of P300 latency. In particular, since evaluation processes control, in part, the activation of response channels, we might find evidence for gain variations by looking at the response activation functions (of the type shown in Figure 5) for trials with different P300 latencies. P300 latency and, by implication, the gain function may be affected by the level of alertness or pharmacological manipulations.

So far, we have considered evidence for two types of state/process interaction, one involving the initial level of activation of the response channels, the other involving the gain of the evaluation process. There is another form of state/process interaction involving the relationship between "response-oriented" and "stimulus-oriented" priming . In the context of our experiment, we have defined priming as the initial level of activation of the response channels. However, the literature on P300 has linked the amplitude of this component to the disconfirmation of a set of stimulus expectancies that are generated by the task situation (Donchin, 1981). How do these two forms of priming, response-oriented and stimulus-oriented, interact? What we find is that the level of response-oriented priming (as revealed by scalp asymmetries) is not related to stimulus-oriented priming (as revealed by P300 amplitude). We reasoned that if response priming was related to the development of an expectancy for the stimulus associated with the primed response, then the amplitude of the P300 to the array should be larger when the target stimulus was that associated with the "un-primed" response. In fact, P300 amplitude did not differentiate between stimuli associated with the "primed" and those associated with the "un-primed" responses. This suggests that response-oriented priming is not related to stimulus oriented priming.

The preceding discussion suggests two conclusions. First, the interaction between states and processes may take several forms. Second, the psychophysiological approach may provide important information about this interaction. Many questions remain to answered. However, in so far as the psychophysiological approach provides the wherewithal to disentangle the multiplicity of influences on human information processing, we have the ability to explore these questions.

FOOTNOTES

1. In all our speed accuracy trade-off analyses, we used the latency of muscle activity, rather than the squeeze response as our measure of reaction time. This is because the muscle response is less subject to the response competition processes to be described below and is a more sensitive measure of response initiation. If both responses occurred, response accuracy was defined in terms of which response occurred first.

REFERENCES

Coles, M. G. H., & Gratton, G. (In press). Psychophysiology and contemporary models of human information processing. In D. Papakostopoulos & I. Martin (Eds.), Clinical and Experimental Neuropsychophysiology. Beckenham, England: Croom Helm.

Coles, M. G. H., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. Journal of Experimental Psychology: Human Perception and Performance, 11, 529-553.

Coles, M. G. H., Gratton, G., & Donchin, E. (1984, September). Flies in the ointment. Paper presented at the IIIrd International Conference on Cognitive Neuroscience, Bristol, England.

Donchin, E. (1979). Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), Evoked potentials and behavior. New York: Plenum Press, pp. 13-75.

Donchin, E. (1981). Surprise! . . . Surprise? Psychophysiology, 18, 493-513.

Donchin, E., Coles, M. G. H., & Gratton, G. (1984). Cognitive psychophysiology and preparatory processes: A case study. In S. N. Kornblum and J. Requin (Eds.), Preparatory States and Processes. Hillsdale, NJ: Erlbaum Associates, pp. 155-178.

Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. Perception & Psychophysics, 25, 249-263.

Gratton, G., Coles, M. G. H., Sirevaag, E., Eriksen, C. W., & Donchin, E. (Manuscript in preparation).

Grice, G. R., Nullmeyer, R., & Spiker, V. A. (1982) Human reaction times: Toward a general theory. Journal of Experimental Psychology: General, 111, 135-153.

Kutas, M., & Donchin, E. (1977). The effect of handedness, the responding hand, and response force on the contralateral dominance of the readiness potential. In J. Desmedt (Ed.), Attention, voluntary contraction and event-related cerebral potentials. Progress in Clinical Neurophysiology, Vol 1, Basel: Karger, pp. 189-210.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. Psychological Review, 86, 287-330.

Meyer, D. E., Yantis, S., Osman, A. M., & Smith, J. E. K. (1985). Temporal properties of human information processing: Tests of discrete versus continuous models. Cognitive Psychology.

Meyer, D. E., & Irwin, D. E. (1982). On the time course of rapid information processing (Tech. Rep. No. 43). Ann Arbor: University of Michigan, Cognitive Science Program.

Miller, J. (1982). Discrete versus continuous stage models of human information processing: In search of partial output. *Journal of Experimental* Psychology: Human Perception and Performance, 8, 273-296.

Rohrbaugh, J. W., Syndulko, K., & Lindsley, D. B. (1976). Brain components of the contingent negative variation in humans. Science, 191, 1055-1057.

Rohrbaugh, J. W., & Gaillard, A. W. K. (1983). Sensory and motor aspects of the contingent negative variation. In A. W. K. Gaillard & W. Ritter (Eds.), Tutorials in event-related potential research: Endogenous components, pp. 269-310. Amsterdam: North-Holland.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. Cognitive Psycyhology, 12, 97-136.

Figure 1

Hypothetical activation functions for correct and incorrect response channels. The effects of response priming (bias) and response competition are illustrated in the function for the correct channel.

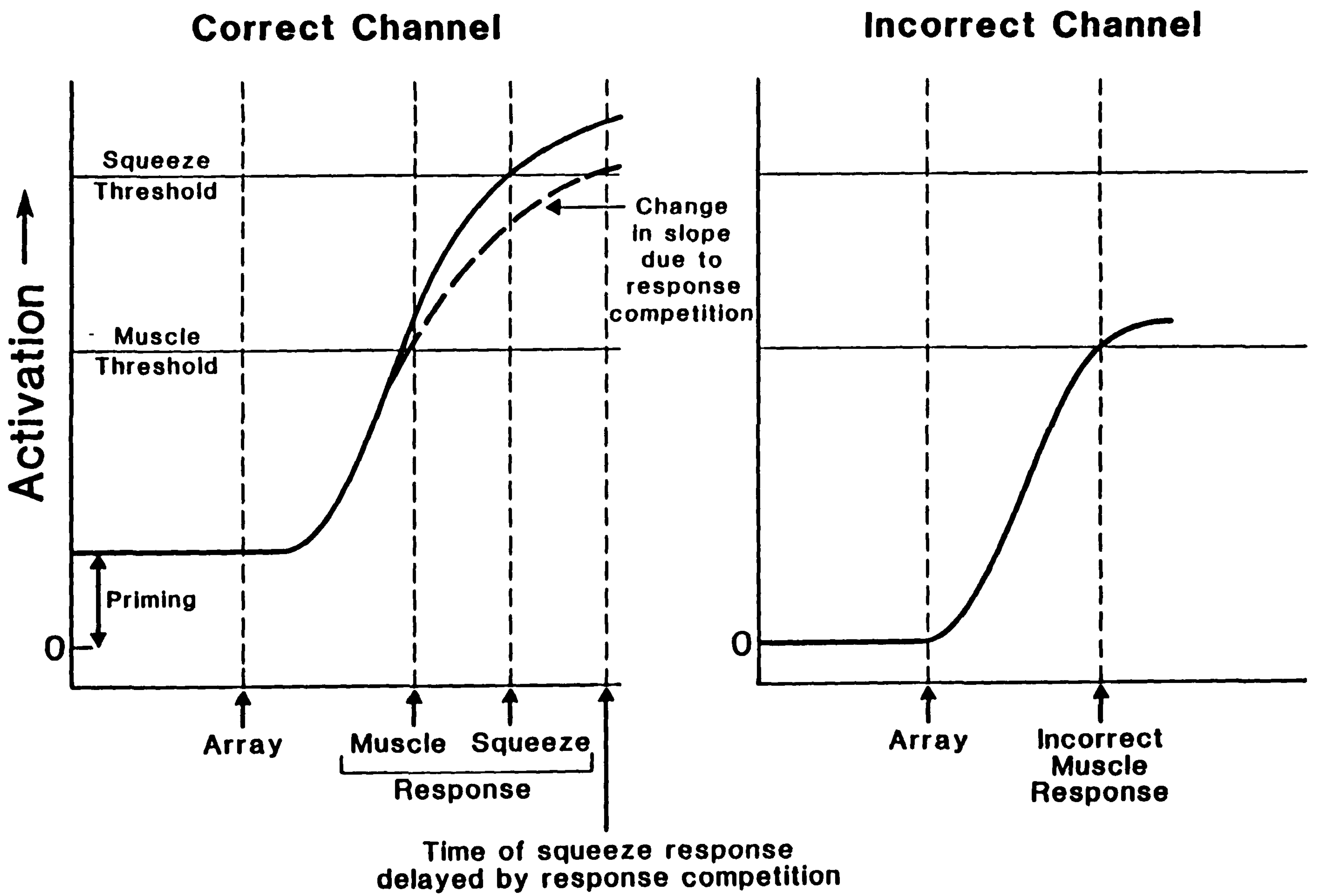

Correct Channel
Incorrect Channel
Activation
Squeeze Threshold
Muscle Threshold
Priming
0
Change in slope due to response competition
Array
Muscle
Squeeze
Response
Incorrect Muscle Response
Time of squeeze response delayed by response competition

Figure 2

Speed-accuracy trade-off functions for compatible and incompatible noise conditions without (2a) and with (2b) adjustment for P300 latency.

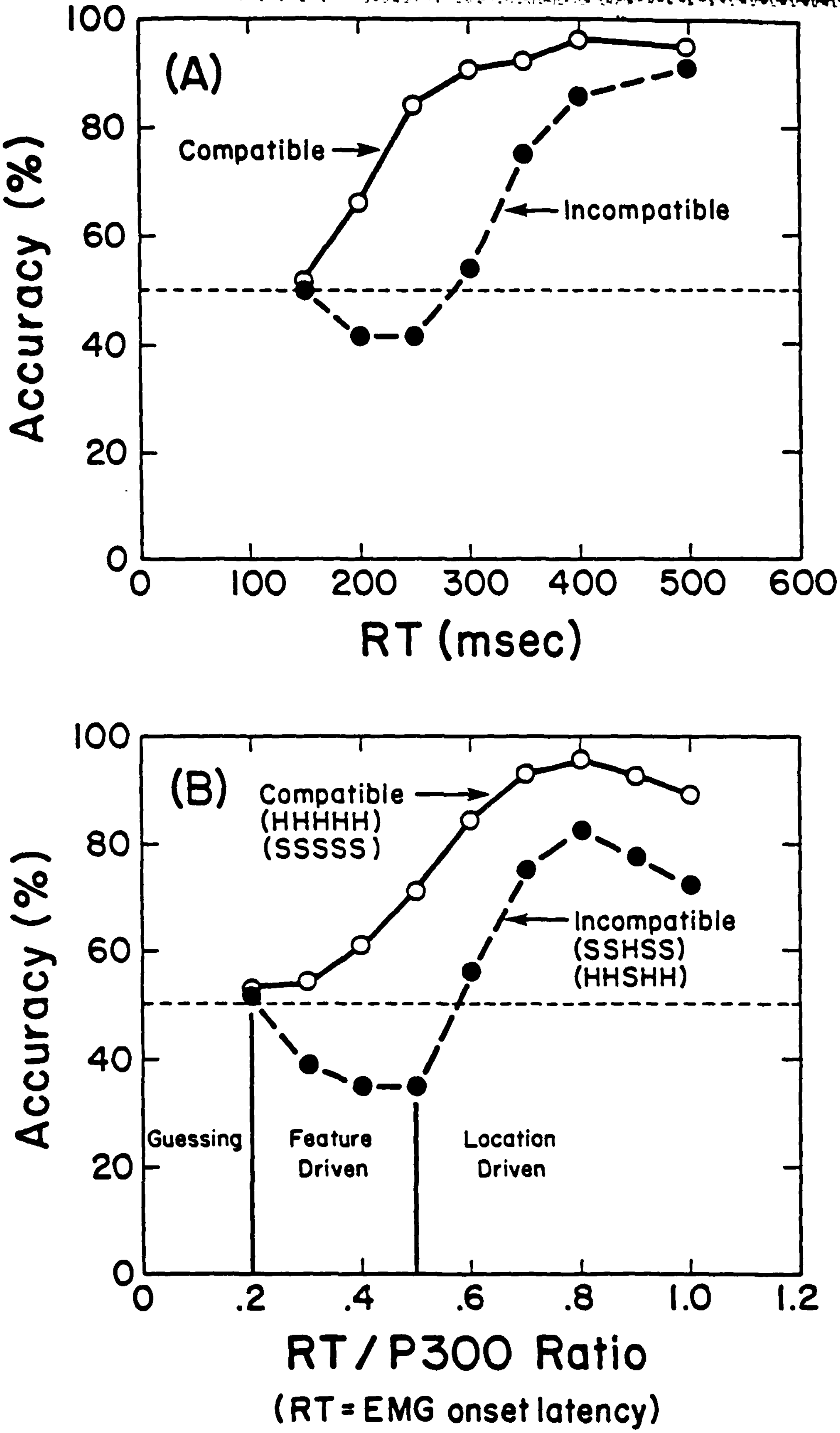
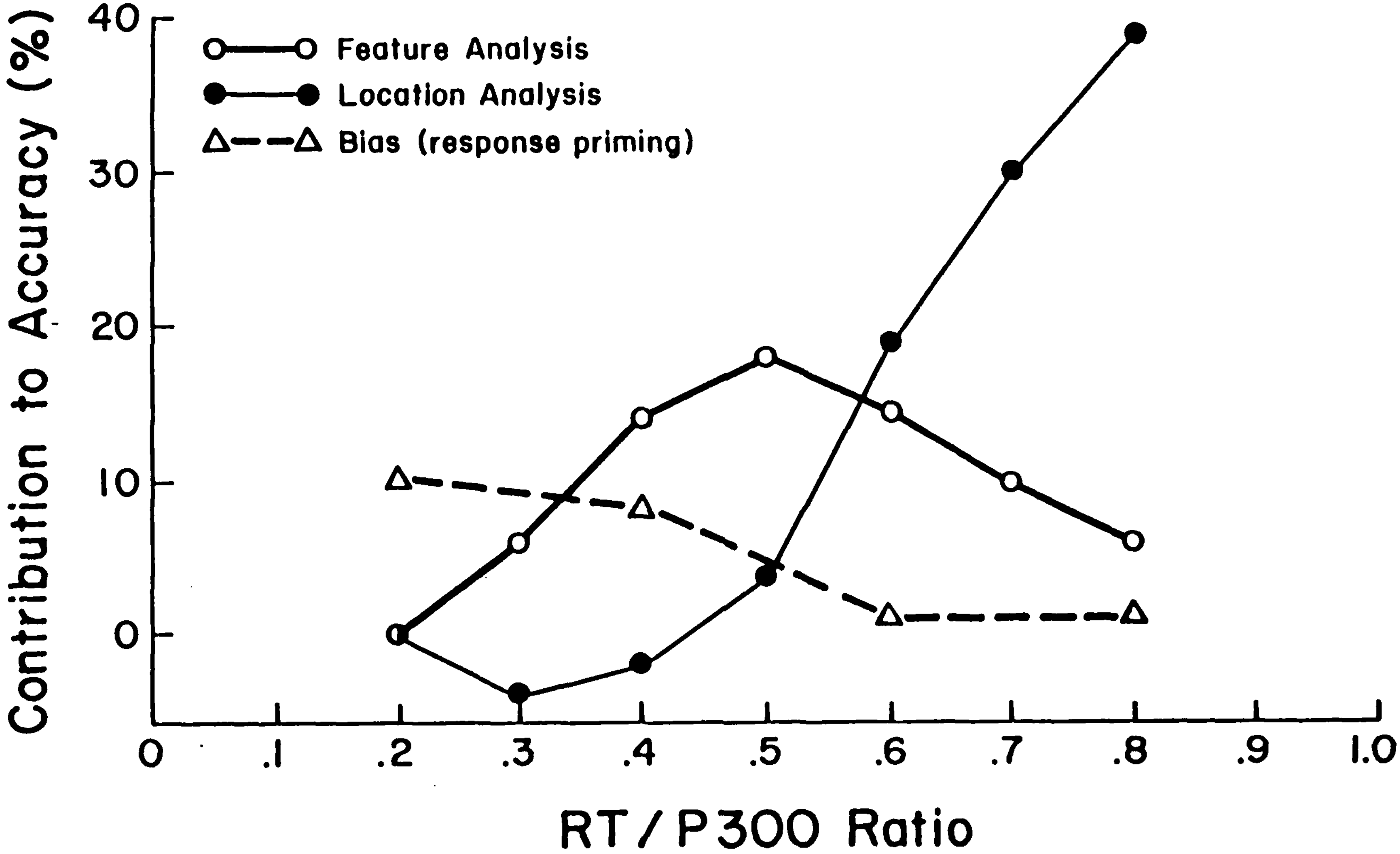

(A)
Compatible
Incompatible
Accuracy (%)
RT (msec)
(B)
Compatible
(HHHHH)
(SSSSS)
Incompatible
(SSHSS)
(HHSHH)
Guessing
Feature Driven
Location Driven
Accuracy (%)
RT / P300 Ratio
(RT = EMG onset latency)

Figure 3

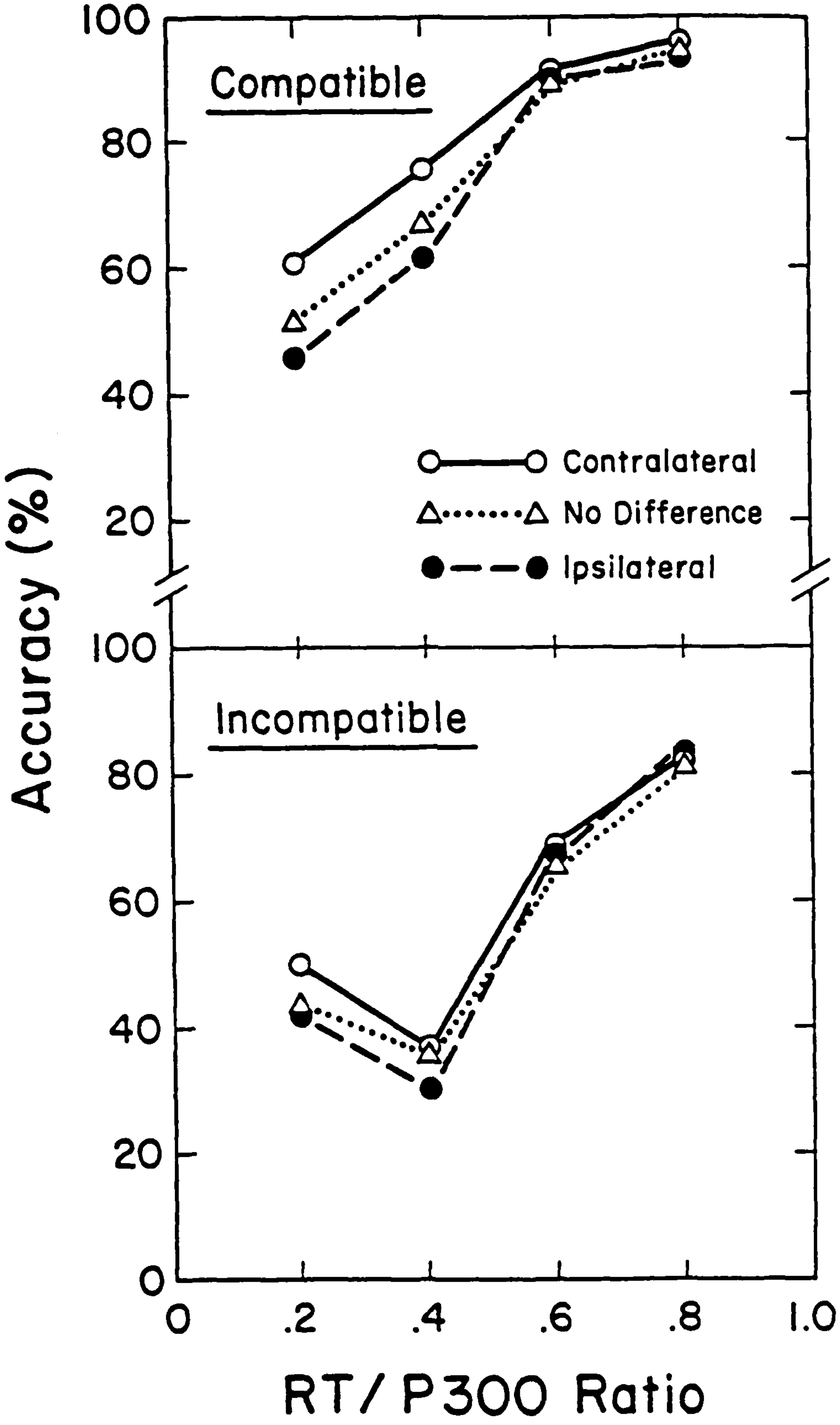
The contributions of response priming (bias), feature analysis, and location analysis to response accuracy as a function of relative response latency.



Contribution to Accuracy (%)
40
30
20
10
0
Feature Analysis
Location Analysis
Bias (response priming)
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
RT / P300 Ratio

Figure 4

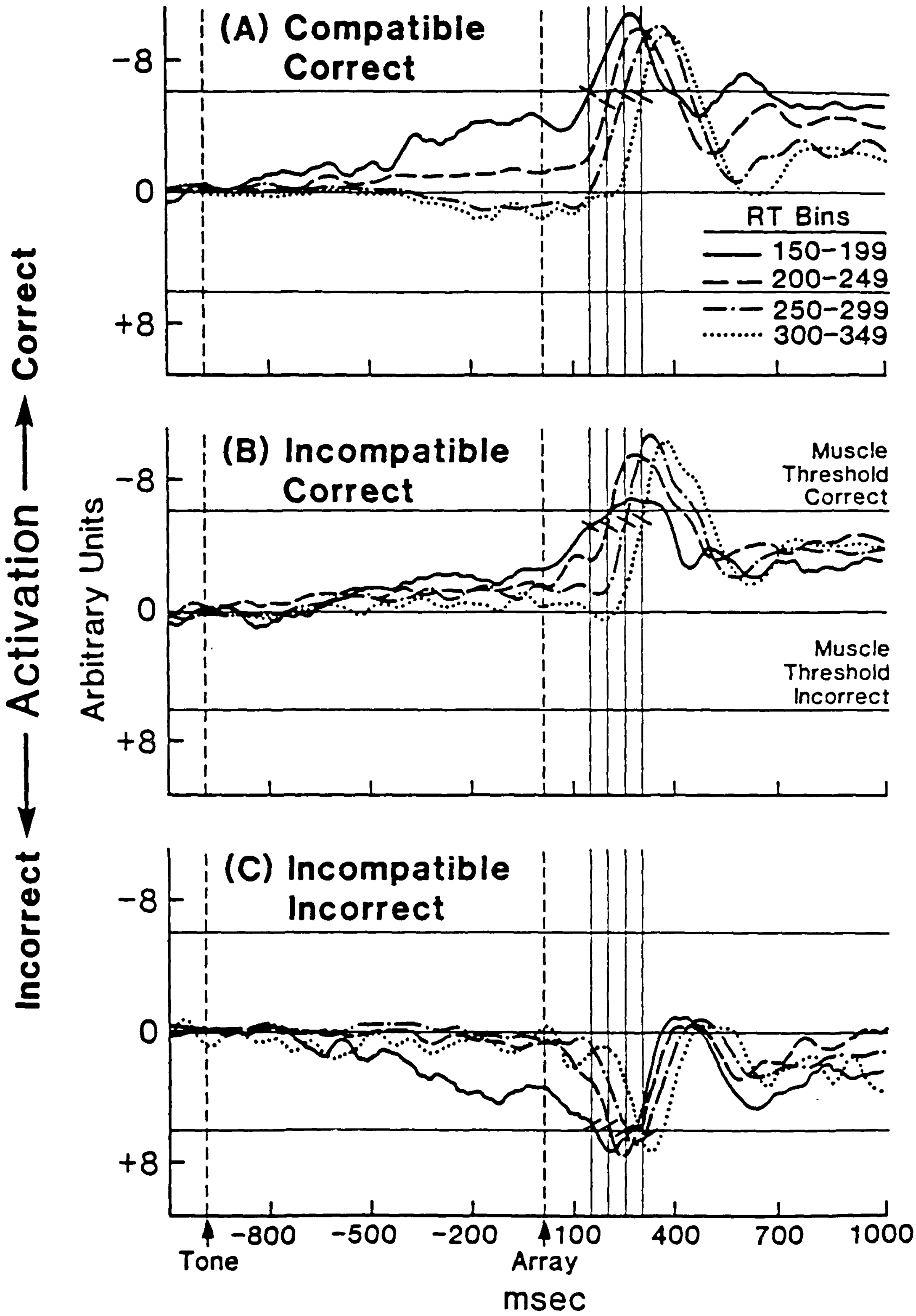
Speed-accuracy trade-off functions (adjusted for P300 latency) for trials with different types of response priming. Contralateral -- correct response primed: Ipsilateral -- incorrect response primed: No difference -- neither response preferentially primed.



Compatible
Contralateral
No Difference
Ipsilateral
Incompatible
Accuracy (%)
100
80
60
40
20
0
0
.2
.4
.6
.8
1.0
RT/ P300 Ratio

Figure 5

Values representing the asymmetry in scalp negativity over the course of a trial for correct compatible trials and for correct and incorrect incompatible trials. The four traces within each panel are derived for trials with different response latencies (defined by EMG). The response latencies are represented by the vertical lines within each panel. The horizontal lines (threshold) indicate the average value of the asymmetry at the time a response is emitted.



(A) Compatible Correct
(B) Incompatible Correct
(C) Incompatible Incorrect
RT Bins
150-199
200-249
250-299
300-349
Muscle Threshold Correct
Muscle Threshold Incorrect
-8
0
+8
Incorrect ← Activation → Correct
Arbitrary Units
-800
-500
-200
100
400
700
1000
Tone
Array
msec

After a Rash Action: Latency and Amplitude of the P300 Following Fast Guesses

by

Emanuel Donchin, Gabriele Gratton, David Dupree,
and Michael Coles

Cognitive Psychophysiology Laboratory
Department of Psychology
University of Illinois

Running title: <u>P300 and fast guesses</u>

Acknowledgements

The study presented in this paper was supported in part by a grant from the Air Force Office of Scientific Research, contract #F49620-83-0144, Al Fregly Program Director. The comments of Monica Fabiani and Art Kramer on early versions of this paper are greatly appreciated.

To appear in: G. Galbraith, M. Klietzman, & E. Donchin (Eds.), <u>Neurophysiology and Psychophysiology: Experimental and Clinical Applications</u>. Hillsdale, NJ: Erlbaum.

Introduction

A persistent theme in Lindsley's writings has been the focus on the temporal characteristics of psychophysiological signals. His second published paper (Herren & Lindsley, 1931) was concerned with the latencies of tendon reflexes. This interest in latencies continually appears in subsequent studies (e.g., Chalupa, Rohrbaugh, Gould, & Lindsley, 1974; Donchin & Lindsley, 1965; Lindsley, 1944; Lindsley, 1954; Lindsley, 1982; Lindsley & Emmons, 1958; Lindsley, Fehmi, & Adkins, 1967; Lindsley & Rubenstein, 1937; Lindsley, Seales, & Wilson, 1973; Schwartz & Lindsley, 1964). This chronometric approach, that has played a critical role in many areas of Psychology (Posner, 1978), makes the seemingly paradoxical assumption that even if we do not know what happened it is useful to know when it happened. The relative timing of events, the rhythm with which they occur and the factors that increase, or decrease, the speed with which the events are triggered, can be useful in the analysis of a system even if we are not quite ready to provide a full account of the processes that underlie the events. It is clear, for example, that when Shakespeare, in the Sixteenth Century, has Leontes in The Winter's Tale (Act I sc. 2) say: "I have a tremor cordis on me: my heart dances; But not for joy; not joy...." he is expressing a recognition of a common relationship between an accelerated heart rate and "joy." It is also evident that Shakespeare is aware that even though the dancing heart is commonly an indicator of "joy" this relationship can break down. The accuracy and utility of this "psychophysiological" implication of Shakespeare's does not depend on his knowing, as he obviously does not, the physiology of the heart and of the control of the heart's rhythms. The validity of such psychophysiological observations depends on the careful observation of the temporal characteristics of the physiological events ("my heart dances"), on the proper definition of the psychological concepts ("not joy"), and on the theory within whose framework these relationships play a role.

In his work on the Excitability Cycles, in studies of backward masking and in the examination of the various rhythms of the EEG, with particular emphasis on the alpha rhythm, Lindsley has shown how it is possible to use the timing of psychophysiological signals in the study of the mind even if the nature of the processes observed by the psychophysiologist are not fully, or even partially, understood. The point is simple. A well defined psychophysiological response is an event. Once we have identified the occurrence of an event, we are in possession of information about its time of occurrence, even if we do not know its causes or its nature. These temporal data can serve as dependent variables in our studies. If our theories generate differential predictions about the variation of these temporal variables as a function of properly selected independent variables, these data can play a critical role in the understanding of the mind.

This point of view is illustrated by the widespread use of the latency of components of the Event-Related Brain Potentials (ERP). Thus, for example, Hillyard and his associates demonstrated that a differentiation between the ERPs elicited by attended and ignored stimuli appears as early as the first 100 ms following the eliciting stimuli (Hillyard, 1984). The observation of Hillyard's that makes a difference for a theory of Attention is the observation about the timing of the ERP component, and this theoretical point is largely unaffected by debates about the precise nature of the ERP components observed (Naatanen, 1982). Another ERP component whose latency has proven a useful tool in the analysis of cognitive function is P300. In this chapter we illustrate the richness of the information the latency of the P300 yields regarding processes that are essentially opaque to the more traditional tools of cognitive psychology. We will, for the purposes of this illustration, provide a partial description of a study that will be reported in full elsewhere (Gratton, Coles, Dupree and Donchin, 1985). The analysis of P300 timing will be made, in this chapter, from a perspective of a theory of P300 that makes specific predictions regarding the consequences of changes in an internal process, changes that are manifested at the scalp by changes in the amplitude of the P300 (Donchin, 1981). As the predictions are confirmed by the data, the study lends support to our interpretation of the P300.

The study we discuss is one in a series employing the "Oddball" paradigm in which the stimuli are names of individuals commonly used in the American culture. In all cases the series are constructed so that 20% of the names were names of males (e.g., David, Henry, Thomas...). The other names used in the series are commonly associated with females (e.g., Nancy, Helen, Susan...). On some occasions, the subject is required to count the number of names that fell in one or another category (a count condition). On other occasions the subject indicates the occurrence of one of the categories by pressing one of two buttons (a Reaction Time, or RT, condition).

The initial study in this series was reported by Kutas, McCarthy and Donchin (1977). Their subjects were presented with 3 different Oddball series. A "Variable Names" series was constructed from names of males and females as described in the previous paragraph. A "Fixed Names" series included just the names "David" and "Nancy." The third series was a sequence of words, 20% of which were synonyms of "prod." In the latter case, the subject's task was to press one button in response to such synonyms and to press another button in response to all other words. The rare events in each series elicited a large P300. This was true regardless of the specific task assigned to the subject.

It turned out that the latency of the P300 varied across the three conditions. This was particularly evident when the subjects were instructed to be accurate. The shortest latency was observed when the subject discriminated between the two names, David and Nancy. A longer latency was seen when the names varied from trial to trial. The longest latency was associated with the need to decide whether each of a rather disparate list of words is a synonym of "prod." These, and a considerable amount of additional data, lead us to suggest that the latency of the P300 depends on the time required for the evaluation of the stimulus and is independent of response selection. Subsequent work (McCarthy & Donchin, 1981) provided strong support to the assertion that the latency of P300 is largely independent of the duration of processes that are involved in the execution of the response. The interesting conclusion from the data reported by Kutas et al. (1977) has been that the latency of P300 is proportional to the time it takes to categorize the stimuli. If this is the case, the P300 latency may be used as a tool in mental chronometry to measure mental timing uncontaminated by "motor" processes (Coles, Gratton, Bashore, Eriksen, & Donchin, 1985; Donchin, 1981; Magliero, Bashore, Coles, & Donchin, 1984; McCarthy & Donchin, 1981). For studies in which P300 latency is indeed utilized in this fashion see Duncan-Johnson and Donchin (1982), Pfefferbaum, Ford, Roth, and Kopell (1980), Goodin, Squires, and Starr (1978), and Pfefferbaum, Ford, Roth, and Kopell (1980).

Kutas et al. (1977) examined the relationship between the latency of P300 and the RT associated with each of the trials in an oddball study using names, sorted according to gender (see also McCarthy, Kutas, & Donchin, 1979). Their analysis capitalized on a filtering technique that allowed the measurement of the latency of P300 on individual trials (Woody, 1967). The principal finding was that the correlation between P300 latency and RT depends on the strategy adopted by the subjects. When the subjects were instructed to be accurate the correlation between P300 latency and RT was larger (.61) than it was when they were instructed to be fast (.48). These data support the suggestion (Donchin, 1979; 1981) that the P300 and the motor response may each be the product of a series of processing activities and that these streams of processing can, in principle, be quite independent of each other.

Since the invocation of P300 is dependent on the evaluation of information conveyed by the stimulus, the latency of P300 must be at least as long as the duration of these evaluative processes. The overt responses, on the other hand, may well be released "prematurely" on the basis of limited information. The correlation between RT and the latency of the P300 will therefore depend on the degree to which the overt responses that define the RT are made contingent on the evaluation of the stimulus. The more inclined is the subject to respond prematurely, the poorer the correlation between the latency of the P300 and the RT.

One striking aspect of the data acquired by Kutas et al. (1977) was observed in the trials in which subjects made errors. These were mostly trials on which the subject responded to a rare event as if it were a frequent event. That is, even though a Male name appeared on the screen, the subject pressed the button associated with Female names. There were but a few such trials in the study reported by Kutas et al. (1977). However, in virtually all these trials, the pattern was the same - the RTs were (relatively) short and the P300 latency was (relatively) long. It was as if on these trials the subjects first acted and than thought! A partial report on these data can be seen in McCarthy (1984). In 10 of the 11 subjects the pattern obtained was identical. The incorrect responses were associated with very short RTs and relatively long P300 latencies. One of several possible interpretations of these data is that, when the information processing system detects an error, the invocation of the P300 is delayed. According to this suggestion, the delay is required to allow further processing of the trial's data. While this interpretation is consistent with the data, it is not the only possible account for the increased latency of the P300 on error trials. Several other possible mechanisms need be considered. Another interpretation of the data is based on the fact that on all the error trials the subject responded rather fast to the stimulus. In other w 's, these are clearly trials on which a variety of factors are injected into the stream of processing. How do we know that it is the recognition of the error, rather than the fact that a very fast response was emitted on the trial, that accounts for the delay? A different, but related possibility is that it is not that P300 is delayed on error trials, but rather that errors may be more likely on trials on which P300 latency is long. Finally, it can be suggested (see McCarthy, 1984) that the positive peak that is observed on the error trials is not a "P300," at least not that elicited by the names, but is rather a different component, perhaps even a different "P300," which may be elicited by the recognition of the error.

The last interpretation raises the issues of how we define "P300." There is no doubt that one of the major difficulties presented by ERP data is associated with the definition and the proper identification of components of the ERP. As we noted above, if our strategy depends on the measurement of the timing of events we must be sure that the timing we measure on different occasions is indeed the timing of the same event. The approach calls for considerable care in the definition of components if the features of the waveform we measure can be affected, as they no doubt are, by different components (Donchin & Heffley, 1978; Fabiani, Gratton, Karis, & Donchin, in press). Consider, for example, each of the positive going peaks observed by Kutas et al. (1977). Each of these positive peaks has been labeled "P300" even though the peaks differ in latency by as much as 100 ms. What leads us to believe that these three peaks are indeed instances of a component whose latency is shifted by the duration of the processing preceding its invocation? How do we know that the peaks with the longer latencies are not entirely new components that are elicited by the presentation of a word, or by the search for a synonym? The issue is generally resolved on the basis of the similarity of waveshapes, of the scalp distribution of the potentials and of the manner in which they respond to experimental manipulations (Donchin, Ritter, & McCallum, 1978; Sutton & Ruchkin, 1984).

To resolve some of the doubts that remained regarding the ERPs elicited on error trials we replicated, and extended, the study reported by Kutas et al. (1977). While in structure the study reported here followed closely the "variable name" phase of the Kutas et al. (1977) study, we expanded the design in a number of ways. To assure that the number of error trials was sufficient to allow the needed comparisons, the sessions were greatly extended and we recorded 800 trials in each of the experimental conditions. As before, each subject performed the task under an "accuracy" and under a "speed" regime. Furthermore, in order to determine the extent to which the observations depend on the imbalance between the probabilities with which the two categories of names appear in the series, the Speed and the Accuracy conditions were run twice. In one case the Male names were rare (P=.20) while in the other case Male and Female names appeared with equal probability. The equal probability condition (which was not present in the Kutas et al. study) allowed us also to assess the parameters (RT, accuracy, P300 latency, etc.) of fast guesses, where there was no particular advantage for either response.

Method

Seven right-handed male students at the University of Illinois were paid $3.50 an hour for their participation in the study. The subject was positioned in front of a PLATO terminal with the fingers of each hand resting on a zero displacement dynamometer. Male and female names were presented on the screen for 200 msec every 2000 msec, and the subject was required to squeeze one of the dynamometers following the presentation of a name.

Subjects were shown the names in blocks of 100 trials. Blocks were composed of either 80 female and 20 male names or 50 of each. In different trial blocks, the subjects were instructed either to respond as quickly as possible ("speed instructions"), or as quickly as possible but without making errors ("accuracy instructions"). Eight-hundred trials were run for each of the probability x instruction conditions, with half the trials run during one session and the remaining half run during a second session.

The EEG was recorded using Burden electrodes at Fz, Cz, Pz, placed according to the International 10-20 system (Jasper, 1958) and referred to linked mastoids. The signals were amplified and filtered on-line (8 sec time constant, and 35 Hz upper half amplitude cutoff point). EOG was recorded for purposes of subtracting out ocular artifact from EEG. The subtraction of the ocular artifact was accomplished by means of a procedure described by Gratton, Coles, and Donchin (1983). The EEG and EOG signals were digitized at 100 Hz for a period of 1400 msec, starting 100 msec before each stimulus presentation.

Average ERP waveforms were computed for each instruction, probability, stimulus category, subject and electrode. P300 latency and amplitude was assessed on each single trial according to a procedure described in Gratton et al. (1985).

Results

The results will be divided into several sections. First, we will present data supporting the claim that the stimulus probability manipulation did indeed affect the subjects' response strategy. Second, we will describe the relationships between P300 latency and RTs, and between these two variables and response accuracy. These data replicate the results obtained by Kutas et al. (1977). Then, we will describe a procedure devised to interpret these relationships. Finally, we will analyze some of the consequences of the processes involved in P300 generation. Note that the present chapter is but a partial report of the stuay. It is intended to illustrate the chronometric use of psychophysiological signals rather than to serve as a comprehensive report of the study. Therefore, we shall ignore, in this discussion, the data obtained in the accuracy instruction condition. Furthermore, we shall ignore many of the detailed analyses of the data that are required to fully support our interpretations. For a full description of the study see Gratton et al. (1985). Note also that some of the analyses were based on five subjects only, because of the small number of error trials in the frequent female condition.

Effects of Manipulations on Response Strategy

The frequency and latency of correct and incorrect overt responses for each probability x stimulus x response condition are shown in figure 1.

Insert Figure 1 About Here

The subjects' overt response was affected by the probability of the stimulus. In fact the response to the rare male stimuli was less accurate than that to the frequent female stimuli (under speed instructions), as revealed by the instruction x stimulus x response interaction, $F(1,4)=9.22$, $p<.05$. The latency of the correct response for male stimuli was slower than that for female stimuli, while the latency of the incorrect response was faster, as revealed by the stimulus x response interaction, $F(1,4)=11.69$, $p<.05$. This was particularly evident for the 20/80 condition, $F(1,4)=91.68$, $p<.001$. In particular, in the 20/80 condition, for male stimuli (rare) the incorrect response was faster than the correct response by 134 ms, while for female stimuli (frequent) the correct response was 50 ms faster than the incorrect response.

These findings support the conclusion that the subjects indeed conformed their response strategy to the probability manipulation. In particular, when the female stimulus was presented more often (the 20/80 condition) the subjects tended to execute fast female responses whatever stimulus was presented. In fact, the error rate for male stimuli under these conditions is 64%, while the error rate for female stimuli is only 5%.

<u>P300</u>, <u>RT</u>, <u>and</u> <u>Accuracy</u>

The grand average waveforms at Pz, for each probability, stimulus, and response condition are shown in Figure 2.

Insert Figure 2 About Here

Inspection of this figure reveals several interesting points. A large positivity is the most dominant feature of the waveforms. We interpret this positivity as P300. The latencies of the P300 peaks (shown in Figure 2) for trials where a correct response was given were 60 ms shorter than for trials where an incorrect response was given, $F(1,4)=24.87$, $p<.01$. The category of the stimulus did not affect P300 latency, $F(1,4)=0.66$, $p>.05$, nor did the probability manipulation, $F(1,4)=1.31$, $p>.05$, or the stimulus x probability interaction, $F(1,4)=0.00$, $p<.05$. Thus, the time the subject takes to emit a P300 does not depend on whether the stimulus is male or female, or, in fact, on whether the probability of the stimulus is manipulated. These results contrast with those obtained for the RT. They indicate that the timing of those processes on which the emission of the P300 depends is not affected by the variations in the criteria for overt response emission, which were introduced by the experimental manipulations. However, the amplitude of the P300 (see Figure 2) was affected by some of these variables. In particular, the male stimuli, when rare, elicited a larger P300 than the female stimuli (in this case, frequent). This produced a significant stimulus x probability interaction, $F(1,4)= 15.34$, $p<.05$. On the other hand, P300 for incorrect responses was only slightly (and not significantly) larger than for correct responses, $F(1,4)=2.99$, $p>.05$.

Summarizing these findings, we note that we have replicated and extended the Kutas et al. (1977) study. Error trials are generally associated with faster RTs, but later P300s than correct trials. The stimulus probability affects RT but not P300 latency.

Error Recognition

The results presented above are consistent with the interpretation that P300 latency is influenced by variables affecting the time required to evaluate the stimulus, but relatively independent of the variations of the response criteria adopted by the subject. However, the observation that the overt response is fast and P300 late on error trials require some explanation. In fact, two explanations are possible. First, it may be that, in analyzing the error trials, we select those trials in which fast guessing and/or delay in the evaluation of the stimulus occur. Thus, errors may occur because stimulus evaluation (i.e., P300) is late in comparison with response activation processes. Second, it may be that processing of the error may delay P300. In this case, the latency P300 would not only reflect the processing of the external stimulus (male or female name), but also the processing of the internal events leading to the response.

To choose between these two interpretations, we focussed our attention on a condition in which errors are particularly frequent (the "speed" condition). Our procedure was based on an analysis of the speed/accuracy functions for this condition. In addition, we were interested in distinguishing among trials with different P300 latency. Speed/accuracy functions for different P300 latency bins are presented in Figure 3, for 20/80 and 50/50 conditions separately.

Insert Figure 3 About Here

For the 20/80 conditions, the functions were computed separately for male and female names. For the 50/50 condition, data from the two name categories were pooled together, since they had the same probability. These functions reveal several interesting points. First, the longer the RT the higher the accuracy. Second, accuracy is higher for trials on which P300 latency is relatively short. These findings suggest that accuracy is largely dependent on the relative timing of P300 latency and RT.

We also note the very low accuracy when the rare male names are presented. This is especially true when RTs are fast and P300 latency is long. In this case, we might speculate that the subject's basic strategy is to emit the "female" response to the "male" stimulus. In fact, virtually no errors can be observed in response to frequent female names. It may be possible to suggest that this pattern of results is due to the fact that errors occur when stimulus evaluation time is for some reason slow so that the male stimulus is not processed fast enough to inhibit the female response. If this explanation is valid then the delay in P300 on error trials is not due to the processing of the consequences of the error.

However, a third observation is not compatible with the interpretation that P300 is solely dependent on the time required to decide whether the stimulus was male or female. In fact, if P300 is sensitive only to the stimulus categorization process, and a delay in P300 indicates only a delay in this categorization process, then responses given before this process is sufficiently established should have a chance level of being correct. In the 50/50 condition this chance level is .50. Thus, this interpretation should predict that, in the speed 50/50 condition, the error rate would never exceed .5 even in cases of long P300 latency and short RTs. Actually, figure 3 reveals that, in this condition, the accuracy for trials with fast RTs and slow P300 latency is lower than the chance level. This indicates that, by looking at trials with fast RTs and long P300 latency, we are "selecting" error trials. We interpret this finding as demonstrating that the association between incorrect response and long P300 latency is not due solely to the fact that errors occur because of a delay in stimulus evaluation. We must also propose that the processing of the incorrect response causes a delay in P300.

The P300 and Future Action

We have demonstrated that P300 is delayed on incorrect trials. This delay indicates that, before emitting the P300, the subject must have not only categorized the name, but also compared the stimulus category with the current response. Presumably, the delay in P300 reflects some process that occurs when the system processes the commission of an error (a recognition that need not reach the subject's awareness). Given the relationship between P300 and schema updating (Donchin, 1981; Karis, Fabiani, and Donchin, 1984), we hypothesized that in the present experiment this process was related to adjustments in the subject's strategy subsequent to the recognition of an error. If this hypothesis is correct then the characteristics of the P300 elicited on the error trials should predict variations in the response criteria in the following trials.

To test this hypothesis we used the following procedure. First, we identified the male trials in which an incorrect response was given. Then, we sorted these trials on the basis of the response made on the following trial on which a male name was presented. The rationale was the following. We assumed that when an incorrect response to a male trial was given, the subject was biased to emit the "female" response regardless of the stimulus. If the subject responded again incorrectly to the following male trial, then we assumed that the subject's bias remained the same. On the other hand, if the subject responded correctly to the following male trial, then we assumed that the subject had revised his strategy. We labelled the latter sequences, "switch" sequences, and the former, "no-switch" sequences. Note that we assumed that the switch in response bias occurred as a consequence of the recognition of an error after the first trial of a sequence. In particular, we predicted that the P300 to the first stimulus of the sequence (incorrect male name) was larger for "switch" than "no-switch" sequences. For both the 20/80 and 50/50 conditions P300 elicited on the first trial of a "switch" sequence was larger than P300 elicited on the first trial of a "no-switch" sequence, $F(1,5)=6.66$, $p<.05$, and $F(1,5)=14.25$, $p<.05$, respectively. To test further the hypothesis that the switch in response bias does indeed occur immediately following the recognition of an incorrect response to a male trial, we examined the female trials which intervened between the first and last male trials of the sequence described above. The prediction was that the response to the intervening female trial should be slower for "switch" than "no-switch" sequences. The RT for these trials are shown in Figure 4, as a function of sequence (switch vs no-switch), lag from the first trial of the sequence, and condition (20/80 vs 50/50).

Insert Figure 4 About Here

Accuracy is higher and RTs are slower for female trials in a "switch" sequence than in a "no-switch" sequence, $F(1,6)=13.36$, $p<.05$, $F(1,5)=7.50$, $p<.05$, respectively. This indicates that the subjects did indeed modify their response strategy at the beginning of a "switch" sequence.

Discussion

The data presented above indicate that when a subject chooses the wrong alternative in a two choice discrimination task, and that error is more than likely due to a bias to respond to the "wrong" stimulus, this recognition tends to introduce a delay of about 100 ms in the invocation of the P300 by that same stimulus. This delay in P300 by the occurence of an error appears to be related to an evaluation by the system of the context within which it operates. The data appear consistent with the suggestion that the magnitude of the P300 can serve as a measure of the degree of revision in the system's biases. This assertion is inferred from the fact that the larger the P300 elicited following the error the less likely the error on the next error-prone trial. Moreover, the larger the P300 that is elicited on error trials the slower will be the subject to respond on the immediately succeeding trial. A shift in response bias, or a tendency to place the response under controlled, rather than automatic, processing mode are plausible interpretations of these data.

These results, and the interpretation proposed above, indicate how the P300 and the study of its latency and amplitude can reveal aspects of the manner in which the human information processing system deals with error trials. These aspects are opaque to the more traditional tools largely because they permit a view of information processing activities that do not have an overt manifestation in performance on the trial in which they are elicited. The view that emerges is one in which at least two information processing streams proceed in parallel. Both depend for their initiation on the initial detection and encoding of the stimulus. However, the processing that leads to the overt response may be completed, yielding the actual response, independently of the evaluative processes whose role is the maintenance of the operating environment.

The metaphor that captures our intent is that of an organization whose operating and administrative arms operate in a highly interactive, but nevertheless independent manner. Actions by the organization's staff are taken in the light of the local interpretation of ongoing events and under the constraints established by the administration's policy. Each event outside the organization, and each action by the organization, trigger in the administration an evaluative process that may long outlast the staff's actions as the administration must optimize its operating policies given the consequences of its own actions and the events in its surround. The time course of the administrative processing may be quite independent of the time course of the processing required by the operating staff before it takes action. Indeed, if staff action was patently erroneous the administration may require additional time before it closed the book on the action, files the reports and makes the necessary adjustments in policy.

The P300 component can, we believe, be viewed as a manifestation of "administrative" rather than "operational" information processing. Donchin et al. (1978) labelled these classes of information processing "strategic" and "tactical" respectively. Evidence is accumulating that the process manifested by the P300 is "future oriented" (see, for example, Donchin, 1981; Fabiani, Karis, Coles, & Donchin, 1983; Karis et al., 1984; Klein, Coles, & Donchin, 1984). The data we reviewed in this chapter are consistent with this view. It seems clear that the magnitude of the P300 elicited on an error trial is related to the performance of the subject on a subsequent trial. Such an effect implies, almost by definition, that the process manifested by the P300 has consequences for future performance. It is, of course, possible that the relationship we observed is fortuitous and both P300 amplitude on trial N, and the subject's performance on trial N+1, are correlated with yet a third factor accounting for both variations. To address this issue we must continue seeking the elucidation of the functional significance of the P300.

In many ways the study described in this report is a direct descendant of the work that Lindsley and his colleagues undertook as the electroencephalographic techniques made their ways from Europe to the United States in the 1930's. The EEG has, of course, become a standard clinical tool and much of the research utilizing the EEG is clinical in nature and in orientation. There is, however, a flourishing research enterprise in which the EEG, and the ERPs embedded in it, are used as tools in the study of cognitive function. Lindsley's work, spanning more than half a century was, and continues to be, an outstanding illustration of the way a scientist bringing the skills and sensibilities of a psychologist can turn the record of a bodily function so it provides a window on the mind. A key element in this enterprise is the chronometric approach that has been so important in Lindsley's research program.

References

Chalupa, L. M., Rohrbaugh, J., Gould, J. E., & Lindsley, D. B. (1974). Evoked potentials and reaction time correlates in monkeys during a simultaneous visual discrimination task. Society for Neurosciences Abstracts, 164.

Coles, M. G. H., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. Journal of Experimental Psychology: Human Perception and Performance, 11, 529-553.

Donchin, E (1979). Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), Evoked Brain Potentials and Behavior (pp. 13-75). New York: Plenum Press.

Donchin, E. (1981). Surprise!... Surprise? Psychophysiology, 18, 493-513.

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 potential research (pp. 555-572). Washington: US Government Printing Office.

Donchin, E., & Lindsley, D. B. (1965). Visual evoked response correlates of perceptual masking and enhancement. Electroencephalography and Clinical Neurophysiology, 19, 325-335.

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

Duncan-Johnson, C. C., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. Biological Psychology, 14, 1-52.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (in press). P300: Theoretical and methodological issues. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.) Advances in Psychophysiology. Volume 2. Greenwich, CT: JAI.

Fabiani, M., Karis, D., Coles, M. G. H., & Donchin, E (1983). P300 and recall in an incidental memory paradigm. Psychophysiology, 20, 439.

Goodin, D. S., Squires, K. C., & Starr, A. (1978). Long latency event-related components of the auditory evoked potential in dementia. Brain, 101, 635-648.

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.

Gratton, G., Dupree, D., Coles, M. G. H., & Donchin E. (1985). The consequence of an error: P300 and future action. Manuscript in preparation.

Herren, R. T., & Lindsley, D. B. (1931). Central and peripheral latencies in some tendon reflexes of the rat. American Journal of Physiology, 99, 167-171.

Hillyard, S. A. (1984). Event-related potentials and selective attention. In E. Donchin (Ed.) Cognitive Psychophysiology: Event-related potentials and the study of cognition. The Carmel Conferences, Vol. I (pp. 51-72). Hillsdale, N.J.: Erlbaum.

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. Electroencephalography and clinical Neurophysiology, 10, 371-375.

Karis, D., Fabiani, M., & Donchin, E. (1984). P300 and Memory: Individual differences in the von Restorff effect. Cognitive Psychology, 16, 177-216.

Klein, M., Coles, M. G. H., & Donchin, E. (1984). People with Absolute Pitch Process Tones Without Producing a P300. Science, 223, 1306-1309.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 197, 792-795.

Lindsley, D. B. (1944). Electroencephalography. In J. McV. Hunt (Ed.) Personality and behavior disorders, Vol. 2 (pp. 1037-1103). New York: Ronald Press.

Lindsley, D. B. (1954). Electrical response to photic stimulation in visual pathways of the cat. Electroencephalography and Clinical Neurophysiology, 6, 690-691.

Lindsley, D. B. (1982). Neural mechanisms of arousal, attention and information processing. In J. Orbach (Ed.) Neuropsychology after Lashley (pp. 315-407). Hillsdale, N.J.: Erlbaum.

Lindsley, D. B., & Emmons, W. H. (1958). Perceptual blanking, evoked potentials and perception time. Electroencephalography and Clinical Neurophysiology, 10, 359.

Lindsley, D. B., Fehmi, L. G., & Adkins, J. W. (1967). Visually evoked potentials during perceptual masking in man and monkey. _Electroencephalography and Clinical Neurophysiology_, _23_, 79.

Lindsley D. B., & Rubenstein, B. B. (1937). Relationship between brain potentials and some other physiological variables. _Proceedings of the Society of Experimental Biology and Medicine_, _35_, 558-563.

Lindsley, D. B., Seales, D. M., & Wilson, G. F. (1973). Changes in the late components of visual evoked potentials with visual information processing. _Society for Neuroscience Abstracts_, 422.

Magliero, A., Bashore, T. R., Coles, M. G. H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. _Psychophysiology_, _21_, 171-186.

McCarthy, G. (1984). Stimulus evaluation time and P300 latency. In E. Donchin (Ed.) _Cognitive Psychophysiology: Event-related potentials and the study of cognition_. The Carmel Conferences, Vol. I. (pp. 254-285). Hillsdale, N.J.: Erlbaum.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. _Science_, _211_, 77-80.

McCarthy, G., Kutas, M., & Donchin, E. (1979). Detecting errors with P300 latency. _Psychophysiology_, _16_, 175.

Naatanen, R. (1982). Processing negativity: An evoked-potential reflection of selective attention. _Psychological Bulletin_, _92_, 605-640.

Pfefferbaum, A., Ford, J. M., Roth, W. T., & Kopell, B. S. (1980). Age differences in P3-reaction time associations. *Electroencephalography and Clinical Neurophysiology*, *49*, 257-265.

Posner, M. I. (1978). *Chronometric explorations of the mind*. Hillsdale, N.J.: Erlbaum.

Schwartz, A. S., & Lindsley, D. B. (1964). Critical flicker frequency and photic following in the cat. *Bol. Inst. Estud. Med. Biol.. Mex.*, *22*, 249-262.

Sutton, S., & Ruchkin, D. S. (1984). The late positive complex: Advances and new problems. In R. Karrer, J. Cohen, & P. Tueting (Eds.) *Brain and information: Event-related potentials* (pp. 1-23). New York: New York Academy of Science.

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

Figure Legends

Figure 1. Reaction time and P300 latency for each probability (20/80 and 50/50), stimulus (male and female) and response (correct - C - and incorrect - I) condition. The frequency of correct and incorrect responses for each condition is also indicated.

Figure 2. Grand average waveforms at Pz for each probability, stimulus and response condition. The solid lines refer to the grand average waveforms for the correct responses, and the dashed lines to the waveforms for the incorrect responses. The average reaction times for correct (solid) and incorrect (dashed) responses are indicated by vertical lines.

Figure 3. Speed-accuracy functions for rare male stimuli, frequent female stimuli and all stimuli in the 50/50 condition. Separate speed-accuracy functions were computed for each of three P300 latency bins (600 to 699 ms, 700 to 799 ms, and 800 to 899 ms).

Figure 4. Reaction time and error rate for female trials in SWITCH and NO SWITCH sequences, as a function of the lag from the incorrect male trial initiating the sequences, for the each probability condition.

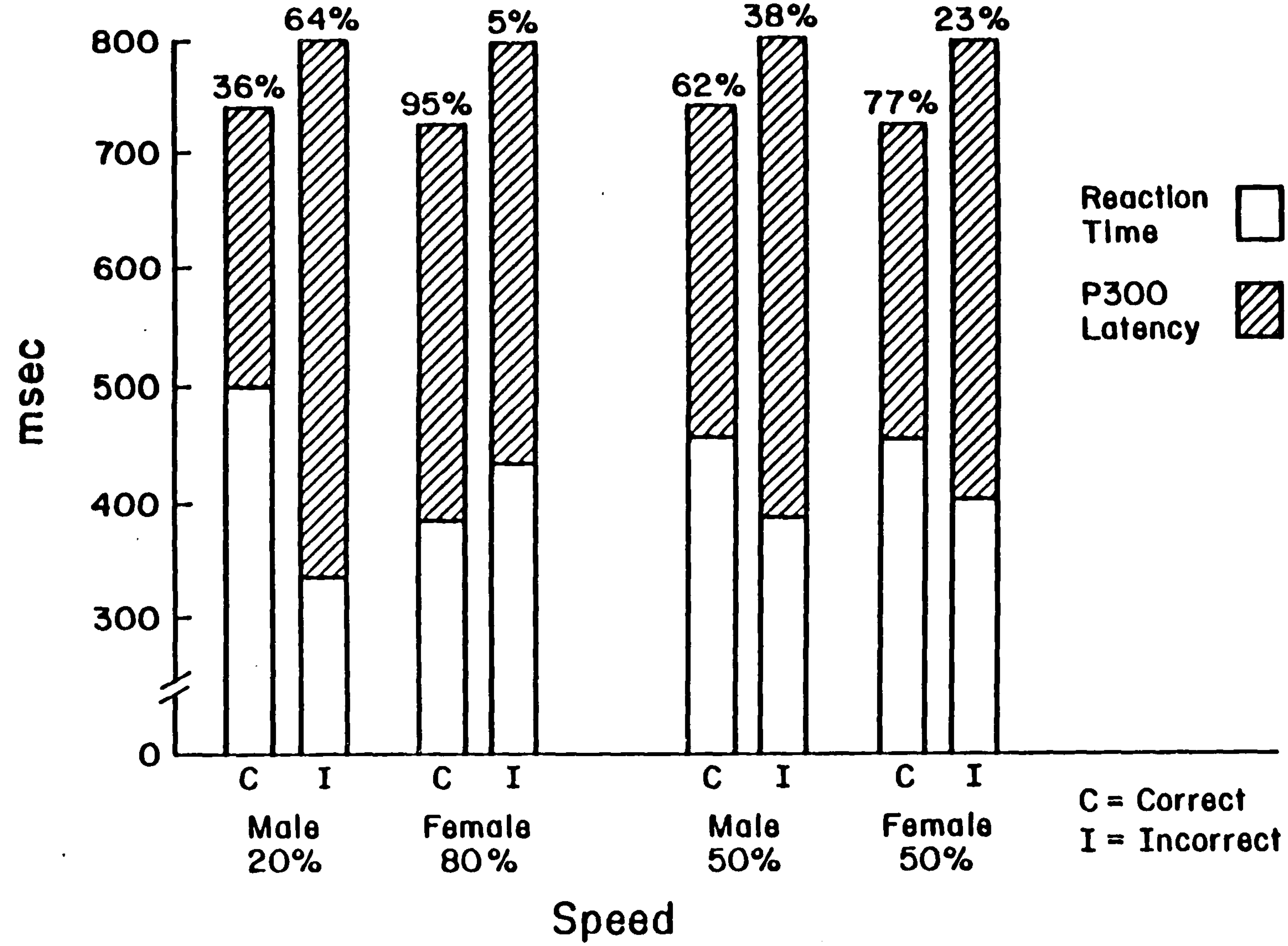

msec
800
700
600
500
400
300
0
36%
64%
95%
5%
62%
38%
77%
23%
C
I
Male
20%
Female
80%
Male
50%
Female
50%
Speed
Reaction Time
P300 Latency
C = Correct
I = Incorrect

Fig. 1

Fig. 2

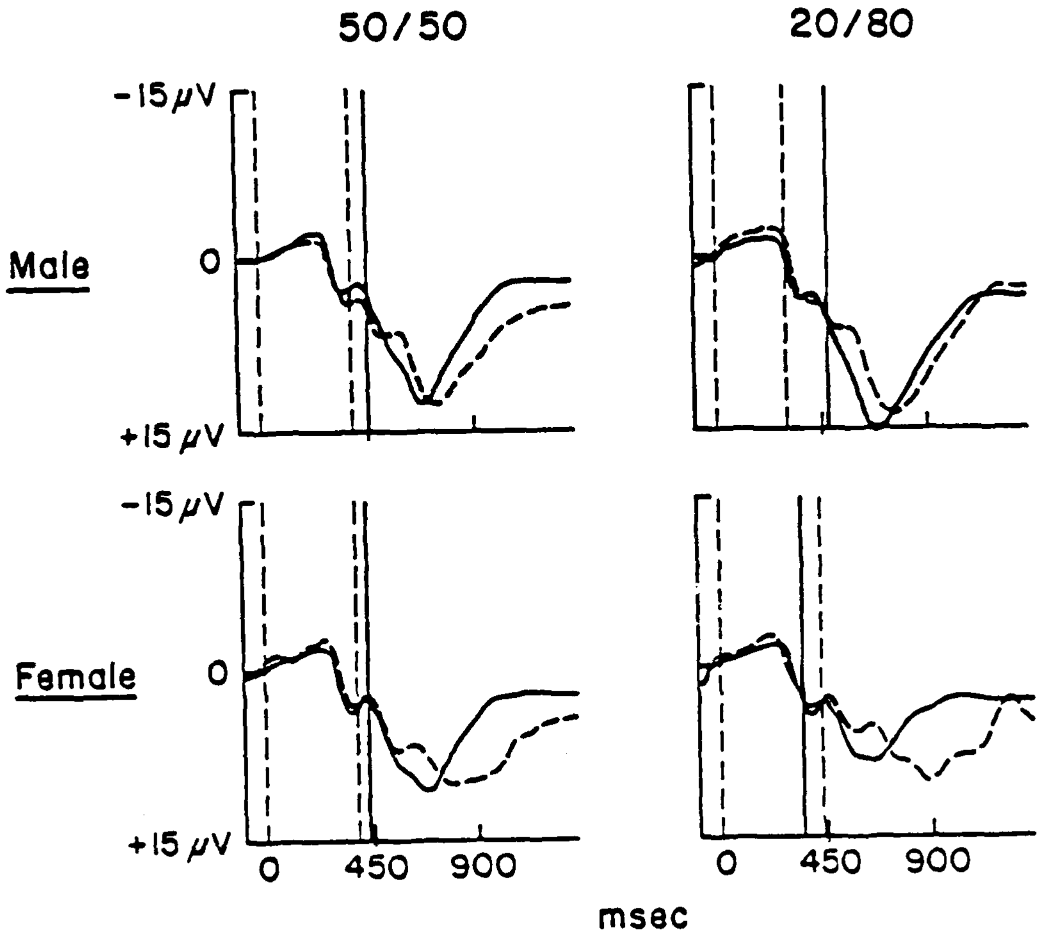
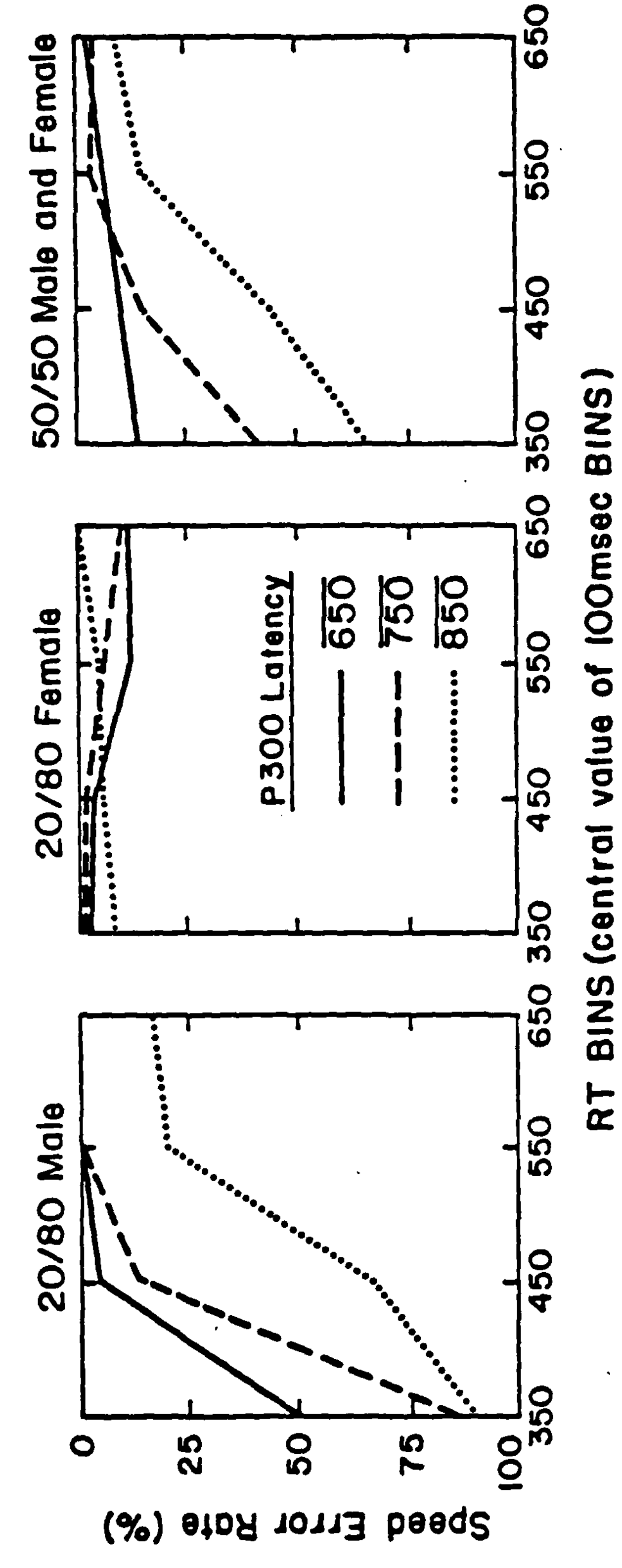
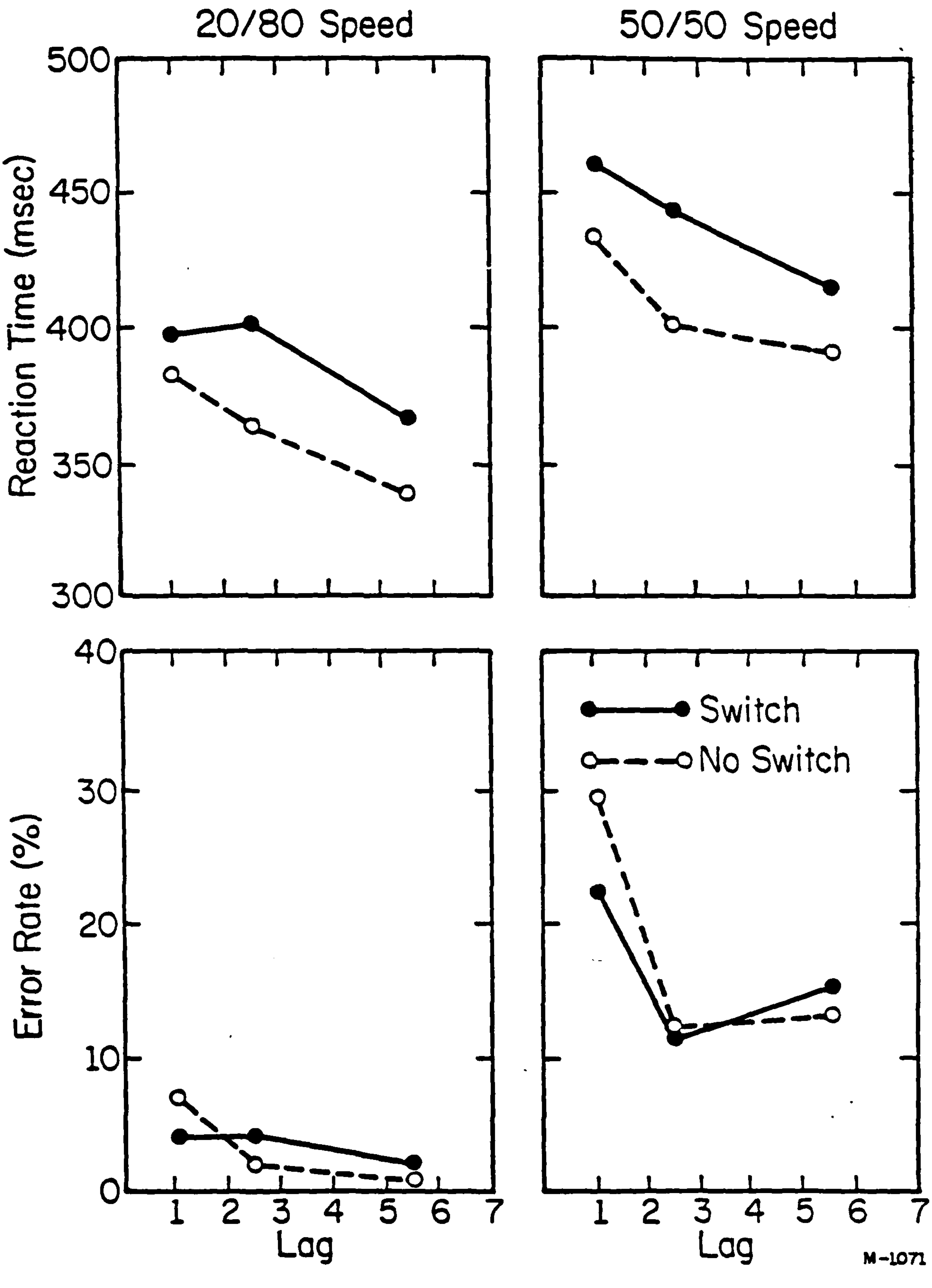


Fig. 3



20/80 Male
20/80 Female
50/50 Male and Female
P300 Latency
650
750
850
Speed Error Rate (%)
0
25
50
75
100
350
450
550
650
RT BINS (central value of 100msec BINS)

Fig. 4


20/80 Speed
50/50 Speed
Reaction Time (msec)
500
450
400
350
300
Error Rate (%)
40
30
20
10
0
Switch
No Switch
1
2
3
4
5
6
7
Lag
M-1071

To appear in M.G.H. Coles, E. Donchin, & S. Porges (Eds.), Psychophysiology: Systems, Processes, and Applications. Middletown, NJ: Till & Till, in press.

Appendix B5

Chapter Twenty-Six

Applications of Brain Event-Related Potentials to Problems in Engineering Psychology

Emanuel Donchin
Arthur F. Kramer
Christopher Wickens

INTRODUCTION

We review in this chapter evidence suggesting that the brain event-related potential (ERP) can be incorporated into the collection of tools of engineering psychology. The utility of the ERP as a tool in the study of cognitive science has been discussed elsewhere (Donchin, 1979, 1981; Donchin, Karis, Bashore, Coles, & Gratton, Chapter 12, this volume; Wickens, 1979). As the human factors that must be addressed by the engineer are increasingly "cognitive" in nature (Rasmussen, 1981; Sheridan, 1981), there is an increasing need for enriching the repertoire of techniques for the assessment of cognitive function. We believe that psychophysiological techniques, in particular ERP-based procedures, can serve this function. We realize that this proposition is not self-evident to the engineering psychology profession. The recording of the ERP is cumbersome. Electrodes must be placed on the subject's scalp. Special equipment is needed for analyzing, digitizing, averaging, and displaying the data. The physiological nature of the signals is essentially unknown, and the functional significance of the ERP components is a subject of controversy. What benefits would accrue to the system designers as they encumber themselves with this exotic technique? Is it likely to help in the assessment of cognitive workload? After all, there is a strong tendency to trust the *subjective* reports of operators in assessing workload. These reports appear to be preferred even to the seemingly simpler techniques proposed by the experimental psychologist. Sheridan and his coworkers concluded (Sheridan, 1980; Sheridan & Simpson, 1979) that it is possible to obtain a reliable and valid measure of workload by administering a rather simple questionnaire. Why should one bother with more costly, elaborate, and indirect measurements of workload?

The question is reasonable, and the answer is clear. If nothing is gained by complicating the measurement process, it is best to avoid the complications. We claim, however, that there are circumstances in which subjective reports may need augmentation, and that in a subset of these circumstances, the ERPs may be very useful.

Consider, for example, the following task. In Figure 26-1 are displayed four pairs of words. The task is to write "yes" next to the pair if the words rhyme, and to write "no" next to the pair if the words do not

Emanuel Donchin, Arthur F. Kramer, and Christopher Wickens. Cognitive Psychophysiology Laboratory, Department of Psychology, University of Illinois, Champaign, Illinois.

Match — Catch

Make — Ache

Catch — Watch

Shirt — Witch

Figure 26-1. A sample of word pairs presented to subjects in a phonological judgment task.

rhyme. Most subjects report that the decision requires the same effort regardless of the pair used, and are quite surprised when they find that their subjective assessment of the workload imposed by these simple judgments does not reflect objective measures of performance.

Note that the four word pairs in Figure 26-1 are instances of four possible relationships between the two words in the pair, as follows:

1. (RO) The two words rhyme and look alike ("Match–Catch").
2. (R-) The two words rhyme but do not look alike ("Make–Ache").
3. (WO) The two words look alike but do not rhyme ("Catch–Watch").
4. (W-) The two words neither rhyme nor look alike ("Shirt–Witch").

We label these pairs with an R to indicate a phonological match, with an O to indicate an orthographic match, with a W to indicate a phonological mismatch, and with a hyphen (-) to indicate an orthographic mismatch. Thus, for the RO and W- pairs, the phonological and the orthographic information agree, and for the R- and WO pairs, there is a conflict between the phonological and the orthographic information. While it is easy to analyze the stimuli in Figure 26-1 and see that they do indeed differ in these attributes, subjects do not usually perceive themselves as having greater difficulty in deciding that the words "Catch–Watch" do *not* rhyme than they do in deciding that the words "Shirt–Witch" do not rhyme.

But these subjective impressions are somewhat misleading. Polich, McCarthy, Wang, and Donchin (1983) and Kramer, Ross, and Donchin (1982) presented subjects with the two words of each pair in succession and required them to indicate their judgments by pressing one of two buttons immediately after the appearance of the second word. The reaction times belied the subjective reports. This can be seen in Figure 26-2, where the reaction time for each of the classes is shown. It is clear that a conflict between phonology and orthography retarded the subjects' reactions by a considerable number of milliseconds. The average delay was about 300 msec when the second word "looked like" the first word but did *not* rhyme with it (the WO pairs, such as "Catch–Watch"). In other words, an individual's subjective assessment may not reveal a processing delay that may cost an operator up to 300 msec in responding to a display change!

What we find, then, is that when tasks place demands on the human information-processing system that affect, or depend on, interactions between or among the automatically activated elements of the

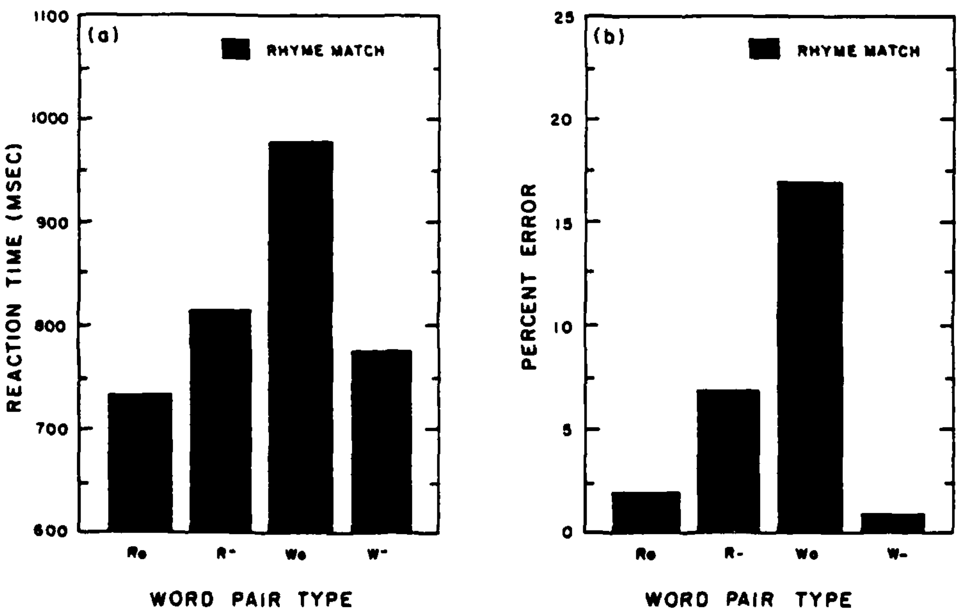


Figure 26-2. Mean reaction times for correct responses and percentage of errors averaged across 40 subjects in the phonological judgment task. (Adapted from Kramer, Ross, & Donchin, 1982.)

processing machinery, loads may be imposed on the system that directly affect its performance, even though they are not available to the internal monitors that yield subjective reports.

This phase of the analysis illustrates the need to supplement subjective reports by accurate and detailed measures of performance. Where, though, can psychophysiology help? We submit that its most effective role is, when properly used, in carrying the analysis beyond the limits imposed by the examination of the more overt responses. Thus, for example, the data in Figure 26-2 indicate that a phonology-orthography conflict delayed the subjects' reactions. But these data do not permit unequivocal conclusions regarding the functional locus of the delay. Did the conflict cause reprocessing of the signal? Were the subjects more cautious when they detected the conflict, or did they require more time to encode the stimuli? Why was the cost of conflict lower for the R- pair than it was for the WO pair? These and similar questions are important not merely for their theoretical significance, but also because our understanding of the nature of the interference is necessary if we are to develop systematic guidelines for improving the design of displays and related systems. The analysis supported by the ERPs may be especially helpful when there is a need to resolve conflicting theories. There are those who suggest that phonological and orthographic codes interact at the encoding stage (Meyer, Schvaneveldt, & Ruddy, 1975; Shulman, Hornak, & Sanders, 1978). Others have suggested that the interference occurs at a response selection stage (Conrad, 1978).

Kramer *et al.*'s (1982) ERP data, shown in part in Figure 26-3, provide information that complements the reaction time data. The waveforms shown in Figure 26-3 are of ERPs averaged over 40 subjects. These data were recorded at the parietal electrode, and each of the lines represents an average over one of the four classes of pairs (RO, R-, WO, W-). As usual, the ERP appears as a sequence of peaks and troughs (often referred to as "components"). It is evident that the waveforms for the four ERPs are congruent until the point of presentation of the second stimulus. The subjects, of course, did not know which of the four pair classes would be used on any trial until the appearance of the second stimulus. Once this happened, the waveforms diverged. It is quite evident that the ERP that was elicited by the WO pair type is different from the other three ERPs. It is characterized by a substantial delay in the elicitation of a large positive (downward-going) component, relative to the appearance of a similar component in the other three ERPs. In our terminology, the latency of this peak, labeled the "P300" for reasons that become apparent later, is increased in the ERP elicited by the WO pairs. Thus, the ERP provides additional data on the two types of orthography-phonology conflict that occurred in this

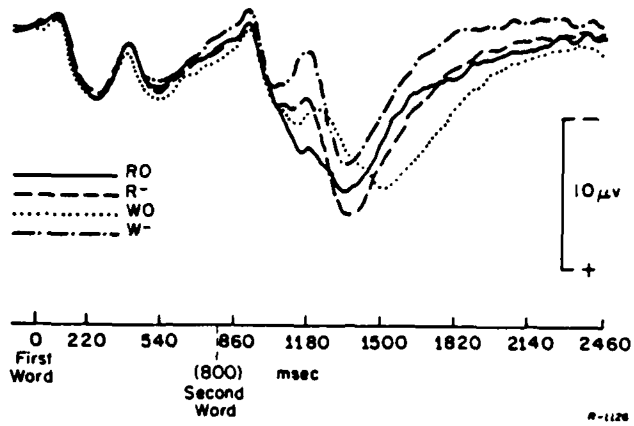


Figure 26-3. Grand average ERPs recorded at the parietal electrode in the phonological judgment task. The data span an epoch that began 100 msec prior to the presentation of the first of a pair of words and ended 1460 msec after the presentation of the second stimulus. The first word was presented at 0 msec, as indicated in the figure. The second word in the pair was presented at 800 msec. Each stimulus was displayed for 200 msec. (Adapted from Kramer, Ross, & Donchin, 1982.)

experiment. If we can interpret these ERP changes, we may be able to gain a better understanding of the process. In fact, since the latency of the P300 component of the ERP provides a measure of mental processing time that is unaffected by response selection and execution processes (McCarthy & Donchin, 1981), the data of Figure 26-3 suggest that at least some of the effect of the orthography-phonology conflict operates prior to the response selection stage. It is interesting to note that the differences obtained in the reaction times were larger than the P300 difference. This suggests that interference, which begins prior to the response selection stage, is amplified during later processing and therefore may reflect a cascading process (McClelland, 1979).

Further examination of Figure 26-3 reveals that the four ERPs differ also int he disposition of a negative (upward-going) peak that just precedes the P300. This peak is labeled "N200." The differences in the amplitude of the N200 component may serve to clarify some issues concerning the detection of orthographic and phonological mismatches. As can be seen in Figure 26-3, the largest N200s were elicited by the W- pairs, in which both the orthography and the phonology of the pair members mismatch. The R- list (phonological match, orthographic mismatch) also elicited a relatively large N200. Thus the R- and W- pair types, which both involve orthographic mismatches, elicited an N200. This suggests that the detection of an orthograpic mismatch may occur automatically. In fact, in an experimental condition not shown here, the subjects were instructed to report "yes" if the words matched visually, regardless of the phonology. The N200 elicited in that condition by the R- and W- pairs was identical to that elicited during the rhyme condition. On the other hand, the WO pairs (orthographic match, phonological mismatch) elicited an N200 only when the subjects were instructed to detect rhymes. This suggests that a phonological mismatch may be detected only when the phonology of the task is relevant. In other words, the ERP data indicate that a phonological comparator was involved solely in the rhyme condition, even though orthographic comparators were involved regardless of the task. Thus both the latency of the P300 component and the amplitude of the N200 component provide information that complements introspection and traditional overt response analysis.

We discuss these data because they illustrate our basic contention: Subjective reports, while valuable, do have limitations. In assessing the demands that a system places on an operator, it is particularly unwise to trust introspective claims that *deny* differences in workload between the systems under comparison. This is especially so when the demands imposed by the system operate at levels of processing that are not normally open to examination by introspection. It is in this domain that the human factors expert is most likely to benefit from the models and techniques of the experimental psychologist. On occasion it will be found that the assessment can be augmented by utilizing ERPs. This is particularly true when there is an interest in developing a theoretical account for the differences between the demands imposed on the operator by different systems. The theoretical models that can be adduced abound in references to internal processing entities. As the ERP components are manifestations of such processing entities, their study is of use.

In the remainder of this chapter, we illustrate these concepts by reviewing a series of studies demonstrating that the amplitude of the P300 can serve as a measure of "workload." We precede this discussion with a brief overview of the study of ERPs. For more details, the reader is referred to Callaway, Tueting, and Koslow (1978), Otto (1978), and Donchin *et al.* (Chapter 12, this volume).

Introductory Comments on the P300 Component

The ERP is a transient series of voltage oscillations in the brain that can be recorded from the scalp in response to the occurrence of a discrete event (Donchin, 1975; Regan, 1972). The ERP is viewed as a sequence of components commonly labeled with an "N" or a "P," which denotes polarity, and a number, which indicates minimal latency measured from the onset of the eliciting event (e.g., "N100" is a negative-going component that occurs at least 100 msec after a stimulus). Since ERPs are relatively small, relative to the ongoing EEG (2–20 μV for the ERPs vs. 50–100 μV for the EEG), their study became practical only after the development of reliable signal averagers (Clynes & Kohn, 1960). These capitalize on the fact that the ERP is, by definition, time-locked to the eliciting event.

It is crucial to recognize the componential nature of the ERP. Early studies of the ERP, which treated the waveform as a unitary entity and measured the amplitude over the entire recording epoch (Satterfield, 1965), were difficult to interpret. The effects of the experimental manipulations tend to be quite specific to a few components, and a combination of the measures of the entire epoch may obscure the relevant variance. There is a degree of controversy as to the proper identification and definition of components (Donchin, Ritter, & McCallum, 1978; Picton & Stuss, 1980). In this chapter, however, we follow Donchin *et al.*'s (1978) definition of an ERP component in terms of the responsiveness of the waveforms to specific experimental manipulations. A component is thus mapped into a cognitive space populated by psychological concepts, such as decisions, expectations, plans, strategies, associations, and memories. Specific components are associated with particular entities in this cognitive space in much the same manner in which cells in the periphery of the visual system are mapped into a field in the visual cortex. The subset of elements in cognitive space associated with a particular component thus contributes to the definition of the ERP component.

The specific attributes of a waveform that are examined in defining a "component" are the amplitude, latency, and scalp distribution. It is the sensitivity of these attributes to experimental manipulations that defines an ERP component. Although no reference has been made to the underlying neural source of components, it is generally assumed that a scalp distribution that is invariant across repeated stimulus presentations implies a specific and fixed set of neural generators (Goff, Allison, & Vaughan, 1978; Wood & Allison, 1981). Thus the scalp distribution, which is related to the underlying neural population responsible for the generation of the component, is assumed to be a crucial defining characteristic.

The ERP components we discuss in this chapter are "endogenous" and are distinct from another class of ERPs called "exogenous" (Donchin *et al.*, 1978; Sutton, Braren, Zubin, & John, 1965). The exogenous components represent an obligatory response of the brain to the presentation of a stimulus. These components are primarily sensitive to such physical attributes of the stimuli as intensity, modality, and rate. The seven peaks or "bumps" that occur in the first 8–10 msec after the presentation of an auditory or somatosensory stimulus are a prototypical example of the exogenous category (Jewett, Romano, & Williston, 1970).

Endogenous components, typically, are not sensitive to changes in the physical characteristics of the eliciting stimuli. On the other hand, these components are very sensitive to changes in the processing demands of the task imposed on the subject. The endogenous components are nonobligatory responses to stimuli. The strategies and expectancies of the subject, as well as other psychological aspects of the task, account for the variance in the endogenous components. A typical example, and one to which we devote the remainder of this chapter, is the P300 component.

This ERP component is elicited by rare, task-relevant stimuli. A task in which it is readily elicited is often called the "oddball" paradigm. In a study by Duncan-Johnson and Donchin (1977), using this paradigm, the subjects were instructed to count covertly the total number of higher-pitched tones in a Bernoulli series. In different blocks of trials, the relative probability of the two tones was manipulated. It can be seen in Figure 26-4 that the amplitude of the P300 increased monotonically as the probability of the stimulus decreased. This occurred regardless of which of the two stimuli was being counted. When the subjects were solving a word puzzle and were not required to process the tones, the P300s were not elicited. Note that the ERPs in Figure 26-4 that were obtained in this "ignore" condition showed no P300 at all levels of probability. Thus, the amplitude of P300 is determined by a combination of the task relevance and the subjective probability of the eliciting event. This basic finding plays a crucial role in the use of P300 in the assessment of workload.

The demonstration that P300 is elicited by unexpected, task-relevant stimuli led Donchin, McCarthy, Kutas, and Ritter (1983) to suggest that "the P300 is a manifestation, at the scalp, of neural action that is invoked whenever the need arises to update the 'neuronal model' (Sokolov, 1969) that seems to underlie the ability of the nervous system to control behavior" (p. 105). The neural or mental model is continually assessed for deviations from inputs and revised when the discrepancies exceed some criterion value. The frequency with which the mental model is revised is based on the surprise value and task relevance of the stimuli. Donchin (1981) also argued that the concept of a subroutine is an appropriate metaphor for the activity of ERP components (Donchin, 1975; Donchin, Kubovy, Kutas, Johnson, & Herning, 1973). In software applications, subroutines represent algorithms that are designed to accomplish a specific task and which can be employed in a variety of different programs. ERP components may be associated with specific information-processing functions that are activated in a variety of different tasks. In the case of the P300, the subroutine may be invoked whenever there is a need to evaluate surprising, task-relevant events. This interpretation of the changes in P300 amplitude is strengthened by the evidence that has accumulated in the past decade regarding the factors that control the latency of the P300. As the use we make of P300 in the analysis of human–machine interaction depends strongly on our theoretical interpretation of the component, it is useful to provide a brief review of the latency data and their interpretation.

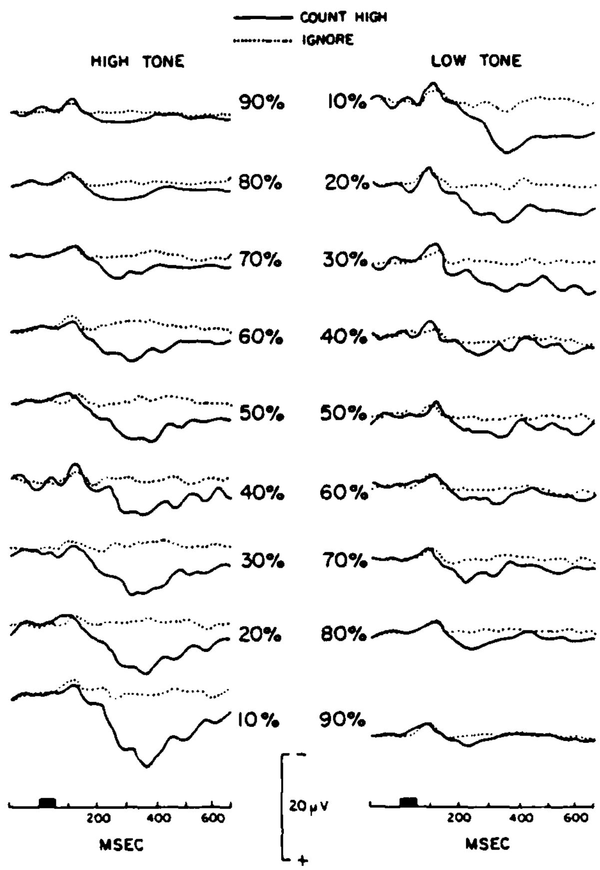


Figure 26-4. Averaged ERPs elicited by high and low tones presented in a Bernoulli series. The waveforms represent experimental conditions in which subjects counted the high tones (solid lines) or solved a word puzzle as the words were presented. (From "On Quantifying Surprise: The Variation in Event-Related Potentials with Subjective Probability" by C. C. Duncan-Johnson and E. Donchin. *Psychophysiology*, 1977, 14, 456–467. Reprinted by permission.)

The Latency of the P300 Component

The peak latency of the P300 component appears to depend on the time required to recognize and evaluate a task-relevant event. The latency ranges between 300 and 750 msec following the presentation of a discrete stimulus. In fairly simple tasks calling, for example, for a discrimination between two tones that differ in pitch (i.e., 1000–1600 Hz), the stimuli elicit relatively short-latency P300s. More difficult discriminations 8.e., semantic analysis) result in increases in the latency of P300.

Assuming that manual or vocal reaction time terminates processing, and that P300 is a manifestation of a process that precedes the response, then it would be expected the P300 latency and reaction time should positively covary. This prediction has been supported by numerous studies (Bostock & Jarvis, 1970; Rohrbaugh, Donchin, & Eriksen, 1974; Wilkinson & Morlock, 1967). Other investigations, however, have failed to detect a relationship between P300 latency and reaction time (Karlin & Martz, 1973; Karlin, Martz, & Mordkoff, 1970).

Donchin *et al.* (1978) proposed an interpretation of the processes underlying the P300 that may reconcile these contradictory findings. They suggested that P300 latency is determined by the time required to evaluate the stimulus, but is largely independent of response selection and execution time. The correlation between reaction time and P300 latency would accordingly vary as a function of the percentage of reaction time variance that is accounted for by stimulus evaluation processes. This percentage would be affected by the strategies employed by the subject. The strategies, therefore, should influence the relationship between P300 latency and reaction time (see also Ritter, Simson, & Vaughan, 1972). Evidence that P300 is determined by the amount of time required to recognize and evaluate a stimulus has been reported by several investigators who have employed Sternberg's (1966, 1969a, 1969b) additive-factors methodology (Ford, Mohs, Pfefferbaum, & Kopell, 1980; Ford, Roth, Mohs, Hopkins, & Kopell, 1979; Gomer, Spicuzza, & O'Donnell, 1976; Kramer, Fisk, & Schneider, 1983). Sternberg's paradigm involves the factorial manipulation of two or more experimental variables that are expected to differentially affect the durations of specific stages of processing. For example, the superimposition of a mask over a display is assumed to influence processing in an early, perceptual stage. On the other hand, reduction of the compatibility between the stimulus and the response would be expected to affect the selection and the execution of the response. In the studies mentioned above, both P300 latency and reaction time increased monotonically with increasing memory load.

Other investigators, employing different paradigms, also report that P300 latency and reaction time are positively correlated when stimulus evaluation time is manipulated. Squires, Donchin, Squires, and Grossberg (1977) found that P300 latency and reaction time covaried with the difficulty of auditory and visual discriminations. Furthermore, P300 latency varied with the manipulation of stimulus discriminability, while reaction time was influenced by both stimulus evaluation and response selection factors. Heffley, Wickens, and Donchin (1978) performed an experiment in which subjects were required to monitor a dynamic visual display for intensifications of one of two classes of targets. P300 latency was found to increase monotonically with the number of elements on the display. Since subjects were not required to make an overt response, the differences in P300 latency were attributed to stimulus evaluation processes.

If P300 latency is determined by stimulus evaluation time and is largely independent of the time required for response selection and execution, then experimental variables that have a different effect on processing time in the two stages should influence the relationship between P300 latency and reaction time. For example, when subjects are instructed to respond quickly with a low regard for accuracy, their responses are probably emitted without full evaluation of the stimulus (Wickelgren, 1977). On the other hand, if subjects are instructed to respond accurately, they are likely to perform a more thorough analysis of the stimuli prior to responding. This analysis leads to the prediction that the correlation between P300 latency and reaction time will vary with the subject's strategies. Specifically, it is predicted that the correlation will be high and positive when the subjects are instructed to be accurate, while low correlations will be observed under speed instructions.

Kutas, McCarthy, and Donchin (1977) tested this hypothesis by requiring subjects to distinguish between two stimuli under both speed and accuracy instructions. In one experimental condition, subjects were required to discriminate between two names, "Nancy" and "David," presented on a CRT (with relative frequencies of 20% and 80%, respectively). In a second condition, female names comprised 20% of the items and male names 80%. In the third condition, subjects were required to discriminate between synonyms of the word "Prod" that occurred with a relative probability of 20% and unrelated words that were presented with the complementary probability. The average P300 latency was shortest for the first condition, intermediate for the second, and longest for the third condition. The more complex the discrimination, the longer the P300 latency. A detailed analysis of the single trials (Woody, 1967) revealed that the correlation between P300 latency and reaction time was larger for the accuracy condition (.617) than the speed condition (.257). Kutas *et al.* (1977) concluded that the data supported the hypothesis that P300 latency reflects the termination of a stimulus evaluation process, while reaction time indexes the entire sequence of processing from encoding to response selection and execution. Thus, under the accuracy condition, when response selection was contingent on stimulus evaluation processes, P300 latency and reaction time were tightly coupled. However, when subjects performed the discrimination under the speed instructions, the processes of stimulus evaluation and response selection were more loosely coupled, and hence the relationship between P300 latency and reaction time was not as high.

Additional evidence bearing on the issue of the P300's sensitivity to the manipulation of stimulus evaluation processes has been obtained in a study by McCarthy and Donchin (1981), who manipulated orthogonally two independent variables in an additive-factors design (Sternberg, 1969a). One factor, stimulus discriminability, has been shown to affect an early encoding stage of processing, while the second factor, stimulus response incompatibility, influences the later stages of response selection and execution (Bertelson, 1963; Sanders, 1970; Schwartz, Pomerantz, & Egeth, 1977). The subjects' task was to decide which of two target stimuli, the words "RIGHT" or "LEFT," were presented in a matrix of characters on a CRT. The characters were either presented within a 4 × 4 matrix of # (number) signs (no-noise condition) or in a 4 × 4 matrix of letters chosen randomly from the alphabet (noise condition). Stimulus response incompatibility was manipulated by preceding the target matrix either with the cue "SAME" or with the cue "OPPOSITE." "SAME" signaled a compatible response. The cue "OPPOSITE" indicated an incompatible response: The right hand would respond to the word "LEFT" and the left hand to the cue "RIGHT." Reaction time increased when the command word was embedded in noise and when the response was incompatible with the stimulus. The effect of the two variables on the reaction time was additive, implying that these manipulations influenced different stages of processing. P300 latency was increased by the addition of the noise to the target matrix, but was not affected by the incompatibility between the stimulus and the response. These results support the conclusion that P300 latency is affected by a subset of the set of processes that affect reaction time. The P300 is elicited only after the stimulus has been evaluated. Subsequent processing required for the selection and execution of the response does not appear to influence the latency of the P300.

The P300 component of the ERP provides a metric for the decomposition of stages of information processing that complements the traditional behavioral measures. In terms of applications to system design and workload evaluation, ERPs used in conjunction with behavioral and subjective measures permit the assessment of stage-specific task interference effects. For example, if two time-shared tasks interfere with each other, it is usually desirable to know the locus of this interaction. Only by discovering the stage at which tasks interact can systems be designed that minimize operator workload.

THE P300 AND HUMAN ENGINEERING

P300 and Perceptual-Central Processing Resources

The studies reviewed above provide evidence that the P300 component is a manifestation at the scalp of a processing entity, or a subroutine, that is involved whenever surprising, task-relevant stimuli are present. The routine appears to be performing a role in the context-updating activities that occur whenever an event calls for the revision of the neuronal model or schema of the environment. It is noteworthy that this subroutine is invoked only if the stimuli are associated with a task that requires that they be processed. Ignored stimuli do not elicit a P300. But what if the stimuli are only partially ignored? What if the subject is instructed to perform the oddball task concurrently with another task? Would the amplitude of the P300 reflect the centrality of the oddball task? Would it, perhaps, change with the amount of resources allocated to the oddball task? Clearly, if it would change in this manner, the P300 might serve as a very useful measure of the amount of resources demanded by the two tasks. It is this series of questions that lies at the core of the usage that can be made of P300 in the assessment of workload.

The study of cognitive workload and of the allocation of processing resources to several tasks performed concurrently is, in fact, the area of research that has profited most from the incorporation of ERP measures. The research reviewed here has been performed within the framework of resource allocation theory. This class of models suggests that it is useful to conceptualize human capacity as represented by a finite pool of "resources" available for sharing among concurrently performed tasks (Kahneman, 1973; Moray, 1967; Norman & Bobrow, 1975). In the Kahneman (1973) model, these processing resources are undifferentiated, implying that all tasks draw resources from the same pool. The general model predicts that when two tasks are time-shared, their levels of performance should decrease relative to single-task levels.

This model underlies the secondary-task technique, a method that is commonly employed in the assessment of the workload associated with a task; the workload is viewed as reflected by the amount of processing resources consumed by a task (Knowles, 1963; Rolfe, 1971; Wickens, 1979). In the secondary-task technique, the subject is assigned two tasks—

a "primary" task, which is to be performed as well as possible, and a "secondary" task, which need be performed only to the extent that primary-task performance remains stable. It is assumed that the demands imposed upon the subjects by the primary task can be assessed by monitoring performance on the secondary task. An easy primary task will require a minimal amount of processing resources, leaving an ample supply for the performance of a secondary task, while a difficult primary task will require the majority of processing resources, leaving an insufficient supply for the performance of the secondary task. Thus, the better the performance of the *secondary* task, the less demanding the primary task.

Although the secondary-task procedure has been extensively used, it presents a number of practical problems (Brown, 1978; Ogden, Levine, & Eisner, 1979). Particularly unfortunate is the fact that secondary-task responses often intrude upon primary-task performance. Of course, fluctuations in primary-task performance make the interpretation of the resource tradeoff extremely difficult. Evidently, it would be useful to have a secondary task that is sensitive to changes in primary task difficulty but that does not require an overt response.

It has been the basic assumption of our research program that the oddball task can be used as a nonintrusive secondary task, since the ERP-eliciting tones occur intermittently, are easily discriminable, and do not require an overt response. Another advantage of this procedure is that it can be applied uniformly across different operational settings. In other words, the oddball task can be inserted into virtually any operational setting without requiring modifications in the system associated with the primary task. Wickens, Isreal, and Donchin (1977) reported one of the first studies in the series, using a compensatory tracking task as the primary task and the oddball paradigm as the secondary task.

Figure 26-5 illustrates the experimental procedures used in this and several other studies to be discussed. The subjects sat in front of a CRT and were instructed to cancel computer-generated cursor movements by keeping the cursor superimposed on a target in the center of the display. This was accomplished by movement of a joystick mounted on the right-hand side of the subject's chair. Levels of tracking difficulty were manipulated by requiring the subject to track in either one or two dimensions (horizontal and/or vertical). The compensatory tracking task was defined as the primary task. In addition to the tracking, the subjects were also instructed to count one of two tones presented in a Bernoulli series of high- and low-pitched tones. Control conditions were also included in which the subjects performed each of the two tasks separately.

The data indicate that the introduction of the tracking task drastically diminished the amplitude of the P300. However, no further reduction in P300 amplitude could be observed as tracking difficulty increased by requiring tracking in two dimensions. Even though tracking difficulty—assessed by root mean square

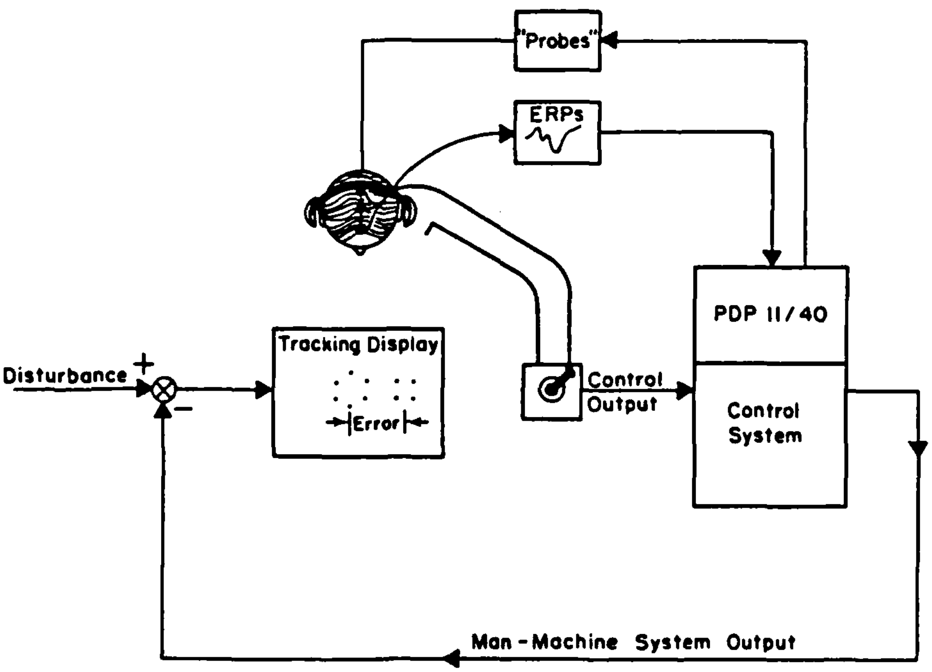


Figure 26-5. An illustration of the experimental paradigm employed in the analysis of the utility of the ERP as a workload measure.

(RMS) error as well as by reaction time to the tones—definitely increased with the addition of a tracking dimension, P300 amplitude did not change.

Isreal, Chesney, Wickens, and Donchin (1980) conducted a similar study requiring subjects to perform a compensatory tracking task concurrently with a counting task. In this case, however, the bandwidth of the random forcing function, rather than the dimensionality of the tracking task, was manipulated. The bandwidth was increased gradually until the cursor's speed reached the highest level the subject could tolerate without exceeding a preset error criterion.

The results are shown in Figure 26-6. Again, P300 amplitude was diminished by the introduction of the tracking task, but increases in the bandwidth of the forcing function did not produce systematic changes in the amplitude of the P300. These results cannot be explained easily within the framework of an undifferentiated capacity theory if we assume that P300 amplitude indexes the demands placed on the subject by the primary task. Increasing the bandwidth clearly affects the performance of overt secondary tasks (McDonald, 1973; Wierwille, Gutmann, Hicks, & Muto, 1977). The fact that P300 did not change, even though a dramatic drop in amplitude was observed with the introduction of the task, requires explanation.

One interpretation of the results is that the P300 is not sensitive to the processing demands of the task, but instead reflects the motor activity required by tracking. This hypothesis was tested by Isreal, Chesney, Wickens, and Donchin (1980) by instructing subjects to manipulate a joystick with one hand concurrently with the oddball task. The amplitude of the P300 component elicited by the tones was not affected by the motor demand. Thus, it would seem that hand movements per se did not decrease the amplitude of the P300.

Another interpretation of the results is that the resources that are tapped when the dimensionality, or the bandwidth of the target, is increased are not the resources reqwuired by the oddball task. Several investigators have proposed that processing resources are not undifferentiated, but, rather, are structured according to various information-processing stages (Kantowitz & Knight, 1976; Kinsbourne & Hicks, 1978; Navon & Gopher, 1979, 1980; Sanders, 1979). Wickens (1980) has identified hypothetical processing structures on the basis of input and output modalities (visual-auditory, manual-vocal), stages of information processing (encoding and central processing, response selection and execution), and codes of processing (verbal, spatial). In this framework, dual tasks are expected to interfere to the extent that they share overlapping resources. For example, two tasks that both require substantial central processing will interfere with each other to a greater extent than a task with central processing demands and another with

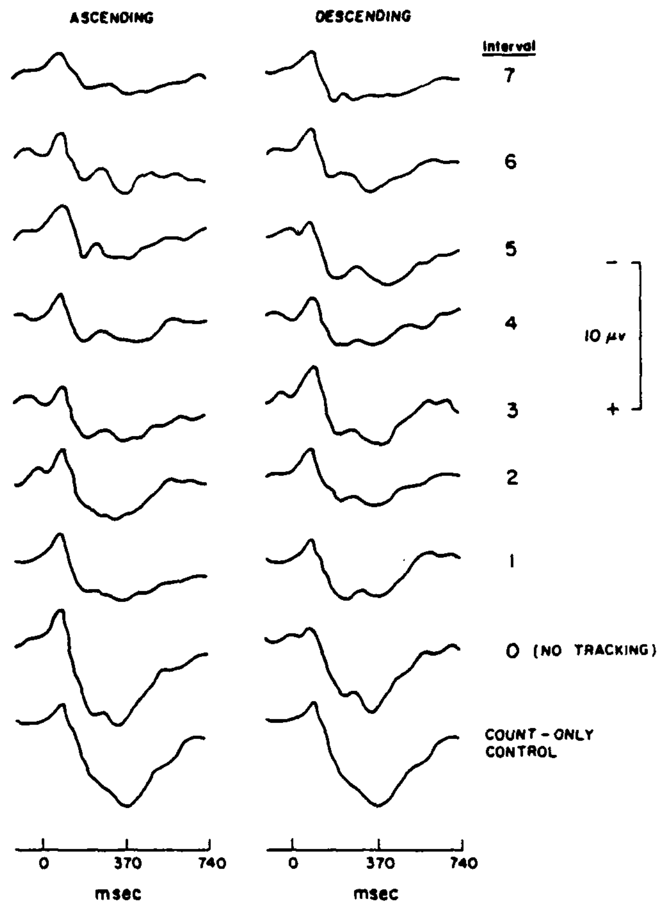


Figure 26-6. Average parietal ERPs, elicited by equiprobable counted tones, for each bandwidth interval and count-only control conditions, for ascending and descending blocks of trials. Bandwidth increases from 1 to 7. (From "P300 and Tracking Difficulty: Evidence for Multiple Resources in Dual-Task Performance" by J. B. Isreal, G. L. Chesney, C. D. Wickens, and E. Donchin. *Psychophysiology*, 1980, 17, 259–273. Reprinted by permission.)

heavy demands for response processes. This view of the allocation of processing resources is consistent with studies that show little or no decrement in performance when two difficult tasks are time-shared (Allport, Antonis, & Reynolds, 1972; North, 1977; Wickens & Kessel, 1979).

The notion that P300 is sensitive to a specific aspect of information processing is consistent with the data, reviewed above, regarding the relation between P300 latency and reaction time, P300 latency appears to be sensitive to a *subset* of the processes that determine reaction time. Furthermore, P300 latency is influenced by manipulations of factors that are assumed to affect relatively early processes of stimulus evaluation, while being insensitive to changes in variables that produce their effect on the later processes of response selection and execution. If the manipulation of the dimensionality and bandwidth of the tracking task demands resources associated largely with response selection and execution processes, then P300 amplitude should not reflect fluctuations in performance. On the other hand, if the perceptual aspects of a task are manipulated, the amplitude of the P300 elicited by a secondary task can be expected to covary with primary-task difficulty.

Isreal, Wickens, Chesney, and Donchin (1980) tested the latter hypothesis by combining the oddball task as a secondary task with a visual monitoring task that served as the primary task. The subjects were instructed to monitor a simulated air traffic control display either for course changes or for intensifications of one of two classes of stimuli (triangles or squares). Primary-task difficulty was manipulated by increasing the number of elements traversing the CRT (Sperando, 1978). The numerosity variable did have a systematic effect on reaction time to the tones when subjects were monitoring for course changes. Reaction time increased monotonically from the control condition to the condition in which subjects were required to monitor eight elements simultaneously. However, in the flash detection condition, reaction time did not increase significantly as a function of the number of elements displayed.

As can be seen in Figure 26-7, the P300 elicited by the counted tones decreased monotonically with increases in difficulty in the monitoring task when subjects were detecting course changes. In the flash detection condition, P300s decreased with the introduction of the monitoring task, but increases in the number of display elements failed to attenuate P300 amplitude further. This result is also consistent with the reaction time data. Since the primary task did not require a response, the data of Isreal, Wickens, Chesney, and Donchin (1980) have demonstrated that P300 amplitude is sensitive to the perceptual demands of a primary task.

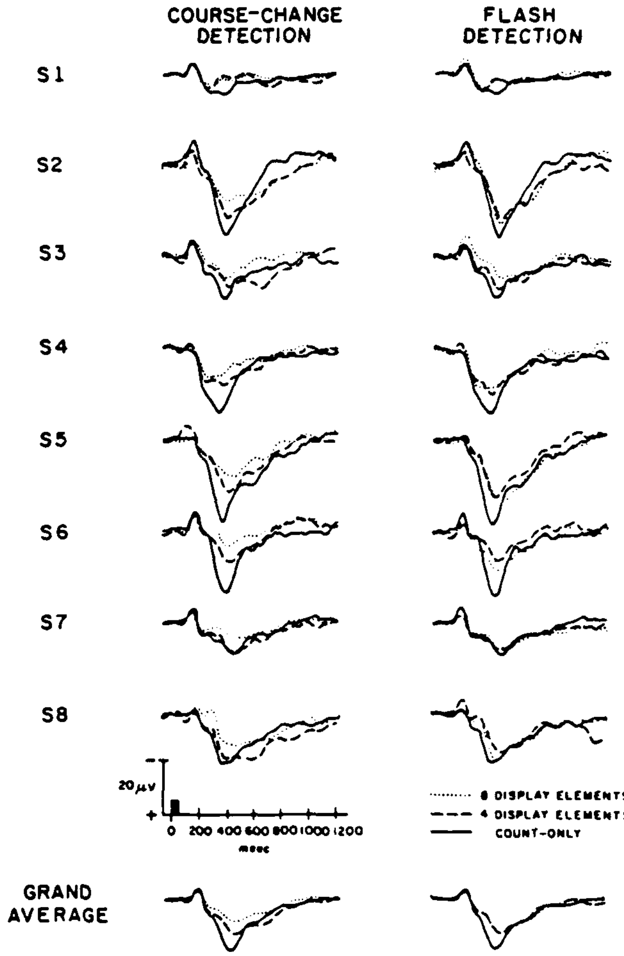


Figure 26-7. Single-subject and average ERPs elicited by infrequent, counted tones presented concurrently with each of two monitoring tasks. Two monitoring conditions as well as a count-only control condition are presented. All waveforms displayed were recorded at the parietal electrode. (From "The Event-Related Brain Potential as an Index of Display Monitoring Workload" by J. B. Isreal, C. D. Wickens, G. L. Chesney, and E. Donchin. *Human Factors*, 1980, 22, 212–224. Reprinted by permission.)

The Use of P300 in Task Analysis

This structure-specific conception of processing resources has several implications for the study of human–machine systems. One area that might benefit from the use of the structure-specific analysis of human information-processing resources is task analysis. Traditionally, the analysis of operator performance in complex systems has been conducted by detailing the observable aspects of tasks and task sequences (Kidd & Van Cott, 1972). This analysis has usually taken the form of elaborate flow charts, which outline such aspects of operator behavior as information input, decisions, and required actions (Coakley & Fucigna, 1955; Folley, Altman, Graser, Preston, & Weislogel, 1960). Although these procedures provide an accurate description of the behavior exhibited by the operators, they do not enable a microanalysis of the task that could provide the system designer with information on the resources required by different subtask sequences. It would be useful to examine a breakdown in performance under high-workload conditions for their relation to resource competition. For example, it would be advantageous to know whether the operator is required to perform tasks that demand a great deal of response processing but little perceptual analysis.

We (Kramer, Wickens, & Donchin, 1983) performed a componential analysis of the demands of controlling higher-order systems. By "order of control," we refer to the number of time integrations of the output of a controller (i.e., joystick) and the output of the system. In a first-order or velocity-driven system, a deflection of the joystick corresponds to a change in the velocity of the controlled element. A second-order or acceleration-driven system produces a change in the acceleration of the controlled element proportional to the movement of the control stick. The increase in system order appears to increase the demand for both perceptual resources (Wickens, Derrick, Micallizi, & Berringer, 1980) and response-related resources (North, 1977; Trumbo, Noble, & Swink, 1967; Vidulich & Wickens, 1981). Effective control over second-order dynamics requires a large degree of perceptual anticipation, as well as a modified response strategy. Assuming that P300 amplitude is sensitive to the perceptual aspects of a task, then a reduction in P300 amplitude by higher-order control should localize some of the influence of the order variable at the earlier processing stages.

Figure 26-8 illustrates the subjects' task. The target appeared on the screen and moved in a straight line at a randomly selected angle. The subjects had to move the cursor into the neighborhood of the target. The time between the appearance of the target and its acquisition by the cursor was called the "acquisition phase." Acquisition was accomplished by manipulating the two-axis joystick mounted on the right side of the chair in which subjects sat. Successful acquisition

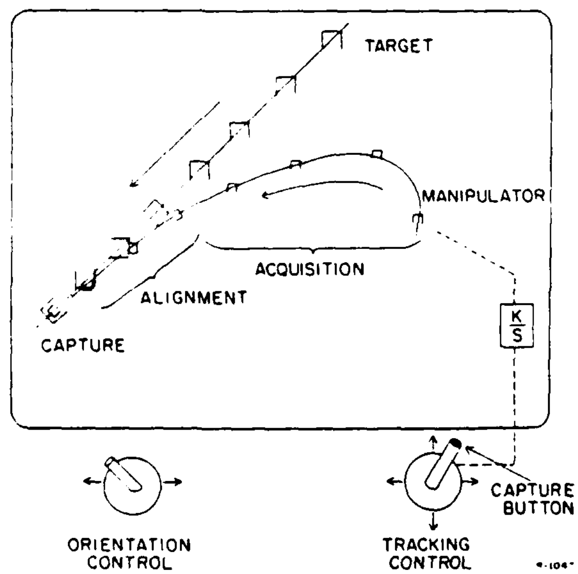


Figure 26-8. The temporal sequence of the target acquisition task (from upper right to lower left). (From "An Analysis of the Processing Demands of a Complex Perceptual-Motor Task" by A. F. Kramer, C. D. Wickens, and E. Donchin. *Human Factors*, 1983, 25. Reprinted by permission.)

initiated the alignment phase. The target began to rotate at a constant velocity in either a clockwise or a counterclockwise direction. The subjects had to rotate the cursor at the same velocity as the target, while also keeping the two elements superimposed. The rotation was accomplished by manipulating the single-axis joystick mounted on the left side of the chair. A deflection of the stick to the right produced a clockwise rotation of the cursor at an angular velocity proportional to the angle of deflection; a deflection to the left produced a counterclockwise rotation. Deviation from the initial acquisition criterion for more than 1000 msec necessitated a realignment of the elements. Once the subjects decided that all of the criteria had been satisfied and that the target and cursor were aligned, they could press a capture button, and the trial was terminated.

We assumed that the alignment phase would be more difficult than the acquisition phase, due to increased perceptual demands imposed by the requirement to control the additional rotational axis. We predicted, therefore, that the P300 elicited by the intensifications of the target and cursor, associated with an oddball task run concurrently with the tracking task, would be larger during the acquisition than during the alignment phase.

The ERP results presented in Figure 26-9 confirm these predictions. The P300 amplitude was attenuated as a function of phase, larger-amplitude P300s elicited in the acquisition phase, and of system order; larger P300s were elicited during the easier, first-order tracking. Another study employing a compensatory tracking task also found a systematic relationship between P300 amplitude and system order (Wickens, Gill, Kramer, Ross, & Donchin, 1981). These studies, along with additive-factors investigators of manual control parameters, have provided converging evidence that system order has a salient perceptual-central processing component (Wickens & Derrick, 1981; Wickens, Derrick, Micallizi, & Berringer, 1980). The results might also be useful in the design and evaluation of complex tracking tasks. If operators are required to perform a tracking task with higher-order system dynamics, then concurrently performed tasks should be designed so as to minimize perceptual-central processing load. We see here, again, how the ERPs provide data that increase the theoretical depth with which one can draw conclsuions about the human information-processing system.

P300 and Resource Reciprocity

The studies cited above have demonstrated a robust relationship between P300 amplitude and the allocation of processing resources in a secondary task. P300s elicited by secondary-task probes decrease in amplitude with increases in the perceptual-central pro-

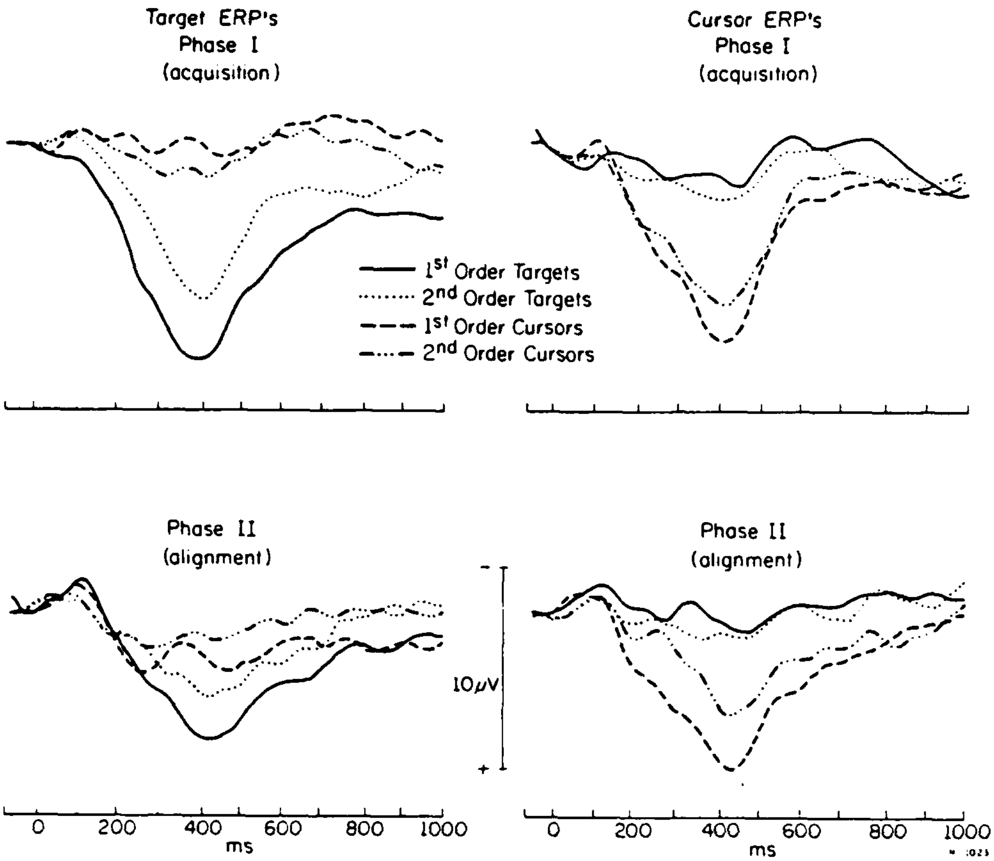


Figure 26-9. Average parietal waveforms elicited by intensifications of the tracking elements. The left panel presents waveforms recorded when the intensity of the target was the relevant event. The right panel displays waveforms collected when the intensity of the cursor was relevant. The top panels display waveforms recorded during the acquisition phase; the bottom panels present waveforms collected during the alignment phase of the target acquisition task. (From "An Analysis of the Processing Demands of a Complex Perceptual-Motor Task" by A. F. Kramer, C. D. Wickens, and E. Donchin. *Human Factors*, 1983, 25. Reprinted by permission.)

cessing difficulty of primary tasks. As outlined previously, one of the basic assumptions of the secondary-task technique is that increases in primary-task difficulty divert processing resources from the secondary task. The decrement in secondary-task performance is believed to reflect this shift of resources from the secondary to the primary task. Thus, it is assumed that there is a reciprocal relationship between the resources allocated to the primary and secondary tasks. If this assumption is correct, then it should be possible to demonstrate that P300s elicited by task-relevant, discrete events embedded within the primary task are directly related to primary-task difficulty.

We (Kramer, Wickens, Vanasse, Heffley, & Donchin, 1981) conducted an experiment in which ERPs were elicited by task-relevant events embedded within a tracking task. The subjects were required to perform a single-axis pursuit step-tracking task with either first-order (velocity) or second-order (acceleration) control dynamics. In this task, the horizontal position of a target was determined by a random series of step displacements occurring at 3-sec intervals. The subjects' task was to keep the cursor superimposed on the target. Difficulty was varied by manipulating two variables: the degree of predictability of the series of steps, and the system order. In the high-predictability condition, the step changes alternated in a regular right-left pattern. In the low-predictability condition, the sequence of step changes was random. The magnitude of the changes was unpredictable in both conditions. The two dimensions of difficulty, system order and input predictability, were crossed to create three conditions of increasing difficulty: first-order control of predictable input, first-order control of unpredictable input, and second-order control of unpredictable input.

Three different types of probes were employed as ERP-eliciting events. In one condition, subjects performed the tracking task while also counting the number of occurrences of a low-pitched tone from a Bernoulli series of high- and low-pitched tones. In the second condition, subjects counted the dimmer of two flashes in a Bernoulli sequence. The flash appeared as a horizontal bar along the path traversed by the target. In the primary-task probe condition, subjects counted the total number of step changes to the left. Two control conditions were also included: one in which the subjects counted the probes but did not track, and a second in which subjects performed the tracking task without counting the probes.

The important findings to note in the data presented in Figure 26-10 are the monotonic relations between the tracking difficulty manipulations and the subjects' perceived ratings of difficulty, as well as those between tracking difficulty and RMS error. Both the subjective and the behavioral indices converge on the same ordering of task difficulty. However, these measures do not provide information concerning the underlying resource structure of the task.

The effect of tracking difficulty on P300 amplitude in the auditory condition (see Figure 26-11) provided results consistent with previous research (Isreal, Wickens, Chesney, & Donchin, 1980; Wickens, Heffley, Kramer, & Donchin, 1980). Thus, in the auditory condition, an increase in the difficulty of the primary task resulted in a decrease in the amplitude of the P300 elicited by the secondary-task probes. In the

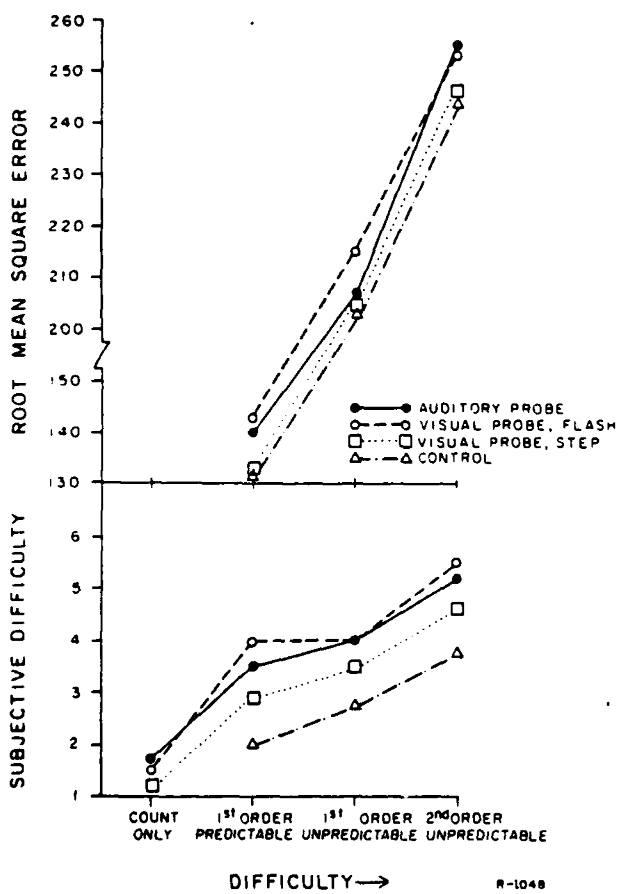


Figure 26-10. Average root mean square error and subjective difficulty ratings recorded for each condition in a pursuit step-tracking task. (Adapted from Kramer, Wickens, Vanasse, Heffley, & Donchin, 1981.)

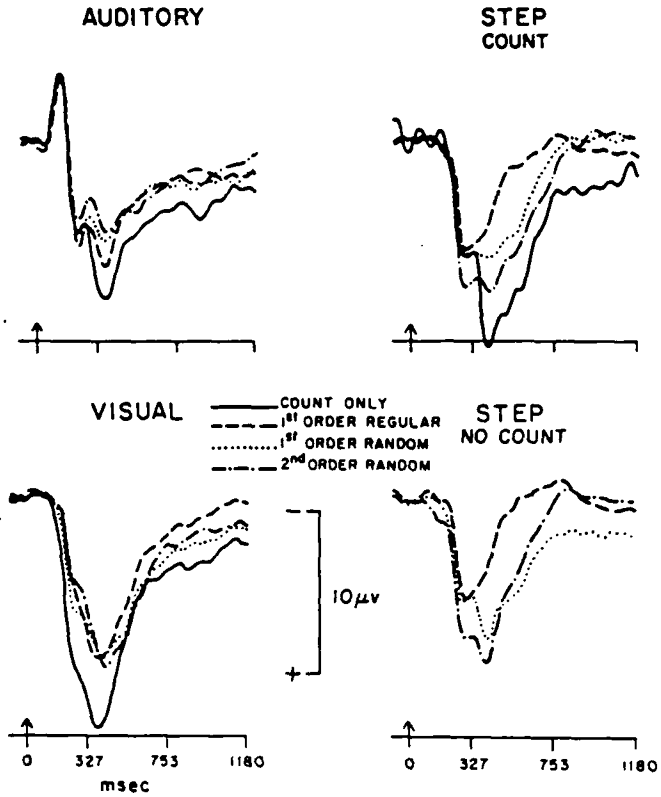


Figure 26-11. Average parietal ERPs elicited by visual, auditory, and spatial probes presented concurrently with a pursuit step-tracking task at each level of difficulty. Also shown are the ERPs elicited during single-task count conditions. (Adapted from Kramer, Wickens, Vanasse, Heffley, & Donchin, 1981.)

visual condition, the introduction of the tracking task resulted in a reduction in the amplitude of the P300. However, increases in tracking difficulty failed to produce any further attenuation. In the step conditions, the amplitude of the P300 elicited by the discrete changes in the spatial position of the controlled element increased with increments in the difficulty of the primary task. Thus, the hypothesis of resource reciprocity between the primary and secondary tasks was confirmed.

One final aspect of the step-tracking study has considerable potential practical utility. The sensitivity of the P300 elicited by visual steps to resource allocation was observed, independently of whether or not the subjects were required to count the stimuli. These data suggest that inferences from the P300 about resource allocation, and therefore about workload, can be made in the total absence of a secondary-task requirement—a considerable advantage if workload is to be assessed unobtrusively in real-time environments.

P300 and Skill Development

Another area in which ERPs (the P300 in particular) have provided useful converging evidence of a hypothetical process is the development of skill. Whether one conceptualizes the development of a skilled behavior as being a discrete two-stage process (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977) or a continuous process (Hirst, Spelke, Reaves, Caharack, & Neisser, 1980; Spelke, Hirst, & Neiser, 1976), the modulation of some hypothetical resources are usually believed to underlie the overt, measurable improvement in performance. Norman and Bobrow (1975) have argued persuasively that there are at least two distinct limits on the performance of complex tasks. The performance of resource-limited processes can benefit from an increase in the amount of processing resources allocated to the task, while the performance level attained with a data-limited process is independent of the quantity of processing resources. The performance of a data-limited process may be improved by increasing the quality of the input stimuli (signal detection limits) or by improving the memory representation of the task (memory data limits). If the P300 amplitude does in fact index the quantity of perceptual resources allocated to the performance of a task, then the modulation of resources presumed to underlie the development of a skilled behavior should be reflected in the amplitude of the P300. With increasing automaticity of a well-practiced task, the structure of the task should change from primarily resource-limited to data-limited. In terms of a single-task situation, P300s elicited by task-relevant events embedded within the primary task should decrease in amplitude, reflecting less of a demand for resources. Rosler (1981) found a sytematic decrease in P300 amplitude as subjects' performance improved during a multiblock stimulus-discrimination-learning task.

In terms of a secondary-task paradigm, P300s elicited by secondary-task probes should increase in amplitude as the primary task becomes progressively more data-limited. This variation in the amplitude of the P300 would presumably result from the increased quantity of processing resources which can be allocated to the secondary task. The target acquisition study outlined previously has found results that are consistent with this hypothesis (Kramer *et al.*, 1983). In this experiment, two groups of subjects received different levels of practice on the task: One group received 120 practice trials and the other 520 practice trials prior to ERP recording.

Behavioral indices of target acquisition performance confirmed that the highly practiced group performed substantially better than the less practiced group. As can be seen from Figure 26-12, when subjects were relatively inexperienced with the task, both manipulations of primary-task difficulty attenuated the amplitude of the P300, with smaller-amplitude P300s elicited in the alignment phase and in the second-order condition. However, when subjects were thoroughly practiced on the task, the P300s elicited by secondary-task probes were the same for different levels of workload. Thus both the behavioral and ERP measures provide evidence for a modulation in the demand for processing resources with practice.

Similar results were obtained in a study of operator workload conducted in a part-task aircraft simulator

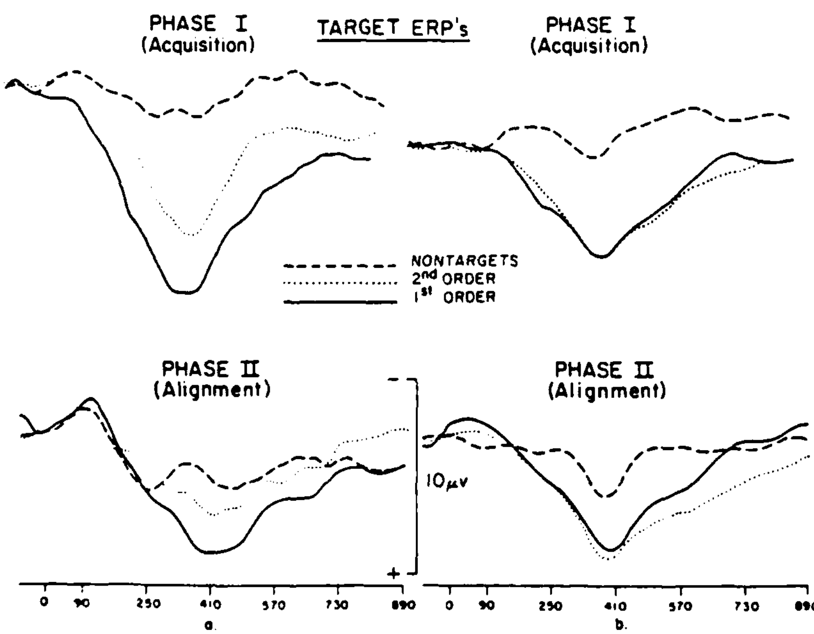


Figure 26-12. Average parietal ERPs elicited by intensifications of the tracking elements. The left panels present waveforms elicited after 120 practice trials, while the right panels present waveforms elicited after 520 practice trials. (Adapted from Kramer, Wickens, & Donchin, in press.)

(Natani & Gomer, 1981). In this case, subjects were required to fly a command flight profile and maintain air speed while concurrently performing threat avoidance and target acquisition tasks. Workload was manipulated by varying the bandwidth of the pitch- and roll-forcing functions. On the first day of the experiment, both behavioral (composite RMS error) and ERP measures indicated a substantial difference between the two workload conditions. P300 amplitude elicited by auditory probes was significantly larger in the low-workload condition. However, on the second day, subjects' performance as well as P300 amplitude indicated no differences between the two workload levels.

In both single-task and dual-task situations, P300 amplitude has been found to react systematically to changes in the skill level of the subjects. The modulation in resource requirements inferred from P300 amplitude may be useful in monitoring the skill development of human operators in complex systems. Furthermore, the sensitivity of P300 to perceptual resources provides information on a selective aspect of resource changes over practice.

SUMMARY AND CONCLUSIONS

The investigations reported above demonstrate that the P300 elicited by a secondary task can diagnostically reflect primary-task workload variations of a perceptual-cognitive nature, uncontaminated by response factors. The absence of overt response requirements provide the P300 eliciting oddball task with a considerable advantage over secondary tasks, which require frequent responses.

As a secondary task, however, the probe task is not entirely unobtrusive, and interpretation of the measures still requires the investigator to make certain assumptions about the nature of the interaction between the primary and secondary tasks in order to make inferences concerning operator workload. It is for this reason that our most recent observations that the P300 elicited by primary-task stimuli also reflects resource allocation are particularly encouraging to the utility of the ERP as a measure of workload in extralaboratory environments.

We have reviewed in this chapter studies of the ERP that have, we think, one characteristic in common. In each case, the ERP has served as a source of information on the timing or the "intensity" of an information-processing activity whose behavior is not easily monitored by means of observations on overt responses. It would seem that a science of engineering psychology that is interested in developing and testing hypotheses about the internal structure and the operating modes of the human operator would benefit from this additional information. We advocate here the use of the ERP as an analytical tool that can usefully aid in deepening our understanding of, and the measurement of, mental workload.

ACKNOWLEDGEMENTS

The work conducted at the Cognitive Psychophysiology Laboratory would not have been possible without the assistance of numerous members of the scientific and technical staff. We would like to thank Mike Anderson, Brian Foote, and Dave McWilliams for their invaluable programming and engineering help. Members of the scientific staff who have been instrumental in the research reported here include Jack Isreal, Daniel Gopher, Greg Chesney, Connie Duncan-Johnson, John Polich, Greg McCarthy, and Earle Heffley.

The research has been supported under the Office of Naval Research (Contract No. N00014-76-C-0002) with funds provided by the Defense Advanced Research Projects Agency; the Air Force Office of Scientific Research under Contract No. F49620-79C-0233; Wright-Patterson Air Force Base under Contract No. F33615-79C-0512; and the U.S. Environmental Protection Agency under Contract No. R805628010. The support of Drs. Craig Fields, Judith Daly, Al Fregley, Don Woodward, Bob O'Donnell, and Dave Otto is gratefully acknowledged.

Part of this chapter was presented at the AIAA Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics, Edwards Air Force Base, California, January 1982.

REFERENCES

Allport, D. A., Antonis, B., & Reynolds, P. On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology*, 1972, 24, 225–235.

Bertelson, P. S-R relationships and reaction time to new versus repeated signals in a serial task. *Journal of Experimental Psychology*, 1963, 65, 478–484.

Bostock, H., & Jarvis, M. J. Changes in the form of the cerebral evoked response related to the speed of simple reaction time. *Electroencephalography and Clinical Neurophysiology*, 1970, 29, 137–145.

Brown, I. D. Dual task methods of assessing workload. *Ergonomics*, 1978, 21, 221–224.

Callaway, E., Tueting, P., & Koslow, S. (Eds.). *Brain event-related potentials in man.* New York: Academic Press, 1978.

Coakley, J. D., & Fucigna, J. T. *Human engineering recommendations for the instrumentation of radar* (Report No. AN FPS-16 (Xn-2)). Moorestown, N.J.: Radio Corporation of America, 1955.

Clynes, M., & Kohn, M. The use of Mnemotron for biological data storage, reproduction, and for an average transient computer. In *Abstracts of the 4th Annual Meeting of the Biophysics Society.* Philadelphia, Pennsylvania. 1960.

Conrad, C. Some factors involved in the recognition of words. In J. W. Cotton & R. W. Klatzky (Eds.), *Semantic factors in cognition.* Hillsdale, N.J.: Erlbaum, 1978.

Donchin, E. Brain electrical correlates of pattern recognition. In G. F. Inbar (Ed.), *Signal analysis and pattern recognition in biomedical engineering.* New York: Wiley, 1975.

Donchin, E. Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), *Evoked brain potentials and behavior.* New York: Plenum Press, 1979.

Donchin, E. Surprise! . . . Surprise? *Psychophysiology*, 1981, 18, 493–513.

Donchin, E., & Isreal, J. B. Event-related potentials and psychological theory. In H. H. Kornhuber & L. Deecke (Eds.), *Motivation, motor and sensory processes of the brain: Electrical Potentials, behavioral and clinical use.* Amsterdam: North-Holland Biomedical Press, 1980.

Donchin, E., Kubovy, M., Kutas, M., Johnson, R., Jr., & Herning, R. I. Graded changes in evoked response (P300) amplitude as a function of cognitive activity. *Perception and Psychophysics*, 1973, 14, 319-324.

Donchin, E., McCarthy, G., Kutas, M., & Ritter, W. Event-related brain potentials in the study of consciousness. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self regulation*. New York: Plenum Press, 1983.

Donchin, E., Ritter, W., & McCallum, C. Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. Koslow (Eds.), *Brain event-related potentials in man*. New York: Academic Press, 1978.

Duncan-Johnson, C. C., & Donchin, E. On quantifying surprise: The variation in event-related potentials with subjective probability. *Psychophysiology*, 1977, 14, 456-467.

Folley, Jr., J. D., Altman, J. W., Graser, R., Preston, H. O., & Weislogel, R. C. *Human factors methods for system design* (Report No. AIR-B90-60-FR-225). Pittsburgh: The American Institutes for Research, 1960.

Ford, M. M., Mohs, R. C., Pfefferbaum, A., & Kopell, B. S. On the utility of P300 latency and reaction time for studying cognitive processes. In H. H. Kornhuber & L. Deecke (Eds.), *Motivation, motor and sensory processes of the brain: Electrical potentials, behavioral and clinical use*. Amsterdam: North-Holland Biomedical Press, 1980.

Ford, J. M., Roth, W. T., Mohs, R. C., Hopkins, W. F., & Kopell, B. S. Event-related potentials recorded from young and old adults during a memory retrieval task. *Electroencephalography and Clinical Neurophysiology*, 1979, 47, 450-454.

Goff, W. R., Allison, T., & Vaughan, Jr., H. G. The functional neuroanatomy of the event-related potentials. In E. Callaway, P. Tueting, & S. Koslow (Eds.), *Brain event-related potentials in man*. New York: Academic Press, 1978.

Gomer, F. E., Spicuzza, R. J., & O'Donnell, R. D. Evoked potential correlates of visual item recognition during memory scanning tasks. *Physiological Psychology*, 1976, 4, 61-65.

Heffley, E., Wickens, C. D., & Donchin, E. Intramodality selective attention and P300—reexamination in a visual monitoring task. *Psychophysiology*, 1978, 15, 269-270.

Hirst, W., Spelke, E. S., Reaves, C. C., Caharack, G., & Neisser, U. Dividing attention without alternation or automaticity. *Journal of Experimental Psychology: General*, 1980, 109, 98-117.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. *Psychophysiology*, 1980, 17, 259-273.

Isreal, J. B., Wickens, C. D., Chesney, G. L., & Donchin, E. The event-related brain potential as an index of display monitoring workload. *Human Factors*, 1980, 22, 212-224.

Jewett, D., Romano, H. W., & Williston, J. S. Human auditory evoked responses: Possible brain stem components detected on the scalp. *Science*, 1970, 167, 1517-1518.

Kahneman, D. *Attention and effort*. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Kantowitz, B. H., & Knight, J. L. On experimenter limited processes. *Psychological Review*, 1976, 83, 502-507.

Karlin, L., & Martz, M. J. Response probability and sensory evoked potentials. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press, 1973.

Karlin, L., Martz, M. J., & Mordkoff, A. M. Motor performance and sensory evoked potentials. *Electroencephalography and Clinical Neurophysiology*, 1970, 28, 307-313.

Kidd, J. S., & Van Cott, H. P. System and human engineering analysis. In H. P. Van Cott & R. G. Kinkade (Eds.), *Human engineering guide to equipment design*. Washington, D.C.: U.S. Government Printing Office, 1972.

Kinsbourne, M., & Hicks, R. E. Functional cerebral space: A model for overflow, transfer and interference effects in human performance. In J. Requin (Ed.), *Attention and performance VII*. New York: Academic Press, 1978.

Knowles, W. B. Operator loading tasks. *Human Factors*, 1963, 5, 155-161.

Kramer, A. F., Fisk, A., & Schneider, W. P300, consistency and visual search. *Psychophysiology*, 1983, 20, 453-454.

Kramer, A., Ross, W., & Donchin, E. A chronometric analysis of the role of orthographic and phonological cues in a non-lexical decision task. *Psychophysiology*, 1982, 19, 330-331. (Abstract)

Kramer, A. F., Wickens, C. D., & Donchin, E. An analysis of the processing demands of a complex perceptual-motor task. *Human Factors*, 1983, 25, 597-621.

Kramer, A., Wickens, C. D., Vanasse, L., Heffley, E. F., & Donchin, E. Primary and secondary task analysis of step tracking: An event-related potentials approach. In R. C. Sugarman (Ed.), *Proceedings of the 25th Annual Meeting of the Human Factors Society, Rochester, New York*. Rochester: Human Factors Society, 1981.

Kutas, M., McCarthy, G., & Donchin, E. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, 1977, 197, 792-795.

McCarthy, G., & Donchin, E. A metric for thought: A comparison of P300 latency and reaction time. *Science*, 1981, 211, 77-80.

McClelland, J. L. On time relations of mental processes: A framework for analyzing processes in cascade. *Psychological Review*, 1979, 86, 287-330.

McDonald, L. B. *A model for predicting driver workload in the freeway environment: A feasibility study*. Unpublished doctoral dissertation, Texas A & M University, 1973.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. Loci of contextual effects on visual word recognition. In P. Rabbit & S. Dornic (Eds.), *Attention and performance V*. New York: Academic Press, 1975.

Moray, N. Where is capacity limited?: A survey and a model. In A. F. Sanders (Ed.), *Attention and performance I*. Amsterdam: North-Holland, 1967.

Natani, K., & Gomer, F. E. *Electrocortical activity and operator workload: A comparison of changes in the electroencephalogram and in event-related potentials* (McDonnell Douglas Technical Report, MDC E2427). St. Louis: McDonnell Douglas Astronautics Company, 1981.

Navon, D., & Gopher, D. On the economy of the human processing system. *Psychological Review*, 1979, 86, 214-255.

Navon, D., & Gopher, D. Interpretations of task difficulty in terms of resources: Efficiency, load, demand and cost composition. In R. Nickerson & R. Pew (Eds.), *Attention and performance VIII*. Hillsdale, N.J.: Erlbaum, 1980.

Norman, D., & Bobrow, D. On data-limited and resource-limited processes. *Cognitive Psychology*, 1975, 7, 44-64.

North, R. Task functional demands as factors in dual-task performance. In *Proceedings of the 21st Annual Meeting of the Human Factors Society, Santa Monica, California*. Santa Monica: Human Factors Society, 1977.

Ogden, G. D., Levine, J. W., & Eisner, E. J. Measurement of workload by secondary tasks. *Human Factors*, 1979, 21, 529-548.

Otto, D. A. (Ed.). *Multidisciplinary perspectives in event-related brain potential research* (EPA 600/9-77-043). Washington, D.C.: U.S. Government Printing Office, 1978.

Picton, T. W., & Stuss, P. T. The component structure of the human event-related potentials. In H. H. Kornhuber & L. Deecke (Eds.), *Motivation, motdor and sensory processes of the brain: Electrical potentials, behavioral and clinical use*. Amsterdam: North-Holland Biomedical Press, 1980.

Polich, J. M., McCarthy, G., Wang, W. S., & Donchin, E. When words collide: Orthographic and phonological interference during word processing. *Biological Psychology*, 1983, 16, 155-180.

Rasmussen, J. Models of mental strategies in process plant diagnosis. In J. Rasmussen & W. B. Rouse (Eds.), *Human detection and diagnosis of system failures*. New York: Plenum Press, 1981.

Regan, D. *Evoked potentials in psychology, sensory physiology and clinical medicine*. London: Chapman & Hall, 1972.

Ritter, W., Simson, R., & Vaughan, H. G., Jr. Association cortex potentials and reaction time in auditory discriminations. *Electroencephalography and Clinical Neurophysiology*, 1972, 33, 547-557.

Rohrbaugh, J. W., Donchin, E., & Eriksen, C. W. Decision making and the P300 component of the cortical evoked response. *Perception and Psychophysics*, 1974, 15, 368–374.

Rolfe, J. M. The secondary task as a measure of mental load. In W. T. Singleton, J. G. Fox, & D. Witfield (Eds.), *Measurements of man at work*. London: Taylor & Francis, 1971.

Rösler, F. Event-related brain potentials in a stimulus discrimination learning paradigm. *Physiological Psychology*, 1981, 18, 447–455.

Sanders, A. F. Some variables affecting the relation between relative stimulus frequency and choice reaction time. In A. F. Sanders (Ed.), *Attention and performance III*. Amsterdam: North-Holland, 1970.

Sanders, A. F. Some remarks on mental load. In N. Moray (Ed.), *Mental workload: Its theory and measurement*. New York: Plenum Press, 1979.

Satterfield, J. H. Evoked cortical response enhancement and attention in man: A study of responses to auditory and shock stimuli. *Electroencephalography and Clinical Neurophysiology*, 1965, 19, 470–475.

Schneider, W., & Shiffrin, R. M. Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review*, 1977, 84, 1–66.

Sheridan, T. B. Mental workload—What is it? Why bother with it? *Human Factors Society Bulletin*, 1980, 23, 1–2.

Sheridan, T. B. Understanding human error and aiding human diagnostic behavior in nuclear power plants. In J. Rasmussen & W. B. Rouse (Eds.), *Human detection and diagnosis of system failures*. New York: Plenum Press, 1981.

Sheridan, T. B., & Simpson, R. W. *Toward the definition and measurement of mental workload of transport pilots* (FTL Report R79-4). Cambridge, Mass.: Massachusetts Institute of Technology Flight Transportation Laboratory, 1979.

Shiffrin, R. M., & Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 1977, 84, 127–190.

Shulman, H. G., Hornak, R., & Sanders, E. The effects of graphemic, phonetic and semantic relationships on access to lexical structures. *Memory and Cognition*, 1978, 6, 115–123.

Shwartz, S. P., Pomerantz, J. R., & Egeth, H. E. State and process limitations in information processing: An additive factors analysis. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, 3, 402–410.

Sokolov, E. W. The modeling properties of the nervous system. In I. Maltzman & K. Cole (Eds.), *Handbook of contemporary Soviet psychology*. New York: Basic Books, 1969.

Spelke, E. S., Hirst, W. C., & Neisser, W Skills of divided attention. *Cognition*, 1976, 4, 215–230.

Sperando, J. C. The regulation of working methods as a function of workload among air traffic controllers. *Ergonomics*, 1978, 21, 193–202.

Squires, N. K., Donchin, E., Squires, K. C., & Grossberg, S. Bisensory stimulation: Inferring decision related processes from the P300 component. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, 3, 299–315.

Sternberg, S. High speed scanning in human memory. *Science*, 1966, 153, 652–654.

Sternberg, S. The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 1969, 30, 276–315. (a)

Sternberg, S. Memory scanning: Mental processes revealed by reaction time experiments. *American Scientist*, 1969, 57, 421–457. (b)

Sutton, S., Braren, M., Zubin, J., & John, E. R. Evoked potential correlates of stimulus uncertainty. *Science*, 1965, 150, 1187–1188.

Trumbo, D., Noble, M., & Swink, J. Secondary task interference in the performance of tracking tasks. *Journal of Experimental Psychology*, 1967, 73, 232–240.

Vidulich, M., & Wickens, C. D. *Time-sharing, manual control and memory search: The joint effects of input and output modality competition, priorities and control order* (Technical Report EPL-81-4/ONR-81-4). Urbana-Champaign: Engineering Psychology Research Laboratory, University of Illinois, 1981.

Wickelgren, W. Speed accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 1977, 41, 67–85.

Wickens, C. D. Human workload measurement. In N. Moray (Ed.), *Mental workload: Its theory and measurement*. New York: Plenum Press, 1979.

Wickens, C. D. The structure of attentional resources. In R. Nickerson & R. Pew (Eds.), *Attention and performance VIII*. Hillsdale, N.J.: Erlbaum, 1980.

Wickens, C. D., & Derrick, W. *The processing demands of higher order manual control: Application of additive factors methodology* (Technical Report EPL-81-1/ONR-81-1). Urbana-Champaign: Engineering Psychology Research Laboratory, University of Illinois, 1981.

Wickens, C. D., Derrick, W. D., Micallizi, J., & Berringer, D. The structure of processing resources. In R. E. Corrick, E. C. Haseltine, & R. T. Durst (Eds.), *Proceedings of the 24th Annual Meeting of the Human Factors Society, Los Angeles, California*. Los Angeles: Human Factors Society, 1980.

Wickens, C. D., Gill, R., Kramer, A., Ross, W., & Donchin, E. The cognitive demands of second order manual control: Applications of the event-related brain potential. In *Proceedings of the 17th Annual NASA Conference on Manual Control* (NASA TM). Washington, D.C.: National Aeronautics and Space Administration, 1981.

Wickens, C. D., Heffley, E. F., Kramer, A. F., & Donchin, E. The event-related brain potential as an index of attention allocation in complex displays. In G. E. Corrick, E. C. Haseltine, & R. T. Durst (Eds.), *Proceedings of the 24th Annual Meeting of the Human Factors Society, Los Angeles, California*. Los Angeles: Human Factors Society, 1980.

Wickens, C. D., Isreal, J. B., & Donchin, E. The event-related cortical potential as an index of task workload. In *Proceedings of the 21st Annual Meeting of the Human Factors Society, Santa Monica, California*. Santa Monica: Human Factors Society, 1977.

Wickens, C. D., & Kessel, D. The effect of participatory mode and task workload on the detection of dynamic system failures. *IEEE Transactions on Systems, Man and Cybernetics*, 1979, SMC-13, 24–34.

Wierwille, W. W., Gutmann, J. C., Hicks, T. G., & Muto, W. H. Secondary task measurements of workload as a function of simulated vehicle dynamics and driving conditions. *Human Factors*, 1977, 19, 557–566.

Wilkinson, R. T., & Morlock, Jr., H. C. Auditory evoked response and reaction time. *Electroencephalography and Clinical Neurophysiology*, 1967, 23, 50–56.

Williges, R. C., & Wierwille, W. W. Behavioral measures of aircrew mental workload. *Human Factors*, 1979, 21, 549–574.

Wood, C. C., & Allison, T. Interpretation of evoked potentials: A neurophysiological perspective. *Canadian Journal of Psychology*, 1981, 35, 113–135.

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

THE ENDOGENOUS COMPONENTS OF THE EVENT-RELATED POTENTIAL - A DIAGNOSTIC TOOL?

Emanuel Donchin, Gregory A. Miller, and Lawrence A. Farwell

Dept. of Psychology and Cognitive Psychophysiology Laboratory

University of Illinois at Urbana-Champaign

Champaign, Illinois (USA)

The research reported here was supported principally by NIA grant AG03151 and also by AFOSR Contract #F49620-79-C-0233, Al Fregly, Project Manager. Theodore R. Bashore, Noel K. Marshall, Ron D. Chambers, Frank Morrell, Leyla DeToledo-Morrell, and Thomas Hoeppner were collaborators during various phases of the project. Kay Strayer's assistance in the execution of the project is greatly appreciated. The authors also wish to acknowledge the subjects who made this work possible.

To appear in: E. Fliers (Ed.), Progress in Brain Research. Amsterdam: Elsevier.

Abstract

The use of the endogenous components of the Event-Related Brain Potential (ERP), and in particular of the P300, in the study of aging is reviewed. We contrast the nosological approach, in which diagnostic tools are developed by seeking reliable differences in the ERPs recorded from different diagnostic groups, with a theory-based approach in which attention is focused on features of the ERPs by an interaction between models of human information processing and models of the ERP. We review, from this perspective, the use of the latency of the P300 in the diagnosis of dementia. We then report finding a group of elderly subjects in whom P300 has proven to be particulalry small. These subjects also display a specific deficit in the maintenance of Working Memory. This study illustrates the process by which psychological and psychophysiological models can play complementary roles as the ERPs are used as tools in the study of cognitive function.

Introduction

Goodin et al. (1978b) reported that a brain response labeled the P300 is elicited with a substantially delayed latency in demented patients. Briefly described, they presented their subjects with a series of tones and recorded from the surface of the scalp an electrical response of the brain to these tones, using a technique requiring extensive computer analysis. This response takes the form of a sequence of peaks and troughs and is known as the Event-Related Brain Potential (ERP). Goodin et al. (1978b) selected one specific peak, the P300, which is positive in polarity and which occurs at least 300 msec following the eliciting tone. This interval, the "latency" of P300, is of particular interest because it seems to be useful as a tool in mental chronometry. There is considerable evidence that the latency of P300 increases with age (Ford et al., 1979; Pfefferbaum et al., 1980a, 1984a; Brown et al., 1983; Mullis et al., 1984; Picton et al., 1984). Goodin et al. (1978a,b) provided additional support for this thesis. However, the more striking aspect of their results was that the latency of the P300 was extraordinarily long only in the demented subjects, not in subjects showing very similar overt symptoms due to depression rather than dementia. These data implied that the P300 may serve as a tool in differential diagnosis.

The degree to which the increase in P300 latency is indeed a specific indicator of senile dementia has proven somewhat controversial (Brown et al., 1982; Pfefferbaum et al., 1984b; Slaets & Fortgens, 1984; Polich et al., in press). It so happens that increases in P300 latency do tend to occur in association with other pathologies (Roth et al., 1979; Baribeau-Braun et al., 1983). Furthermore, conflicting results have been reported by other investigators (Slaets & Fortgens, 1984). However this matter is resolved, there is an aspect of Goodin et al.'s (1978b) contribution that is noteworthy and on which we wish to focus this review of the role that Cognitive Psychophysiology can play in the study of aging. We refer to the theory-based rather than the nosological foundation of the study.

The nosological approach is a rather common, and in many cases beneficial, procedure for developing diagnostic applications of psychophysiological and other measures (see for example Halliday, 1978). The starting point for a nosological study is the availability of groups of clinically diagnosed patients as well as an adequate control group of non-patients. The strategy is empirical. Given that the studied groups are known to be proper representatives of the diagnostic classes of interest, then any measure that discriminates between the groups is potentially useful. This approach has, for example, been of great benefit in developing a diagnostic measure for multiple sclerosis based on the latency of brain responses elicited by moving checkerboards (Halliday, 1973). However, the success of the nosological approach is contingent on the specificity of the deficit and the certainty with which patients can be diagnosed clinically.

Matters have proven considerably more complex when the same approach has been applied in the analysis of components of the ERP which are not simple, obligatory perceptual responses (exogenous components) but reflect higher-order processes (endogenous components). Consider, for example, the P300, with which this paper is primarily concerned. Evidence has accumulated in the past two decades that the P300 is smaller in amplitude, and generally longer in latency, in most nosological groups investigated. Low amplitude P300 has been reported for the mentally retarded (N. Squires et al., 1979), for the schizophrenic (Pfefferbaum et al., 1984b), for the depressed (Roth et al., 1981; Pfefferbaum et al., 1984b), for the alexic (Neville et al., 1979), and for the alcoholic (Porjesz & Begleiter, 1982), to name but a few. Thus, to observe a difference between any nosological group and a control group does not necessarily imply that the primary diagnostic criterion according to which the groups were constructed is indeed responsible for the observed difference in the dependent measure. There may well be a non-specific deficit that the group studied shares with many other daignostic groups.

The nosological approach can be substantially enriched if its empirical observation of differences can be augmented by a theoretical understanding of the functional significance of the psychophysiological observations and of the underlying neurophysiology. The choice of the latency of the P300 as a measure for examination in dementia has been driven by the extensive evidence that this latency can serve as an index of mental timing that is not contaminated by motor factors (Donchin, 1975, 1981; Kutas et al., 1977; McCarthy & Donchin, 1979, 1981; Donchin & McCarthy, 1980). Within this conceptual framework, Goodin et al.'s (1978a,b) work assumes increased value, because it goes beyond asserting merely that a group of demented patients are deviant on yet another measure. Rather, the data are interpreted as supporting an assertion about the nature of the deficit in a manner that specifies directions for further study. A similar use of the theory presented by Donchin and associates to account for variance in P300 latency has been made by Ford et al. (1979), who inferred from an analysis of the P300 latency in a test of short term memory that elderly subjects are slowed, in such a test, by motor rather than by cognitive factors.

Our purpose in this chapter is to introduce the conceptual foundations of the theory-based approach to the use of ERPs, in contrast to the nosological approach. In the course of this review we will survey currently available information on the P300 that is recorded in aged subjects. We will conclude this chapter with the description of a series of studies of P300 in aged subjects conducted in our laboratory that serves to illustrate our approach to the use of ERPs in Cognitive Psychophysiology.

The Status of P300 in Psychological Theory

As is well known, the ERP is a sequence of voltage oscillations, recorded from the scalp, that are time-locked to an event. It is extracted from the electroencephalographic record by means of Signal Averaging (see Regan, 1972, or Callaway et al., 1978, for a technical introduction to ERPs; Hillyard & Kutas, 1983, for a review of current cognitive research with ERPs; and Coles et al., in press, for a general treatment of this methodology). The ERP is generally parsed into different "components". Components are defined in terms of their polarity (positive or negative voltage), latency (temporal relationship to the event), and topography (variation in voltage with electrode location on the scalp), as well as by their relationship to experimental variables. Components can be quantified using simple magnitude measures or through the application of more elaborate techniques such as Principal Component Analysis (PCA) and Vector Analysis (Gratton et al., 1984; see also Gratton et al., submitted for publication). Component labels are generated by a polarity descriptor and a characteristic latency descriptor. Thus, the P300 is a positive ERP component with a modal latency of 300 msec. In some cases, as with Contingent Negative Variation (CNV) and Slow Wave (SW), the latency descriptors are omitted.

The assumptions and the model underlying our study of ERPs have been presented elsewhere (Donchin, 1979, 1981). In brief, we assume that the voltages we record at the scalp are the result of synchronous activation of neuronal ensembles whose geometry allows their individual fields to summate to a field whose strength can affect scalp electrodes (Galambos & Hillyard, 1981). As explained above, it is useful to parse the ERP into a set of components. The component, in our scheme of things, is characterized by its consistent response to experimental manipulations (see Donchin et al., 1978, for a discussion of components). We further assume that each component is a manifestation at the scalp of an intracranial processing entity. We are not implying that each ERP component corresponds to a specific neuroanatomical entity or that the activity manifested by the component corresponds to a distinct neural process. Rather, we assume that a consistent information processing need, characterized by its eliciting conditions, activates a collection of processes which, for perhaps entirely fortuitous reasons, have the biophysical properties that generate the scalp-recorded activity. As a working hypothesis, we postulate that ERP components are manifestations of functional processing entities that play distinct roles in the algorithmic structure of the information processing system. In other words, we believe that it is possible to describe in detail the transformations that the processing entity applies to the information stream. The goal of Cognitive Psychophysiology, within this framework, is to provide such detailed descriptions. This may be achieved by developing comprehensive descriptions of the conditions governing the elicitation and attributes of the components (the "antecedent" conditions). These descriptions can be used to support theories that attribute certain functions to the "subroutine" manifested by a component. In turn, the theories should lead to predictions regarding the consequences of the elicitation of the "subroutines", predictions which can be tested empirically.

The ensemble of ERP components is rich in members, from the early brain stem potentials through general components such as N100, P200, N200, P300, the Slow Wave, and several event-preceding negativities. Each of these merits a detailed review, and with respect to each some work pertaining to aging has been done. However, both because our own work has focused on the P300, and because the scope of this chapter must be limited, we will restrict ourselves to the P300 component.

The most commonly used experimental context in the study of the P300 is the oddball paradigm (Donchin et al., 1978). The subject is presented with a series of stimuli that can be classified into one of two categories. The instructions are either to count, or to respond in some other manner, to items from one of the categories. An alternate response (which may be "ignore") is required to elements from the other category. It is typically the case that occurrences in one of the categories are rare (the "oddball"). There is an abundance of evidence that rare events elicit a large P300, whose amplitude tends to be inversely related to the subjective probability of the eliciting stimulus. However, probability is neither a necessary nor a sufficient condition for the elicitation of a P300. The task relevance of events is at least as important a determinant of the amplitude of P300 (Donchin, 1981). Figure 1 presents an example of ERPs elicited in young and elderly subjects in four different oddball tasks (Marshall et al., 1983; described in detail in the section below on P300 and Working Memory in the Elderly). The P300 is the positive deflection (downward in the figure) between 300 and 600 msec post-stimulus. Note the comparatively large P300s in the Target conditions, which are either rare, task-relevant, or both, and the smaller and longer-latency P300s in the elderly subjects.

In any usage of the ERP, or of a specific component of the ERP, these

---

Insert Figure 1 About Here

---

complex waveforms need to be converted in some manner into summarizing measures. The two attributes of the P300 which are generally quantified in this manner are the "amplitude" and the "latency" of the component. We place these terms in quotes because there is a need to distinguish between the use of these terms as referents to conceptual attributes and their use as referents to self-evident features of the waveform. In most studies, "latency" refers to the interval between stimulus onset and the peak of the wave which is identified with the P300. This is a self-evident feature which can be measured with relative ease. Conceptually, however, this feature is taken as a function of the interval between the event eliciting the P300 and the occurrence of the information processing manifested in P300. This conceptual latency can be represented by many different measures. The onset of the positive deflection, for example, is as obvious a measure of latency as is the peak of the positivity. The choice of the peak derives from the larger reliability with which it can be measured. Such choices in measurement must be made in any investigation. Yet it must be remembered that these are in many ways arbitrary choices. As the different measures are not perfectly correlated it is conceivable that studies will conflict because they have used different features of the waveform to represent the conceptual attributes.

With this caveat in mind, it is possible to discuss the data that have accumulated to date, mostly from samples of young adults, regarding the behavior and possibly the functional significance of the P300. Detailed reviews can be found elsewhere (Pritchard, 1981; Hillyard and Kutas, 1983; Donchin, in press). Here, we shall only note that the amplitude can be interpreted as a measure of the activation of the intracranial processor whose activity is manifested by the P300 (Donchin et al., 1973; Donchin, 1981). The amplitude appears to increase with the task relevance and the improbability of the eliciting event (Pritchard, 1981). The relationship between task relevance and P300 amplitude has led to an extensive series of studies assessing the degree to which P300 can be used as a measure of Mental Workload (Isreal et al., 1980; Kramer et al., 1983; Sirevaag et al., 1984; see also Gopher & Donchin, in press).

The latency of the P300, on the other hand, has been shown to depend on the duration of information processing activities that must precede the elicitation of the P300. In the main, an entire class of variables that are known to either retard or to speed up reaction time also have a similar effect on P300 latency (Ritter & Vaughan, 1969; Kutas et al., 1977; Ragot & Renault, 1981; McCarthy, 1984). However, it turns out that there is a set of variables known to have a very large effect on reaction time that do not seem to affect, or at most have a microscopic effect, on P300 latency (McCarthy & Donchin, 1981; Duncan-Johnson & Donchin, 1981, 1982; Magliero et al., 1984; Coles et al., submitted for publication). In the main the variables that affect both P300 latency and reaction time are clearly associated with the evaluation of stimulus and response contingencies. Those variables that affect primarily RT and do not affect P300 latency appear to be related primarily to motor processes. It is this distinction and the functional interpretation it implies that underlies the use of P300 latency in studies of aged subjects.

P300 in Healthy and Demented Adults: Basic Findings

A number of studies have investigated the relationship between P300 latency and aging. The consensus of these studies seems to be that, in adults, P300 latency in both auditory and visual sensory modalities increases with age. A number of studies comparing groups of young vs. elderly adults have found longer latencies in the elderly (e.g., Ford et al., 1979, 1982a; Pfefferbaum et al., 1980a, 1984a). Several studies have indicated that P300 increases linearly in latency from early adulthood through old age, with a slope of between 1 and 2 msec/year (e.g, Syndulko et al., 1982, 1.1 msec/year; Brown et al., 1983, 1.12 msec/year; Picton et al., 1984, 1.36 msec/year; K. Squires et al., 1980, 1.64 msec/year; Goodin et al., 1978a, 1.8 msec/year). Some studies have found that the slope of the P300 latency curve increases with age. Beck et al. (1980) found twice the slope in the 63-79 age range as in the 28-63 age range (1.6 vs. 0.8 msec/year). Brown et al., (1983) fitted the latency/age function with a curvilinear first- and second-degree orthogonal polynomial, indicating a positively accelerating function. Mullis et al. (1985) fitted data on 108 subjects aged 8 to 90 with a similar function (the function included decreasing latencies with age in subjects between 8 and 22 years).

Although the increase of P300 latency with age appears to be a robust observation, this finding has not been accepted without controversy. Some investigators have failed to observe such an increase in P300 latency with age. Podlesny and Dustman (1982) found no age effects on P300 latency in visually signaled simple and choice reaction time tasks. Snyder and Hillyard (1979) report a constant latency, with age, of P300 elicited by visual targets or novel visual stimuli. Michaelewski et al. (1982) and Picton et al. (1984) found no significant increase with age in P300 latency in response to an omitted stimulus. Moreover, researchers have found a considerable variability in P300 latency within different age groups. Typically, the range of latency within each age group has been as large or nearly as large as the range across groups. Thus, no particular latency range specifically identifies a particular age group, although Goodin et al., (1978b) did report that their demented patients fell outside the normal range for the elderly.

Results on age-related changes in P300 amplitude are less clear than those relating to latency. Some investigators report decreases in P300 amplitude with age (Goodin et al., 1978a; Podlesny & Dustman, 1982; Brown et al., 1983; Picton et al., 1984, Podlesny et al., 1984; Mullis et al., 1985), while others report that P300 amplitude remains constant with age (Beck et al., 1980; Pfefferbaum et al., 1980a, 1984a; Ford et al., 1982b).

Several studies have found changes in the scalp distribution of P300 with age (Smith et al., 1980; Pfefferbaum et al., 1980b, 1984a; Picton et al., 1984). The P300 is parietally maximal in younger subjects but seems to shift frontally with age. Wickens et al. (1985; see also Braune et al., in press; Strayer et al., 1985), report a study of 60 subjects ranging in age from 20 to 64 years, in which the amplitude of P300, averaged over three scalp recording sites--Fz (frontal), Cz (central), and Pz (parietal)--did not change with age. However, the difference between Pz amplitude and Fz amplitude decreased significantly. Both an increase in Fz amplitude and a decrease in Pz amplitude with advancing age contributed to this result.

As mentioned above, it has been proposed (Goodin et al., 1978b; Squires et al., 1979, 1980) that prolonged P300 latency may be a clinically useful tool in the differential diagnosis of dementia. The initial studies of P300 latency and dementia, as well as several subsequent studies (Syndulko et al., 1982; Brown et al., 1982; Polich et al., in press) used a two-stimulus oddball paradigm, with the subjects required to count or otherwise respond to the rare events. All of these studies found prolonged P300 latency in demented patients. Polich et al. (in press) also reported a positive correlation between P300 latency and the degree of cognitive impairment.

Pfefferbaum and colleagues (Pfefferbaum et al., 1984a,b) conducted an extensive study comparing 135 normal controls ranging in age from 18 to 90 with demented patients as well as non-demented, cognitively impaired patients, schizophrenics, and depressives. Both auditory and visual modalities were employed. Both rare target (subjects responded with a button press) and rare non-target (no response required) as well as frequent stimuli were included. P300s were elicited by both types of rare stimuli.

This study found that dements did exhibit significantly prolcnged P300 latencies when compared with normal subjects for both target and non-target rare stimuli in both auditory and visual modalities. However, schizophrenics also showed significantly prolonged P300s. Non-demented, cognitively impaired patients and depressives seemed to exhibit somewhat longer P300 latencies than normals, but did not show as great an effect as the demented patients. Latency variability was significantly greater in demented patients and also in schizophrenics.

Amplitude effects were also observed in this study. Demented patients showed significantly diminished P300 amplitude for almost all conditions. Schizophrenics and depressives who were not under medication appeared to show a similar but less pronounced effect. Both demented and schizophrenic patients also were slower than controls in RT.

This important study illustrates the difficulties in attempting to employ ERPs to distinguish clinically defined populations. Although significant relationships were found, the lack of specificity of the effects makes application of these results in differential diagnosis problematic at best.

Moreover, not all studies have found P300 latency differences in demented patients. In a study of 42 demented elderly patients, 29 non-demented elderly patients, and 10 healthy young controls, Slaets and Fortgens (1984) found no significant difference in P300 latency between demented and non-demented patients performing an auditory oddball counting task.

Of course, in considering this seemingly conflicting literature, it is necessary to note that all too often studies designed as putative replications of previous work differ in important details. The pattern of ERPs one obtains in any experimental paradigm is enormously sensitive to subjects' perception of the task and the range of strategies that are employed by the subject. It is crucial, therefore, that investigators take pains to ensure that when they claim to replicate a study they have indeed done so in the formal sense of the word "replicate." Thus, for example, if an original study reports a pattern of results that was obtained when a sequence of stimuli was randomly selected, then any study that imposes constraints on the stimulus sequence is, by definition, not a replication. Failure to confirm the original results in such cases is not entirely surprising. It is in such a context that one must, for example, evaluate the implication of Pfferfferbaum et al.'s (1984b) failure to replicate Goodin et al. (1978b).

We cannot resolve the dispute regarding P300 and dementia in this chapter. Clearly, more data are needed. To be useful, such data must be obtained consistently and reliably. Yet methodological purity, while necessary, is not sufficient to ensure the utility of the results. There remains a need for a theoretical framework within which it may be possible to make more sense of the data. Although Goodin et al. (1978a,b) proceeded from a theoretical interpretation of P300 latency, their procedure was largely nosological. Subjects were classified according to age and medical history, and differences between the groups were sought. The debate in the literature boils down to a dispute about the degree to which groups labeled clinically do indeed show a consistent pattern of ERPs. However, the importance of the issue depends on the degree to which one accepts that the groups compared are indeed comparable and the procedures used with the different groups commensurate. This approach assigns much weight to standard classification and diagnostic criteria, or to chronological age. We are not persuaded that this is a good strategy. Indeed, in our own work with the aged, which we will now discuss in some detail, we found it impossible to organize the results in a meaningful way by relying on standard tests and commonly used diagnostic criteria. We were forced to rely on our theory of the functional significance of the P300, and it was this theoretical approach that allowed us to organize the data (Marshall et al., 1983; Farwell et al., 1985).

P300 and Working Memory in the Elderly

The study reviewed in this section was undertaken in the context of an investigation designed originally to use P300 latency in an attempt to identify the locus of mental slowing in the aged. The intent was to determine the stages of processing at which P300 latency is particularly lengthened in the elderly. As the study depended on our ability to run elderly subjects in oddball paradigms, we began, merely as a feasibility study, by asking a group of elderly subjects to participate in several variations of the oddball paradigm. We chose the following four tasks:

(1) Count: The subject counted one of two easily discriminable, equally probable tones (1000 and 1500 Hz). At the end of each block of 100 trials, the subject reported the number of target stimuli. This task provided a pure assessment of the target effect of P300, uncontaminated by rareness factors or overt response demands.

(2) Choice Reaction Time (CRT): The second oddball task required a choice reaction time response. Two tones (1000 Hz, 20% probability, and 1500 Hz, 80%) corresponded to two microswitches, one under each thumb. This task highlighted the probability effect, holding task relevance and motor demands constant.

(3) Omitted Stimuli: In this task, one of the two classes of "stimuli" was actually the *non-occurrence* of a tone. Thus, the tone was omitted on 10% of the trials. A 1000 Hz tone occurred on the remaining trials. The subject was asked to count the number of omitted stimuli. Such a paradigm facilitates study of endogenous components (such as P300) without troublesome overlap from exogenous components.

(4) Names: Male (20%) and female (80%) names appeared on a computer-controlled video display, and the subject was asked to count the number of male names. This provided an assessment of P300 in a second modality.

Fifty-three generally healthy individuals (38 females) living in the community, aged 60 through 82, were paid for participating in each of the four studies. In the main, the results indicated that these subjects yielded an orderly data set in which rare stimuli elicited a clear and "normal-looking" P300, albeit with a somewhat longer latency than we have observed in young adults. Perusal of the data, however, revealed one striking phenomenon. In about one sixth of these individuals the P300 appeared to be absent. The remarkable aspect of these data was that the P300 was absent or very small in all four tasks for this subset of subjects. This was especially surprising as, in general, it is very easy to observe a P300 in such experiments, based on the hundreds of young adults we have run in similar studies. Occasionally, a subject will show no P300, or a very small P300, in some experimental condition. However, invariably these subjects will produce a perfectly normal P300 in another experimental paradigm. Yet, here we had individuals who seemed to perform the tasks normally and yet had little if any P300.

To test the reliability of these observations, we repeated the four oddball tests with half of the original sample (n=27). Nine of these were among the subjects who displayed very low P300s (averaged across the four rare or target conditions), 9 were those with the highest P300s, and 9 were randomly chosen from the remaining subjects. In addition, we ran a sample of unselected young adults (n=19, 18-30 years) through the same tasks. It turned out that the low-P300 elderly subjects whose ERPs were consistent in lacking a P300 across tasks were also quite consistent across time. Test-retest correlations for P300 amplitude to target stimuli for the four tasks were as follows: Count, .51; CRT, .73; Omitted Stimuli, .44; Name, .66.

Our comparison of 19 young adults with the full elderly sample yielded findings consistent with those recorded in other laboratories, reviewed earlier. There was a highly significant P300 latency difference between our aged (539 msec) and young (455 msec) groups, averaged across tasks and target/non-target. This is a slowing of 1.77 msec/year, close to values found by other investigators.

On the average, P300 amplitude was also found to be lower in the elderly. Across tasks and conditions, the aged group's P300 was only 63% as large as that of the young group. (Note that this finding was not consistent across subjects: the average of the elderly subjects' P300 amplitude includes those elderly individuals who exhibited little or no P300, as well as those with a more normal response.) Young subjects also showed a larger enhancement to target stimuli than did the elderly. P300s to target stimuli were 49% larger than P300s to non-target stimuli in the young subjects but only 37% larger in the elderly.

Reaction time data were obtained in the 20/80 auditory choice RT oddball task. While the target/non-target P300 latency effect in the elderly group was 20 msec (463 vs. 443), the RT effect was more than twice as large (434 vs. 385). Under an additive factors model, the extra delay in RT indicates that the RT effect reflects not only a delay in stimulus categorization (the P300 effect) but an additional delay in response selection and execution (by "response selection and execution" we refer specifically to motor processes).

Thus, our preliminary data appeared to display a latency pattern that was consistent with the literature. Yet we had a subset of subjects with especially small and in some cases entirely absent P300. It was obviously necessary, if these data were to be interpretable as more than an empirical curiosity, that we determine what, if anything, functionally distinguished the low-P300 subjects from the rest of the sample. The standard operating procedure when one is confronted with individuals who consistently deviate from some norm is to assess what other variables may distinguish this group from the others. We tried a variety of metrics in the attempt to differentiate the two groups. Medical records were examined, data on educational background were collected, and numerous psychometric, sensory and other tests were administered to the subjects. The exercise proved futile. None of these conventional approaches could serve to characterize the subjects who had no P300. It seemed that the only way in which these subjects differed from the norm is in the size of their P300s.

As we despaired of making sense of the data using the standard empirical approach there remained one other possibility. We had tried a wide range of clinical tests, for none of which was there an a priori basis to believe that it would predict a reduction in P300 amplitude. This state of affairs was a reflection of the fact that we had no theoretical framework from which to derive predictions regarding the effect that the given values of any of these measures of individual attributes will have on the the amplitude of the P300. An alternate approach would begin from a theory of the P300 and ask what subject attributes should correlate with its presence. In other words, if the P300 is a manifestation of a subroutine with some specific information processing mission, then one should be able to predict what functional difference should appear in subjects in whom this subroutine appears to be abnormally quiescent. In the next section we review briefly the evidence supporting a hypothesis concerning the functional significance of the P300. This theoretical structure led to an experiment yielding data that do indeed indicate a difference in the information processing system of the subjects with and those without the P300.

The Functional Significance of P300

Our attempt to identify the manner in which the low-P300 subjects differed from the subjects who displayed a normal P300 was guided by the proposition that the P300 is a manifestation of the processes associated with the updating of a cognitive schema or model (Donchin, 1981). This model of the P300 had developed within the context of a more general theory, proposed by several authors, ranging from Sokolov (1969) to Baddeley (1974, 1981) which assumes that some aspects of the information in memory are more readily available to processing than others. The more readily available segment is shaped by the tasks performed by the organism and can be viewed as a model of the current environment, or as a scratch pad in which currently important information is kept. This "Neuronal Model" of the current context and its associated representations, to use Sokolov's term, appears equivalent, at least in part, to the concept of "Working Memory". In adopting this concept in the interpretation of the P300, we need concentrate on only one aspect of this Working Memory, namely that the system must continually update and revise this model. If Working Memory, or the Neuronal Model, is to be useful in the performance of the current task under current circumstances, then, even as tasks and circumstances change, so must the model change. Thus, it is plausible to assume that if such a Working Memory exists there must be a set of processes that maintain it. In other words, regardless of our view of the organization and processes that characterize Working Memory, we have to assume that its representations are continually revised. A context-updating process must be included in any system that is context-sensitive. We suggest that it is this updating process which is manifested by the P300.

We are assuming here that the system maintains an activated representation of its schema and that this representation is utilized whenever action is required. Mismatches between actual inputs and schema-based expectations generally drive the organism to action which is, in some sense, reflected by the complex of activities referred to as the Orienting Response. Thus, for example, whenever a mismatch is detected between the schema and an input, it is necessary that this very fact be registered in the schema. Repeated mismatches between expectations and events must be capable of forcing a revision of the schema. Otherwise, the schema will soon fail to be a realistic representation of the environment. That such updating and revision do occur is evidenced by the process of Habituation.

Several studies showing that variables which affect P300 amplitude also influence the strength of a representation in Working Memory support this view. The effect of stimulus probability on the P300 has been modeled by K. Squires et al. (1976) as a result or the summation of decaying traces of past occurrences of each stimulus. This model implies that the amplitude of the P300 elicited by a task-relevant event depends on the interval between repetitions of that event. Presumably, the time period between repetitions of task-relevant events influences the strength of the representation available in Short-Term Memory (STM). If the representation is weak, more updating must be done following target presentation. From these studies, then, it may be inferred that P300 amplitude is related to the amount of updating which is required by a task-relevant event.

Also consistent with this hypothesis are the results of study by Heffley et al. (1978). In this study, target and non-target stimuli were presented while subjects monitored a continuously moving display. Target probability and the interval between stimuli (Inter-Stimulus Interval, ISI) were manipulated. At the longer ISIs (6 seconds), target probability had no effect on P300 amplitude. All stimuli elicited large P300s. It is only when the ISI was shortened to approximately 2000 msec that low-probability stimuli elicited larger P300s than high-probability stimuli. This is consistent with the STM updating hypothesis of P300. With long ISIs, all targets regardless of probability will need updating of their representation because of the longer interval between repetitions. At the short ISIs, only rare targets need updating upon presentation. Frequent targets are most likely to occur while the previous occurrence is still held in STM.

Karis et al. (1984) reported a study that demonstrates rather clearly the effects that the elicitation of a P300 has on the representations created by the eliciting stimuli. They employed the von Restorff paradigm. The subject was instructed to recall a series of words, and a deviant item (an "isolate") was embedded in the series. Von Restorff (1933), the Gestalt psychologist who created this paradigm, demonstrated that the isolates were better recalled by subjects than were comparable non-deviant items. This enhanced recall of the isolates is the von Restorff or isolation effect.

As the isolates are both rare and task-relevant, they are apt to elicit a P300. The context-updating model of the P300 predicts that the larger the P300 elicited by an isolate the better it will be recalled. To test this hypothesis, Karis et al. (1984) examined the relationship between the amplitude of the P300 elicited by an isolate and whether or not it was subsequently recalled. Subjects were presented with word lists in which isolates were presented with a larger or smaller font than the other words, ERPs were recorded for each word, and after each list of 15 words subjects were asked to recall as many words as possible. The main experimental hypothesis was that isolated words recalled in the subsequent test should elicit larger P300s when initially presented than isolated words later not recalled.

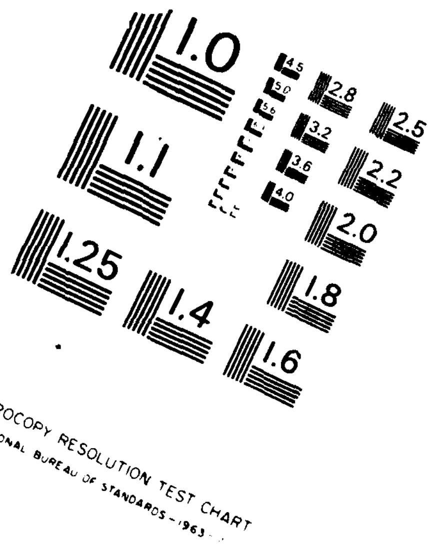
Striking individual differences were found in the degree to which subjects showed the von Restorff effect. These differences were surprising, because this effect has always been described as very robust. Some subjects reported using a rote memorization strategy and showed a large von Restorff effect (i.e., recalled proportionately more isolates), though their overall recall was quite poor. For these subjects, isolates that were recalled did elicit larger P300s on their original presentation than non-recalled isolates. Other subjects, however, showed a very different pattern. They were good memorizers and used elaborate memorization strategies. These subjects showed no von Restorff effect; that is, they recalled the non-isolates as well as they recalled the isolates. In these subjects P300 amplitude was unrelated to recall.

These data are consistent with the suggestion that P300 amplitude is a manifestation of an updating process in Working Memory. The data confirm that representation of a word in Working Memory is affected, in some manner, when a P300 is elicited. The change in the representation aids recall in rote memorizers. All subjects produced equally large P300s, and we believe the same updating process occurred in all our subjects. If no further processing occurs, as in our group of rote memorizers, then P300 amplitude will be related to recall. However, if cognitive activity continues after the initial processing reflected by P300, then this additional activity may obscure the relationship between P300 and recall. When, for example, subjects link words together as part of an elaborate memorizaton strategy, the recall of any individual word becomes less dependent on its initial encoding and more dependent on its relationships to other words.

We will not review here subsequent work in the Cognitive Psychophysiology Laboratory in which the validity of this finding was confirmed (see Fabiani et al., 1984, in preparation; Klein et al., 1984).

AD-A170 794 THE EVENT-RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING A..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 3/8

UNCLASSIFIED 20 FEB 86 CPL-86-1 F49620-85-C-0041 F/G 5/10 NL


MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Suffice it to say that we conclude from the work reviewed here and from other work that the amplitude of the P300 is proportional to the degree of activation of a process invoked when Working Memory is updated and organized. We proceed now to illustrate how this conceptual interpretation of the P300 can be used in illuminating the difference, discussed above, between elderly subjects in whom the P300 is abnormally small and other subjects.

The Low-P300 Subjects and Recall

The striking finding that a group of elderly subjects consistently lacked a P300 led us to a further investigation of the relationship between P300 and memory in the aged. We noted that the low-P300 subjects performed the oddball task as well as the other subjects. Comparing the low-P300 group with the high-P300 group revealed no differences in reaction time. The low-P300 subjects were 12 msec faster in mean RT than the high-P300 subjects for the rare stimuli (452 vs. 464 msec) and 38 msec slower on the frequent stimuli (412 vs. 374 msec); neither difference is statistically significant. Nor were the groups different in P300 latency or response accuracy.

If P300 is, as we believe, a manifestation of an intracranial process that is associated with the specific information processing task of updating representations in Working Memory, then it is possible that subjects who lack a P300 will show a deficit in performing tasks that depend on the viability of those short-term representations. For this reason, we designed a task that was intended to determine the subject's ability to maintain the status of his/her Working Memory. Two years after the initial oddball testing, we presented this new task to the subjects who had been originally recalled for the reliability retest. (Not all subjects were available for further testing by this time; nine high-P300 subjects and seven low-P300 subjects were fully tested.)

The memory test was patterned after the Hebb-Corsi test used by Milner (1978) and her colleagues in evaluating the effects of severe temporal lobe damage on memory. We designed a somewhat simpler task than the original Hebb-Corsi test, one in which it would be possible to record ERPs to each stimulus. On each trial, the subject was presented with a digit followed by a second digit 1000 msec later. The subject was instructed to indicate by a button press whether or not the pair had appeared previously in that trial block. In each of four blocks of 60 trials, one specific pair was repeated approximately every three trials. All other pairs were different. Reaction time and error rate served as dependent variables.

In accord with our theory, we found significant differences in performance between the high- and low-P300 subjects. Both groups were able to accomplish the task with high accuracy (88% for the "lows", and 93% for the "highs"; not significantly different). The low-P300 subjects, however, were significantly slower in RT to the non-repeating digit pairs; that is, they were slow to report that a pair had not been seen before. Mean RTs to non-repeating pairs for the high-P300 and low-P300 groups respectively were 850 and 1009 msec. Spearman rank-order correlation between RT to non-repeating digit pairs and P300 amplitude in the original oddball study, for the high and low groups combined, was -.49 ($p<.05$). The non-parametric Jonckheere test yields a significance level of $p<.025$ for these data.

The low-P300 subjects were somewhat slower (658 msec) than the high-P300 subjects (591 msec) on the repeating pairs (the one pair in each block that was repeated about every third trial). This difference between groups was not significant. A possible explanation for this differential deficit is that, for the low-P300 group, recall was impaired, whereas recognition was spared.

In order to test the reliability of these results, a new sample (n=32) of elderly subjects was recruited and tested on the auditory choice reacton time and visual names oddball tasks and in the memory paradigm. The essential findings of the first study were replicated.

Of the 32 new subjects, five had little or no P300, and eight were classified as high-P300 subjects. (Classification was by visual inspection of waveforms for the oddball tasks, blind to task performance results.) Once again, the low-P300 subjects did not differ from the high-P300 subjects on RT or accuracy in the oddball or memory tasks. (In auditory oddball CRT, low-P300 subjects were trivially faster than high-P300 subjects for the frequent stimuli--432 vs. 435 msec--and slightly slower for the rare stimuli--508 vs. 475 msec; neither difference is statistically significant.)

Like the initial group, the new low-P300 subjects were about 160 msec slower than the high-P300 subjects in RT to the non-repeating digit pairs in the memory task. (With the smaller sample size in the replication study, this difference fell short of statistical significance when analyzed using statistics that collapse all RTs for each subject to a single point--mean or median. An analysis of these data utilizing more powerful statistics is in progress.) The low-P300 subjects also appeared to be somewhat slower on the repeating pairs in all blocks. (This second group of subjects had previously been presented with two practice blocks of the memory task in a screening study.) Again as in the original study, the slowness of the low-P300 subjects was much more marked on the novel pairs than on the pairs they had already seen in that trial block (the repeating or target pairs). Figure 2 presents the median reaction times, following the onset of the second digit in a pair, for the high-P300 and low-P300 subjects, for each block of each of the two experiments.

Insert Figure 2 About Here

In summary, we found that elderly subjects who lack a P300, but are otherwise normal, are slower in indicating that an item is novel; we interpret this as indicating that they are slower either in searching Working Memory or in reporting the results of such a search, when the search is unaided by cues provided by the externally presented stimuli. The specificity of this finding is underscored by the fact that the high- and low-P300 subjects were equally fast in the choice-RT oddball task, which does not require the short-term maintenance of a memory set. Thus, an exceptionally low characteristic P300 reflects a deficit specific to recall from Working Memory, rather than a generalized deficit so often seen in special populations. These data support the interpretation of P300 as a manifestation of the updating of Working Memory and suggest that it is useful to examine the deterioration of memory performance in the elderly and the demented in terms of specific properties such as the maintenance of Working Memory.

Final Note

In the previous section we saw an example of the power of theory-driven research. Beginning with a hypothesis about the functional significance of the P300 as a manifestation of a process invoked in the updating of Working Memory made possible the design and implementation of a systematic series of experiments. This resulted in the discovery of a specific cognitive deficit that was predicted on the basis of a specific ERP deficit in a particular population. It is our view that such a theory-driven approach will prove to be a fruitful way to apply the knowledge and methods of psychophysiology in the clinical realm. Formulation and testing of hypotheses about information processing functions, their implementation by neural tissue, and the manifestations of this neural activity in psychophysiological measures such as ERPs may provide a key to unraveling the specific deficits in brain functioning which underlie dementia and other mental disorders.

References

Baddeley, A.D. (1981) The concept of working memory: A view of its current state and probable future development. Cognition, 10: 17-23.

Baddeley, A.D., and Hitch, G.J. (1974) Working memory. In G.A. Bower (Ed.), The psychology of learning and motivation: Vol. 8. Academic Press, New York.

Baribeau-Braun, J., Picton, T. W. and Gosselin, J.-Y. (1983) Schizophrenia: a neurophysiological evaluation of abnormal information processing. Science, 219: 874-876.

Beck, E. C., Swanson, C. & Dustman, R. E. (1980) Long latency components of the visually evoked potential in man: Effects of aging. Experimental Aging Research, 6: 523-542.

Braune, R., Wickens, C. D., Strayer, D. & Stokes, A. (In press) Age-dependent changes in information processing abilities between 20 and 60 years. Human Factors (Abstract).

Brown, W. S., Marsh, J. T. & LaRue, A. (1982) Event-related potentials in psychiatry: differentiating depression and dementia in the elderly. Bull. Los Angeles Neurol. Soc., 47: 91-107.

Brown, W. S., Marsh, J. T. & LaRue, A. (1983) Exponential electrophysiological aging: P3 latency. Electroencephalography and Clinical Neurophysiology, 55: 277-285.

Callaway, E., Tueting, P. & Koslow, S. H. (Eds.), (1978) Event-related brain potentials in man. Academic Press, New York.

Coles, M.G.H., Gratton, G., Bashore, T.R., Eriksen, C.W., & Donchin, E.

(submitted for publication) A psychophysiological investigation of the continuous flow model of human information processing. Submitted to Journal of Experimental Psychology: Human Perception and Performance.

Coles, M. G. H., Gratton, G., Kramer, A., Miller, G. A., & Porges, S. W. (in press) Principles of signal acquisition and analysis. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), Psychophysiology: Systems, Processes, and Applications -- A Handbook. New York: Guilford Press.

Donchin, E. (1975) On evoked potentials, cognition, and memory. Science, 190: 1004-1005.

Donchin, E. (1979) Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), Evoked Brain Potentials and Behavior, New York, Plenum, pp. 13-75.

Donchin, E. (1978) Use of scalp distribution as a dependent variable in event-related potential studies: Excerpts of preconference correspondence. In D. Otto (Ed.), Multidisciplinary perspectives in event-related brain potential research, (EPA-600/9-77-043), U.S. Government Printing Office, Washington, D.C., pp. 501-510.

Donchin, E. (1981) Surprise!. . . Surprise?, Psychophysiology, 18: 493-513.

Donchin, E. (In press) Hypothetical constructs, intervening variables, and ERP components. In Proceedings of the Third Annual Carmel Conference.

Donchin, E., Kubovy, M., Kutas, M., Johnson, R., Jr. & Herning, R. I (1973) Graded changes in evoked response (P300) amplitude as a function of cognitive activity. Perception & Psychophysics, 14: 319-324.

Donchin, E. & McCarthy, G. (1980) Event-related brain potentials in the study of cognitive processes. In C. Ludlow & M. E. Doran-Quine (Eds.), Proceedings of Symposium on Neurological Bases of Language Disorders in

Children: Methods and Direction for Research, NINCDS Monograph #22, U.S. Government Printing Office, Washington, D.C., pp. 109-128.

Donchin, E., Ritter, W. & McCallum, C. (1978) Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting & S.H. Coslow (Eds.), Event-related brain potentials in man, Academic Press, New York.

Duncan-Johnson, C. & Donchin, E. (1981) The relation of P300 latency to reaction time as function of expectancy. In H. H. Kornhuber & L. Deecke (Eds.), Motivation, motor and sensory processes of the brain: Electrical potentials, behavior and clinical use, Progress in brain research, Vol. 54, Elsevier, Amsterdam, pp. 697-715.

Duncan-Johnson, C. C. & Donchin, E. (1982) The P300 component of the event-related brain potential as an index of information processing. Biological Psychology, 14: 1-52.

Fabiani, M., Karis, D. & Donchin, E., (1984) Paper presented at the Third International Conference on Cognitive Neuroscience, Bristol, England.

Fabiani, M., Karis, D. & Donchin, E., (In preparation) "Gee, that was my aunt Mathilda": P300 and recall in an incidental memory paradigm.

Farwell, L.A., Chambers, R.D., Miller, G.A., Coles, M.G.H., & Donchin, E. (1985) A specific memory deficit in elderly subjects who lack a P300. Psychophysiology, 22: 589 (Abstract; extended paper in preparation).

Ford, J.M., Duncan-Johnson, C.C., Pfefferbaum, A. & Kopell, B.S. (1982a) Expectancy for events in old age: stimulus sequence effects on P300 and reaction time. Journal of Gerontology, 37: 696-704.

Ford, J.M., Pfefferbaum, A., Tinklenberg, J.R. & Kopell, B.S. (1982b) Effects of perceptual and cognitive difficulty on P3 and RT in young and old adults. Electroencephalography and Clinical Neurophysiology, 54:

311-321.

Ford, J.M., Roth, W.T., Mohs, R., Hopkins, W. & Kopell, B.S. (1979) Event-related potentials recorded from young and old adults during a memory retrieval task. Electroencephalography and Clinical Neurology, 47: 450-459.

Galambos, R. & Hillyard, S. A. (1981) Electrophysiological approaches to human cognitive processing. Neurosciences Research Program Bulletin, 20: 145-265.

Goodin, D., Squires, K., Henderson, B. & Starr, A. (1978a) Age related variations in evoked potentials to auditory stimuli in normal human subjects. Electroencephalography and Clinical Neurophysiology, 44: 447-458.

Goodin, D. S., Squires, K. C. & Starr, A. (1978b) Long latency event-related components of the auditory evoked potential in dementia. Brain, 101: 635-648.

Gopher, D. & Donchin, E., (In press) Workload - An examination of the concept. In K. Boff & L. Kaufman (Eds.), Handbook of Perception and Human Performance, Wiley and Sons, New York.

Gratton, G., Coles, M.G.H. & Donchin, E. (Submitted for publication) A multivariate approach to the analysis of scalp distribution of Event Related Potentials: Vector Analysis. Submitted to Electroencephalography and Clinical Neurology.

Gratton, G., Coles, M.G.H. & Donchin, E. (1984) Component identification with vector analysis. Paper presented at the Third International Conference on Cognitive Neuroscience, Bristol, England.

Halliday, A.M. (1978) Clinical applications of evoked potentials. In W.B. Matthews & G.H. Glaser (Eds.), Recent advances in clinical neurology, Churchill Livingstone, Edinburgh, pp. 47-73.

Halliday, A.M. (1973) Visual evoked response in diagnosis of multiple sclerosis. British Medical Journal, 4: 661-664.

Heffley, E., Wickens, C.D. & Donchin, E. (1978) Intramodality selective attention and P300: A reexamination in a visual monitoring task. Psychophysiology, 15: 269-270.

Hillyard, S.A. & Kutas, M. (1983) Electrophysiology of cognitive processing. In M. R. Rosenzweig & L. W. Porter (Eds.), Annual Review of Psychology, 34: 33-61.

Isreal, J.B., Wickens, C.D., Chesney, G.L. & Donchin, E. (1980) The event-related brain potential as an index of display monitoring workload. Human Factors, 22: 211-224.

Karis, D., Fabiani, M. & Donchin, E. (1984) "P300" and memory: individual differences in the von Restorff effect. Cognitive Psychology, 16: 177-216.

Klein, M., Coles, M.G.H. & Donchin, E. (1984) People with absolute pitch process tones without producing a P300. Science, 223: 1306-1309.

Kramer, A., Wickens, C.D. & Donchin, E. (1983) An analysis of the processing demands of a complex perceptual-motor task. Human Factors, 25: 597-622.

Kutas, M., McCarthy, G. & Donchin, E. (1977) Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 197: 792-795.

Magliero, A., Bashore, T.R., Coles, M.G.H. & Donchin, E. (1984) On the dependence of P300 latency on stimulus evaluation processes. Psychophysiology, 21: 171-186.

Marshall, N.K., Bashore, T.R., Miller, G.A., Coles, M.G.H., & Donchin, E. (1983) ERP consistency in young adult and elderly subjects. Symposium presentation at the annual meeting of the Society for Psychophysiological Research, Asilomar, CA, September, 1983.

McCarthy, G. (1984) Stimulus evaluation time and P300 latency. In E. Donchin (Ed.), Cognitive Psychophysiology: ERPs and the Study of Cognition, The Carmel Conferences, Vol. I, Lawrence Erlbaum, Hillsdale, N.J., pp. 254-301.

McCarthy, G. & Donchin, E. (1979) Event-related potentials--Manifestations of cognitive activity. In F. Hoffmeister & C. Muller (Eds.), Bayer-Symposium VII, Brain function in old age, Springer-Verlag, New York, pp. 318-335.

McCarthy, G. & Donchin, E. (1981) A metric for thought: A comparison of P300 latency and reaction time. Science, 211: 77-80.

Michalewski, H.J., Patterson, J.V., Bowman, T.E., Litzleman, D.K., & Thompson, L.W. (1982) A comparison of the emitted late positive potential in older and young adults. Journal of Gerontology, 37: 52-58.

Milner, B. (1978) Clues to the cerebral organization of memory. In Buser & Rongl-Buser (Eds.), Cerebral correlates of conscious experiences: INSERM Symposium No. 6, pp. 139-153.

Mullis, R.J., Holcomb, P.J. & Dykman, R.A. (1985) The effects of aging on the P3 component of the visual event-related potential. Electroencephalography and Clinical Neurology, 62: 141-149.

Neville, H.J., Snyder, E., Knight, R. & Galambos, R. (1979) Event related potentials in language and non-language tasks in patients with alexia without agraphia. In D. Lehmann & E. Callaway (Eds.), Human Evoked Potentials, Plenum, New York, pp. 269-283.

Pfefferbaum, A., Ford, J., Roth, W.T. & Kopell, B.S. (1980a) Age differences in P3-reaction time associations. Electroencephalography and Clinical

Neurophysiology, 49: 257-265.

Pfefferbaum, A., Ford, J., Roth, W.T. & Kopell, B.S. (1980b) Age-related changes in auditory event-related potentials. Electroencephalography and Clinical Neurophysiology, 49: 266-276.

Pfefferbaum, A., Ford, J.M., Wenegrat, B., Roth, W. & Kopell, B. (1984a) Clinical application of the P3 component of event-related potentials. I. Normal Aging. Electroecephalography & Clinical Neurophysiology, 59: 85-103.

Pfefferbaum, A., Ford, J.M., Wenegrat, B., Roth, W. & Kopell, B. (1984b) Clinical application of the P3 component of event-related potentials. II. Dementia, depression, and schizophrenia. Electroecephalography & Clinical Neurophysiology, 59: 104-124.

Picton, T.W., Stuss, D.T., Champagne, S.C. & Nelson, R.F. (1984) The effects of age on human event-related potentials. Psychophysiology, 21: 312-325.

Podlesny, J.A., Dustman, R.E. & Shearer, D.E. (1984) Aging and respond-withhold tasks: Effects on sustained potentials, P3 responses and late positivity. Electroencephalography and Clinical Neurophysiology, 58: 130-139.

Podlesny, J.A. & Dustman, R.E. (1982) Age effects on heart rate, sustained potential, and P3 responses during reaction-time tasks. Neurobiology of Aging, 3: 1-9.

Polich, J., Ehlers, C.L., Otis, S., Mandell, A., & Bloom, F.E. (In press) P300 latency reflects the degree of cognitive decline in dementing illness. Electroencephalography and Clinical Neurophysiology.

Porjesz, B. & Begleiter, H., (1982) Evoked brain potential deficits in alcoholism and aging. Alcoholism, Clinical and Experimental Research, 6:

53-63.

Pritchard, W.S. (1981) Psychophysiology of P300. Psychological Bulletin, 89: 506-540.

Ragot, R. & Renault, B. (1981) P300 as a function of S-R compatibility and motor programming. Biological Psychology, 13: 289-294.

Regan, D. (1972) Evoked Potentials in Psychology, Sensory Physiology and Clinical Medicine, Chapman and Hall, London.

Restorff, H. von (1933) Uber die Wirkung von Bereichsbildungen im Spurenfeld, Psychologische Forschung, 18: 299-342.

Ritter, W. & Vaughan, H.G., Jr. (1969) Averaged evoked responses in vigilance and discrimination: A reassessment. Science, 164: 326-328.

Roth, W.T., Horvath, T.B., Pfefferbaum, A.Tinklenberg, J.R., Mezzich, J.E., & Kopell, B.S. (1979) Late event-related potentials and schizophrenia. In H. Begleiter (Ed.), Evoked Brain Potentials and Behavior, Vol. 2, Downstate Series of Research in Psychiatry and Psychology, Plenum, New York, pp. 499-515.

Roth, W.T., Pfefferbaum, A., Kelly, A.F., Berger, P.A. & Kopell, B.S. (1981) Auditory event-related potentials in schizophrenia and depression. Psychiatry Research, 4: 199-212.

Sirevaag, E., Kramer, A., Coles, M.G.H. & Donchin, E. (1984) P300 amplitude and resource allocation. Psychophysiology, 21: 598-599 (Abstract; extended paper in preparaton).

Slaets, J.P. J. & Fortgens, C. (1984) On the value of P300 event-related potentials in the differential diagnosis of dementia. British Journal of Psychiatry, 145: 652-656.

Smith, D.B.D., Michalewski, H.J., Brent, G.A. & Thompson, L.W. (1980) Auditory averaged evoked potentials and aging: Factors of stimulus, task and topography. Biological Psychology, 11: 135-151.

Snyder, E. & Hillyard, S.A., (1979) Changes in event-related potentials in older persons. In F. Hoffmeister (Ed.), Bayer-Symposium VII: Brain function in old age, Springer-Verlag, Berlin, pp. 112-125.

Sokolov, E.N. (1969) The modeling properties of the nervous system. In I. Maltzman & M. Cole (Eds.), Handbook of Contemporary Soviet Psychology, Basic Books, New York, pp. 671-704.

Squires, K.C., Chippendale, T.J., Wrege, K.W., Goodin, D.W., & Starr, A. (1980) Electrophysiological assessment of mental function in aging and dementia. In L. Poon (Ed.), Aging in the 1980s, American Psychological Association, Washington, D.C., pp. 125-134.

Squires, K., Goodin, D. & Starr, A. (1979) Event related potentials in development, aging and dementia. In D. Lehmann & E. Callaway (Eds.), Human Evoked Potentials, Plenum, New York, pp. 383-396.

Squires, K.C., Wickens, C.D., Squires, N.K. & Donchin, E. (1976) The effect of stimulus sequence on the waveform of the cortical event-related potential. Science, 193: 1142-1146.

Squires, N., Galbraith, G. & Aine, C. (1979) Event-related potential assessment of sensory and cognitive deficits in the mentally retarded. In D. Lehmann & E. Callaway (Eds.), Human Evoked Potentials, Plenum, New York, pp. 397-413.

Strayer, D., Wickens, C.D. & Braune, R. (1985) Mental chronometry and aging: an additive factors/psychophysiological approach. Psychophysiology, 22: 615-616 (Abstract).

Syndulko, K., Hansch, E.C., Cohen, S.N., Pearce, J.W., Goldberg, Z., Montana, B., Tourtellotte, W.W., & Potvin, A.R. (1982) Long-latency event-related potentials in normal aging and dementia. In J. Courjon, F.

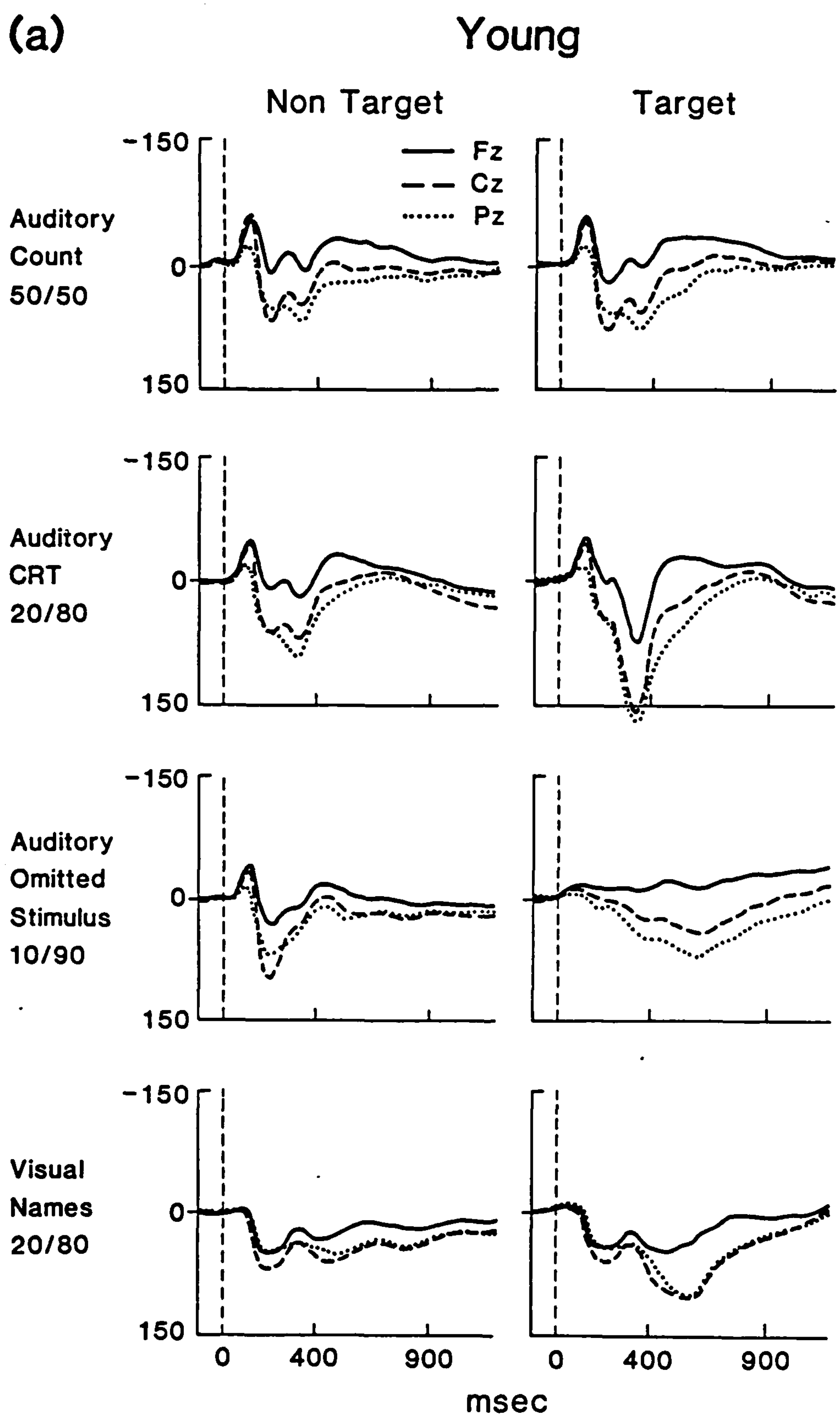
Mauguiere & M. Revol (Eds.), Clinical Applications of Evoked Potentials in Neurology, Raven, New York, pp. 279-285.

Wickens, C.D., Braune, R., Stokes, A. & Strayer, D. (1985) Individual differences and age-related changes: refinement and elaboration of an information processing performance battery with aviation relevant task structures. Final Report, University of Illinois Engineering Psychology Research Laobratory Technical Report EPL-85-1/NAMRL-85-1, February.

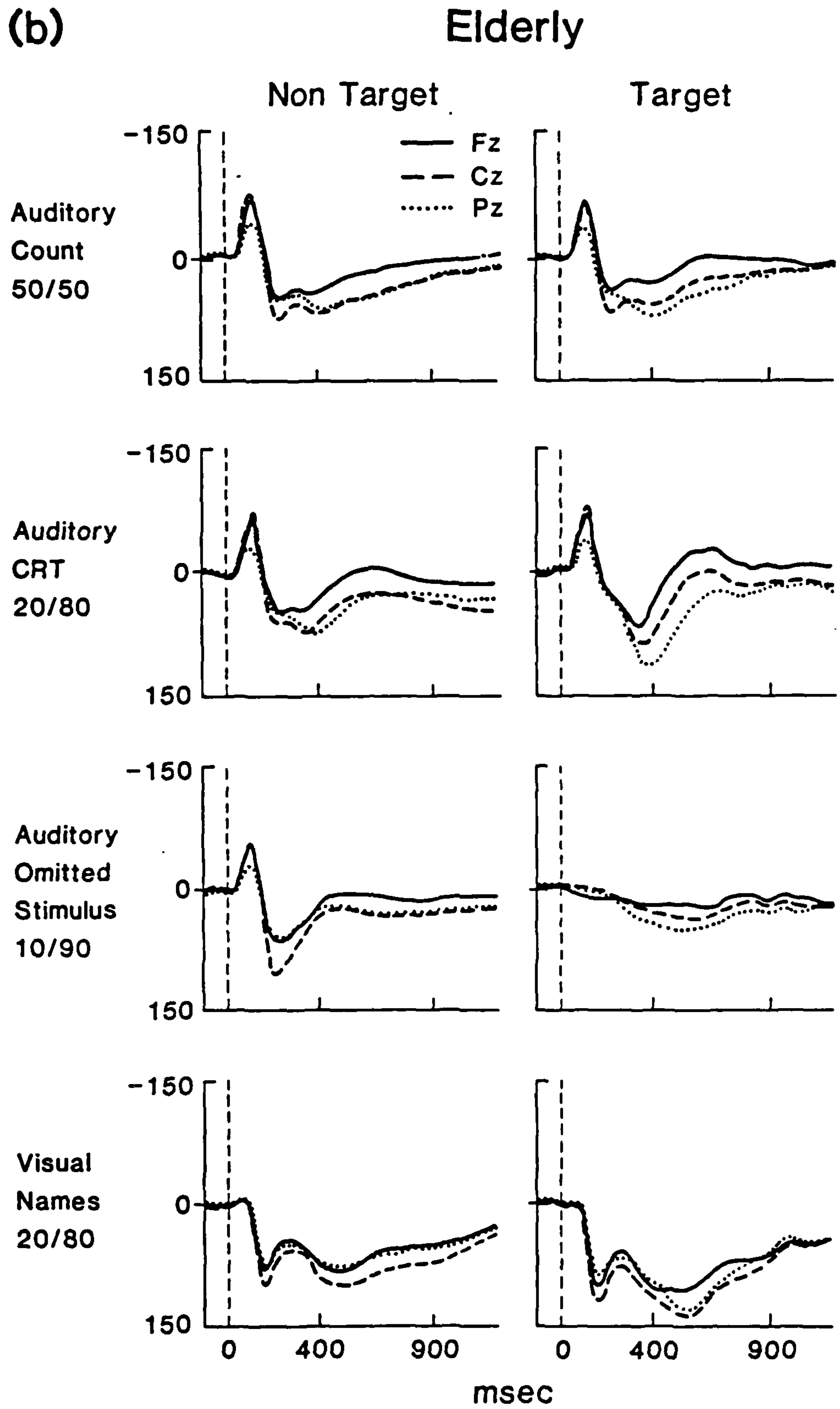
Figures

Figure 1. Grand average ERP waveforms for (a) 19 young and (b) 53 elderly subjects in four oddball tasks. Stimuli presented in Bernoulli series with the probabilities indicated. ERPs recorded at Fz, Cz, and Pz. Note the P300, a positive deflection (downward in the figure) between 300 and 600 msec post-stimulus, larger in the Target conditions.

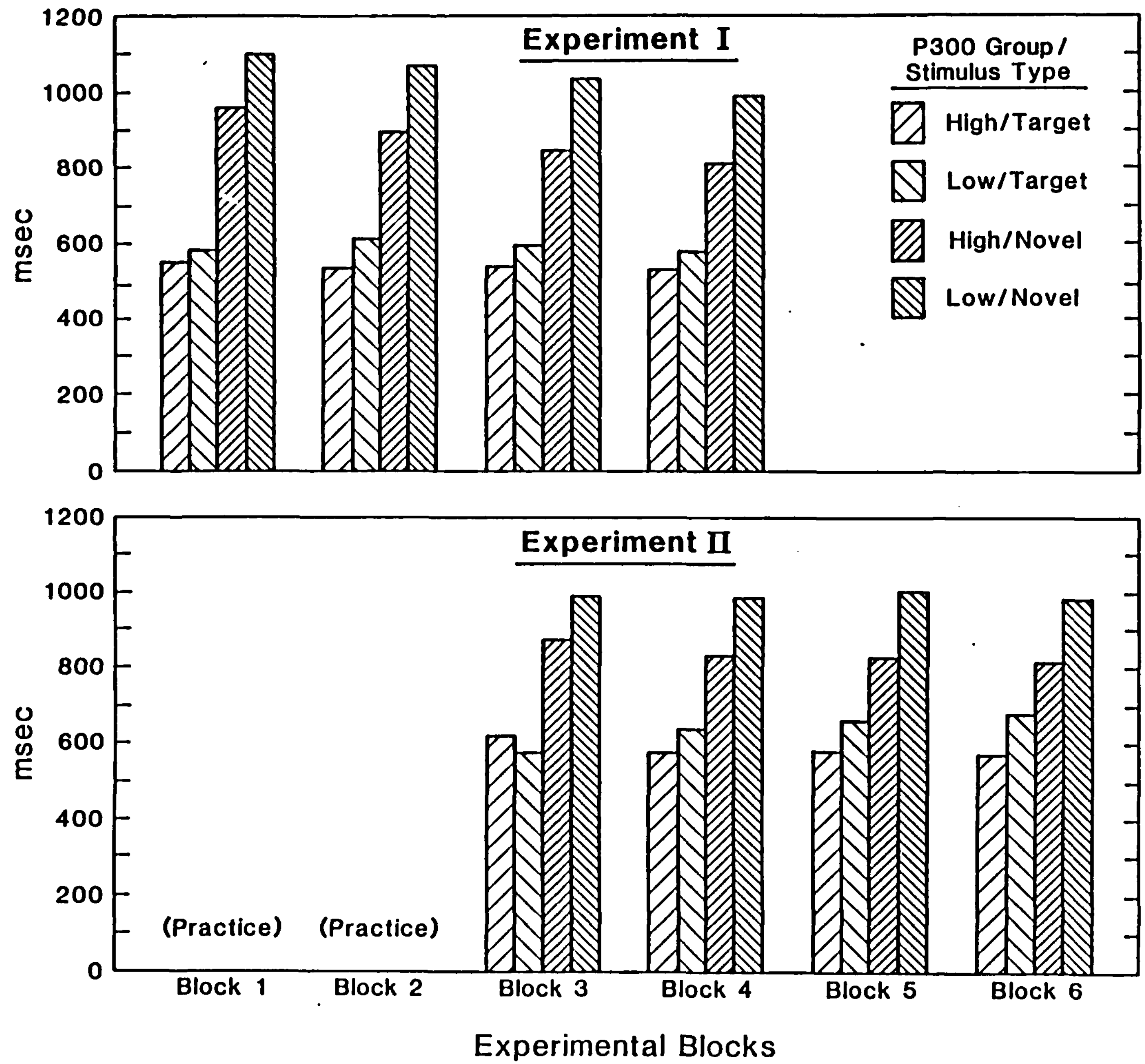
Figure 2. Median Reaction Times for 9 high-P300 and 7 low-P300 elderly subjects to Target (Repeating) and Non-target (Non-repeating) digit pairs. (Blocks 1 and 2 in Experiment II were practice blocks; RTs not recorded.)



(a)
Young
Non Target
Target
Fz
Cz
Pz
Auditory Count 50/50
Auditory CRT 20/80
Auditory Omitted Stimulus 10/90
Visual Names 20/80
-150
0
150
0
400
900
msec


(b)
Elderly
Non Target
Target
Fz
Cz
Pz
Auditory Count 50/50
Auditory CRT 20/80
Auditory Omitted Stimulus 10/90
Visual Names 20/80
-150
0
150
0
400
900
msec

Median Reaction Times



Appendix B7

THE DEFINITION, IDENTIFICATION, AND RELIABILITY OF MEASUREMENT OF THE P300 COMPONENT OF THE EVENT-RELATED BRAIN POTENTIAL

Monica Fabiani, Gabriele Gratton, Demetrios Karis, and Emanuel Donchin[1]

Cognitive Psychophysiology Laboratory

University of Illinois

603 East Daniel

Champaign, Il. 61820

Running head: P300: DEFINITION, IDENTIFICATION, RELIABILITY

To appear in: P. K. Ackles, J. R. Jennings & M. G. H. Coles (Eds.), Advances in Psychophysiology, Volume 2. Greenwich, CT: JAI Press, Inc.

INTRODUCTION

In 1965, Sutton, Braren, Zubin, and John reported that stimuli that "reduce a subject's uncertainty" elicit an event related brain potential (ERP) characterized by a a large positive-going deflection that peaks some 300 ms after the eliciting event. In this chapter we shall use the nomenclature proposed at the Brussels Congress on ERPs (Donchin et al., 1977), and refer to this deflection as the "P300 component." The original observations by Sutton et al. (1965) triggered a considerable amount of research (for a review, see Pritchard, 1981), and there is now an extensive literature on the P300. The literature is rife with controversies, ranging from debates on the proper measurement of the component, to disputes among rival theories regarding the functional significance of the component.

In this chapter we address issues relevant to many of these controversies. We present data on the reliability of various measurement procedures that are used to identify, and measure, the P300. However, the attempt to measure an entity presupposes the existence of a measurable entity. It further assumes the existence of consensual criteria on which a definition of the component, and its measurements, can rest. It turns out that even these criteria are disputatious matters. Therefore, we precede the report on the study of reliability with a review of theoretical and methodological issues concerning the P300 component of the ERP. Our intent is to develop, on the basis of this discussion, a framework within which to conduct the interpretation and assessment of the research on P300. We shall begin our review by considering the concept of a "component" which is, of course, central to the discussion of the P300. We shall show how the decomposition of ERPs into components is an instance of the use of decomposition techniques in psychophysiological measurement. This will be followed by a discussion of the methods used, and of the problems encountered, in identifying and measuring components. This discussion will reveal that, to a degree, the measurement of a component depends on the experimental paradigm in which it is elicited. We are led therefore to attempt a taxonomy of experimental paradigms that have been used in studies of the P300. This theoretical discussion will be followed by a report of a study on the reliability of various techniques used to measure P300.

This chapter is written, in part, for individuals who are not familiar with ERP research. This review will, we hope, ease their entry into a quickly expanding literature by clarifying the problems encountered in data analysis, and by reviewing some of the theoretical controversies. Three appendices contain more detailed discussions of the paradigms used in the study of P300, of some factors influencing P300 parameters, and of the relationship between P300 and other late positivities. For readers familiar with research in the field this chapter offers new data on the reliability of P300 measurement techniques. However, in addition we intend the discussion to serve as a contribution to the debate regarding the "existence" of the P300 and its relation to other "positivities." Several investigators have drawn attention to a number of issues pertaining to the P300. For example, there seems to be some debate on the number of components that happen concurrently with the P300. It is not clear if P300 represents a unitary process. Some of these questions are legitimate and interesting. Others are "pseudo-issues" where concern with irrelevant details elevates them to the status of major theoretical conundrums. The following discussion should bear on these issues.

THE DECOMPOSITION OF PSYCHOPHYSIOLOGICAL MEASUREMENTS INTO COMPONENTS

The concept of psychophysiological components

Psychophysiological measures are records of the activity of some organ system such as the heart, the stomach or the brain. These measures become "psychophysiological" when the measurements are made in the context of experiments in which the independent variables are "psychological" in nature. The term "psychological" in the previous sentence is used in the sense of "commonly used by psychologists." The measures are obtained, however, for psychological reasons where by "psychological" we mean the development of theories regarding the structure and the function of the mind. The basic assumption of Psychophysiology is that the measured physiological activity is of interest because its variation is due in part to variance induced in the organism by the psychological manipulations. Thus the psychophysiological measures are seen as physiological manifestations of psychological events. It is possible, especially in medical applications, to use the psychophysiological measures to learn about the state of the organs from which the recordings are obtained. However, in this chapter we are focusing on the use of these measures as tools in the study of cognition and affect. The primary task, in this case, is to analyze the variance in the psychophysiological measures, so that the share of the variance attributable to specific experimental manipulations can be ascertained. As it is unlikely that any given experimental manipulation will affect the same set of psychological processes, it has become the main task of psychophysiologists to develop procedures that permit a decomposition of the measures into their components. Each component of the variance carries in its modulations information about specific aspects of the underlying interactions between the processing activity of the mind and its implementation by the bodily systems. Thus, for example, psychophysiologists have long isolated heart rate as a specific component of cardiac activity because they could establish lawful relationships between acceleration and deceleration of the heart rate and information processing variables (e.g. Coles, 1984; Lacey, 1959). Again, the argument for the isolation is made in terms of the specificity of the relationships between the variables controlling information processing activity and the behavior of the measure.

Thus, the term "component" seems to be used to refer to a portion of the observed variance that can be viewed as an element useful in the development of a psychophysiological theory. Clearly, the physiological events usually occur within a continually varying context. For any psychological process we wish to study, the psychophysiological data reflect both activity associated with the process of interest (the signal) and activity associated with other processes which are not of interest (noise). Furthermore, both these sources of psychophysiological variance may be very complex in nature. It is for this reason that the definition of a component is a matter for so much controversy, and the isolation of a psychophysiological event as a useful component must be carefully validated.

Ideally, components represent discrete and interpretable units of the psychophysiological response. Each component should manifest the activity of one, and only one, psychological process. The attributes of a component may vary as a function of independent variables, and these variations may reflect changes in the "strength" and "timing" of the underlying psychological process.

This interpretation of psychophysiological components makes them particularly attractive for the investigation of psychological processes. In fact, they may be used as indices of the occurrence of processes otherwise not directly observable. However, the investigator must demonstrate: (a) the existence of a component, (b) its independence from other components and indices of psychological processes, (c) that the component is unitary and not multiple (no further decomposition is possible), (d) that the component is the manifestation of a specific psychological process, and (e) that an identifiable portion of the variability of the psychophysiological component reflects an identifiable portion of the variability of the psychological process. Of course, to address these problems, instruments capable of measuring the amplitude and time course of the component are needed.

In practice, the satisfaction of all these criteria (particularly points b and c) is almost impossible. However, in most cases a partial description of the psychophysiological events underlying a certain experiment may be sufficient for practical purposes. For instance, we do not need to know whether a component is a composite of several sub-components (and therefore not the smallest possible unit) when we are incapable of decomposing it further. This may change, of course, with the implementation of more sophisticated techniques.

Theoretical and Operational Definitions of a Component

To separate the component of interest from other psychophysiological activities, we must provide a definition of that component. It is important to stress a distinction proposed by Donchin et al. (1977) between observational and theoretical reference to components. The distinction was made with regard to the use of labels such as P300 in the ERP literature. The observational reference to components was presumed "...to serve to describe the data collected and (its) primary function is to serve as a descriptive shorthand." However, the authors go on to note that:

> "We do, however, use the nomenclature in an altogether different form when the label is used to identify a theoretical entity. Thus, when one talks of the 'CNV' or the 'P300,'... one is talking of an entity which one believes characterizes the evoked response and represents some essential physiological, psychological or hypothetical construct whose properties are under study. The conflict between the observational and the theoretical nomenclature arises from the fact that the theoretical 'P300' may observationally appear as P250, P300, P350, or perhaps P400. One assumes then that all these observational components are realizations of the unique theoretical process referred to as 'the P300'" (Donchin et al., 1977, p. 10).

The distinction, while it may help to clarify the nomenclature assigned to components, is not as sharp as it may appear. Evidently, one does not proceed to develop a theoretical definition of a component before a wealth of observations suggest that there is indeed variance in the data that can be interpreted by assuming the existence of a component. On the other hand, the studies of a component, once the data point to its existence, are driven to a large extent by the view one has of its likely functional significance. It would, for example, have been rather surprising if researchers studying the auditory brainstem potentials sought to establish the effects of emotional stress on these components before analyzing the effects of the spectral composition of the stimulus, or of such variables as stimulus intensity and the repetition rate.

The distinction between observational and theoretical definition of a component is useful, because once the component has been identified as a target for study, the experimental designs are driven by the theoretical concepts, but the actual data acquisition is determined by the observational definition. Indeed, it may be more appropriate to refer to the latter definition as operational because it specifies the operations an investigator undertakes when acquiring data about a component.

An operational definition is the set of practical operations used to identify and measure a component, and consists of a subset of notions derived from the theoretical definition. In general, the operational definition of a component is based on some specific features of the data, such as its spatial and temporal characteristics, or its response to particular experimental manipulations. When these features are used it is assumed that they are invariant characteristics of the component itself. Therefore, when this invariance assumption is violated, the existence of additional components is usually inferred.

In discussing the theoretical definition of a component it is again necessary to clarify the multiple meanings of the usage of the term component. Indeed, in the ERP literature, there are at least three senses in which the word component is used. The term may refer to morphological features of the waveforms. It may refer to a theoretical entity that exists within the algorithmic description of the information processing system provided within a cognitive analysis of the system. Finally, the term can be used to refer to the intracranial activity that underlies the component. We arrive at the idea that it is useful to decompose the ERP into components because observations on the morphology of the waveforms suggest that there are peaks and troughs in the waveform that seem to move together in response to experimental manipulations. Thus, one defines a component as a uniform morphological feature of the data. It quickly becomes evident that there is considerable fluidity in the morphological description of a component. Specific morphological features may be used in the operational definition of a component. For example, the P300 may be defined, in part, by the latency of its peak ("...with a latency of about 300 ms."). Yet, aspects of the P300 are often accepted as representations of a component even though the morphological features have deviated from the defining range. Thus, peaks with latencies far longer than 300 ms are accepted as "P300s." We return below to the degree to which the theoretical definition can tolerate variation in observational, or operational, definition.

In the theoretical definition of a component, an important role must be given to its function. "A component is a set of potential changes that can be shown to be functionally related to an experimental variable or to a combination of experimental variables" (Donchin, Ritter, & McCallum, 1978, p. 353). On the other hand, the relationship between a component and its neural substrate must also be considered. Donchin et al. (1978) write:

"...an ERP component is a subsegment of the ERP whose activity represents a functionally distinct neuronal aggregate. Functionally distinct aggregates need not be anatomically distinct neuronal populations. But it is assumed that neuronal aggregates whose activity will be represented by an ERP component have been distinctly affected by one or more experimental variables" (p.353). An ERP component is thus not synonymous with a peak or deflection in the waveform. Donchin et al. (1978) go so far as to say that "a single ERP waveform can never reveal components" (p.354), because no variability between conditions can be observed.

This conception of an ERP component varies radically from others, such as that expressed by Sams, Alho, & Näätänen (1984), whose definition of a component is based on the underlying neural generator rather than on its function. They write that,

> The terms deflection and component should be distinguished. An ERP component should be understood as a contribution of a single generator process (such as transient activation of some brain center or region) to the total ERP waveform. ERP deflections (e.g., N1, P2, N2, P3) are usually products of two or more partially or totally overlapping components. (p.434)

However, given the present difficulty (or impossibility) of determining if only a single neural generator is involved (see below for a discussion of this subject), one is left with a series of "deflections". Therefore, following this reasoning, Sams et al. (1984) take mean amplitudes at each electrode in 50 ms bins throughout the epoch, instead of identifying components and measuring their parameters. This is a descriptive approach which does not allow any inferences about the psychological processes underlying endogenous components,[2] and P300 in particular.

As data on the morphology of a component accumulate, and as its response to experimental manipulations is elucidated, a concept of the information processing activity underlying the component emerges. It is at this point that the component as a theoretical entity begins to emerge. At this level the observed morphological changes are viewed as manifestations of an entity that can be described in information processing terms. One may view in this context the P300 as a manifestation of context updating (Donchin, 1981), or of closure (Desmedt, 1980). The N400 may be viewed as a manifestation of a mechanism invoked to process semantic ambiguities (Kutas & Hillyard, 1980a) and the N200 as reflecting a mismatch detector (Näätänen, 1982). This level of theoretical definition expands the range of the operational definition because it predicts the effects of certain experimental manipulations on the component, and, if these predictions are confirmed, these very effects become part of the definition of a component. In this manner, the sensitivity of P300 latency to categorization time, once established, becomes a defining characteristic of the P300.

While the algorithmic specification of a component may be adequate for many purposes, it is clear that the scalp recorded ERP is generated intracranially and that the information processing activity it manifests is implemented in the brain. Thus, a complete theoretical definition of a component must ultimately include a specification of the intracranial sources that generate the recorded activity. The theoretical definition of a component may, in the end, include statements about specific brain structures whose activity is responsible for the appearance of the component on the scalp. As the theoretical definition is thus expanded, it will again determine the development of additional elements of the operational definition.

Note that, in a specific study, we may employ in the operational definition of a component only a small subset of the features that are implied by the theoretical definition. Typically, investigators focus on a few features of the component, such as its latency or scalp distribution, and these are used in identifying the component within the study. This approach must be applied with care, as two different components may sometimes share features of an operational definition. That is, morphological similarities between different components may impair component identification. It is critical, therefore, that the appropriateness of the operational definition of a component be evaluated in any study. It is evident that statements of value about ERPs are statement regarding the theoretical entities identified as the components rather than the morphological features of a particular waveform. Therefore, the degree to which a given operational definition yields unique measures of a specific theoretical component should be evaluated in drawing conclusions from experimental data.

Problems Associated with Multiple Overlapping Components

Once the occurrence of a component has been determined, its parameters can be measured. This task is relatively easy when the component is isolated, but becomes arduous when several components overlap in time and/or in space. Indeed one of the more vexing problems in ERP research is "component overlap." While the ERP may appear as a sequence of peaks and troughs there is no escaping the fact that many processes are simultaneously active in the brain, and that several such processes may have scalp manifestations concurrently. Furthermore, components do not act in an instantaneous fashion. Rather, they appear to be activated, and to subside, over considerable time intervals. This being the case, it is quite likely that the voltage recorded at the scalp represents, at each instance, the activity of multiple components. Evidently, the morphology of the ERP is determined by the joint action of several components, and therefore the operational estimates of these components, and the interpretation of their parameters, may be in error because the effects of one theoretical entity is confused with that of another.

Investigators have different solutions to this problem. Several procedures attempt to decompose the observed ERP and thus allow the comparison of records obtained under different spatial, temporal, and experimental conditions. A very common approach is the computation of "difference" waveforms. These are presumed to isolate overlapping components. Two waveforms are subtracted point by point and the resulting waveform is considered an estimate of a component. This procedure assumes that only one component varies between the two conditions, and that, therefore, the difference between the two waveforms is an estimate of that component. A further assumption is that only the amplitude (but not the latency) of the component varies between two conditions.

More complex procedures are based on the analysis of the variances and covariances of multiple records. The most common is Principal Component Analysis (PCA). The use of PCA in ERP analysis involves the assumption that the temporal characteristics of a component are invariant across conditions (Donchin, 1966, 1969a). Other assumptions about components and noise are also required (see Coles, Gratton, Kramer, & Miller, in press). In general, the use of any decomposition procedure involves some assumption of "invariance." The validity of these assumptions need to be evaluated on a case-by-case basis.

THE P300 COMPONENT OF THE ERP

Donchin et al. (1978) operationally defined P300 as a component with a latency longer than 275 ms, positive in polarity at all midline electrode locations (in comparison with a "neutral" reference), with maximum positivity at parietal and central locations, elicited by task relevant stimuli, and whose amplitude was affected by the subjective probability and task relevance of the stimulus.

Thus, Donchin et al. (1978) used four defining features to characterize P300: (a) P300 has a particular polarity (i.e., positive), (b) P300 has particular latency characteristics (i.e., in excess of 275 ms), (c) P300 has a particular scalp distribution (i.e., maximum positivity at centro-parietal scalp locations),[3] and (d) P300 has a particular pattern of response to experimental manipulations (i.e., larger for rare and task relevant stimuli than for frequent and unattended stimuli).

This definition of P300 attempted to summarize the consensus regarding operations to be used in defining the theoretical entity "P300." These operations had all appeared in the literature available in 1978. However, in most cases, the operational definition of P300 employed only a subset of these operations. For instance, in most studies the morphology, rather than the response of P300 to specific experimental manipulations, is used in the operational definition. The scalp distribution is often ignored, as some investigators use only a single electrode site (usually Cz; electrode locations refer to the 10-20 System, Jasper, 1958). As the latency of the P300 is frequently a dependent variable in the experiments, it can not be used as a defining variable.

The development of a theoretical definition of P300 is still under way. The specificity of the scalp distribution of the P300, and therefore its utility as a defining attribute have been questioned. Several investigators have noted the appearance of other positive deflections in the same time region as the P300. While it is not clear which of these deflections represent components, there is a distinct possibility that multiple components are active in this epoch. The existence, classification, and definition of these components are still uncertain.

The operational definition of P300 that we recommend corresponds closely to the guidelines given by Donchin et al. (1978). We recommend identifying P300 on the basis of (a) polarity (positive), (b) latency (which, however, is influenced by modality, task requirements, and stimulus complexity and clarity), (c) morphology (it must have an identifiable peak), (d) scalp distribution (Pz and Cz > Fz, Pz usually > Cz), and, when possible, (e) well established relationships with experimental manipulations (in particular, probability and target effects).

In the remaining sections of this chapter, we will discuss several issues related to the theoretical definition of P300. We will then describe the operational definitions used for the study of P300, their usage by researchers in the ERP field, and their reliability. The definition of P300 will be clarified by identifying the conditions under which P300 is usually observed. We have therefore developed a tentative classification of the paradigms commonly used in the study of P300. A detailed description of these paradigms, and of factors affecting the attributes of the P300 elicited in these paradigms, is given in Appendices A and B.

THEORETICAL ISSUES

Scalp Topography and Neural Generators

When the scalp topography of ERPs varies between two conditions, we may infer that the biophysics of the generating processes is different. It is generally assumed that the appearance of differences in topography imply that different generators are responsible for the ERPs. "Evoked potential components with significantly different scalp distributions must derive from different sources. Either different cells are involved in the generation of the scalp-recorded potential or the active cells are differentially responsive" (Picton, Woods, Stuss, & Campbell, 1978, pp. 518-519).

The converse, however, is not true - neural generators may be different in two conditions, yet scalp topography may remain constant. That is, for any of a number of reasons the electric fields associated with distinct neural activities are indistinguishable.

It is very difficult to elucidate generator sources from scalp distributions. The biophysics involved, and individual differences in skull, brain and meninges make this a very difficult enterprise. Several investigators, most noteworthy Vaughan, Ritter and Simson (1980), have been able to derive rather useful inferences on the origin of ERP components. When successful, such techniques capitalize on the known neuroanatomy of candidate generators of the component. Alternate hypotheses regarding the generator lead to competing predictions on the likely scalp distribution of the potentials. These predictions can be tested empirically and thus confirm one of the competing hypotheses. While this approach has proven to hold considerable value in the analysis of exogenous components of the ERP, it is very difficult to apply in the case of the endogenous components as there is little, if any, knowledge on which to base neuroanatomical hypotheses regarding their origin.

The elucidation of the origin of the P300 must depend, therefore, on the study of intracranial analogs of these components and on magnetoencephalographic techniques. Nevertheless, even though it is unlikely to reveal intracranial sources, the scalp distribution is important not only for identifying components, but also for trying to understand more precisely the variations in information processing in different tasks. This is particularly true in the difficult task of isolating a "component." As we noted above, component overlap makes it very difficult to determine whether a morphological change in the waveform represents a change in a specific component from one condition to another, or whether a new component has emerged overlapping with the target component, or whether several components are changing simultaneously. Scalp distribution is a key aide in component identification.

Several methods have been applied to the study of scalp topography. The simplest technique is the direct comparison of the voltage of a designated peak across different electrodes. A version of this technique involves the computation of the ratio between the voltage obtained at each electrode and that of the vertex (Cz) (several examples are presented in Appendix A). This reduces variability among subjects, but necessitates assumptions that are probably unwarranted. These were pointed out by Wood and Wolpaw (1982) with respect to earlier components, but are equally valid for P300.

> This approach assumes that the scalp distribution over the entire duration of each deflection is accurately represented by the distribution at the peak. Such an assumption is justified if the deflection is identical in morphology and latency across electrode locations, and varies only in amplitude. However, if there are changes in morphology or latency across electrode locations, then the scalp distribution at the peak does not adequately characterize the potential fields in question (Wood and Wolpaw, 1982, p. 25).

When an array of electrodes is used, isopotential maps can be created by interpolating between electrodes at one time point (e.g., Simson, Vaughan, & Ritter, 1976; Simson, Vaughan, & Ritter, 1977a, 1977b); or, one can map changing contours over time using equipotential lines (spatio-temporal mapping, "chronotopogram"; Remond, 1961). This technique presents much more information than ratios of the voltage recorded at Cz to other electrode sites. An array of electrodes along the midline, or perpendicular to it, is presented on the ordinate, while amplitude variations over time are displayed on the abscissa (see, for example, Ragot, 1984, and Renault, 1983). This is an efficient way to present a large amount of data. The problem is that no new methods of quantifying the data have emerged, and these complex isopotential maps are used primarily for visual inspection and comparison. A discussion of problems related to the quantification of scalp distribution has been presented by McCarthy and Wood (in press; see also Donchin, 1979.)

Vector analysis (Gratton, Coles, & Donchin, 1983b, 1985) attempts to partially solve the problem of quantifying topographic data. This technique allows the investigator to separate the contributions of particular patterns of scalp distribution. Thus, "filters" for the scalp distribution of a particular component can be used to separate this component from other overlapping components. Alternatively, the contributions of different scalp distributions to an observed component can be dissociated.

Variations in the scalp distribution of P300 may be due to differential activity of overlapping components, or to variations in the population of cells involved in the generation of P300. The latter explanation is related to questions about the neural generators of P300. In recent years, several investigators have proposed that P300 does not arise from a unitary generator, but rather from multiple sources with multiple orientations (Wood, Allison, Goff, Williamson, & Spencer, 1980). This does not necessarily mean that P300 represents multiple psychological processes, as several neural generators may need to be simultaneously activated to carry out what would be considered a unitary psychological process. It does, however, suggest that if the same set of multiple sources can be differentially activated in order to accommodate minor changes in the processing required by a particular task, then this may lead to slight differences in scalp topography, and consequently provide an explanation for the differences observed among paradigms and, perhaps, among subjects. This speculation may blend into another: Do these multiple generators represent subprocesses, with P300 the result of their combination? This view has slightly different theoretical consequences, because it implies that more atomistic psychological formulations are possible. These questions also bring up the issue of whether P300 represents a unitary process. Is there a "core" P300 process that can be combined with various other processes, producing different scalp distributions? Or are there several distinct independent processes that summate on the scalp to produce P300?

To collect data relevant to these questions there have been attempts to record P300 intracranially in animals, and in patients who are about to undergo neurosurgery. With intracranial techniques, several investigators have recorded "P300-like" activity originating in several different brain regions (for a review, see Wood, McCarthy, Squires, Vaughan, Woods, and McCallum, 1984). For instance, Halgren, Squires, Wilson, Rohrbaugh, Babb, and Randall (1980), using intracranial recording, observed electrical activity from the medial temporal region, in proximity to the hippocampus and amygdala. This activity responded to probability and target manipulations in a fashion similar to the scalp-recorded P300. Okada, Kaufman, & Williamson (1983), using magnetoencephalographic recordings, also claimed to have observed P300-like activity originating in the hippocampal region. However, data have also been reported that appear incompatible with the hippocampal hypothesis. For instance, Curry and McCallum (unpublished data reported in Wood et al., 1984) recorded P300-like activity originating from other brain regions in epileptic patients, including the posterior parietal cortex and thalamic structures. Wood, McCarthy, Allison, Goff, Williamson, and Spencer (1982) found that the scalp distribution of P300 was not affected by unilateral temporal lobe excisions which included the hippocampus. Similar results have been reported by Paller, Zola-Morgan, Squire, & Hillyard (1984) on monkeys. Thus, although P300-like activity has been observed from the hippocampal region, it has also been observed from other structures and, because the scalp-recorded P300 is not affected by hippocampectomy, it is unlikely that the scalp P300 is actually generated only in the hippocampus.

An alternative explanation is that a single neural generator, projecting to several brain structures, may be responsible for the activity recorded at the scalp. The scalp-recorded P300 would then be associated with diffuse cortical post-synaptic activation, concurrent with similar activation in the hippocampal region. This explanation is compatible with the position that P300 is a unitary phenomenon.

Even if P300 is a complex electrical phenomenon, with multiple "independent" generators, it may still behave as a functional unit. In fact, as we mentioned above, the activity in different regions may all be necessary to perform a single unitary cognitive operation. The question of whether P300 represents a unitary process is an important issue, and one to which we return.

P300: A Unitary Process?

Several positive components have been described as being active in the same time range as P300 (e.g., "P3a", a frontal P300, Slow Wave). A detailed discussion of these components is presented in Appendix B.

Pointing to the existence of these multiple late positivities, several investigators claim that P300 is not a unitary process (Courchesne, Hillyard, & Galambos, 1975; Friedman, Vaughn, & Erlenmeyer-Kimling, 1981; Picton & Stuss, 1980; Rösler & Sutton, in press). For instance, Courchesne et al. (1975) write, "...'the P3' wave is not a unitary phenomenon but should be considered in terms of a family of waves, differing in their brain generators and in their psychological correlates" (p. 142). Similarly, Friedman et al. (1981) write, "It is evident that the identification of multiple overlapping late components that differ in their relationships to task variables makes it impossible to regard P300 as a unitary phenomenon and obviates the effort to relate 'P300' to a single psychological construct" (p.647). However, even though the presence of multiple overlapping components makes measurement more difficult, it does not follow that there is no unitary P300 component. In fact, as Friedman et al. (1981) go on to say, "The presence of multiple positive components within the P300 latency time range allows the possibility for finer functional distinctions between the classical P300 component and other positivities that have recently been reported" (p.647).

In part, the problem results from a lack of clarity about exactly what is meant by "P300." ERP waveforms are complex, and the late positive activities reported in many studies differ widely. The P300 often overlaps with other components, such as the Slow Wave. Several of these components are sometimes called P300, but have different scalp topographies and react differently to experimental variables. We suggest keeping these components distinct from the "classic" P300. The claim that the late parietal or central P300 is not unitary, in the sense that it represents independent psychological processes, loses most of its impact when these other components are treated separately. The question should be, "Is there a unitary process in all this babel of late positivities?" We believe that the answer is yes. In fact, "P3a" and the frontal P300 of Courchesne et al. (1975) should be considered separate components, provided that they can be observed and measured reliably. Small differences in scalp distribution of the "classic" P300 may be due to the activation of auxiliary processes or to overlapping components. If we operationally define P300 as Pz or Cz maximum, with latency greater than 300 ms, and acknowledge an overlap with various slow wave components, we are left with a component with basic similarities across paradigms. There is a communality in practically all the basic theoretical formulations, and we believe that in most paradigms common neural generators, performing identical information processing, are activated.

Early research focusing on the existence of a "unitary P300" examined whether P300 was modality specific, whether P300s to omitted stimuli were similar to physically present targets, and whether scalp topographies were similar in different experimental paradigms. The general conclusion from these studies has been that the scalp distribution of P300 is not modality specific (e.g., Snyder, Hillyard & Galambos, 1980; Picton, Stuss, Champagne, & Nelson, 1984), and that there are similar topographies when stimulus omissions are targets (Simson et al., 1976, Ruchkin, Sutton, Munson, Silver, & Macar, 1981).

Examining P300 in different experimental paradigms is a valuable exercise, and it would be very useful to extend work such as that of Hillyard, Courchesne, Krausz, & Picton, (1976). In that study scalp topographies were compared in six different experimental contexts. We need studies which challenge subjects with a variety of complex information processing tasks, and address the problems of changes in scalp distributions and overlapping components as tasks change. Only when we have the variability yielded by a wide variety of paradigms will we be able to clarify the relationships between ERP components and various cognitive subprocesses.

We believe that a unitary P300 may still exhibit small variations in scalp distribution across tasks (and subjects). It is difficult to define "small" precisely, but we would not, for example, label a positivity with a frontal maximum as P300. We would interpret a frontal positivity as a different component, or the product of an overlapping component. Since there are differences among paradigms in the nature of their information processing demands, in the cognitive "effort" required, and in the activation of particular motor and sensory systems, it is understandable that there is variability in scalp distribution. We need to understand the meaning of these differences, and identify the cognitive subprocesses that shift or alter scalp distribution. Ideally, one would like to be able to perform a very fine grained task analysis, and then create a model able to predict scalp topography given details about the subprocesses required, and other dimensions of the task. In the following section we propose a taxonomy of P300 paradigms to facilitate such model building. The ability to identify basic individual differences, and incorporate these into the model, would of course be desirable. Such a model is too ambitious at the present, but we can start to detail some basic factors that will have to be incorporated into the model. These include differences in P300 that depend on the modality of the eliciting stimulus, and age and sex differences. The effects of these variables on P300 are reviewed in Appendix C.

A TAXONOMY OF P300 PARADIGMS

Many different paradigms are used in the study of P300 and other late positivities. In Appendix A we describe the most commonly used paradigms, and note how the choice of paradigms affects the variance of P300, particularly in terms of scalp distribution. It would be useful to organize this diversity of paradigms in a manner that would identify critical dimensions that influence P300 and other endogenous components.

A taxonomy of paradigms will help in clarifying the specific information processing activities that are manifested by various components. There are many names for late positivities, and researchers keep "discovering" more. How are these deflections related, and do they share common processes? Are these deflections merely variants of the same component whose morphology changes due to overlap with other components? It should be possible to analyze any paradigm, fit it into our taxonomy, and predict the ERP components that are likely to be recorded.

To create a taxonomy of ERP paradigms we begin with the working hypothesis that virtually all paradigms in which P300 is observed are instances of one canonical paradigm. The paradigm is allowed to vary along a number of different dimensions. Each class of experiments in which P300 can be recorded can be positioned in the space defined by the dimensions along which the experimenter is free to make choices. Thus, to obtain a useful taxonomy we must identify dimensions that can be varied independently. Ideally, these dimensions should: (a) be exhaustive (i.e., thoroughly describe past and, possibly, future experimental paradigms), (b) have classificatory power (i.e., allow comparisons and clustering of several different paradigms), and (c) be based on well validated physiological and psychological constructs.

In virtually all P300 studies a series of events occurs; each event can be classified into one of several categories, and usually the subject must respond in some fashion to events in one of the categories. This "oddball" experiment is one simple version of the canonical paradigm. We propose the following six dimensions along which versions of this paradigm can vary, and outline some of their subcomponents:

1. Sequence Generating Rule
    - A. random or constrained
    - B. probability
2. Trials
    - A. number of stimuli and temporal relationships
        1. among stimuli in a trial
        2. among trials
    - B. stimuli
        1. modality
        2. physical properties (e.g., intensity) and quality
        3. complexity, structure and content
3. Mapping Rule (stimulus to response)
    - A. memory load requirements

B. nature of processing and transformation required

C. complexity of response selection

4. Response Execution

A. covert or overt

B. discrete or continuous

C. ballistic, or involving feedback

5. Consequences (e.g., payoffs)

6. Subject's characteristics (e.g., age, sex, handedness, neurological and psychiatric status)

The sequence generating rule determines how different types of trials and events are combined, and the probabilities that will be employed. The second dimension, trials, is quite broad, because it includes not only the nature of the stimuli in terms of modality and complexity, but also the number of stimuli that define an event, and the temporal relationships among them. For example, an event may comprise several stimuli, as in S1 - S2 paradigms, and there may be more than two categories, although usually these can be reduced to two (e.g., targets and non-targets). The mapping rule directs the categorization of stimuli so that the proper response can be made. Considerable processing and transformations are often required before categorization can occur, and these operations involve the use of varying amounts of information held in working memory. The subsequent response selection is also included here. Response execution may involve overt behavior such as a button press, or covert activity such as counting instances of a particular class of events. In these cases responses are discrete, but in other paradigms they may be continuous. For example, pursuit tracking is continuous and overt, whereas memorizing a list of words is often continuous and covert. Consequences usually include payoffs or rewards, and may also involve modifications of a subject's processing in the future course of the experiment. Subject characteristics include individual differences that may account for some part of the variance in P300. The effect of aging is one of the most extensively studied, although ERPs have also been recorded in several clinical populations (see Donchin & Bashore, in press). A description of some of these factors is presented in Appendix C.

An advantage of this taxonomy is the possibility of comparing experiments that differ both in their theoretical framework (e.g., signal detection, recognition, classic oddball) and in the experimental variables used. As an example, we will compare Duncan-Johnson and Donchin's (1977) experiment, in which the a priori probability of stimuli in an oddball task was varied (we will refer to this as experiment A) with McCarthy and Donchin's (1981) experiment, where stimulus discriminability and stimulus-response compatibility were varied (experiment B). We can describe the paradigms as follows, using the dimensions of the taxonomy:

1. Sequence generating rule: In both experiments trials were selected at random. However, in experiment A probability varied widely, while in B all conditions were equiprobable. We thus expect that the largest P300s may be elicited in the low probability conditions of experiment A. This is indeed the case.

2a. Trials - temporal relationships: Experiment A was a typical oddball task, with one event presented every 1.5 seconds, while B involved a pair of stimuli on each trial (what we call S1-S2 paradigms in Appendix A). The word "SAME" or "OPPOSITE" was presented at S1, and "RIGHT" or "LEFT" was presented 1000 ms later at S2. We expect a CNV in the S1-S2 interval, and one is clearly present in their Figure 1. Such a CNV may have an influence on the scalp distribution of the P300 generated after S2.

2b. Trials - stimuli: In A, simple auditory stimuli are used (binaural tone bursts). Tones are either low (1000 Hz) or high (1500 Hz). In B words are presented visually, in a matrix of letters (noise condition) or # signs (no noise condition). Since the stimuli in the two experiments differ in complexity, discriminability, and modality, there should be large differences in P300 latency. This is the case, with P300 latencies in B several hundred ms longer than in A.

3. Mapping rule: Both experiments have a low memory load. In A there are also minimal processing demands, while in B the subject must decide the meaning of S2 based on S1.

4. Response execution: In both experiments there are discrete responses, but in A they are covert (increase an internal count), while in B they are overt (press one of two buttons).

5. Consequences: In A there are bonuses for reporting the correct number of high tones, while in B there is no mention of any bonuses.

Using the taxonomy we have been able to identify both similarities and differences between these two experiments. Of course, a comparison between these two experiments was possible even without the taxonomy. However, in dealing with a large number of experiments a taxonomy is very helpful for organizing the relevant information. Most researchers develop an internal model of such a taxonomy, but it is rarely made explicit. As the number and complexity of experiments increases it becomes more difficult to develop and maintain an accurate model, and to use it to catalogue experimental findings. Making a taxonomy explicit can have many functions. It can:

a. increase accuracy in communication by providing an accepted framework.

b. help identify the critical variables in ERP experiments, and their interrelationships, by making clear the multidimensional nature of experiments and the necessity of considering several factors simultaneously.

c. aid in experimental design by providing a list of variables that should be considered.

d. give perspective to a body of research by making clear exactly what values are used on each dimension.

e. draw attention to "details" missing from published papers that may be important (e.g., was there a system of bonuses in experiment B?).

Several typical ERP paradigms are described in Table 1 using this taxonomy. Within most of these paradigms there is substantial variability; Table 1 presents the most common cases. Changes in the dimensions of the

----

Insert table 1 About Here

----

taxonomy may influence P300 amplitude, latency, and scalp distribution. This is presented in Table 2. (Many of the experimental studies that provide the data for this Table are reviewed in Appendices A and B.)

Insert table 2 About Here

This table represents only a tentative summary of the effects of changes in the dimensions of the taxonomy on P300 parameters, just as Table 1 was only a beginning of classification of paradigms. We must further elaborate the categories of importance within each dimension of the taxonomy. In particular, there must be a more sophisticated analysis of the task demands required by various mapping rules. However, we believe that providing such instruments of reference may greatly improve the ability to interpret experimental data. In fact, we could compare the paradigm used in a particular experiment with the dimensions of the taxonomy and the typical paradigms described in Table 1, and the parameter of P300 observed in the experiment with those provided in Table 2. By means of these comparisons we could evaluate whether our data confirm or disconfirm the predictions about P300 parameters derived from the taxonomy, and eventually refine the taxonomy by identifying new relevant dimensions. However, to increase the accuracy of the predictions, we will probably have to add some additional columns in Table 2 devoted to Slow Wave and other late components, and then determine how they interact with P300.

OPERATIONAL DEFINITIONS OF P300

Having described the experimental settings in which P300 has been recorded, it remains the case that the waveforms obtained in each experiment need to be assessed for the presence of the P300. Once identified, the component's latency and amplitude need to be measured. Many procedures have been used to operationally define the P300. In this section we review these procedures.[4]

Investigators have used a variety of combinations of the four features proposed by Donchin et al. (1978) to operationally define P300. Such variety reflects, in part, disparity of opinions about the invariance of the physical characteristics of P300. For instance, the invariance of P300 scalp distribution has been questioned, and, as already mentioned, P300s with maxima at the parietal, central, or frontal electrode have been described. Another reason for this variety is that the different operational definitions of P300 each have their own advantages and disadvantages.

The operational definitions of P300 adopted by investigators vary in complexity, from those making minimal assumptions and requiring little computational load, to those making several assumptions about the invariance of the characteristics of P300 and requiring large computational loads.

A minimal operational definition of P300 may simply state that "P300 is a positive component with a latency in excess of 300 ms." With this definition, whatever positivity occurs after the 300 ms landmark is automatically attributed to P300. This definition makes only two assumptions about P300: that P300 has a positive polarity, and that P300 has a latency longer than 300 ms. The addition of a terminal point, after which the activity cannot be attributed to P300, identifies a time window for the P300. It also allows the experimenter to estimate the magnitude of P300 by integrating all the activity recorded in the time window (an area measure).

By narrowing the analysis to a particular scalp location, many investigators introduce a further assumption about P300 (i.e., that the scalp location chosen is the most representative one).

A more complex definition assumes that "P300 is the largest positive peak after 300 ms, or in a particular time window" (for example, between 300 and 800 ms). This definition corresponds to a peak identification procedure. This definition (as the preceding one) has often been modified with the further restriction that the peak has to occur at a particular scalp location (most often at Pz, though sometimes at Cz). Note that we have introduced the concept that P300 is characterized by a peak. Peak measures have the advantage of permitting measures of both amplitude and latency, coupled with the use of very simple detection algorithms. However, the validity and reliability of peak-picking, in particular for amplitude estimations, has been questioned, mainly on the basis of two arguments (Donchin and Heffley, 1979): (a) peak measures do not allow a discrimination between overlapping components (this criticism also apply to area measures); (b) peak measures may be unreliable because they are based on single point estimates. The use of area measures was proposed to alleviate the supposed unreliability of peak-picking in amplitude estimates. Donchin and Heffley (1979) present peak, area and other measures for the estimation of P300 amplitude as instances of linear filters. A linear filter is a linear combination of the voltage values recorded at each timepoint, with a vector of weights. These weights (one for each timepoint) characterize the type of linear filter used. Examples of the vectors of weights corresponding to several operational definitions of P300 are shown in figure 1. This

----

Insert figure 1 About Here

----

conceptualization provides a common ground for the comparison of several procedures (e.g., peak and area measures, Principal Component Analysis, Stepwise Discriminant Analysis). An extension of this conceptualization will be presented later.

A further problem is the appropriate choice of the time window. In some cases, other positive components may occur at the limits of the time window. If the choice of the time window is inappropriate, some of these components may be mistakenly identified as P300. This will produce large errors, particularly in latency estimates.

Complex operational definitions of P300 use linear filters obtained to produce an optimal solution for some specific criteria. Two such procedures have been frequently applied, PCA and Stepwise Discriminant Analysis (SWDA).

Principal Component Analysis is a multivariate procedure which identifies axes of maximal variance in a space defined by a set of variables. The axes are chosen to explain most of the variance and covariance between the variables with the smallest possible number of orthogonal axes. These axes, labelled components, are described by the cosines with the original variables. Such cosines are labelled component loadings (the set of cosines for each component is labelled an "eigenvector"), and express the "similarity", or correlation, between the components and the original variables. An orthogonal rotation may be used to maximize (or minimize) the loadings, in order to satisfy a simple structure criterion. Varimax rotation has been the most commonly used, but other rotational procedures are also possible, and may in some cases be preferable (see, for example, Tatsuoka, 1971, for a detailed description of PCA and other multivariate procedures).

PCA has been applied to ERPs to identify independent components of variance (Donchin, 1966, 1969a). In this case, timepoints are used as variables. Each component can therefore be described by a set of loadings, one loading for each timepoint. Donchin and Heffley (1979) describe the use of the component loadings as linear filters, to obtain estimates of component magnitude. Two other applications of PCA to ERP analysis were described by John, Ruchkin, & Villegas (1964; see also John, Ruchkin, & Vidal, 1978), and Skrandies & Lehmann (1982). These investigators propose the use of subjects or scalp locations as variables, instead of time points.

We would like to stress two important considerations in the use of PCA to identify and quantify components. Both of them are based on the distinction between PCA components and ERP components. First, PCA defines components by means of a fixed set of component loadings, one for each time point. The assumption is that the time characteristics (latency and waveshape) of P300 are fixed, at least for all the waveforms used in the analysis. If the latency (or the waveshape) of P300 varies across the waveforms used for the analysis, than variance produced by these latency shifts will be a source of some of the components. In other words, the components extracted by PCA in this case are likely to produce an unfaithful picture of the component structure underlying the waveforms (see Donchin & Heffley, 1979).

Second, PCA identifies components which maximally contribute to the variance/covariance matrix. Components are defined on the basis of the covariances between timepoints. A series of timepoints which varies together will produce a PCA component. Therefore the set of components extracted by PCA depends on the variability of the set of waveforms entered in the analysis. In particular, two ERP components will be dissociated by PCA only if they vary independently in the original waveforms.

Thus, it is clear that there are limitations associated with the use of PCA to identify and quantify ERP components. However, the use of PCA in the study of P300 has two advantages. First, PCA permits some discrimination between overlapping components, which is not possible with peak-picking or area measures. Such an advantage has been clearly illustrated by Donchin and Heffley (1979). Recently, Wood and McCarthy (1984), in a simulation study, reported cases in which PCA dissociated overlapping components, but misallocated variance and inflated the probability of type I error. However, in the absence of procedures able to produce a better discrimination between overlapping components, PCA remains the first (though not optimal) choice in cases of overlapping components and fixed component latency.

The second advantage of using PCA is that the estimates of P300 amplitude are based on several timepoints. Furthermore, the weight for each timepoint (component loading) reflects the contribution of the component to the variance of that timepoint. As already mentioned, the component loadings obtained with PCA might be considered as a sort of linear filter adapted to the component. This is true only in the case that all the assumptions of PCA are met. There are other assumptions of PCA we have not yet discussed. For instance, PCA assumes that the error at each timepoint is not only random, but also uncorrelated with the error at any other timepoint. Such an assumption is usually not met in the case of ERPs, and there has been some debate on the seriousness of this violation (Hunt, 1980). The amplitude estimates of ERP components obtained with PCA are supposed to be more "reliable" than the magnitude estimates obtained with peak-picking. A comparison of the reliability of the estimates of P300 magnitude obtained with PCA, peak-picking, area measure, as well as other procedures, will be presented later in this chapter.

Step Wise Discriminant Analysis is a stepwise regression procedure which identifies a subset of variables that will optimally discriminate between two (or more) sets of data. The procedure is iterative. Initially, the variable which best discriminates between the sets of data is identified, and the appropriate regression weight is computed. Then the variance attributed to this variable is partialled out, and the procedure is repeated, until a certain criterion is reached. The criterion may be a certain F value, or number of iterations, asymptotic level, etc. (for a more detailed description of SWDA as applied to ERPs, see Donchin, 1969b; Donchin & Herning, 1975; Horst & Donchin, 1980; Squires and Donchin, 1976).

SWDA has been applied to ERPs to identify subsets of timepoints able to distinguish between two or more experimental conditions. In particular, Squires and Donchin (1976), and Johnson and Donchin (1980, 1982) applied SWDA to find which timepoints best classified trials from an oddball paradigm into "rare" and "frequent" categories. The weights obtained with SWDA can be applied to obtain magnitude estimates (e.g. Squires, Wickens, Squires, & Donchin, 1976), both on the waveforms used for the analysis and/or on an independent sample of waveforms. Donchin and Heffley (1979) showed how such a set of weights can be conceptualized as a linear filter for the magnitude of a component.

The definition of component obtained by SWDA is based on two features of a component (in our case, P300): (1) the ability to distinguish between two (or more) experimental conditions, and (2) the invariance of its time characteristics. Note that the definition of component obtained with SWDA does not consider as relevant the contiguity of the timepoints belonging to the component. Actually, a component, as defined by SWDA, may be made up of timepoints belonging to very different sections of the waveform. Note also that the number of identifiable components cannot be larger than the number of experimental groups, minus one. As a consequence, the applicability of SWDA to estimates of amplitude is limited. An advantage of SWDA is that it provides a description of those timepoints that best discriminate between conditions. However, such a discrimination criterion could also be obtained with a Multivariate (Fisherian) Discriminant Analysis (MDA). This procedure is based on a "simultaneous" (rather than "stepwise") solution that may provide a structure more interpretable than SWDA. In fact, given the hierarchical nature of SWDA, too much reliance on this description could lead to misinterpretation. For example, two very correlated timepoints will not likely be chosen as representative of the same component. Note also that, given the "a posteriori" nature of both SWDA and MDA, these procedures require a cross-validation.

Other procedures have been proposed to operationally define P300. We will mention a few of them, including Cross-correlational techniques and Woody filter, and Vector filter.

Cross-correlational techniques are based on an analysis of the similarities between the ERP waveform and a wave segment, called "template." The template is chosen in order to have a waveshape similar to the component of interest, P300 in our case (knowledge about P300 waveshape may be derived from other studies, PCA, difference waveforms, etc.). The procedure determines which segment of the ERP waveform has the maximum correlation with the template. This ERP segment is classified as P300. A minimum correlation criterion can be set, and ERP waveforms for which no segment reach the criterion level are said not to contain the component. An estimate of P300 magnitude can be obtained by computing the cross-product between the template and the segment of ERP waveforms classified as P300. Such estimates can again be conceptualized as a linear filter, where the template corresponds to the vector of weights of the linear filter.

Since cross-correlation defines P300 by means of its shape, the shape of P300 is assumed to be fixed. Therefore, the choice of the correct template becomes critical for the validity of the procedure. Woody (1967) proposed a technique to generate a template in the case in which the shape of the component is not known a priori. This technique, called Woody filter, is based on an iterative procedure. Initially, an arbitrary template is chosen (in the original paper, Woody proposed the use of an average waveform). Then, the waveforms are shifted, in order to align the wave segments with maximum correlation with the template. Then an average waveform is computed, and used as a template for a new iteration. The procedure is repeated until the template does not show any change from the previous iteration. The advantage of the iterations (and therefore of Woody filter in comparison with cross-correlation) has been investigated by means of simulation studies (Wastell, 1977).

Cross-correlation and Woody filter have been particularly useful for the estimation of single trial latency, in cases where other problems (e.g., very small signal-to-noise ratio) make the use of techniques such as peak-picking particularly unreliable. Recent data (Gratton, Kramer, Coles, & Donchin, 1985) indicate that in such cases cross-correlation is particularly advantageous, provided that a correct template has been used.

Vector filter is a procedure proposed by Gratton, Coles, and Donchin (1985; see also Gratton, Coles & Donchin, 1983b). This technique defines components by means of their scalp distribution. A set of weights, one for each scalp location, is used to estimate the magnitude of the component at each timepoint. This procedure can be conceptualized as a linear filter, where values obtained at different scalp locations are appropriately weighted. It contrasts with the traditional procedure of channel selection, in which a weight of one is attributed to the channel selected, and a weight of 0 to all the other channels. Vector filter can be used in conjunction with other procedures, such as area measures, peak-picking, SWDA, cross-correlation or Woody filter. The set of weights can be chosen a priori, or in such a way so as to satisfy particular criteria (ability to discriminate between sets of data, to explain maximum amount of variance, etc.). Vector filter defines P300 by means of its scalp distribution. However, if coupled with other techniques, the assumptions of these other techniques should also be considered for the definition of P300.

General Comparison Between Operational Definitions

The procedures and the corresponding operational definitions presented above will be subdivided into two categories: (a) those using features of the time series (i.e., amplitude at different timepoints) to define P300, and (b) those using scalp distribution features to define P300. The first group includes area measures, peak-picking, PCA, SWDA, and cross-correlation techniques, while the second group includes Vector filter and channel selection. The comparison between the latter two techniques has been presented in the previous section. We will attempt to provide a general framework under which all the procedures in the first group (and the corresponding operational definitions) can be classified. Such a framework is intended to facilitate the comparison between the different procedures, and provide an adequate basis for the analysis of the corresponding operational definition of P300 and of their assumptions.

As we discussed above, Donchin and Heffley (1979) considered different procedures for P300 amplitude estimation as different instances of linear filters. We intend to expand this analysis by considering the different procedures used to detect and quantify P300 as particular instances of pattern recognition techniques,[5] based on the analysis of the cross-products of segments of the ERP waveform with a template (i.e., vector of weights).

Table 3 presents a comparison of the different procedures on the basis of such a framework. (For a graphic representation see also figure 1.) The

----

Insert table 3 About Here

----

procedures may differ at each of three levels: data preparation required prior to the analysis, the type of template, and whether or not the segment of the waveform under consideration is shifted.

Data preparation. Some of the procedures imply a transformation of a segment of the waveform, before the actual cross-product is computed. For instance, before applying cross-correlation, each segment of waveform is normalized. This normalization implies that size is not a defining characteristic of P300, but shape is.

Template. The procedures differ in the choice of the template used for the computation of the cross-products. Area measure is based on the use of a rectangular template. Peak-picking is based on a template consisting of only one point (or, only one point has a value different from 0). Cross-correlation allows the experimenter to choose any template. Woody filter adapts the template to the average waveform. PCA and SWDA optimize the template in order to satisfy some criteria. The choice of a given template reflects some a priori assumptions about the nature of P300. Such assumptions were presented above. However, comparing templates may help in understanding the differences among procedures.

Shifting of the waveform segment. Some of these techniques involve the comparison of the cross-products obtained with segments progressively shifted by one time unit. Examples of such cases are peak-picking (where the template is reduced to a single point), cross-correlation and Woody filter. These procedures do not assume that P300 time characteristics are fixed, and in fact permit their measurement. On the other hand, area measure, PCA, and SWDA, use only one segment of the ERP waveform, without any shifting. These procedures assume that P300 time characteristics are fixed.

To conclude this section, we note that simple operational definitions of P300 need fewer assumptions, and allow for more phenomena about P300 to be studied (e.g., not only amplitude, but also latency and scalp distribution can be used as dependent variables). On the other hand, simple definitions do not always provide the investigator with a tool powerful enough for discriminating overlapping components, or components from background EEG noise. These discriminations may be improved dramatically by more sophisticated definitions. Since the influence of the measurement error may differ for different operational definitions of P300, their reliability may also be very different.

In the next section we will present data on the frequency of various operational definitions in the recent literature. Then, we will present data on the reliability of P300 estimates obtained with some of the procedures presented above.

Operational Definitions of P300 in the Recent Literature

We examined four and a half years of Psychophysiology (January, 1980, 17(1) to September, 1984, 21(5)) to determine how P300 has actually been defined. This is not intended to be an exhaustive literature review, but a representative sample of the papers published during this time period. There were 34 articles in which P300 was measured, excluding special presentations (SPR Presidential Addresses, Young Psychophysiologist Award Addresses) and methodological notes.

The types of measurements and the electrode placements used are presented in Table 4a and 4b.

Insert table 4 About Here

There are several notable aspects to these tables. First, baseline-to-peak (base-to-peak) measures predominate for the measurement of P300 amplitude. This is the simplest technique, but also the one that uses the least information, that is, a single value from a single electrode for the peak amplitude, measured from the baseline. The second most common technique is PCA, and five of the eleven studies which use PCA also use another technique. In all, only 6 of the 34 studies use more than one technique. Peak-to-peak, area, and the use of discriminant functions are all fairly rare, which seems unfortunate, since these techniques all have certain advantages. Although area measures are not commonly used for P300, area (along with PCA) is frequently used to measure Slow Wave. Slow Waves rarely have a prominent peak and may continue for many hundreds of milliseconds, so base-to-peak measures are clearly undesirable. Six studies (18%) use only a single electrode at vertex; this is unfortunate because information on scalp distribution is very important. Sixty-two percent of the studies (21) include three midline electrodes, Fz, Cz, and Pz. These three electrodes are essential not only for identifying P300, but also for separating it from

Slow Waves and other late positivities that may be present.

As a general guideline, the use of multiple analyses is recommended, because different techniques often reveal different aspects of the data. (It is, of course, not acceptable to use several techniques in an attempt to reach a certain level of statistical significance.) Tukey (1978), in a commentary on data analysis, made several points that remain valid. The emphasis in ERP analysis, he feels, is on single measures, with investigators asking themselves, "...which single (and simple) derivation does the best?" (p.142). He points out that other fields have not progressed until techniques were developed for more thorough data analysis. It is also important to look at more than one "expression of the data." He made a special point (p.142) of suggesting that "much can be gained" by combining the information from several electrodes. The development of vector analysis is an initial attempt to meet this need (Gratton, Coles, & Donchin, 1983b, 1985).

Problems arise when investigators do not explicitly define P300. This is particularly clear when PCA is used, because the problem arises of how to identify which PCA component is the "P300". We suggest examining not only peak latency, but also time course and scalp distribution. In some cases, response to experimental manipulations can be used to help identify P300. This is practical when the paradigm includes conditions which vary in probability, or other dimensions along which P300 amplitude is known to vary. In most cases P300 will be one of the first few components of PCA, and account for a large percentage of the variance.

Two other serious problems with PCA are overinterpretation and misidentification. Overinterpretation occurs when investigators insist on interpreting every component produced by their PCA. Choosing the number of components to rotate, and deciding how many to interpret, is not easy or straightforward. Even though one may have rotated the "proper" number of components statistically, this is no guarantee that all these components will be meaningful on a psychological or a biological level. Given the small number of waveforms recorded in nearly all ERP experiments, the component structure is highly unstable. One should examine the time course of the component, its scalp distribution, and its response to experimental manipulations. PCA alone should never be used to "discover" new components, but only to measure existing components that the investigator can identify with confidence (for example, by applying the criteria we have indicated above). This leads to the second problem, misidentification. In some experiments a component is said to be P300 merely on the basis of latency, and then it is stated that this P300 has an unusual scalp distribution, and does not behave like a P300. The authors may then state that this is additional evidence that P300 is not a unitary phenomenon. A more likely interpretation is that P300 was misidentified in the component structure of the PCA.

THE PROBLEM OF RELIABILITY OF P300 MEASURES

This section is devoted to the presentation of data on the reliability of several estimates of parameters of P300. We emphasize that although reliability is a desirable feature for an operational definition of P300, it is not the only criterion for judging the relative merit of different procedures. In general, an operational definition of P300 should satisfy these ideal criteria: (a) feasibility (it should be applicable to actual data), (b) reliability (if applied to data obtained under the same conditions from the same subject, it should reproduce the same results), and (c) validity (it should correspond to the "real" characteristics of P300).

It is difficult to test all these criteria. In particular, the validity criterion is not directly testable, given that we do not know what the "real" characteristics of P300 are. Therefore, a discussion about the validity of an operational definition of P300 remains speculative, and depends on the theoretical approach chosen in the analysis of ERPs. Researchers in the field use different approaches. If a psychophysiological approach is adopted (Donchin, Coles, & Gratton, 1984), validity would be inferred by the convergence of the findings about P300 with a plausible psychological model. If, instead, a neurophysiological approach is adopted (see Wood & Allison, 1981), validity would be inferred by directly measuring P300 at its source, by means of intracranial recording or the use of magnetoencephalography. Ideally, the evidence from these two sources should converge. However, since this convergence has not yet taken place, uncertainty over the validity of operational definitions of P300 remains.

The feasibility criterion is easily verifiable. In fact, it consists of the possibility of implementing an algorithm corresponding to the definition. Note, however, that the possibility of implementation may vary among laboratories, because of the computational and instrumental facilities available. Also, the benefit obtained with very complex algorithms may not justify their cost in time and money. Such choices remain entirely with the investigator.

The question of the reliability of an operational definition of P300 can be, at least in part, answered. Reliability can be assessed by comparing the parameters estimated with procedures corresponding to some operational definition of P300 from two samples of trials recorded from the same subjects under comparable conditions. Next we present data concerning the reliability of P300 parameters estimated with some of the operational definitions of P300 presented earlier in this paper.

The Reliability Study

The data for this study were collected within the general framework of a project concerned with predicting subjects' performance in a complex perceptual/motor task on the basis of ERP measures (Karis, Coles, & Donchin, 1984). A fundamental question of this study was to determine whether differences in P300 observed in different subjects were attributable to "true" between-subject variability in P300, or to random fluctuations. Therefore, an assessment of the reliability of the between-subject variability in P300 was required. As several operational definitions of P300 are available, we were interested in comparing the reliability of these measures of P300.

We focus here on the reliability of the P300 parameters (latency and amplitude) estimated within one experimental session (within-session reliability) and over two sessions, one to two weeks apart (between-session reliability). The within-session reliability assesses the amount of variance which may be attributed to systematic differences between subjects, rather than to random fluctuations. The between-session reliability assesses the stability over time of the estimates obtained for each subject.

Although this latter form of reliability does not directly address the question of the reliability of the P300 estimates (and therefore, of the operational definition used to obtain them), it is relevant to the question of individual differences in P300 parameters.

Several oddball tasks were used in the study. They included simple auditory and visual discrimination tasks, as well as a more complex visual task. The visual tasks required the subject to count one of two classes of stimuli, while the auditory task involved choice reaction time (CRT). This range of tasks allowed us to compare the reliability of the P300 estimates across several experimental conditions. One operational definition of P300 might be more appropriate for certain experimental conditions, and another more appropriate for others. Such apparent contradictions may reflect different characteristics of the P300 (amplitude, latency, scalp distribution, duration, waveshape, signal-to-noise ratio, etc.) due to different experimental conditions.

As in many oddball paradigms, our tasks consisted of a random series of rare (20%) and frequent (80%) stimuli, with the subject instructed to either count one stimulus, or differentially respond to both. Rare stimuli elicit large P300s, and frequent stimuli elicit small P300s (see Duncan-Johnson & Donchin, 1977, for a parametric study of stimulus probability effects).

In most oddball tasks, the rare stimuli are also the targets, confounding probability and target effects. In this study, we included conditions in which the subjects were instructed to count or respond to the frequent stimuli. These conditions allowed us to assess separately the effects of probability and task relevance on P300. Both manipulations have been shown to affect P300 amplitude (see Johnson & Donchin, 1978)

Most research on P300 has focused on variations in P300 amplitude and latency. However, the study of P300 scalp distribution has also gained attention. Most researchers (see, for instance, Donchin et al., 1978) consider P300 scalp distribution to be unaffected by experimental manipulations, in particular by stimulus modality. P300 scalp distribution is usually maximum at the parietal electrode, and progressively smaller at the central and frontal electrode, but still positive over all midline sites. There are, however, variations in this modal pattern, and reports of dramatic differences, which we have mentioned above. In this study, we were interested in assessing the use of scalp distribution information for the detection and measurement of P300. In particular, we evaluated the merits of Vector filter, a procedure recently developed by Gratton, Coles, & Donchin (1983b, 1985). This procedure involves combining information provided by several scalp electrodes by using a set of weights, one for each electrode. Vector filter was compared with the information obtained from a single electrode (channel selection). In particular, we compared Vector filter with estimates obtained at the parietal electrode, where P300 is usually maximum.

Method

Subjects

Fifty males between the ages of 18 and 31 served as subjects. All were run for one session, and twenty of them also participated in an additional session. However, several records were lost or had to be discarded,[6] and sample sizes for several analyses were reduced. All subjects were right handed, with normal or corrected to normal vision and hearing, and were paid for their participation.

Tasks

Each session consisted of a series of five oddball tasks. The first oddball task used auditory stimuli, the other four tasks visual stimuli. Tones for the Auditory oddball task were produced by a Schlumberger sine-square audio generator (model SG-18A), and administered binaurally through headphones. All the visual stimuli were presented on a DEC VT-11 display.

In the first oddball task (Auditory Oddball), the subjects were presented with a series of 150 fifty ms tones (145 in session 2). The tones were of two pitches, either high (1500 Hz) or low (1000 Hz). One tone occurred 80% of the time (frequent) and the other 20% of the time (rare). The subject's task was to respond by pressing a button with one hand to the high tone and with the other hand to the low tone. Rare tone and response hand were counterbalanced across subjects.

In each of the second and third oddball tasks (H/S oddballs), subjects were presented with series of 120 letters (115 in session 2), each lasting for 100 ms. The letters H and S were used, each subtending a visual angle of .5 degrees. In both these oddball tasks, one of the letters occurred on 80% of the trials (frequent), and the other letter appeared on the remaining 20% (rare). The subject's task was to keep a running count of one letter in the second oddball and of the other letter in the third oddball. The target letter was always rare in the second oddball task and frequent in the third.

In the fourth and fifth oddball tasks (Name oddballs), subjects were presented with a list of 105 names (100 in session 2). Each name was presented for 200 ms. The length of each name varied between 3 and 9 letters, with each letter subtending a visual angle of .5 degrees. The names were either of males or of females, and no name was ever repeated. In both oddball tasks one of the two categories occurred 80% of the time, the other 20%. The subject's task was to count the number of names presented in one of the categories. Subjects counted a different category in each oddball task; in one they counted the rare names, and in the other the frequent names. The order in which they counted the stimuli (count rare first or second) and the rare stimulus were counterbalanced across subjects.

For all the oddball tasks, the interval between two consecutive stimuli was fixed at 2000 ms. All aspects of the experiment were controlled by a PDP 11/40 minicomputer.

Recording Apparatus

EEG was recorded at Fz, Cz, and Pz, referred to linked mastoids, and vertical EOG was recorded from above and below the right eye. Beckmann Biopotential Ag/AgCl electrodes were used for all these placements and for the ground electrode (placed on the subject's forehead). Impedance was always below 10 KOhms. Both EEG and EOG were amplified and filtered with model 7P122 Grass amplifiers. The data were filtered on-line with a 35 Hz half cut-off low pass filter and an 8 second time constant, and were digitized on line at 100 Hz. The recording epoch began 100 ms before stimulus onset, and lasted 1500 ms. The data were monitored on-line by the experimenter, and recorded on tape for further processing. The baseline level was estimated by computing, for each single trial and channel, the average activity of the 100 ms period preceding the stimulus. The baseline was subtracted from each record before further processing. This has the effect of eliminating differences in the baseline level among conditions and electrodes. Trials during which saturation of the analog/digital converter occurred were detected with an off-line algorithm and discarded. Recordings from one session were available for 49 subjects, and recording from two sessions were available for 17 subjects.

Ocular Artifact

In order to eliminate ocular artifacts, an off-line correction procedure was applied (EMCP, Gratton, Coles, & Donchin, 1983a). This procedure is based on the estimation of the propagation of ocular potentials to the scalp electrodes. The activity recorded at scalp electrodes attributable to ocular artifacts is then subtracted from the records. The propagation of the ocular potentials to the scalp was estimated by means of a regression procedure. The propagation factors were computed on the activity not time-locked with the stimulus in both EOG and EEG records. However, the correction was applied considering both the event-related and the "not-event-related" activity in both channels. The procedure was applied separately for each subject, session, and task, and separate propagation factors were computed for each condition.

The experimental design also allowed for a study of the consistency of the propagation factor computed by EMCP. The data indicated a high consistency across subjects and conditions. In most cases, the propagation factors were within a small range. For saccades, 77% of all the propagation factors computed for Fz were between .17 and .29, for Cz, 75% were between

.05 and .14, and for Pz 79% were between .02 and .09. Narrower ranges were observed for the propagation factors for blinks. For Fz, 72% of all the propagation factors were between .15 and .22, for Cz, 83% were between .05 and .11, and for Pz, 76% were between .03 and .06.

While the range of the propagation factors was narrow, the correlations between the propagation factors computed for each subject across tasks in the same session (session 1) ranged between .37 and .77, median r = .60 (n = 50). Thus, it would seem that the propagation factor is quite unstable across tasks. It is even less stable across sessions. The correlations across sessions between the propagation factors for the same task ranged between -.10 and .60, median r = .29 (n = 19). This instability is probably due to the many variables that influence the propagation factor that are likely to vary with time, such as electrode placements, wetting conditions of the eye, and so on. In fact, the propagation of the ocular potential to the scalp can hardly be considered a trait. These data are in accord with the findings reported by Gratton et al. (1983a), and suggest that separate propagation factors should be computed for different sessions. It is also desirable to compute separate propagation factors for each task within a session, provided enough trials are available (100 is a minimum).

Typical ERP waveforms

In figure 2 we present average waveforms from the first session of one representative subject in the five oddball tasks (Pz only). In figure 3 we present the corresponding grandaverage waveforms for all the 49 subjects. The P300 component is clearly visible in all these waveforms. Differences in P300 latency across tasks are also evident in these figures, with the shortest latency for the auditory oddball and the longest one for the name oddball. P300 amplitude is larger for rare and target (counted) stimuli than for frequent and non-target (non counted) stimuli (probability and target effects).

Insert figure 2 About Here

Insert figure 3 About Here

In figure 4 we present the average Pz waveforms for each individual subject in the H-S oddball. Waveforms for the "count rare" condition are presented in figure 4a and for the "count frequent" condition in figure 4b. For the "count rare" condition, all the subjects showed a larger P300 for the rare (and target) stimulus than for the frequent (and non-target) stimulus. For the "count frequent" condition, where probability and target effects on P300 amplitude are dissociated, some subjects show a larger P300 for the rare stimuli, and other subjects a larger P300 for the frequent stimuli.

Insert figure 4 About Here

In figure 5 we present examples of average Pz waveforms from the first and second sessions of three subjects. In the upper panel we present an example of a subject whose waveforms are very similar across sessions. In the middle panel we present a subject whose waveforms in the two sessions are rather dissimilar (particularly for rare stimuli; the number of trials in both sessions is approximately equal). In the lower panel we present a subject whose waveforms from the two sessions have similar waveshapes, but different P300 amplitudes.

Insert figure 5 About Here

Averages and Single Trials

P300 estimates were obtained both on average and single trial waveforms. For the study of the reliability of P300 estimates between sessions, separate averages were obtained for each subject, session, task, stimulus and electrode. The data were then digitally low-pass filtered (-3 dB at 6.29 Hz, zero at 14.29). For the study of the reliability of P300 estimates within a single session, a "split-half" approach was used. Each trial (for each subject, task, and stimulus) was randomly assigned to one of two groups (.5 probability for each group). For the study of the reliability of the estimates taken over averages, average ERPs for subject, task, stimulus, group, and electrode, were then computed. The average waveforms were then digitally low-pass filtered (-3 dB at 6.29 Hz, zero at 14.29 Hz). For the study of the reliability of P300 estimates obtained on single trials (from a single session), the single trial waveforms were digitally low-pass filtered (-3 dB at 3.14 Hz, zero at 7.15 Hz). The estimates obtained at each single trial were then averaged according to the same trial assignment used above. Thus averages of single trial estimates were obtained for each subject, task, stimulus, and group. In addition, a quarter of the frequent trials were also assigned at random to one of two sub-groups, so that the frequency of trials belonging to one of these subgroups was comparable to the frequency of the trials belonging to one of the two rare groups. Average waveforms and averages of single trial estimates were also obtained for these two subgroups, following the same procedure explained above.

Measures

The measures we took differed in their use of spatial and time series information.

Procedures Using Spatial Information

We adopted two approaches to the use of spatial information.

Channel selection. In this approach one recording channel is selected for the estimation of P300 parameters (Pz in this study). This corresponds to a linear filter which arbitrarily gives a weight of 1 to one electrode (Pz), and a weight of 0 to the other electrodes (Fz and Cz), whose information is therefore ignored.

Vector filter. This approach consists of weighting the information obtained at several electrodes so as to optimize the discrimination between the target component (in our case P300) and the other sources of electrical brain activity. The vector filter we adopted corresponded to the following set of weights: Fz = .1611, Cz = -.5335, Pz = .8210. In a simulation study on techiques for estimating P300 latency (Gratton, Kramer, Coles & Donchin, 1985), this set of weights was found to produce an average reduction of 20% in error for P300 latency estimation in comparison with the use of the parietal location alone. Note, however, that this set of weights does not correspond to the distribution of potentials usually recorded at these scalp electrodes for the P300. A set of weights directly reflecting P300 scalp distribution would have had a positive value for the central electrode. Gratton, Kramer, Coles, and Donchin (1985) showed that a set of weights reproducing the P300 scalp distribution would not discriminate as well between P300 and noise as the set of weights we adopted in this study. They attributed this observation to the high correlation, in the background EEG, between the parietal and the central electrode. Thus a better discrimination between P300 and noise is obtained by giving opposite weights to these two locations.

Procedures for the Analysis of Time Series

We adopted several approaches to the analysis of the time series (i.e., the waveforms). In particular we used an area measure, peak-picking, PCA, and cross-correlation. These procedures can be conceptualized as different instances of a cross-product function. All of these procedures allow the amplitude of P300 to vary (and are therefore suitable for a study of P300 amplitude), while only some of them allow the latency of P300 to vary (and are thus suitable for studying P300 latency). The algorithms used to derive estimates of P300 parameters with each procedure are given below. The same time windows were used for area, peak and cross-correlation measures. The time window for the auditory and H-S oddball tasks started 250 ms and terminated 700 ms after stimulus onset. For the name oddball task, the time window went from 400 ms to 900 ms post stimulus onset. Area, peak and cross-correlation measures were taken at Pz and on waveforms obtained by combining the electrodes with Vector filter.

Area measures. Area measure estimates were obtained by integrating the activity over the entire time window.[7] Only estimates of P300 amplitude were obtained with area measures.

Peak measures. The P300 peak was identified by an algorithm that searched for the most positive point in the time window. Amplitude estimates were obtained by measuring the difference between the value recorded at the P300 peak and the prestimulus baseline level (base-to-peak amplitude). Peak latency measures were also obtained.

Cross-correlation measures. According to this procedure, P300 was identified by determining which segment of the waveform was maximally correlated with a template. The template was a full cycle of a 2 Hz inverted cosinusoidal waveform. The central point of the segment of ERP waveform considered had to belong to the time window. Latency estimates were obtained by considering the central point of the segment with maximal cross-correlation (CC) with the template. Amplitude estimates were obtained in two ways. One procedure (labelled CC-Amplitude) was based on the computation of the difference between the value on the ERP waveforms at the central point of the segment with maximal cross-correlation and a prestimulus baseline level. The other procedure was based on the computation of the covariance of the whole ERP segment with the template (labelled CC-Covariance). This procedure is independent of the baseline setting, and considers values recorded over a wide range of points, rather than on a single point. On the other hand, it is sensitive to the presence of other negative, as well as positive, peaks in the ERP segment.

PCA. The use of PCA to identify and measure ERP components has been described by Donchin (1966), and Donchin and Heffley (1979), and reviewed elsewhere in this chapter. The use of PCA to obtain estimates of the amplitude of ERP components is based on the computation of the cross-products of the ERP waveforms (which are subtracted from the grand average) with the component loadings for each component. These cross-products are labelled component scores. Thus, reliable component loadings are required to obtain reliable component scores. The reliability of the component loadings is low when the ERP waveforms entered in the PCA contain high levels of auto-correlated noise (Hunt, 1980; Molenaar, 1983). This is particularly the case when the ERP waveforms entered are averages based on a small sample of trials. Therefore, we only applied PCA to the between session analysis. For this analysis, we applied PCA in two different ways: by computing separate PCAs for each session, or by computing PCAs using averages from both sessions combined. In each case, separate PCAs were run for the Auditory, H/S and Name oddball tasks.

The PCAs run separately for each session were based on 102 waveforms for the Auditory oddball task (17 subjects x 2 stimuli x 3 electrodes) and 204 waveforms for the H/S and Name oddball tasks (17 subjects x 2 tasks x 2 stimuli x 3 electrodes). The component loadings for the PCAs run on each session and on each of the three oddballs are shown in Figure 6. The

---

Insert figure 6 About Here

---

component structures of the first and second session do not completely overlap. In particular, substantial differences are evident between the component structures of the two sessions for the Name oddball task. (We acknowledge, however, that this is to some extent a subjective judgment, as there is no easy way to determine what is a "substantial" difference.) The unreliability of the component structures might be due to the relatively low number of waveforms entered in the analysis (see Picton & Stuss, 1980, for a discussion of this subject). Given the low stability of the component structures across sessions, we considered a comparison between the component scores (which are derived from the component loadings) in the two sessions to be inappropriate, especially because of the difficulty in identifying the "correct" P300 component.

The PCAs run on waveforms from both sessions were based on 204 waveforms for the Auditory oddball task (17 subjects x 2 sessions x 2 stimuli x 3 electrodes), and 408 waveforms for the H/S and Name oddball tasks (17 subjects x 2 sessions x 2 tasks x 2 stimuli x 3 electrodes). The component loadings for each oddball task are shown in Figure 7.

For the auditory and name oddball tasks only one component met the "requirements" of P300 (appropriate latency, parietal scalp distribution, response to probability and the target effect), while in the H/S oddball task two components met some of these requirements, making the choice of which to call P300 difficult. For the Auditory oddball, component 1 peaked

----

Insert figure 7 About Here

----

380 ms after the stimulus, and was interpreted as P300. It was maximally positive at the parietal electrode ($F(2, 32) = 68.08$, $p < .001$), and larger for rare than for frequent stimuli ($F(1, 16) = 32.32$, $p < .001$). For the H/S oddball, component 2 peaked 280 ms after the stimulus and component 3 peaked 440 ms after the stimulus. Both these components were most positive

at the parietal electrode ($F(2, 32) = 35.44$, $p < .001$ for component 2, and $F(2, 32) = 92.31$, $p < .001$ for component 3). Component 2 was larger for the target stimuli, either rare or frequent, than for the non-target stimuli (task x stimulus interaction, $F(1, 16) = 27.63$, $p < .001$), but the probability effect was present only in interaction with the electrode location ($F(2, 32) = 4.52$, $p < .02$), with a larger separation among electrodes for rare than frequent stimuli. Component 3 was larger for rare than frequent stimuli ($F(1, 16) = 44.07$, $p < .001$), but was not significantly larger for the target letter (task x stimulus interaction, $F(1, 16) = 3.96$, $p >.05$). We decided to interpret component 2 as "P300", since it exhibited both probability and target effect, although we believe that both PCA components may represent different aspects of the P300 component. For the Name oddball, component 3 peaked 540 ms after stimulus onset, and was interpreted as P300. It was most positive at the parietal electrode ($F(2, 32) = 11.05$, $p < .001$), larger for rare than frequent stimuli ($F(1, 16) = 12.97$, $p < .01$), and larger for target stimuli (task x stimulus interaction, $F(1, 16) = 32.16$, $p < .001$).

Estimates of P300 latency are not available with PCA. In fact, PCA requires the assumption that the latencies of ERP components are constant for all waveforms (from different subjects and conditions) entered in the analysis (see Donchin and Heffley, 1979).

Likelihood estimates. We also used estimates of P300 parameters intended to simulate estimates that could have been obtained with an "eye-ball" inspection of the waveforms. We included this procedure to test whether a simple eyeball inspection was as reliable as more complex procedures. This methodology assumes that experienced researchers can identify P300 by visual inspection of the waveforms by applying some criteria (see Kramer, 1985). We assumed that the visual inspection process involves the following steps. First, selection of all the positive peaks in the waveform. Second, identification of which positive peak is P300 by means of an amplitude criterion, a latency criterion (P300 should have a latency of approximately "x" milliseconds, and should be within a certain range) and a scalp distribution criterion (P300 should have a parietal maximum, or at least a central one). Third, measurement of the parameters (amplitude and latency) of the point selected as P300. Thus, we created a computer algorithm based on these three steps. First, all positive peaks were identified. Second, the amplitude of each peak was weighted for its latency and scalp distribution. The latency weights were different for each oddball task and are shown in Figure 8. The scalp distribution weights

----

Insert figure 8 About Here

----

were given by the "closeness" of the scalp distribution observed at each peak point to an "ideal P300" scalp distribution having a maximum at the parietal electrode, a medium value at the central electrode, and the least positive value at the frontal electrode. The "closeness" of the observed scalp distribution to the theoretical one was measured with a Pearson product-moment correlation coefficient. The base-to-peak amplitude of each peak was then "weighted" by multiplying it by the latency and scalp distribution weights. The weighted amplitude of all the peak points were then compared, and the point with maximum value was selected as the "P300" peak point. The final step was the computation of P300 parameters.

Amplitude was estimated by measuring the difference between the amplitude of the point selected as P300 peak and a prestimulus baseline level. Latency was the latency of the P300 peak point. The "likelihood" procedure was used only for the between session analysis.

Table 5 presents a summary of the P300 measures we obtained with each

----

Insert table 5 About Here

----

procedure for the between session and the within session analysis.

Results

Within Session Reliability

The analysis of the within session reliability of P300 estimates was based on 49 subjects. One subject's data were discarded because only one rare trial was free of recording artifacts (saturation of the A/D converter) and therefore a split-half analysis was not possible. As mentioned above, the analysis was based on the random assignment of each trial (for each subject, task, and stimulus) to one of two groups. Note that the random assignment of each single trial to the two split-half groups once done was maintained in all the analyses.

Recently, Callaway, Halliday, Naylor, and Thouvenin (1984) noted that estimates of the latency of P300 were very different when they were taken on average waveforms or by averaging single trial parameters. Thus, to investigate the reliability of parameters obtained on average ERP waveforms, or by averaging single trials parameters, two types of procedures were applied. The first procedure was based on three steps: (a) computation of averages for each of the "split-half" groups, (b) estimation of P300 parameters on these averages, and (c) computation of reliability estimates.

For the second procedure, the order of the first two steps was reversed. The second procedure was the following: (a) estimation of P300 parameters on each single trial, (b) computation of two averages of P300 parameters, and (c) computation of reliability estimates. We label the P300 estimates obtained after the first two steps "Average parameters" for the first procedure, and "Average of single trial parameters" for the second procedure. Thus, we will present the reliability of average parameters and the reliability of the average of single trial parameters. Note, however, that the two sets of estimates are based on exactly the same trials, and a comparison between the two kinds of reliabilities is therefore legitimate.

Reliability of average parameters. Reliabilities of average parameters were computed for each task and stimulus. The computation was based on a Pearson product-moment correlation. The estimates were then corrected for the reduced number of trials (one half) with the "Spearman-Brown Prophetic Formula". Separate reliability indices were computed for each of the ten task x stimulus conditions (i.e., auditory rare target, auditory frequent target, H/S rare target, H/S rare non-target, H/S frequent target, H/S frequent non-target, Name rare target, Name rare non-target, Name frequent target, and Name frequent non-target). The overall range and median of the reliability indices are shown in Table 6. Also shown are the medians of the reliabilities for each oddball task (Auditory, H/S, and Name) and stimulus class (rare or frequent, and target or non-target).

Insert table 6 About Here

Table 6 shows that amplitude measures have higher reliability than latency measures. Given that the amplitude estimates (apart from the area measures) are subordinate to the latency estimation (i.e., the latency of P300 is first determined, and then amplitude is measured), this result may appear surprising. A possible explanation may be the following. If an error in P300 peak detection occurs, the latency estimates may be any value within the time window considered, and the distribution of the error of estimation may not follow a normal distribution. That is, small errors are not more likely than large errors. In fact, the distribution may very well follow a rectangular distribution (i.e., any value has the same probability of being chosen). However, the amplitude estimates may not follow the same rule. In fact, the amplitude estimate may still be approximately correct even after an incorrect point has been chosen as the P300 peak point. In general, for amplitude estimates, small errors are more likely than large errors. This would also explain the smaller range of reliabilities for amplitude estimates.

The most reliable amplitude estimates are obtained with base-to-peak measures on Vector filtered waveforms. This procedure yields an overall median reliability of .88, which is rather high, if we consider that several of the samples used contained less than 30 trials (in particular, all the rare samples had less than 30 trials each). However, reliability appeared to be sensitive to the task and stimulus characteristics. In particular, the Auditory oddball task yielded higher reliabilities (for all measures a value close to .90) than the H/S and Name oddball tasks. For the latter task, the highest reliability was .71 (with Base-to-peak and CC-covariance on Vector filtered waveforms). Another observation is that the reliability was higher for frequent than for rare stimuli and that the reliability for target stimuli was higher than the reliability for non-target stimuli. The first effect can be attributed to the number of trials used to compute the averages. When the reliability for frequent stimuli was computed on averages based on a number of trials comparable to those used for the averages for the rare stimuli, this effect was reversed. In this case, all measures had higher reliability for rare stimuli than for frequent stimuli. The target effect can be explained in terms of signal-to-noise ratio. Target stimuli elicited a larger P300 than non-target stimuli (the effects of probability and of number of trials were compensated).

The highest overall reliability for latency measures was obtained with the Peak-picking procedure applied to Vector filtered waveforms. However, the reliability obtained with this procedure (.70) is not particularly high. In particular, the reliabilities of all procedures appear rather low in the Name oddball task. It is relevant to note here that this task, involving a rather complex discrimination, may produce jitter in P300 latency from trial to trial. If this is the case, P300 amplitude may be reduced in the averages, producing a reduction of the signal-to-noise ratio, and the estimation of P300 latency using averages may be inappropriate.

Reliability of the average of single trial parameters. The reliabilities of the average of single trial parameters were obtained in the same way as those of the parameters directly estimated from the average waveforms. Ranges, overall medians, and medians for each task and stimulus class (rare vs. frequent, target vs. non-target) are shown in table 7.

Insert table 7 About Here

It may be seen in Table 7 that the highest reliability for the amplitude estimates were obtained by using the CC-covariance procedure. Slightly better reliabilities were obtained when these measures were taken after application of Vector filter. This is particularly evident for the Name oddball task. The reliabilities obtained by measuring P300 amplitude with CC-covariance procedure on single trials and then averaging were very high (overall median of .92). Furthermore, particularly when applied on single trial waveforms passed through Vector filter, they were very consistent, with a minimum value of .77. The reason why CC-covariance does so well when applied to single trials is not known. However, we recall that CC-covariance measures are (a) independent of the baseline estimation (which in single trials may not be satisfactory), and (b) based on a large number of points rather than only one. When compared with measures taken on averages, the reliabilities obtained with CC-covariance on single trials were clearly higher.

The gain in reliability by using single trial estimates is particularly evident for the latency measures. In fact, all procedures have an overall median reliability larger than .80. The highest reliabilities for latency estimates were obtained with peak-picking applied to single trials passed through Vector filter. However, the reliabilities obtained with all measures are within a small range. The largest separation is present for the Name oddball task. This is interesting, because larger variations between trials in P300 latency might be expected in this oddball task than in the Auditory or H/S oddball tasks.

As for measures taken on average waveforms, large differences are visible for the reliabilities for stimulus classes. Again, higher reliabilities are obtained for frequent stimuli, and for target stimuli. The latter difference may be attributed to P300 amplitude (larger for target), whereas the higher reliability for frequent stimuli are probably attributable to the number of trials used for the computation of mean estimates. As in the analyses on average waveforms above, the effect reverses for all measures when comparable sample sizes are used for rare and frequent stimuli. The effect of the number of trials on the reliability of the estimates can be clearly seen in Figure 9, for amplitude estimates, and in Figure 10, for latency estimates.

Insert figure 9 About Here

These figures illustrate the reliabilities of the mean estimates when the first 4, 8, 12, 16 and 20 trials from each trial group are used for the

----

Insert figure 10 About Here

----

computation of the mean estimates. Only frequent trials were used for this analysis. Six subjects were not used because they had less than 20 trials in some frequent group (total sample = 43). Different mean values were computed for each task, and median reliabilities were then computed on comparable data. The reliabilities were not corrected with the Spearman-Brown Prophetic Formula. The values on the extreme right indicate the reliabilities obtained when the full sample of frequent trials was used for the analysis (the number of trials in the complete sample varied from 20 to over 50).

Reliability of the probability effect. Donchin et al. (1978) proposed that one of the defining characteristics of P300 is the particular way in which it responds to experimental manipulations. One of the most commonly studied effects on P300 amplitude is the "probability effect", defined above. Given its potential use in the definition of P300, we were interested in determining the reliability of this effect.

Estimates of the probability effect for each subject and task can be obtained by subtracting the estimates for frequent trials from the estimates for rare trials, and this can be done for both single trial and average estimates. Alternatively, the estimates can be obtained from difference waveforms produced by subtracting the average waveforms for frequent trials from the average from rare trials. Thus, we compared three different procedures for the estimation of the probability effect: (a) the difference between average scores computed on single trials for rare and frequent stimuli, (b) the difference between estimates obtained from the average waveforms for rare and frequent stimuli, and (c) estimates obtained on difference waveforms generated by subtracting average waveforms (rare - frequent). We will call the first method difference scores on single trial estimates; the second, difference scores on average estimates; and the third, estimates on difference waveforms. The reliability estimates were obtained with the same split-half procedure described above. Pearson's product-moment correlation coefficients were later corrected with the Spearman-Brown Prophetic formula to yield the reliability coefficients.

Table 8 shows the median reliability estimates obtained with each measure for any procedure.

Insert table 8 About Here

The reliabilities of estimates obtained on difference waveforms are higher than the reliabilities of the difference between estimates, both computed on averages and on single trials. This result indicates that the use of difference waveforms may be methodologically correct, at least in the case of comparable waveforms. "Comparable" in this sense means cases in which only one manipulation was used (for example, probability), and this manipulation does not affect the latency of ERP components.

The highest reliabilities were obtained with CC-covariance estimates. A slight advantage was given by passing the waveforms through Vector filter. Note, however, the highest reliability is only .64. Thus, the probability effect is not, in general, very reliable. Therefore, using a probability manipulation alone to define P300 may result in a rather unreliable definition.

Between Session Reliability

The between session reliability maps the consistency over time (the time lapsing between the two sessions) of the P300 parameters, as measured by several procedures. In general, the reliability between sessions cannot exceed the reliability within session, obtained with the same procedure and under the same conditions. However, it may be much lower. Such a decline in reliability should not necessarily be considered as a sign of inconsistency of the procedure used to measure P300. It may very well be the case that the time lapse affects subjects to varying degrees. This "subject x session" interaction may be an "interesting" effect (e.g., it may provide information about learning, adaptation, etc.), and yet still contribute to decrease the between session reliability.

On the other hand, good consistency of P300 parameters over time is important for their use in studies of individual differences. It is also very important in cases where subjects are run in multiple sessions. This is particularly the case in very easy tasks where learning should not be an important factor. Given that these data were collected in the context of an individual difference study, we were particularly interested in assessing the consistency of P300 estimates over time.

The assessment of between session reliability was based on a sample of 17 subjects. The records of two subjects for the second session were lost, and another subject was rejected because too few trials were free of recording artifacts. The two sessions were recorded 7 to 15 days apart and were approximately equal, but the number of trials for the H/S and Name oddball tasks was slightly reduced for the second session. This allowed us to be sure the subject really accomplished the count task, and did not merely remember the correct number of trials from the first session.

Average waveforms were obtained for each subject, session, task, stimulus, and electrode. Amplitude and latency parameters were then computed, according to the procedures described above, on the average waveforms. The between session reliability of mean single trial estimates was not tested. The reliability of the estimates was then computed separately for each task and stimulus, with a Pearson product-moment correlation. The Spearman-Brown correction was not applied. Thus, these reliability indices are comparable to those obtained above in the within session analysis.

Table 9 shows the range and overall median of the reliability indices obtained with each procedure. Median reliabilities for each task (Auditory, H/S, and Name), and stimulus (rare or frequent, and target or non-target) are also shown.

Insert table 9 About Here

There are several interesting differences with respect to the within-session reliability. In general, the between session reliabilities are lower than the within session reliabilities. This is particularly true for the CC-covariance measures and for the latency estimates. The highest between session reliabilities are obtained by using PCA to obtain an appropriate set of weights. However, PCA needs a large body of data to obtain stable estimates of the component loadings, used as weights. This was evident in our analysis. In fact, the loadings obtained by separate PCAs for the first and second session did not overlap. Furthermore, even the reliabilities obtained with PCA are lower than those obtained (with other procedures) in the within-session analysis. This indicates that part of the individual differences in P300 observed in a single recording session may not be observed again one or two weeks later.

The second highest reliability for P300 amplitude estimates is obtained with base-to-peak estimates on Vector filtered waveforms. Base-to-peak estimates obtained at the P300 peak selected by cross-correlation procedure yield similar reliabilities values. In both cases, the reliability is close to .80.

As observed for the within-session analysis, the reliabilities of amplitude estimates are larger than the reliabilities of latency estimates. Again, we believe that this phenomenon might be explained in terms of error distribution. The error of estimation would tend to follow a normal distribution for the amplitude estimates, and a rectangular distribution for the latency estimates. This would result in a higher probability of "outliers" (values very far from the sample mean) for latency than for amplitude estimates. To investigate this hypothesis, we studied the scatter plots of the amplitude and latency estimates. The plots were obtained by plotting the estimates for the first session against the estimates for the second session. To obtain comparable plots, the estimates from each session, procedure, task, and stimulus, were transformed into "t" scores. Some examples of the scatter plots obtained in this way are shown in Figure 11.

Insert figure 11 About Here

As exemplified in some of these plots, very high or very low reliability indices may be due to the effect of a small sub-sample of subjects with very deviant values (outliers). To detect outliers, we applied an algorithm to the data entered in each scatter plot. We defined outliers as P300 estimates corresponding to a "t" score larger (in absolute value) than 2.57 (corresponding to a probability of .02, two-tailed) in either of the two sessions. Outliers were four times more frequent for latency than for amplitude estimates. To assess the effect of these outliers on the reliability indices, we computed the reliabilities excluding the outliers. While the reliabilities for amplitude estimates did not change significantly, those for the latency estimates did improve markedly, as shown in Table 10.

Insert table 10 About Here

The highest between-session reliability for latency estimates are obtained with cross-correlation on Vector Filtered waveforms. However, latency reliabilities are sensitive to the task, being lower for the Name oddball. We attribute this lower reliability to P300 latency jitter between single trials, which we believe is particularly present for the Name task.

Discussion

A summary of the results is shown in Table 11. As can be seen in Table

---

Insert table 11 About Here

11, the reliability of P300 parameters (amplitude and latency) usually exceed .60 for most techniques. However, different procedures yield different reliability indices, that range from .92 to .34. The procedure that was intended to simulate visual assessment of ERP data (likelihood) has some of the lowest reliabilities in Table 11. This suggests that much can be gained in the accuracy of P300 detection and measurement by the use of appropriate procedures.

By averaging single trial estimates we may obtain more reliable values than by deriving the estimates from average waveforms. This is particularly evident for latency estimates, but also for amplitude, when the CC-covariance procedure is applied. The advantage of measuring latency and amplitude from single trials may appear to be a violation of the assumption of the averaging technique. In fact, by measuring P300 parameters on each single trial it is not necessary to assume that P300 (and, in general, ERP) latency is constant over trials as when averaging procedures are used.

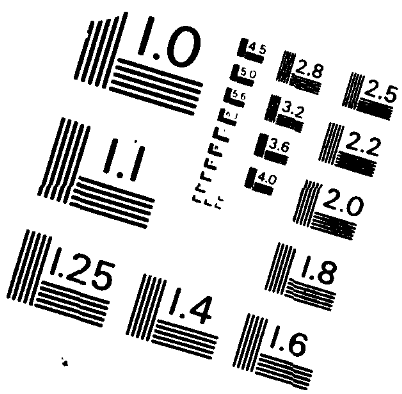
A possible problem in estimating P300 amplitude from single trials may be due to the confounding effects of background EEG noise, whose phase, but not amplitude, should be random. Subjects may differ in the amplitude of this background EEG activity, and thus high reliability of these estimates may not correspond to high validity. However, this does not explain the advantage of single trial measures for latency estimates. Furthermore, CC-covariance, being a sort of "peak-to-peak" measure should be maximally sensitive to variation in amplitude of background EEG noise. This procedure is the most reliable for assessing the probability effect. This result may indicate that this procedure is valid. Thus, the influence of background EEG noise on our amplitude estimates is probably small. A relevant point is that single trial waveforms were heavily filtered (half amplitude at 3.14 Hz).

A possible explanation of the advantage of measuring latency from single trials may also be found in the distribution of the error of estimation for average and single trial estimates. For each waveform considered, if an error is committed, any point in the time window may be selected as P300 peak point. Thus, for averages such errors may produce very deviant values. On the other hand, for each single trial these errors are likely to produce different values, and therefore, in the long run, they compensate. This speculation is supported by an analysis of the outliers (as defined above). The means of single trial estimates were nine times less likely to produce an outlier value than estimates on average waveforms. This result is also in accord with the well known limit theorem: sample means tend to be distributed normally even when the distribution of the single values deviate from normality. Thus, parametric statistics will be in general more appropriate for means based on single trial estimates than for estimates based on averages.

The use of spatial information (by means of the Vector filter procedure) produced, in general, an improvement in the reliability of both amplitude and latency estimates. However, the advantage was most evident in the between session analysis, and for the Name task. These two cases correspond to the lowest reliability indices. Thus, the use of spatial information may be particularly advantageous in cases of low reliability, but it may not be advantageous in those cases in which the reliability of the measures is already high when measures are taken at Pz.

We should note here that the weights for each electrode used for the Vector filter were chosen on the basis of theoretical speculations. They were chosen to produce a good discrimination between the signal (P300) and the noise (other sources of electrical activity, like background EEG noise, N200, Slow Wave, etc.). However, the choice was not based on a particular

AD-A170 794 THE EVENT-RELATED BRAIN POTENTIAL AS AN INDEX OF 4/8
INFORMATION PROCESSING A..(U) ILLINOIS UNIV CHAMPAIGN
COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL.
UNCLASSIFIED 20 FEB 86 CPL-86-1 F49620-85-C-0041 F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

mathematical algorithm, but rather on the experience of the investigators. The weights used do produce some improvement in comparison with traditional techniques. However, as shown in a simulation study by Gratton, Kramer, Coles, and Donchin (1985), variations in the weights assigned to each electrode result in variations in the accuracy of P300 detection. It may very well be the case that weights more appropriate than those used in the present study might have been chosen, and that higher reliabilities would have been obtained. Furthermore, different weights may be more appropriate for different tasks or for different subjects.

An approach to the appropriate choice of weights is the computation of a set of weights which optimally discriminates between P300 and various sources of noise. This can be accomplished by "estimating" the covariances between electrodes attributable to P300 and to noise. Alternatively, the weights may be estimated with a Multivariate Discriminant Analysis (MDA) between the scalp distributions of "rare" and "frequent" trials on an independent sample of trials. In fact, the weights we used approximated those obtained with a discriminant analysis of the P300 scalp distributions of two groups of subjects in a separate study (none of these subject participated to the reliability study).

A similar caveat should be considered in selecting a template for use in the cross-correlation estimates. Template choice (a full cycle of an inverted cosinusoidal waveform) was arbitrary, and based on the experience of the investigators, rather than on a specific algorithm. As with the weights for the Vector filter, the template used in cross-correlation was not necessarily the most appropriate. We note here that the best theoretical template for cross-correlation does not necessarily correspond to the P300 waveshape. Rather, it should be that template that yields the best discrimination between P300 and noise.

This goal (discrimination between signal and noise) should be the real target of any signal detection technique, and of any procedure devoted to identifying P300. For this reason, the set of weights, or linear filters, used for the detection of P300 should not only reflect our knowledge of P300, but also our knowledge (at present rather approximate) of the noise in which P300 is embedded.

GUIDELINES

In this paper we have focused primarily on methodology, but we do not wish to imply that there is any substitute for theoretical rigor. Methodological sophistication is often necessary to carry out good research, but it is never sufficient. Creative and well defined hypotheses within a solidly based theoretical framework are also required. We shall not attempt to specify how one should construct theories in the realm of ERP research. We will, however, present two sets of methodological guidelines. The first set includes fairly general methodological criteria that are essential for progress in understanding the endogenous ERP components. The second set of guidelines includes more detailed suggestions concerning ERP analysis. They arise, primarily, from the reliability study presented here, and related simulation studies (Gratton, Kramer, Coles, & Donchin, 1985). Additional elaboration and verification will be required before a more comprehensive set of generally accepted guidelines can be developed; we present these as a starting point for further research and discussion.

The recommendations in our first set of guidelines have been selected from the "Publication Criteria for Studies of Evoked potentials (EP) in Man" (Donchin et al., 1977), with some additional comments and suggestions. We will quote freely; "we" in the following paragraphs refers to the opinions of the present authors, not the committee. Since these criteria were published in 1977 there have been numerous examples of published papers that have not followed some of the major points. It is for this reason that we feel the need to repeat them here. We will summarize the points especially relevant to studies on P300 and other endogenous components.

General Guidelines

1. Electrodes: "EP investigators...should use as many electrodes as they can in any one experiment, because different EP components may have different scalp distributions" (p. 4). We recommend EOG, Fz, Cz, and Pz as a minimum for P300. This is essential for identifying components.

2. Ocular artifact: "The possible contamination from eye movements should be a major concern to all EP investigators and the measures taken to deal with this problem should be considered in any published report" (p.5). In the past, trials contaminated by EOG activity were usually discarded. There are now several techniques that can be used to correct the EEG channels when eye movements occur (Gratton et al., 1983a; Quilter, MacGillivray, & Wadbrook, 1977; Verleger, Gasser, & Mocks, 1982; Whitton, Lue, & Moldofsky, 1978).

3. Recording epoch: since P300 may be followed by a Slow Wave that lasts for hundreds of milliseconds, we recommend a recording epoch of at least one second from stimulus onset. Appropriate amplification and system bandwidth are of course required (pp. 5-6).

4. Raw records: "All the members of the committee strongly agree that it should be an absolute acceptance criterion for all EP papers submitted for publication that they include actual records of average EPs" (p.8). These should include not only superaverages, but also "typical" waveforms of single subjects. Since most ERP experiments use fewer than 20 subjects, it is often possible (and desirable) to present an average waveform for each individual subject. It is desirable to present, at least for some averages, waveforms from each electrode. The presentation of waveforms are of paramount importance. "Only by inspecting such records can one get an idea of the care which the investigator may have taken in eliminating artifacts or of the degree to which the EP recorded in a given laboratory are similar or different to those recorded in other laboratories" (p. 9).

5. Polarity: since a convention on display of polarity has not been developed, it should "... be required that in all published figures the polarity convention ...be indicated by the '+' and '-' sign by the calibration signal. Reporting the polarity convention in the text or legends only would be considered inadequate" (p. 9). This may seem a trivial point, but many figures do not have polarity indicated on them.

6. Amplitude measurement: since there was no general agreement on how to measure the peak amplitudes of ERP components (this is still true today), the methods used should be specified in detail. It is also often desirable to use more than one method, a procedure many investigators now follow.

7. Experimental reports: as in any experimental research, independent variables should be carefully identified and described, and detailed information provided on the subjects, stimuli, and procedure (pp. 3 & 4). "Atypical" subjects should not be eliminated without justification. We feel that unusual morphology is not a good reason for rejection, unless it results from artifacts.

8. Debriefing: it is very important to debrief subjects after any experiment. This provides first, a check that the equipment was functioning properly; second, a check that instructions were fully understood; and third, information on the subject's strategies, motivation, and general attitudes toward the experiment. As paradigms become more complex, and demand more elaborate cognitive processing, the importance of the third point increases because individual differences may become very large. For example, in our recent memory experiments (Fabiani, Karis, & Donchin, in press; Karis, Fabiani, & Donchin, 1984) mnemonic strategies played a major role, and without debriefing the subjects, valuable information would have been lost.

In this chapter we have emphasized the large variability across subjects at several points. All aspects of P300 may vary dramatically, and there are some rare subjects who produce practically no P300 at all. It is thus always possible that "unusual" P300s may result not from any of the experimental manipulations, but from the particular selection of subjects. In our experience this problem is particularly serious with respect to scalp distribution. To solve it we suggest including a simple oddball task in every P300 experiment. A simple two-element auditory or visual oddball can be presented, with probabilities of .80 and .20 (or other fairly extreme frequent-rare probabilities), and the subject should be instructed to count the rare stimuli and report a running total at the end. This can be performed quickly, and helps in interpreting the waveforms obtained in more complex paradigms (see, for example, Magliero, Bashore, Coles, & Donchin, 1984). If a subject does not show a large probability/target effect on P300 amplitude in an oddball task (with a parietal maximum), then one should be cautious about using the data from that subject to make gen alizations about P300 (or other "new" components) obtained in more complex manipulations performed subsequently.

Specific Guidelines

The results of the reliability study lead us to a more specific set of guidelines and suggestions.

1. Single trials vs averages. ERP researchers have traditionally been diffident of single trial estimates. In fact, several investigators do not even maintain a record of single trials, and analyze only averages. We do not recommend this procedure. Other investigators (see Callaway et al., 1984) have shown that the analysis of single trials are more robust, and reveal aspects of the data not visible in the average waveforms. Our belief is that single trial analysis can be very useful for at least three reasons. First, single trials give more reliable estimates of P300 than averages, particularly for latency estimates. Second, with single trials it is not necessary to meet the assumptions of signal invariance; and third, it is useful to test whether the assumptions of signal averaging have been met.

2. Filters. The use of appropriate filters can be particularly useful to increase the signal-to-noise ratio, and thus improve the detection of P300. Gratton, Kramer, Coles, and Donchin (1985) found that appropriate frequency filters may be particularly helpful. For P300 analysis, appropriate frequency filters may have an upper cutoff point as low as 3 Hz for single trials and 6 Hz for averages. The use of filters for a particular scalp distribution (Vector Filter) may also be advantageous. Filters are particularly useful for latency analysis. Emphasis should be given to the discrimination between signal and noise. Thus, not only the signal (P300), but also the characteristics of the noise (other sources of electrical activity) must be considered.

3. Signal detection algorithms. It appears obvious that the best results should be obtained with algorithms that use the most information in defining P300, and this is indeed the case. For example, the two best procedures (in terms of reliability) are PCA and Vector Filter with Cross-correlation, and it is these procedures that use information from many time points and several electrodes. Optimizing techniques, such as PCA and SWDA, are also useful. However, caution should be used with these techniques. In fact, they usually require a validation procedure. Of particular importance is the interpretation of results obtained with PCA. They should never be accepted uncritically. Most confusion about PCA results from inappropriate interpretation of the findings and violations of the assumptions. Investigators must determine whether the assumptions of PCA are violated before performing a PCA. The choice of the rotational procedure, the number of component rotated, and the rationale for their interpretation must be described and justified.

4. Analysis of scalp distribution. Scalp distribution information may be very useful, given its relationship with the ERP generators. However, as already observed by McCarthy and Wood (in press), the measurement procedure should be chosen carefully. The use of an univariate ANOVA model is inappropriate. To interpret different electrode locations as experimental manipulations is unwarranted, and may lead to errors of interpretation. We believe that a multivariate model is more appropriate for the analysis of scalp distribution. Vector analysis (Gratton, Coles, & Donchin, 1985) adopts such a model. Furthermore, the use of a particular application of this approach, the Vector filter (Gratton et al., 1983b), may improve the detection and measurement of P300.

5. Multiple analysis. The use of several measurement procedures should be encouraged. In fact, the convergence of findings obtained with different procedure may strenghten the conclusions. Furthermore, different procedures may shed light on different aspects of the data.

6. Multiple sessions. When subjects are run in multiple sessions in order to get a sufficient number of trials, the sessions should be comparable. Preferably, each session should contain all the experimental and control conditions. If this is not possible, counterbalanced or random experimental designs should be used. Data from different sessions should be analyzed separately and pooled together only when yielding similar results.

REFERENCES

Allison, T., Hume, A. L., Wood, C. C., & Goff, W. R. (1984). Developmental and aging changes in somatosensory, auditory and visual evoked potentials. Electroencephalography and Clinical Neurophysiology, 58, 14-24.

Becker, D. E., & Shapiro, D. (1980). Directing attention toward stimuli affects the P300 but not the orienting response. Psychophysiology, 17, 385-389.

Brown, W. S., Marsh, J. T., & LaRue, A. (1983). Exponential electrophysiological aging: P3 latency. Electroencephalography and Clinical Neurophysiology, 55, 277-285.

Campbell, K. B., Courchesne, E., Picton, T. W., & Squires, K. C. (1979). Evoked potential correlates of human information processing. Biological Psychology, 8, 45-68.

Callaway, E., Halliday, R., Naylor, H., & Thouvenin, D. (1984). The latency of the average is not the average of the latencies. Psychophysiology, 21, 571 (Abstract).

Coles, M. G. H. (1984). Heart rate and attention: the intake-rejection hypothesis and beyond. In M. G. H. Coles, J. R. Jennings, & J. A. Stern (Eds.) Psychophysiological perspectives: Festschrift for Beatrice and John Lacey (pp. 276-294). N.Y.: Van Nostrand Reinhold.

Coles, M. G. H., Gratton, G., Kramer, A. F., & Miller, G. A. (in press). Principles of signal acquisition and analysis. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), Psychophysiology: Systems, processes, and applications. N.Y.: Guilford Press.

Courchesne, E. (1977). Event-related brain potentials: Comparison between children and adults. Science, 197, 589-592.

Courchesne, E., Courchesne, R. Y., & Hillyard, S. A. (1978). The effect of stimuli deviation on P3 waves to easily recognized stimuli. Neuropsychologia, 16, 189-199.

Courchesne, E., Hillyard, S. A., & Courchesne, R. Y. (1977). P3 waves to the discrimination of targets in homogeneous and heterogeneous stimulus sequences. Psychophysiology, 14, 590-596.

Courchesne, E., Hillyard, S. A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. Electroencephalography and Clinical Neurophysiology, 39, 131-143.

Desmedt, J. E. (1980). P300 in serial tasks: An essential post-decision closure mechanism. In H. H. Kornhuber & L. Deecke (Eds.) Motivation, motor, and sensory processes of the brain. Progress in Brain Research, vol. 54 (pp. 682-686). Amsterdam: Elsevier-North Holland.

DeSwart, J. H., Kok, A., & DasSmall, E. A. (1981). P300 and uncertainty reduction in a concept-identification task. Psychophysiology, 18, 619-629.

Donchin, E. (1966). A multivariate approach to the analysis of average evoked potentials. IEEE Transactions on Bio-Medical Engineering, BME-13, 131-139.

Donchin, E. (1969a). Data analysis techniques in averaged evoked potential research. In E. Donchin & D. B. Lindsley (Eds.), Averaged evoked potentials: Method, result, evaluations. (NASA SP-19, pp. 199-236). Washington, DC: U.S. Government Printing Office.

Donchin, E. (1969b). Discriminant Analysis in average evoked response studies: the study of single trial data. Electroencephalography and Clinical Neurophysiology, 27, 311-314.

Donchin, E. (1979). Use of scalp distribution as a dependent variable in event-related potential studies: Excerpts of preconference correspondence. In D. Otto (Ed.), Multidisciplinary perspectives in event-related brain potential research (pp. 501-510). (EPA-600/9-77-043). Washington, D.C.: U.S. Government Printing Office.

Donchin, E. (1981). Surprise? ... Surprise! Psychophysiology, 18, 493-513.

Donchin, E., & Bashore, T. R. (in press). Clinical versus psychophysiological paradigms in the study of event-related potentials. Brain and Behavioral Sciences.

Donchin, E. Callaway, E., Cooper, R., Desmedt, J. E., Goff, W. R., Hillyard, S. A., & Sutton, S. (1977). Publication criteria for studies of evoked potentials (EP) in man. In J. E. Desmedt (Ed.), Attention, voluntary contraction and event-related cerebral potentials. Progress in clinical Neurophysiology, vol.1 (pp. 1-11). Basel: Karger.

Donchin, E., Coles, M. G. H., & Gratton, G. (1984). Cognitive psychophysiology and preparatory processes: A case study. In S. Kornblum & J. Requin (Eds.), Preparatory state and processes. Hillsdale, NJ: LEA.

Donchin, E., & Heffley, E. (1979). Multivariate analysis of event-related potential data: A tutorial review. In D. Otto (Ed.), Multidisciplinary perspectives in event-related brain potential research (pp. 555-572). (EPA-600/9-77-043). Washington, D.C.: U.S. Government Printing Office.

Donchin, E., & Herning, R. I. (1975). A simulation study of the efficacy of

Step-Wise Discriminant Analysis in the detection and comparison of event-related potentials. Electroencephalography and Clinical Neurophysiology, 38, 51-68.

Donchin, E., Karis, D., Bashore, T. R., Coles, M. G. H., & Gratton, G. (in press). Cognitive psychophysiology and human information processing. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), Psychophysiology: Systems, processes, and applications. New York: Guilford Press.

Donchin, E., Kramer, A. F., & Wickens, C. (1982). Probing the cognitive infrastructure with event-related brain potentials. In M.L. Frazier & R.C. Crombie (Eds.), Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics . AFFTC-TR-82-5. Edwards Air Force Base, California.

Donchin, E., Kramer, A. F., & Wickens, C. (in press). Applications of event-related brain potentials to problems in engineering psychology. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), Psychophysiology: Systems, processes, and applications. New York: Guilford Press.

Donchin, E., Ritter, W., and McCallum, C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting & S.H. Coslow (Eds.), Event-related Brain Potentials in man (pp. 349-411). New York: Academic Press.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: the variation of event-related potentials with subjective probability. Psychophysiology, 14, 456-467.

Duncan-Johnson, C. C., & Kopell, B. S. (1981). The stroop effect: Brain potentials localize the source of interference. Science, 214, 938-940.

Fabiani, M., Karis, D., & Donchin, E. (in press). P300 and memory. In F.

Rösler & S. Sutton (Eds.), On how to interbreed endogenous ERP components with cognitive constructs. Electroencephalography and Clinical Neurophysiology.

Fitzgerald, P. G., & Picton, T. W. (1981). Temporal and sequential probability in evoked potential studies. Canadian Journal of Psychology, 35, 188-200.

Ford, J. M., & Pfefferbaum, A. (1985). Age-related changes in event-related potentials. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), Advances in Psychophysiology, Volume 1. Greenwich, CT: JAI Press, Inc.

Ford, J. M., Roth, W. T., Mohs, R. C., Hopkins, W. F. III, & Kopell, B. S. (1979). Event-related potentials recorded from young and old adults during a memory retrieval task. Electroencephalography and Clinical Neurophysiology, 47, 450-459.

Friedman, D. Brown, C., Vaughan, H. G., Cornblatt, T., & Erlenmeyer-Kimling, L. (1984). Cognitive brain potential components in adolescents. Psychophysiology, 21, 83-96.

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

Gomer, F. E., Spicuzza, R. J., & O'Donnell, R. D. (1976). Evoked potential correlates of visual item recognition during memory-scanning tasks. Physiological Psychology, 4, 61-65.

Goodin, D., Squires, K., Henderson, B., & Starr, A. (1978). Age related variations in evoked potentials to auditory stimuli in normal human subjects. Electroencephalography and Clinical Neurophysiology, 44, 447-458.

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

Gratton, G., Coles, M. G. H., & Donchin, E. (1983b, abstract). Filtering for scalp distribution: A new approach (Vector Filter). Psychophysiology, 20, 443-444.

Gratton, G., Coles, M. G. H., & Donchin, E. (1985). A multivariate approach to the analysis of the scalp distribution of event-related potentials: Vector Analysis. Unpublished manuscript.

Gratton, G., Kramer, A. F., Coles, M. G. H., & Donchin, E. (1985). A comparative study of measures of the latency of event-related brain potentials. Unpublished manuscript.

Halgren, E., Squires, N. K., Wilson, C. L., Rohrbaugh, J. W., Babb, T. L., & Randall, P. H. (1980). Endogenous potentials generated in the human hippocampal formation and amygdala by infrequent events. Science, 210, 803-805.

Hansen, J. C., & Hillyard, S. A. (1980). Endogenous brain potentials associated with selective auditory attention. Electroencephalography and Clinical Neurophysiology, 49, 277-290.

Hansen, J. C., & Hillyard, S. A. (1984). Effects of stimulation rate and attribute cuing on event-related potentials during selective auditory attention. Psychophysiology, 21, 394-405.

Heffley, E. F. (1981, April). Event-related brain potentials in visual monitoring tasks: Selective attention, display format, and event frequency. Paper presented at the Eastern Psychological Association annual meeting, New York.

Hillyard, S. A., Courchesne, E., Krausz, H. I., & Picton, T. W. (1976). Scalp topography of the P3 Wave in different auditory decision tasks. In W. C. McCallum & J. R. Knott (Eds.), The Responsive Brain (pp. 81-87). Bristol: John Wright and Sons, Limited.

Hillyard, S. A., & Hansen, J. C. (in press). Attention: Electrophysiological Approaches. In M. G. H. Coles, E. Donchin, & S. W. Porges, Psychophysiology: Systems, processes, and applications. New York: Guilford Press.

Hillyard, S. A., & Kutas, M. (1983). Electrophysiology of cognitive processing. In M. R. Rosenzweig & L. W. Porter (Eds.), Annual Review of Psychology, vol. 34 (pp. 33-61). Palo Alto, CA: Annual Reviews, Inc.

Horst, R. L.,& Donchin, E. (1980). Beyond averaging II: single trials classification of exogenous event-related potentials using Step-Wise Discriminant Analysis. Electroencephalography and Clinical Neurophysiology, 48, 113-126.

Hunt, E. (1980). The validity of assumptions underlying the use of PCA and ANOVA in ERP research. Paper presented at the meeting of the Society for Psychophysiological Research, Vancouver, Canada.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual task performance. Psychophysiology, 17, 259-273.

Jasper, H. H. (1958). Report of the committee on methods of clinical examination in electroencephalography. Electroencephalography and Clinical Neurophysiology, 10, 370-375.

John, E. R., Ruchkin, D. S., & Vidal, J. J. (1978). Measurement of event-related potentials. In E. Callaway, P. Tueting, & S. H. Koslow (Eds.) Event-related brain potentials in man (pp. 93-138). Academic Press: NY.

John, E. R., Ruchkin, D. S., & Villegas, J. (1964). Signal analysis and behavioral correlates of evoked potential configurations in cats. Annals of the New York Academy of Sciences, 112, 362-420.

Johnson, R., Jr. (in press). Multiple P300s in feedback tasks. In F. Rösler & S. Sutton (Eds.), On how to interbreed endogenous ERP components with cognitive constructs. Electroencephalography and Clinical Neurophysiology.

Johnson, R., Jr., & Donchin, E. (1985). Second thoughts: Multiple P300s elicited by a single stimulus. Psychophysiology, 22, 182-194.

Johnson, R., Jr., & Donchin, E. (1978). On how P300 amplitude varies with the utility of the eliciting stimuli. Electroencephalography and Clinical Neurophysiology, 44, 424-437.

Johnson, R., Jr., & Donchin, E. (1980). P300 and stimulus categorization: Two plus one is not so different from one plus one. Psychophysiology, 17, 167-178.

Johnson, R., Jr., & Donchin, E. (1982). Sequential expectancies and decision making in a changing environment: An electrophysiological approach. Psychophysiology, 19, 183-200.

Johnston, V. S., & Holcomb, P. J. (1980). Probability learning and the P3 component of the visual evoked potential in man. Psychophysiology, 17, 396-400.

Karis, D., Bashore, T., Fabiani, M., & Donchin, E. (1982, abstract). P300 and Memory. Psychophysiology, 19, 328.

Karis, D., Chesney, G. L., & Donchin, E. (1983). "...'twas ten to one; And yet we ventured...": P300 and decision making. Psychophysiology, 20, 260-268.

Karis, D., Coles, M. G. H., & Donchin, E. (1984). The use of event-related brain potentials for prediction of performance quality in pilots of advanced tactical jet aircraft (Tech. Rep. No. CPL 83-1-F33615-82-K-0611). Champaign: University of Illinois, Cognitive Psychophysiology Laboratory.

Karis, D., Druckman, D., Lissak, R., & Donchin, E. (1984). A psychophysiological analysis of bargaining: ERPs and facial expressions. In R. Karrer, J. Cohen, & P. Tueting (Eds.), Brain and information: Event-related potentials (pp. 230-235). New York: The New York Academy of Sciences.

Karis, D., Fabiani, M., & Donchin, E. (1984). P300 and memory: Individual differences in the von Restorff effect. Cognitive Psychology, 16, 177-216.

Kerkhof, G. A. (1982). Event-related potentials and auditory signal detection: Their diurnal variation for morning-type and evening-type subjects. Psychophysiology, 19, 94-103.

Klorman, R., & Ryan, R. M. (1980). Heart rate, contingent negative variation, and evoked potentials during anticipation of affective stimulation. Psychophysiology, 17, 513-523.

Kramer, A. F. (1985). The Interpretation of the component structure of event-related brain potentials: An analysis of expert judgements. Manuscript submitted for publication.

Kutas, M., & Hillyard, S. A. (1980a). Reading senseless sentences: Brain potentials reflect semantic incongruity. Science, 207, 203-205.

Kutas, M., & Hillyard, S. A. (1980b). Reading between the lines: Event-related brain potentials during natural sentence processing. Brain and Language, 11, 354-373.

Kutas, M., & Hillyard, S. A. (1980c). Event-related brain potentials to semantically inappropriate and surprisingly large words. Biological Psychology, 11, 99-116.

Kutas, M., & Hillyard, S. A. (1982). The lateral distribution of event-related potentials during sentence processing. Neuropsychologia, 20, 579-590.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 197, 792-795.

Lacey, J. I. (1959). Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In E. A. Rubinstein & M. B. Parloff (Eds.), Research in psychotherapy (pp. 160-208). Washington: APA.

Magliero, A., Bashore, T. R., Coles, M. G. H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation process. Psychophysiology, 21, 171-186.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction times. Science, 211, 77-79.

McCarthy, G., & Wood, C. C. (in press). Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. Electroencephalography and Clinical Neurophysiology.

Molenaar, P. C. M. (1983, personal communication). Component analysis of ERPs.

Näätänen, R. (1982). Processing negativity: an evoked potential reflection of selective attention. Psychological Bulletin, 92, 605-640.

Okada, Y. C., Kaufman, L., & Williamson, S. J. (1983). The hippocampal formation as a source of the slow endogenous potentials. Electroencephalography and Clinical Neurophysiology, 55, 417-426.

Paller, K. A., Zola-Morgan, S., Squire, L. R., & Hillyard, S. A. (1984). Monkeys with lesions of hippocampus and amygdala exhibit event-related brain potentials that resemble the human P300 wave. Paper presented at the Third International Conference on Cognitive Neuroscience, Bristol, England.

Parasuraman, R., & Beatty, J. (1980). Brain events underlying detection and recognition of weak sensory signals. Science, 210, 80-83.

Parasuraman, R., Richer, F., & Beatty, J. (1982). Detection and recognition: Concurrent processes in perception. Perception & Psychophysics, 31, 1-12.

Peters, J. F., Billinger, T. W., & Knott, J. R. (1977). Event-related potentials of brain (CNV and P300) in a paird associate learning paradigm. Psychophysiology, 14, 579-585.

Pfefferbaum, A., Ford, J., & Johnson, R., Wenegrat, B., & Kopell, B. S. (1983). Manipulations of P3 latency: Speed vs. Accuracy instructions. Electroencephalography and Clinical Neurophysiology, 55, 188-197.

Pfefferbaum, A., Ford, J., Roth, W. T., & Kopell, B. S. (1980). Age-related changes in auditory event-related potentials. Electroencephalography and Clinical Neurophysiology, 49, 266-276.

Picton, T. W., & Hillyard, S. A. (1974). Human auditory evoked potentials. II: Effects of attention. Electroencephalography and Clinical Neurophysiology, 36, 191-199.

Picton, T. W., & Stuss, D. T. (1980). The component structure of the human event-related potentials. In H. H. Kornhuber & L. Deecke (Eds.), Motivation, Motor and Sensory Processes of the Brain, Progress in Brain Research (vol. 54) (pp. 17-49). Elsevier: North-Holland Biomedical Press.

Picton, T. W., Stuss, D. T., Champagne, S. C., & Nelson, R. F. (1984). The effects of age on human event-related potentials. Psychophysiology, 21, 312-325.

Picton, T., Woods, D., Stuss, D., & Campbell, K. (1978). Methodology and meaning of human evoked potential scalp distribution studies. In D. Otto (Ed.), Multidisciplinary perspectives in event-related brain potential research (pp. 515-522). (EPA-600/9-77-043). Washington, D. C.: U.S. Government Printing Office.

Polich, J., Howard, L., & Starr, A. (1983). P300 latency correlates with digit span. Psychophysiology, 20, 665-669.

Poon, L, Thompson, L. W., Williams, R. B., & Marsh, G. R. (1974). Changes of antero-posterior distribution of CNV and late positive component as a function of information processing demands. Psychophysiology, 11, 660-673.

Pritchard, W. S. (1981). Psychophysiology of P300. Psychological Bulletin, 89, 506-540.

Quilter, P. M., MacGillivray, B. B., & Wadbrook, D. G. (1977). The removal of eye movement artifact from EEG signals using correlation techniques. In _IEEE Conference on Random Signal Analysis_. Conference publication No. 159, 93-100.

Ragot, R. (1984). Perceptual and motor space representation: An event-related potential study. _Psychophysiology_, _21_, 159-170.

Remond, A. (1961). Integrated and topographical analysis of the EEG. _Electroencephalography and Clinical Neurophysiology_, _20_, 64-67.

Renault, B. (1983). The visual emitted potentials: Clues for information processing. In A. W .K. Gaillard & W. Ritter (Eds.), _Tutorials in event related potential research: Endogenous components_ (pp.159-176). Amsterdam: North-Holland Publishing Company.

Rösler, F. (1981). Event-related brain potentials in a stimulus-discrimination learning paradigm. _Psychophysiology_, _18_, 447-455.

Rösler, F., & Sutton, S. (Eds.) (in press). On how to interbreed endogenous ERP components with cognitive constructs. _Electroencephalography and Clinical Neurophysiology_.

Ruchkin, D. S., Munson, R., & Sutton, S. (1982). P300 and Slow Wave in a message consisting of two events. _Psychophysiology_, _19_, 629-642.

Ruchkin, D. S., & Sutton, S. (1983). Positive slow wave and P300: Association and disassociation. In A. W .K. Gaillard & W. Ritter (Eds.), _Tutorials in event related potential research: Endogenous components_ (pp. 233-250). Amsterdam: North-Holland Publishing Company.

Ruchkin, D. S., Sutton, S., Kietzman, M. L., & Silver, K. (1980). Slow wave and P300 in signal detection. _Electroencephalography and Clinical Neurophysiology_, _50_, 35-47.

Ruchkin, D. S., Sutton, S., Munson, R., Silver, K. & Macar, F. (1981). P300 and feedback provided by absence of the stimulus. Psychophysiology, 18, 271-282.

Sams, M., Alho, K., & Näätänen, R. (1984). Short-term habituation and dishabituation of the mismatch negativity of the ERP. Psychophysiology, 21, 434-441.

Sanquist, T. F., Rohrbaugh, J. W., Syndulko, K., & Lindsley, D. B. (1980). Electrocortical signs of levels of processing: Perceptual analysis and recognition memory. Psychophysiology, 17, 568-576.

Schwent, V. L., Snyder, E., & Hillyard, S. A. (1976). Auditory evoked potentials during multichannel selective listening: Role of pitch and localization cues. Journal of Experimental Psychology: Human Perception and Performance, 2, 313-325.

Simson, R., Vaughan, H. G., & Ritter, W. (1976). The scalp topography of potentials associated with missing visual or auditory stimuli. Electroencephalography and Clinical Neurophysiology, 40, 33-42.

Simson, R., Vaughan, H. G., & Ritter, W. (1977a). The scalp topography of potentials in auditory and visual discrimination tasks. Electroencephalography and Clinical Neurophysiology, 42, 528-535.

Simson, R., Vaughan, H. G., & Ritter, W. (1977b). The scalp topography of potentials in auditory and visual go/nogo tasks. Electroencephalography and Clinical Neurophysiology, 43, 864-875.

Skrandies, W., & Lehmann, D. (1982). Spatial principal components of multichannel maps evoked by lateral visual half-field stimuli. Electroencephalography and Clinical Neurophysiology, 54, 297-305.

Snyder, E., Hillyard, S. A., & Galambos, R. (1980). Similarities and differences among the P3 waves to detected signals in three modalities. Psychophysiology, 17, 112-122.

Squires, K. C., & Donchin, E. (1976). Beyond averaging: the use of discriminant functions to recognize event related potentials elicited by single auditory stimuli. Electroencephalography and Clinical Neurophysiology, 41, 449-459.

Squires, K. C., Donchin, E., Herning, R. I., & McCarthy, G. (1977). On the influence of task relevance and stimulus probability on event-related-potential components. Electroencephalography and Clinical Neurophysiology, 42, 1-14.

Squires, K. C., Hillyard, S. A., & Lindsay, P. H. (1973). Cortical potentials evoked by confirming and disconfirming feedback following an auditory discrimination. Perception and psychophysics, 13, 25-31.

Squires, K. C., Squires, N. K., & Hillyard, S. A. (1975). Decision-related cortical potentials during an auditory signal detection task with cued intervals. Journal of Experimental Psychology: Human Perception and Performance, 1, 268-279.

Squires, K. C., Wickens, C., Squires, N. K., & Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. Science, 193, 1142-1146.

Squires, N. K., Donchin, E., Squires, K. C., & Grossberg, S. (1977). Bisensory stimulation: Inferring decision-related processes from the P300 component. Journal of Experimental Psychology: Human Perception and Performance, 3, 299-315.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive wave evoked by unpredictable stimuli in man.

Electroencephalography and Clinical Neurophysiology, 38, 387-401.

Strayer, D. L., Karis, D., Coles, M. G. H., & Donchin, E. (1984, abstract). P300 latency and reaction time are dissociated in a Sternberg task. Psychophysiology, 21, 600.

Stuss, D. T., & Picton, T. W. (1978). Neurophysiological correlates of human concept formation. Behavioral Biology, 23, 135-162.

Stuss, D. T., Sarazin, F. F., Leech, E. E., & Picton, T. W. (1983). Event-related potentials during naming and mental rotation. Electroencephalography and Clinical Neurophysiology, 56, 133-146.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Information delivery and the sensory evoked potential. Science, 150, 1187-1188.

Sutton, S., Tueting, P., Zubin, J., & John, E. R. (1967). Information delivery and the sensory evoked potential. Science, 155, 1436-1439.

Tatsuoka, M. M. (1971). Multivariate analysis. N.Y.: Wiley.

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

Vaughan, H. G., Ritter, W., & Simson, R. (1980). Topographical analysis of auditory event-related potentials. In H. H. Kornhuber & L. Deecke (Eds.), Motivation, motor, and sensory processes of the brain. Progress in Brain Research, vol. 54 (pp. 279-285). Amsterdam: Elsevier-North Holland.

Verleger, R., Gasser, T., & Mocks, J. (1982). Correction of EOG artifacts in event-related potentials of the EEG: Aspects of reliability and validity. Psychophysiology, 19, 472-480.

Warren, L. R. (1980). Evoked potential correlates of recognition memory. Biological Psychology, 11, 21-35.

Warren, L. R., & Marsh, G. R. (1979). Changes in event related potentials during processing of stroop stimuli. International Journal of Neuroscience, 9, 217-223.

Wastell, D. G. (1977). Statistical detection of individual evoked responses: an evaluation of Woody's adaptive filter. Electroencephalography and Clinical Neurophysiology, 42, 835-839.

Whitton, J. L., Lue, F., & Moldofsky, H. (1978). A spectral method for removing eye movement artifacts from the EEG. Electroencephalography and Clinical Neurophysiology, 44, 735-741.

Wood, C. C., Allison, T. (1981). Interpretation of evoked potentials: a neurophysiological perspective. Canadian Journal of Psychology, 35, 113-135.

Wood, C. C., Allison, T., Goff, W. R., Williamson, P., & Spencer, D. C. (1980). On the neural origin of P300 in man. Progress in Brain Research, 54, 51-56.

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.

Wood, C. C., McCarthy, G., Allison, T., Goff, W. R., Williamson, P. D., & Spencer, D. D. (1982). Endogenous event-related potentials following temporal lobe excision in humans. Bulletin of the Society for Neuroscience, 8, 976.

Wood, C. C., & McCarthy, G., Squires, N. K., Vaughan, H. G., Woods, D. L., & McCallum, W. C. (1984). Anatomical and physiological substrates of event-related potentials. In R. Karrer, J. Cohen, & P. Tueting (Eds.), _Brain and Information: Event-related Potentials_ (pp. 681-721). New York: The New York Academy of Sciences.

Wood, C. C., & Wolpaw, J. R. (1982). Scalp distribution of human auditory evoked potentials. II. Evidence for overlapping sources and involvement of auditory cortex. _Electroencephalography and Clinical Neurophysiology_, _54_, 25-38.

Woody, C. D. (1967). Characterization of an adaptive filter for the analysis of variable latency neuroelectrical signals. _Medical and Biological Engineering_, _5_, 539-553.

Appendix A

Typical P300 Paradigms

In this appendix we describe typical P300 paradigms, focusing on the nature and time course of the stimuli, as well as the probabilities employed, the nature of the subject's task, and the scalp distribution of P300. In particular, we want to stress the use of scalp distribution as a dependent variable in the way amplitude has been used, and examine how scalp distribution varies with changes in information processing (see Donchin, 1979).

All ERP paradigms can be divided into two groups on the basis of the nature of the stimuli: those in which an event is composed of a single stimulus and long sequences of stimuli are presented (long sequential presentations), and those in which several stimuli define an event. The first group includes the "classic" oddball paradigm, and the second a variety of what we call S1-S2 paradigms. These include signal detection and feedback paradigms.

Long Sequential Presentations

The Classic Oddball Paradigm

The "classic oddball task" is a paradigm in which stimuli that can be classified into two categories are presented sequentially. The two categories are presented with complementary probabilities. Often, the mix of probabilities is an independent variable. Usually one of the categories is <u>rare</u>, while the other is <u>frequent</u>. Rare-frequent stimulus probabilities are generally between .33-.67 and .5-.95. Both fixed and variable inter-stimulus intervals (ISIs) are used. In most oddball tasks, ISIs are between one and two seconds, although ISIs between two and three seconds are also common. Only rarely are there ISIs longer than three seconds or shorter than one.

In the oddball task subjects are usually required to classify each stimulus and give a discriminative response. The subject's response may consist of increasing an internal count whenever a stimulus from a designated category (a <u>target</u>) is presented, or responding to stimuli from one (go-nogo) or from both categories (choice reaction time). Note that a category can be composed of only one stimulus, presented repeatedly, or of a variety of stimuli, all of which can be classified as belonging to that category. Note also that the subject must attend to each stimulus in order to determine whether or not it is a target. The distinction between target and non-target can be based on physical differences between two stimuli (e.g., pitch, intensity, or duration). Semantic classifications are also used (e.g., male versus female names, or animals versus nonanimals.) To give an example, a simple classic auditory oddball paradigm might involve presenting 1000 and 1500 Hz tones randomly, with the 1500 Hz tone occurring 20% of time. The subject would be instructed to count the high (1500 Hz) tones and give a running total at the end of a block of trials (usually 100 or more).

Thus, the term "oddball" covers a wide range of paradigms that vary in the type of stimuli used, the probability levels, and the response required. Although the name "oddball" originally referred to sequences in which one of the categories appeared with a low probability (an oddball), it is now also used when there are equiprobable stimuli (this is the case in dual task paradigms, in particular; see Isreal, Chesney, Wickens, & Donchin, 1980; Donchin, Kramer & Wickens, 1982, in press).

In a classic oddball paradigm, P300 amplitude increases as the probability of the eliciting event decreases (probability effect), and events that require a response, or extra processing, elicit larger amplitude P300s than other events (target effect). Usually, the rare event is also the target, and the difference between rare and frequent events is a function of both probability and the target effect, since both contribute to P300 amplitude. It is also common to have one frequent stimulus and two rare stimuli, only one of which is the target. When two rare stimuli are used, it is possible to partially disentangle the target effect from the probability effect by comparing the ERPs to rare targets with those to rare non-targets (target effect), and the ERPs to rare non-targets with those to frequent non-targets (probability effect). Another way of isolating the probability effect is to instruct the subject to count or respond to all stimuli. There is now evidence that the ISI modulates the effect of probability on P300 amplitude (see Heffley, 1981; Fitzgerald & Picton, 1981).

The latency of P300 has been shown to reflect the time required for stimulus evaluation and categorization, and to be relatively independent of response selection and execution processes (Kutas, McCarthy, & Donchin, 1977; McCarthy & Donchin, 1981; Magliero et al., 1984). For example, in the reliability study mentioned above the tasks required different amounts of processing. In Figure 2 we presented average waveforms from one representative subject in a simple auditory task (count one of two tones), a relatively simple visual task (count "H"s or "S"s), and a relatively complex visual task (count male or female names). Differences in P300 latency are clearly visible in this figure. As discussed above, P300 latency is a defining attribute in the sense that it must be greater than a minimum. This is usually more than 300 ms, except in simple auditory paradigms, where it may be shorter.

The scalp distribution of P300 in oddball tasks is usually parietally maximum, with Pz > Cz > Fz. Unfortunately, extensive information on scalp distribution is only rarely reported. Moreover, it is usually impossible to determine from published articles the variability among subjects in scalp distribution or if there are differences in scalp distribution between target and non-target conditions. We think that data on scalp distribution is important, both for identifying P300, and for understanding the relationships among the many late positivities.

The omitted stimulus paradigm. The omitted stimulus paradigm is an oddball task in which the rare "stimulus" is not a physical event, but rather the omission of a regularly occurring stimulus. For example, a subject may be presented with an easily detectable tone or visual pattern once every second, and instructed to count how often the tone or pattern fails to appear. Since there is no actual physical stimulus present, ERPs elicited by the counted omission can contain only endogenous components. The results of these studies were important in the early arguments that P300 was indeed an endogenous component. Comparisons between missing stimulus potentials (emitted potentials) and stimulus evoked potentials were taken as evidence that P300 was not modality specific, but instead represented a particular class of information processing. Scalp topography is generally similar to that found in classic oddball tasks, with a parietal maximum (e.g., Simson et al. 1976; Friedman, Brown, Vaughan, Cornblatt, & Erlenmeyer-Kimling, 1984), although Hillyard et al. (1976), and Picton and Hillyard (1974) reported Cz equal to Pz.

Selective attention. Most experiments investigating the effects of selective attention on ERPs have used variations of the oddball paradigm (see Hillyard & Hansen, in press, and Hillyard & Kutas, 1983, for reviews). Unlike traditional oddball tasks, these experiments typically present two simultaneous sequences, each with a rare and frequent stimulus. Auditory stimuli are almost exclusively used, and the sequences are usually distinguishable on the basis of pitch or spatial location. (In early papers each sequence was described as a separate "channel", e.g., Schwent, Snyder, & Hillyard, 1976). The subject's task is to detect targets from only one of these sequences. In general, ISIs are much shorter than in the classic oddball task (less that one second, sometimes less than 500 ms), and the focus is usually not on P300, but on an early negativity that was at first assumed to be N100, but was later determined to be an endogenous component, called Nd. Most analysis is on the non-targets, and focuses on examining the difference between non-targets on the attended channel versus non-targets on the non-attended channel(s). Recording epochs are often very short, and there is frequently no measurement of P300, which is elicited only by the targets. When P300 is measured there is often little information on scalp distribution. Furthermore, only infrequently are recordings made from Fz, Cz, and Pz. Cz is almost always used, often with

C3 and C4 lateral placements. When data on P300 scalp distribution are presented, P300 is maximum at the parietal electrode (e.g., Hansen & Hillyard, 1980, 1984).

Variations on the Classic Oddball Paradigm

These are paradigms in which sequential stimuli appear, superficially, not to fit the pattern of a classic oddball. Often, the two categories used to construct the sequence are presented with equal probability. Each of the categories may be represented by many different stimuli, and the categorization rule applied before the category membership can be determined may be very abstract. Furthermore, at times no overt response is required to each stimulus. Yet all these paradigms can be classified as oddball tasks in the sense that the subject is presented with a sequence of events that must be categorized, and a specific response, not necessarily overt, is requested for each category. In recognition paradigms, for example, ERPs may be elicited by words in a test phase in which subjects must classify each of a series of words as "old" (previously seen) or "new" (Warren, 1980; Karis, Bashore, Fabiani, & Donchin, 1982). Similarly, in memory search (Sternberg) paradigms with multiple probes (Gomer, Spicuzza, & O'Donnell, 1976; Strayer, Karis, Coles, & Donchin, 1984) subjects decide whether or not each of a series of probes is, or is not, a member of a previously presented memory set. In other memory paradigms, there is less similarity to oddball tasks. For example, subjects may be instructed to attend to a sequence of words in order to be able to recall them subsequently (Karis, Fabiani, & Donchin, 1984). Subjects may also be presented with sentences, one word at a time, and told that they should pay attention in order to answer questions ("about their contents") at the end of the experiment. In most of these studies the focus has been on the last word of the sentence, and on differences in ERPs (usually a late negative peak, N400) as a function of semantic or orthographic incongruity (Kutas & Hillyard, 1980a, 1980b, 1980c, 1982), but a P300 is also recorded.

Other paradigms in this group are more similar to oddball tasks. In experiments using the Stroop effect, for example, subjects may be required to name the color of a printed word, either by a keypress (Warren & Marsh, 1979), or vocally (Duncan-Johnson & Kopell, 1981). Ragot (1984) had subjects respond differentially to an equiprobable sequence of green or red lights. In these studies, unlike most of the memory experiments, the same small set of stimuli is typically presented repeatedly with equal probability. In the cases where the nature of the stimuli are similar to an oddball there is usually some manipulation that is intended to alter the information processing required for responding correctly. In the Stroop experiments there is a conflict between the color word (e.g., "red") and the color of ink used to print the word (e.g., blue), whereas in Ragot's (1984) experiment, the light could be on the same side as the response hand or on the other, and there were conditions with and without the hands crossed.

In Pfefferbaum, Ford, Johnson, Wenegrat, and Kopell (1983) two horizontal lines were presented on each trial, and the subject judged whether their lengths were the same or different. Difficulty was then manipulated by varying the difference between the lines. To the extent that the subject is uncertain about the decision this resembles a signal detection experiment, or manipulations aimed at increasing "equivocation" (see the section on Slow Waves in Appendix B). In Friedman et al. (1981)

sequences of numbers were presented, from 02 to 19, with a target defined in one condition as the repetition of the immediately preceding number. To respond accurately the subject must hold each item in working memory until after the presentation of the following item. There are many cases like this in which there are great similarities to a classic oddball task, but enough differences in the processing required to categorize events to justify a separate classification.

In many cases data on scalp distribution are not presented, or electrode placements at Fz, Cz, and Pz are not all used. However, from the data that are available, it appears that in all these paradigms with long sequential presentations P300 is parietally maximum with a classic distributional pattern (Pz > Cz > Fz).

S1-S2 Paradigms

In contrast to the paradigms reviewed above, the "S1-S2" paradigms involve a series of trials, each consisting of two or more stimuli. We label them S1-S2 because they all have some similarity to the simple paradigm consisting of a warning followed by a second, imperative, stimulus that indicates a response should be made. There is often a short sequence of stimuli, some conveying task information, some providing feedback, and others warning that information or feedback will occur. We will divide these paradigms into three groups based on the nature of the eliciting stimulus: "normal" S1-S2 paradigms, signal detection paradigms, and feedback paradigms. In the first group, ERPs may be recorded to the warning and/or the imperative stimulus, while in the latter two paradigms ERPs are recorded to threshold level stimuli or to feedback stimuli (concerning either a prediction or task performance). In all these paradigms a CNV may be generated between the warning and imperative stimuli, or over the course of several stimuli. This CNV, and its resolution, can sometimes exert a strong influence on the morphology and scalp distribution of P300.

"Normal" S1-S2 Paradigms

There is great diversity in what we are calling normal S1-S2 paradigms. The modal pattern is a warning stimulus followed by a more complex stimulus requiring a response. Often, however, the first stimulus provides information concerning the processing or response to the second stimulus, or two comparable stimuli are presented and some comparison must be made between them. The interval between the two stimuli is generally between one and two seconds, although it ranges from 250 ms to over five seconds. The interval between trials is usually between four and seven seconds, but it is sometimes longer (e.g., 35 seconds in Klorman & Ryan, 1980). There may actually be more than two stimuli. Sometimes there is some feedback indicating whether or not the response was correct, and ERPs to the event providing feedback may also be recorded (e.g., Stuss & Picton, 1978). Sometimes there is a warning stimulus followed by two or more complex stimuli. These paradigms are primarily visual, using words or complex visual patterns or pictures. ERPs may be elicited by the first stimulus, the second, or both. In general, P300s are usually larger when elicited by S2, because it is only after that point that a decision can be made and a response initiated. The probability of different types of trials or stimuli, and of responses, is usually equal, so that probability is not a confounding variable.

The tasks, and the processing required, differ radically. Subjects may have to learn the relationship between the two stimuli over the course of the experiment, as with paired associates, or in terms of probability relationships, or concept learning (Peters, Billinger, & Knott, 1977; Johnston & Holcomb, 1980; Rösler, 1981). Subjects may have to determine whether or not a probe letter or number was a member of a memory set (Ford, Roth, Mohs, Hopkins, & Kopell, 1979), decide whether two words are orthographically, phonologically, or semantically similar or dissimilar (Sanquist, Rohrbaugh, Syndulko, & Lindsley, 1980), name a picture, read a word, or perform a mental rotation (Stuss, Sarazin, Leech, & Picton, 1983), anticipate a neutral or affectively charged picture (Klorman & Ryan, 1980), or respond on the basis of the combination of words presented at S1 and S2 (McCarthy & Donchin, 1981; Magliero et al., 1984). This is by no means an exclusive list, but it gives an idea of the diversity involved.

The scalp distribution of P300 is generally parietally maximum, although once again, many studies do not use several midline electrodes, or do not report these data. There are interesting variations in scalp distribution in go/no-go paradigms. In these paradigms there are generally two possible imperative stimuli, each occurring on 50% of the trials. One indicates a response should be made, while the other indicates that no response should be made. Hillyard et al. (1976) found that P300 on the no-go trials, compared to the go trials, became smaller at Pz and larger at Fz. Simson et al. (1977b), on the other hand, found that most change in amplitude occurred at Cz. On no-go trials, compared with go trials, Cz became much more positive, while Pz remained constant (and Cz thus became equal to Pz). In a go/no-go experiment in progress with 47 of the subjects used in the reliability study described above, we found results similar to Simson et al. (1977b). Fz and Pz remained constant across all trials, but there was a dramatic increase in P300 at Cz during no-go trials. An explanation centering around differences in motor potentials between go and no-go trials is not without problems. In our study subjects were instructed to respond as quickly as possible after S2, while Simson et al. (1977b) required a delayed motor response. They suggest that the "nogo P3 presents a central extension that could represent a contribution from CNV resolution" (p. 871). This is plausible, but other explanations are also possible.

Signal Detection Paradigms

ERPs have been recorded to weak auditory, visual, and somatosensory stimuli in detection paradigms, although auditory paradigms are the most common. The typical paradigm involves a warning stimulus, followed less than a second later by a threshold-level stimulus on half of the trials. A few seconds later (1.5 to 3.5, usually) a response cue is presented and the subject must respond, by indicating whether or not a stimulus was present. Sometimes the subject is also asked to identify the stimulus from among several possibilities, and then provide confidence ratings on these decisions. Stimulus intensities are adjusted individually and detection accuracy ranges from about 75 to 95%. Inter-trial intervals are generally 3 to 5 seconds, but range from two to over ten. ERPs are recorded to the threshold level stimulus, or its absence, and a marker light may be used to time lock responses on trials where no stimulus is presented (e.g., K. Squires, Squires, & Hillyard, 1975). ERPs have also been recorded to the feedback (e.g., Squires, Hillyard, & Lindsay, 1973), in which case ERPs are likely to be more similar to those described in the section on feedback paradigms below. In signal detection paradigms the relationship between Cz and Pz is often altered. Cz is about equal to Pz in some studies (Kerkhof, 1982; Parasuraman, Richer, & Beatty, 1982, using their mean amplitude values, see also Parasuraman & Beatty, 1980; Hillyard et al., 1976), or even larger than Pz in others (Snyder et al., 1980; K. Squires et al., 1975). Snyder et al. (1980) expressed the amplitude at Fz and Pz as a percentage of Cz. Pz amplitudes were 86% of Cz for auditory and somatic stimuli, and 89% for visual. In Hillyard et al. (1976) Fz equalled 77% of Cz, and Pz 99%. Picton and Hillyard (1974) combined an oddball and detection task by presenting a click every second and at intervals ranging from 5 to 30 seconds slightly lowering the intensity of a single click. Subjects were able to detect between 80 and 95% of these low intensity clicks, and were instructed to count them. In this case, of course, the detection is not of the presence of a stimulus, but of a slight difference between the standard and target. The scalp distribution of the P300 to the targets, expressed as a percentage of Cz, was 54% for Fz and 165% for Pz. This is a distributional pattern more typical for oddball tasks than for signal detection. Does this mean that the information processing was more similar to that in an oddball? Would the distribution change if the discrimination became more difficult? Are there likely to be trial by trial variations in distribution as a function of confidence? When we develop a more detailed understanding of scalp topography we hope to be able to answer such questions.

Feedback and Guessing Paradigms

In these paradigms, each trial consists of two phases. First, a subject predicts which of several stimuli will occur (guessing paradigm), or performs some task (feedback paradigm). Second, feedback is provided by presenting one of the several possible stimuli, or by one of two signals signifying that the subject was correct or incorrect in the previous task. ERPs are recorded to the feedback stimulus and the differences between ERPs elicited by confirming and disconfirming feedback in various conditions are usually examined. The time sequence of these paradigms is hard to characterize due to the variability of the event sequence in a single trial. The sequence may be as simple as a keypress initiated by the subject indicating a prediction, followed immediately (after 500 ms) by the feedback (e.g., Karis, Chesney, & Donchin, 1983) or as complex as in Ruchkin et al. (1981), where a pair of clicks, 600 ms apart, was presented after the subject placed his finger on the response key. The subject then lifted his finger after estimating another 600 ms, afterwhich he replaced his finger on the key. Two to three seconds later two events were presented 5.5 to 6 seconds apart, with each event consisting of either one or two clicks. Feedback was sometimes presented by the second event alone, and sometimes by the combination of both events.

Feedback is usually presented by a single stimulus, although in some experiments, as in the Ruchkin et al. (1981) experiment just described, the nature of the feedback depends on the combination of two successive events (see also Ruchkin, Munson, & Sutton, 1982). In most prediction experiments the subject has no way to accurately predict which stimulus will actually appear (e.g., Sutton, Tueting, Zubin, & John, 1967), which is why these are sometimes called guessing paradigms. There are, however, exceptions. In situations where learning is taking place, the subjects may eventually learn to judge their performance and anticipate the nature of the feedback (Karis, Druckman, Lissak, & Donchin, 1984; DeSwart, Kok, & Das-Smaal, 1981; Stuss & Picton, 1978), or to learn the pattern of stimuli and thus be able to accurately predict which will be presented (Poon, Thompson, Williams, & Marsh, 1974). In other cases (time estimation, for example, Ruchkin et al., 1981) the criterion for correct performance may be revised throughout the experiment to maintain performance at a constant level. P300 scalp distribution is typically maximum at Pz, although Cz is sometimes equal to Pz (Johnson & Donchin, 1978; Campbell, Courchesne, Picton, & Squires, 1979, see experiments 1 and 6). Campbell et al. (1979) also reported that activity at Fz increases, relative to Cz, as the probability of being correct decreases.

As an historical aside, we should point out that the first published report on P300 used a guessing-feedback paradigm (Sutton et al., 1965). After a warning stimulus subjects guessed whether a click or light flash would occur, and ERPs were recorded to the subsequent feedback (a click or flash). There were two different warning stimuli - one was always followed by a flash (or a click in other runs), while the second could be followed by either a click or a flash, with unequal probabilities. When the subject had to guess which stimulus would occur, and was uncertain as to which feedback stimulus would follow, P300 was much larger than when subjects could predict accurately.

Johnson and Donchin (1985; Johnson, in press) point out that feedback paradigms involve two stages of information processing. First, there must be stimulus evaluation and categorization, and then the subject must make multiple decisions. In order to improve performance the subject must compare the feedback with internally generated information, judge accuracy, and modify plans and strategies for future performance. Because processing continues for an extended period, and may involve a series of decisions, Johnson and Donchin (1985) suggest that additional P300s may be emitted. Since these P300s are not time locked to the feedback, they may "smear" together in averages and produce what has often been labeled a Slow Wave. Slow Waves reported in many experiments may thus be composed of multiple P300s.

Appendix B

Late Positivities

Frontal "P300"s, and P3a

In recent years the existence of frontal P300s, and the dichotomy between "P3a" and "P3b", have been used in arguments about the unitary nature of P300. In this appendix we will review the paradigms used to record these positivities, and try to reconcile their existence with our views about P300.

Courchesne et al. (1975; see also Courchesne, 1977, Courchesne, Hillyard, & Courchesne, 1977, and Courchesne, Courchesne, & Hillyard, 1978) presented four types of visual stimuli to their subjects at regular intervals of 1.3 seconds. These were the number 2, the number 4, "novel" stimuli, and "simple" stimuli. Each simple stimulus consisted "...of an easily recognized black and white pattern (e.g., the word "THE", a simple line drawing of a face, a black and white grid, geometric figures, etc.)" (p. 132), while novel stimuli were "...completely novel (i.e., complex, colorful abstract-type drawings which were unrecognizable)" (p. 131). In the condition of interest here 4s were targets and occurred 10% of the time. Instructions were to count the number of 4s. Eighty percent of the stimuli were 2s. The other 10% were novels for some subjects, or novels (5%) and simples (5%) for others. Neither novels or simples were ever repeated. The novels elicited what have been referred to as "frontal P300s". The maximum was actually at the vertex, but amplitudes were larger at Fz (93% of Cz) than Pz (72% of Cz) (in Courchesne, 1977, amplitude in adults was 15.2 μv at Fz, 15.5 μv at Cz, and 12.7 μv at Pz). Latency was in the P300 range, with averages between 360 and 450 ms. Simples also elicited P300s largest at Cz, but Pz was almost as large, and Fz was smaller. The counted 4s elicited a traditional parietally maximum P300. When subjects were instructed to count novels, or simples, P300 amplitude at Pz increased. While this experiment is of considerable interest, it would have been easier to accept the existence of this frontal positivity if there had been reports of this component from other laboratories.

N. Squires, Squires, and Hillyard (1975) used an oddball paradigm with loud and soft tones, or tones of two different frequencies, and an ISI of 1.1 seconds. Probability was varied, as was the task: subjects sometimes counted loud tones, sometimes soft tones, and sometimes ignored both and read a book. During the ignore condition the rare (10%) stimulus elicited an ERP with a small early positive peak (220 - 280 ms, and approximately 6 μv in amplitude) that was largest at the frontal or central sites. This was labeled P3a. During attend conditions the large positivity was later (310 to 380 ms) and similar to a traditional P300 in scalp topography. This was labeled P3b. In the attend condition there were often both a P3a and a P3b, although it was often hard to identify the P3a. It is now common to identify any tiny deflection immediately before the P300, or during its initial positive deflection, as a "P3a." Usually little concern is given for scalp topography, or for the fact that in the original report P3a was seen primarily in the ignore conditions. Little attention has been focused on the functional significance of P3a.

The existence of the frontal P300 does not mean, as Courchesne et al. (1975) argue, that P300 is not a unitary phenomenon. Another interpretation is that this is a separate component, elicited in a very specific paradigm, or that it results from the overlap with other components. We argue that the same may be true for the "P3a" reported by N. Squires et al. (1975). It is unfortunate, we feel, that the term "P3a" has become "institutionalized", against the advice of N. Squires et al. (1975). They wrote, "We do not intend for the labels 'P3a' and 'P3b' to become institutionalized and only used them as a shorthand notation for the present paper" (p.399). With institutionalization comes an acceptance that may not be warranted. Indeed, as Squires, Donchin, Herning, and McCarthy (1977) have noted, there is a confusion between the morphological statement that P300 is sometimes preceeded by a positive peak, that they labeled P3a, and the theoretical statement that a component, labeled P3a, appears between P200 and P300. Squires et al. (1977) did observe a weak, and highly variable morphological P3a. But, it was also clear in their data that this peak was associated with the N200 coomponent. That is, the only component extracted in their PCA between N100 and P300 was the N200 component. The P3a may therefore be the morphological result of the positive-going segment of the N200. Indeed, Snyder, Hillyard, and Galambos (1980), in studies following the Squires et al. (1977) paper, consistently refered to the "N2-P3a" component, as if these are two features of the same component. The institutionalizing of the P3a is one of the more unfortunate consequences of the failure of some investigators to distinguish between morphological descriptions and theoretical analysis.

In comparing P3a with the frontal P300, the only similarity is the more frontal distribution than a classical P300. As Courchesne et al. (1975) point out, these two "components" share little else in common, differing in latency, peak amplitude (P3a was much smaller), modality and complexity of the eliciting stimulus, susceptibility to habituation, and effect of attention. Although some studies label peaks as P3a (e.g., Polich, Howard, & Starr, 1983), others fail to find it (e.g., Simson et al. 1976, 1977a).

Ruchkin et al. (1981) found that feedback signals in a complex S1-S2 paradigm elicited two P300s and a late positive Slow Wave (the experiment is described in Appendix A in the section on Feedback paradigms). They suggested that the first P300 (P300E, E for early) might be a P3a, because it had a fronto-central distribution and occurred between 240 and 320 ms. The second P300 (P300L, L for late) was a traditional parietally maximal P300, with a latency range of 400 - 600 ms. Possible functions of P300E were not discussed, however, and the meaning of the differences between P300E and L were never explained. Stuss and Picton (1978) measured both P300 (in a time window from 275 to 500 ms poststimulus) and a subsequent P4 (500 to 800 ms poststimulus). No PCA was performed, but from the waveforms presented their P4 appears to be what would now be called a Slow Wave.

Slow Wave

N. Squires et al. (1975) identified a slow deflection that they called "Slow Wave." The Slow Wave, measured in their paper as the mean amplitude from 400 to 500 ms post-stimulus, was negative frontally and positive at Cz and Pz, with a Pz maximum. Since that time, emphasis on the Slow Wave has grown rapidly. PCA is often used to distinguish Slow Wave from P300, because Slow Waves have been reported starting as early as 100 ms after stimulus onset (see K. Squires et al., 1977, for a replication involving PCA of the N. Squires et al., 1975, experiment). Slow waves may often last over 1000 ms. In average waveforms Slow Wave usually appears as a continuing positivity after P300, an interruption in the return to baseline or, where there is not a prominent P300, a very broad positivity. When PCA is not used, Slow Wave is usually measured as the average amplitude over several hundred ms in a time period after P300 (the window chosen is quite variable). Stuss et al. (1983) measured two Slow Waves, one between 500 and 800 ms, the other between 1000 and 1200 ms. Their reasoning for choosing these two particular intervals, however, is not presented.

Ruchkin, Sutton, and their colleagues have focused on Slow Wave in a number of studies (Ruchkin et al., 1982; Ruchkin et al., 1981; Ruchkin, Sutton, Kietzman, & Silver, 1980; see Ruchkin & Sutton, 1983, for a review) and argue that Slow Wave represents "additional" or "further" processing activities continuing after P300. When task difficulty increases, especially when a clear cut decision is impossible, Slow Wave will increase. For example, by increasing the perceptual difficulty of a task there is a loss of information which they call "equivocation", "due to the subject's _a posteriori_ uncertainty concerning which stimulus occurred in the event" (p.630). Ruchkin et al. (1982) examined the effects of equivocation on both P300 and Slow Wave. They increased the perceptual difficulty of a same-different judgment by varying the intensity of clicks (e.g., in the equivocation conditions, discriminate a 7 dB click from no click, or one at 33 dB from 40 dB) and found that P300 amplitude decreased with equivocation (increased perceptual difficulty), while Slow Wave amplitude increased. Increased equivocation, they argue, requires a "mobilization of effort" and "increased processing", although, "The nature of the putative additional processing is not clear at this time" (Ruchkin & Sutton, 1983, p. 242).

In Sanquist et al. (1980) subjects were required to judge whether two words were the same or different based on either orthography, phonology, or semantics. One would expect that more processing would be required for a semantic comparison than an orthographic one. As Ruchkin and Sutton would predict, the Slow Wave was affected by the type of comparison required, while P300 was influenced only by the type of judgment, words requiring "same" responses eliciting larger P300s than words requiring "different" responses. Similarly, Karis, Fabiani, and Donchin (1984) found a Slow Wave only in subjects who used complex associative strategies in a free recall task, and not in subjects who used simple rote strategies. They speculated that while the P300 reflected processing associated with the initial encoding of a word and the activation of its representation in memory, Slow Wave reflected the processing that continued long after P300, and was related to the subject's strategies of combining and associating the present word with previous words. The Slow Wave reported by Karis, Fabiani, and Donchin (1984), however, was positive frontally, not negative, as is usually reported. Stuss and Picton (1978) also reported a frontally positive Slow Wave under conditions where a similar explanation may apply. They found a Slow Wave (which they called a "sustained potential") that was larger while the subject was learning a concept formation task (during what they called "preinsight") than after learning (Fz was not recorded; this finding was observed primarily at F4). Parasuraman et al. (1982) recorded a Slow Wave with an "unconventional" distribution (Cz maximum instead of Pz). Indeed, there is evidence that the Slow Wave may consist of two functionally independent components. Friedman et al. (1984) divided Slow Wave into two parts, one frontal and one parietal. They found that P300 and the frontally negative Slow Wave were large to targets, while parietal positive Slow Wave was not. It was P300 and parietal Slow Wave, however, that did not vary with age, while the negative Slow Wave did. Others have also found differences between frontal and parietal Slow Wave (Fitzgerald & Picton, 1981; Ruchkin et al., 1980). Ruchkin and Sutton (1983) sum up two major unresolved issues with respect to Slow Wave. First, how many different Slow Waves are there, and second, "What are the unique functional correlates of each of them" (p.249)? The measurement of Slow Wave, and the separation of Slow Wave from P300, is still problematical. This is especially true when paradigms are used in which late or variable latency P300s are generated. This is because the separation of P300 and Slow Wave is often based on procedures, such as PCA and area measures, that assume P300 latency remains constant. When latency varies, or when both components substantially overlap in time, problems may arise and measurements may contain contributions from both components.

Appendix C

Stimulus Modality and Subject Factors

The large majority of ERP researchers use visual and auditory stimuli, and it is consistently found that P300s elicited by visual stimuli are both larger and later than those elicited by auditory stimuli. The findings with somatic stimuli vary with respect to their relationship to visual and auditory stimuli. The latency difference between visual and auditory presentations may be as large as 100 ms or more. Using an oddball paradigm Simson et al. (1977a) found an average P300 latency of 350 ms for auditory stimuli and 465 ms for visual. In an omitted stimulus paradigm they found a similar difference of 100 ms; there was an average P300 peak of 465 ms for omitted auditory stimuli, and 565 ms for omitted visual stimuli (Simson et al., 1976). Snyder et al. (1980), in a signal detection paradigm, also found that visual latencies are longer, but the difference was only 22 ms (auditory average = 449 ms, visual average = 471 ms). Somatic stimuli elicited the shortest latency (389 ms). In oddball paradigms Picton et al. (1984) found P300 an average of 89 ms longer for visual than auditory stimuli, but there was no statistically significant difference in amplitude. Somatic stimuli also elicited P300s that were longer than those elicited by auditory stimuli (by an average of 64 ms). In the experiment on reliability discussed above we also found differences between auditory and visual stimuli in P300 latency (see Figure 4).

N. Squires, Donchin, Squires, and Grossberg (1977) also found large differences in P300 latency elicited by visual and auditory stimuli in two oddball tasks (500 ms versus 360 ms). They argued, however, that the auditory oddball task, which was composed of 1000 and 1500 Hz tones, might have required an easier discrimination than the visual oddball task, which was composed of left and right-pointing arrows. To test this hypothesis they simplified the visual oddball task by using flashes of colored light as stimuli, and developed both easy and difficult oddball tasks within each modality (easy: 1100 versus 1000 Hz, orange versus blue; difficult: 1100 versus 1060 Hz, orange versus yellow). Now there was a large effect of difficulty on P300 latency, but not modality (in the easy condition, 359 and 371 ms for auditory and visual; in the difficult conditions, 419 and 420 ms for auditory and visual). Since simple tones are often used in auditory paradigms, but a variety of quite complex visual stimuli are used, many of the differences reported between auditory and visual stimuli may be related primarily to the complexity of the stimuli and the difficulty of the discrimination required.

There are few reports of sex differences in P300 amplitude, but when differences are examined it is usually reported that amplitudes are larger for females. In Picton et al. (1984) females had larger P300s than males, both in an auditory oddball task (12.9 μv versus 10.4 μv) and in a visual oddball task (17.4 μv versus 11.3 μv). Becker and Shapiro (1980) also found that females had significantly larger amplitude P300s in a study on the orienting response. Picton et al. (1984) argue that sex difference in P300 most likely "represents some physical difference in head size or skull-thickness rather than any cognitive differences between the sexes" (p.321).

There are many studies that report changes in P300 latency and amplitude with age. Most emphasis has been on P300 latency, which has been found to increase with age from young adults to elderly people. The amount of change varies, but it is usually between 1 and 2 ms per year (Brown, Marsh, & LaRue, 1983; Ford & Pfefferbaum, 1985; Picton et al., 1984; Goodin, Squires, Henderson, & Starr, 1978). The reasons for the age-related changes in P300 latency are not completely clear, although since nerve conduction time does increase with age (Allison, Hume, Wood, & Goff, 1984), this must account for part of the change in P300 latency. The change in amplitude may be related to a shifting scalp distribution. Several studies report that amplitude differences among the midline electrodes decrease with age. P300 amplitude at Fz increases, while it decreases at Pz (Pfefferbaum, Ford, Roth, & Kopell, 1980). There is some evidence that this may be a result of Slow Wave changes with age, and not change in P300 (Pfefferbaum et al., 1980).

ACKNOWLEDGMENTS

The reliability study reported here was performed at the Cognitive Psychophysiology Laboratory, and was supported by the Air Force (USAFSAM/NG, Brooks AFB, Contract #F33615-82-K-0611; scientific monitor, Dr. Bryce Hartman), and by the Air Force Office of Scientific Research, Contract F49620-79-C-0233. We are indebted to Pat Ackles, Dick Jennings, and Art Kramer for helpful comments on an earlier draft.

FOOTNOTES

1. The contributions of the first three authors were equal; they are listed in alphabetical order.

2. ERP components are subdivided into two broad categories: exogenous and endogenous. Exogenous components are early components (0-100 ms after stimulus presentation), and are responsive to the physical characteristic of the eliciting stimuli (e.g., modality, intensity, etc.). Endogenous components (including P300) are late components occurring 100 ms to several seconds after stimulus presentation. They are related to psychological processes, and are independent of the physical characteristics of the eliciting stimuli. In fact, the absence of an expected stimulus can elicit these components. See also Donchin et al., 1978, and Donchin, Karis, Bashore, Coles, & Gratton, in press, for a discussion of the endogenous-exogenous distinction.

3. Scalp distribution of a component is the amplitude variability displayed by that component at different electrode locations.

4. For reason of space, we will not provide details about the procedures used in the various operational definitions described in this section. We invite the interested reader to consult the original papers to obtain this information.

5. Pattern recognition techniques are procedures aimed at detecting signals characterized by a particular pattern, and extracting them from background noise. Linear filters can be considered instances of pattern recognition techniques.

6. These were primarily trials where saturation of the A/D converter

occurred.

7. Note that since the baseline was subtracted from each record, the baseline was always zero.

Table 1

Examples of Common ERP Paradigms

Dimensions of the Taxonomy	Oddball	Selective Attention	Word Lists (To Remember or Judge)	Priming and N400 Paradigms	Signal Detection	Recognition
1. Sequence generating rule						
A. random or constrained	random	random	random	constrained	random	random
B. probability	rare/frequent	2 rare, 2 frequent	rares in some experiments	equiprobable	equiprobable	equiprobable
2. Trials						
A. number of stimuli and temporal relationships						
1. *number of stimuli per trial*	1	1	1	*2 or more*	*2 or more*	1
2. among trials (sec)	1-3	< 1	2	1 to > 5	3 to > 10	2-3
B. stimuli						
1. *modality*	*aud & vis*	*auditory*	*visual*	*visual*	*auditory*	*visual*
2. physical properties	easily discriminable or readable	easily discriminable or readable	easily discriminable or readable	easily discriminable or readable	low intensity; detection difficult	easily disc.
3. *complexity, structure, and content*	usually simple (wide variety of stimuli)	simple (tones)	moderately complex (words)	relationships present among stimuli (primarily words)	simple (tones)	moderately complex (words)

Table 1 (continued)

Examples of Common ERP Paradigms

3. Mapping rule						
A. memory load	low	low	usually low	low	low	high
B. nature of processing and transformation required	usually simple	selective attention	complex	priming in semantic memory	simple	memory, search & comparison
C. complexity of response selection	choice RT or go/nogo	go/nogo	choice RT or nothing	choice RT	choice RT	choice RT
4. Response execution						
A. covert or overt	either	usually covert	either	overt	overt	overt
B. discrete or continuous	--discrete--					
C. ballistic, or involving feedback	ballistic	no response	ballistic	ballistic	ballistic	ballistic

Table 2

Experimental Variables that Affect P300

Dimensions of the Taxonomy	P300 Amplitude	P300 Latency	P300 Scalp Distribution
1. Sequence generating rule			
A. random or constrained	constrained smaller	--	?
B. probability	5-15 uv effect (rare larger)	50-80 ms difference (rare longer)	no difference
2. Trials			
A. number of stimuli and temporal relationships			
1. number of stimuli per trial	--	--	Cz is more negative when there is a warning (and thus a CNV)
2. among trials	probability disappears at long ISIs (>3sec)	--	--
B. stimuli			
1. modality	visual larger (0-5 uv)	visual longer (100ms)*	Cz usually more positive for vis?
2. physical properties	--	50-150 ms longer with degraded stimuli or in noise conditions	Intensity may have an effect; Cz may be = or > Pz at low levels
3. complexity, structure and content	smaller for complex, but confounded with latency	longer for more complex stimuli (may exceed 100 ms)	--
3. Mapping rule			
A. memory load requirements	--	--	--

Table 2 (continued)

Experimental Variables that Affect P300

B. nature of processing and transformation required	selective attention required; positive (old) items 5 uv or more larger in recognition or Sternberg paradigms; reduction (5-10 uv) with increase in workload (in dual tasks) 0-5 uv memory effect	memory search (slope 0-25 ms)	may be changes due to feedback depending on their evaluation; dual tasks - Pz decreases relative to Cz
C. complexity of response selection	--	--	Increase in Cz on nogo trials?
4. Response execution			
A. covert or overt	--	--	--
B. discrete or continuous	--	--	--
C. ballistic, or involving feedback	--	--	--
5. Consequences	0-5 uv increase with bonus for correct guess/ prediction	--	--
6. Subject characteristics	decrease with age from maturity	increases with age from maturity (1-2 ms/year)	more frontal with age? smaller differences among Fz, Cz, Pz?

*In most studies visual stimuli are more complex than auditory stimuli, and there is a strong effect of complexity on P300 latency (see 2.B.3).

Table 3

Comparison Among Operational Definitions of P300

Procedure	Data Preparation	Template	Time Shift
Area Measure	no	square wave	no
Peak-Picking	no	single point	yes
PCA *	subtraction of grand-average waveform	satisfaction of maximum variance criterion (+ Varimax rotation)	no
SWDA	standardization for each time point	satisfaction of maximum discrimination criterion	no
Cross-correlation	standardization for each ERP segment	free	yes
Woody Filter	standardization for each ERP segment	ERP average after preceding iteration	yes

* as proposed by Donchin (1966).

Table 4a

Number of Studies Using Various Techniques for P300 Measurement (N = 34)

Base-to-Peak only (or Feak for latency)	15
Peak-to-Peak only	1
Area only	4
Discriminant Score only	2
PCA only*	6
Base-to-Peak plus PCA	4
Area plus PCA	1
Area plus Base-to-Peak	1

* one of these (Ruchkin, Munson, & Sutton, 1982) also measured latency by "computing the mean of the latencies of the half-of-peak amplitude points on the rising and falling edges of the component peak (Tukey, 1978, p. 143)" (p. 632).

Table 4b

Number (and Percentage) of Studies with Particular Placements ($N = 34$)

Cz only:	6	(18%)
Fz, Cz, Pz only:	7	(21%)
Fz, Cz, Pz + others*:	14	(41%)
Other**:	7	(21%)
Cz included?	33	(97%)
Fz, Cz, Pz included?	21	(62%)
Lateral Placements?	8	(24%)

* six of these consisted of Fz, Cz, Pz, Oz

** these often consisted of Cz plus lateral placements

Table 5

Summary of the Procedures Used in the Reliability Study

	Between session	Within session
P300 amplitude	Area measure at Pz	Area measure at Pz
	Area measure on VF*	Area measure on VF*
	Base-to-peak at Pz	Base-to-peak at Pz
	Base-to-peak on VF*	Base-to-peak on VF*
	CC-amplitude at Pz	CC-amplitude at Pz
	CC-amplitude on VF*	CC-amplitude on VF*
	CC-covariance at Pz	CC-covariance at Pz
	CC-covariance on VF*	CC-covariance on VF*
	PCA comp. scores for Fz	
	PCA comp. scores for Cz	
	PCA comp. scores for Pz	
	Likelihood	
P300 latency	Peak-picking at Pz	Peak-picking at Pz
	Peak-picking on VF*	Peak-picking on VF*
	Cross-correl. at Pz	Cross-correl. at Pz
	Cross-correl. on VF*	Cross-correl. on VF*

Likelihood

---

* waveforms obtained by combining data from different electrodes with Vector filter.

Table 6

Within Session Reliabilities of Average Parameters

		Medians							
Parameter	Range	All	Aud.	H/S	Name	Rare	Freq.	Targ.	NTarg.
Amplitude: Base-to-peak									
at PZ	.50-.92	.85	.92	.87	.66	.75	.90	.89	.74
on VF	.57-.92	.88	.90	.89	.71	.68	.89	.90	.74
Amplitude: Area									
at Pz	.35-.92	.80	.91	.81	.62	.75	.82	.81	.78
on VF	.39-.89	.80	.88	.77	.66	.69	.80	.82	.74
Amplitude: CC-Amplitude									
at Pz	.52-.91	.84	.91	.86	.67	.75	.88	.87	.73
on VF	.58-.92	.83	.89	.84	.63	.60	.84	.86	.62
Amplitude: CC-Covariance									
at Pz	.29-.92	.85	.89	.90	.61	.77	.90	.88	.70
on VF	.23-.93	.86	.90	.90	.71	.80	.91	.88	.77
Latency: Peak-Picking									
at Pz	.35-.97	.67	.80	.58	.62	.48	.71	.67	.65
on VF	.48-.89	.70	.69	.74	.58	.67	.73	.69	.70
Latency: Cross-correlation									
at Pz	-.18-.88	.57	.79	.66	.48	.72	.57	.64	.57

on VF	.41-.84	.61	.80	.53	.61	.54	.67	.70	.54

Table 7

Within Session Reliabilities of Averaged Single Trials

Parameters

		Medians							
Parameter	Range	All	Aud.	H/S	Name	Rare	Freq.	Targ.	NTarg.
Amplitude: Base-to-peak									
at PZ	.56-.94	.88	.93	.91	.73	.80	.91	.91	.81
on VF	.54-.93	.85	.92	.87	.72	.73	.86	.88	.77
Amplitude: Area									
at Pz	.39-.92	.79	.92	.80	.62	.72	.80	.80	.75
on VF	.46-.90	.80	.88	.79	.65	.71	.80	.82	.75
Amplitude: CC-Amplitude									
at Pz	.36-.90	.73	.85	.70	.59	.61	.78	.75	.69
on VF	.31-.84	.76	.79	.75	.65	.56	.81	.77	.65
Amplitude: CC-Covariance									
at Pz	.65-.97	.92	.94	.94	.83	.89	.96	.92	.90
on VF	.77-.96	.92	.94	.95	.87	.89	.96	.93	.90
Latency: Peak-Picking									
at Pz	.64-.91	.81	.82	.81	.74	.71	.88	.81	.79
on VF	.57-.90	.83	.82	.84	.79	.72	.88	.83	.79
Latency: Cross-correlation									
at Pz	.39-.86	.82	.82	.82	.62	.66	.82	.82	.71

on VF	.57-.90	.82	.83	.82	.73	.67	.86	.82	.77

Table 8

Within Session Reliabilities of the Probability Effect

Parameter	Difference scores on single trial estimates	Difference scores on average estimates	Estimates on difference waveforms
Amplitude: Base-to-peak			
at Pz	.36	.33	.61
on VF	.44	.26	.50
Amplitude: Area			
at Pz	.20	.21	.35
on VF	.33	.26	.55
Amplitude: CC-Amplitude			
at Pz	.11	.31	.52
on VF	-.15	.35	.59
Amplitude: CC-Covariance			
at Pz	.43	.55	.63
on VF	.52	.55	.64

Table 9

Between Session Reliabilities

		Medians							
Parameter	Range	All	Aud.	H/S	Name	Rare	Freq.	Targ.	NTarg.
Amplitude: Base-to-peak									
at PZ	.51-.89	.72	.81	.68	.77	.70	.79	.72	.72
on VF	.53-.89	.79	.81	.74	.75	.66	.83	.79	.76
Amplitude: Area									
at Pz	.50-.90	.68	.80	.64	.65	.69	.61	.70	.65
on VF	.53-.90	.71	.80	.63	.79	.69	.76	.71	.71
Amplitude: CC-Amplitude									
at Pz	.52-.86	.74	.81	.68	.73	.72	.74	.75	.63
on VF	.51-.90	.77	.82	.72	.74	.70	.78	.77	.67
Amplitude: CC-Covariance									
at Pz	.20-.91	.70	.77	.77	.40	.62	.79	.74	.49
on VF	.37-.90	.65	.70	.76	.53	.53	.80	.72	.53
Amplitude: PCA									
at Fz	.52-.89	.71	.71	.81	.62	.70	.73	.71	.72
at Cz	.66-.87	.76	.79	.73	.78	.73	.81	.75	.78
at Pz	.54-.91	.83	.78	.78	.84	.69	.84	.83	.78
Amplitude: Likelihood									
	.57-.90	.67	.79	.61	.76	.67	.70	.66	.68
Latency: Peak-Picking									
at Pz	.43-.84	.56	.48	.75	.54	.52	.69	.56	.58
on VF	-.15-.89	.47	.50	.23	.50	.50	.19	.33	.50

Latency: Cross-correlation									
at Pz	-.07-.99	.34	.21	.55	.39	.34	.33	.38	.34
on VF	-.13-.95	.63	.57	.79	.59	.62	.63	.69	.59
Latency: Likelihood									
	.23-.68	.43	.59	.38	.20	.45	.24	.48	.20

Table 10

Between Session Reliability for Latency Estimates

(Outliers Excluded)

		Medians							
Parameter	Range	All	Aud.	H/S	Name	Rare	Freq.	Targ.	NTarg.
Latency: Peak-Picking									
at Pz	.43-.84	.51	.48	.75	.48	.52	.49	.51	.58
on VF	.20-.89	.57	.83	.55	.50	.50	.64	.72	.50
Latency: Cross-correlation									
at Pz	.21-.82	.47	.46	.72	.39	.34	.61	.53	.41
on VF	-.13-.95	.70	.78	.79	.59	.62	.83	.80	.59

Table 11

Summary of Reliability Results

Parameter	Within session		Between session
	Average	Single trials	Average
Amplitude: Base-to-peak			
at Pz	.85	.88	.72
on VF	.88	.85	.79
Amplitude: Area			
at Pz	.80	.79	.68
on VF	.80	.80	.71
Amplitude: CC-amplitude			
at Pz	.84	.73	.74
on VF	.83	.76	.77
Amplitude: CC-covariance			
at Pz	.85	.92	.70
on VF	.86	.92	.65
Amplitude: PCA			
at Fz			.71
at Cz			.76
at Pz			.83
Amplitude: Likelihood			.67
Latency: Peak-picking			
at Pz	.67	.81	.56

on VF	.70	.83	.47
Latency: Cross correlation			
at Pz	.57	.82	.34
on VF	.61	.82	.63

FIGURE CAPTIONS

Figure 1. Graphic representation of the vectors of weights corresponding to several operational definitions of P300. The arrows indicate the possibility of shifting the vector of weights. In these cases P300 latency can be measured.

Figure 2. Average Pz waveforms for rare (20%) and frequent (80%) stimuli from the first session of one subject (#1) in the five different oddball tasks, (a) auditory RT task involving pitch discrimination, (b) visual count rare task involving letter discrimination, (c) visual count frequent task involving letter discrimination, (d) visual count rare task involving name gender discrimination, and (e) visual count frequent task involving name gender discrimination.

Figure 3. Grandaverage Pz waveforms (n=49) for rare (20%) and frequent (80%) stimuli from the first session in the five different oddball tasks, (a) auditory RT task involving pitch discrimination, (b) visual count rare task involving letter discrimination, (c) visual count frequent task involving letter discrimination, (d) visual count rare task involving name gender discrimination, and (e) visual count frequent task involving name gender discrimination.

Figure 4. Average Pz waveforms of each individual subject in the letter discrimination (H-S) task for rare (20%, solid line) and frequent (80%, dashed line) stimuli from the first session. Waveforms from the count rare condition are presented in figure 4a, and waveforms for the count frequent condition in figure 4b.

Figure 5. Examples of average Pz waveforms for rare and frequent

stimuli for three different subjects and two sessions. The first session is represented by solid lines, the second by dashed lines.

Figure 6. Reliability study: component loadings for the PCA's run separately for each session (left, session 1 - right, session 2). Component loadings for the Auditory oddball are at the top, the H/S oddball is in the middle, and the Name oddball is at the bottom.

Figure 7. Reliability study: component loadings for the PCA's run on both sessions. Component loadings for the Auditory oddball are at the top, the H/S oddball is in the middle, and the Name oddball is at the bottom.

Figure 8. Reliability study: "time weights" used for the "simulated visual inspection" of the waveforms.

Figure 9. Reliability study: reliability of four P300 amplitude measures (base-to-peak on Vector filtered data, base-to-peak at Pz, CC-covariance on Vector filtered data, CC-covariance at Pz) as a function of the number of trials used for the analysis. The measures were taken on single trials and then averaged. Only frequent trials were used for the analysis.

Figure 10. Reliability study: reliability of four P300 latency measures (peak-picking on Vector filtered data, peak-picking at Pz, cross-correlation on Vector filtered data, cross-correlation at Pz) as a function of the number of trials used for the analysis. The measures were taken on single trials and then averaged. Only frequent trials were used for the analysis.

Figure 11. Examples of scatter plots of P300 parameters (t-scores) measured at the first and second session on 17 subjects. The plots refer to data obtained in the H/S oddball, counted frequent condition. The upper

left plot corresponds to estimates of P300 amplitude obtained with CC-amplitude on Vector filtered waveforms. The upper right plot corresponds to estimates of P300 latency obtained with peak-picking on Vector filtered waveforms. The lower left plot corresponds to estimates of P300 latency obtained with cross-correlation at Pz. The lower right corresponds to estimates of P300 amplitude obtained with base-to-peak at Pz. Between session reliability values are reported below the plots. The presence of outliers and their effect on reliability is clearly visible for the latency measures.

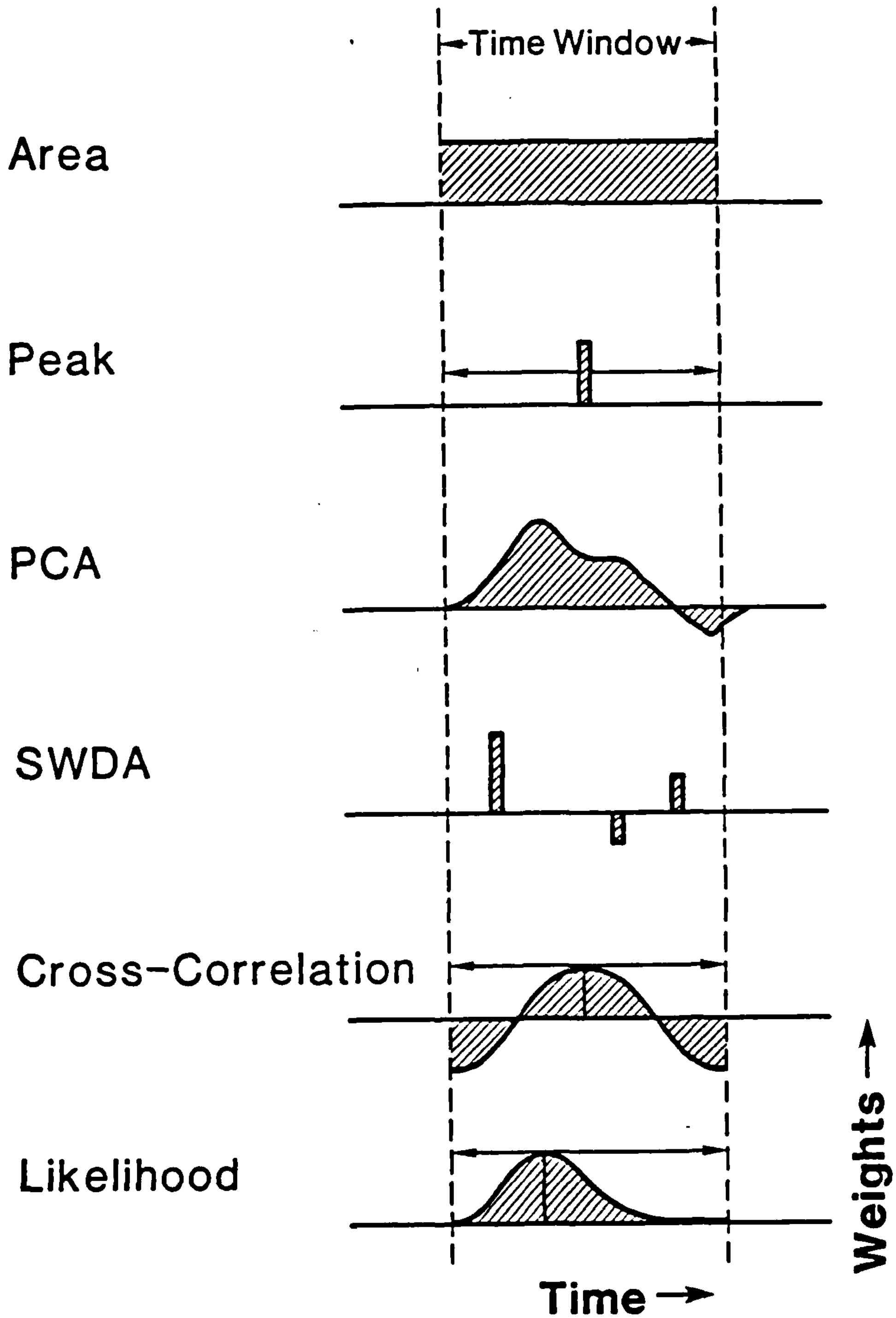

Time Window
Area
Peak
PCA
SWDA
Cross-Correlation
Likelihood
Time →
Weights →

FIG 1

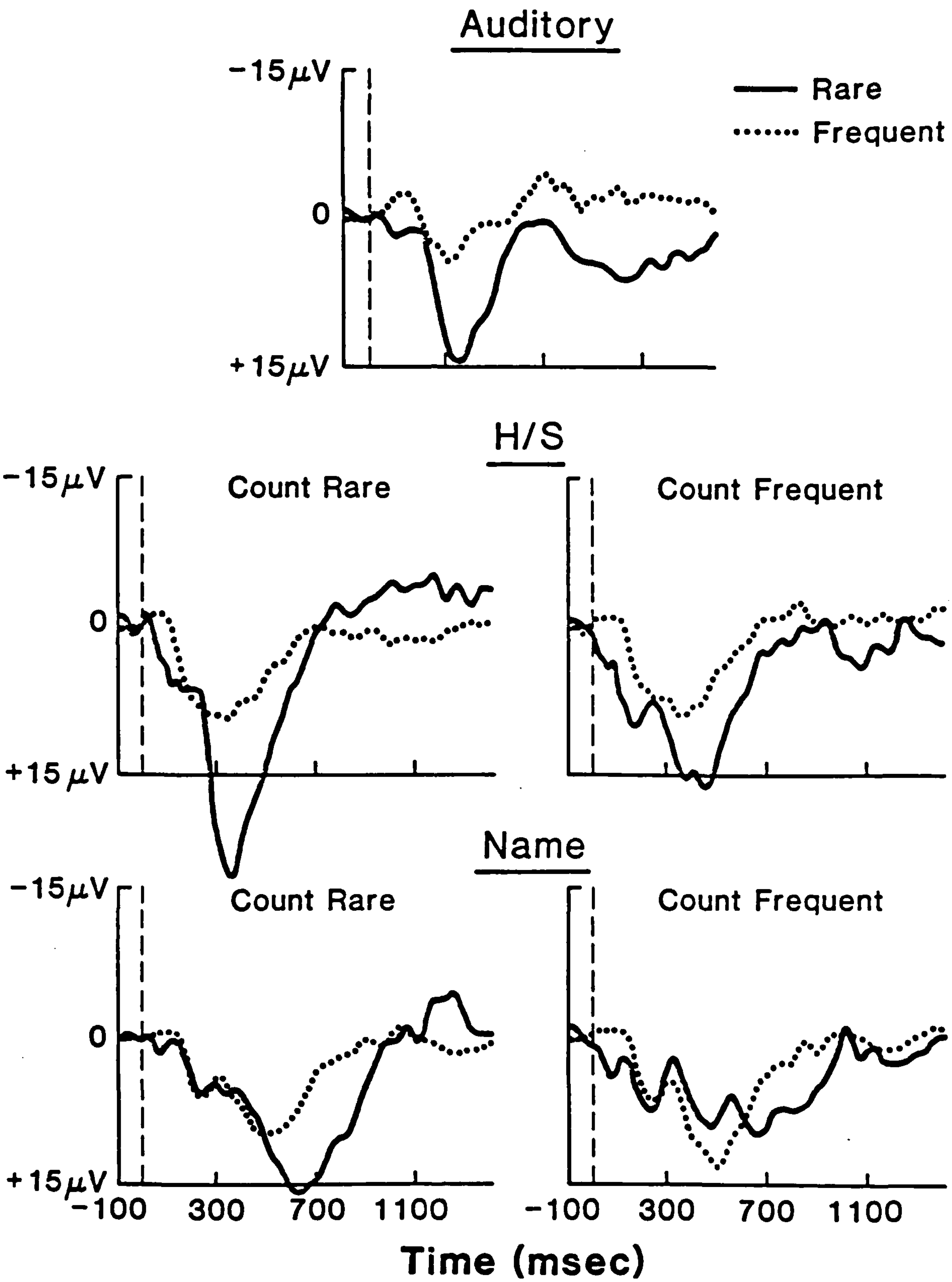
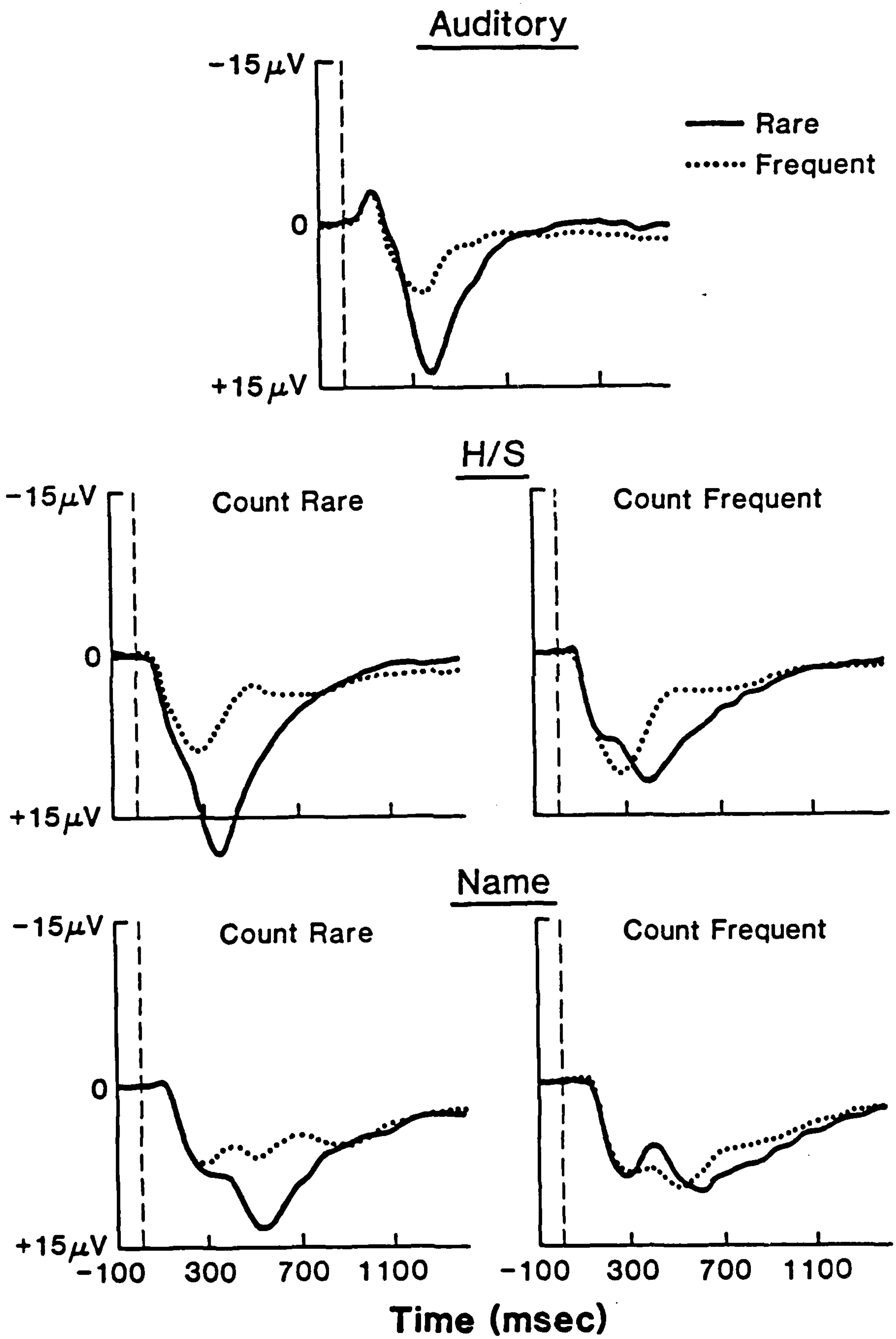

Auditory
−15µV
0
+15µV
Rare
Frequent
H/S
Count Rare
Count Frequent
Name
Count Rare
Count Frequent
−100
300
700
1100
Time (msec)

FIG 2



Auditory
−15μV
0
+15μV
Rare
Frequent
H/S
Count Rare
Count Frequent
Name
Count Rare
Count Frequent
−100
300
700
1100
Time (msec)

FIG 3

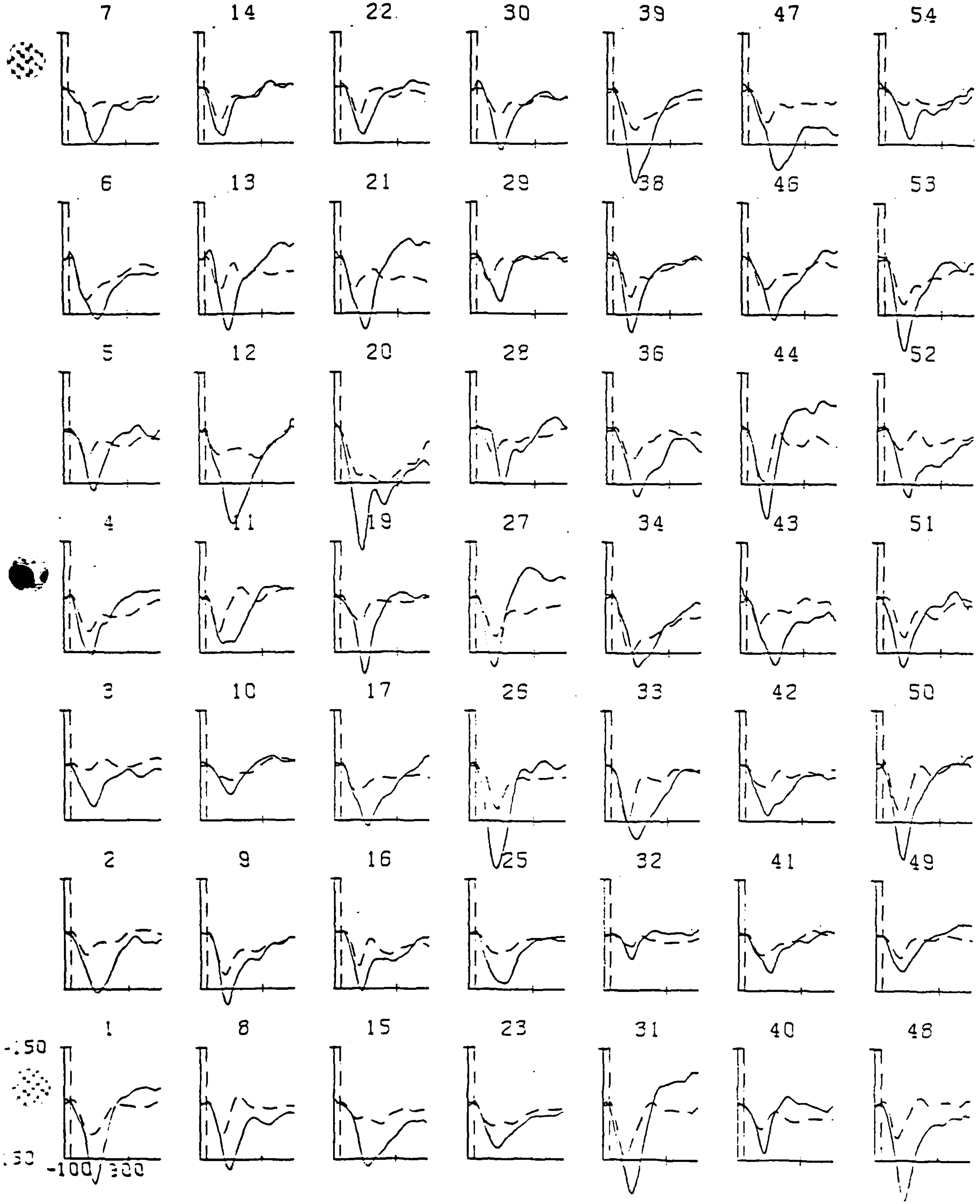

7
14
22
30
39
47
54
6
13
21
29
38
46
53
5
12
20
28
36
44
52
4
11
19
27
34
43
51
3
10
17
26
33
42
50
2
9
16
25
32
41
49
1
8
15
23
31
40
48
-100/300

FIG 4a

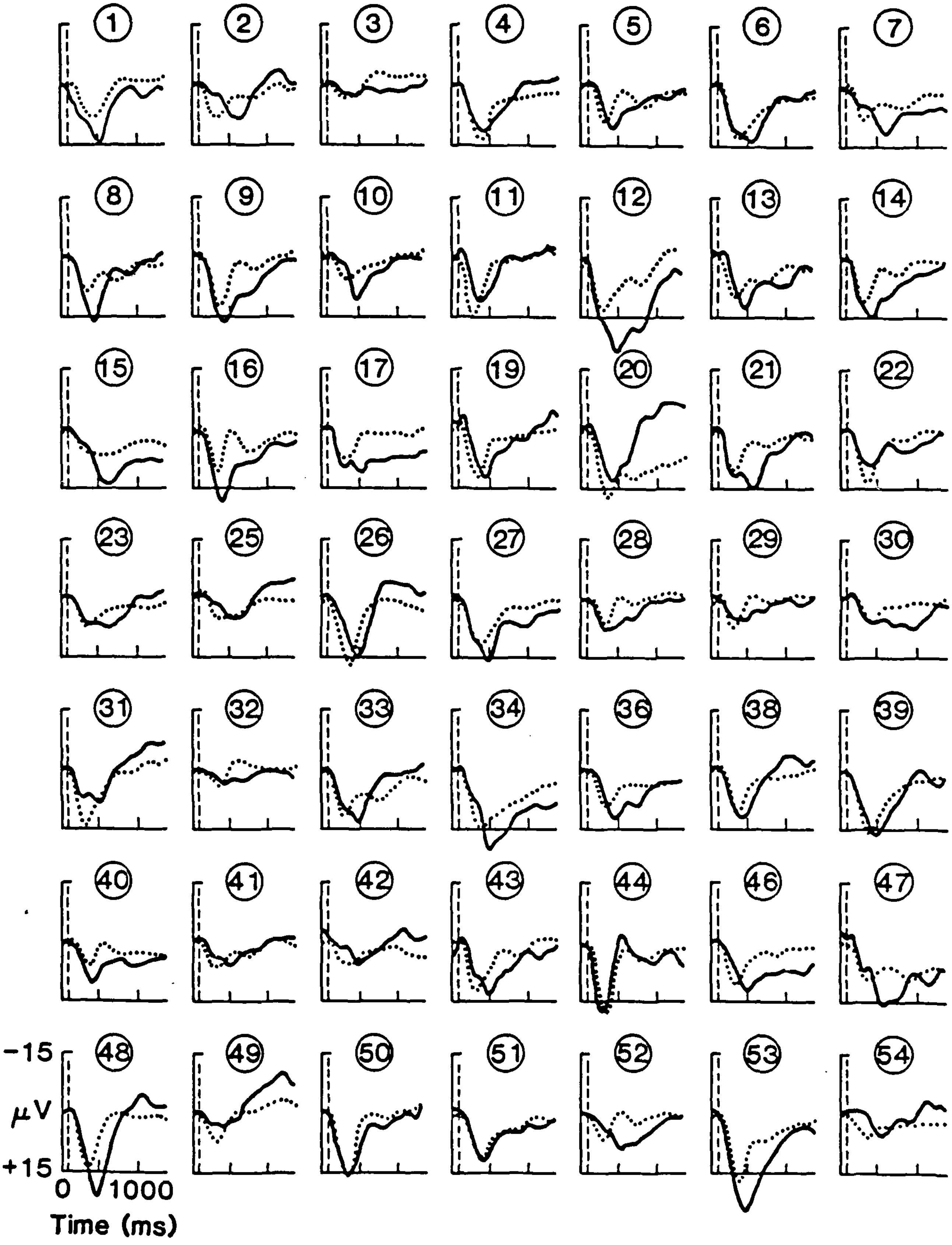
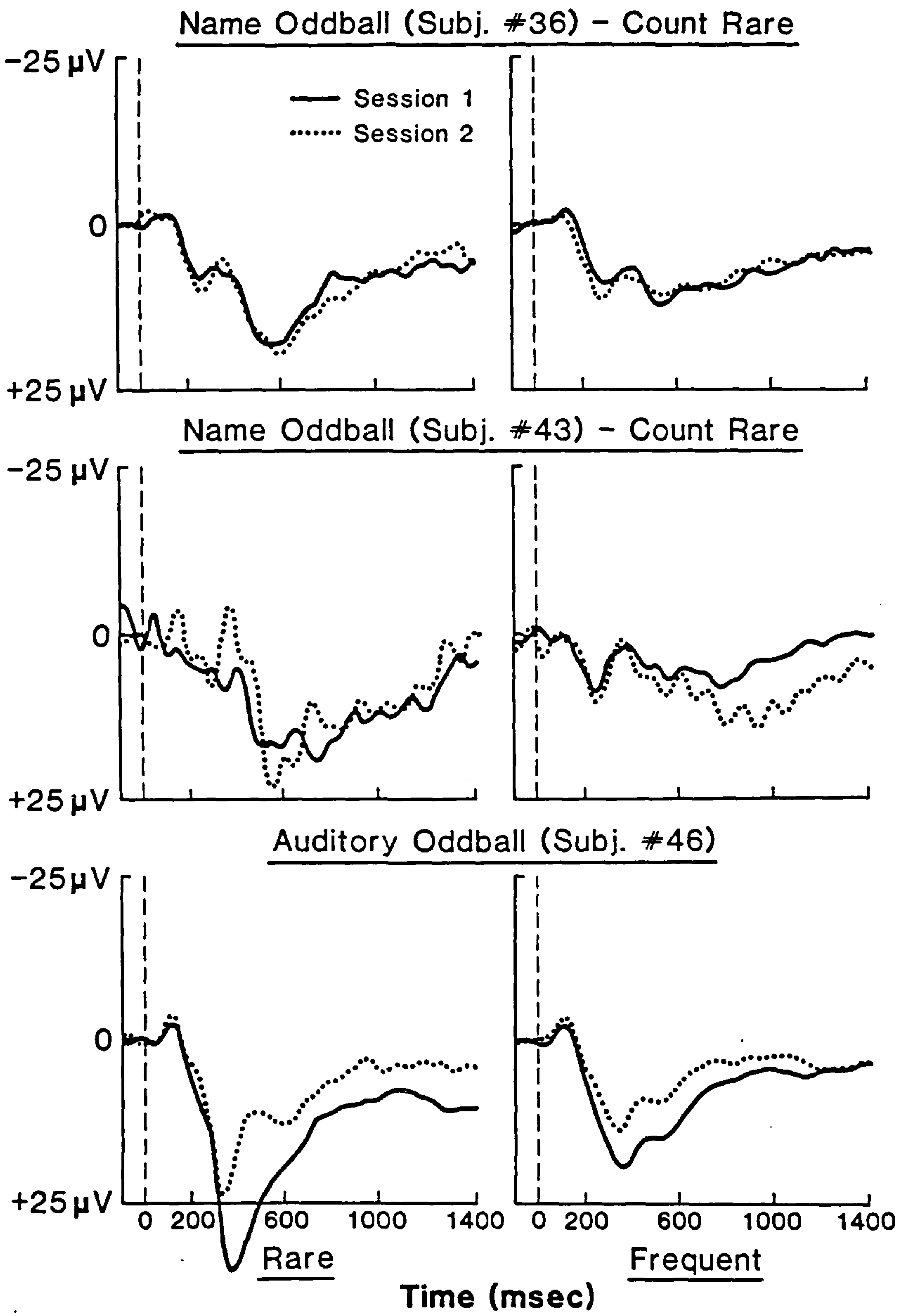

1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 19 20 21 22
23 25 26 27 28 29 30
31 32 33 34 36 38 39
40 41 42 43 44 46 47
48 49 50 51 52 53 54
−15
μV
+15
0
1000
Time (ms)

Fig 4b



Name Oddball (Subj. #36) - Count Rare
Session 1
Session 2
Name Oddball (Subj. #43) - Count Rare
Auditory Oddball (Subj. #46)
-25 µV
0
+25 µV
0 200 600 1000 1400
Rare
Frequent
Time (msec)

Fig 5

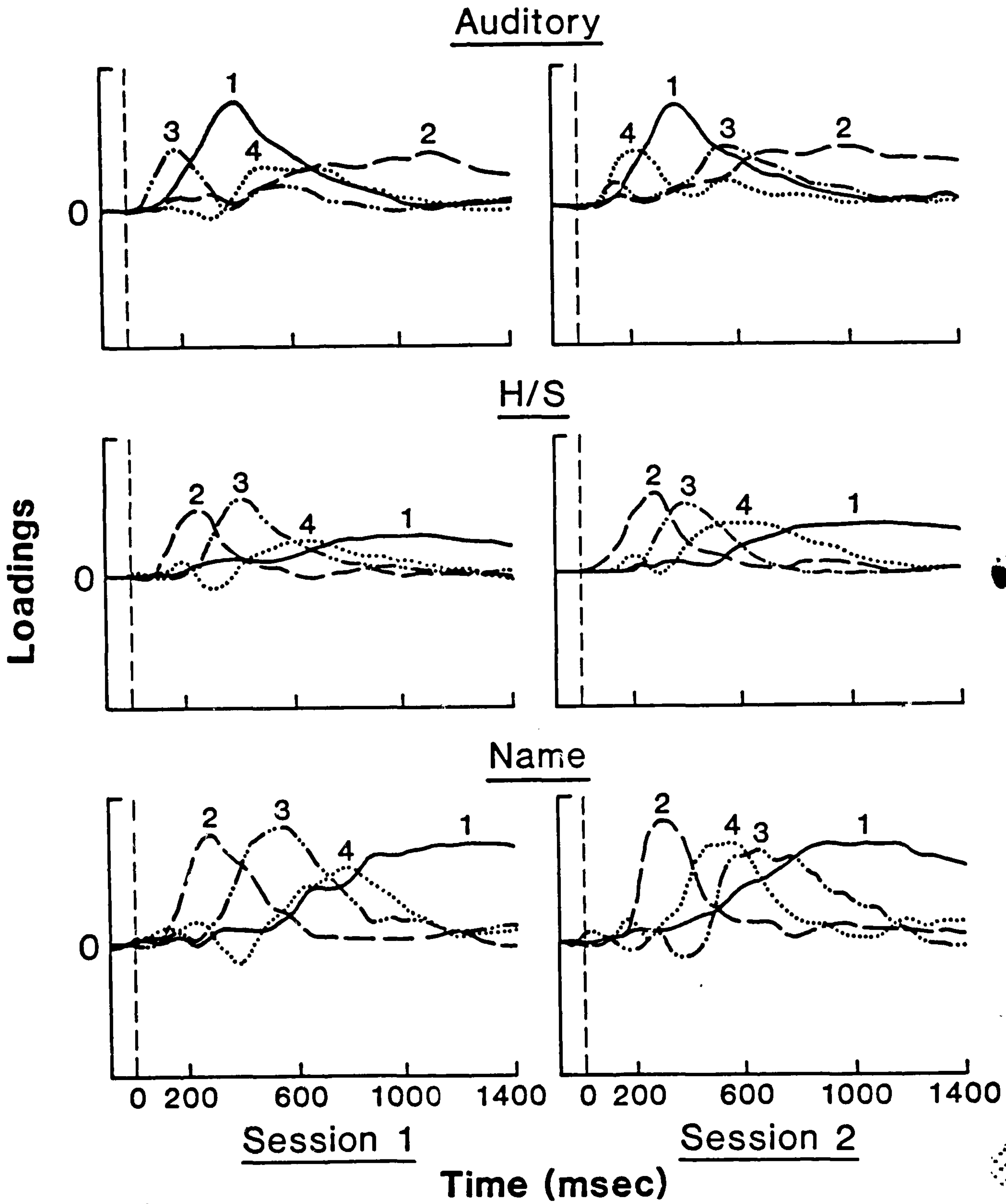

Auditory
H/S
Name
Loadings
0
1
2
3
4
0 200
600
1000
1400
Session 1
Session 2
Time (msec)

Fig 6

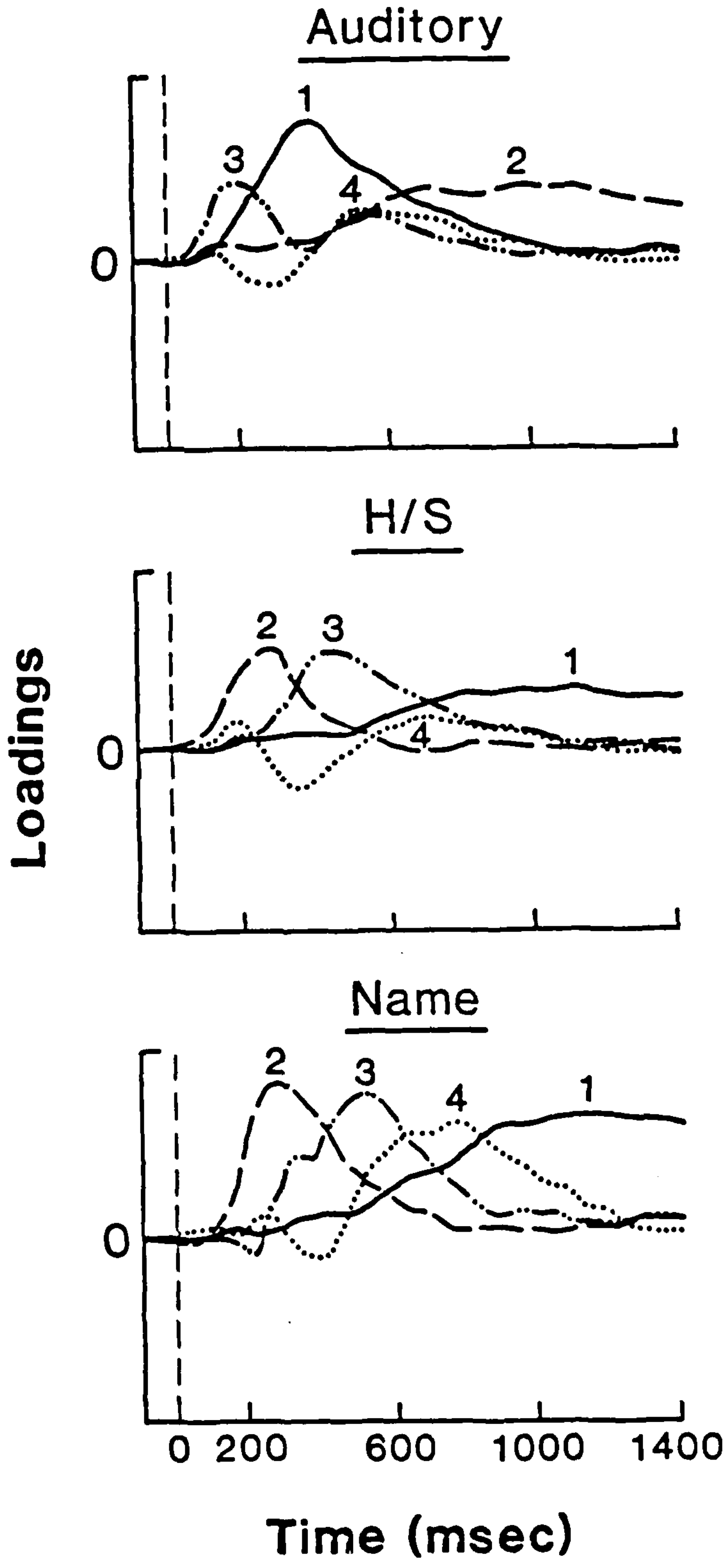

Auditory
H/S
Name
Loadings
0
1
2
3
4
0 200 600 1000 1400
Time (msec)

Fig 7

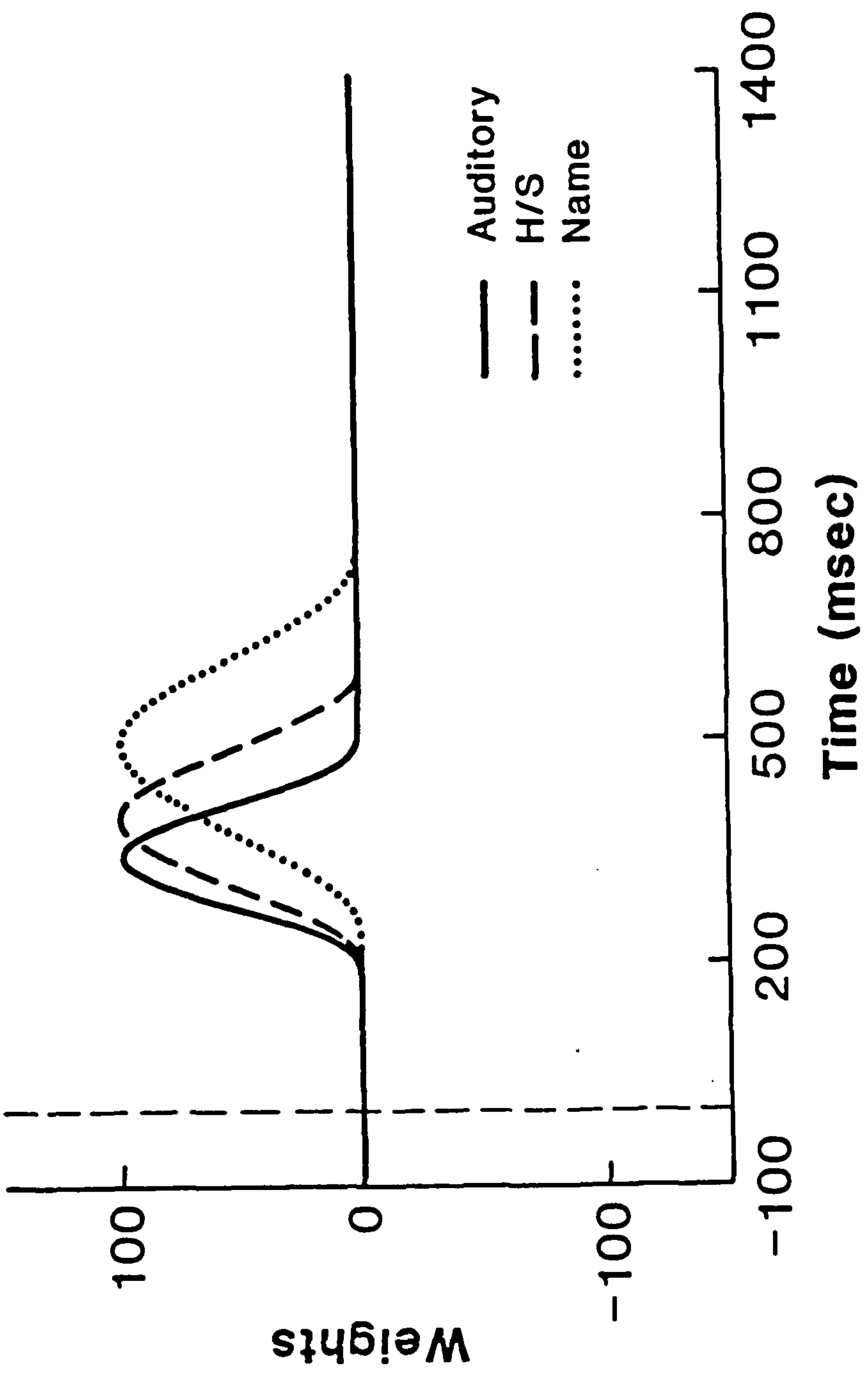

Weights
100
0
−100
−100
200
500
800
1100
1400
Time (msec)
Auditory
H/S
Name

Fig 8

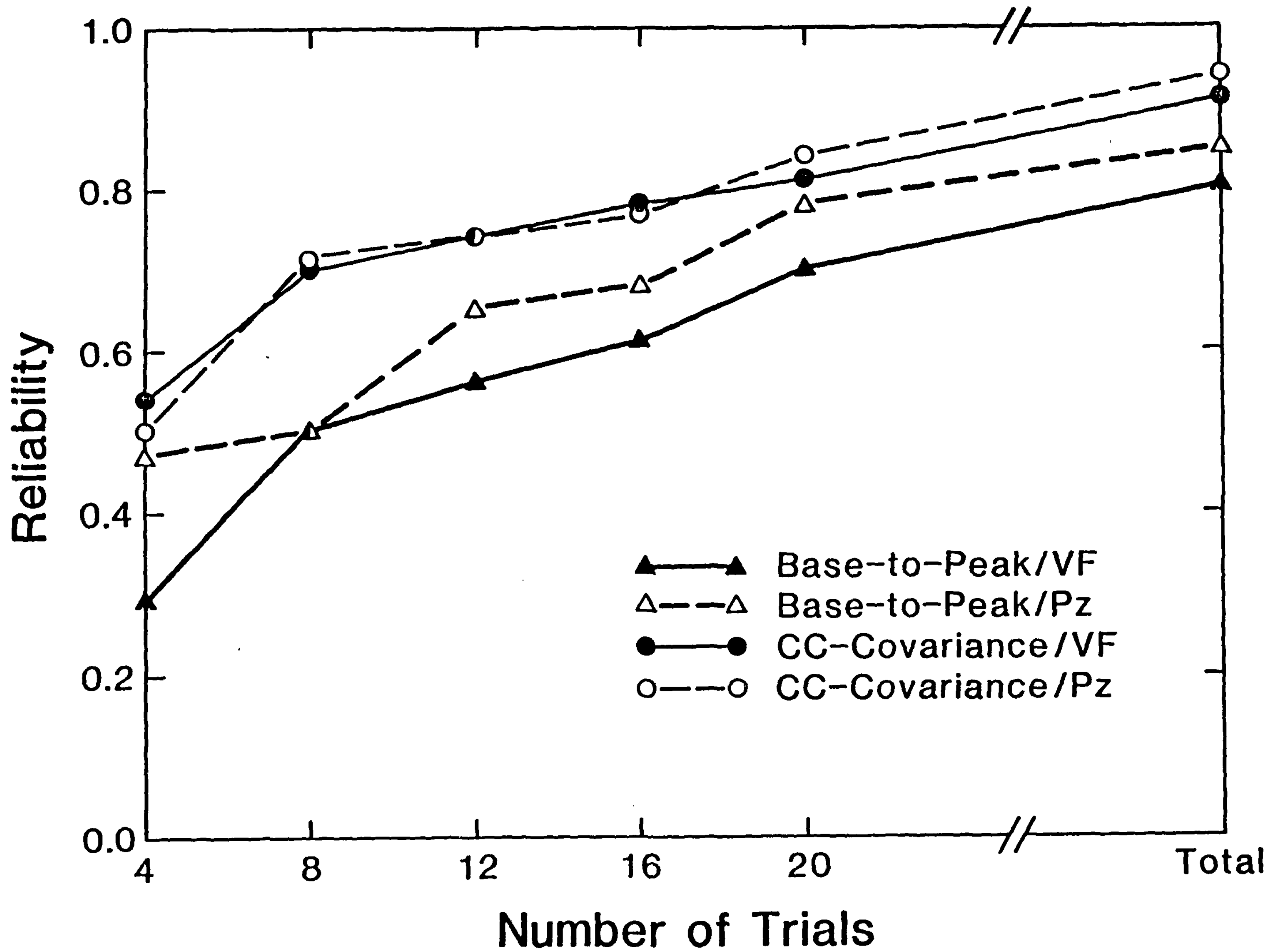

1.0
0.8
0.6
0.4
0.2
0.0
Reliability
4
8
12
16
20
Total
Number of Trials
Base-to-Peak/VF
Base-to-Peak/Pz
CC-Covariance/VF
CC-Covariance/Pz

Fig 9

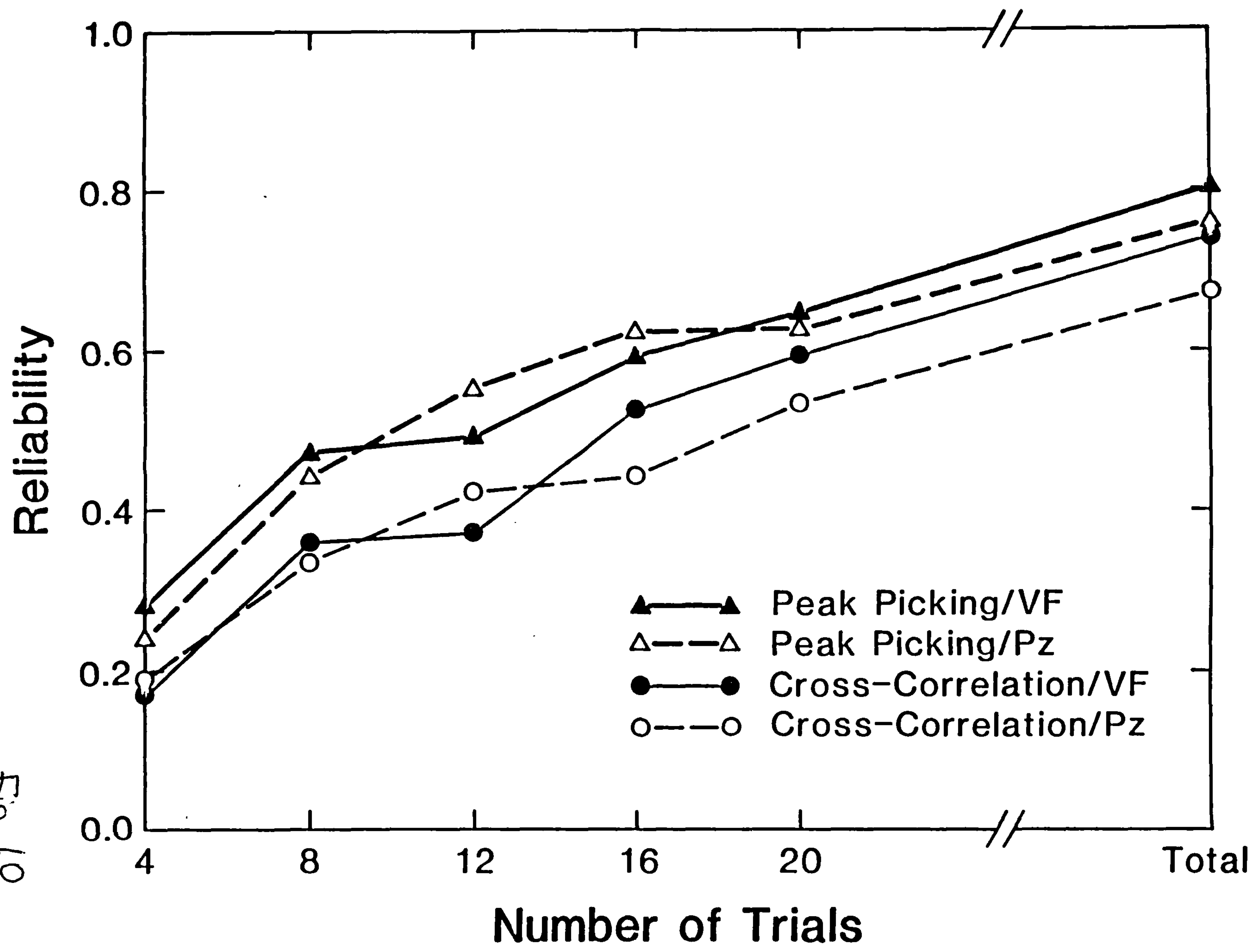

Reliability
1.0
0.8
0.6
0.4
0.2
0.0
4
8
12
16
20
Total
Number of Trials
Peak Picking/VF
Peak Picking/Pz
Cross-Correlation/VF
Cross-Correlation/Pz

Fig 10

H/S Task: Frequent Target

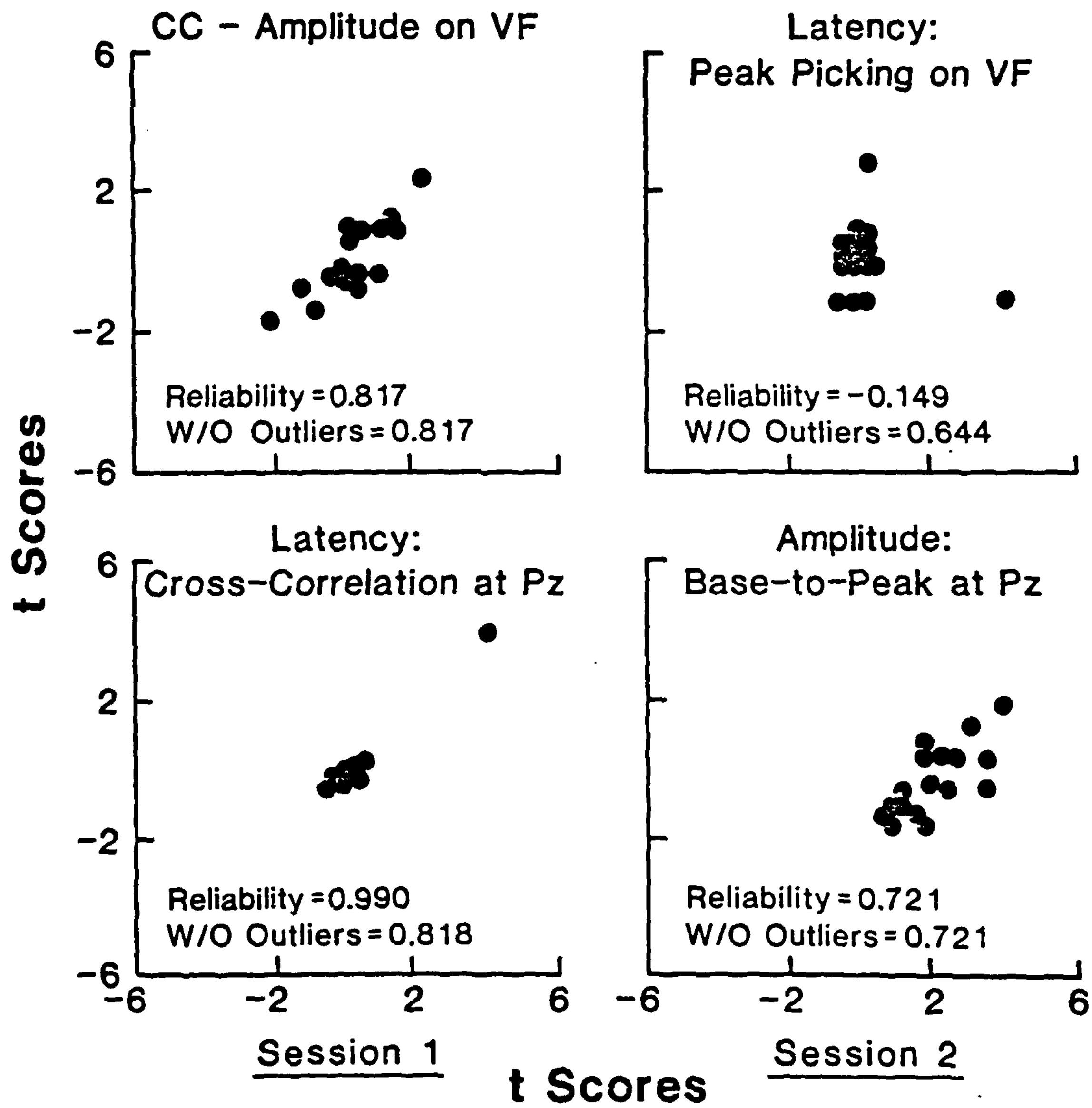


Fig 11

P300 and Recall in an Incidental Memory Paradigm

Monica Fabiani, Demetrios Karis,
and Emanuel Donchin

Cognitive Psychophysiology Laboratory
University of Illinois
Champaign, Illinois

A partial report of this study was presented at the EPIC VII, Florence (Italy), September 4-9, 1983, and at the 23rd annual meeting of the Society for Psychophysiological Research, Asilomar (CA), September 25-28, 1983. The research described was supported by the Air Force Office of Scientific Research, contract #F33615/82/K/0611. The support of Al Fregly is appreciated. D.K. is currently with the Grumman Aerospace Corporation, MS C04-14, Bethpage, NY, 11714. We wish to acknowledge the help of Ron Chambers, Kay Strayer, Dan Tepper and Neil Walch in the collection of the data. We are also indebted to Gabriele Gratton for helpful comments and statistical advice, and to Mike Coles, Art Kramer and Dave Strayer for reviewing an earlier version of this manuscript.

Address requests for reprints to: E. Donchin, University of Illinois, Psychology Department, 603 East Daniel, Champaign, Illinois, 61820.

Running head: P300 and Memory

Abstract

In previous research we found a relationship between the amplitude of the P300s elicited by words and subsequent recall performance (Karis, Fabiani, & Donchin, 1984). Words later recalled elicited larger P300s than words later not recalled. However, this relationship was dependent on the mnemonic strategies used by the subjects. There was a strong relationship between P300 amplitude and recall when rote strategies were used, but when subjects used elaborative strategies the relationship between P300 amplitude and recall was not evident. In the present experiment we employed an incidental memory paradigm to reduce the use of mnemonic strategies. An "oddball" task consisting of a series of names was presented, and subjects were required to count either the male or the female names. Event-related brain potentials were recorded to the presentation of each name. Following the oddball task, subjects were asked, unexpectedly, to recall as many names as possible. The names that were recalled had elicited, on their initial presentation, larger P300s than names not recalled. Thus, these results confirm our hypothesis: when elaborative strategies are not used, the relationship between P300 and memory emerges more consistently. Our data provide support for a "context updating" hypothesis of the functional significance of the P300.

DESCRIPTORS: P300, incidental memory, oddball paradigm.

P300 and Recall in an Incidental Memory Paradigm

Monica Fabiani, Demetrios Karis,
and Emanuel Donchin

Introduction

Karis, Fabiani and Donchin (1984) compared the Event Related Potential (ERP) elicited by words subsequently recalled or not recalled. They reported that, in subjects who used rote mnemonic strategies, the P300 component elicited by words subsequently recalled was larger than that elicited by words that were not recalled. However, this relationship between P300 amplitude and recall was not observed in subjects who used elaborative mnemonic strategies. This paper confirms the existence of a relationship between P300 amplitude and subsequent recall. We examined this relationship under conditions in which subjects were unlikely to use elaborative strategies.

The P300 component of the ERP is a positive-going component first described by Sutton, Braren, Zubin, and John (1965). It peaks 300 ms, or more, after the eliciting event and is maximal at the parietal electrode (Pz - International 10-20 System, Jasper, 1958). Considerable information has accumulated on the antecedent conditions required to elicit the P300 (for a review, see Pritchard, 1981). We assume that ERP components, such as the P300, are manifestations at the scalp of the activity of specific intracranial processors. These need not be specific neuroanatomical entities, but rather processors that represent the activation of internal "subroutines" (Donchin, 1979; 1981). When a P300 is generated in response to some event, we assume that, somewhere in the brain, ensembles of neurons are activated synchronously as they participate in an activity that implements some specific information processing transaction. Several lines of evidence converge to support the view that P300 is elicited whenever there is a need to revise the internal model of the environment (Donchin, 1981; see also Donchin, Coles, & Gratton, 1984).

Two of the factors known to control P300 amplitude are the subjective probability and the task relevance of the eliciting event (Duncan-Johnson & Donchin, 1977; Johnson & Donchin, 1978). The dependence of P300 amplitude on subjective probability suggests that the process manifested by P300 is involved when novelty is encountered. The fact that this process is elicited only if the information is relevant and usable by the subject also supports this view. Johnson and Donchin (1982), for example, demonstrated that the amplitude of P300 increased gradually as the subject attempted to detect a change in the rule generating a sequence of trials. The P300 was largest just prior to the trial on which the subject announced the detection.

The evidence on hand suggests that the processing manifested by P300 is used in the service of future actions, rather than in the execution of the specific responses to the eliciting event (Donchin, 1979; Donchin, Ritter & McCallum, 1978; Gratton, Dupree, Coles, & Donchin, 1985). Indeed, Gratton, Dupree, Coles and Donchin (1985) have shown that the subject's response bias, in a choice reaction time study, varied as a function of the amplitude of the P300 elicited on trials on which the subject committed an error. The larger the P300 elicited by an error, the less likely was the subject to err on the next trial.

These lines of evidence support Donchin's (1981) argument that P300 represents a process of revision of representations in working memory (or context updating). Studies that focused on changes in P300 amplitude as a function of the previous sequence of stimuli, and on variations in inter-stimulus interval (ISI), are consistent with this context updating hypothesis. For example, Squires, Wickens, Squires and Donchin (1976) demonstrated that the amplitude of the P300 elicited by a stimulus depends on the sequence of the preceding stimuli. Specifically, if a Bernoulli series composed of two stimuli, A and B, is presented, then a stimulus "A" will elicit a larger P300 when it is preceded by a sequence of B stimuli than when it is preceded by stimuli of its own kind. Squires et al. (1976) suggested that the expectation that a stimulus will be repeated depends on the strength of a decaying memory trace.

Investigations of the effects that varying the ISI have on the P300 are consistent with the view that P300 amplitude is inversely proportional to the strength of a decaying memory representation. In fact, as ISI is lengthened, the amplitude of the P300 elicited by the frequent events increases to the point at which the difference between rare and frequent events disappears (Fitzgerald & Picton, 1981; Heffley, 1981). These findings also support the hypothesis that P300 reflects a process involved in the updating, or "refreshing", of representations in working memory. When the ISI is short, working memory needs to be updated only when a rare event is presented, because when a frequent event is presented it is likely to be already represented in working memory. Instead, when the ISI is long, working memory must be updated for all the events, because the representation of either event is likely to have decayed by the time the next event is presented.

Thus, rare or unexpected events will usually lead to a restructuring or updating of working memory, and this activity is part of the ongoing process of maintaining accurate schemas of the environment. The updating process may lead to an "activation" of the representation, or to the "marking" of some attribute of the event that was crucial in determining the updating process. This restructuring of the representation of an event is assumed to facilitate the subsequent recall of the event, by providing valuable retrieval clues, so that the greater the restructuring that follows an individual event, the higher the probability of later recalling that event. If P300 amplitude represents the degree of restructuring in working memory, then P300 amplitude should also predict later recall.

We evaluated the context updating hypothesis using a von Restorff paradigm (Karis, Fabiani, & Donchin, 1984). Words were presented sequentially, using lists composed of 15 unrelated words. After each list, the subjects were asked to recall as many words as possible. Seventy-five percent of the lists contained one word that was "isolated" from the other words by a change in the size of the word. Von Restorff (1933) had found previously that isolated items are usually recalled better than comparable non-isolated items. This enhanced recall of the isolates is called the "von Restorff," or "isolation," effect. In the Karis, Fabiani, and Donchin (1984) study, subjects varied dramatically in their general recall performance, in the degree to which they showed a von Restorff effect, and in the relationship between the P300 amplitude elicited by an isolated word and the subsequent recall of that word. These differences appeared to depend on the mnemonic strategies that the subjects reported upon debriefing. Subjects who reported using rote mnemonic strategies (e.g., repeating the words) recalled far fewer words than subjects who combined the words into complex images, sentences, or short stories. However, these "rote-memorizing" subjects exhibited a larger von Restorff effect than the elaborators. Furthermore, it was only in the subjects using rote strategies that Karis, Fabiani and Donchin (1984) observed a strong relationship between P300 amplitude and recall.

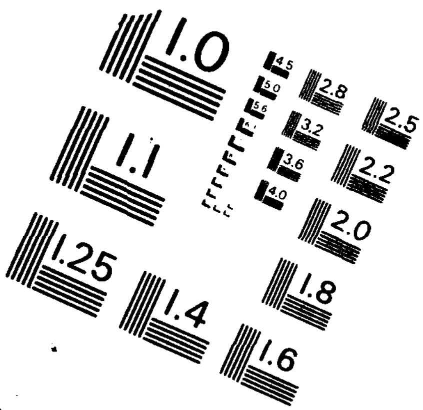
It is important to emphasize that rote memorizers and elaborators did not differ in the amplitude of the P300 elicited by the isolated words, but did differ with respect to the interaction between recall and P300 amplitude. It appears, therefore, that the initial processing of the words was similar for the two groups. In general, we take these data to suggest that the elicitation of a P300 is associated with a change in the internal representation of a word. This change makes it easier to recall the word only for rote memorizers, who rely largely on the activation or marking of the original representation. The elaborators, whose recall depends on the networks of associations formed as the words are presented, gain no benefit from any effect that P300 may have had on the representations.

In the Karis, Fabiani, and Donchin (1984) study the subjects' use of strategies was not controlled. Subjects selected strategies according to their personal proclivities. In the present experiment, the subjects did not know that recall would be tested until they were, without previous warning, asked to recall a list of names presented as part of another task. We reasoned that, in this way, we would minimize the use of elaborative strategies.

This experiment was part of a larger study in which several "oddball" series were presented in sequence.[1] In each oddball task the subject was presented with a Bernoulli series of events and instructed to count (or to respond to) one of the events. After three oddball tasks, in which the series were composed of tones, and of the letters H and S, subjects were

AD-A170 794 THE EVENT-RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING A..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 5/8

UNCLASSIFIED 20 FEB 86 CPL-86-1 F49620-85-C-0041 F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

presented with an oddball series created by randomly mixing male and female names, and were instructed to count names of one gender. The subjects had no reason to expect that they would be asked to recall these names. Therefore, we assumed that they would not develop, and use, elaborative strategies to facilitate recall. When they were subsequently asked to recall the names, the relationship between P300 and recall could then be evaluated in the absence of elaborative strategies.

Method

Subjects

Forty-one male subjects (age range 18 to 31) were recruited by means of advertisements in a local newspaper, and were paid $3.50 per hour, plus bonuses as described below. All were right-handed with normal or corrected to normal vision and hearing.

Data Collection

Ag-AgCl Beckman Biopotential electrodes were affixed at Fz, Cz, and Pz by means of Grass EC-2 electrode cream, and to both mastoids by adhesive collars. Linked mastoids were used as references. Beckman Biopotential electrodes were also used as ground and electrooculographic (EOG) electrodes. The subject was grounded on the forehead. Sub- and supra-orbital electrodes were used to record the vertical EOG. Electrode impedance did not exceed 10 KOhm. The EEG was amplified with model 7P122 Grass amplifiers (time constant 8 seconds, upper half-amplitude frequency 35 Hz, 3dB/octave roll-off) and was digitized at the rate of 100 samples/s for 1500 ms, beginning 100 ms prior to stimulus onset.

All aspects of experimental control and data collection were controlled by a PDP-11/40 computer system (see Donchin & Heffley, 1975). Average waveforms and single-trial records were monitored on-line using a GT40 display. Eye movement artifacts were corrected off-line using a procedure described in Gratton, Coles and Donchin (1983a).

Procedure

The subject sat in an air conditioned, unshielded, room in front of a DEC VT-11 display. The recording and control apparatus was located in an adjacent room. Five oddball tasks were administered to each subject, according to the sequence depicted in figure 1.

Insert Figure 1 About Here

Subjects were given 10 to 20 trials of practice before each new task. For all tasks, the ISI was 2000 ms, and the probability of the rare stimulus was always .20. The relationship between the responding hands and the stimuli, as well as the probability of each stimulus category, were counterbalanced across subjects.

The five oddball tasks were, in the order of occurrence:

a) One block of an auditory oddball (150 trials), in which tones of different pitch (1000 and 1500 Hz) were each presented for 50 ms, and the subject indicated, by pressing one of two buttons, whether the tone was high or low.

b) Two blocks of an H and S oddball (120 trials each), in which a series composed of the letters H and S was presented, and the subjects kept a mental count of the rare letter (in the first block) or of the frequent letter (in the second block). Each letter was displayed for 100 ms.

Results from these first three oddball tasks were presented elsewhere (Fabiani, Gratton, Karis, & Donchin, in press; Karis, Coles, & Donchin, 1984), and will not be discussed further in this paper. These series were essential, however, in providing a context for the "name oddball" series. As far as the subject knew, this was yet another counting task. The request for recall of the names was quite unexpected.

c) The <u>name oddball</u> task consisted of two blocks of 105 trials each. Two lists of names were constructed for each subject by a computer program that randomly selected names from a master list composed of 336 names, half male and half female. Names in the master list were chosen from those given in Battig and Montague (1969), and from a book of baby names. Unusual names, and names that could be both male and female, were excluded. A list of 14 names (7 male and 7 female names) was used as a practice list for all the subjects. The names contained in the practice list were excluded from the master list.

The length of the names varied between 3 and 9 letters, and, for a given subject, no name was ever repeated. Each name was displayed for 200 ms. The letters used to display the names subtended a visual angle of .5 degrees. For each subject one of the categories, either male or female names, was rare (p = .20), with gender of the rare names counterbalanced across subjects.

The subjects were instructed to count a different category in each block, and to report the running total at the end. A $.50 bonus was given to the subjects whenever their count was correct. Since the same category was rare in both blocks, they counted the rare stimuli in one block and the frequent stimuli in the other. The block (first or second) during which the subject had to count the rare stimuli was counterbalanced across subjects.

After the first block of the name oddball was completed, the subjects were asked to perform the incidental memory task described below. A debriefing and the second block of the name oddball followed the memory task. Before starting the second block of the name oddball the subjects were specifically warned that another recall task would not be given.

d) An Incidental free recall task was assigned to the subjects immediately after the completion of the first block of name oddball. They were told that they should try to recall as many names as they could (both male and female names), and were given 5 minutes to write them down. This task was unexpected and all the subjects expressed surprise. To discourage guessing, subjects were awarded $.10 for each name correctly recalled and penalized $.10 for each name incorrectly reported.

At the end of the incidental memory task the subjects were debriefed about the strategies they had used in recalling the names.

Results

The 41 subjects were divided into two groups, depending on whether they counted the rare or the frequent names in the first block of the name oddball. Group 1 was composed of 23 subjects who counted the rare names in the first name oddball. We will refer to this group as the "count-rare group." For 11 of these subjects male names were rare, and for the remaining 12 female names were rare. Group 2 was composed of 18 subjects who counted the frequent names in the first name oddball ("count-frequent group"). Male names were rare for half of these subjects. Subjects were generally very accurate in the count task. Only two subjects missed the correct count by more than one in the first block of the name oddball.

Recall Performance

The percentage of names correctly recalled was computed separately for rare and frequent names for each subject. Note that the subjects were asked to recall both male and female names even though their task was to count just one group.

The 41 subjects recalled, on the average, approximately 17 names (16% of the names presented), of which 4 were rare names (20% of the rare), and 13 frequent (15% of the frequent). The low number of names recalled by the subjects is due to the incidental nature of the memory task and the length of the name list (105 names). The two groups (count-rare and count-frequent) were not significantly different in their average recall scores, $F(1,39) = 3.51$, $p > .05$. The average number of confabulations (names reported which were not in the list) was less than 2 per subject, and the two groups did not differ in this respect, $F(1,39) = 0.00$, $p > .05$. As can be seen in figure 2, subjects recalled more counted (target) names (20.4%) than non-counted (non-target) names (14.5%), $F(1,39) = 16.72$, $p < .001$, and more rare (19.4%) than frequent names (15.5%), $F(1,39) = 7.15$, $p < .02$.

Insert Figure 2 About Here

The strategies that subjects reported using during the recall of the names were divided into four categories. These are listed in table 1, along with their frequency of occurrence.

Insert table 1 About Here

Subjects reported, on average, to have used two types of strategies. Thirty-nine out of the 41 subjects reported thinking about people they knew with the same names.

Memory effect

Our primary interest was in the ERPs elicited by the names in the first block of name oddball, and whether the amplitude of the P300 elicited by these names was related to their subsequent recall. For this purpose, average ERPs were computed for each subject, after sorting the trials according to the stimulus presented (rare or frequent) and its recall (recalled or not recalled in the subsequent test). Relatively few names were recalled by the subjects (see section on memory performance). This was particularly true for the rare names, so that the averages of rare names later recalled are based on very few trials. On the other hand, the large number of subjects that were run in this study allowed us to obtain reliable grand average waveforms. Average ERPs at three electrode locations over all subjects are presented for the count-rare and count-frequent groups in figures 3 and 4 respectively.

Insert figures 3 & 4 About Here

These waveforms are characterized by a broad component, positive at all electrode sites, and largest at the parietal electrode, with latency of about 500 ms. We interpret this component as a P300, on the basis of its latency, polarity, and scalp distribution (Donchin et al., 1978). This interpretation is also supported by the analyses described in this section. In general, the P300 appears larger for the names that were subsequently recalled.

The breadth of the P300 observed in the grand average waveforms of figures 3 and 4 suggested variability in the latency of P300 of different subjects, which may in part obscure the differences between conditions. This conclusion was further supported by the visual inspection of the single subject averages. Therefore, average waveforms sorted on the basis of recall were latency adjusted for each subject and condition. In order to filter out the activity unrelated to P300, data from the three EEG electrodes (Fz, Cz, and Pz) were combined according to the Vector Filter procedure described by Gratton, Coles, & Donchin (1983b; 1985). The weights (-.1 for Fz, .8 for Cz and .8 for Pz) were chosen to maximize the contribution of a centro-parietal scalp distribution, and a positive polarity. A cross-covariance procedure was used to identify the P300 component on the waveforms obtained by combining the electrodes (time window from 400 to 900 ms; see Coles, Gratton, Kramer & Miller, in press). The template adopted was a 2 Hz cosinusoidal wave (1 cycle). The value of maximum cross-covariance between waveform and template was considered the measure of P300 amplitude. This procedure was chosen because it minimizes the variability due to the small number of trials in each average, and because it is insensitive to errors in baseline definition. The use of a cross-covariance procedure on the waveforms obtained with Vector Filter is also supported by a study on the reliability of P300 measures (Fabiani, Gratton, Karis, & Donchin, in press) in which this procedure proved to be the most reliable of the several measures of P300 amplitude evaluated, for this particular task. Further support for this procedure comes from a simulation study of P300 latency measures (Gratton, Kramer, & Coles, 1984; Gratton, Kramer, Coles, & Donchin, 1985). The latency adjusted grand averages are shown in Figure 5 for the count-rare group and in figure 6 for the count-frequent group.[2]

Insert figures 5 & 6 About Here

The difference in P300 amplitude between names subsequently recalled or not recalled is clearly evident in these figures for both groups and stimuli.

An analysis of variance was performed on the amplitude estimates derived using the cross-covariance procedure described above. An "unequal N" design with one nested factor (task: count-rare or -frequent) and repeated measures was used (ALICE statistical package, program "UNEN," Grubin, Bauer, & Walker, 1976). The prediction that P300 amplitude would be greater for names subsequently recalled than for names not recalled was confirmed (main effect of memory: $\underline{F}(1,39) = 19.44$, $\underline{p} < .001$). As expected, rare stimuli elicited significantly larger P300s than frequent stimuli (main effect of stimulus: $\underline{F}(1,39) = 44.14$, $\underline{p} < .001$). However, this effect was much larger for subjects counting the rare stimuli than for subjects counting frequent stimuli, for whom the P300s elicited by the rares were only slightly larger than the P300s elicited by the frequents ("target effect" - task x stimulus interaction: $\underline{F}(1,39) = 7.15$, $\underline{p} < .02$).[3]

A second "peak," occurring after P300, is visible in the latency adjusted waveforms, and it also discriminates between recalled and unrecalled names (figures 5 and 6). This peak is particularly evident for the subjects who were counting the rare stimuli (figure 5). The polarity and scalp distribution of this peak do not differ from those of the preceding P300 (i.e., positive polarity and maximum amplitude at Pz). Note that in the latency adjusted waveforms, which are adjusted on the basis of P300 latency, this peak is more evident than in the unadjusted waveforms, suggesting that its temporal relationship to the P300 is constant. We will comment on the possible significance of this peak in the discussion.

An analysis of the single trials was performed to ensure that the effects observed on the means were not an artifact due to the small number of trials in some of the averages. Given that the average ERPs for recalled names were based on a smaller number of trials than the averages for unrecalled names, the averages for recalled names may contain more noise, which may bias the peak amplitude estimates obtained from the average waveforms. On the other hand, amplitude estimates obtained on single trials are not biased in favor of the "recalled" category. Thus, P300 amplitude estimates were obtained for each trial, subject, stimulus and task, using the cross-covariance procedure described above. This procedure appears to be very reliable in the estimation of P300 latency and amplitude from single trials (Fabiani, Gratton, Karis, & Donchin, in press). The single trials were filtered prior to analysis using a 6.29 Hz half cut-off low-pass filter. An analysis of variance was then performed on the median amplitude estimates, using the same factorial design described above. The results were comparable to those obtained with the analysis on the averages. The effect of recall was replicated (main effect of memory: $F(1,39) = 5.37$, $p < .05$). In addition, the difference in P300 amplitude between recalled and unrecalled names was larger for subjects in the count-frequent group (task x memory interaction: $F(1,39) = 6.03$, $p < .02$). The effect of probability on

P300 amplitude and the target effect were also replicated (main effect of stimulus: $F(1,39) = 21.25$, $p < .001$; task x stimulus interaction: $F(1,39) = 21.68$, $p < .001$).

Probability and target effects

As described above, the two groups of subjects (count-rare and count-frequent) showed different patterns of recall in the first block of the name oddball, as well as differences in their waveform of the ERP elicited by the names. The count-rare group showed a large effect of probability; this was mitigated by a large target effect in the count-frequent group. Therefore, it is important to determine if the two groups of subjects produce similar ERP waveforms when given the same instructions. To this purpose we compared the ERP waveforms obtained in the first block of the name oddball for group 1 with those obtained in the second block for group 2, and vice versa. In this comparison the two groups are matched with respect to the task (both groups are counting rare names or both are counting frequent; see figure 1).

Average ERPs were computed for each subject and each block of the name oddball (count-rare and count-frequent) according to type of stimulus (rare or frequent). Grand average ERPs at the parietal electrode, for rare and frequent names, for both groups of subjects, and for both blocks of name oddball are presented in figure 7.

Insert figure 7 About Here

As expected, the rare names elicited larger P300s than the frequent names. However this probability effect was tempered in the count-frequent condition by a large target effect: the frequent names, when counted, elicited P300s almost as large as the uncounted rare names. Target rare names showed the largest P300 and non-target frequent names the smallest. Note that ERPs recorded under the same instructions for the two groups are very similar, even though the subjects in the two groups were different, and were given the instructions in reverse order.

Finally, it is interesting to note that the recall performance (see figure 2) in the first block varies consistently with the absolute amplitude of P300. That is, rare targets are best recalled and elicit, on average, the largest P300, while frequent non-targets are infrequently recalled and elicit, on average, the smallest P300. The recall of rare non-targets and of frequent targets falls between these extremes and so does their average P300 amplitude.

Discussion

We have argued that P300 amplitude is a manifestation of the updating of representations in working memory (context updating - Donchin, 1981; Karis, Fabiani, & Donchin, 1984). If this updating process alters the representation of words in memory, and this change increases the probability of recall, then variations in P300 amplitude should be related to memory performance. Exactly how this updating process alters a representation in memory is not clear. It may increase the "activation" level of the representation, or the distinctiveness of the representation may be increased by marking it in some way. The data we reported previously (Karis, Fabiani, & Donchin, 1984) suggest that the process manifested by the P300 has its effect primarily on the representation of the actual stimulus presented to the subject rather than on its more global network of associations. This conclusion is supported by the fact that the use of elaborative mnemonic strategies during learning can eliminate the relationship between P300 amplitude and recall.

The purpose of the present study was to test the strength of the relationship between P300 and recall when the use of elaborative strategies can be assumed to be minimal. We have created an incidental memory paradigm by camouflaging the list to be recalled as a series of items to be counted. The name oddball series was embedded in a sequence of other oddball tasks. Subjects counted the male or female names, with no expectation that recall would be required. During the presentation of names subjects categorized each name as "male" or "female" in order to maintain an accurate count of one category. Our main prediction was that names that were subsequently recalled would have elicited larger P300s than names that were not recalled. This prediction was confirmed in a series of analyses using a variety of measurement techniques, based on both averages and single trials. In addition, our prediction was also supported by the observation that the conditions in which the amplitude of the P300 is largest are those in which the recall performance is highest.

Even when the use of mnemonic strategies is minimized by using an incidental memory paradigm, subjects can still use a variety of strategies during recall (as listed in table 1). It is evidently common for the presentation of a name to evoke a recollection of a person known by that name. These recollections were then relied on as a mnemonic strategy. It is possible that the choice of strategy at recall can also influence the magnitude of the P300-recall relationship. For example, since there is a finite set of names, a subject can mentally run through a list of common names alphabetically, and thus change the task from recall to recognition. This may reduce the advantage of words transformed by the updating process, and overshadow the relationship between P300 amplitude and recall.

A positive peak, with a latency of about 900 ms, is clearly evident in figures 5 and 6. It is tempting to interpret this peak as a second P300. Precise measurement of this peak independently of P300 is difficult, given the similarity in polarity, scalp distribution, time window, and the sensitivity to recall-related processes. Actually, this second peak does not seem to discriminate between rare and frequent stimuli (see figure 7). Johnson and Donchin (1985) reported observing multiple P300s following a single stimulus when the subjects may engage "second thoughts" about their performance. In this vein, one possible explanation of this second peak is that it is actually a second P300, that is related to an additional categorization of the name. This second peak is most evident when subjects are counting the rare stimuli. From the single-trial analysis it also appears that, for these subjects, the "first" P300 is less related to memory than for subjects who count the frequent stimuli. In the case of the count-rare task, therefore, the subjects may first realize that the name is a member of the category of names presented rarely, and which they have to count (first P300), and then they may also realize that it is the name of a known person, or the title of a song, etc. (second P300).

When the subjects are counting the frequent names, they have to increase their count on almost every trial. In this case, it may be that the two processes (categorization of the name as a member of the target category, and of the "known-people category") may occur concurrently, and only one P300, predicting later recall, will be emitted.

The present results confirm the conclusion drawn by Karis, Fabiani &

Donchin (1984) that when the use of elaborative strategies is minimized the amplitude of the P300 elicited by an event is correlated with the probability that this event will be recalled. These data are also consistent with the theory that the P300 is a manifestation of processes invoked when events occur and create a need to revise the current representations in working memory (context updating - Donchin, 1981). We have shown that, when subjects are not likely to use mnemonic strategies to memorize material, a strong relationship does emerge between P300 amplitude and subsequent recall.

References

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology, 80, 1-46.

Coles, M. G. H., Gratton, G., Kramer, A. F., & Miller, G. A. (in press). Principles of signal acquisition and analysis. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), Psychophysiology: Systems, processes, and applications. New York, NY: Guilford Press.

Donchin, E. (1979). Event-related potentials: a tool in the study of human information processing. In H. Begleiter (Ed.), Evoked potentials and behavior (pp.13-75). New York, NY: Plenum Press.

Donchin, E. (1981). Surprise? ... Surprise! Psychophysiology, 18, 493-513.

Donchin, E., Coles, M. G. H., & Gratton, G. (1984). Cognitive psychophysiology and preparatory processes: a case study. In S. N. Kornblum & J. Requin (Eds.) Preparatory states and processes (pp. 155-178). Hillsdale, NJ: Erlbaum Associates.

Donchin, E., & Heffley, E. (1975). Minicomputers in the signal-averaging laboratory. American Psychologist, 38, 449-461.

Donchin, E., Ritter, W., and McCallum, C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting & S. H. Coslow (Eds.), Event-related Brain Potentials in man (pp. 349-411). New York: Academic Press.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. Psychophysiology, 14, 456-467.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (in press). The definition, identification and reliability of measurement of the P300 component of the event-related brain potential. In P. K. Ackles, J. R. Jennings & M. G. H. Coles (Eds.) Advances in Psychology, (Vol. 2). Greenwich, CT: JAI Press, Inc.

Fitzgerald, P. G., & Picton, T. W. (1981). Temporal and sequential probability in evoked potential studies. Canadian Journal of Psychology, 35, 188-200.

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

Gratton, G., Coles, M. G. H., & Donchin, E. (1983b, abstract). Filtering for scalp distribution: A new approach (Vector Filter). Psychophysiology, 20, 443-444.

Gratton, G., Coles, M. G. H., & Donchin, E. (1985). A multivariate approach to the analysis of the scalp distribution of event-related potentials: Vector Analysis. Unpublished manuscript.

Gratton, G., Dupree, D., Coles, M. G. H., & Donchin, E. (1985). The consequences of an error: P300 and future action. Unpublished manuscript.

Gratton, G., Kramer, A. F., & Coles, M. G. H. (1984, abstract). A comparative study of measures of the latency of event-related brain components. Psychophysiology, 21, 578-579.

Gratton, G., Kramer, A. F., Coles, M. G. H., & Donchin, E. (1985). A simulation study of latency measures of components of event-related brain potentials. Unpublished manuscript.

Grubin, M. L., Bauer, J. A., Jr., & Walker, E. C. T. (1976). ALICE user's guide. Natick, MA: Alice Associates.

Heffley, E. F. (1981, April). Event-related brain potentials in visual monitoring tasks: Selective attention, display format, and event frequency. Paper presented at the Eastern Psychological Association annual meeting, New York, NY.

Jasper, H. H. (1958). Report of the committee on methods of clinical examination in electroencephalography. Electroencephalography and Clinical Neurophysiology, 10, 370-375.

Johnson, R., Jr., & Donchin, E. (1985). Second thoughts: Multiple P300s elicited by a single stimulus. Psychophysiology, 22, 182-194.

Johnson, R., Jr., & Donchin, E. (1978). On how P300 amplitude varies with the utility of the eliciting stimuli. Electroencephalography and Clinical Neurophysiology, 44, 424-437.

Johnson, R., Jr., & Donchin, E. (1982). Sequential expectancies and decision making in a changing environment: An electrophysiological approach. Psychophysiology, 19, 183-200.

Karis, D., Coles, M. G. H., & Donchin, E. (1984). The use of event-related brain potentials for prediction of performance quality in pilots of advanced tactical jet aircraft (Tech. Rep. No. CPL 83-1-F33615-82-K-0611). Champaign: University of Illinois, Cognitive Psychophysiology Laboratory.

Karis, D., Fabiani, M., & Donchin, E. (1984). P300 and memory: Individual differences in the von Restorff effect. Cognitive Psychology, 16, 177-216.

Pritchard, W. S. (1981). Psychophysiology of P300. Psychological Bulletin, 89, 506-540.

Restorff, H. von (1933). Uber die Wirkung von Bereichsbildungen im Spurenfeld. Psychologische Forschung, 18, 299-342.

Squires, K. C., Wickens, C., Squires, N. K., & Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. Science, 193, 1142-1146.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Information delivery and the sensory evoked potential. Science, 150, 1187-1188.

Footnotes

1. A series of events that can be divided into at least two discrete classes is called an "oddball paradigm" when one event (or class of events) is rarer than the other (although sometimes even series with .5/.5 probability are called oddballs).

2. Note that the waveforms shown in figure 5 and 6 are not the waveforms combined using Vector Filter. They are the grand averages obtained by aligning the averages for each subject at each electrode on the basis of P300 latency (measured on the combined waveforms by a cross-covariance procedure).

3. To validate our cross-covariance procedure for estimating P300 amplitude on Vector filtered waveforms, we also used other measures, including peak-picking at Pz and on Vector filtered waveforms, and cross-covariance at Pz. These techniques gave comparable results, although the F values were slightly smaller.

Table 1

Percentage of Subjects Using Various Retrieval Strategies

	Experimental Conditions		
Strategy	Count-Rare (N=23)	Count-Frequent (N=18)	Total (N=41)
Think of known people with these names.	96	94	95
Unusual, long or funny names.	57	39	49
Go through alphabet thinking of common names and decide whether they were in the list or not.	30	33	32
Miscellaneous (e.g., movies or songs with the same names, names from the Bible, mental visualization of the screen with the name on it, etc.)	13	28	20

Figure captions

Figure 1. Experimental design. Note that the first block of the name oddball, after which subjects were unexpectedly asked for free recall, was preceded by three oddball tasks, where subjects responded to, or counted, the rare or frequent stimuli, but were <u>not</u> asked to recall any of the stimuli.

Figure 2. Bar graph depicting the recall performance for each group (count-rare and count-frequent) and type of stimulus (rare and frequent). The shaded bars indicate the counted (target) stimuli.

Figure 3. Grand average waveforms from the first block of the name oddball of subjects in the count-rare group (N=23). Rare names (left column) and frequent names (right column) were sorted on the basis of their subsequent recall.

Figure 4. Grand average waveforms from the first block of the name oddball of subjects in the count-frequent group (N=18). Rare names (left column) and frequent names (right column) were sorted on the basis of their subsequent recall.

Figure 5. Latency adjusted grand average waveforms from the first block of the name oddball of subjects in the count-rare group (N=23). Rare names (left column) and frequent names (right column) were sorted on the basis of their subsequent recall. The two vertical dashed lines indicate the limits of the segment of waveform identified as P300.

Figure 6. Latency adjusted grand average waveforms from the first block of the name oddball of subjects in the count-frequent group (N=18). Rare names (left column) and frequent names (right column) were sorted on the basis of their subsequent recall. The two vertical dashed lines

indicate the limits of the segment of waveform identified as P300.

Figure 7. Grand average waveforms at Pz, for the two groups of subjects in the two blocks of the name oddball. Waveforms for group 1 (N=23 - subjects who counted the rare names in the first block) are on the left. Waveforms for group 2 (N=18 - subjects who counted the frequent names in the first block) are on the right. Solid lines indicate rare stimuli, dashed lines indicate frequent stimuli.

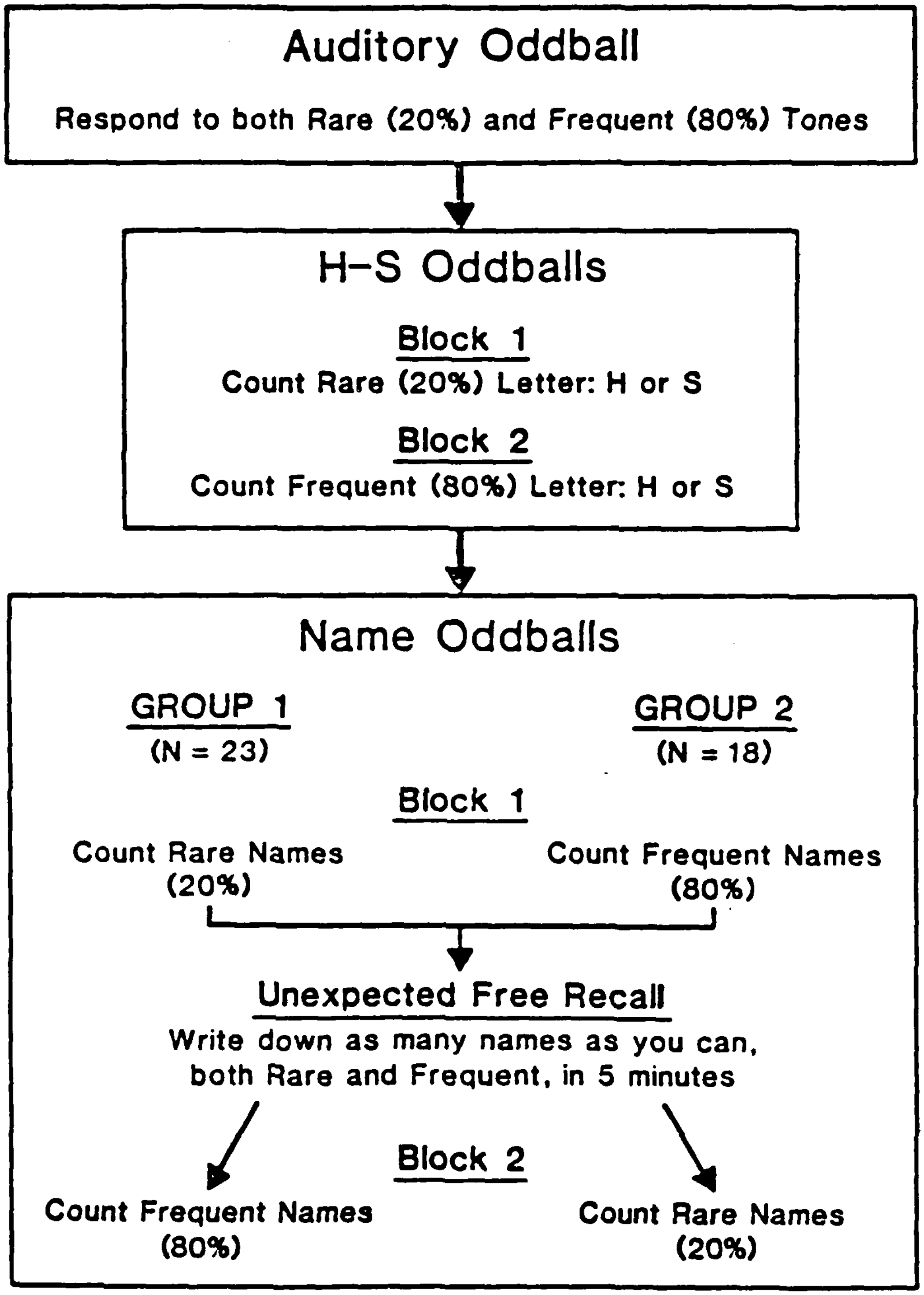

Auditory Oddball
Respond to both Rare (20%) and Frequent (80%) Tones
H–S Oddballs
Block 1
Count Rare (20%) Letter: H or S
Block 2
Count Frequent (80%) Letter: H or S
Name Oddballs
GROUP 1
(N = 23)
GROUP 2
(N = 18)
Block 1
Count Rare Names (20%)
Count Frequent Names (80%)
Unexpected Free Recall
Write down as many names as you can, both Rare and Frequent, in 5 minutes
Block 2
Count Frequent Names (80%)
Count Rare Names (20%)

Fig. 1

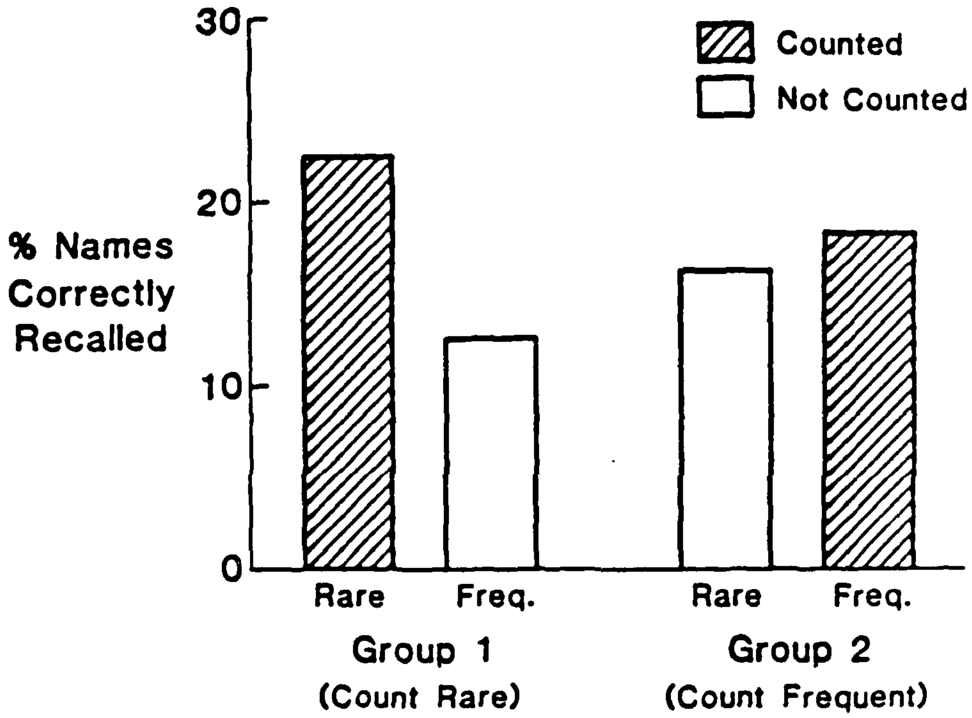
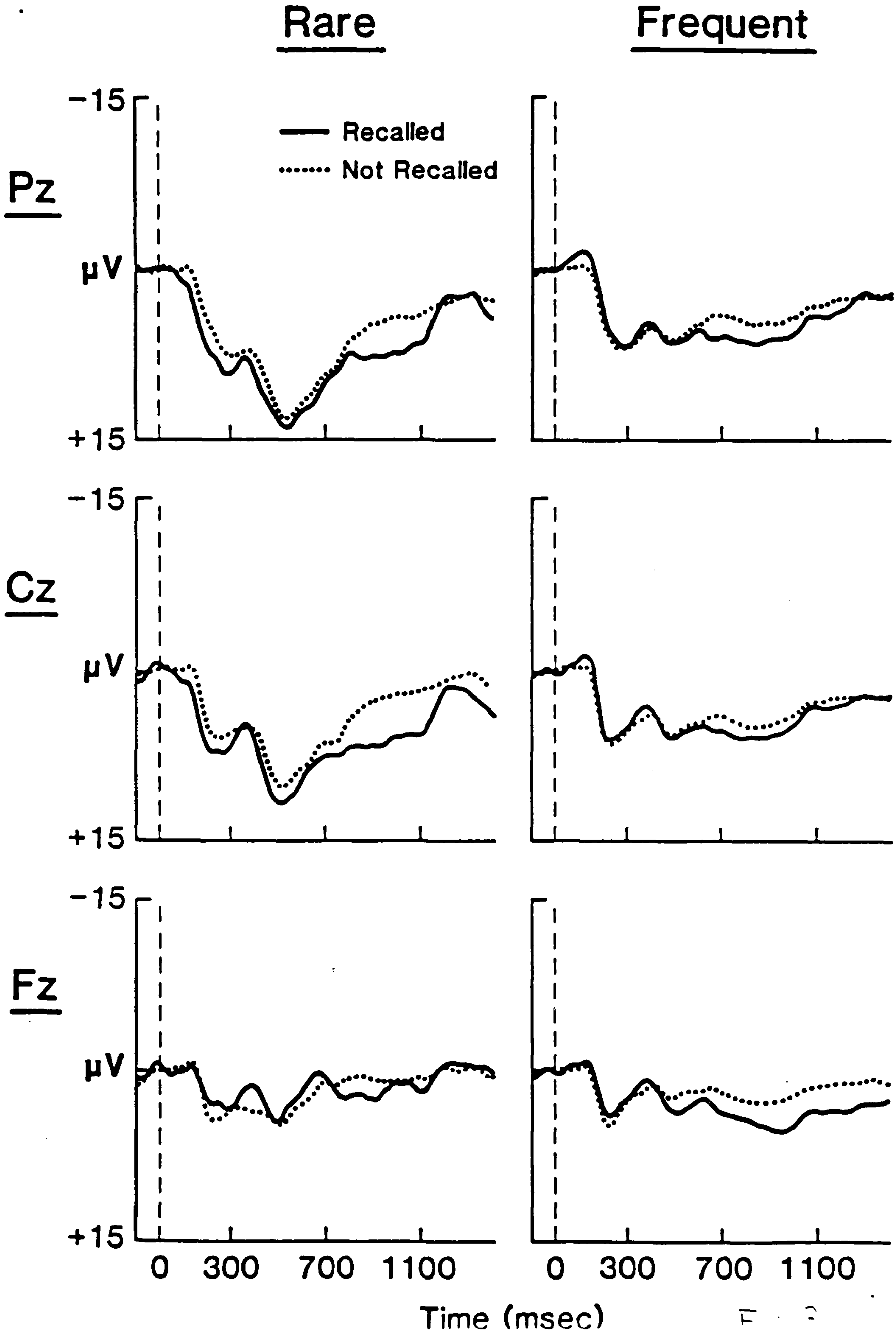
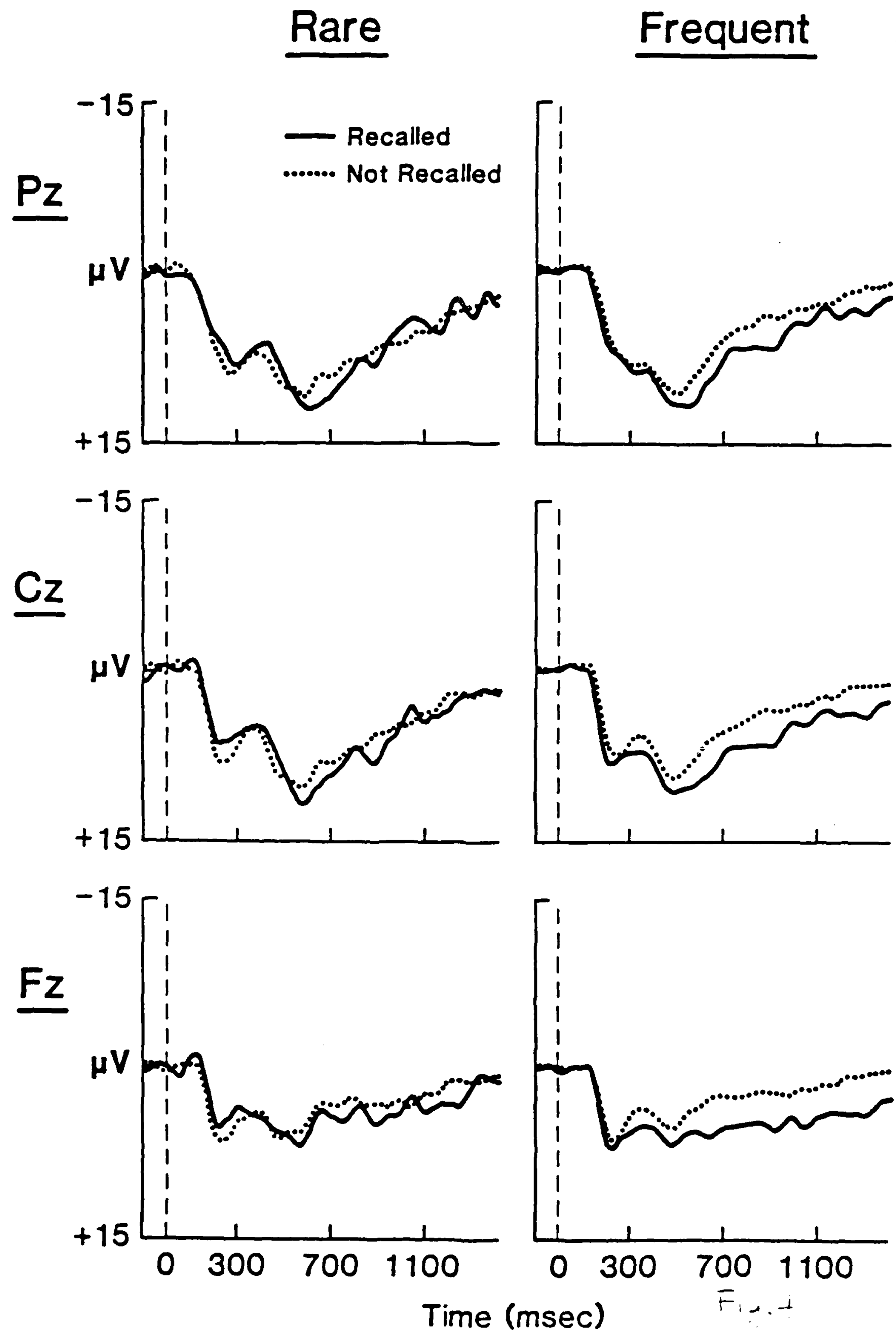

30
20
10
0
% Names Correctly Recalled
Counted
Not Counted
Rare
Freq.
Rare
Freq.
Group 1
(Count Rare)
Group 2
(Count Frequent)

Fig 2

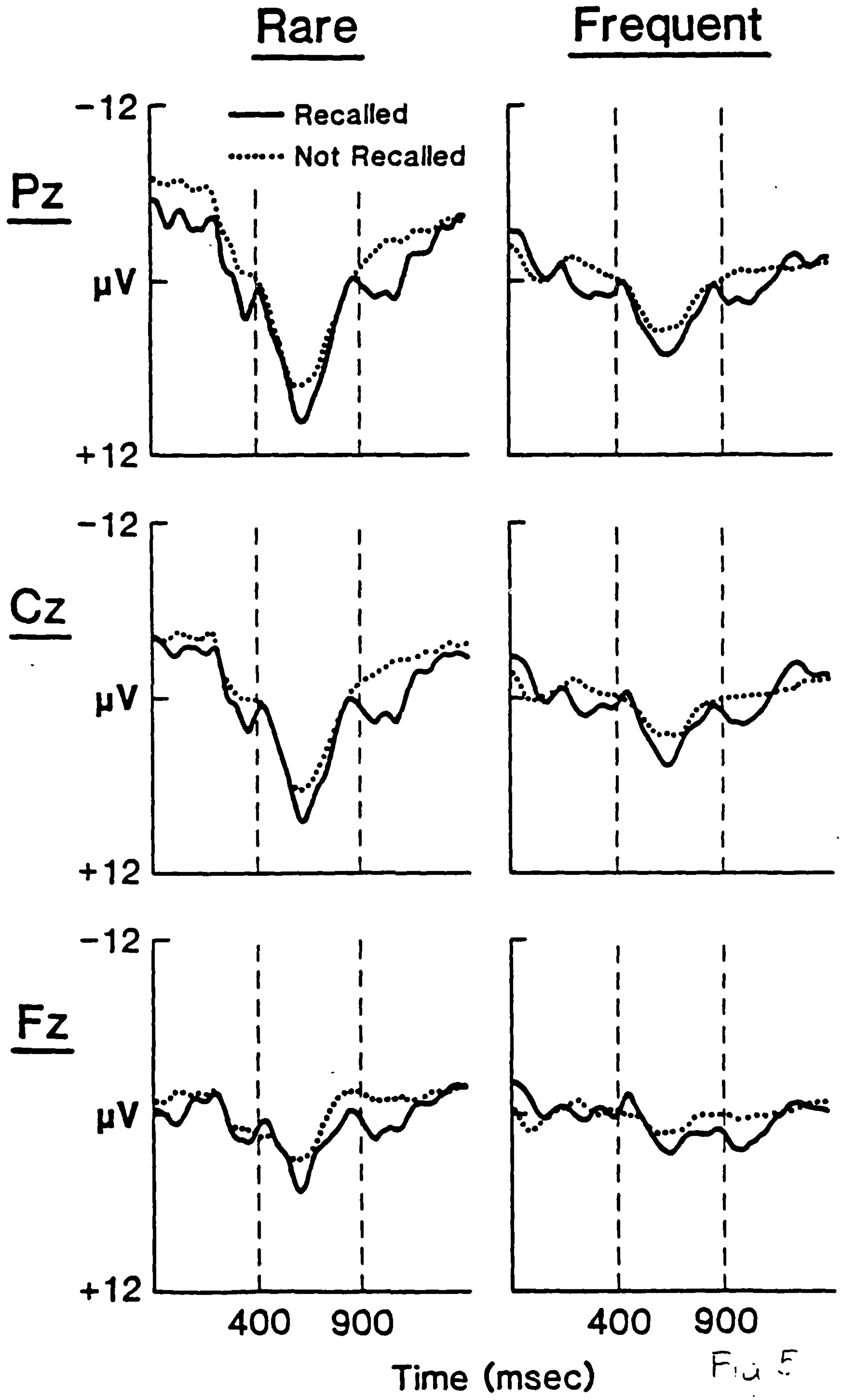


Rare
Frequent
Recalled
Not Recalled
Pz
Cz
Fz
μV
-15
+15
0
300
700
1100
Time (msec)



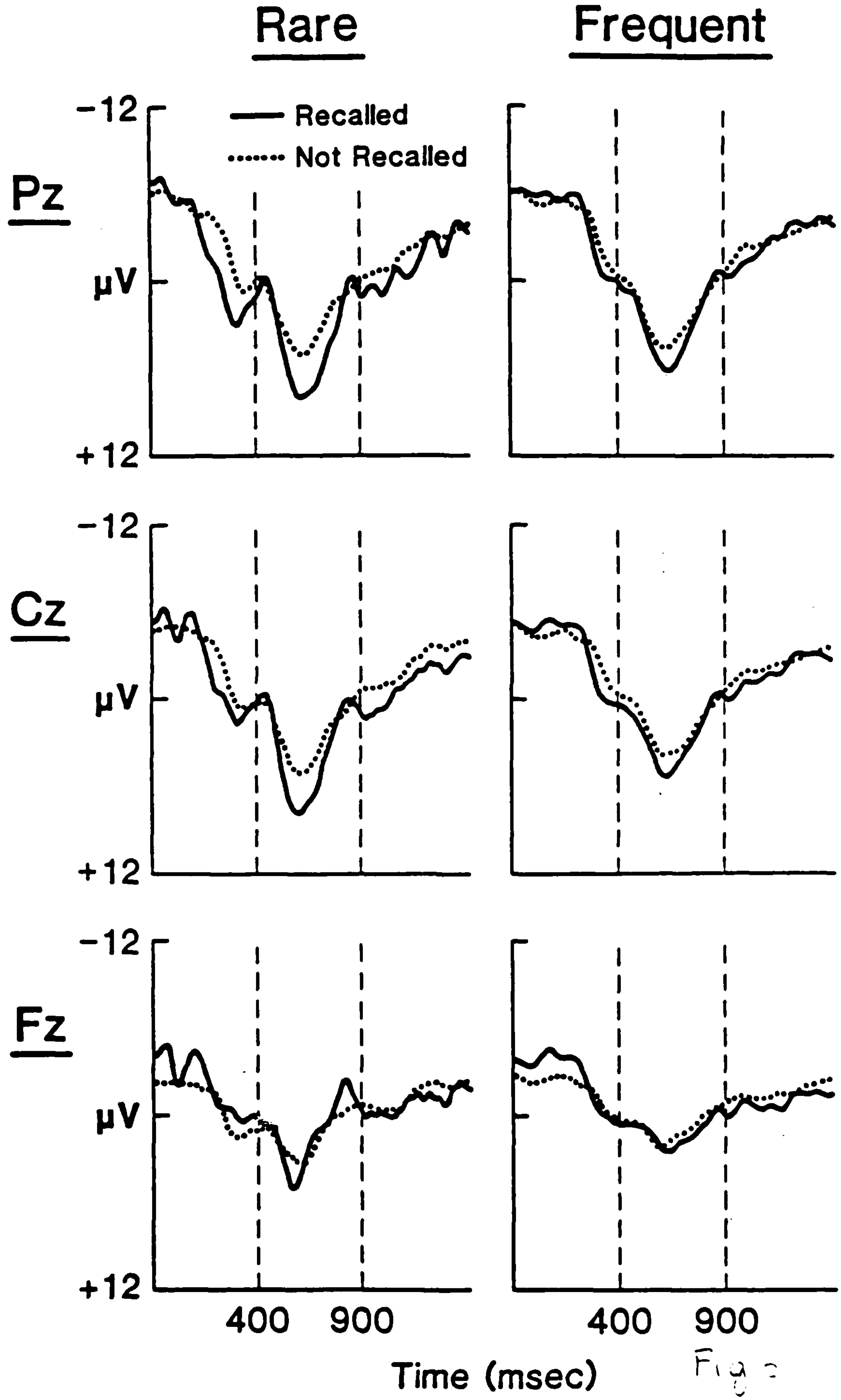
Rare
Frequent
Recalled
Not Recalled
Pz
Cz
Fz
−15
μV
+15
0
300
700
1100
Time (msec)

Fig. 4

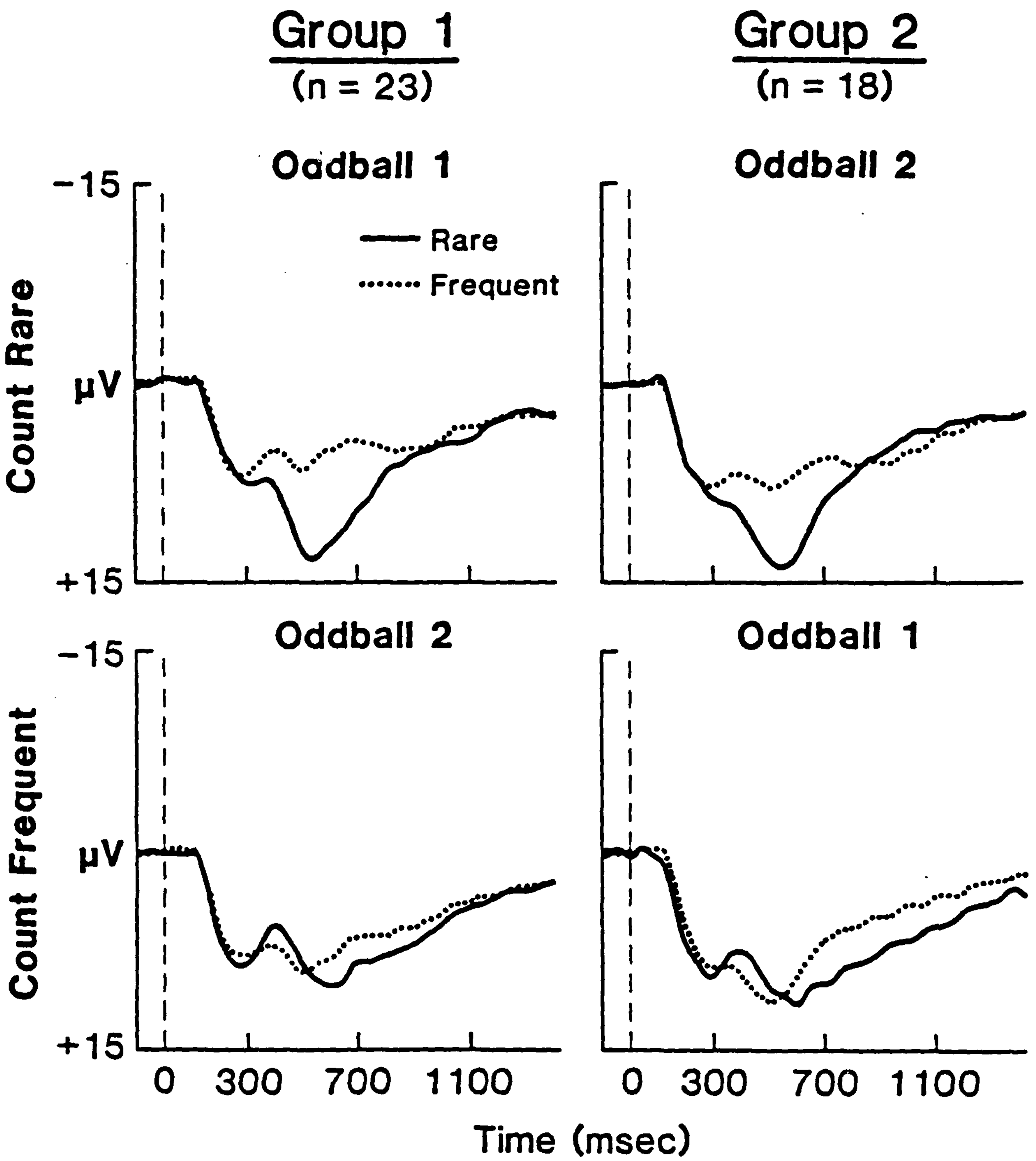


Rare
Frequent
Recalled
Not Recalled
Pz
Cz
Fz
−12
μV
+12
400
900
Time (msec)

Fig 5



Rare
Frequent
Recalled
Not Recalled
Pz
Cz
Fz
−12
μV
+12
400
900
Time (msec)



Group 1
(n = 23)
Group 2
(n = 18)
Oddball 1
Oddball 2
Rare
Frequent
Count Rare
Count Frequent
−15
μV
+15
Oddball 2
Oddball 1
0
300
700
1100
Time (msec)

Fig 7

Effects of Mnemonic Strategy Manipulation in a von Restorff Paradigm

Monica Fabiani, Demetrios Karis

and Emanuel Donchin

Cognitive Psychophysiology Laboratory

University of Illinois

Champaign, Illinois

Partial reports of this study were presented at the Third International Conference on Cognitive Neuroscience (ICON III), Bristol (England), September 16-21, 1984, and at the 25th annual meeting of the Society for Psychophysiological Research, Houston (TX), October 16-20, 1985. The research described was supported by the Air Force Office of Scientific Research, contract #1-5-25288, Al Fregly program director. We are indebted to Gabriele Gratton for helpful comments and statistical advice. D. K. is currently with the Human Factors Group, Grumman Aerospace Corporation, MS C04-14 Bethpage, NY, 11714.

Address requests for reprints to: Emanuel Donchin, Psychology Department, University of Illinois, 603 East Daniel, Champaign, Illinois, 61820.

Running head: P300 and Memory

Abstract

Donchin (1981) proposed that the P300 component of the Event-Related Brain Potential (ERP) manifests the updating of schemas in working memory ("context updating"). In a previous study Karis, Fabiani and Donchin (1984) tested the hypothesis that words recalled had elicited, on their initial presentation, a larger P300 than words subsequently not recalled. They found that this relationship was evident only for those subjects who used rote mnemonic strategies, while it was not evident in subjects who used elaborative strategies. This conclusion was drawn by capitalizing on different strategies used by different subjects.

In the present experiment this relationship between P300 amplitude and recall was confirmed within subjects, by having each subject use both strategies. Ten subjects participated in a three-session experiment. We recorded ERPs elicited by words in series that contained a deviant word (an "isolate"). In general, isolated items are better recalled than comparable non-isolated items (the von Restorff effect). Subjects were instructed to use either rote or elaborative strategies to memorize the words. When instructed to use rote strategies, subjects displayed a significantly higher von Restorff effect and lower performance than when instructed to use elaborative strategies. Furthermore, analysis of the ERP data supported the hypothesis that the amplitude of P300 is related to subsequent recall only when subjects use rote strategies, while this relationship is not present when subjects continue their processing well beyond the P300 time-range, by using elaborative strategies.

Descriptors: Event-related brain potentials, P300, memory, mnemonic strategies, von Restorff effect.

Introduction

Karis, Fabiani, and Donchin (1984) reported the existence of an interaction between processes occurring at the moment of stimulus encoding and manifested by a component of the event related brain potentials (ERP) (footnote 1), commonly labelled P300, rehearsal strategies and recall performance. The encoding-related processes, manifested by the P300, predicted the subsequent recall of words for subjects who used rote rehearsal strategies, but not for subjects who used elaborative strategies. Subjects in this experiment used the strategies of their choice, and reported them in a post-experimental debriefing. However, in order to demonstrate that the relationship between encoding processes and recall performance is indeed due to rehearsal strategies, strategies need to be manipulated in a within subject design. This was the purpose of the experiment reported in this paper.

The P300 component of the ERP.

The P300 component of the ERP was first described by Sutton, Braren, Zubin, and John (1965). It peaks 300 ms or more after the eliciting event, is positive at all the midline electrode sites and is maximal at the parietal electrode (Pz - 10-20 International System, Jasper, 1958).

The P300 is elicited only by events that are relevant to the task the subject is performing, and its amplitude is inversely related to the subjective probability of the eliciting event (Duncan-Johnson & Donchin, 1977; Johnson & Donchin, 1978). The dependence of P300 amplitude on these variables suggests that the process manifested by P300 is associated with the processing of novel and relevant stimuli.

One of the critical clues to the functional significance of the P300 is that the processing represented by P300 appears to be used in the service of future actions, rather than in the execution of the specific responses to the eliciting event (Donchin, 1979; Donchin, Gratton, Dupree, & Coles, in press; Donchin, Ritter, & McCallum, 1978; Kutas, McCarthy, & Donchin, 1977; Munson et al., 1984). For instance, Donchin, et al. (in press) found that the subject's response bias, in a choice reaction time study, varied as a function of the amplitude of the P300 elicited on trials on which the subject committed an error. The larger the P300 elicited by an error, the less likely was the subject to err on the next trial.

These data, and data from studies that focused on changes in P300 amplitude as a function of the previous sequence of stimuli (Squires, Wickens, Squires, & Donchin, 1976) (footnote 2) and on variations in inter-stimulus interval (Heffley, 1981; Fitzgerald & Picton, 1981), are consistent with the hypothesis that P300 represents a process of revising the representations in working memory (or "context updating" - Donchin, 1981) (footnote 3). When a rare event occurs, or an error is made, or conditions change, this information must be incorporated into schemas. These schemas will both govern the perception and action taken on future trials, and also affect the recall of the information on the specific trial.

The role of novelty in learning: P300 and memory.

In recent years, the role of novelty in learning has been enphasized by several theories of animal learning (MacKintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972; Wagner, 1976; 1978; see also Dickinson, 1980, for a general review). No one theory has yet received universal acceptance, but all agree that changes in a stimulus representation following a learning experience depend upon the extent to which the predictor and/or the predicted event are jointly processed in working memory. A critical condition for this joint processing is that the stimulus should be surprising or unexpected. Similar theories of human memory (Ohman, 1979; Sokolov, 1963; 1969; 1975) have also argued that a short term memory is used to maintain an internal model of a dynamic environment, and that deviations from this internal model require an updating process.

Rare or unexpected events should lead to a restructuring or updating of the current memory schemas, because only by doing so an accurate representation of the environment can be maintained. The updating process may involve an "activation" of the memory representation of the event, or the "marking" of some attribute of the event that was "distinctive", and therefore crucial in determining the updating process. This restructuring of the memory representation of an event is assumed to facilitate its subsequent recall, by providing valuable retrieval clues, so that the greater the restructuring that follows an individual event, the higher the probability of later recalling that event. P300 amplitude is assumed to be proportional to degree of restructuring of the memory representation of the event. Therefore, P300 amplitude should also predict the *subsequent* recall of the eliciting event.

The use of P300 in the investigation of working memory.

The Karis et al. (1984) paper. Karis, et al. (1984) recorded ERPs to words in a von Restorff paradigm (von Restorff, 1933). Series composed of unrelated words were presented and after each series the subjects were asked to recall as many words as possible. Most of the series contained a deviant word (an "isolate"). The isolation was achieved by changing the size of the word. When one item in a series is distinctly different from the others (e.g., because of color, size, meaning, or class) the probability that it will be recalled increases. The label von Restorff, or isolation effect, refers to the enhanced recall of an "isolated" item, with respect to comparable, non-isolated items (for a review, see Cimbalo, 1978; Wallace, 1965).

Measures of the magnitude of the von Restorff effect, of the general recall performance, and of the amplitude of the P300 component of the ERP were computed for each subject.

Striking individual differences emerged on all measures, and subjects were placed into three distinctly different groups based on the magnitude of their von Restorff effect in the free recall. In group 1 subjects' overall performance was low, but "isolating" a word by changing its size increased recall dramatically (high von Restorff effect). These subjects reported using primarily rote strategies (e.g., repeating the words). Furthermore, it was only in these subjects that Karis et al. (1984) observed a positive relationship between P300 amplitude and recall. At the other extreme, subjects in group 3 exhibited high overall performance, and there was no effect of isolation on recall. These subjects reported using complex elaborative strategies (e.g., making up sentences, stories or images). The association between recall and P300 was absent in subjects of this group. A "frontal-positive slow wave" (footnote 4) was also sensitive to the probability of recall, and subjects in group 3 (good memorizers with a low von Restorff index) exhibited more evidence of this component than the other two groups. Karis et al. (1984) speculated that the slow wave could be associated with the beginning of elaborative organizational processes.

The intriguing aspect of the data reported by Karis et al. (1984) is the modulating role that rehearsal strategies played on the relationship between P300 and recall. P300 provides information about the cognitive processing of an event occurring during the first second after its presentation, while processes that influence recall often continue for an extended period. The relationship between P300 and recall will thus depend on the nature of this extended mnemonic processing.

The results of the Karis et al. (1984) study are consistent with a three-phase model of the information processing leading to the subjects' recall. Phase 1 is driven by the processes that are invoked as the words are encoded and categorized. The process manifested by the P300 is consistently activated by the words that are displayed in a deviant (isolated) font. This processing affects the representation of the word and seems to occur with equal frequency in all the subjects of the Karis et al. (1984) study. In fact, the average amplitude and the amplitude distribution of the P300s elicited by isolated words was the same for rote memorizers and elaborators. Karis et al. (1984) interpreted these data as evidence that all subjects "noticed" the isolated words, and reacted upating their memory representations and producing large P300s. The differences between the subjects emerge in phase 2 - when subjects are trying to memorize the stimuli, by using different types of rehearsal strategies. The subjects' mnemonic strategies continue to play a crucial role in phase 3, when subjects are trying to retrieve the words. Changes in the stimulus representation, induced by the isolation and manifested by P300, make it easier to recall the word only for rote memorizers, who rely largely on the original representation, activated or marked when the word was first presented. The elaborators, whose recall depends on the networks of associations formed as the series were presented, gain no benefit from any effect that the process manifested by P300 might have had on the representations.

An incidental memory experiment. In a second experiment Fabiani, Karis, and Donchin (1986) tried to minimize individual differences in the subjects' mnemonic strategies by using an incidental memmory paradigm. In this study, several "oddball" series were presented in sequence (footnote 5). In each oddball task the subject was presented with a Bernoulli series of events and instructed to count (or to respond to) one of the events. After three oddball tasks, in which the series were composed of tones, and of the letters H and S, subjects were presented with an oddball series created by randomly mixing male and female names, and were instructed to count names of one gender. The subjects had no reason to expect that they would be asked to recall these names. Therefore, Fabiani et al. (1986) assumed that they would not develop, and use, complex associative strategies to facilitate recall and therefore the relationship between P300 and recall could then be evaluated in the absence of elaborative strategies. Our main prediction was that names that were subsequently recalled would have elicited larger P300s than names that were not recalled. This prediction was confirmed: when complex elaborative strategies are minimized by the use of an incidental memory paradigm, the relationship between P300 and memory emerges consistently.

The present study.

Although the Fabiani et al. (1986) data are consistent with the predictions derived from the Karis et al. (1984) study, they are mostly based on a negative finding (i.e., the reduction of individual differences due to rehearsal strategy when an incidental memory test is employed).

However, the individual differences observed by Karis et al. (1984) could still be attributed to the subjects' idiosyncrasies rather than to the use of particular mnemonic strategies. Therefore, it is crucial to determine if an interaction between the updating process (manifested by P300) and mnemonic strategies in determining recall performance exists within rather than between subjects. Thus, we devised a paradigm in which we directly manipulated strategies by instructions. The subjects were run in a von Restorff paradigm, similar to that used by Karis et al. (1984). Instructions to use "rote" strategies required the subject to repeat each word as it was presented, while "elaborative" instructions required the subject to combine words into images, sentences, or stories. We expected that the same subject would behave like a rote memorizer of the Karis et al. (1984) study when given rote instructions, and like an elaborator when given elaborative instructions.

In addition we reasoned that, if it is indeed the case that the deviant size is a "distinctive" attribute of the memory representation of the isolates, such as to help recall in the case of rote instructions, subjects should have a better memory of the word size when they had used rote strategies than when they had used elaborative strategies. To test this hypothesis we constructed a "size-recall test".

Method

Subjects.

Ten right-handed female subjects were run individually in an experiment consisting of three sessions. All the subjects were undergraduate students at the University of Illinois (age range = 18 to 21, median = 19.5). They were paid $3.50 per hour, with a $5 bonus when they completed the third session.

Data Collection.

Ag-AgCl Beckman Biopotential electrodes were affixed along the midline of the scalp at frontal, central, and parietal sites (Fz, Cz, and Pz) by means of Grass EC-2 electrode cream. Ag-AgCl Beckman Biopotential electrodes, affixed by means of adhesive collars, were used as grounds, electrooculogram (EOG) and references. The subject was grounded on the forehead. Sub- and supra-orbital electrodes were used to record the vertical EOG. Linked mastoids were used as references. Electrode impedance did not exceed 10 KOhm. The EEG was amplified with Van Gogh Model 50000 amplifiers (time constant 10 seconds, upper half-amplitude frequency 35 Hz, 3dB/octave roll-off) and was digitized at the rate of 100 samples/sec.

All aspects of experimental control and data collection were controlled by a PDP-11/40 computer system interfaced with an Imlac graphics processor (Donchin & Heffley, 1975). Average waveforms and the single-trial records were monitored on-line using a DEC VT-11 display processor. Eye movement artifacts were corrected off-line using a procedure described by Gratton, Coles, and Donchin (1983).

The data were filtered before being analized, by using an off-line moving average (1 iteration) corresponding to a 3.14 Hz filter (Ruchkin & Glaser, 1978).

Word List.

Five word lists were constructed. The first list was used for all the subjects during the first session (list F). The remaining 4 lists were used for the second and third session (a pair per session - Lists A-B and C-D). Each word list was composed of several series of 15 words. Seventy-five percent of the series contained an isolated word (Isolated series), and 25% did not (Control series). The isolated word, displayed in large size, occurred, at random, from position 6 through 10 of each Isolated series.

Words in each list were selected at random by a computer program, from a master list composed of all the actual words with 3 to 7 letters in Toglia and Battig (1978). Each word could appear in only one list, so that it was presented no more than once to each subject. The computer program also determined, at random, which series in each list were to contain an isolate, and the position of the isolate (from 6 through 10).

List F contained 40 series of 15 words. The first 10 series were used for the "No strategy instruction" condition. Eight of them contained an isolate. The next 15 series were used for the "Rote strategy" condition, and 11 of them contained an isolate. Finally, the last 15 series were used for the "Elaborative strategy" condition, and 11 of them also contained an isolate. Lists A, B, C, and D, were composed of 20 series each (15 of which were isolated). The pair of lists (A-B or C-D) used in each session (second or third), and the order of strategy instructions were counterbalanced among subjects.

Six practice series were also constructed, according to the same rules.

Words in the practice series were not included in any of the experimental lists. Two practice series (one Isolated and one Control) were used for each of the instruction conditions.

Words in each series were presented sequentially, for 250 ms each, with a 2 s interval between words. Non-isolated words were formed by 12mm X 12mm letters (word length 36mm to 84mm, visual angle 2.25 to 5.25 degrees). Isolated words were larger, and were formed by 20mm X 20mm letters (word length, 60mm to 140mm, visual angle 3.75 to 8.75 degrees). Size differences were easily discriminable, and all the subjects reported they could easily read words displayed in either size in the time provided.

Procedure.

The subject was seated in an air conditioned unshielded room in front of a Hewlett Packard (HP) CRT display (#1310A). The recording and control apparatus were located in an adjacent room.

We designed an experiment with 3 sessions around the mnemonic strategy manipulation. The first session was run from 9 to 26 days (Median = 17 days) before the second session. The second session was run 1 to 5 days before the third (Median = 2 days). The first session was intended to be a baseline and training session. For this reason it had a different experimental design, and there was a longer interval between the first and the second session than between the other two sessions. The general procedure and instructions were similar to those used by Karis et al. (1984), with the exception of the strategy instructions. During each session the subjects were presented with 40 series of 15 words each, and were asked to memorize as many words as they could and to write them down at the end of each series (free recall task). The subject was given a clipboard and recall sheets (one per series) on which to write the words after each 15-word series was completed. A 5-second pause was interposed at the end of each series, during which the subject was instructed to get the recall sheet for that series. At the end of this pause a small light attached to the clipboard was turned on, signaling the subject to pick up the pen and start writing. Removal of the pen from its holder activated a switch monitored by the experimenter, so that the subject could not begin writing prematurely. Fifty-five seconds were provided for the free recall, and all subjects reported that this interval was sufficient. The writing light was then turned off to indicate that the recall period had ended. After a verbal warning ("ready?") from the experimenter (via an intercom), another series was presented. This paradigm is depicted in figure 1.

Insert figure 1 about here

In the <u>first session (baseline and training session)</u> ERPs were not recorded. Two practice series were given to the subject at the beginning of the session. The first one contained an isolated item, while the second was a control list. After the isolated series, the subjects were asked if they had noticed that one word was larger than the others, and were told that occasionally a word would appear larger, but that they should attend to all the words, and ignore size differences between words. No strategy instructions were given to the subject before the first 10 word series. At the end, the subjects were interviewed, and this provided information about their spontaneous use of mnemonic strategies. The next 15 series were preceded by instructions to use rote strategies, and by two additional practice series. The final 15 series were preceded by elaborative strategy instructions and by two more practice series. This manipulation provided information about the subject's ability to use both strategies, and ensured that all the subjects would start the experimental sessions with the same familiarity with both strategies. The subjects were given approximately 5 minutes rest after the first 10 series and after the next 15 series.

In both the second and third session (experimental sessions) ERPs were recorded to each word, and subjects were instructed to use one strategy (either rote or elaborative) during the first half of the session (20 series), and the other during the second half (20 series). The order of instructions was counterbalanced across subjects. The subjects were given short (approximately five minutes) rest periods after every ten series. After each group of series memorized using the same strategy, the subject was asked to describe the strategies she used.

In addition to the free recall task, the third session included a counting task (oddball task), and a size-recall test. The oddball task and the size-recall test were unexpected. The size-recall test is described in more detail below. The oddball task will not be further discussed, because it served mainly as a "filler" task between the free-recall and the size-recall.

The EEG data were acquired whenever a stimulus word was presented on the HP screen (i.e., during the free recall and the oddball tasks). For the free recall, the stimulus duration was 250 ms, the inter-stimulus interval was 2000 ms and the recording epoch was also 2000 ms, beginning 100 ms prior to stimulus onset.

Strategy Instructions for the Free Recall Task. The strategy instructions were the crucial independent variable in this experiment. To test the effectiveness of our strategy instructions we ran a pilot study with 26 subjects. The subjects were divided in two groups, and each group was run in a collective session (no ERPs were recorded in this session). Eleven series of 15-words each were read out loud by D. K., at the rate of 1 word every two seconds. The reading rate was maintained constant with the help of a metronome. A 5-second pause followed the reading, and then 50 seconds were allowed for free recall. Before the first three series, the subjects were instructed to recall as many words as they could, and no particular mnemonic strategy was suggested. Then, before the next four series, instructions were given to use either rote or elaborative strategies. Finally, the subjects were instructed to use the other strategy before the last four series. The order of instructions was counterbalanced between the two groups of subjects. Note that the interest of this pilot study was to test the effectiveness of our instructions to manipulate recall performance. Therefore, no isolated items were included in the lists.

We found that our strategy instructions were effective in influencing recall performance. The subjects recalled, on average, 51% of the words when no strategy instructions were given. They recalled 42% of the words under rote instructions, and 55% under elaborative instructions. The difference between rote and elaborative instructions was significant, $F(1,25) = 25.67$, $p < 0.0001$.

In the present experiment the instructions were recorded on tape, in order to minimize variations from session to session and from subject to subject. The subjects were told that they could interrupt the tape if they had questions, and were debriefed at the end of the recording to make sure they understood the instructions correctly.

The strategy intructions are reported in the Appendix. Rote strategy instructions required the subjects to repeat the words silently to themselves, in any way they chose. Elaborative strategy instructions required the subjects to connect or organize the words, by making sentences, or forming images or pictures with them. The strategy instructions and the practice series were repeated in each session, as a reminder to the subjects, and as a way to keep the sessions as standard as possible.

Even though our instructions had proven effective, we felt that an ongoing report on the subjects' use of strategies would be useful, in order to spot temporary shifts in strategy use. Therefore we added an additional check of the subjects' use of strategies. On top of each recall sheet there were two boxes, one with a "Y" and one with an "N." The subjects were asked to indicate, by checking one of the boxes for each 15-word series, whether or not they had been able to use the required strategy. In session 2 and 3 only those series for which the subjects reported (by marking the "Y" box) having used the strategy required by the instructions were used in the computation of the recall performance and the VRI. For the 10 subjects, there were 31 series out of 800 presented in which the "N" box was marked (3.9%). Eighteen of them were series for which elaborative instructions had been given (4.5% of the series presented under elaborative instructions) and 13 of them were series presented under rote instructions (3.3% of all the series presented under rote instructions). Five of the 10 subjects did not check any "N" box. Of the remaining 5 subjects only one (subject 3) checked a high number of "N" boxes (18 series). Note that the number of lists checked "N" by this subject amounts to more than half of the total number of series checked "N" by the entire group of 10 subjects (31 series). The 18

"N" series marked by subject 3 correspond to the 22.5% of all the series presented to her (25% under elaborative instructions and 20% under rote). It is also interesting to note that, for the "N" series, the overall recall performance of this subject was higher for rote (55%) than for elaborative (40%) instructions, thus suggesting that the subject was indeed using the opposite strategy from the one suggested. Debriefing confirmed this interpretation. For the remaining five subjects the percentage of lists excluded never exceeded 8% of the total number of lists presented in the two experimental sessions.

Size-Recall Test. In order to determine whether subjects could recall which words were isolated, we devised a size-recall test. The subject was presented with a printed list of all the isolates presented in that session (half under rote and half under elaborative instructions) randomly interspersed with an equal number of non-isolated words (half from each strategy instruction). Isolates and non-isolates presented in the size-recall test were matched for position. The subject was told to indicate whether or not each word had been originally displayed in large size. An oddball task was interposed between the free recall task and the size-recall test.

Results

Data from the first session (baseline and training) were analyzed separately from those from the second and third session (experimental sessions). Data from the second and third session were also examined separately at first, to make sure that combining them was not artifactual. Given that the data obtained in these two sessions were comparable, they were combined for all the analyses.

Analysis of Recall.

As in the Karis et al. (1984) study, we computed two indices to summarize the subjects' performance in the free recall task: a measure of the magnitude of the von Restorff effect (von Restorff Index, or VRI) and an index of the overall recall performance (P). Both indices were computed using the words the subject recalled during the free recall tests. The two indices were computed separately for each strategy condition. VRI and P were computed as follows:

VRI = percentage of isolated words recalled (position 6-10) - percentage of non-isolated words recalled (position 6-10)

P = overall percentage of words recalled from all positions (isolates and non-isolates)

Non-isolated words from both Isolated and Control series were used to compute the VRI. Only words originally presented in position 6 through 10 were used in the denominator of the VRI (in order to match the positions of isolates and non-isolates). In the computation of the overall recall performance all the words were used.

First session. The VRI and P for each subject in the first session as well as means and SDs for each instruction condition are presented in Table I.

Insert table I about here

When strategies were not manipulated, there was little variability in the overall performance of the subjects, less than that reported in the Karis et al. (1984) study. On the other hand, the VRI was highly unstable in this condition, perhaps because it is computed from only 8 isolates, with a difference in the recall of one isolate accounting for a 12% difference in recall.

Both performance and the VRI were affected when strategy instructions were given to the subjects. Subjects recalled significantly fewer words for rote than for elaborative strategies, $F(1,9) = 30.65$, $p < .001$. The VRI was higher for rote strategies than for elaborative, even though this difference does not reach significance, $F(1,9) = 4.69$, $p = .059$.

Sessions 2 and 3. The performance and VRI of each individual subject in session 2 and 3, as well as the mean and SD for each strategy instruction are reported in table II.

Insert table II about here

The two strategy instructions conditions differed significantly in both performance and VRI. Recall performance was lower for rote than for elaborative instructions, $F(1,9) = 18.84$, $p < .002$. The VRI was higher for rote than for elaborative instructions, $F(1,9) = 16.54$, $p < .003$.

Table II reveals that only one subject (subject 8) showed a lower VRI under rote than under elaborative instructions. Given that the behavior of this subject was, for some reason, different from that of the other subjects, the waveforms of this subject might reflect different processes from those of the other subjects. Therefore, the analyses presented in the following paragraphs were performed on the nine remaining subjects.

However, the analyses were also performed on the full sample of ten subjects and any differences will be discussed.

The serial position curves averaged over nine subjects are shown in Figure 2 for both instruction conditions.

Insert figure 2 about here

These curves show the average percentage of words recalled in session 2 and 3 for each of the 15 positions. Data for isolated and non-isolated words are plotted separately. Note that the isolates are represented as a horizontal line, which actually indicates the average percent recall of the isolates from position 6 through 10. For both strategies there is a "primacy," as well as a "recency" effect in recall performance. The magnitude of the von Restorff effect in each condition is represented by the elevation of the curve for the isolated items relative to the elevation of the curve for the non-isolates in comparable positions (6 through 10). Clearly, the von Restorff effect is large under rote instructions, while it is almost absent under elaborative instructions, while performance is higher under elaborative than under rote instructions.

Figure 3 compares the results obtained in the present study with the results reported by Karis et al. (1984).

Insert figure 3 about here

The average von Restorff index is plotted in figure 3 against the average overall performance for each strategy condition. The circles represent rote strategies, and the triangles elaborative strategies. The means from the

Karis et al. (1984) study are indicated by filled symbols, connected by a solid line. The means for the subjects in the present study are indicated by empty symbols, connected by a dashed line. Note that the means from the Karis et al. (1984) study were computed on two different groups of subjects, who, during a post-experimental debriefing, reported having used different strategies, while in the present study the same subjects performed in both conditions. The two studies yield very similar data, even though the differences between mean values are smaller in the present study than in the Karis et al. (1984) one. In the Karis et al. (1984) study the subjects chose the strategies they preferred, so one would in fact expect more extreme values than in the present study.

From the data reported so far we can therefore conclude that our strategy manipulation was successful in affecting both recall performance (more words were recalled by using elaborative strategies) and the VRI (which was higher under rote instructions), in the manner predicted by the Karis et al. (1984) model. In the section on ERP waveforms we will examine the interaction between the strategy instructions and the amplitude of P300.

Session Effects. The data reported above were combined across the second and the third session. The data were also examined separately for the two experimental sessions, and yielded comparable results. There was a main effect of strategy instructions on performance. In both sessions subjects recalled more words when using elaborative strategies than when using rote, $F(1,9) = 17.24$, $p < .003$. There was also a main effect of strategy instructions on the VRI: in both sessions the VRI was higher for rote than for elaboative instructions, $F(1,9) = 16.37$, $p < .003$. In addition, there was a main effect of session on performance, with subjects' overall performance improving from session 2 to session 3, $F(1,9) = 7.05$, $p <$ .03. However, there were no significant interactions between session and strategy instructions, thus supporting the legitimacy of combining the values from the two sessions.

Analysis of ERP Waveforms.

The ERP data for each subject were also combined across the two experimental sessions (session 2 and 3).

Isolation effect. In order to find out whether isolates elicited larger P300s than non-isolates, and whether isolates and non-isolates elicited similar waveforms for the two strategy instruction conditions, we first compared the waveforms elicited by the isolates with the waveforms elicited by the non-isolates in the same positions (6 through 10). The EEG records of each subject were sorted according to strategy instructions (rote or elaborative), to the word class (isolates, non-isolates in isolated series and non-isolates in control series) and to their position (position 6-10 only). The grand average waveforms at three electrode locations (Fz, Cz, and Pz) are plotted in figure 4. In this figure, non-isolates coming from series containing an isolate and from control series are plotted separately.

Insert figure 4 about here

A positive peak is visible in all the waveforms. We label this positive peak "P300" because its peak latency (in excess of 300 ms) and its scalp distribution are characteristic of the P300 component of the ERP. Amplitude values are positive at the three electrode locations, with Pz more positive than Cz, and Cz more positive than Fz. Isolates elicited larger P300s than non-isolates, while non-isolates from isolated and control series elicited very similar waveforms. This result is consistent with the general observation that task relevant, distinct stimuli elicit a larger P300 than do companion stimuli that are common (e.g., Duncan-Johnson & Donchin, 1977). It can also be noted that the waveforms for rote and elaborative instructions are similar.

Memory effect. Given that isolates did elicit P300s we proceeded to determine if there was the expected relationship between the strategy instructions, the amplitude of the P300 and the subsequent recall. For this analysis the EEG records of each subject were sorted for averaging on the basis of strategy instructions (rote or elaborative), word type (isolates, non-isolates in isolated series and control series), position (position 6-10, other positions) (footnote 6) and subsequent recall (recalled, not-recalled).

Average waveforms at Fz, Cz and Pz for isolates recalled and not recalled are presented in Figure 5 for both strategy instructions.

Insert figure 5 about here

We note two primary aspects of these waveforms. First, under rote instructions there is a larger difference in P300 between isolates recalled and not recalled than under elaborative instructions, which is most evident at the central and parietal electrodes. Second, there is a difference between isolates recalled and not recalled at the frontal electrode, most evident for elaborative instructions, which suggests the presence of a frontal-positive slow wave.

These impressions were supported by means of statistical analysis on measures of P300 amplitude, and of area measures of the frontal-positive slow wave, taken on the waveforms of each individual subject. The amplitude of P300 was estimated by using a cross-covariance procedure (Fabiani, Gratton, Karis, & Donchin, in press). First, the P300 component was identified at Pz as the segment with maximal cross-correlation with a template (2 Hz, 1 cycle, inverted cosinusoidal wave), in a time window ranging from 350 to 800 ms. Then, P300 amplitude was assessed by computing the covariance between the template and the segment of Pz waveform identified as P300. Fabiani et al. (in press) demonstrated that the use of cross-covariance is particularly appropriate when relatively few trials are entered in the computation of the averages. An analysis of variance was applied to the P300 amplitude measures to test the differences in amplitude over different experimental conditions. A repeated measures design was used (ALICE statistical package, program "ANOVA," Grubin, Bauer and Walker, 1976). This analysis confirmed that isolated words elicited larger P300s than other word types (main effect of word position), $F(14,112) = 3.99$, $p < .0001$. In addition, words that were recalled under rote instructions (regardless of their type, and position) elicited larger P300s than words not recalled, while there was no difference between words recalled and not recalled under elaborative instructions (strategy x memory interaction), $F(1,8) = 5.72$, $p < .05$ (Footnote 7). The grand averages of the waveforms for all subjects and positions, latency adjusted on the basis of the P300 peak are shown in figure 6 for both strategy conditions. An mplitude difference between words recalled and not recalled is noticeable under rote, but not under elaborative instructions.

Insert figure 6 about here

As we mentioned, a positivity at the frontal electrode was evident in the waveforms elicited by the words subsequently recalled. This "frontal-positive slow wave" is more positive frontally than centrally and parietally and it is more visible when subjects are given elaborative strategy instructions. To measure this slow positivity we used area measures taken at Fz for each subject, word type and condition. Area measures were particularly appropriate given the absence of a clearly defined peak for this component.

An analysis of variance, with the same design used for the P300 amplitude measures, was applied to the area measures of the frontal-positive slow wave. This component was larger when subjects were instructed to use elaborative strategies than when they were instructed to use rote strategies (main effect of strategy) $F(1,8) = 13.09$, $p < .01$. This component was also larger for recalled than not recalled words (main effect of memory) $F(1,8) = 57.90$, $p < .001$.

Analysis of the size-recall test.

The size-recall test was devised to test the hypothesis that the subjects would have a better memory of the size at which the word was originally displayed when the words had been memorized under rote than under elaborative instructions. Even though the subjects were told that there were an equal number of words originally displayed in regular and large size on the printed list they received, most subjects were biased toward responding "regular" rather than "large." Therefore, an unbiased measure of accuracy was used in order to test our hypothesis. The accuracy in identifying words originally displayed as large (isolates) was given by:

> Percentage of large-size words correctly identified divided by the total number of "large" responses given.

The accuracy in identifying words originally displayed in "regular" size (non-isolates) was computed in an analogous fashion by:

> Percentage of regular-size words correctly identified divided by the total number of "regular" responses given.

These accuracy indices were computed separately for the two strategy instructions conditions. Our hypothesis was confirmed: subjects were more accurate in recalling the original size of the words presented to them under rote instructions (72%) than under elaborative instructions (58%) - main effect of strategy, $F(1,8) = 7.28$, $p < .03$. The subjects also tended to be more accurate in identifying large words (isolates) than regular-size words (non-isolates), even though this result was not significant (71% of the isolates vs. 59% of the non isolates) - $F(1,8) = 4.81$, $p = .06$. The results of the size-recall test are depicted in figure 7.

Insert figure 7 about here

Discussion and Conclusions

In this paper we have reported data from a study concerning the interaction between encoding, rehearsal strategies and retrieval of verbal material. Traditionally, these processes have been investigated by studying the effects of some experimental manipulations (supposed to selectively affect some of these processes) on final-output measures, such as recall performance, accuracy and speed of recognition, subjective reports, etc. Even though this approach has shown great power in clarifying the nature, and the course of processes affecting memory, it remains the case that these procedures provide but an indirect observation of the processes that intervene between the stimuli and the responses.

The data presented in this paper serve to illustrate the mannner in which the endogenous components of the ERPs, and in particular the P300, may be used in conjunction with observations and overt responses to augment the analysis of human cognitive functions. The use of the endogenous components of the ERP in the study of human cognitive functions is based on the assumption that ERP components are the manifestation of specific psychological processes (Donchin, 1979). In particular, Donchin (1981) proposed that the P300 component of the ERP manifests the updating of schemas in working memory. Recent studies indicate that there are aspects of the encoding and storage system that can be revealed by augmenting final-output studies with process-oriented measures like the P300 (Fabiani, et al., 1986; Johnson, et al., 1985; Karis, et al., 1984; Klein, Coles, & Donchin, 1984; Paller, et al., 1985).

References

Baddeley, A.D. (1981). The concept of working memory: A view of its current state and probable future development. Cognition, 10, 17-23.

Baddeley, A.D., & Hitch, G.J. (1974). Working memory. In G.A. Bower (Ed.), The Psychology of Learning and Motivation, (Vol. 8). New York: Academic Press.

Cimbalo, R.S. (1978). Making something stand out: The isolation effect in memory performance. In M.M. Grunnenberg, P.E. Morris, and R.N. Sykes (Eds.), Practical aspects of memory . New York: Academic Press.

Dickinson (1980).

Donchin, E. (1979). Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), Evoked Brain Potentials and Behavior . New York: Plenum Press, 13-75.

Donchin, E. (1981). Surprise!...Surprise?, Psychophysiology, 18, 493-513.

Donchin, E., & Heffley, E. (1975). Minicomputers in the signal-averaging laboratory. American Psychologist, 38 , 449-461.

Donchin, E., Gratton, G., Dupree, D., & Coles, M. G. H. (in press). After a rash action: Latency and amplitude of the P300 following fast guesses. In G. Galbraith, M. Klietzman, & E. Donchin (Eds.) Neurophysiology and Psychophysiology: Experimental and Clinical Applications. Hillsdale, NJ: Erlbaum.

Donchin, E., Karis, D., Bashore, T. R., Coles, M. G. H., & Gratton, G. (in press). Cognitive psychophysiology and human information processes. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), Psychophysiology: Systems, Processes, and Applications. Volume I: Systems. New York: Guilford Press.

Donchin, E., Ritter, W., & McCallum, C. (1978). Cognitive Psychophysiology: the endogenous components of the ERP. In E. Callaway, P. Tueting, and S. Koslow (Eds.), Brain Event-Related Potentials in Man . New York: Academic Press, 349-441.

Duncan-Johnson, C.C., & Donchin, E. (1977). On quantifying surprise: The variation in event-related potentials with subjective probability, Psychophysiology, 14 , 456-467.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (in press). P300: Theoretical and methodological issues. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.) Advances in Psychophysiology. Volume 2. Greenwich, CT: JAI press, Inc.

Fabiani, M., Karis, D., & Donchin, E. (1986). P300 and recall in an incidental memory paradigm. Submitted for publication.

Fitzgerald, P.G., & Picton, T.W. (1981). Temporal and sequential probability in evoked potential studies. Canadian Journal of Psychology, 35, 188-200.

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.

Gratton, G., Coles, M. G. H., & Donchin, E. (1986). A multivariate approach to the analysis of the scalp distribution of event-related potentials: Vector Analysis. Submitted for publication.

Grubin, M.L., Bauer, J.A., Jr., Walker, E.C.T. (1976). Alice User's Guide, Alice Associates, 29 Wellesley Avenue, Natick, MASS, 01760.

Heffley, E.F. (1981, April) Event-related brain potentials in visual monitoring tasks: Selective attention, display format, and event frequency. Paper presented at the Eastern Psychological Association Annual Meeting, New York City, NY.

Jasper, H.H. (1958). Report of the committee on methods of clinical examination in electroencephalography, Electroencephalography and Clinical Neurophysiology, 10 , 370.

Johnson, R.E. Jr., & Donchin, E. (1978). On how P300 amplitude varies with the utility of the eliciting stimuli. Electroencephalography and Clinical Neurophysiology, 44, 424-437.

Johnson, R., Pfefferbaum, A., & Kopell, B. S. (1985). P300 and long-term memory: latency predicts recognition performance. Psychophysiology, 22, 497-507.

Karis, D., Fabiani, M., & Donchin, E. (1984) "P300 and memory: individual differences in the von Restorff effect. Cognitive Psychology, 16, 177-216.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time, Science, 197 , 792-795.

Mackintosh (1975).

Mandler, G. (1979). Organization and repetition: Organizational principles with special reference to rote learning. In L.G. Nilsson (Ed.), Perspectives on Memory Research. Hillsdale, N.J.: Erlbaum.

Mandler, G. (1980). Recognizing: The judgment of pervious occurrence. Psychological Review, 87, 352-271.

Munson et al. (1984).

Ohman, A. (1979). The orienting response, attention, and learning: an information processing perspective. In H. D. Kimmel, E. H. Van Olst, and J. F. Orlebeke The orienting reflex in humans. Hillsdale, N.J.: Erlbaum.

Paller, K. A., Kutas, M., & Mayes, A. R. (1985). An investigation of neural correlates of memory encoding in man. Psychophysiology, 22, 607.

Pearce & Hall (1980).

Pritchard, W. (1981). Psychophysiogy of P300. Psychological Bullettin, 89, 506-540.

Rescorla & Wagner (1972).

Restorff, H. von (1933). Uber die Wirkung von Bereichsbildungen im Spurenfeld, Psychologische Forschung, 18 , 299-342.

Ruchkin, D. S., & Glaser, E. M. (1978). Some simple digital filters for examination of CNV and P300 on single trial basis. In D. Otto (Ed.), Multidisciplinary perspectives in event-related brain potentials (pp. 579-581). Government Printing Office.

Ruchkin, D.S., & Sutton, S. (1983). Positive slow wave and P300: Association and disassociation. In A.W.K. Gaillard & W. Ritter (Eds.), Tutorials in event related potential research: Endogenous components (pp. 233-250). Amsterdam: North-Holland Publishing Company.

Sokolov, E.N. (1963). Perception and the conditioned reflex. New York: Macmillan.

Sokolov, E.N. (1969). The modeling properties of the nervous system. In I. Maltzman and M. Cole (Eds.), Handbook of Contemporary Soviet Psychology. New York: Basic Books, 671-704.

Sokolov, E.N. (1975). The neuronal mechanisms of the Orienting Reflex. In E.N. Sokolov and O.S. Vimogradova (Eds.), Neuronal Mechanisms of the Orienting Reflex . Hillsdale: LEA, 217-235.

Squires, Squires & Hillyard (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. Electroencephalography and Clinical Neurophysiology, 38, 387-401.

Squires, K.C., Wickens, C., Squires, N.K., & Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. Science, 193 , 1142-1146.

Sutton, S., Braren, M., Zubin, J., & John, E.R. (1965). Evoked-potential correlates of stimulus uncertainty. Science, 150, 1187-1188.

Toglia, M.P., & Battig, W.F. (1978). Handbook of Semantic Word Norms, Hillsdale: LEA.

Wagner, A.R. (1976). Priming in STM: An information processing mechanism for self-generated or retrieval-generated depression in performance. In T.J. Tighe, and R.N. Leaton (Eds.), Habituation perspectives from child development, animal behavior, and neurophysiology . Hillsdale: LEA.

Wagner, A. R. (1978).

Wallace, W.P. (1965). Review of the historical, empirical, and theoretical status of the von Restorff phenomenon, Psychological Bulletin, 63 , 410-424.

Appendix

Strategy instructions.

General instructions. This time we want you to use particular strategies to memorize the words. There are many strategies people can use to remember lists of words in experiments like this one. We want everyone to use the same strategies, because if everyone does something different we often get confusing data and don't have any idea why the results turn out the way they do. This is why we asked you what strategies you used during the first part. So, it's very important that you try to follow my instructions, and if you have trouble doing this, it's important that you tell me when I ask you afterwards. There are two boxes on top of the recall sheets, one with a Y and one with a N. This is so you can indicate whether or not you used the appropriate strategies, Y for yes, N for no. You must choose one. Pick the one that best represents what you did. If you used primarily the appropriate strategy, check the Y box. Otherwise, check the N box. I want you to remember as many words as you can, of course, but it's even more important that you follow the strategies I describe. If this makes it harder for you to recall the words, that's okay, because the most important thing is to use the correct strategy. You'll be using two different strategies, and when I describe the second you may want to keep on using the first, or you may want to use some other strategy that you think is better. Please don't! Try very hard to do what I suggest. If that's very hard I want you to tell me, because that's another source of information that may be helpful to us when we analyze the data. Honesty, of course, is crucial, and it's important that you accurately describe what you do. One more thing: sometimes you may recall a word but not be sure whether it came from the last list or from some earlier list. Or you may recall a word from a previous list that you didn't write down before. In these cases you should write down the words on the sheet you are using.

Rote Rehearsal Strategy Instructions. I want you to repeat the words silently to yourself. You can repeat each word until the next appears, or you can repeat both the word presented and some of the previous words. For example, if the first four words in a list are kid, green, fit, and bank, you could repeat each word three times to yourself before the next was presented - kid, kid, kid ... green, green, green ... fit, fit, fit .... Or, if you wanted, you could also repeat the previous words: kid, kid, kid; green, kid, green, kid; fit, green, kid, fit, green, kid... Of course, after several words have been presented, you will not be able to repeat them all before the next appears. That's OK, of course. There are many other ways to repeat the words. For example, repeating the last word, along with just one of the previous words: kid, kid, kid ... green, green, kid ... fit, fit, green ... bank, bank, fit .... You don't have to stick to one of these methods, of course. You may want to combine them, or change from one to another. The important thing is to repeat the words, in whatever way you choose. You should not talk out loud, however, or even move your lips and mouth, because it's possible that such movements could interfere with our recording.

Elaboration rehearsal strategy instructions. I want you to try to connect or organize the words in some way. You can make sentences out of them, you can form images or pictures with them, you can make a story out of them, you can assign them to different categories, or use other methods that occur to you. It is very difficult, of course, to use all 15 words together, e.g., in the same story, but you don't have to. You can have several groups of a few words, or one or two larger groups, or whatever seems easiest. You may not be able to use every word. That's okay, just try to do the best you can. You don't have to combine words that were presented together. It's okay to combine early words with later words, or words from one part of the list with words from another. Let's say a list started like this: kid, green, fit, bank, horse, during, final, sun, feed, mend, took, spend, tend, fez. (Experimenter holds up card with all the words on it.) Now, I didn't pick these words at random. I chose them to illustrate some of the many methods that can be used to remember words. The lists you will see, however, are composed of words picked at random from a master list.

There are no correct or incorrect ways to do this. Whatever helps is okay. Here are some examples. You might create a sentence: The kid rode the green horse to the bank during the sun-day fair. Or you could imagine that scene, or use both the sentence and the scene. Combining this many words is difficult. You could just imagine a green horse or a kid having a fit, or a kid turning green during a fit, and so on. You also might combine words that start with the same letter or letters: fit, final, feed, fez, or words that rhyme: mend, spend, tend, although, if possible, it is better to connect or organize based on meaning, rather than sound or letters. Some people find it helps to form images or pictures of scenes containing the objects that must be remembered. Many words, of course, are not easy to visualize. These might be combined with other words that are, or made into a sentence, or grouped in ways that don't require visualization. For example, the word "guilty" is hard to visualize by itself. However, if the words "dog" and "steak" were also presented somewhere in the list, you could think of, or visualize, a dog, looking guilty after stealing a steak. The important point should be clear - if it's not, tell me. I want you to do more than just repeat the words to yourself. By doing more I mean such things as organizing or combining the words as I've described, or visualizing them, or combinations of these various techniques, or other similar techniques.

Footnotes

1. ERP components are subdivided into two broad categories: exogenous and endogenous. Exogenous components are early components (0-100 ms after stimulus presentation), and are responsive to the physical characteristic of the eliciting stimuli (e.g., modality, intensity, etc.). Endogenous components (including P300) are late components occurring 100 ms to several seconds after stimulus presentation. They are related to psychological processes, and are independent of the physical characteristics of the eliciting stimuli. In fact, the absence of an expected stimulus can elicit these components. See also Donchin et al. (1978), Donchin, Karis, Bashore, Coles, & Gratton (in press), and Fabiani et al. (in press) for a discussion of the endogenous-exogenous distinction.

2. For example, with two equiprobable events, A and B, the P300 elicited by the last A in a sequence will be larger in the sequence BA than AA. Similarly, in third order series, P300 amplitude will decrease from BBA to ABA to BAA to AAA.

3. "Working memory refers to the role of temporary storage in information processing" (Baddeley, 1981, p. 17; see also Baddeley & Hitch, 1974). The concept of working memory emphasizes function, while short term memory has often been used to refer to a hypothetical structure.

4. We label this component a "frontal positive slow wave" to distinguish it from the more typical slow wave distribution reported in the past (negative frontally, becoming more and more positive as one moves back across the scalp; see Squires, Squires, & Hillyard, 1975; and Ruchkin & Sutton, 1983).

5. A series of events that can be divided into discrete classes is called an "oddball" when one event (or class of events) is much rarer than the other (although sometimes even series with 50-50 probability are called oddballs).

6.

7. Both the position effect and the strategy by memory interaction were replicated with the entire 10-subject sample, even though the F values were slightly different. In addition, other measurement procedures were applied: Vector Filter (Gratton, Coles, & Donchin, 1985) was applied in conjunction with peak-picking and cross-covariance, and peak-picking measures were taken at Pz, for both the 10- and the 9-subject sample. All these analyses yielded similar results.

Table I

Performance and von Restorff Index (Percentages) in session 1 (baseline and training).

	Strategy instructions					
	No		Rote		Elaborative	
S#	P	VRI	P	VRI	P	VRI
1	58	15	40	17	64	12
2	53	37	68	4	65	-8
3	46	17	39	33	56	1
4	48	-47	28	36	46	2
5	51	10	45	55	65	22
6	42	36	32	33	50	-13
7	43	11	37	-8	41	34
8	48	24	44	20	63	12
9	49	24	38	25	63	-15
10	50	-1	47	20	59	12
M(SD)	49(5)	13(24)	42(11)	23(17)	57(9)	6(16)

von Restorff Index (VRI): % isolates recalled minus % non-isolates recalled (position 6-10)

Performance (P): % recalled from all positions

Table II

Performance and von Restorff Index (Percentages) in session 2 and 3 (expermental sessions, combined values).

	Strategy instructions			
	Rote		Elaborative	
S#	P	VRI	P	VRI
1	54	31	72	4
2	65	10	69	-1
3	40	36	49	8
4	30	16	43	-3
5	45	28	68	1
6	38	19	36	-7
7	38	12	40	10
8	41	16	64	23
9	38	15	64	10
10	48	22	67	-3
M(SD)	44(10)	20(9)	57(14)	4(9)

von Restorff Index (VRI): % isolates recalled minus % non-isolates recalled (position 6-10)

Performance (P): % recalled from all positions

Figure Captions

Figure 1. Experimental design.

Figure 2. Average serial position curves for rote strategy (2a) and elaborative strategies (2b). Data for isolated and non-isolated words are plotted separately. The isolates are represented as solid horizontal lines, indicating the average percent recall of the isolates from position 6 through 10. The curves for the non-isolates are represented by dashed lines.

Figure 3. The results obtained in the present study are compared with the results reported by Karis et al. (1984). The average VRI is plotted against the average recall performance for each strategy condition. The circles represent rote strategies, and the triangles elaborative strategies. The means from the Karis et al. (1984) study are indicated by filled symbols, connected by a solid line. The means for the subjects in the present study are indicated by empty symbols, connected by a dashed line.

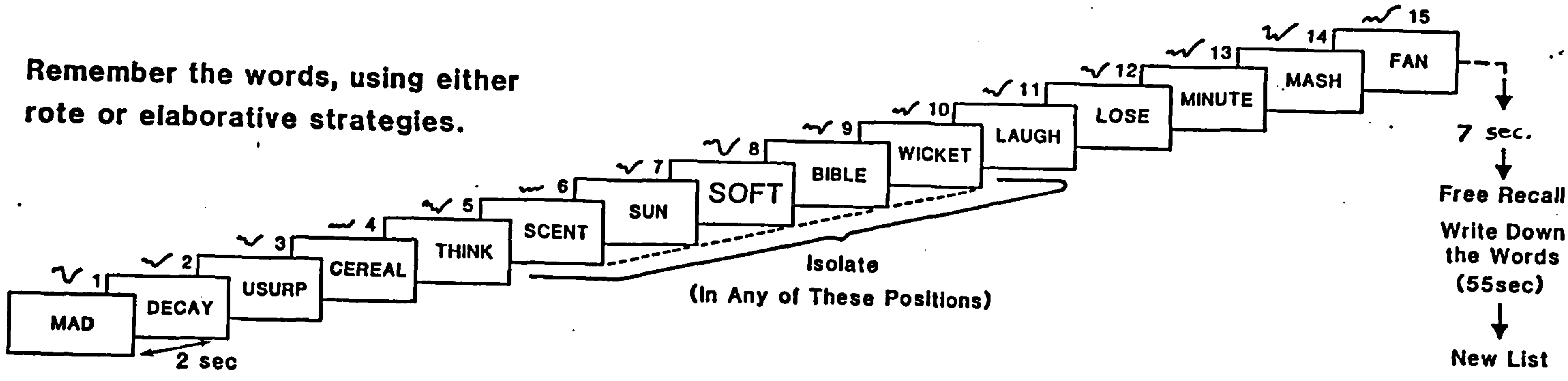
Figure 4. Grand average waveforms at three electrode locations (Fz, Cz, and Pz) for isolates (solid line), non-isolates coming from isolated series (dashed line) and from control series (dotted line). The two instruction conditions are plotted separately.

Figure 5. Grand average waveforms at Fz, Cz and Pz for isolates recalled (solid line) and not recalled (dashed line) for both strategy instructions.

Figure 6. Grand average waveforms at Fz, Cz and Pz for all subjects and positions, latency adjusted on the basis of the P300 peak for both strategy conditions. Recalled words are represented by a solid line, not recall words by a dashed line. The vertical dashed lines represent the limits of the time window in which P300 amplitude was assessed.

Figure 7. Bar graph representing results of the size-recall test. The size-recall accuracy is plotted on the ordinate. The light-gray bar represents the accuray for rote strategy, the dark-gray the accuracy for elaborative. The horizontal dashed line indicates a chance level of accuracy.

Free Recall



Session 1

(R) 15 exp. lists
5 control lists

(E) 15 exp. lists
5 control lists

Session 2

(E) 15 exp. lists
5 control lists

(R) 15 exp. lists
5 control lists

2 sec. ISI
2 sec. epoch
250 msec Stim. dur.

(R) = Rote
(E) = Elaborative

FIG. 1

Rote Strategy

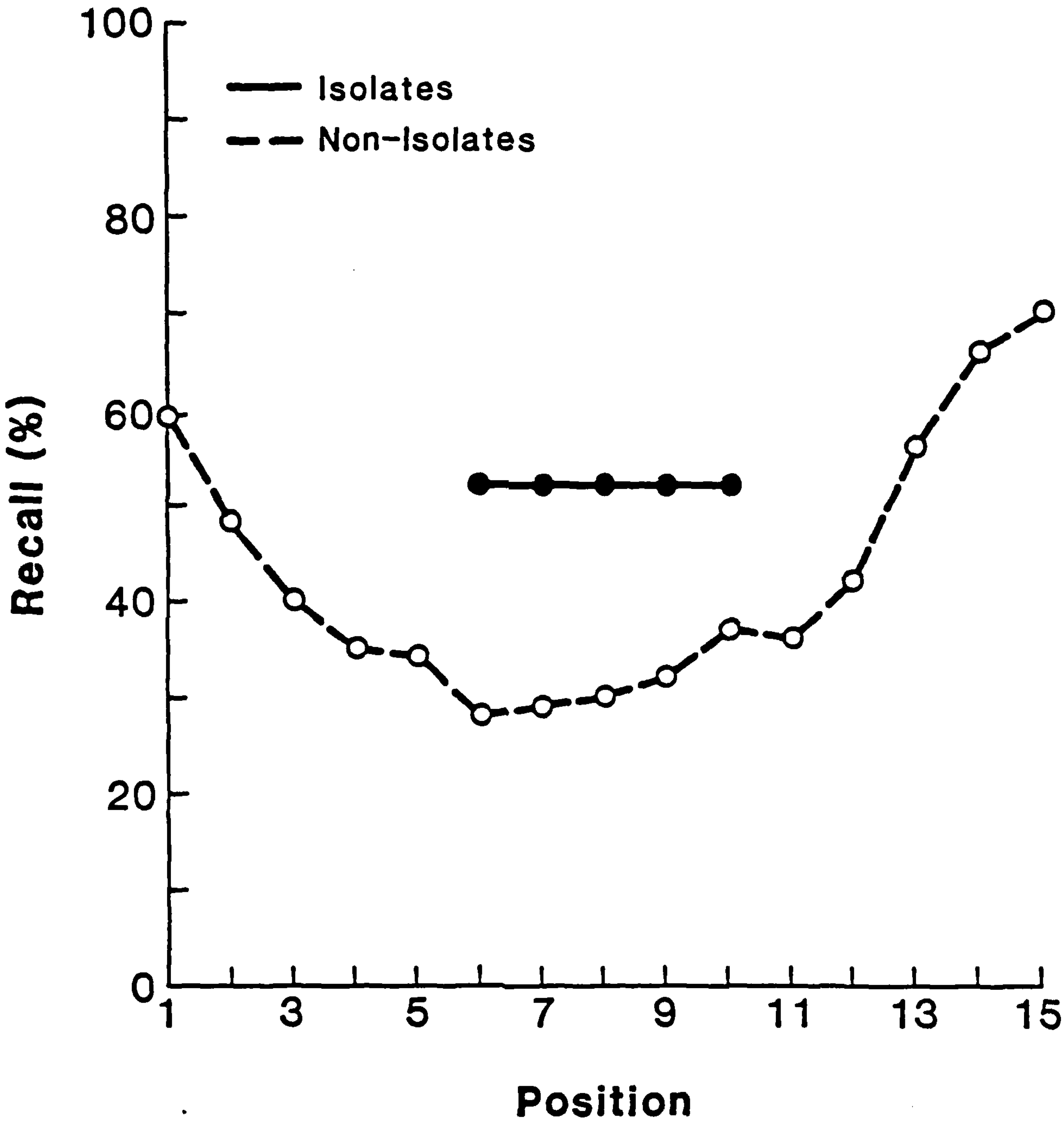


FIG 2a

Elaborative Strategy

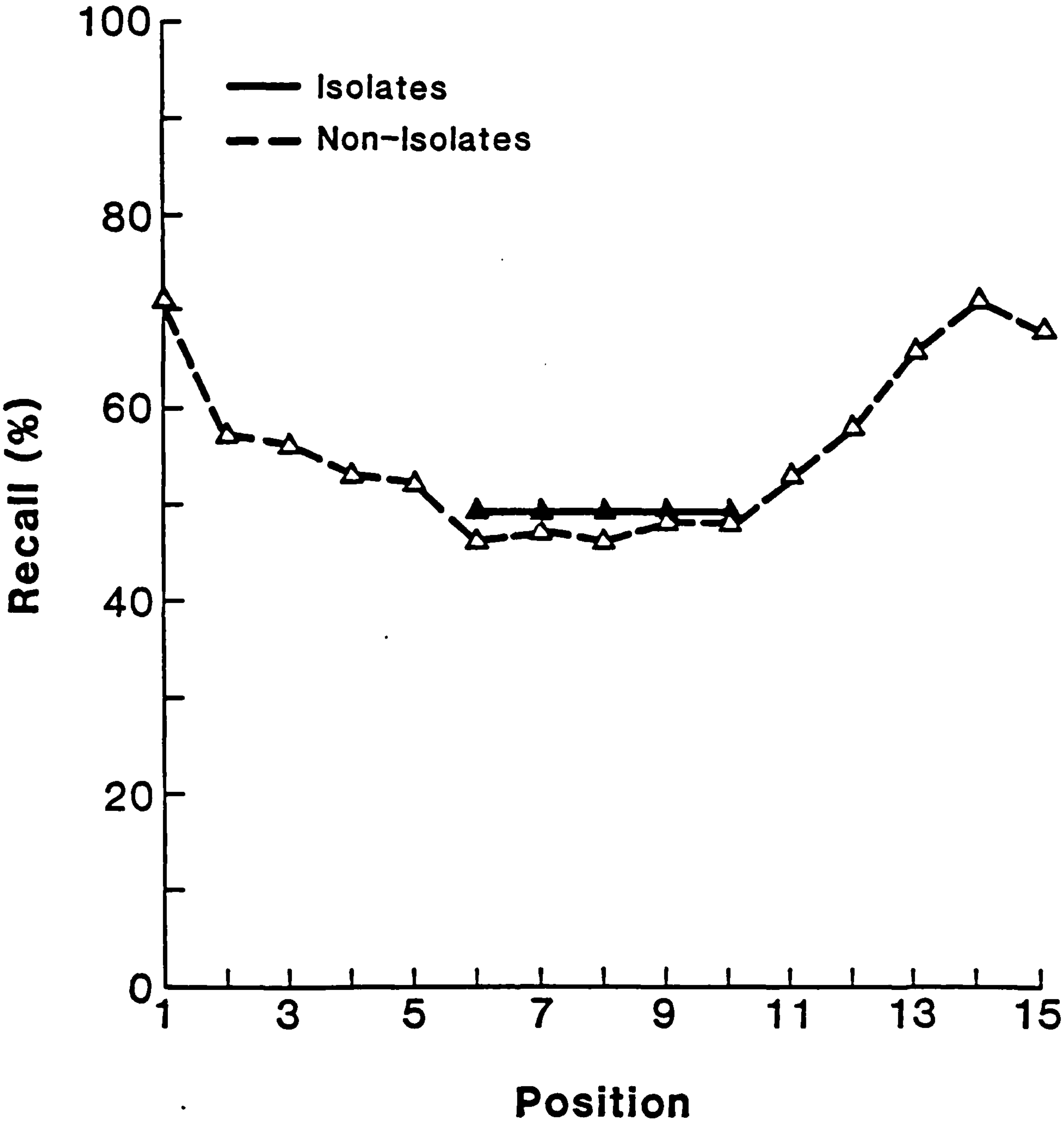
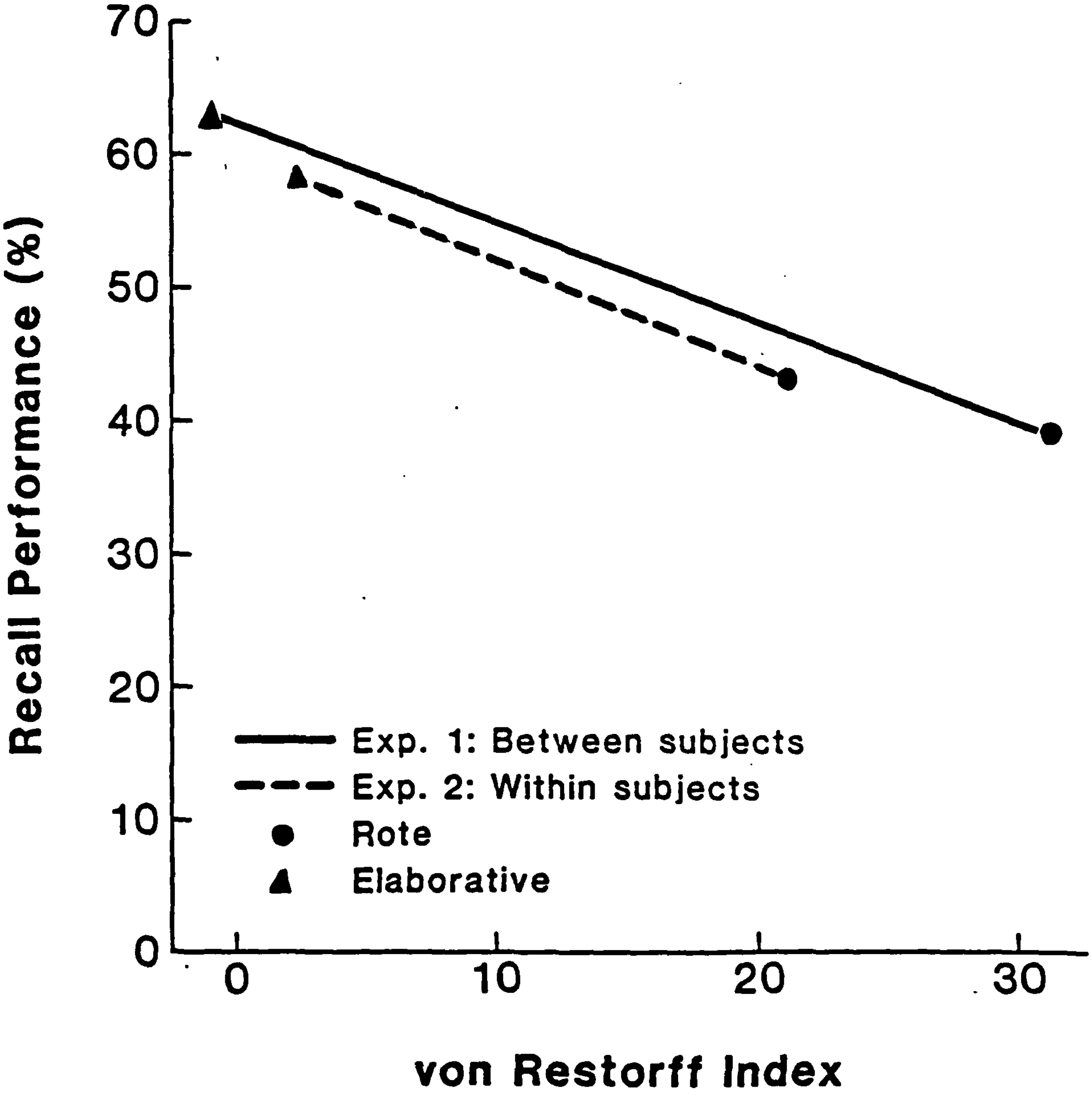
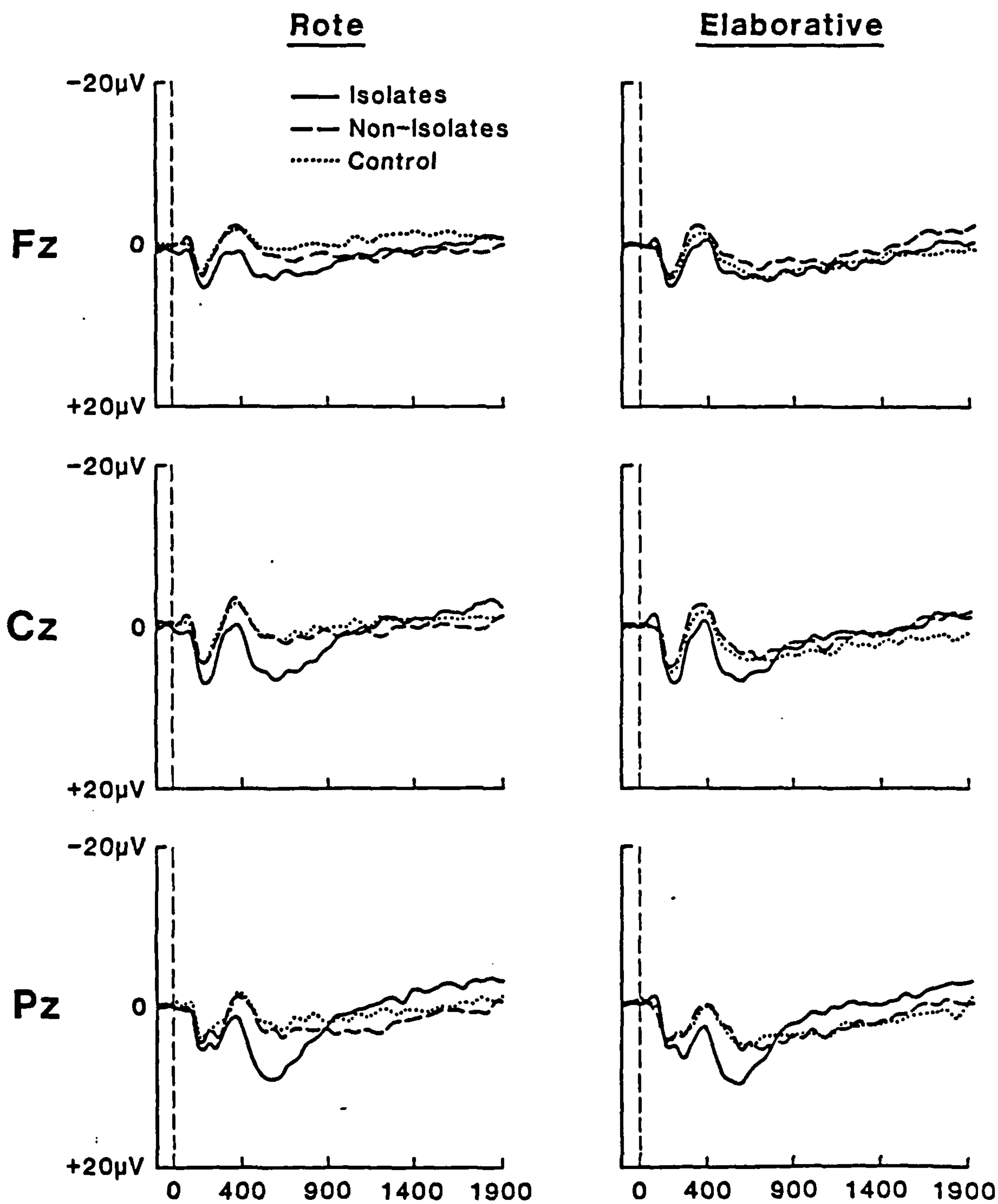


FIG 2b



70
60
50
40
30
20
10
0
Recall Performance (%)
Exp. 1: Between subjects
Exp. 2: Within subjects
Rote
Elaborative
0
10
20
30
von Restorff Index

FIG 3



Rote
Elaborative
Isolates
Non-Isolates
Control
Fz
Cz
Pz
-20μV
0
+20μV
0
400
900
1400
1900

FIG. 4

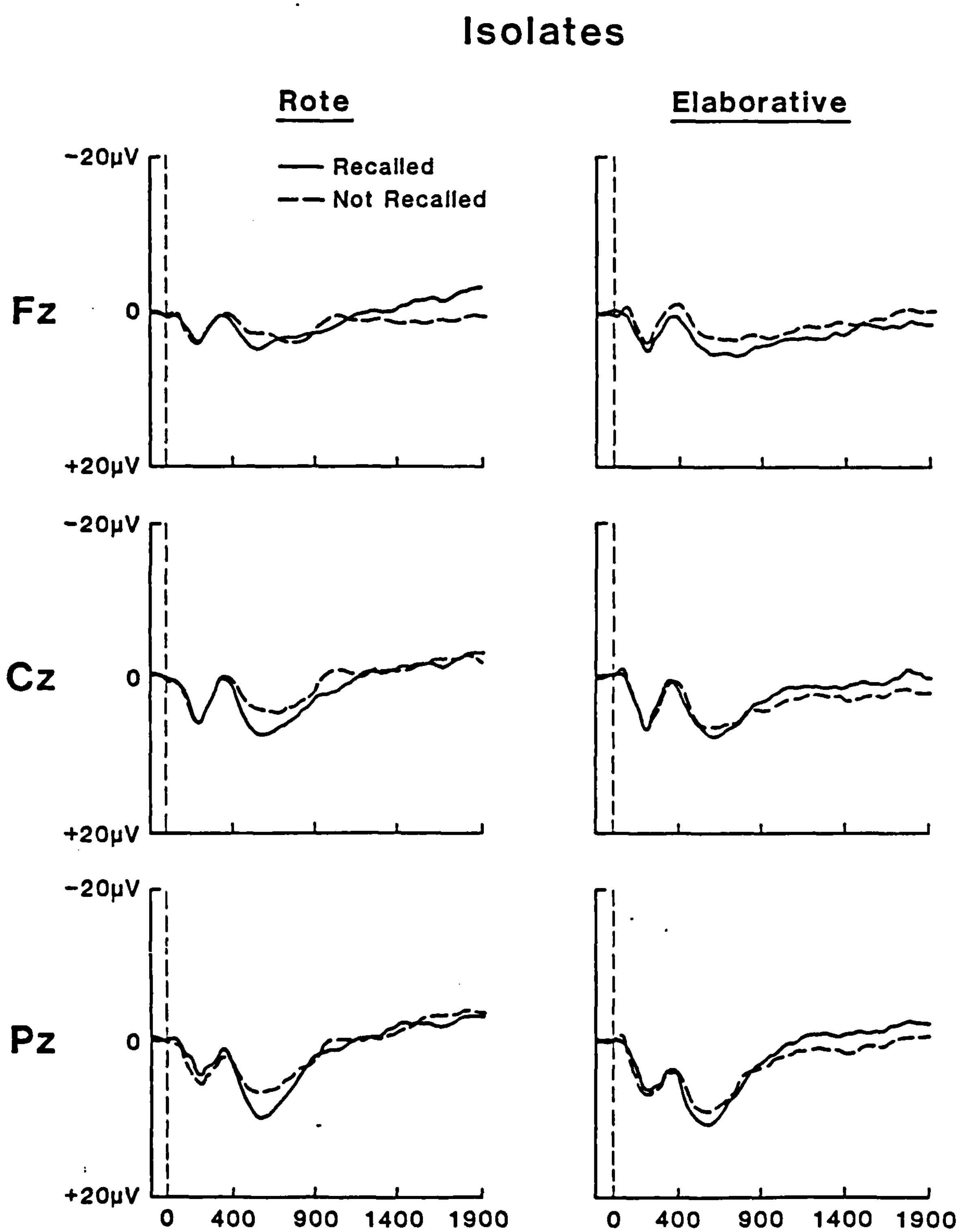
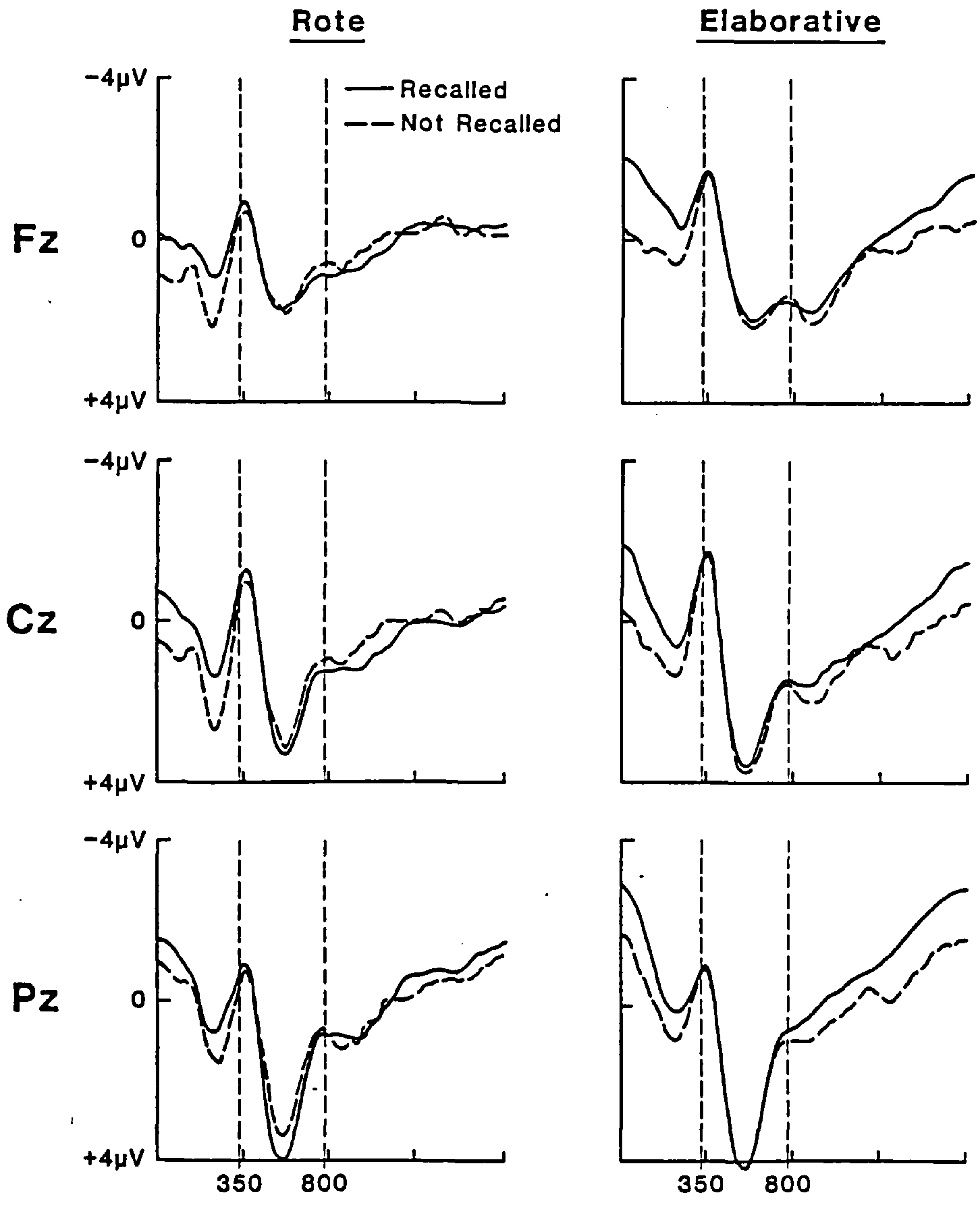
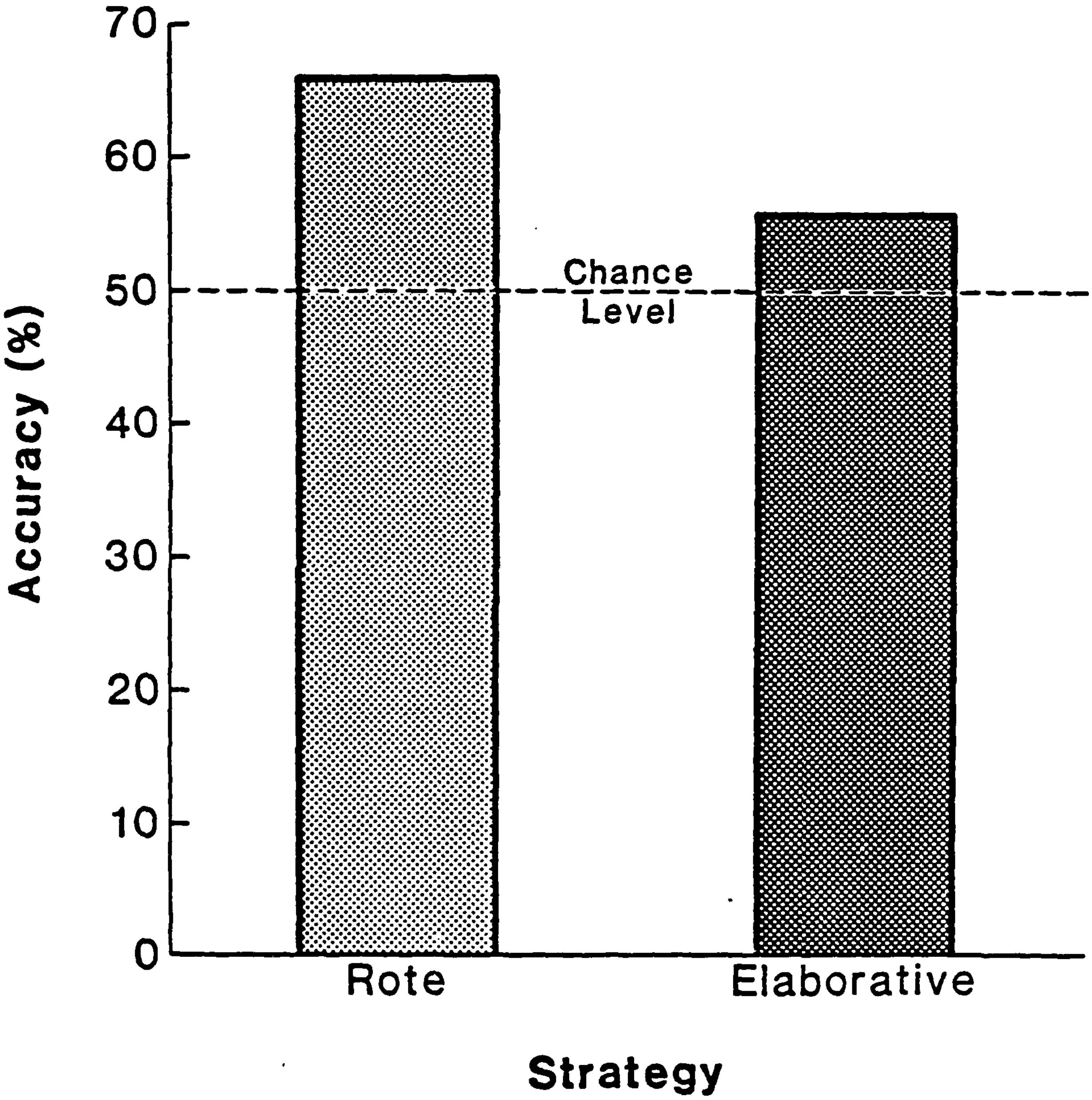

Isolates
Rote
Elaborative
Recalled
Not Recalled
Fz
Cz
Pz
−20µV
0
+20µV
0
400
900
1400
1900

FIG. 5



Rote
Elaborative
Recalled
Not Recalled
−4μV
0
+4μV
Fz
Cz
Pz
350
800

FIG. 6



70
60
50
40
30
20
10
0
Accuracy (%)
Chance
Level
Rote
Elaborative
Strategy

FIG. 7

Submitted for publication
Do not quote without permission

A MULTIVARIATE APPROACH TO THE ANALYSIS OF THE SCALP DISTRIBUTION OF EVENT-RELATED POTENTIALS: VECTOR ANALYSIS

GABRIELE GRATTON, MICHAEL G. H. COLES, AND EMANUEL DONCHIN 1

Cognitive Psychophysiology Laboratory

University of Illinois at Urbana-Champaign

Running title: Vector Analysis

Send all correspondence to:

Gabriele Gratton

University of Illinois

Psychology Department

603 E. Daniel

Champaign (IL) 61820, U.S.A.

A Multivariate Approach to the Analysis of the Scalp Distribution of Event-Related Potentials: Vector Analysis

Gabriele Gratton, Michael G. H. Coles, and Emanuel Donchin

Introduction

The Event-Related Brain Potential (ERP) is generated by the synchronous activity of ensembles of neurons. For an ERP to be recorded at the scalp, the spatial organization of the neurons must allow their individual fields to summate, relative to the scalp electrode (see Wood & Allison, 1981), and their temporal relationship to external or internal events must allow their consistent recording over different trials.

In general, it has been found useful to consider the ERP as an ensemble of "components." The concept of component derived originally from an analysis of the morphology of the waveforms of the ERP. The peaks and troughs that constitute ERP waveforms appear to have a distinct identity in the sense that the fate of some features of the waveforms appear to be independent of the fate of other features. As data about ERPs accumulated, it became increasingly plausible to assume that these morphological features represent different underlying components.

A component may be viewed solely as a functional entity, in which case it is defined largely in terms of its morphology and sensitivity to experimental variables (Donchin, Ritter, & McCallum, 1978). However, it is also plausible to assume that a component is a manifestation of a unique set of neuronal activities. Indeed, there is, at present, a very active search for the generating sources of ERP components (see, for instance, Wood, McCarthy, Squires, Vaughan, Woods, & McCallum, 1984). The tendency to view components in terms of their neural generators is enhanced by the fact that different components are characterized by a specific scalp distribution. That is, the amplitude of the component is different at different electrodes. As scalp distribution is a consequence of the location of the generating sources, as well as their strength and orientation, it is reasonable to assume that components with different scalp distribution are indeed different. For this reason, scalp distribution has become one of defining characteristic of an ERP component (e.g., Donchin et al., 1978).

In this report, we present a method for the identification of ERP components by means of their scalp distribution. We assume that ERP components have scalp distributions that are constant over time and occasions. We also assume that the scalp distribution of a given ERP component may be distinguished from that of other components and from the background electroencephalographic (EEG) activity. Therefore, we argue that the analysis of scalp distribution may help in identifying components, and enhance the discrimination between an ERP component and other sources of electrical brain activity.

Note that, while we assume that each component has a distinguishable distribution (because each component reflects the activation of different neural structures), we do not assume that we can locate the underlying neural structures from information about scalp distribution. Thus, our interest, in this report, is distinct from that of those who have tried to use distributional information to infer the intracranial location of the component generators (e.g. Lehmann, Darcey, and Skandies, 1980; Schneider and Gerin, 1970; Sidman, Kearfott, and Smith, 1980; Sidman, Pitblado, and Giambalvo, 1981; Simson, Vaughan, & Ritter, 1976, 1977a, 1977b; Vaughan & Ritter, 1970; Wolpaw & Wood, 1982; Wood & Wolpaw, 1982). While this is an important, though difficult, enterprise, we argue that it is possible to consider the scalp distribution mainly as a defining attribute that characterizes components (Donchin, 1978).

Scalp distribution is studied by comparing the potentials recorded at several electrode locations. The voltage observed at an electrode is defined as the difference in potential between the scalp and a reference site. Even though both electrodes are necessary to the recording, it is commonly assumed that the reference site is "neutral," or "inactive." Moreover, the possible influence of "resting" potentials is eliminated from the data by subtracting an estimate of the baseline level, before any further analysis of the data (Donchin et al., 1977).

These commonly used procedures are based on two assumptions:
1. That the reference is indeed inactive during the period under study;
2. That a reliable estimate of the "resting" potential has been obtained.
Both these assumptions require careful evaluation (Donchin et al., 1977), and they have been questioned by several investigators.

First, it has been argued that there is no such thing as an inactive reference site (Lehmann & Skrandies, 1980; in press; Lehmann, Skrandies, & Adachi-Usami, in press; Wolpaw & Wood, 1982). Although attempts have been made to solve this problem by using an average reference (Lehmann & Skrandies, 1980, in press; Offner, 1950; Wolpaw & Wood, 1982), most investigators attempt to circumvent it by (a) acknowledging that their reference is active and (b) adopting a the same reference within and between experiments.

Second, the concept of "resting" potential implies that there exists a time when no component is active. The problem is that we do not know whether such a situation does exist and, if so, when it exists. A partial solution to this problem is provided by high-pass filters, which remove slow variation in potential. Some investigators (Lehmann & Skrandies, in press; Lehmann et al., in press) also argue that the average potential across the entire recording epoch can be assumed to be zero. However, this may not be true, since there may be event-related activity in the epoch which will not average to zero. Another, more popular, solution involves the computation and subtraction of a "baseline" level, that is generally given by the mean activity over a time period preceding the presentation of the stimulus (Donchin et al., 1977). In this case, careful attention should be paid to eliminate the possibility that components associated with preparatory activity are present in the foreperiod. Such components could clearly falsify our claim that we are estimating "resting" activity.

Information about scalp distribution can be obtained by "eye-ball" comparison of the records from several electrode locations. A more informative procedure involves the visual inspection of "isopotential maps," first used in the analysis of EEG activity (Lilly, 1950; Remond, 1962; Remond & Offner, 1952; Walter & Shifton, 1951). Since then, isopotential maps have been repeatedly applied to the study of ERPs (e.g. Ragot & Remond, 1978; Remond & Ragot, 1975; Simson et al. 1976, 1977a, 1977b; Wolpaw & Wood, 1982; Wood & Wolpaw, 1982). However this procedure has several disadvantages. First, it requires the use of a large number of electrode locations. Second, separate plots must be obtained for each data-point, as well as for each experimental condition. To remedy this second problem, Remond (1968) proposed the use of "chronotopograms." These are a type of isopotential maps where the spatial information is reduced to one dimension (e.g., front to back, left to right), and the other dimension is given by time.

Isopotential maps provide a representation of the distributions of the potentials recorded at the scalp. In this sense, they provide descriptive rather than analytic information (Coles, Gratton, Kramer, & Miller, in press). In fact, they do not distinguish between signal and noise. Rather, they are based on the assumption that all the information available is given by the signal (Coles et al., in press), that is, that noise-free data have been obtained.

A recent extension of the mapping procedure, "significant probability mapping," does attempt to provide inferential information. This method, which was developed by Bartels and Subach (1976) to deal with information from light microscopic and Sonar images, has been applied to ERP data by Duffy and his colleagues (Duffy, Bartels, & Burchfiel, 1981; Duffy, Burchfiel, & Lombroso, 1979; Duffy, Denckla, Bartels, Sanding, & Kiessling, 1980). The voltage values are transformed to a measure of the deviation of the voltages from some standard. The deviation values (e.g. the "t" statistic) indicate the statistical significance of the difference between two groups or between an individual and a previously established norm. These significant probability maps consider the voltage at each electrode location and each data-point as independent observations. However, this assumption is not particularly tenable, given the high intercorrelation between the values observed for two adjacent scalp locations or successive data-points.

In this paper we describe an approach we call Vector Analysis. This approach is based on the use of scalp distribution information as an aid in disentangling the component structure of the ERP. Vector Analysis is an approach rather than a specific technique. In fact, several existing techniques may be used when we apply the Vector Analysis approach, as well as new procedures which have been specifically devised for the analysis of scalp distribution. For the purpose of this paper, the Vector Analysis label will include a collection of procedures that share a common model and methodology. The model considers an ERP as given by the vectorial sum of components with specific scalp distributions and noise. The methodology of analysis is based on a multivariate representation of the scalp distribution information, and, therefore, it allows the exploitation of a large number of existing multivariate techniques for inferential testing. Vector Analysis allows also dealing with the problem of inactive reference in an original way, by distinguishing between the information provided by the relative values observed at different electrode sites (distributional information) and the information given by the common difference between the electrodes and the reference (polarity information).

Other investigators have described procedures which share several elements with Vector Analysis, including the use of a multivariate approach (e.g., Skrandies & Lehmann, 1982) and principles of vectorial decomposition of ERPs into components (e.g., Scherg & Von Cramon, 1985). These techniques have generally been concerned with the problem of identifying the sources of the ERPs. Thus, they face problems related to biophysical assumptions about the generation and propagation of the components of ERPs. The approach we are describing does not intend to address the questions of the localization of the component generators. Rather, Vector Analysis capitalizes on the observation that different ERP components are characterized by specific scalp distributions to help in the analysis and decomposition of ERPs into components.

In the remainder of this paper we will first illustrate some basic principles of the Vector Analysis approach. We will then describe two of its applications, Vector Filtering and Modelling.

Vector Analysis: General Principles

The purpose of Vector Analysis is to obtain a description of the ERP in terms of underlying components, where components are defined as having specific patterns of scalp distribution and a specific polarity (positive or negative). The ERP is analyzed on the basis of its scalp distribution at each moment in time, and the relative contributions of different components are separated using analytic techniques to be described below (and in the Appendix). Vector Analysis considers the values observed at different scalp locations as multiple observations of a phenomenon, the ERP, characterized by the presence of a small number of underlying components. The number, magnitude, and spatial properties (distribution) of these underlying components determines the scalp distribution of the ERP observed at any point in time. In fact, we assume that the observed scalp distribution of the ERP is given by the "vectorial" sum of all the active components, plus noise. It is further assumed that the scalp distribution of a particular component is invariant, at least within similar experimental situations and similar subject populations. Of course, these assumptions are consistent with a multivariate interpretation of scalp distribution data (see also Skrandies & Lehmann, 1982).

The obvious consequence of this interpretation is that an analysis of scalp distribution should be oriented toward a decomposition of the observed scalp distribution into its underlying components. This may be accomplished by applying one of several existing multivariate techniques, such as Principal Component Analysis, Discriminant Analysis, Canonical Analysis, or

Multivariate Analysis of Variance. All of these procedures have the property of providing a new description of the scalp distribution data which (a) uses components rather than electrodes as descriptive elements, and (b) meets particular criteria established by the investigator. These existing techniques are concerned with minimizing or maximizing parameters derived from the data themselves, such as variance, ratio between variances, etc. Since components derived in this way are data-dependent, the problem of their interpretation remains - that is, the investigator must decide whether a particular component is, or is not, a P300, a CNV, etc. Such a decision is based on a subjective comparison between the obtained component and the investigator's "template" of how a particular component should behave.

An alternative approach to the description of components is to use a procedure that allows the investigator to establish the characteristic distribution of a component _a priori_ and to derive a description of the observed scalp distribution of the ERP in terms of the degree to which it fits the pre-determined distribution. One of the features of the Vector Analysis procedure is that it allows one to determine the contribution of a component with a known distribution to an observed set of data. It should be emphasized that this approach is _not_ incompatible with the idea of deriving information about components on an _a posteriori_ basis. Rather, the approach leaves investigators free to satisfy that criterion that best suits their experimental question. Later in this paper, we will review the application of Vector Analysis to cases in which we assume (Vector Filter) or do not assume (Modelling) a priori knowledge about scalp distribution. The first application is particularly appropriate when the investigators wish to separate the contribution of a particular component of known scalp distribution (e.g. the P300) from data containing several different sources of electrical activity. The second application is more suited to the case in which the investigators wish to study differences in scalp distribution between experimental conditions or between different subject populations.

Description of the procedure

In this section, we provide a brief, narrative description of some basic elements of the procedure. A more analytic description is given in the Appendix.

As noted above, the basic model of Vector Analysis assumes that the values recorded at all the scalp electrodes at a particular data-point are given by the sum of several components and noise. It also assumes that each component is characterized by a specific scalp distribution and polarity, indicated by a particular set of weights, one for each electrode. For example, for the P300 component, the weights for both Pz and Fz will be positive, but the weight for Pz will be larger than that for Fz since P300 is a positive component characterized by a parietally maximal scalp distribution (Donchin et al., 1978).

Traditionally, multivariate analysis represents a set of dependent measures as a "space" whose axes correspond to the variables. With Vector Analysis, we identify a space whose axes correspond to the different electrodes. Thus, for any data-point, we can represent the voltage values at several electrodes as a point in this space. This is shown in Figure 1 for a case in which only two electrodes are used, and where the obtained voltage value at Fz is 1, and that at Pz is 8. Note that we can represent any particular scalp distribution (i.e., component) as a line in this space, if that scalp distribution is defined in terms of a fixed relationship among the values for different electrodes. In Figure 1, we show the line corresponding to a scalp distribution for which the ratio between two electrodes (Pz:Fz) is constant (2:1). (That is, all the ERPs for which the amplitude at Pz is precisely twice the amplitude of Fz will fall on the component line.) Note that, as one moves away from the origin along the component line, the corresponding values for the electrodes increase. Thus, we define the amplitude of a component in terms of its distance from the origin, along the component line.

Insert Figure 1 About Here

Points not lying on the component line are associated with different scalp distributions. Note also that any scalp distribution (i.e., fixed relationship among the electrodes) can be represented by a line in the space defined by the electrodes. This line can be described in terms of the angles (alpha and beta) between the line and the electrode axes.

So far, we have seen that any set of electrode values can be represented by a point in a space whose axes correspond to the electrodes, and that any distribution can be represented by a line in that space. Note also that it is possible to express the extent to which data-point (8,1) is similar to the component represented by the line. The projection of the data point onto the line yields an amplitude which can be interpreted as the contribution of the component to the observed data-point. Indeed, using simple geometrical principles, it is possible to project any point in the space onto any line in the space and, in this way, to define the contribution of any component (scalp distribution) to any data-point. In Figure 1, the length of the projection of the data-point onto the distribution line is 7.6 (arbitrary units). Projecting the data-point on the component line is equivalent to rotating the axes that define the space so as to align one axis with the line corresponding to the component distribution.

As we have noted, any distribution can be represented by a line in the space. If we consider lines that are orthogonal (i.e., at right angles to each other), we can derive a new set of reference axes that will represent components (i.e., scalp distributions) to describe a data-point. These new axes can then be used in place of the original electrodes axes. The advantage of doing so is that it is possible to describe a data-point in terms of the contribution of specific components.

Out of all the possible set of axes, one has special significance (see Figure 2). One of the axes of this particular set is the line corresponding to a distribution that weights all the electrodes equally (mean axis). The projection of any data-point onto this axis is proportional to the mean of the electrode values. If the value of this projection is positive, then the component has a positive polarity (and vice versa if the projection is negative). The axis (or axes if there are more than two electrodes; see Appendix) orthogonal to the mean axis describes the variance between electrodes (variance axis).2 In fact, the projection of the data-point onto the variance axis is proportional to the standard deviation among electrodes (in the case in which only two electrodes are used - see Appendix for an extension to the "n-electrode" case). The mean and variance axes, as well as the data-point and the component line already shown in Figure 1, are shown in Figure 2.

Insert Figure 2 About Here

Note that the line representing the component can also be described in terms of its relationship to the new set of axes. This description is based on the angles between the component and the new axes. Furthermore, if we draw a line connecting the origin to the data-point, we can also describe that line in terms of angles. In fact, we can locate a particular data-point by specifying (1) the angle between this line and the mean axis, and (2) the distance of the data-point from the origin (length of the line). The complementary value of the angle is labelled "polarity angle,"[3] while the value of the distance is labelled "magnitude." In fact, the polarity angle is a function of the ratio between the mean and the standard deviation of the electrode values. The value of the magnitude measure is a function of the absolute voltage values obtained at each electrode (in fact, it is equal to the square root of the sum of squares of the individual electrode values). Note that two data-points, having different mean values and different standard deviations, can have the same polarity angle but different magnitudes. This procedure allows us to deal with the problem of magnitude variation of a component separately from variations in scalp distribution. That is, a given component can vary in magnitude as long as the ratio among electrodes (i.e., the scalp distribution) remains constant. As McCarthy and Wood (1985) have recently argued, variation in component magnitude will result in unequal effects on the voltage values at different electrodes, although the ratio among electrodes will remain constant. Thus, Vector Analysis represents magnitude variation in a manner that is consistent with this view.

In conclusion, we can derive three different descriptions of an observed set of electrode values for a particular data-point:

1. A description based on the voltage values obtained at each electrode. That is, we can locate the data-point in a space defined by the electrodes.
2. A description based on the projection of the data-point onto lines associated with specific components. That is, we can describe the data-point in terms of the contribution of different components.
3. A description based on the projection of the data-point onto axes associated with mean and variance. In the two-electrode case, we can also describe a data-point in terms of the distance of the data-point from the origin (magnitude) and the angle with the variance axis (polarity angle). If there are more than two electrodes, more than one angle will be required for the description.4

By using Vector Analysis we can also derive two different descriptions of a component (i.e., scalp distribution):

1. A description based on the angles between the component line and the electrode axes;
2. A description based on the angle between the component line and the mean (and variance) axis.

Note that, so far, the procedures described are a way of "re-describing" the information provided by the observed values simultaneously at different electrode locations. As such, the procedures are atheoretical. Theoretical notions of the distribution of a particular component are important when we wish to interpret a particular observation. We will now describe two applications of the principles of the procedure: Vector Filter and Modelling.

Analysis of Component Contribution:

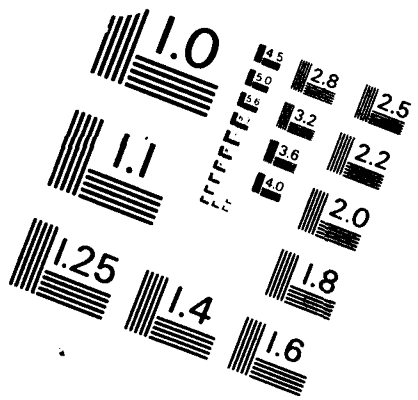
Vector Filter

Vector Filter is an application of the principles described in the previous section, designed to extract information about a particular component (Gratton, Coles, & Donchin, 1983a). The component is characterized by a target scalp distribution defined a priori. Vector Filter weights the values observed at each data-point according to the degree to which the observed scalp distribution conforms to the distribution of the target component. In this sense, Vector Filter is a way of "filtering" for a particular scalp distribution (i.e., component). This filtering is accomplished by rotating the space defined by the electrodes, and by aligning one of the axes with the line corresponding to the target distribution. The projections of each data-point onto the newly defined axis are then computed. These values constitute the Vector filtered data.[5] An example of Vector Filter is shown in Fig. 3.

Insert Figure 3 About Here

In Figure 3 (panel a), the waveforms for two scalp electrodes (Fz and Pz) are shown. For a given data-point (indicated by the vertical line), we extract the values for the two electrodes (panel a) and represent them in the space defined by the two electrodes (panel b). In the same space we identify an axis corresponding to the target scalp distribution (component). The projection of the data-point onto the target scalp distribution (component) gives us the filtered value for that data-point. The procedure

AD-A170 794   THE EVENT-RELATED BRAIN POTENTIAL AS AN INDEX OF   6/8
INFORMATION PROCESSING A..(U) ILLINOIS UNIV CHAMPAIGN
COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL.
UNCLASSIFIED 20 FEB 86 CPL-86-1 F49620-85-C-0041   F/G 5/10   NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

may then be repeated for all data-points in the waveform, yielding a vector filtered waveform (panel c).

The value obtained by filtering may be interpreted as the amount of activity at a given data-point which may be attributed to a particular component, defined in terms of scalp distribution. In this vein, Vector Filters can be used to determine the independent contribution of several components to the same data-point. However, such an application requires that the components whose contribution we wish to identify have scalp distributions whose axes are orthogonal. Although this may not be always possible, in several cases an appropriate choice of the scalp electrode locations may produce conditions where this assumption may be met. For example, in the three electrode case, with Fz, Cz, and Pz, the P300 and the Contingent Negative Variation have roughly orthogonal scalp distributions, as one is Pz maximal, the other is Cz maximal, and both are small at Fz.

Another possible application of Vector Filter is to prepare data for an analysis of latency and amplitude. In this case, Vector Filter can be used to improve the discrimination between signal and noise, in association with other "traditional" filtering techniques. Note that, in this case, Vector Filter is not a signal detection technique - that is, Vector Filter is used to reduce the noise rather than to detect the signal. Vector Filter should be therefore compared with other procedures used for the selection of spatial information. For example, analysis of P300 latency usually involves a selection of the information provided by the Pz electrode. Such a selection is, in fact, a filtering procedure that ascribes a weight of unity to the Pz electrode and zero to all other electrodes.

Because the Vector Filter procedure allows the use of the information provided by more than one electrode, it is, at least in principle, superior to procedures that involve selection of a single electrode. In fact, Gratton, Kramer, Coles, and Donchin (1985) conducted a simulation study in which P300 latency estimates obtained from data prepared with Vector Filter were compared with estimates obtained from data prepared with more traditional methods, such as selection of the parietal electrode site. They adopted a Vector Filter which, rather than using the P300 scalp distribution as the target component, used a distribution which provided a good discrimination between P300 and noise. This choice is based on the fact that the scalp distribution of the noise partially overlaps that of P300. In the same study, they also compared the accuracy of P300 latency estimates obtained with peak-picking and cross-correlational techniques.

The results of this study clearly indicated the advantage of preparing the data with Vector Filter before applying a signal detection algorithm. Gratton et al. (1985) found that Vector Filter produced an improvement in the latency estimates, with reductions of the error ranging from 0 to 50%. Vector Filter produced greater improvement at relatively low signal-to-noise ratios (about 1) than at very high signal-to-noise ratios. The improvement given by Vector Filter proved to be very robust when components with a very different scalp distribution from P300 were introduced but was impaired by the presence of overlapping components with a scalp distribution that is similar to that of P300 (such as Slow Wave). This latter finding could be predicted from the theoretical arguments outlined above. However, even in the worst case, the application of Vector Filter never yielded larger errors in the estimated latencies than the traditional procedure used as a comparison.

Vector Filter relies on a model which assumes that the values observed at the scalp electrodes at a given point in time may be attributed to a particular component (scalp distribution), designated as signal, and to noise. Note that noise is here defined as all that information which cannot be attributed to the target component. Although this definition of noise may be acceptable for several applications of Vector Filter (including the use of Vector Filter to prepare data for latency estimation, as shown above), the assumptions underlying the model may be empirically tested. This is illustrated in the next section.

Analysis of scalp distribution: Model building and testing

One possible problem in the interpretation of scalp distribution is to determine whether or not an observed distribution conforms to that of a particular component, as it is described in the literature, or as it is present in a "normative" group. The Vector Filter procedure, described in the previous section, addresses this issue.

Secondly, we may be interested in determining whether two groups differ in scalp distribution, or whether scalp distribution varies as a function of experimental manipulations. We shall illustrate below how this problem can be addressed.

A third problem is to determine the degree to which the distributional information can be explained in terms of only one component, or whether more than one component is necessary. In fact, if more than one component is present, we should distinguish among the independent contributions of each component. Again, we shall illustrate below how this problem may be addressed.

The database used for this illustration comes from a larger study of the effect of aging on the P300 (Miller, Bashore, Marshall, Gratton, Coles, & Donchin, in preparation). In this study, 46 subjects took part in a series of four "oddball" experiments.6 The subjects were all volunteer, in normal health, with normal or corrected to normal vision and hearing. The subjects were divided into two age groups: "old" group and "young" group. The subjects of the "old" group ranged in age between 60 and 82, while the subject of the "young" group ranged in age between 18 and 23. For the demonstration to be given below, ten subjects were selected from each group, on the basis of the size of their P300 (the subjects selected were those with the larger amplitude of the P300 peak for each age group). Note, however, that the series of analyses was also repeated on the complete samples of 27 and 19 subjects, yielding similar results.

The four oddball experiments differed in terms of stimulus modality (three were auditory and one was visual), task (three required the subject to count the occurrence of one of the stimuli, and one was a choice reaction time task), and cognitive demands (two required tone discrimination, one detection of stimulus omission, one semantic classification). The series consisted of:

a. An auditory count task, where two tones (frequency 1,000 and 1,500 Hz respectively, duration 50 ms) each had a .50 probability of being presented. The subject's task was to count the number of occurrences of one of the tones.

b. An auditory choice reaction time task, where the same two tones used in the auditory count task were presented with different probabilities (.20/.80). The subjects were required to discriminatively respond to each tone, by pressing one of two buttons.

c. An auditory omitted stimulus task, where 10% of the stimuli were omitted. The subject was required to count the number of omissions. The tones presented had a frequency of 1,000 Hz (or 1,500 Hz, in half of the subjects), duration of 50 ms.

d. A visual count task, where male and female names were presented with a .20/.80 probability. The subject counted the occurrences of the rare stimuli. The names and the procedure used for this task are similar to those described in Kutas, McCarthy, and Donchin (1977) for the "Variable Name" condition.

All aspects of stimulus generation and data collection were controlled by a Pearl Laboratory System. The auditory stimuli were presented binaurally through headphones. The visual stimuli subtended a visual angle ranging between 4 and 6 degrees. For each task, 200 trials were presented with an interstimulus interval of 2 seconds, divided into two blocks of 100 trials. EEG was recorded from Fz, Cz, and Pz (International 10-20 System, Jasper, 1958) by Burden Ag/AgCl electrodes attached with collodion and referenced to linked mastoids. Vertical EOG was recorded from above and below the right eye. Beckman Biopotential Ag/AgCl electrodes were used for the reference and EOG electrodes. Impedence was below 10 KOhm. The EEG signal was amplified and conditioned by Grass amplifiers (time constant 8 seconds, upper half/cut off frequency 35 Hz). The signal was digitized on-line at 200 Hz, for 1280 ms starting 100 ms before the stimulus. Ocular artifacts were corrected off-line according to a procedure described by Gratton, Coles, and Donchin (1983b).

The experimental questions that we will address here are (a) whether the two groups differ in the scalp distribution of the ERPs elicited during these tasks, and (b) whether the scalp distribution of the two groups respond in a different way to the particular demands of the different tasks.

If we also accept the notion that scalp distribution tells us something about the use of internal brain structures in performing the tasks, we may actually consider the more general question about the differential use of brain structures in response to the task demands in the two age groups (cf. Pfefferbaum, Ford, Roth, & Kopell, 1980).

Average ERPs for each subject, task, and stimulus, were obtained at Fz, Cz, and Pz, referred to linked mastoids. Given that this research was mainly concerned with the P300 component of ERPs, we choose to focus our analysis on the point of maximum positivity (independent of electrode) in a time window between 300 and 900 ms after the occurrence of the stimulus. For each electrode, the mean value of the 100 ms immediately preceding the stimulus was used as an estimate of the "resting" potential and subtracted from the appropriate data before any further analysis. Visual inspection of the data confirmed that the point selected by a computer algorithm for each subject and condition actually corresponded to the most positive peak in the window. Peak picking was used because it is most commonly employed to define the P300 component in the ERP literature (cf. Fabiani, Gratton, Karis, & Donchin, in press; Kramer, 1985). Furthermore, since no electrode was selected, we do not expect the procedure to produce any bias in favor of a particular scalp distribution. It also allows us to take into account any systematic differences in latency between groups and tasks.

Figure 4a and 4b show the grand average waveforms for each group according to task, stimulus, and electrode. From an inspection of these figures, we see that the pattern of scalp distribution at the point of maximum positive peak is generally different in the two groups. In fact, the young group (panel a) shows a larger difference among the three electrodes than the old group (panel b). However, the average amplitude across electrodes is similar in the two groups.

Insert Figure 4 About Here

Thus, one of the differences in scalp distribution at P300 peak between the young and the old group is in the ratio between the mean across electrodes and the standard deviation among the electrodes.7 As we already noted, this ratio can be expressed in terms of the tangent of the "polarity" angle. This angle has a positive value when the mean of the electrodes is positive, a value close to zero when the mean across electrodes is close to zero, and a negative value when the mean is negative. Extreme values of the polarity angle (close to + or -90 degrees) correspond to cases in which the standard deviation among the electrode values is very small (relative to the mean) or 0. The mean tangents of the polarity angle (obtained as described above - see also equation 10 in the Appendix) as a function of group, task and stimulus are shown in figure 5.

Insert Figure 5 About Here

Overall, the polarity angle had larger values for the old group than for the young group, $F(1,18)=12.45$, $p<.01$. However, this difference was most evident for rare stimuli in the name oddball. The interaction of group x task x stimulus was significant, $F(3,54)=4.54$, $p<.01$. Further scrutiny of the polarity values revealed that the young group had a constant polarity across tasks and stimuli, while the old group exhibited more variable polarity.

These results enable us to draw some preliminary conclusions about the differences between groups and tasks. First, the young group exhibits a larger difference among electrodes. Second, this difference (adjusted for mean amplitude variations) remains relatively constant across tasks and stimuli, while the differences among electrodes for the old group is more variable.

To obtain a more detailed description of the differences between the two groups, we analyzed the variance across electrodes at the moment of P300 peak. Since we used three scalp electrodes (Fz, Cz, and Pz), the variance across electrodes can be described by means of two axes, corresponding to two orthogonal decomposition of the variance. These two axes correspond also to two particular scalp distribution, characterized by, respectively, a parietal maximal and a central maximal distribution. These two axes were chosen because they corresponded, respectively, to the scalp distribution most frequently associated with P300 (the parietal distribution axis), and to a scalp distribution that stresses the contribution of the central electrode (the central distribution axis). For each subject, task, and stimulus, we can plot a point in the space defined by these two axes. Thus, for each group (and task and stimulus) we can obtain a "cluster" of points, corresponding to the scalp distribution of each subject at the moment of P300 peak. Note that, for each point, the distance from the origin indicates the standard deviation among electrodes for that subject, task and condition, while the orientation in the space reflects the relative contribution of the parietal and central distribution.

The clusters for each group, task and stimulus, can be described by ellipses, encompassing 90% of the variance among subjects. These ellipses are presented in figure 6. Note that this figure shows the variance plane and should not be confused with the space defined by the electrodes. Note also that the size of the ellipses indicates the inter-subject variability, while the distance of the ellipses from the origin indicates the inter-electrode variability. The distance of the ellipses from the origin represent the standard deviation among electrodes, and the orientation of the ellipses in the space provides information about the pattern of scalp distribution at the moment of P300 peak. Finally, note that the major axis of each ellipse represents the axis of largest variability (in scalp distribution) across subjects.

Insert Figure 6 About Here

As explained above, the distance from the origin in these graphs is equal to the standard deviation among electrodes. The young group shows larger variance among electrodes than the old group. A second important point to note from these graphs is the angle subtended by each ellipse. This angle represents the range of scalp distributions in which we are likely to find the mean of the group. These "confidence" angles are shown in figure 7.

Insert Figure 7 About Here

This figure indicates that the young group tends to have similar patterns of scalp distribution across tasks and stimuli, while the patterns of scalp distribution of the old group vary as a function of these variables.

So far, we have seen that our approach to the analysis of scalp distribution allows us to detect differences between groups, tasks, and stimuli and their interactions. Note that we have made but a few assumptions about models or about components. Even the use of ANOVA to test for differences is not mandatory - distribution-free statistics could have been used to obtain similar results. However, the interpretation of the differences revealed by this analysis requires the choice of a model and some assumptions.

One possible interpretation of the differences visible in figures 5, 6, and 7 is that only one component (i.e., P300) is responsible for the scalp distribution observed in the young group regardless of task and stimulus. The old group, on the other hand, perhaps produces more than one component in response to different tasks and stimuli. An alternative explanation is that the variability in scalp distribution shown by the old group is attributable to random variations, which happen to be particularly visible because of the attenuation of the main component (i.e., P300).

To determine which of these two interpretations is correct, we must propose a hypothesis about the scalp distribution of the P300 and determine whether the residual variance after removal of the P300 scalp distribution is significantly different from zero. In other words, instead of filtering for the target component, as we did with the Vector Filter, we remove, by filtering, the target component and study the residuals.

This analysis is a test of a particular model, which assumes that a single component (P300) is able to explain the scalp distribution under all conditions. The main problem, in this case, is the definition of the scalp distribution of the P300. In fact, the scalp distribution of the P300 may differ in the two groups. This difference may be due to physiological or anatomical differences between the groups. Therefore, we need a definition of the scalp distribution of P300 which is adapted to each group. We chose to use the scalp distribution for the rare omitted stimulus as the P300 scalp distribution, because this condition should be free of exogenous components. An interesting observation is that the ellipse for this condition in the young group is particularly narrow and is oriented toward the origin (see Figure 6), suggesting that most of the variance (both across electrodes and across subjects) can be attributed to a single component.

The procedure, adopted separately for each group, was to rotate the distribution space so as to align one of the axes with the target distribution (i.e., the scalp distribution obtained for the omitted stimulus) and then study the residuals. Note that we re-introduced in the analysis information about the mean of the electrode values. Therefore, given that we used three electrodes, the residuals can be plotted still on a plane (see Appendix). The axes in these graphs are arbitrary. They indicate pairs of scalp distributions that are orthogonal to the P300 target distribution in the space defined by the electrodes. They are different for the two groups, since the distribution of the "P300" for the omitted stimulus is different in the two groups. As before, we obtained clusters of points (corresponding to each single subject) for each group, task and stimulus, that can be summarized by ellipses encompassing 90% of the inter-subject variability. The ellipses containing the confidence region for the mean of the residuals are shown in figure 8, for each group, task, and stimulus.

Insert Figure 8 About Here

The real point of interest in this figure is to see whether or not the ellipses encompass the origin - that is, whether or not the residuals after filtering out the P300 are consistently different from 0. The presence of consistent residuals would indicate that a second component, with scalp distribution different from the P300, must be invoked to explain the data, and that the single-component model must be rejected. A numerical analogue to this geometrical form of representation is obtained through a one-sample Hotelling T-square test (Tatsuoka, 1971). The results of this test are given in Table I.

Insert Table I about here

Figure 8 indicates that the P300 component (defined as above) is sufficient to explain most of the data of the young group. In fact, in five out of the eight stimulus by task conditions, the ellipses corresponding to the residuals encompass the origin. In the visual name task, another component is consistently present for both the rare and the frequent stimuli, since, for these conditions, the ellipse does not encompass the origin. For the rare stimulus in the auditory reaction time task consistent residuals are also present.

In contrast to the young group, the old group tends to have much larger residuals, particularly for the non-target and frequent stimuli, indicating the presence of activity which cannot be attributed to the P300 scalp distribution. Therefore, we must invoke other scalp distributions to explain these data.

These results help evaluate the relative merits of the two alternatives discussed above. The variability in scalp distribution presented by the old group cannot be attributed to random variations, but to the presence of ERP components overlapping with the P300. The impact of additional components is much larger in the old group than in the young group.

On the basis of these data, one could speculate about differences between the two groups in their use of brain structures to perform the different tasks. Perhaps, more structures are utilized by the old group because of an impaired "accessibility" of the routine manifested by P300.

This analysis demonstrates the possibility of using scalp distribution data to test hypotheses about the presence or absence of overlapping ERP components. The procedure we have presented could be repeated with other components, testing multi-component models. The only limitation is presented by the number of electrodes used.

Conclusion

In this paper, we have proposed that scalp distribution can be treated as a dependent variable by using a collection of procedures we label "Vector Analysis." There are two important features of Vector Analysis that differentiate it from the procedure customarily used:

a. The use of a multivariate approach, by means of which electrode locations are regarded as a series of dependent variables rather than as different levels of an independent variable.

b. The emphasis on the concept of "component," with component defined in terms of scalp distribution.

We have demonstrated that Vector Analysis is useful in tackling three common problems in ERP research. First, the question of whether or not an observed scalp distribution conforms to that of a particular component, as it is described in the literature or is present in a "normative" group. Second, the question of whether two groups differ in the scalp distribution of their ERPs, or whether scalp distribution varies as a function of experimental manipulations. Third, the question of the degree to which distributional information can be explained in terms of only one component, or whether more than one component is necessary.

We have also shown that procedures based on Vector Analysis (e.g. Vector Filter) can be brought to bear on problems for which some solution had already been proposed (e.g. P300 latency estimation on single trials). In this case, our procedures prove to be superior to those in common use (Gratton et al., 1985).

We emphasized in the introduction that an interest in distributional information does not imply any assumptions about the relationship between particular neural generators and scalp electrical activity. We consider a component to be the manifestation at the scalp of the activity of unknown brain structures. However, we do assume that, whenever we observe a difference in scalp distribution as a result of an experimental manipulation, different brain structures are involved. An important qualification of this assumption is that it refers only to those cases in which we may assume that anatomical and physiological (in contrast to psychological) conditions remain constant. In the previous section, we presented a case in which age differences in scalp distribution were found. It could be argued that they are due to differences in the anatomical and physiological conditions of the two groups. In this case, scalp distribution of an ERP component may vary, for instance, as a consequence of differences in calcification of the skull, or size of the ventricles, or brain lesions.

Although individual differences of this type may be interesting to some investigators, we have focussed on the psychological (functional) interpretation of differences in scalp distribution. Such a focus is justified if there is an interaction between group and task - in other words, if the difference between the groups is not constant over tasks.

Summary

This paper presents a procedure, Vector Analysis, for the analysis of the scalp distribution of ERPs. The interest in scalp distribution is based on the assumption that the electrical activity of certain brain structures involved in psychological processes is manifested at the scalp by particular components. The scalp distribution of each component reflects anatomical and physiological properties of the structures involved in the generation of the component, as well as conductive characteristics of the interposed media. Most investigators agree that ERP components are characterized by specific scalp distributions. This is probably due to the invariance of the underlying source(s). Hence, the use of information about scalp distribution may be helpful in the identification and analysis of ERP components.

Vector Analysis considers the values observed at several electrode sites as the sum of several components, each characterized by a specific pattern of scalp distribution, and of background noise. Each ERP component is characterized by a specific scalp distribution. This specific scalp distribution may be expressed by a series of weights, one for each electrode. Each component may vary in amplitude as a function of time and experimental manipulation. Our goal is to obtain estimates of the set of weights describing the scalp distribution of the components and of the variations in amplitude as a function of time and experimental manipulations. Within the framework of our model, such estimates would provide a complete description of the ERPs under study. The adoption of a multivariate approach in obtaining these estimates is emphasized.

One possible application of Vector Analysis consists of filtering for a

particular scalp distribution (Vector Filter). Filtering for scalp distribution improves the discrimination between signal (component) and noise (background EEG and overlapping components). The power of Vector Filter is illustrated by a simulation study, in which we compared the accuracy of P300 latency estimates obtained with several techniques. Particular attention has been given to the problem of enhancing the discrimination between signal and noise. The best discrimination, and the most accurate latency estimation, are obtained when the characteristics of both the signal and the noise are considered.

To demonstrate the ability of the procedure to discriminate between overlapping components, the results obtained by applying our approach to a study of P300 in aging were illustrated. The data suggest that the scalp distribution of the P300 peak is different for young adults and elderly people over a variety of tasks. The two groups showed both overall differences and task-related differences. The analysis of the ERPs observed in each task revealed that these differences in scalp distribution cannot be entirely attributed to a generally different distribution of P300 in the two groups, to variations in P300 amplitude, or to a combination of these two factors. Rather, they should be attributed, at least in part, to the presence of overlapping component(s). This is particularly evident for the subjects in the old group.

Acknowledgements

The authors wish to acknowledge the contribution of Greg Miller, Noel Marshall, Ted Bashore, Kay Strayer, and Ron Chambers to data collection and analysis. The comments of Monica Fabiani, Demetrios Karis, Art Kramer, Greg Miller, and Dave Strayer on early versions of this paper are greatly appreciated.

References

Bartels, P.H., & Subach, J.A. Automated interpretation of complex scenes. In: E. Preston & M. Onoe (Eds.) Digital Processing of Biomedical Imagery, Academic Press, New York, 1976: 101-114.

Coles, M.G.H., Gratton, G., Kramer, A.F., & Miller, G.A. Principles of signal acquisition and analysis. In: M.G.H. Coles, E. Donchin, & S.W. Porges (Eds.) Psychophysiology: Systems, Processes, and Applications. Vol. I: Systems. New York: Guilford press, in press.

Donchin, E. A multivariate approach to the analysis of average evoked potentials. IEEE Transactions in Biomedical Engineering, 1966, 13: 131-139.

Donchin, E. Use of scalp distribution as a dependent variable in event-related potential studies: excerpts of preconference correspondence. In: D. Otto (Ed.), Multidisciplinary Perspectives in Event-Related brain Potentials research (EPA-600/9-77-043). US Govt. Printing Office, 1978, 501-510.

Donchin, E., Callaway, E., Cooper, R., Desmedt, J.E., Goff, W.R., Hillyard, S.A., & Sutton, S. Publication criteria for studies of evoked potentials in man. In: J.E. Desmedt (Ed.) Attention, Voluntary Contraction and Event-Related Cerebral Potentials. Prog. clin. neurophysiol., 1977, vol. 1, 1-11.

Donchin, E., Ritter, W., & McCallum, C. Cognitive psychophysiology: The endogenous components of the ERP. In: E. Callaway, P. Tueting, & S. H. Coslow (Eds.) Event-related brain potentials in man. Academic Press, New York, 1978, 349-411.

Duffy, F.H., Bartels, P.H., & Burchfiel, J.L. Significant probability mapping: An aid in the topographical analysis of brain activity. Electroencephal. clin. Neurophysiol., 1981, 51: 455-462.

Duffy, F.H., Burchfiel, J.L., & Lombroso, C.T. Brain electrical activity mapping (BEAM): a new method for extending the clinical utility of EEG and evoked potential data. Ann. Neurol., 1979, 5: 309-321.

Duffy, F.H., Denckla, M.B., Bartels, P.H., Sanding, G., & Kiessling, L.S. Dislexia: automated diagnosis by computerized classification of brain electrical activity. Ann. Neurol., 1980, 7: 421-428.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. The definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. To appear in P.K. Ackles, R. Jennings, and M.G.H. Coles (Eds.) Advances in psychophysiology, vol. II. JAI, Greenwich (CT), in press.

Gratton, G., Coles, M.G.H., & Donchin, E. Filtering for spatial distribution: A new approach (Vector Filter). Psychophysiology, 1983a, 20, 4: 443-444 (Abstract).

Gratton, G., Coles, M.G.H., & Donchin, E. A new method for the off-line removal of ocular artifact. Electroencephal. clin. Neurophysiol., 1983, 55: 468-484

Gratton, G., Kramer, A.F., Coles, M.G.H., & Donchin, E. A simulation study of latency measures of event-related potentials. In preparation.

Jasper, H.H. The ten-twenty electrode system of the International Federation. Electroencephal. clin. Neurophysiol., 1958, 10: 371-375.

Kramer, A.F. The interpretation of the component structure of the event-related brain potential: An analysis of expert judgement. Psychophysiol., 1985, 22: 334-344.

Kutas, M., McCarthy, G., & Donchin, E. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 1977, 197: 792-795.

Lehmann, D., Darcey, T.M., & Skrandies, W. Intracerebral and scalp fields evoked by hemiretinal checkerboard reversal, and modeling of their dipole generators. Proceedings International Symposium on Clinical Applications of Evoked Potentials in Neurology, Oct. 16-17, 1980, Lyon, France: 31-32.

Lehmann, D., & Skrandies, W. Reference-free identification of components of checkerboard-evoked multichannel potential fields. Electroencephal. clin. Neurophysiol., 1980, 48: 609-621.

Lehmann, D., & Skrandies, W. Spatial Analysis of Event-related potentials. 1983, in press.

Lehmann, D., Skrandies, W., & Adachi-Usamy, E., The baseline for the measurement of evoked potentials. In: A. Rothenberger (Ed.) Event-related Potentials in Children, Amsterdam: Elsevier, in press.

Lilly, J.C. A method for recording electrical potential gradients in the brain: The 25-channel Bavatron and electro-iconograms. In: Conference of Electronics in Nucleonics and Medicine. Amer. Rev. of Electr. Eng., New York, 1950: 37-43.

McCarthy, G., & Wood, C.C. Scalp distribution of event-related potentials: An ambiguity associated with analysis of variance models. Electroencephal. Clin. Neurophysiol., 1985.

Miller, G.A., Bashore, T.B., Marshall, N., Gratton, G., & Donchin, E. Study of variation of event-related brain potentials with aging. in preparation

Offner, F.F. The EEG as potential mapping: the value of the average monopolar reference. Electroencephal. clin. Neurophysiol., 1950, 2: 215-216

Pfefferbaum, A., Ford, J., Roth, W.T., & Kopell, B.S. Age-related changes in auditory event-related potentials. Electroencephal. clin. Neurophysiol., 1980, 49: 266-276.

Ragot, R.A., & Remond, A. EEG field mapping. Electroencephal. clin. Neurophysiol., 1978, 45: 417-421.

Remond, A. Construction et adjustment des enregistrements cartographiques et spatio-temporels des EEG. Rev. Neurol., 1962, 106: 135-136.

Remond, A. The importance of topographic data in EEG phenomena, and an electrical model to reproduce them. Electroencephal. clin. Neurophysiol. Supplement, 1968, 27: 29-49.

Remond, A., & Offner, F. A new method for EEG display. Electroencephal. clin. Neurophysiol., 1952, 7: 453.

Remond, A., & Ragot, R. Un capteur topologique en EEG. Actes du Colloque International sur les Capteurs Biomedicaux, Paris, 1975.

Scherg, M., and Von Cramon, D. A new interpretation of the generators of BAEP waves I-V: Results of a spatio-temporal dipole model. Electroencephal. clin. Neurophysiol., 1985, 62: 290-299.

Schneider, M., & Gerin, P. Une methode de localisation des dipoles cerebraux. Electroencephal. clin. Neurophysiol., 1970, 28: 69-78.

Sidman, R.D., Kearfott, B., & Smith, D.B. A method for identifying noise-free evoked potential components - application of DLM (dipole localization method) to these components. IEEE 1980 Frontiers of Engineering in Health Care, 137-140.

Sidman, R.D., Pitblado, C., & Giambalvo, V. Mathematical models for deducing the asymmetry of brain response during verbal processing. Proceeding of Ninth Annual Northeast Bioengineering Conference, Pergamon Press, 1981.

Simson, R., Vaughan, Jr., H.G., & Ritter, W. The scalp topography of potentials associated with missing visual stimuli. Electroencephal. clin. Neurophysiol., 1976, 40: 33-42.

Simson, R., Vaughan, Jr., H.G., & Ritter, W. The scalp topography of potentials in auditory and visual discriminatory tasks. Electroencephal. clin. Neurophysiol., 1977a, 42: 528-535.

Simson, R., Vaughan, Jr., H.G., & Ritter, W. The scalp topography of potentials in auditory and visual go-nogo tasks. Electroencephal. clin. Neurophysiol., 1977b, 43: 864-875.

Skrandies, W., & Lehmann, D. Spatial principal components of multichannel maps evoked by lateral visual half-field stimuli. Electroencephal. clin. Neurophysiol., 1982, 54: 297-305.

Tatsuoka, M. M. Multivariate analysis. Wiley, New York, 1971.

Vaughan, Jr., H.G., & Ritter, W. The sources of auditory evoked responses recorded from the human scalp. Electroencephal. clin. Neurophysiol., 1970, 28: 360-378.

Walter, W.G., & Shifton, H.W. A new toposcopic display system. Electroencephal. clin. Neurophysiol., 1951, 3: 281-292.

Wolpaw, J.R., & Wood, C.C. Scalp distribution of human auditory evoked potentials. I. Evaluation of reference electrode sites. Electroencephal. clin. Neurophysiol., 1982, 54: 15-24.

Wood, C.C., & Allison, T. Interpretation of evoked potentials: A neurophysiological perspective. Canad. J. Psychol., 1981, 35(2): 113-135.

Wood, C.C., McCarthy, G., Squires, N.K., Vaughan, H.G., Woods, D.L., & McCallum, W.C. Anatomical and physiological substrates of event-related potentials. In R. Karrer, J. Cohen, & P. Tueting (Eds.), Brain and Information: Event-related Potentials. Annals of the New York Academy of Science, New York, 1984, 681-721.

Wood, C.C., & Wolpaw, J.R. Scalp distribution of human auditory evoked potentials. II. Evidence for overlapping sources and involvement of auditory cortex. Electroencephal. clin. Neurophysiol., 1982, 54: 25-38.

Appendix

The basic model assumes that the values recorded at all the scalp electrodes are given by the sum of several components, and noise. The components are characterized by a specific scalp distribution, indicated by a particular set of weights, one for each electrode. The components may vary in amplitude as a function of time and of the experimental manipulations. Therefore, the values recorded at any scalp electrodes at a given point in time may be described by the following equation:

$$e_{jk} = \sum_{i=1}^{n} a_{ik} w_{ij} + \text{noise} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where:

e_{jk} is the value recorded at the electrode j at time k;

a_{ik} is the amplitude of the component i at time k;

w_{ij} is the weight of the component i at the electrode j.

n is the number of components with a "significant" contribution to time k;8

We can consider the data-point k as an m-dimensional space (m being the number of electrode placements). We assume that this m-dimensional space may be described by n components. We call this space, the "Distribution Space." Of course, if n (number of components) is larger than m, an indefinite number of different combinations of component weights and amplitudes are able to reproduce the values observed at all m electrodes. Therefore, the number of components, whose contribution can be estimated, is limited to the number of electrodes used.

Skrandies and Lehmann (1982) proposed the use of Principal Component Analysis (PCA) to determine which components contribute to the variance and covariance between electrodes at any given point in time. Equation (1) can be used as the basis for a PCA.9

An important property of the Distribution Space should be emphasized. A particular line (passing through the origin) in this space defines a fixed pattern of electrode values - that is, a relationship among electrode values that is independent of magnitude.

Insert Figure 9 About Here

For example, in Figure 9a we represent a space defined by two electrode locations. All the values lying on the dashed line have the same ratio between the values at the two electrodes. The points on this line differ by a multiplicative factor that is equal for the two electrodes. This factor, which is proportional to the distance from the origin, we label "amplitude". Each line that passes through the origin in the distribution space is composed of all the points for which the relative pattern of values observed at each electrode is constant. Conversely, points not lying on the same line are associated with different patterns. If we define scalp distribution as a particular pattern of values observed at each electrode, then any particular scalp distribution has a unique correspondence to a line passing through the origin in the distribution space.

The lines corresponding to the scalp distributions can be defined in terms of their angles with the axes corresponding to the electrode locations. In Figure 9a, these are the angles alpha and beta. The cosines of these angles correspond to the weight of a particular scalp distribution for each electrode, and are equal to w(ij) in the equation (1). The values for each electrode at a particular data-point identify an unique point in the space. One and only one line can be drawn that passes through the point and the origin. This line corresponds to the scalp distribution at that point in time.

In figure 9b, we consider a specific two electrode case, for which the value of Cz is 3 and Pz is 4. The point (4,3) represents the values for the two electrodes and the line "l" corresponds to the scalp distribution of this point. Note that the line can be defined in terms of the angle between it and the Pz or Cz axes (37 and 53 degrees).

As with all spaces, the distribution space may be rotated and defined by new axes. For any point, the projection on any of the new axes is given by the usual formula:

$$p_{ik} = \sum_{j=1}^{m} (e_{jk} c_{ij}) \quad \text{.....................................(2)}$$

where (as before):

p_{ik} is the projection on the new axis i for time k;

m. is the number of the electrodes.

e_{jk} is the value recorded at the electrode j at time k;

c_{ij} is the cosine of the angle between the new axis and the old axis corresponding to the electrode j; note that when we rotate the axes, the values of c must be such that

$\sum_{j=1}^{m} c_{ij}^2 = 1$ (see Tatsuoka, 1971).

Of all the possible rotations, one presents particular advantages. This rotation involves aligning one of the axes in a direction for which all the angles between the old axes (corresponding to the electrodes) and the new axis are equal. Now, the projection on this new axis of a point will be proportional to the mean of the values over electrodes. This can be seen by inspection of formula (2). If all the angles are equal, the formula reduces to:

$$p_k = c \sum_{j=1}^{m} (e_{jk}) \qquad (3)$$

where (as before):

p_k is the projection on the new axis for time k;

c is the cosine of the angle between the new axis and all the old axes corresponding to the electrodes;

e_{jk} is the value recorded at the electrode j at time k;

m is the number of the electrodes.

The value of p is proportional to the mean because formula (3) is equivalent to the formula for the mean with the replacement of (1/m) by another constant c. Note that c corresponds to the cosine of the angle between each of the electrode axes and the new axis. Note also that this angle must be such that the square of its cosine is equal to 1/m (m being the number of electrodes). One of the consequences of this is that the value of the projection of p onto each electrode axis is equal to the mean of the values across electrodes (see Figure 9c).

This can be shown as follows:

$$p'_k = c\, p_k \quad \text{.......................................(4)}$$

where p' is the projection of p onto an electrode axis

Substituting from (3), we get:

$$p'_k = c^2 \sum_{j=1}^{m} (e_{jk}) \quad \text{.......................................(5)}$$

$$\text{but } c^2 = 1/m \quad \text{... ....(6)}$$

So from (5) and (6), we obtain:

$$p'_k = 1/m \sum_{j=1}^{m} (e_{jk}) \quad \text{.......................................(7)}$$

In the two electrode case, this new axis will have an angle of 45o with the original axes corresponding to the two electrodes (Figure 9c).

While the projection of the point on the new axis will be proportional to the mean over the electrodes, the distance between the point and its projection on the new axis will be related to the standard deviation among electrodes.

For each electrode axis, the distance from the projection of the new point (p_k) to the projection of the original point (e_k) is given by:

$$d'_{jk} = e_{jk} - p'_k \text{ ....(8)}$$

for all electrodes, then, the total distance is given by

$$d_k = \sqrt{\sum_{j=1}^{m} d'^2_{jk}} \text{ ....(9)}$$

Note that, since d is equivalent to the deviation between the observed value and the mean, formula (9) will yield a value that is proportional to the standard deviation among electrodes. We illustrate this for the two-electrode case in Figure 9d.

Note that this rotation procedure allows us to isolate the information provided by the mean of the values over all electrode locations from the remaining information. As we have seen, this remaining information refers to the variance among the electrodes.

In the case of ERPs, this is particularly advantageous. In fact, as we discussed in the introduction, the mean of the values across all electrode locations corresponds to the mean difference with the reference that is common to all the electrodes. We may therefore analyze this information (relating to the comparison with the common reference) separately from the remaining information (corresponding to the comparison among electrodes).

As we indicated earlier, the Vector Analysis procedure considers scalp distribution <u>and</u> polarity to be defining characteristics of an ERP component. We incorporate polarity into our procedure by defining it as follows:

$$pol_k = \arctan (p_k / d_k) \text{ .... (10)}$$

where pol_k is the polarity at data-point k

Since p can be positive or negative, depending on the value of the mean across the electrodes relative to the reference, the value of pol_k can also be positive or negative. Furthermore, since d_k is related to the standard deviation of the electrode values, pol_k is an index of the ratio of the mean to the standard deviation. It is, therefore, independent of the "magnitude" of the waveform as this is traditionally measured.

The polarity value thus obtained is related to the angle between the axis corresponding to the observed scalp distribution and the axis corresponding to the mean of the electrodes. In fact, this last angle is complementary to the polarity angle (the sum of the two angles is equal to 90 degrees). This relationship, as applied to the two-electrode case, can be seen in Figure 9e.

As we have seen, the information that is contained in a set of values recorded at a particular data-point from m electrodes can be decomposed into information about the mean and information about the variance. It is also the case that the information about the variance can be decomposed further. In particular, there will be m-1 independent pieces of information contained in the variance. In the two electrode case, there is only one variance dimension. However, in the three electrode case, the number of possible independent components of variance is equal to 2. In this case, we can represent the variance across electrodes as a two-dimensional space (plane), where the two dimensions represent orthogonal decompositions of the variance.

Figure 9f represents an example of the application of this procedure. From the values recorded at three electrode locations, the mean value was first computed. Then, by computing variance and standard deviation, the value of the polarity angle was computed. Finally, two orthogonal decompositions of the variance were set-up, providing the values for each of the new dimensions. Note that the axis corresponding to the mean exists in a third dimension, orthogonal to the page.

Lines passing through the origin are made up of all those points for which the ratios among the values for each orthogonal decomposition are equal. They correspond to different patterns of scalp distribution, independent of the mean of the electrode values (i.e. independent of polarity). In the case of three electrodes, these patterns may be described by a single angle with one of the axes. We call this angle "orientation".

Since we have only performed an orthogonal rotation on the distribution space, the length of the vector corresponding to the data-point remains unaffected. The length of this vector is therefore independent of the values of the angles used to describe scalp distribution. We therefore consider this as a measure of the activity recorded at a given data-point, independently of its scalp distribution, and call this measure "amplitude". It is defined as follows:

$$a_k = \mathrm{sqrt} \sum_{j=1}^{m} e_{jk}^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(11)$$

By means of the transformations mentioned above, we have obtained a description of the values observed at a particular point in time given by a series of angles, and a length measure. This description corresponds to vector (or polar) notation, and contrasts with the usual cartesian notation. The advantage of the vector notation is that it allows us to separate the information about scalp distribution (provided by the angles) from the information about the amplitude (provided by the length of the vector). We have also been able to separate the information provided by the relative weight of the mean of the electrodes (polarity) from the information relative to the pattern of values at different electrodes (orientation in the case of three electrodes, several angles in the case of more than three electrodes).

Footnotes

1. The study presented in this paper was supported in part by a grant from the Air Force Office of Scientific Research, contract #F49620-83-0144, Al Fregly, Program Director. Portions of this study were presented at the 23rd Annual Meeting of the Society for Psychophysiological Research, Asilomar, California, September 26-29, 1983, and at 3rd International Conference on Cognitive Neuroscience, Bristol, England, September 17-20, 1984. Some of the data described in this report were collected in connection with a study of aging, supported by a grant from NIA, #03151.

2. Note that all the information conveyed by a series of observations can be partitioned into two parts: the common trend of all the observations (mean), and the deviation of each observation from the common trend (variance). Thus, mean and variance are "statistically" independent, and can be graphically represented as orthogonal axes. Furthermore, the variance can be partitioned into a number of components equal to the number of observations minus one, according to the well-known Fisherian approach. Each of these components of variance (labelled "orthogonal decompositions of the variance") will be orthogonal to the mean.

3. We use the complementary value of the angle because we want our measure of polarity to vary from positive through zero to negative as the value of the mean across electrodes varies in the same way.

4. Note that the description of the data-point in terms of distance and angles corresponds to a description in terms of polar notation.

5. In this way, Vector Filter is "defined" by the target scalp distribution. The target scalp distribution may be selected on the basis of previous knowledge of the scalp distribution of a particular component, or

so to satisfy some criteria, (e.g., discrimination between signal and noise, etc.). The amplitude (i.e., length of the vector in the space defined by the electrodes) of each data-point is scaled for a factor determined by the similarity between the observed scalp distribution and that of the target component. This similarity can be quantified by computing the cosine of the angle between the component line and the line connecting the data-point to the origin (see Figure 3). Note that Vector Filters with greater selectivity may be obtained by using uneven powers of this cosine; however, such filters cannot be considered as cases of orthogonal rotation, and the space described by such filters is distorted.

6. By "oddball" we mean an experiment in which the subject is confronted with a series of stimuli, each belonging to one of two categories, where the probability of each of the two categories may be manipulated.

7. The use of this ratio is justified by the assumption that when a component increases in amplitude, not only the mean amplitude across electrodes increases, but the differences among the electrode values increases (see McCarthy & Wood, 1985, for a discussion of this subject).

8. Significant is here intended in a statistical sense. The contribution of a component may be tested in comparison with the contribution of the noise or as proportion of explained variance, etc.

9. This use of PCA is different from that proposed by Donchin (1966) and most commonly adopted in ERP research. Donchin considers data-points as variables, while Skrandies and Lehmann consider scalp locations as variables.

Table I

Analysis of Residuals after P300 Subtraction: Hotelling T-Square (n=10).

Task	Stimulus	Old group	Young group
Auditory count (50/50)	Target	5.92	.80
	Non-target	8.98+	2.10
Auditory RT (20/80)	Rare	.62	9.20+
	Frequent	2.49	2.72
Auditory omitted-stimulus	Rare	.01	.01
(10/90)	Frequent	26.79**	4.23
Visual names (20/80)	Rare	34.97**	14.76*
	Frequent	85.31**	20.62**

+ $p < .10$

* $p < .05$

** $p < .01$

Figure legends

Fig. 1. The voltage values obtained at Pz and Fz for an imaginary data-point are plotted in a space defined by the two electrode locations. A line corresponding to a particular scalp distribution (i.e., that of "P300") is also plotted, as well as the projection of the data-point onto this line.

Fig. 2. The voltage values obtained at Pz and Fz for an imaginary data-point are plotted in a space defined by the two electrode locations. A line corresponding to a particular scalp distribution (i.e., that of "P300") is also plotted, as well as the projection of the data-point onto this line. The axes corresponding to the mean and variance of the electrodes, as well as the respective projection of the data-point on these axes are also plotted.

Fig. 3. Graphic representation of the Vector Filter procedure. Waveforms recorded at Pz and Fz are shown in panel (a), as well as a particular data-point (vertical line). In panel (b), the data-point is represented on the plane defined by the electrode locations; the scalp distribution corresponding to the target component is indicated by a line; the projection of the data-point on the component line is also shown. Panel (c) shows the waveform obtained by repeating the procedure for each data-point.

Fig. 4. Grand average waveforms at Fz (solid), Cz (dashed), and Pz (dotted) for each group, task and stimulus (n=10). Waveforms from the YOUNG group are shown in panel (a), and those from the OLD group are shown in panel (b).

Fig. 5. Polarity (i.e. tangent of polarity angle, see Appendix) as a function of task for the old (circles) and the young (triangles) group. Target stimuli are indicated with solid lines, and not-target stimuli with dashed lines.

Fig. 6. Representation of the variance among electrodes for each group, task, and stimulus. The axes correspond to orthogonal decomposition of the variance. The ellipses represent the 90% confidence regions for each estimate (n=10).

Fig. 7. Representation of confidence limits for the scalp distribution (i.e., orientation angle, see Appendix) for each group, task, and stimulus. The axes correspond to orthogonal decomposition of the variance.

Fig. 8. Representation of residual activity after subtraction of the P300 for each group, task, and stimulus. The axes for the old group and for the young group are not equal. The ellipses represent 90% confidence regions for the mean estimate.

Fig. 9. Graphic representation of the basic steps of Vector Analysis (see Appendix).

(a.) Example of distribution space defined by two electrode locations (Pz and Cz). An axis corresponding to a specific scalp distribution is also shown. The angles describing the weight of each electrode location in the scalp distribution are indicated with alpha and beta.

(b.) Example of distribution space defined as in (a.) The values at the two electrode locations (e_{1k} and the e_{2k}) indicated.

(c.) Same as (b.), but the axis corresponding to the mean of the electrodes (labelled new axis) is also shown. The projection of e onto the new axis (p) and the angle of the observed scalp distribution with the new axis are also shown.

(d.) Same as (c.), but the projection of (p_{1k}) of p onto the Pz axis is indicated. The distances between p and e_k (d) and between p'_{1k} and e_{1k} (d'_{1k}) are shown.

(e.) An axis orthogonal to that corresponding to the mean of the electrodes is indicated. The statistical significance of this axis, and of the projection of e_k onto it are explained in the text. The polarity angle is also shown.

(f.) Representation of the variance among electrodes in the three electrode cases. A point (e), corresponding to three deviation values, is indicated. The corresponding orientation angle is also shown. For explanation about the significance of this angle, see text.

Fig. 1

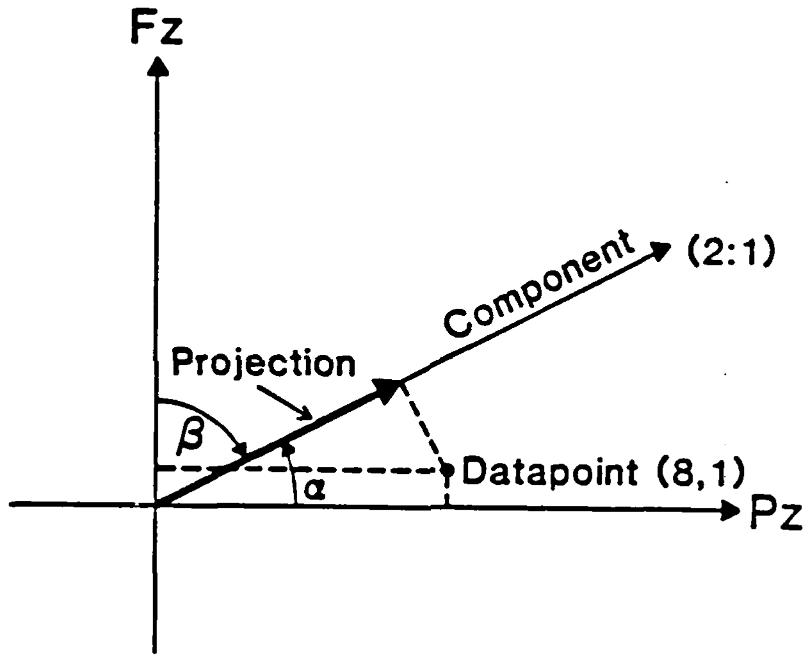
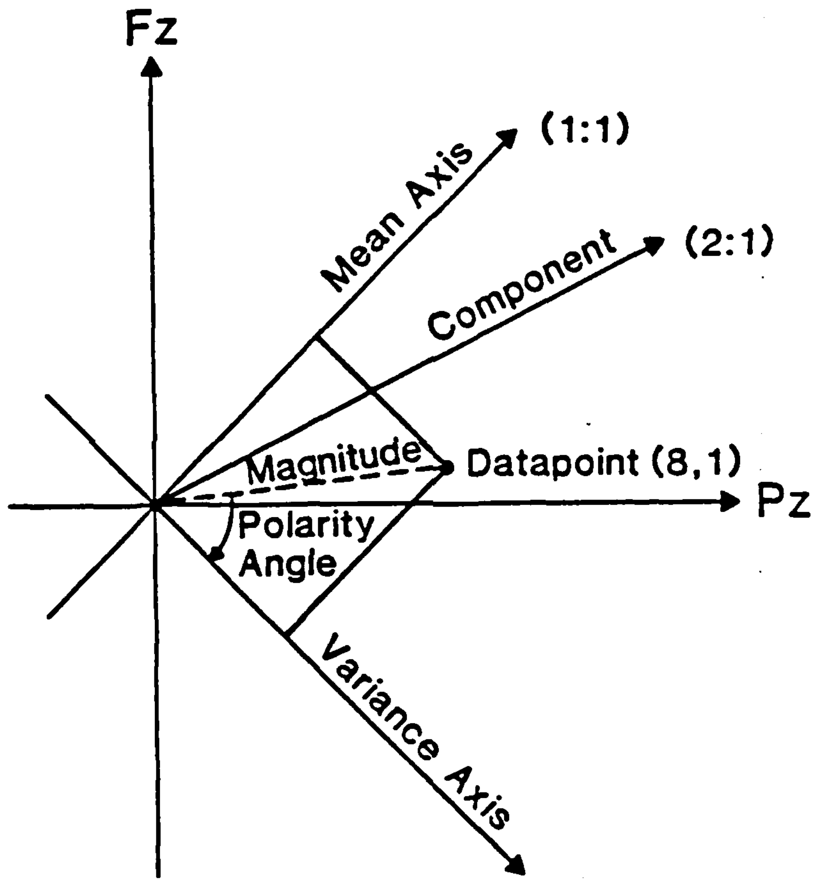
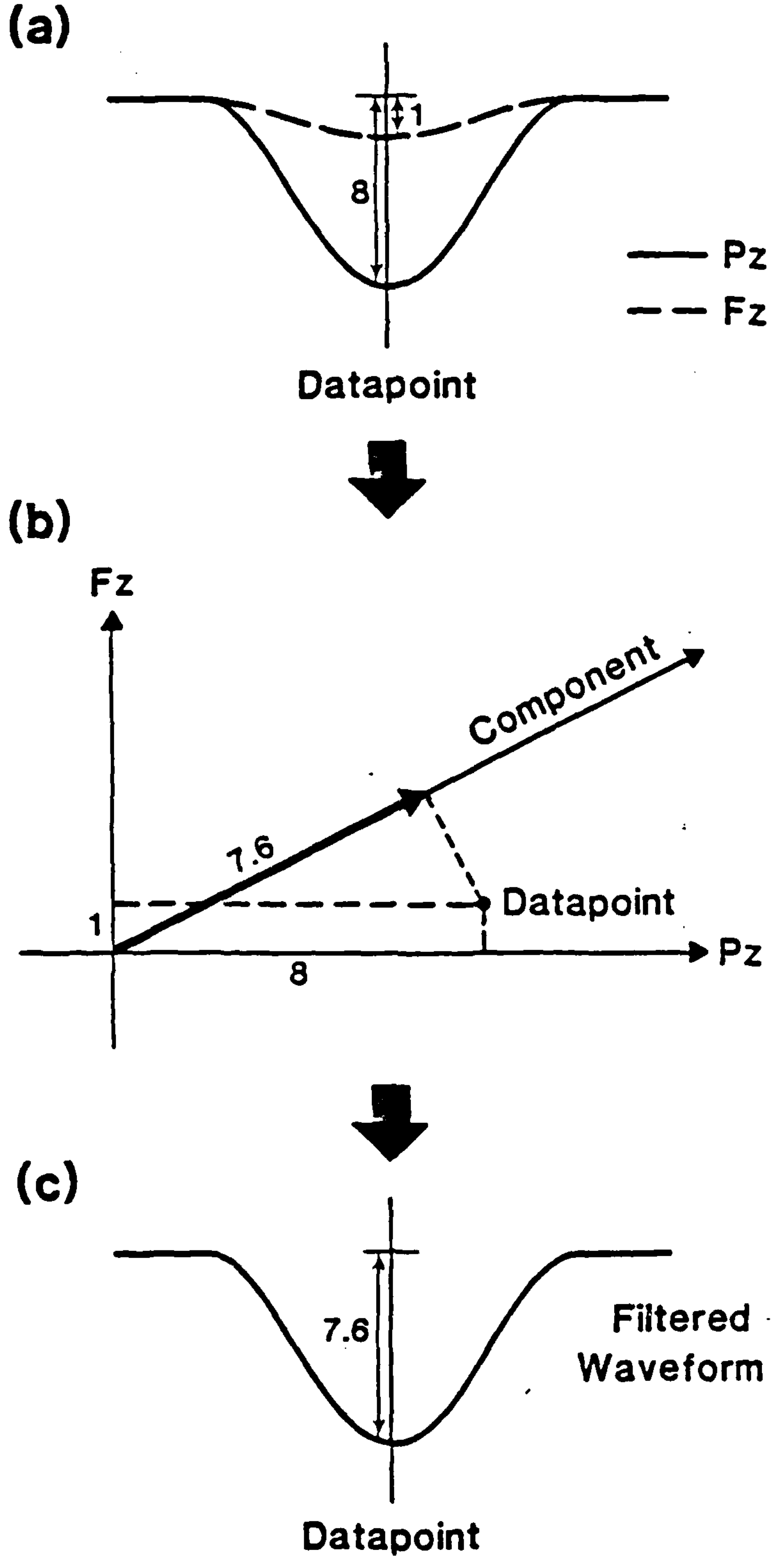


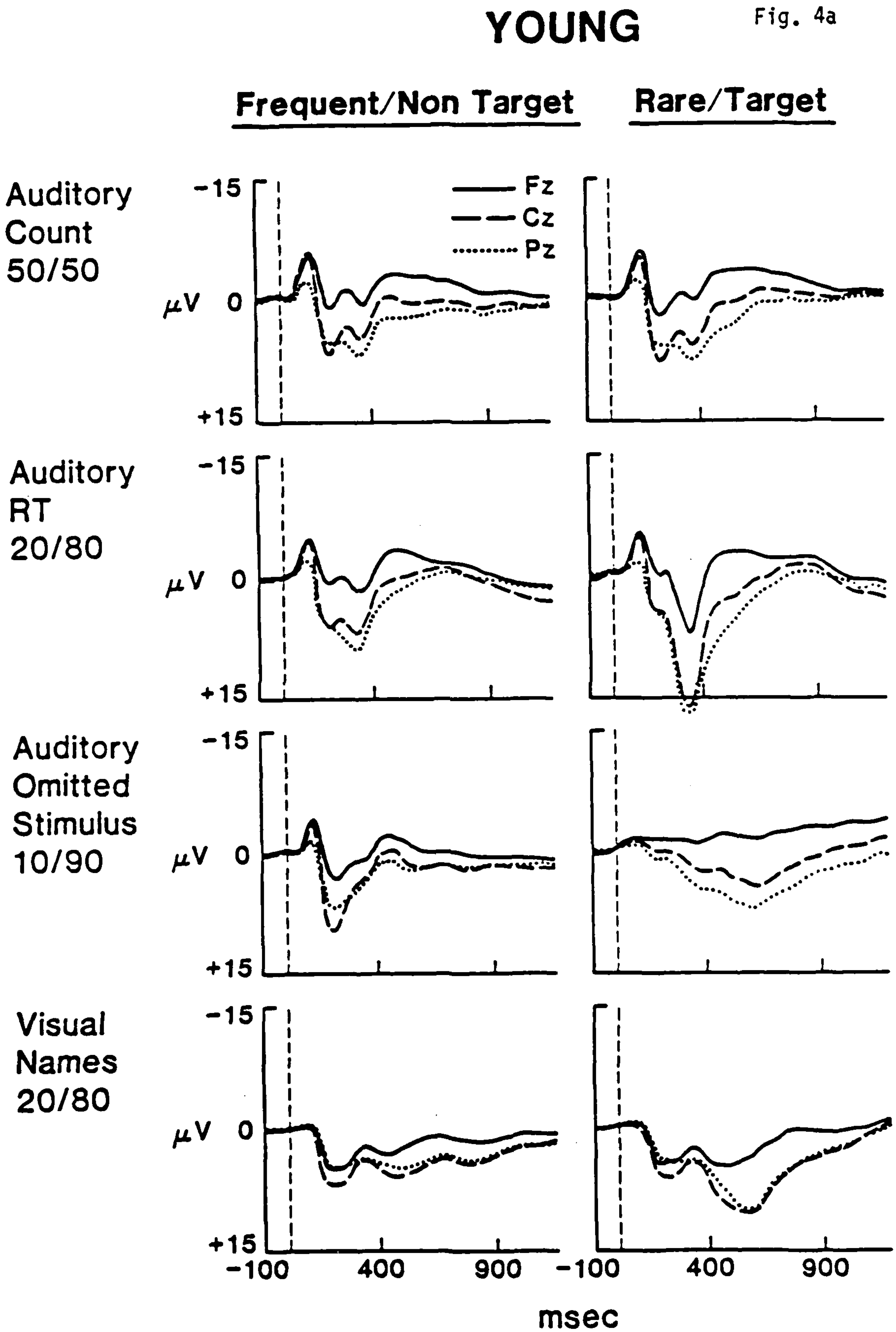
Fig. 2



Fz
Mean Axis
(1:1)
Component
(2:1)
Magnitude
Datapoint (8,1)
Pz
Polarity
Angle
Variance Axis

Fig. 3


(a)
1
8
Pz
Fz
Datapoint
(b)
Fz
Component
7.6
Datapoint
1
8
Pz
(c)
7.6
Filtered
Waveform
Datapoint



YOUNG
Fig. 4a
Frequent/Non Target
Rare/Target
Auditory Count 50/50
Auditory RT 20/80
Auditory Omitted Stimulus 10/90
Visual Names 20/80
Fz
Cz
Pz
μV
-15
0
+15
-100
400
900
msec

Fig. 4b

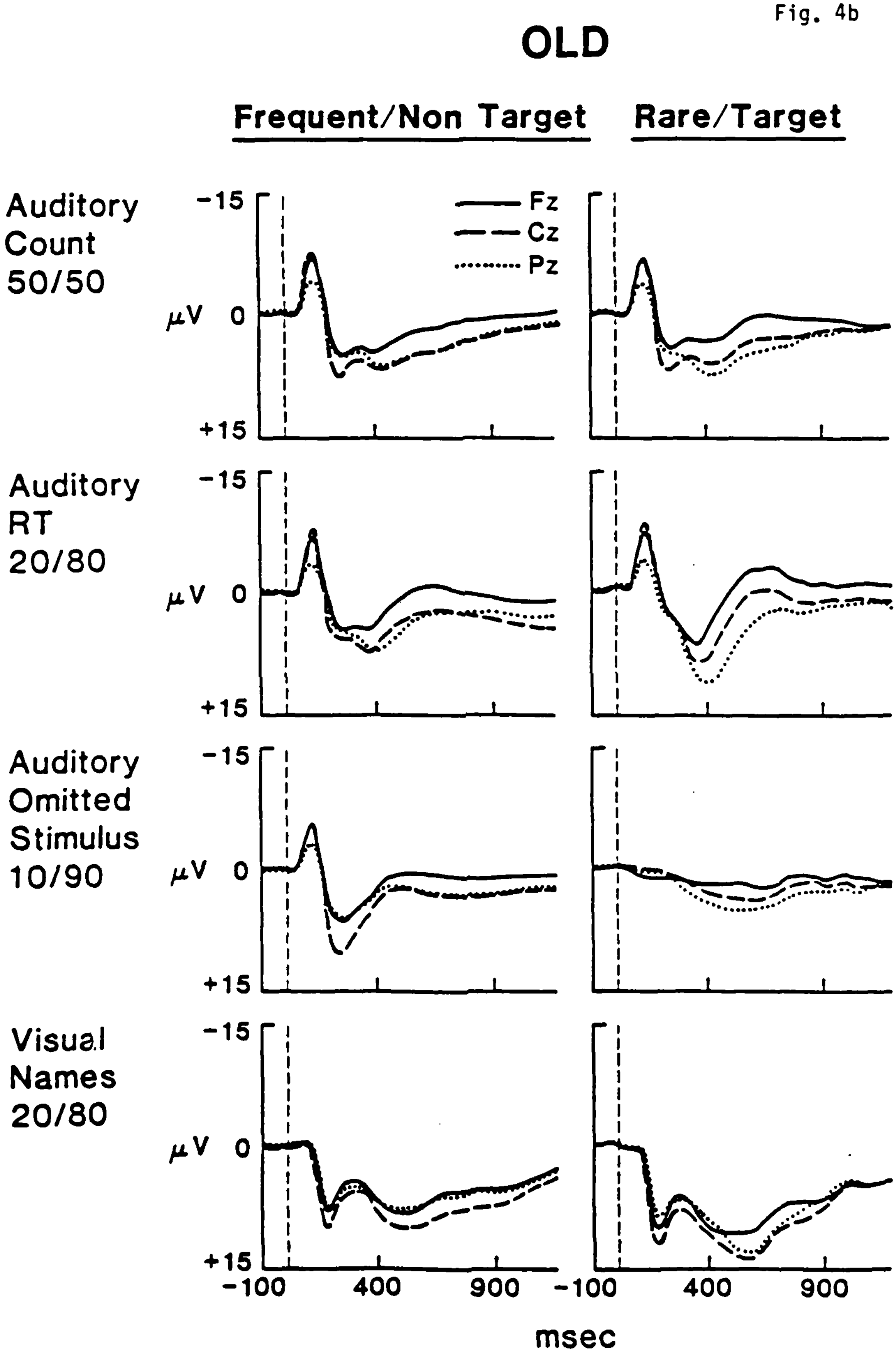
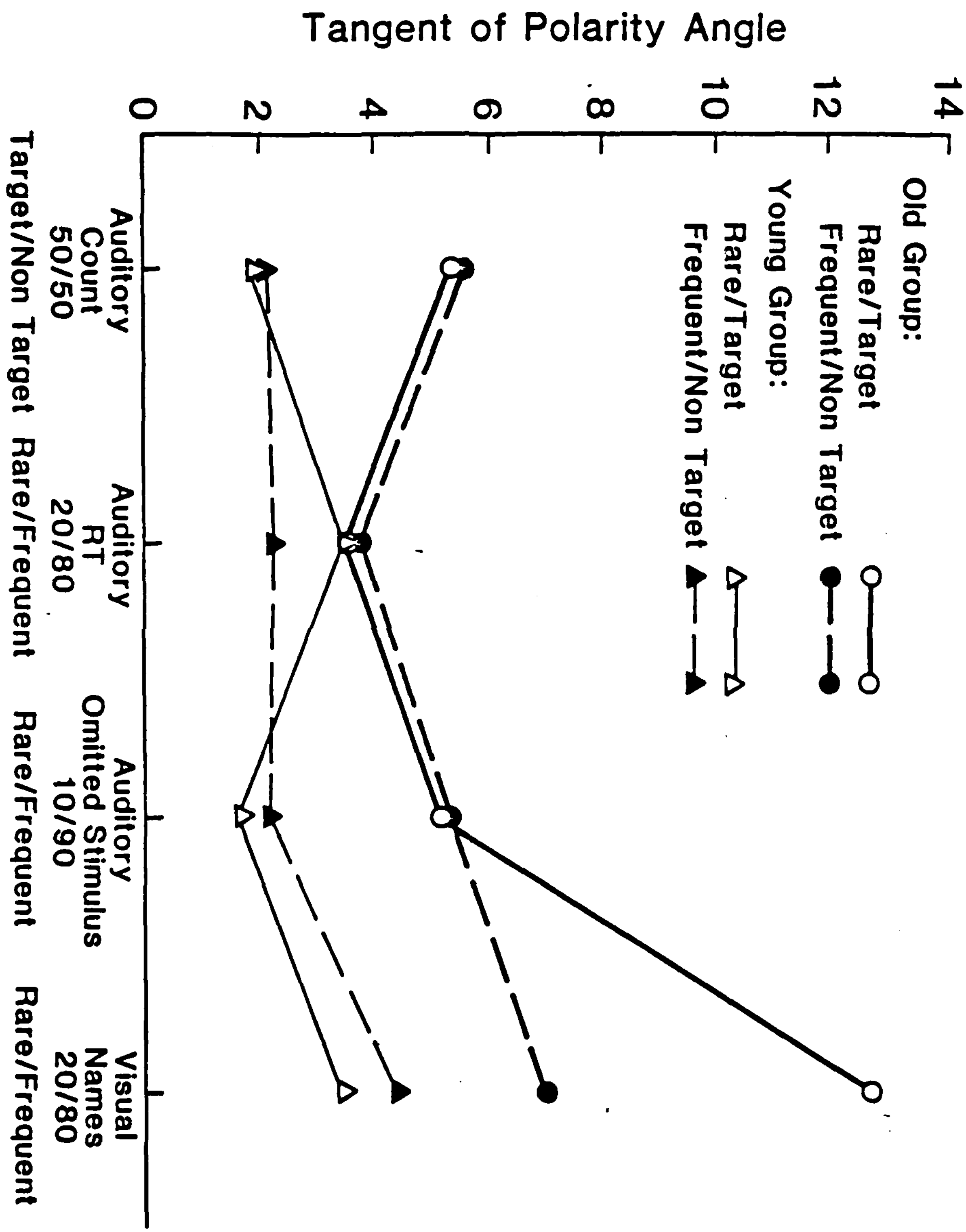
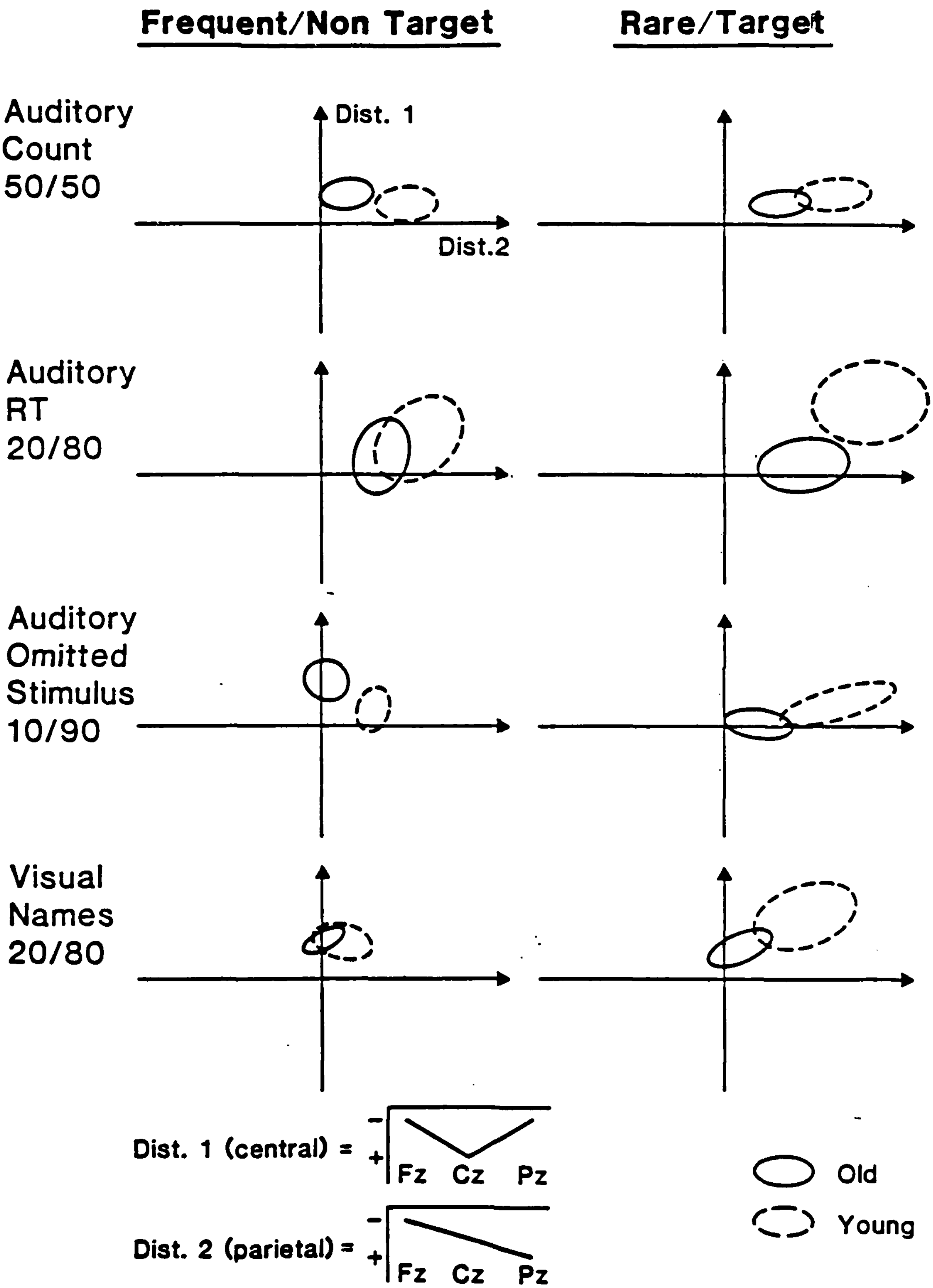

OLD
Frequent/Non Target
Rare/Target
Auditory Count 50/50
Auditory RT 20/80
Auditory Omitted Stimulus 10/90
Visual Names 20/80
Fz
Cz
Pz
-15
μV 0
+15
-100
400
900
msec

Fig. 5

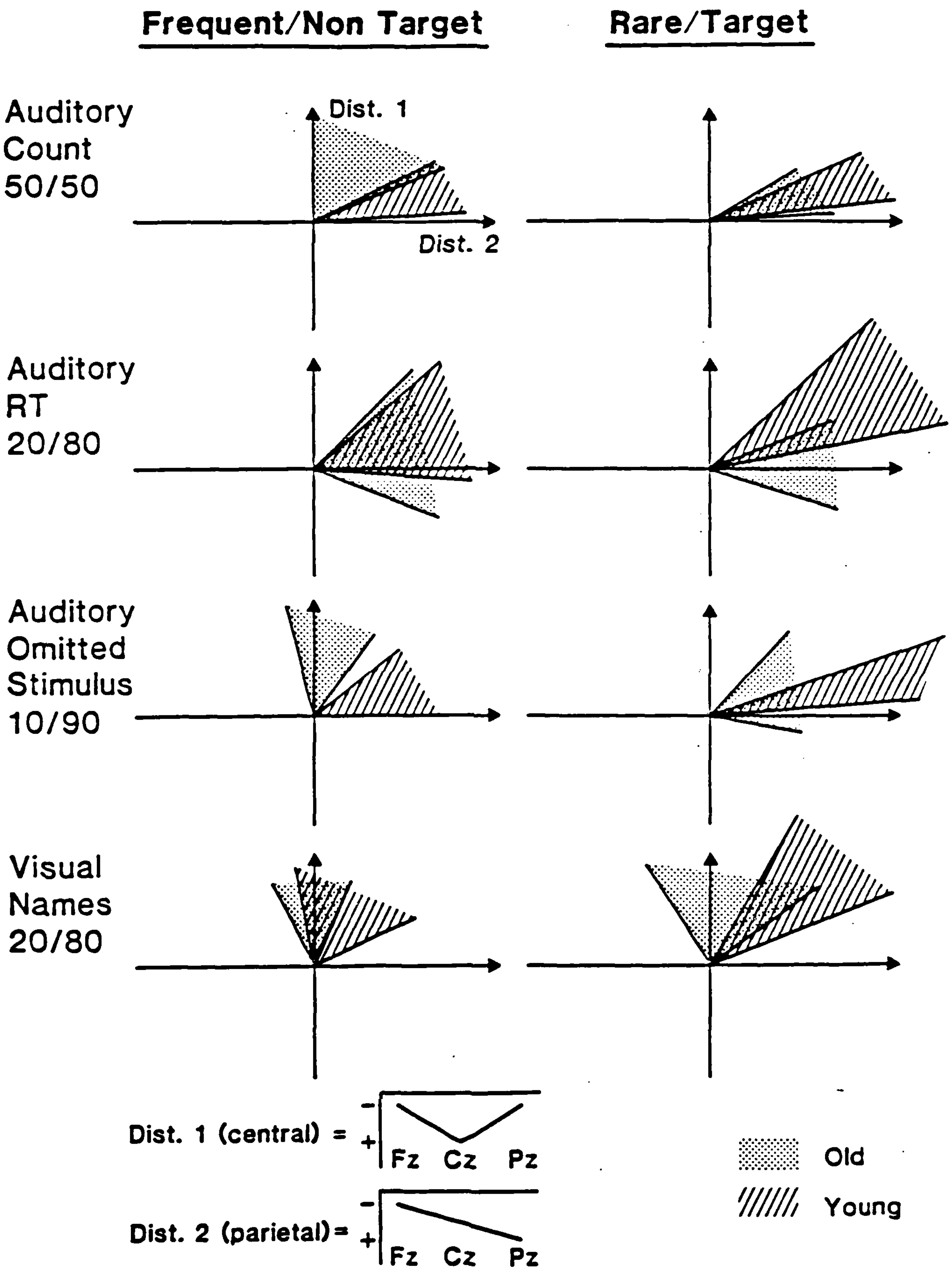


Tangent of Polarity Angle
0
2
4
6
8
10
12
14
Old Group:
Rare/Target
Frequent/Non Target
Young Group:
Rare/Target
Frequent/Non Target
Auditory
Count
50/50
Target/Non Target
Auditory
RT
20/80
Rare/Frequent
Auditory
Omitted Stimulus
10/90
Rare/Frequent
Visual
Names
20/80
Rare/Frequent

Fig. 6

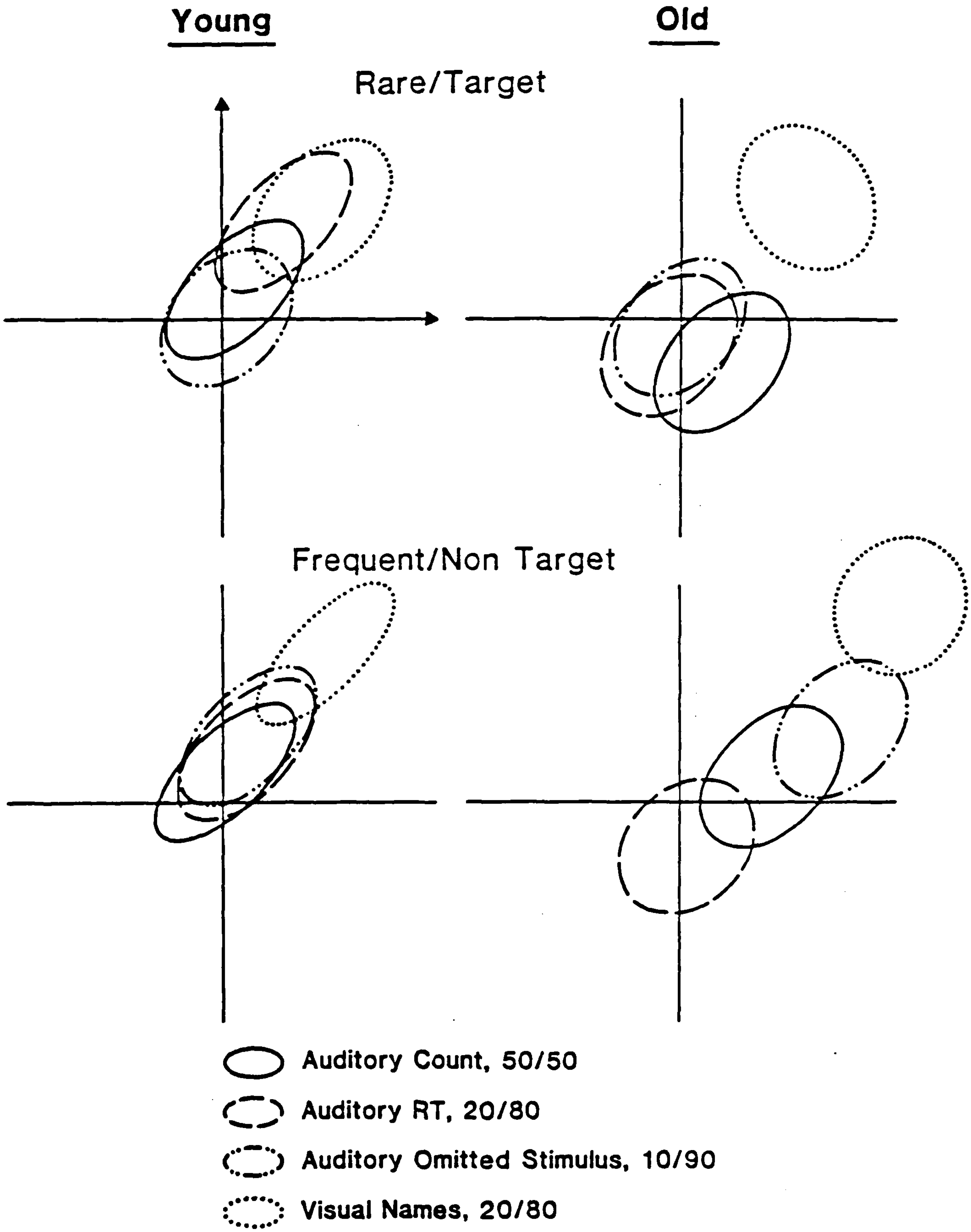


Frequent/Non Target
Rare/Target
Auditory Count 50/50
Dist. 1
Dist.2
Auditory RT 20/80
Auditory Omitted Stimulus 10/90
Visual Names 20/80
Dist. 1 (central) =
-
+
Fz Cz Pz
Dist. 2 (parietal) =
-
+
Fz Cz Pz
Old
Young



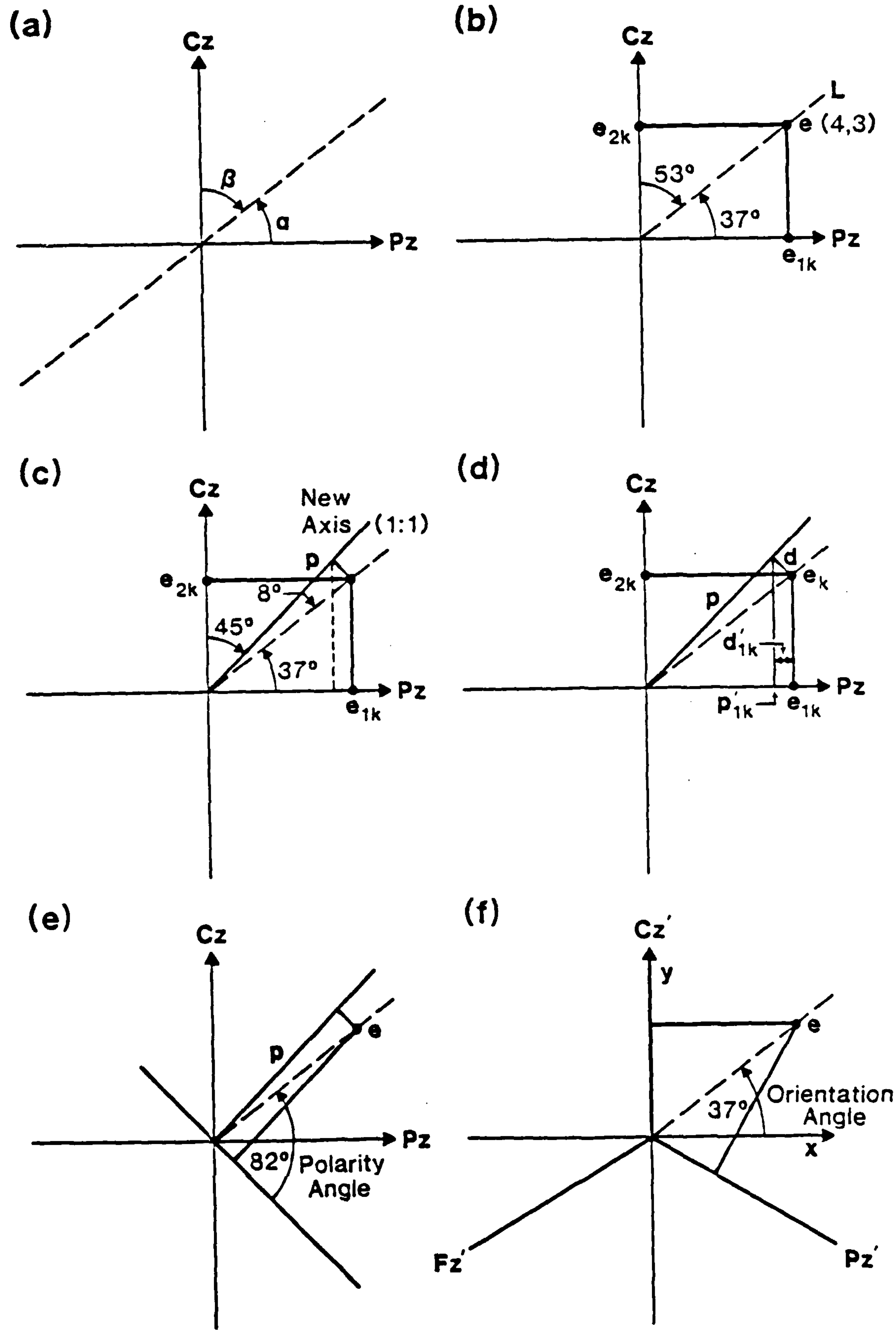
Frequent/Non Target
Rare/Target
Auditory Count 50/50
Dist. 1
Dist. 2
Auditory RT 20/80
Auditory Omitted Stimulus 10/90
Visual Names 20/80
Dist. 1 (central) =
Fz Cz Pz
Dist. 2 (parietal) =
Fz Cz Pz
Old
Young

Fig. 8



Young
Old
Rare/Target
Frequent/Non Target
Auditory Count, 50/50
Auditory RT, 20/80
Auditory Omitted Stimulus, 10/90
Visual Names, 20/80

Fig. 9



(a)
Cz
β
α
Pz
(b)
Cz
L
e2k
e (4,3)
53°
37°
Pz
e1k
(c)
Cz
New
Axis
(1:1)
p
e2k
8°
45°
37°
Pz
e1k
(d)
Cz
d
e2k
ek
p
d'1k
Pz
p'1k
e1k
(e)
Cz
p
e
Pz
82°
Polarity
Angle
(f)
Cz'
y
e
37°
Orientation
Angle
x
Fz'
Pz'

Appendix B11

Submitted to Electroencephalography and Clinical Neurophysiology

A SIMULATION STUDY OF LATENCY MEASURES OF COMPONENTS OF EVENT-RELATED BRAIN POTENTIALS

Gabriele Gratton, Arthur F. Kramer, Michael G.H. Coles, and Emanuel Donchin 1

Cognitive Psychophysiology Laboratory
Department of Psychology
University of Illinois, Champaign, Illinois

Running title: A simulation study of latency measures

Send all correspondence to:

Gabriele Gratton

University of Illinois

Psychology Department

603 E. Daniel

Champaign, Illinois, 61820

A Simulation Study of Latency Measures of Components of Event-Related Brain Potentials

Gabriele Gratton, Arthur F. Kramer, Michael G.H. Coles, and Emanuel Donchin

Introduction

In this paper we compare, by means of a simulation study, several of the techniques that are commonly used for the measurement of the latency of components of the Event-Related Brain Potential (ERP). It is common for investigators to consider the ERP as a series of "components," and to specify the experimental results in terms of the relationship between independent variables and some attributes of a component. Components, in turn, are interpretable as a manifestation of the activity of functional units in response to, or in preparation for a particular event (Donchin, Coles, & Gratton, 1984). Components are customarily defined in terms of polarity, latency, scalp distribution, and sensitivity to experimental manipulations (Donchin, Ritter, & McCallum, 1978; Sutton & Ruchkin, 1984).

Theoretical speculation (e.g., Coles & Gratton, in press; Coles, Gratton, Bashore, Eriksen, & Donchin, in press; Donchin, 1981) and empirical evidence (e.g., Duncan-Johnson, 1981; Duncan-Johnson and Donchin, 1982; Kutas, McCarthy, & Donchin, 1977; McCarthy & Donchin, 1981; Ragot, 1984) suggest the utility of the latency of components of the ERP (and of P300 in particular) in the study of cognition. However, since ERP components are estimated from data containing substantial amounts of noise, the measures of latency necessarily yield but approximations to the "true" latency of the components. In this study, we compare the accuracy of latency estimates when they are obtained by means of several different latency detection procedures. We also compare the impact of different signal-to-noise ratios on the estimates. We simulated ERP data so that the known ERP was embedded in noise of different amount and quality. The degree to which different procedures accurately estimates the known parameters of the ERP components could then be assessed.

The need for such a comparison is evident. While the importance of measuring latencies is generally acknowledged, there is little consensus regarding the methods (see Fabiani, Gratton, Karis, & Donchin, in press, for an overview of different procedures used for the definition of one particular ERP component - P300). Seemingly conflicting results appear, and the conflict may be due entirely to the incommensurability of the methods. Thus, our results can be used as a source of guidelines for the choice of an appropriate procedure for the estimation of component latency. Of course, such guidelines are confined to the domain explored by this study.

A procedure intended to estimate the latency of an ERP component should take into account the characteristics of both the signal (i.e., the ERP component) and the noise (i.e., all the other electrical activity recorded by the scalp electrode). However, most component detection algorithms focus solely on the characteristics of the signal. We will show that, if the characteristics of the noise are considered, the accuracy of the latency estimation can be improved.

From the point of view of an investigator of the ERP, there are at least two categories of noise. The electrodes record, in addition to the ERP, the background EEG (random noise) and the activity of concurrent ERP components, other than the object of study (systematic noise). Previous studies simulating ERP data tended to focus solely on the effect of random noise (intended to simulate background EEG) on the accuracy of latency estimates (see Pfefferbaum, 1983; Woody, 1967). A thorough evaluation of latency estimation procedures can be obtained only when both of these sources of noise are considered.

The term "background EEG" labels the electrical brain activity which is not time-locked to the external event. It is generally assumed that the background EEG has a random phase in relation to the triggering event. However, this activity is technically not random, as some frequency bands may be dominant. The presence of strong auto-correlation functions in the background EEG activity may affect signal detection procedures which are based on the auto-regressive properties of the signal (cf. auto-correlation procedures). In fact, the background EEG activity may share some characteristics with the signal (for instance, frequency). Such noise may lead to a detection of a component where, in fact, there is only noise (Pfefferbaum, 1983). To allow the sources of such "false alarms" to play a role in the simulation, the frequency characteristics of the background EEG must be reproduced in the simulated waveforms.

Unfortunately, the characteristics of the background EEG activity occurring during an ERP experiment are not well known. Several studies have concurrently examined ERP and EEG frequency power spectra (for example, see McCarthy & Donchin, 1978). However, the presence of ERPs makes the frequency power spectra that are obtained with such procedures poor estimates of the background EEG activity (see Ungar & Basar, 1976, for a related discussion on the utility of "Wiener's Filter," Wiener, 1949).

Several procedures have been adopted to simulate the background EEG noise. Two criteria for the choice of such procedures have been stressed:

face validity of the noise, and ability to control and manipulate noise characteristics. However, the characteristics of the background EEG noise are known only in part and are not necessarily constant over different experimental conditions. Therefore the simulation of background noise is difficult. An example of a first order auto-regressive noise simulation procedure is illustrated in Donchin and Herning (1975).

In the present study, we simulated the background noise by using the deviation of single-trial ERPs from the average ERP. This definition of noise corresponds to that implicit in signal averaging technique. It may be considered valid in those cases in which the latency, amplitude, and morphology of a component are reasonably constant over trials.

The second source of noise is provided by those ERP components which are active in the same time segment as the target component. In this case, since ERP components are believed to be invariant across trials (again, following the assumptions of signal averaging), the average value of the noise over a large number of trials represents an estimate of the amplitude of the overlapping component(s). Note that this value is not equal to zero as is the case with averages of random noise. The error induced by such systematic noise is particularly insidious because it may vary as a function of experimental manipulations. Furthermore, latency detection procedures might be differentially affected by overlapping components. To study the effects of overlapping components on the accuracy of latency estimates in the present study, we added a set of components to the waveforms. The components varied in the degree of temporal and spatial overlap with the target component. The amplitude and latency of the components were systematically varied.

The relative amplitudes of the component (signal) and of the noise are important in determining the accuracy of the detection of the component. Several studies have demonstrated that detection accuracy increases monotonically with increases of the signal-to-noise ratio (Nahvi, Woody, Ungar, & Sharafat, 1975; Pfefferbaum, 1983; Wastell, 1977; Woody, 1967). However, different procedures may be differentially affected by the same increase in the signal-to-noise ratio, and a procedure that is more accurate at one signal-to-noise ratio may be less accurate at another signal-to-noise ratio (see a related discussion about Wiener's Filter in Wastell, 1981).

Investigators have used several procedures to deal with the problem of the low signal-to-noise ratio in ERP research. One of these procedures involves deriving several EEG epochs, each of which is time-locked to the same external triggering event. Isolation of the ERP is accomplished by averaging over epochs (repetitions). Of course, averaging is based on the assumption of invariance of the signal (ERP) over trials. If this assumption is met, the signal-to-noise ratio will vary as a function of the number of trials considered (one in the case of single-trials estimates, and a variable number in the case of averages).

However, several investigators have described instances in which the amplitudes (e.g., Squires, Wickens, Squires, & Donchin, 1976) and/or latencies (e.g., Kutas et al., 1977) of ERP components vary from trial to trial. Such variability suggests caution in the use of average waveforms to estimate component latency. In fact, a trial associated with a large component amplitude will have more weight in determining the average waveform than a trial associated with a small component amplitude. Therefore, the latency of the average waveform will be mostly determined by those trials associated with large component amplitudes. Latency variability may also affect amplitude estimates, by reducing the amplitude of the component peak (see Donchin & Heffley, 1978, for a discussion). This problem is usually labelled "latency jitter."[2] Latency jitter is a serious problem in the analysis of ERPs, because the waveshape of the average waveform may not resemble the "real" ERP, however large is the number of trials. Furthermore, when latency jitter is present, there may be a difference between computing the latency on the average waveform, or computing it on each single trial and then averaging the single-trial latency estimates, depending on the distribution of the single-trial latencies, amplitudes, and their intercorrelation.[3]

In the present study, we distinguish between two classes of techniques which are used in the estimation of the latency of ERP components: (a) preparation, or filtering, procedures, and (b) signal detection procedures. The purpose of the filtering techniques is to increase the signal-to-noise ratio, by capitalizing on the differences in the physical properties of the signal and of the noise. These techniques prepare the data for the latency estimation, that is carried out by the signal detection procedures. We will consider two kinds of filtering techniques, frequency filters and spatial filters.

The function of frequency filters is to eliminate electrical activity of undesirable frequencies. However, the effect of frequency filters on latency estimates has not been thoroughly explored. Both on-line analog and off-line digital filters are used (for a discussion of both kinds of filters see Coles, Gratton, Kramer, & Miller, in press). On-line analog filters generally introduce phase shifts which result in distortions of the latency estimates. The magnitude of the phase distortion depends on the band-pass characteristics of the filter (see Duncan-Johnson and Donchin, 1977, for a discussion of the effect of high-pass filters on ERP waveforms). For this reason most researchers use broad-band filters in the collection of ERP data. Off-line digital filters can be designed in such a way as to avoid the introduction of phase shift. Such filters may be used during the preparation of the ERP data for signal detection techniques. In this study, we will focus on low-pass off-line digital filters with no phase distortion (Ruchkin & Glaser, 1978).

Scalp distribution information has not often been used in the preparation of data for latency estimation (see Nahvi et al., 1975, for an attempt to use multichannel information for improving signal detection). In general, researchers have simply selected one electrode location to use for further analysis (we will label this procedure "channel selection"). One of the goals of this study is to evaluate the use of scalp distribution information to improve the detection of a component at different signal-to-noise ratios. Scalp distribution information can be used by adopting a procedure (Vector Filtering, VF), described by Gratton, Coles, and Donchin (submitted for publication). VF assumes that scalp distribution is a defining characteristic of an ERP component. Therefore information on scalp distribution can be employed to discriminate among ERP components and between ERP components and noise. We will compare the accuracy of the latency estimates obtained from data prepared with VF and with channel selection. VF can be thought of as a linear filter where the weights for each channel are chosen in order to optimize the detection of the target component. Channel selection can be thought of as a linear filter giving a weight of 1 to the selected channel, and a weight of 0 to all the other channels.

The task of a signal detection technique is to detect the signal under noisy conditions (for a review of signal detection techniques in ERP research, see Coles, Gratton, Kramer, & Miller, in press). Two types of signal detection techniques are commonly used by ERP researchers in the study of a component's latency: peak-picking and cross-correlation. These techniques differ in the way a signal is defined. Peak-picking identifies a component as the maximum or minimum value in a certain time window. Note that only the time-point at which the peak occurs is used to estimate the parameters (amplitude and latency) of the component. Cross-correlation techniques define a component in terms of its waveshape (Derbyshire, Driessen, & Palmer, 1967; Palmer, Derbyshire, & Lee, 1966). They define a component as that segment of waveform whose shape maximally "resembles" an externally defined segment of waveform, labelled the "template." "Resemblance" is assessed by means of a correlation or covariance measure.

The difference between peak-picking and cross-correlation may be conceptualized in terms of the number of data points considered for component detection. Peak-picking techniques use only a single point, while cross-correlation techniques use a set of points (a segment of the waveform).

Woody (1967) proposed a particular variant of the cross-correlation procedure, where the template is "adapted" to the average waveform through several iterations. Woody's procedure (labelled "Woody Filter") has been applied to ERPs (e.g., Kutas et al., 1977) with the main purpose of overcoming the problem of latency jitter. In fact, the waveshape of the target component (in Kutas et al.'s study, the P300) was not known, and could not be estimated from the average waveform. The justification for the Woody Filter procedure was therefore to obtain estimates of the waveshape of P300 that were progressively more accurate at each iteration.

Wastell (1977) investigated the utility of the iterating procedure proposed by Woody (1967). He reported that, if an appropriate template has been selected, iterations do not improve the detection of the signal. Furthermore, Pfefferbaum (1983) found that Woody's iterations may extract spurious, artifactual components that are indistinguishable from real components when the signal-to-noise ratio is very low. He concluded that cross-correlational techniques (and Woody Filter in particular) produce the best results at relatively high signal-to-noise ratios. These studies suggest that the reliability of a signal detection procedure should be evaluated at several signal-to-noise ratios. Since filtering procedures affect the signal-to-noise ratio, the impact of these procedures on different signal detection procedures will also be considered in the present study.

Method

There were three phases to the present study. First, simulated waveforms were generated according to a particular model of ERPs. Second, procedures were applied to these waveforms to obtain latency estimates. These procedures included filtering and signal detection techniques. Finally, the accuracy of the latency estimations was computed, and the merits of each procedure were evaluated.

General design

We should note the differences between the basic design and the control conditions we ran to investigate particular problems. The basic design consisted of a factorial manipulation of several simulated conditions and analysis procedures. The simulated conditions yielded 220 different conditions (11 P300 conditions x 20 overlapping component conditions). For each condition we obtained 100 repetitions by adding background EEG noise, sampled at random and with reselection from 100 noise trials (note that relationship between waveform and recording electrode - Fz, Cz, and Pz - was maintained). This yielded 22,000 waveforms (for each electrode). Each of these waveforms was filtered using five filtering conditions. This produced a total of 110,000 waveforms for each electrode, each of which was filtered spatially using either channel selection or VF. Finally, we applied four signal detection algorithms (peak-picking, cross-correlation, Woody Filter with 2 iterations, and Woody Filter with 3 iterations) to obtain latency estimates for each of the 2 x 110,000 waveforms.

The comparisons between signal detection algorithms and spatial filtering techniques (as well as their interaction) were based on repeated measures, while the comparison between frequency filtering procedures was based on independent measures, resulting in a nested design. The component manipulation yielded independent measures (different noise components were added to each condition). A list of the experimental conditions for the basic design is shown in Table I.

Insert Table I about here

Model

The study was based on the following model:

$$E_{it} = a_i C_t + \sum (k_{ji} S_{jt}) + R_{it}$$

where:

E_{it} is the potential recorded at the electrode i at time t;

a_i is the weight of the target component at the electrode i;

C_t is the amplitude of the target component at time t;

k_{ji} is the weight of the overlapping component j at the electrode i;

S_{jt} is the amplitude of the overlapping component j at time t;

R_{it} is the background EEG noise at the electrode i at time t.

In adopting this model, we assume that each component is characterized by a particular scalp distribution, defined by a series of weights, one for each electrode location. This assumption is similar to that proposed for the VF procedure (Gratton et al., submitted for publication).

Single trials

Each assessment of the accuracy of the latency estimates was based on 100 repetitions, at three different electrode locations (labelled Fz, Cz, and Pz). Each repetition was called a "trial" and was obtained by adding several time series (vectors) point by point. The trials were constructed by adding a different noise vector to each of 100 identical component vectors. The vectors consisted of 128 data points that we considered to have been recorded at 100 Hz digitizing rate, starting 200 msec before a hypothetical stimulus. The average of the first 20 points was considered as an estimate of the "pre-stimulus" baseline level and was subtracted from the data.

Simulation of ERP components

One target and four non-target components were obtained by adding together five cosinusoidal waves. The amplitude, latency and duration (wavelength) of the cosinusoidal waves could be varied, and each component was simulated by using different parameters. The scalp distribution of each component was simulated by multiplying the vector by a different scaling factor for each electrode. The target component simulated the "P300" component, and the non-target components simulated the "N100", "P200", "N200", and "Slow Wave" (SW) components. The parameters (amplitude, latency, duration, and scalp distribution) of each component do not correspond to data obtained in a particular experiment. However, an attempt was made to reproduce the parameters of the components described in the ERP literature (see Donchin et al., 1978). The parameters of N100 and P200 were not varied systematically, since they do not overlap temporally with P300, and, therefore, should not affect the P300 latency estimates. The amplitude and latency of N200 and SW were systematically varied to simulate different degrees of overlap with P300 that may be present in real data. A control condition in which only the P300 component was present was also included in the study. The parameters adopted for each component are presented in Table II.

Insert Table II about here

P300 amplitude was varied systematically from 0 (absence of the component) to 500 units, with increments of 50 units. Each unit was intended to be equivalent to .1 microvolts, so that P300 amplitude varied from 0 to 50 microvolts. This manipulation of P300 amplitude allowed us to evaluate the different procedures over a wide range of signal amplitude conditions. Given the complex procedure we adopted to simulate noise (described below), and in particular the presence of systematic noise, we could not express the true signal-to-noise ratio as an absolute value.

However, we could compute the ratio between the amplitude of the signal and the root mean square amplitude of the background EEG noise that was fixed at 100 (as shown later). We chose to label this value "signal-to-noise ratio." This ratio varied systematically from 0 to 5, in half unit increments. Both N200 and SW partially overlapped with P300. However, the scalp distribution of these components had different degrees of overlapping with that of P300. N200 had a frontally maximum scalp distribution that was clearly different from that of P300. The scalp distribution of the SW was maximally positive at Pz, and was therefore more similar to that of P300. Thus, the manipulations allowed us to study the impact of different levels of component overlap (with manipulations of latency, amplitude, and scalp distribution) on the P300 latency estimates. The five components were added to obtain complex ERP waveforms. An example of these waveforms is presented in the upper panel of Figure 1.

Insert Figure 1 About Here

Component amplitudes and latencies were factorially combined. The design included 2 x 2 (amplitude and latency) N200 manipulations, 2 x 2 SW manipulations, and four repetitions of the control condition, with a total of 20 conditions of component overlap. As there were 11 levels of P300 amplitude, 220 basic ERP waveforms were obtained for each electrode.

Background EEG simulation

Background EEG activity was simulated by obtaining non-event-related activity from a set of 100 trials recorded from an individual subject in an oddball experiment (Fabiani et al., in press).4 The non-event-related activity was obtained by subtracting the average appropriate for each event from each single-trial record.

This procedure yielded a set of waveforms whose average is a flat line. However, the variability from trial to trial is not equal for all time-points and electrodes. In particular, larger intertrial variance was observed at a latency of approximately 300 msec, and smaller variance during the pre-stimulus period. Under these conditions, the characteristics of these waveforms could not be considered stationary over the whole epoch, and, therefore they could not be considered as good estimates of the background EEG noise. A further disadvantage was that the impact of noise could vary as a function of the latency.

Therefore we chose to standardize each time-point (and electrode), with a mean of 0, and a standard deviation of 100 units (10 microvolts). We considered the resulting 100 waveforms for each electrode as our simulated background EEG activity. An analysis of the frequency characteristics of the simulated background noise after the standardization did not reveal any difference from the "non-standardized" noise. However, we also ran a control condition with non-standardized waveforms.

An example of single-trial waveforms obtained by adding the complex ERP waveform and the standardized background noise is shown in the lower panel of Figure 1. Note that the waveforms obtained by adding the noise are different in several aspects from those not containing noise: the relationship among electrodes is altered, double peaks are noticeable, etc.

Off-line Frequency Filtering

The study included a comparison between five off-line, low-pass frequency filtering procedures. All of them were based on a moving average filter (Ruchkin & Glaser, 1978). The procedure differed in the number of consecutive time-points used for the smoothing (length), and in the number of iterations of the procedure adopted. Two length levels (7 and 13 points, roughly equivalent to a 6.29 and a 3.14 Hz half amplitude cut-off filter respectively), and two iteration levels (1 and 2 iterations) were used.5 As a control, we used a condition where no off-line frequency filter was applied.

We should emphasize that the comparison between filtering procedures described above does not exhaust all of the off-line frequency filters available to the investigator. We intend only to evaluate the effects of several frequency filters on latency estimations, and to determine whether, and to what extent, the general practice of smoothing waveforms improves the component latency estimation.

Spatial Filtering

Two spatial filtering procedures were compared: channel selection and VF.

Channel selection. This procedure consists of the selection of one electrode for further analysis. Given that our P300 component was maximum at Pz, we chose this electrode for the analysis.

Vector Filtering. This procedure, described by Gratton et al. (submitted for publication), is based on the assumption that ERP components are characterized by a specific scalp distribution that can be expressed by a series of weights, one for each electrode. Gratton et al. (submitted for publication) showed how sets of weights describing specific scalp distributions correspond to lines with different orientations in a space defined by the scalp locations.6 The orientations corresponding to the scalp distribution of the components used in the present study are shown in Figure 2. Note that the axes in this figure do not correspond to the electrodes, but to specific scalp distributions (respectively, that of the simulated P300 and N100 components in our study). Note also that, since three electrodes were used, a complete description of any possible scalp distribution would need a three-dimensional plot. The third dimension, not shown in Figure 2, corresponds to the mean over electrodes, while the two dimensions shown in Figure 2 describe the variance among electrodes (cf. Gratton et al., submitted for publication).

Insert Figure 2 About Here

The VF allows the experimenter to estimate the contribution of a specific, target component (defined in terms of scalp distribution) to a particular time-point. The scalp distribution of the target component also corresponds to a particular orientation in the space defined by the electrode locations. The result of VF is an estimate of the amplitude of the target component for each datapoint. However, rather than considering this procedure as a "signal extraction" technique, we prefer to consider it as a filtering technique used to prepare the data for signal detection. By applying VF, we eliminate from our data the part of the electrical activity that, because of its scalp distribution, does not represent the target component.

The scalp distribution we filter for may not necessarily be that of the target component (in our case, P300). In fact, filtering for a different scalp distribution might produce even better results (i.e., improvement of the signal-to-noise ratio) than filtering for the target component. This may particularly be the case when some patterns of scalp distribution are more likely than others to be represented in the noise. This condition is generally true for both systematic and background noise. We observed that in our case noise activity at the central and parietal electrodes was strongly correlated (the correlation was above .80). For this reason, a larger reduction of noise and consequently an improvement of the signal-to-noise ratio, will be obtained by filtering for a scalp distribution where noise activity will not be strongly represented. In particular, filtering for a scalp distribution which dissociates the activity of electrodes with strong noise coherence functions is advisable. This rationale was at the basis of our choice of VF parameters (i.e., angles describing its orientation). In fact, these parameters produced a dissociation between the activity of the parietal and central electrodes.

As parameters for our VF we chose a polarity angle of 15 degrees and an orientation angle of 300 degrees (see Footnote 6 for a description of these parameters). These values do not correspond to the scalp distribution of P300. The relative scalp distribution associated with an orientation angle of 300 degrees is also shown in Figure 2, and corresponds to a parietal maximum, but central minimum scalp distribution. The choice of the specific parameters was based on a previous study on real data in which they were found to produce on optimal discrimination between two groups of trials (rare and frequent trials in an oddball paradigm) with different P300 amplitude. As a control for our choice, we ran a condition in which we parametrically varied the parameters of the VF (polarity and orientation), to determine which parameters resulted in the best improvement in the estimation of P300 latency.

Signal detection techniques

The signal detection algorithms included peak-picking and cross-correlation techniques. The peak-picking algorithm is based on the detection of the maximum value in a prespecified temporal window. Cross-correlation techniques involve the computation of a series of correlations between segments of the ERP waveform progressively shifted by one data point and a template, which represents the ideal signal to be detected. The segment of the ERP associated with the largest correlation is considered to contain the component (signal). Woody (1967) proposed the use of an adaptive template, obtained with an iterative procedure (Woody Filter). The template for the first iteration is usually the average ERP, but sometimes arbitrary templates are used, like a sinusoidal wave (Pfefferbaum, 1983). However, in subsequent iterations, the template is always extracted from the data by averaging the segments of ERP with a maximum correlation with a template derived from a previous iteration.

In this study we adopted a procedure which allowed us to evaluate both cross-correlation and Woody Filter. We used a template equivalent to the P300 component we entered in the simulated waveforms as the first iteration of the Woody Filter procedure. Thus, the first iteration corresponded to a cross-correlation algorithm, while the subsequent iterations corresponded to successive iterations of the Woody Filter. The estimate of P300 latency was obtained by selecting the central value of the ERP segment with the maximum correlation with the template.

Since the template we used for the cross-correlation technique was the target component itself, the detection of P300 obtained with this algorithm may be more accurate than that which could be obtained in real (non-simulated) conditions, when the "true" P300 waveshape is not perfectly known. Therefore, the basic design does not really address the question of the advantage of iterating with Woody Filter. To address this question we added cases in which we did not have a good representation of the P300 waveshape. We ran two control conditions:

1. A condition in which the duration of the simulated P300 was parametrically varied between 100 and 1000 msec, while the duration of the cosinusoidal wave adopted as the template was fixed at 500 msec.

2. A condition in which, at each single trial, the latency of the simulated P300 was sampled from a random distribution with a mean of 500 msec and a standard deviation of 83 msec. There where three levels of P300 duration (300, 500, and 700), and the duration of the template for cross-correlation was fixed at 500 msec.

For each signal detection procedure, the temporal window began 300 msec post-stimulus and ended 800 msec post-stimulus (respectively, 250 msec before and 250 msec after the P300 peak). The duration of the template used for the cross-correlation algorithm was 500 msec.

Error (accuracy) estimation

As described above, 100 repetitions were obtained for each condition. To assess the accuracy of latency estimation obtained with each procedure under each condition, the root mean square error (MSE) value was calculated. This value was obtained as follows:

$$MSE = \sqrt{\sum_{i=1}^{n} (l_i - L)^2 / n}$$

where:

MSE is the root mean square error of latency estimate;

n is the number of trials;

l_i is the latency estimate at trial i;

L is the P300 peak latency (550 msec).

Most of the figures presented in the following sections of this paper show variations of the MSE value as a function of variations of the signal-to-noise ratio (i.e., P300 amplitude). In the plots presented here, a logarithmic scale is used.7

Another dependent variable we used was an approximate estimate of the number of trials required to reduce the standard error of estimate to 3 msec. This measure was used to assess the relative power of the different procedures and was obtained as follows. The MSE can be considered an estimate of the standard deviation of the population of single-trial P300 latency estimates for each condition (note that, in this case, the population mean is known). However, the mean of the single-trial estimates of the sample may not correspond to the mean of the population (550 msec). If the normality assumption is met, we can compute the theoretical distribution of the population of sample means from which the mean of our sample is extracted. Following the theorem of central tendency, this distribution will have a width (measured by the standard error of estimate) proportional to the MSE (standard deviation) and inversely related to square root of the number of trials used to compute the mean. By increasing the number of trials we may theoretically reduce the standard error of estimate to any desired value. Thus, by appropriately setting the sample size, we may in theory obtain a standard error of estimate of 3 msec, given a specific value of the MSE. In fact, we can compute the sample size required with the following equation:

$$N = (MSE / 3)^2 + 1$$

where N is the number of trials required to obtain a standard error of 3 msec. Note that this value is only an approximation. In fact, it requires (a) that the distribution of the single-trial estimates is normal, and (b) that the sample mean is not systematically different from 550 msec. The first assumption is violated, since only values inside the time window (300 to 800 msec) are possible. However, the distribution of the single-trial estimates is approximately normal when the signal-to-noise ratio is larger than 1. Examples of distributions of single-trial estimates for different signal-to-noise ratios will be shown later. The second assumption may also be violated in some cases, but it holds in most cases. Since the number of trials required to obtain a standard error of estimate of 3 msec are related to the MSE, we simply added a scale reporting the corresponding values for this dependent variable in most of the figures in which MSE (or log MSE) is presented.

Results and Discussion

This section is divided into two parts: first, we present the results obtained from the basic design; second, we discuss a series of additional analyses we ran to investigate the effect of non-standardized background noise, variations of P300 duration and latency, and variations of the parameters of VF, on the accuracy of the latency estimates.

Basic design

Non-overlapping component condition. The basic design was devised to permit a comparison of the accuracy of several procedures over a wide variety of signal and noise conditions. As a reference point we will first present the data obtained in the condition in which no overlapping components were present. The MSE values (averaged across 400 repetitions) for this condition are shown in Fig. 3.

Insert Figure 3 About Here

As a reminder, in this and most of the following figures the abscissa represents the signal-to-noise ratio (i.e. P300 amplitude), while the ordinate represents the MSE (logarithmic transform). Since the basic design was not intended to specifically investigate the validity of the iterative procedure proposed by Woody (1967), the results obtained with the second and third iterations of the Woody Filter are not shown in this figure. However, the MSE values obtained with the Woody Filter were similar to those obtained with the cross-correlation procedure.

Several important effects are apparent in figure 3. First, variations of the signal-to-noise ratio produced the largest effects on the accuracy of estimation. As expected, at a signal-to-noise ratio of 0, all the procedures produce about the same results. The MSE at this signal-to-noise ratio is close to that which would be obtained by picking points at random in the temporal window. In fact, the log MSE obtained in this way is 2.2. By increasing the signal-to-noise ratio, exponential decreases of the MSE can be observed (the functions approximate a straight line in the figure because of the logarithmic scale used for the ordinate). At a signal-to-noise ratio of 5, the MSE value is one tenth of that found at a signal-to-noise ratio of 0. Second, the use of cross-correlation as a signal detection procedure yielded lower MSE values than peak-picking. The gain in accuracy obtained with cross-correlation may be as high as 50%, at very high signal-to-noise ratios, and when no frequency filter is applied. Third, the use of VF as a spatial filtering technique reduced the error in latency estimation in comparison with channel selection (Pz). This advantage is evident at middle and high signal-to-noise ratios. At low or middle signal-to-noise ratios (.5 to 2.0), the advantage of VF is comparable to that of cross-correlation. However, the advantage of VF rarely reaches the 50% level and is usually about 25%. The advantages of VF and of cross-correlation appear to be independent. Fourth, low-pass frequency filters produced marked improvements of the accuracy of latency estimation. The largest improvement was obtained with a a 3.14 Hz low-pass filter iterated twice. Filters with a wider band-pass produced less improvement. However, this effect was particularly evident when a peak-picking algorithm was used for signal detection. The gain for cross-correlation was small. In fact, the effect of the frequency filters was to bring peak-picking to the same level of accuracy as cross-correlation. The gain obtained with VF was unaffected, and in fact the smallest MSE values were obtained by the joint use of frequency filters, VF, and cross-correlation.

Histograms of the latency estimates for each single trial in the non-overlapping component condition, without frequency filter, are shown in Figure 4.

Insert Figure 4 About Here

The distribution at a signal-to-noise ratio of 0 was approximately rectangular, indicating that no point was more likely to be chosen than any other when no signal was present (apart for a small preference for the first point in the epoch). At higher signal-to-noise ratios, central values becomes progressively more represented. The mode tends to correspond to the actual P300 latency. An exception to this general rule can be observed at a signal-to-noise ratio of 2.5 for the peak-picking algorithm on Pz waveforms. A skewed distribution indicates the presence of systematic error (the average estimated latency does not correspond with the real P300 latency).

Latency-adjusted average waveforms obtained with cross-correlation and second and third iterations of Woody Filter for the non-overlapping component condition, at extreme levels of the signal-to-noise ratio and no frequency filter, are shown in Figure 5.

Insert Figure 5 About Here

Inspection of this figure reveals that, even when no ERP component is present (signal-to-noise ratio is equal to 0), the latency adjustment procedure "creates" one. When the component is large, the distortion produced by the latency adjustment is negligible. This finding is in agreement with the results reported by Pfefferbaum (1983). The artifactual component created by the latency adjustment appears larger when the signal detection algorithm is applied to Pz waveforms, than for waveforms obtained with VF. This effect is confounded in part with an overall reduction in amplitude produced by VF. The artifactual component appears also to have the same amplitude if the latency adjustment is obtained after the cross-correlation procedure, Woody Filter with one iteration, or Woody Filter with two iterations. Thus, the problem does not seem to be related to the number of iterations, but rather to the latency adjustment procedure per se.

Overlapping component conditions. Logarithms of MSE (and number of trials required to obtain an error of 3 msec) for four different component overlap conditions with two spatial filtering and two signal detection procedures (without frequency filtering) are shown in Figure 6.

Insert Figure 6 About Here

The "no-component overlap" condition is shown in the upper left panel for comparison. The other three conditions shown in the figure were "small component overlap" (N200 amplitude = 50 units, N200 latency = 300 msec, SW amplitude = 100 units, SW latency = 1280 msec), "large N200 overlap" (N200 amplitude = 100 units, N200 latency = 400 msec, SW amplitude = 100 units, SW latency = 1280 msec), and "large SW overlap" (N200 amplitude = 100 units, N200 latency = 300 msec, SW amplitude = 200 units, SW latency = 1000 msec). Inspection of this figure reveals that component overlap impaired the accuracy obtained with each procedure to a different degree. In particular, procedures based on channel selection (Pz) were markedly affected by component overlap, both of N200 and SW. Procedures based on Vector filtered data were also affected by SW overlap, but were not affected by N200 overlap. We should note here that the scalp distribution of SW was close to that of P300, while the scalp distribution of N200 was different. Thus, estimates obtained on Vector filtered data are not affected by overlapping components with a scalp distribution different from that of P300, but are affected by an overlapping component with a scalp distribution similar to that of P300. However, even in the worst case (large SW overlap) estimates obtained on Vector filtered data are no worse than those obtained on Pz waveforms.

For reasons of space, we cannot present here the results obtained with all the other combinations of component overlap, signal-to-noise ratio, frequency filtering, spatial filtering, and signal detection algorithm. However, these results confirm the observations we have presented so far.

In summary, the following conclusions can be drawn:

1. The error of latency estimation decreases exponentially as the signal-to-noise ratio increases.
2. Cross-correlation provides a more accurate estimate than peak-picking.
3. Woody Filter with 2 or 3 iterations is comparable to cross-correlation (although this result may be due to the fact that the template for cross-correlation was equal to the target component).
4. Frequency f ters improve markedly the accuracy of estimates obtained with peak-picking and, to a lesser extent, with cross-correlation.
5. VF yields estimates that are more accurate than channel selection (Pz).
6. Overlapping components impair the accuracy of estimates of

channel selection, while the latency estimates of Vector filtered data are impaired only if the scalp distribution is similar to that of P300 (e.g. SW).

7. The accuracy improvements obtained with cross-correlation, VF, and increases in signal-to-noise ratio appear to be independent (additive). The improvements in accuracy obtained with VF and frequency filtering are also independent.
8. The accuracy improvements obtained with cross-correlation and frequency filtering are not additive, that is, the combined use of both these procedures does not produce much better results than either of them alone.
9. Latency-adjustment procedures may "create" artifactual components. This is especially apparent at small signal-to-noise ratios. However, this phenomenon is less evident when Vector filtered, rather than Pz, data are considered.

Additional analyses

Some of the findings described above may be related to the particular conditions used in this study. The procedures adopted were largely arbitrary (although we did attempt to simulate veridical conditions), and variations of some of the parameters may have a crucial impact on the accuracy of latency estimates. In order to generalize these findings to a wider variety of situations, four additional analyses were performed. These analyses explored the effects of four variables on the accuracy of latency estimates: (a) standardizing background EEG noise, (b) varying the duration of P300, (c) varying the latency of P300, and (d) varying the parameters used for VF.

Effect of standardizing background EEG noise. Our simulation of background EEG noise included the standardization, across trials and separately for each time-point, of single-trial deviations. The purpose was to obtain comparable variance across the whole epoch. However, this procedure might alter the veridicality of our simulation procedure. Analyses of the frequency characteristics of the deviation waveforms before and after standardization did not reveal any particular alteration after standardization. To investigate further the effect of the standardization procedure, we ran part of the basic design of the study on non-standardized waveforms. The replication was exact, apart from the absence of frequency filtering. However, all the other manipulations were replicated. Note also that the signal-to-noise ratio for non-standardized waveforms could not be exactly determined. However, P300 amplitude was manipulated as in the basic design, and the level of noise in the P300 region was roughly comparable to that of the basic design. Thus, the same scale was adopted for the signal-to-noise ratio manipulation.

Some of the results obtained with non-standardized noise are presented in Figure 7.

Insert Figure 7 About Here

Comparison of the accuracy of latency estimation with standardized (Figure 6, upper left panel) and non-standardized background noise (Figure 7) indicates that the standardization procedure did not significantly alter the results. All the findings were replicated. Thus, the standardization procedure did not impair the veridicality of the simulation.

Effect of P300 duration. The results from the basic design indicate that cross-correlation provides more accurate latency estimates than peak-picking, and that no further advantage is obtained by iterating the Woody Filter. However, the relatively high accuracy obtained with cross-correlation may be due to the fact that the template we used for this procedure is the same as that used to simulate the P300. This may also explain why adaptive templates, such as those used in the second and third iteration of Woody Filter, do not produce any improvement. However, this situation is not normally encountered. In reality the exact P300 duration is not known. Cross-correlation may be very sensitive to small differences between the template and the waveshape of the real component. On the other hand, it might be that the template which best discriminates between the target component and other sources of brain electrical activity does not mirror exactly the duration of the target component but has some additional features which reduce its affinity with noise. Thus, we studied the accuracy of latency estimates obtained in conditions in which the template used for cross-correlation does not exactly correspond to the target component. We obtained this dissociation by varying the duration (i.e., wavelenght) of P300.

To study the effect of P300 duration, we varied systematically the duration of the simulated P300 between 200 and 800 msec, with increments of 100 msec. However, we did not vary the wavelength of the template used for the cross-correlation procedure. P300 amplitude was fixed at 250 units (corresponding to a signal-to-noise ratio of 2.5). Background EEG noise was simulated through standardized waveforms, but no overlapping components were added. No frequency filtering was applied.

MSE (and number of trials required to obtain a standard error of 3 msec) as a function of P300 duration for two spatial filtering procedures (Pz selection and VF) and four signal detection algorithms (peak-picking, cross-correlation, Woody Filter with two iterations, and Woody Filter with three iterations) are shown in figure 8.

Insert Figure 8 About Here

An inspection of this figure reveals several noteworthy findings. First, the accuracy of latency estimation depends on the duration of P300. The sharper the P300, the more accurate the estimate. This is particularly true for the peak-picking procedure (especially if used in conjunction with VF) and Woody Filter. For cross-correlation, the most accurate estimation is obtained when the duration of the P300 is slightly shorter (400 msec) than that of the template. When the duration of the component is shorter than that of the template, peak-picking and Woody Filter produce estimates equal to or more accurate than cross-correlation. However, when the duration of the component is longer than the duration of the template, cross-correlation yields better estimates. These results suggest that peak-picking and Woody Filter produce accurate estimates in cases of sharp components. For peak-picking, this is not surprising. For Woody Filter, it may be that this procedure produces sharper templates at each iteration. Thus, iterating with Woody Filter may be advantageous when the original template has a longer wavelength than the target component, but disadvantageous when the original template has a shorter wavelength than the target component. It is interesting to note that, in cases of latency jitter, the averages tend to be "smooth" and the components "widened." Therefore, using the "unadjusted" average waveform as template for the first iteration of Woody Filter may be appropriate in cases in which the duration of the target component is not known.

Effect of variability of P300 latency. This condition simulated the phenomenon of latency jitter. To this purpose, the latency of P300 was randomly varied from trial to trial, by sampling from a random distribution with a mean of 500 msec and a standard deviation of 83 msec. To assess the interaction between the effect of latency jitter and inaccuracy in the estimation of the template, the duration of P300 was varied, with three levels, 300, 500, and 700 msec. The duration of the template for cross-correlation was fixed at 500 msec.

The results of this analysis are presented in Table III.

Insert Table III About Here

These results indicate that, in conditions of latency jitter, the advantage of VF over channel selection remains, while the advantage of cross-correlation over peak-picking is dependent on the correspondence between the template and the target component. The most interesting comparison is between cross-correlation and Woody Filter, that is, the effect of iteration. Woody Filter with two iterations yields latency estimates more accurate than cross-correlation when the component is sharper than the template, while no advantage for the iteration procedure is evident when the template is equal to or sharper than the component.

Effect of variation of VF parameters. The parameters for VF used in the basic design were chosen to discriminate between the target component (P300) and various sources of noise. We presented above the rationale for our choice. However, other parameters could have been chosen, and some of them might have produced better results than those selected. To evaluate the consequences of the choice of VF parameters, we varied them systematically and studied their impact on the accuracy of latency estimation.

For this study, we used a P300 amplitude of 250 units (signal-to-noise ratio = 2.5). Background EEG noise was simulated with standardized waveforms, no overlapping components were added, and no frequency filtering was introduced. The parameters of VF (polarity and orientation angles; see footnote 6) were systematically varied by increments of 30 degrees, with three levels of polarity (0, +30, +60 degrees) and twelve levels of orientation (-180, -150, -120, -90, -60, -30, 0, +30, +60, +90, +120, and +150 degrees). The relative patterns of scalp distribution corresponding to some of these orientation angles, and the ratio between the mean and the standard deviation of the electrodes, corresponding to the polarity values, are shown in Figure 9. The effect of varying the parameters of VF was evaluated by comparing it with the results obtained with channel selection (Pz). Thus, each of the 36 parameter combinations was classified as yielding estimates "clearly superior" to Pz selection (MSE more than 30% lower), "superior" to Pz selection (MSE 0 to 30% lower), "inferior" to Pz selection (MSE 0 to 30% higher), and "clearly inferior" to Pz selection (MSE more than 30% higher). These values are shown in Figure 9.

Insert Figure 9 About Here

Figure 9 shows that a specific region (combination of VF parameters) yields the best results. This region does not correspond to the scalp distribution of P300. In fact, in this region the central electrode is more negative (or less positive) than the frontal electrode. The parameters of VF we used for the basic design (polarity of 15 and orientation of 300 degrees) are within this region, although not at the center. Thus, the parameters we chose were "good." The MSE was lower at the center of the region than at the point corresponding to the parameters we used for the basic design. Thus, the parameters we chose on the basis of a previous empirical study were not the "best" possible for the simulated data.

Note that several combinations of VF yield very low accuracy (regions where VF is clearly inferior to Pz). This is not surprising. In fact, these combinations correspond to scalp distributions which do not enhance the discrimination between signal and noise. Rather, they enhance the noise or reduce the signal.

Conclusions and Guidelines

The results obtained in this study indicate that the accuracy of latency estimation is affected by several variables, including the signal-to-noise ratio, characteristics of signal and noise, the use of preparatory (filtering) procedures, and the choice of the signal detection algorithm.

The signal-to-noise ratio appears to be the main factor. In general, the error of estimation decreases exponentially with increases in the signal-to-noise ratio. Thus, any methodology which enhances the signal-to-noise ratio is very valuable. However, the effect of the signal-to-noise ratio does not appear to interact with other effects. Procedures which yield the most accurate estimates at a high levels of signal-to-noise ratio tend to produce the most accurate estimates at low levels of signal-to-noise showed that the heavier the filtering, the more accurate the signal detection. Of course, there should be a level at which further filtering produces impairment of the signal detection because the signal itself is degraded. Such a level of filtering was not reached in our study. It should be noted that our findings relate only to moving averages, since we did not compare the effect of filters with different band-pass characteristics, or the effect of high-pass filters.

The choice of the signal detection algorithm may also affect the accuracy of our estimations. Cross-correlation produces better results than peak-picking, at least when the wavelength of the template is comparable to, or shorter than, the wavelength (duration) of the signal. The difference between the two procedures may also be reduced by the use of appropriate frequency filters. Two or three iterations of Woody Filter do not yield significant improvement over cross-correlation alone in cases in which the template for cross-correlation has a wavelength comparable to, or shorter than, that of the signal. However, the Woody Filter iterations produced a marked improvement in accuracy in those cases in which the wavelength of the template was much longer (2 times or more) than that of the signal. Thus, cross-correlation alone appears the best choice when the duration of the target component is known (at least approximately). When no information is available, cross-correlation with iterative Woody Filter should be used. The use of peak-picking should be restricted to the detection of sharp (duration equal or less than 300 msec) components, and, even in these cases, its use may be justified mainly on the basis of its simplicity and low computational load.

As a general commentary, the results of this study emphasize that the characteristics of both signal and noise must be considered for the choice of procedures for the estimation of the latency of ERP components. The interaction between signal and noise characteristics was particularly evident for the choice of spatial filtering procedures. However, we believe that this is merely an instance of a general principle, and that ERP signal detection algorithms should be based on those characteristics of the signal which allow its discrimination from the noise in which it is embedded.

Summary

We compared the accuracy of P300 latency estimation obtained with different procedures under different signal and noise conditions. The procedures included preparatory and signal detection techniques. Preparatory techniques included frequency filters and spatial filters (channel selection and Vector Filter). Signal detection techniques included peak-picking, cross-correlation, and Woody Filter. Different signal-to-noise ratios were simulated by multiplying the signal (P300) by a scaling factor. Two kinds of noise were added: event-related noise (overlapping components), and non-event-related noise (background EEG).

Accuracy in the latency estimation increased exponentially as a function of the signal-to-noise ratio. Cross-correlation provided better estimates than peak-picking. The results with Woody Filter paralleled those obtained with cross-correlation. Vector Filter provided better estimates than channel selection. The use of frequency filtering reduced the advantage of cross-correlation. Large component overlap impaired the accuracy of the estimates obtained with channel selection, but it impaired the accuracy of the estimates obtained with Vector Filter only when the overlapping component had a scalp distribution similar to that of the signal component. The effects of varying noise characteristics, P300 duration and latency, as well as the parameters of Vector Filter were also investigated.

These results indicate an advantage of using all the available information in the estimation of the latency of ERP components. Characteristics of the noise are also relevant to the choice of the appropriate procedure. Some guidelines for this choice are provided.

Acknowledgements

We wish to thank Monica Fabiani, Demetrios Karis and Greg Miller for their helpful comments on earlier versions of this paper.

References

Callaway, E., Halliday, R., Naylor, H., & Thouvenin, D. The latency of the average is not the average of the latencies. Psychophysiol., 1984, 21: 571.

Coles, M.G.H., & Gratton, G. Psychophysiology and contemporary models of human information processing. In D. Papakostopoulos, and I. Martin, (Eds.), Clinical and Experimental Neuropsychophysiology. Croom Helm Ltd., Beckenham, England, in press.

Coles, M.G.H., Gratton, G., Bashore, T., Eriksen, C.W., & Donchin, E. A psychophysiological investigation of the continuous flow model of human information processing. J. of Exp. Psychol.: Human Perc. and Perf., in press.

Coles, M.G.H., Gratton, G., Kramer, A.F., & Miller, G.A. Principles of signal acquisition and analysis. In M.G.H. Coles, E. Donchin, and S.W. Porges (Eds.), Psychophysiology: Systems, Processes, and Applications. Guilford Press, New York, in press.

Derbyshire, A.J., Driessen, G.J., & Palmer, C.W. Technical advances in the analysis of single acoustically evoked potential. Electroencephal. clin. Neurophysiol., 1967, 22: 467-481.

Donchin, E. Surprise?...Surprise! Psychophysiol., 1981, 18: 493-513.

Donchin, E., Coles, M.G.H., & Gratton, G. Cognitive psychophysiology and preparatory processes: a case study. In S. Kornblum and J. Requin (Eds.), Preparatory state and processes. LEA, Hillsdale, NJ, 1984: 155-178.

Donchin, E., & Heffley, E.F. Multivariate analysis of Event-Related Potential data: a tutorial review. In D. Otto (Ed.), Multidisciplinary perspectives in Event-Related brain Potentials. Government Printing Office, 1978: 555-572

Donchin, E., & Herning, R.I. A simulation study of the efficacy of stepwise discriminant analysis in the detection and comparison of Event-Related Potentials. Electroencephal. and Clin. Neurophysiol., 1975, 38: 51-68.

Donchin, E., Ritter, W., & McCallum, C. Cognitive psychophysiology: the endogenous components of the ERP. In E. Callaway, P. Tueting, and S.H. Koslow (Eds.) Event-Related brain Potentials in man. Academic Press, New York, 1978.

Duncan-Johnson, C.C. P300 latency: a new metric for information processing. Psychophysiol., 1981, 18: 207-215.

Duncan-Johnson, C.C., & Donchin, E. The time constant in P300 recording. Psychophysiol., 1977, 16: 53-55.

Duncan-Johnson, C.C., & Donchin, E. The P300 component of the Event-Related brain Potential as an index of information processing. Biol. Psychol., 1982, 14: 1-52.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. P300: methodological and theoretical issues. In P.K. Ackles, J.R. Jennings, and M.G.H. Coles (Eds.), Advances in psychophysiology. Vol. II. JAI Press, Greenwich, CT, in press.

Gratton, G., Coles, M.G.H., & Donchin, E. A new method for off-line removal of ocular artifact. Electroencephal. and Clin. Neurophysiol., 1983a, 55: 468-484.

Gratton, G., Coles, M.G.H., & Donchin, E. A multivariate approach to the analysis of scalp distribution of Event-Related Potentials: Vector Analysis. Submitted for publication.

Kutas, M., McCarthy, G., & Donchin, E. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 1977, 197: 792-795.

McCarthy, G., & Donchin, E. Brain potentials associated with structural and functional visual matching. Neuropsychologia, 1978, 18: 571-585.

McCarthy, G., & Donchin, E. A metric for thought: A comparison of P300 latency and reaction time. Science, 1981, 211: 77-80.

Nahvi, M.J., Woody, C.D., Ungar, R., & Sharafat, A.R. Detection of neuroelectrical signals from multiple data channels by linear filter method. Electroencephal. clin. Neurophysiol., 1975, 38: 191-198.

Palmer, C.W., Derbyshire, A.J., & Lee, A.W. A method for analyzing individual cortical responses to auditory stimuli. Electroencephal. clin. Neurophysiol., 1966, 204-206.

Pfefferbaum, A. A simulation study of latency adjustment procedures. Paper presented at III Workshop on ERP Methodology, Annual Meeting of the Society of Psychophysiological Research, Asilomar, California, 1983.

Ragot, R. Perceptual and motor space representation: An event-related potential study. Psychophysiol., 1984, 21: 159-170.

Ruchkin, D.S., & Glaser, E.M. Some simple digital filters for examination of CNV and P300 on single trial basis. In D. Otto (Ed.) Multidisciplinary perspectives in Event-Related brain Potentials. Government Printing Office, 1978: 579-581.

Squires, K.C., Wickens, C., Squires, N.K., & Donchin, E. The effect of stimulus sequence on the waveform of the cortical Event-Related Potential. Science, 1976, 193: 1142-1146.

Sutton, S., & Ruchkin, D.S. The late positive complex: Advances and new problems. In R. Karrer, J. Cohen, & P. Tueting (Eds.) Brain and information: Event-related potentials. New York Academy of Science, 1984: 1-23.

Ungar, P., & Basar, E. Comparison of Wiener filtering and selective averaging of evoked potentials. Electroencephal. clin. Neurophysiol., 1976, 40: 516-520.

Wastell, D.G. Statistical detection of individual evoked responses: an evaluation of Woody's adaptive filter. Electroencephal. clin. Neurophysiol., 1977, 42: 835-839.

Wastell, D.G. When Wiener filtering is less than optimal: an illustrative application to the brainstem evoked potential. Electroencephal. clin. Neurophysiol., 1981, 51: 678-682.

Wiener, N. Extrapolation, interpolation and smoothing of stationary time series. John Wiley, New York, 1949.

Woody, C.D. Characterization of an adaptive filter for the analysis of variable latency neuroelectrical signals. Medical and Biol. Eng., 1967, 5: 539-553.

Footnotes

1. The study presented in this paper was supported in part by a grant from the Air Force Office of Scientific Research, contract #F49620-83-0144, with Al Fregly as technical monitor. A partial report of the data was presented at the 24th Annual Meeting of the Society for Psychophysiological Research, Milwaukee, Wisconsin, October 18-21, 1984.

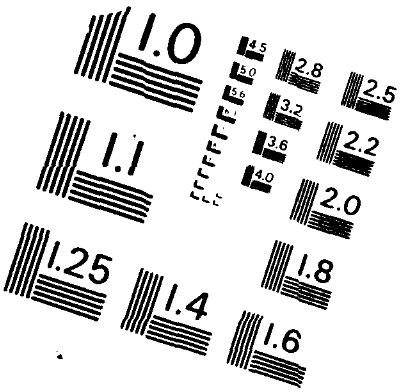
2. In this case, an unbiased estimate of component amplitude can be obtained by aligning each single trial on the component peak, and then computing the average. However, this implies the knowledge of the latency of the component peak at each single trial.

3. Recent investigations, both on simulated (Callaway, Halliday, Naylor, & Thouvenin, 1984) and real data (Fabiani, Gratton, Karis, and Donchin, in press) point to the greater validity and reliability of estimates of component latency obtained by averaging single-trial estimates rather than by estimating the latency directly from average waveforms. Furthermore, since this procedure yields estimates of the latency of an ERP component at each single trial, more information is available. For instance, the distribution of single-trial latencies can be computed, and comparisons with other measures of mental chronometry (e.g. reaction time) can be performed on each trial.

4. In this experiment, the subject was presented with one of two tones on any given trial. The tone probabilities were .2 and .8. ERPs were recorded at Fz, Cz, and Pz. The on-line filtering procedure included a low-pass filter with a half amplitude cut-off point at 35 Hz, and a high-pass filter with a time constant of 8 sec. Vertical EOG was recorded from above and below the right eye, and ocular artifacts were corrected with a procedure described in Gratton, Coles, and Donchin (1983a). Separate

AD-A170 794 THE EVENT-RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING A..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 7/8

UNCLASSIFIED 20 FEB 86 CPL-86-1 F49620-85-C-0041 F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS - 1963 - A

averages were obtained for frequent and rare trials and for each electrode.

5. Moving average filters cannot be perfectly described in terms of "half amplitude cut-off frequency", because their frequency function is quite complex, as shown by Ruchkin and Glaser (1978).

6. The orientation in the space defined by the electrode locations can be described by a series of angles. These angles reflect the weight given to each electrode. However, this method of describing a scalp distribution is impractical because (a) one of these angles is redundant, and (b) recovery of information about the overall polarity of the distribution is difficult. Hence, Gratton et al. (submitted for publication) proposed a different way of describing a scalp distribution. This description is based on the separation of the information related to the common trend across electrode sites (polarity) and to the relative patterning of the electrodes (orientation). In the three-electrode case this approach allows us to describe a scalp distribution (and therefore a Vector Filter) by means of two angles:

a. A "polarity" angle, reflecting the relative value of the mean of the electrodes in comparison with their standard deviation. A negative value of the polarity angle indicates an overall negative scalp distribution, a positive value, an overall positive scalp distribution. The 0 value expresses the condition in which the mean of the electrodes is equal to 0.

b. An "orientation" angle describing the relative distribution of the potentials over the electrodes. In the three-electrode case, the deviations of each electrode from their mean can be plotted on a plane as three non-orthogonal axes (the angles between the axes will be equal to 120 degrees). This plane corresponds to a description of the variance across electrodes (Gratton et al, submitted for publication). Any pattern of deviations can be described by an axis in this plane. We label "orientation" the angle between this axis and an arbitrary reference axis. Thus, any pattern of relative distribution will be associated with an orientation angle. A reference axis which corresponded to the relative distribution of electrode values was used for our simulated P300 (maximum at Pz, minimum at Fz, with the Cz exactly a half distance between the two). Any other relative scalp distribution could be described by an "orientation" angle with this axis. A procedure to obtain orientation values for any scalp distribution is described in Gratton et al. (submitted for publication).

7. The logarithmic scale was chosen because we assumed the variability in the MSE obtained with different procedures to be proportional to its absolute value. With a logarithmic scale, similar percentage differences at different absolute levels of MSE are represented equally.

Table I

Basic Design

(A) Trials per condition (100)

(B) Signal detection algorithms (4)

peak-picking
cross-correlation
Woody with two iterations
Woody with three iterations

(C) Scalp distribution filtering techniques (2)

channel selection (Pz)
Vector Filtering

(D) Target component amplitude levels (11)

signal-to-noise ratio varying from 0 to 5 by .5 increments

(E) Amplitude levels of overlapping components (5)

no overlap
N200: 50, 100
SW: 100, 200

(F) Latency levels of overlapping components (4)

N200: 300, 400
SW: 800,1080

(G) Frequency filter conditions (5)

no filter
6.29 Hz, 1 iteration
6.29 Hz, 2 iterations
3.14 Hz, 1 iteration
3.14 Hz, 2 iterations

(H) Dependent Variables (2)

MSE
number of trials required to obtain 3 msec error

Table II

Parameters of ERP Components

Components	Amplitude (units)	Latency (msec)	Duration (msec)	Scalp Distribution Weights Fz	Cz	Pz
N100	100	100	100	-.6	-1.8	-.6
P200	150	250	150	1.0	1.0	.5
N200	50/100	300/400	250	-1.2	-.8	-.4
SW	100/200	800/1080	800	-.4	.0	.4
P300	0 to 500	550	500	.4	.8	1.2

Table III

MSE in Case of Latency Jitter

Measures	P300 duration 300 msec	500 msec	700 msec
Peak-Picking			
Pz	40	55	75
Vector Filter	24	49	68
Cross-correlation			
Pz	61	47	75
Vector Filter	26	35	67
Woody Filter (2 iterations)			
Pz	56	41	74
Vector Filter	20	37	66
Woody Filter (3 iterations)			
Pz	49	42	74
Vector Filter	20	43	67

Components' parameters:

P300 amplitude = 250 units (signal-to-noise ratio = 2.5)

N100 amplitude = 100 units	N100 latency = 100 msec
P200 amplitude = 150 units	P200 latency = 250 msec
N200 amplitude = 50 units	N200 latency = 300 msec
SW amplitude = 100 units	SW latency = 800 msec

Figure Legends

Figure 1. Examples of simulated single-trial waveforms. Waveforms without EEG noise are shown at the top, waveforms with EEG noise are displayed at the bottom. The amplitude is expressed in arbitrary units (simulated P300 amplitude = 250 units).

Figure 2. Orientation angles corresponding to the scalp distributions of the components used in the study. The orientation corresponding to the Vector Filter is also shown (300 degrees).

Figure 3. Log MSE and number of trials required to obtain a Standard Error of 3 msec, as a function of signal-to-noise ratio, for different frequency filters, signal detection algorithms, and spatial filtering conditions.

Figure 4. Histograms of latency estimates for four different signal-to-noise ratios from two detection algorithms and two spatial filtering techniques. The vertical lines indicate the latency of the simulated P300.

Figure 5. Latency-adjusted average waveforms over 100 trials. P300 peak latency was computed with cross-correlation (upper panels), Woody 2-iterations (middle panels), Woody 3-iterations (lower panels). The left column refers to waveforms obtained with a signal-to-noise ratio of 0 (no P300 was present), the right column refers to waveforms obtained with a signal-to-noise ratio of 5. The solid lines indicate Pz waveforms, the dashed lines indicate Vector filtered waveforms. The vertical lines indicate the latency of the simulated P300. The amplitude is expressed in

arbitrary units.

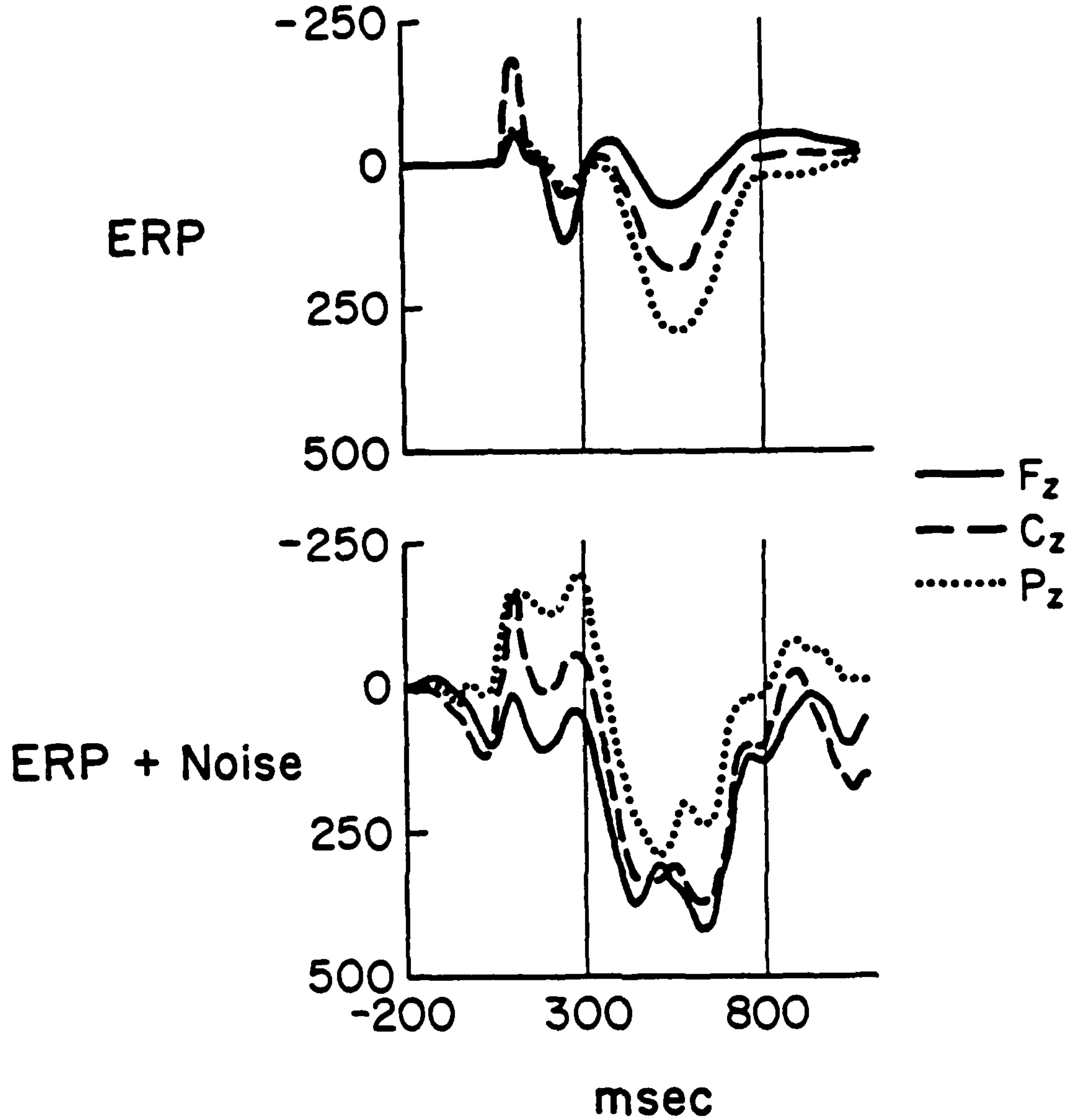
Figure 6. Log MSE and number of trials required to obtain a Standard Error of 3 msec, as a function of signal-to-noise ratio for four different component overlap conditions. Two detection algorithms and two spatial filtering techniques are shown for each condition.

Figure 7. Log MSE and number of trials required to obtain a standard error of 3 msec, as a function of the amplitude of P300 for the non-standardized noise condition. Two detection algorithms and two spatial filtering techniques are shown.

Figure 8. Effect of P300 duration on the accuracy of latency estimation with peak-picking, cross-correlation and Woody Filter with two and three iterations.

Figure 9. Effect of the manipulation of Vector Filter parameters on the accuracy of latency estimation as compared to the use of a single electrode (Pz).

Fig. 1



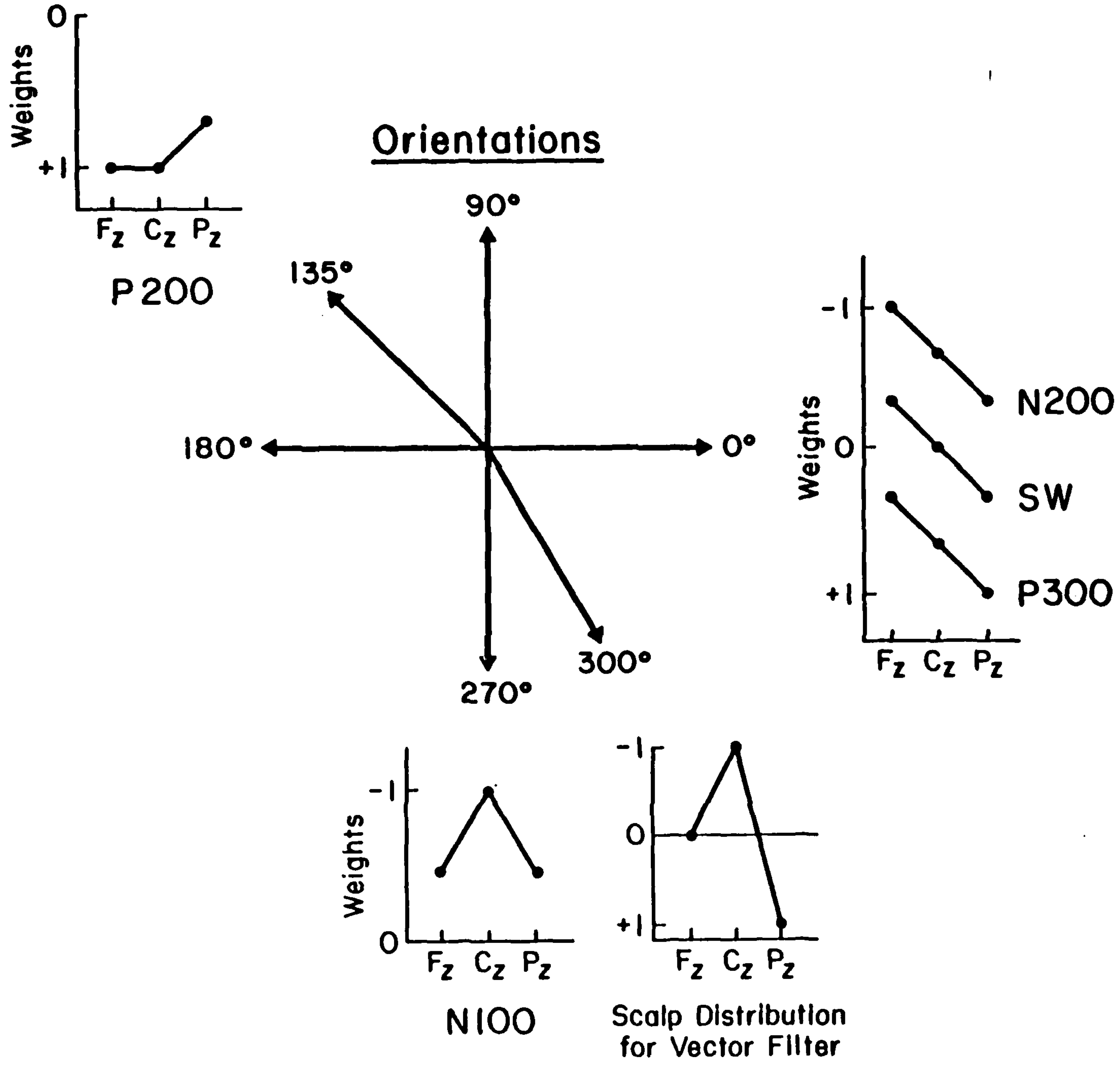

Orientations
90°
135°
180°
0°
270°
300°
Weights
0
+1
Fz Cz Pz
P200
-1
0
+1
N200
SW
P300
N100
Scalp Distribution
for Vector Filter

Fig. 2

Fig. 3

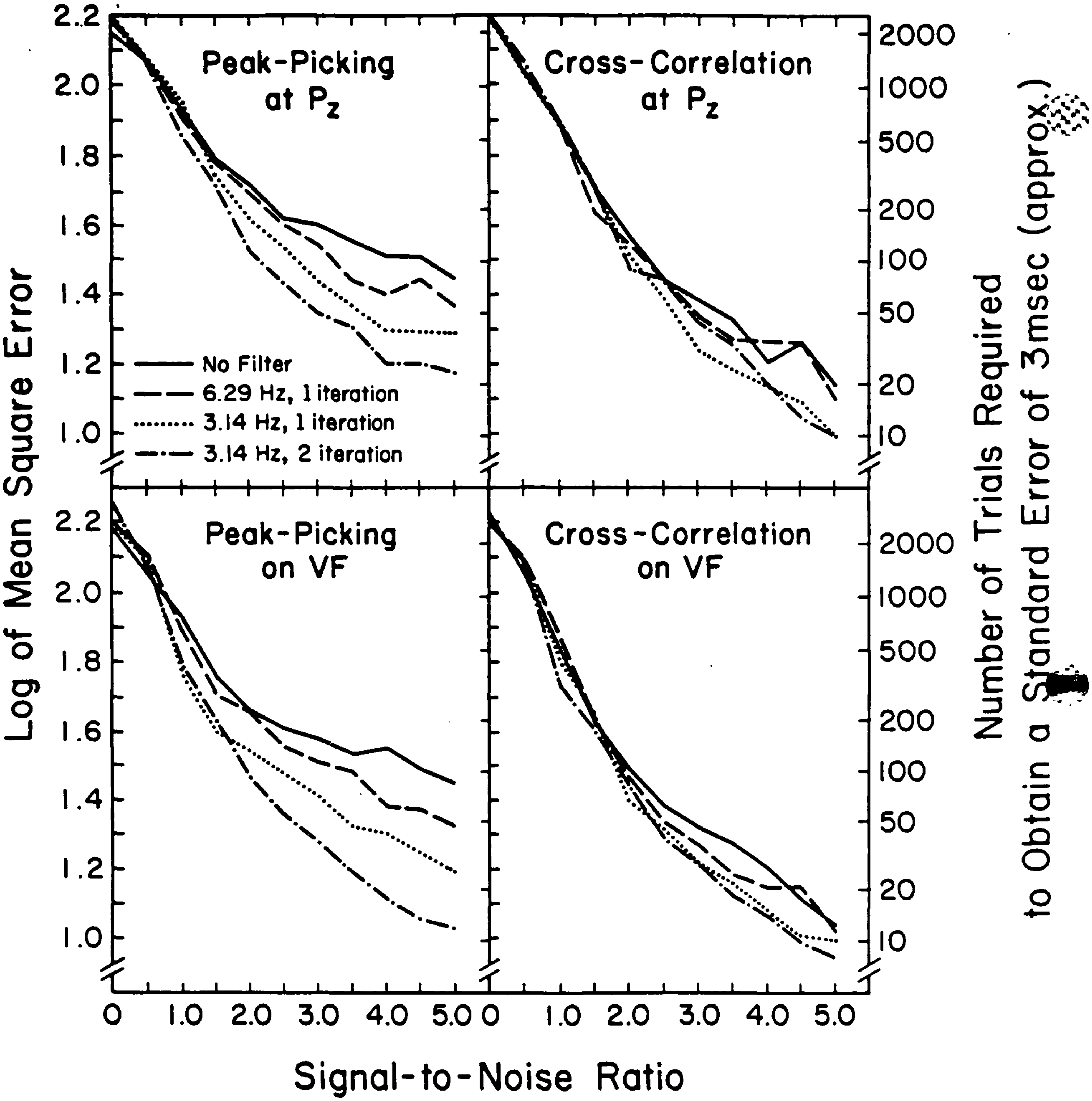
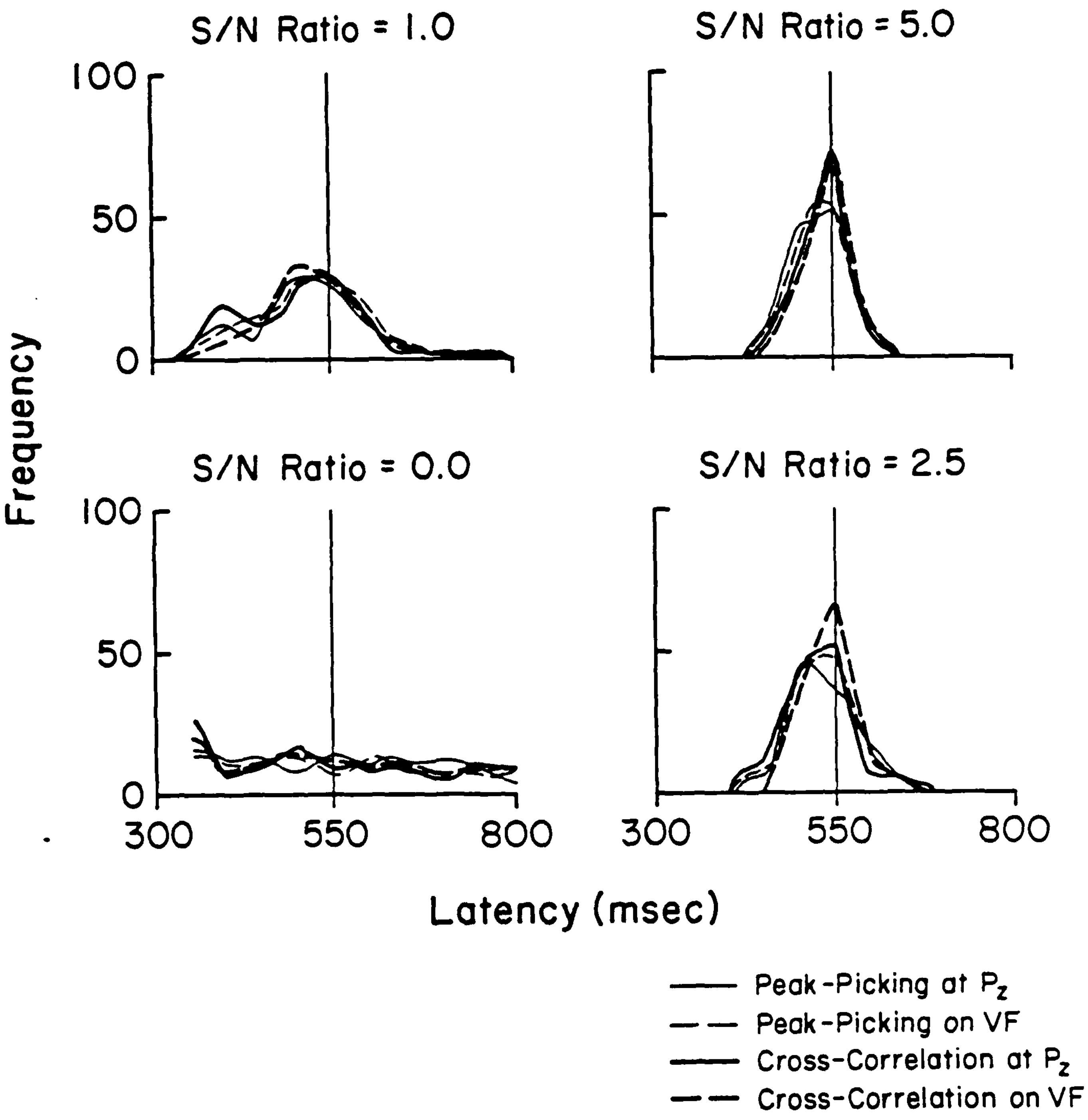
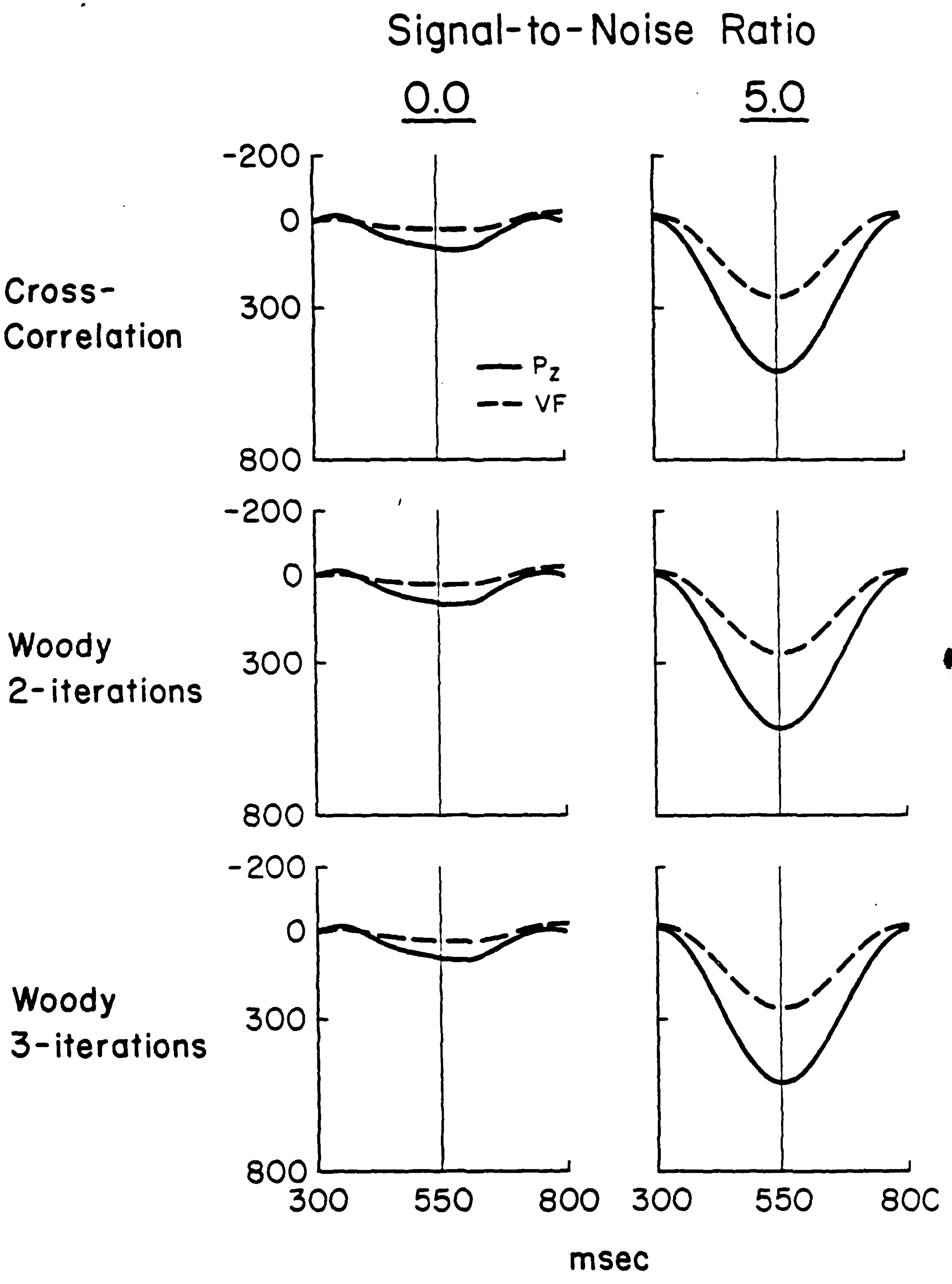

Peak-Picking at P_z
Cross-Correlation at P_z
Peak-Picking on VF
Cross-Correlation on VF
No Filter
6.29 Hz, 1 iteration
3.14 Hz, 1 iteration
3.14 Hz, 2 iteration
Log of Mean Square Error
Number of Trials Required to Obtain a Standard Error of 3msec (approx.)
Signal-to-Noise Ratio
2.2
2.0
1.8
1.6
1.4
1.2
1.0
2000
1000
500
200
100
50
20
10
0
1.0
2.0
3.0
4.0
5.0

Fig. 4



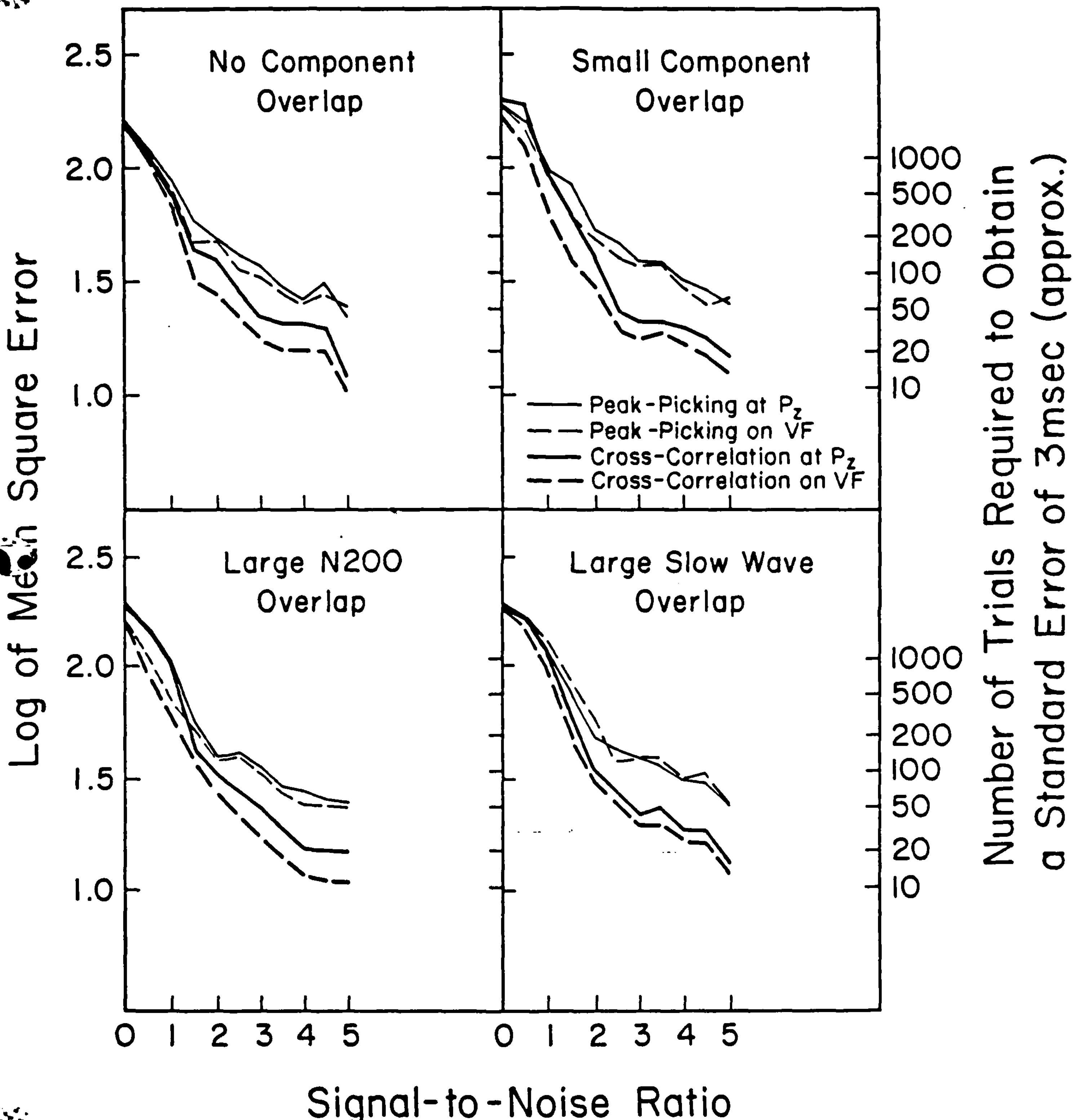
S/N Ratio = 1.0
S/N Ratio = 5.0
S/N Ratio = 0.0
S/N Ratio = 2.5
Frequency
100
50
0
100
50
0
300
550
800
300
550
800
Latency (msec)
Peak-Picking at P_z
Peak-Picking on VF
Cross-Correlation at P_z
Cross-Correlation on VF

Fig. 5



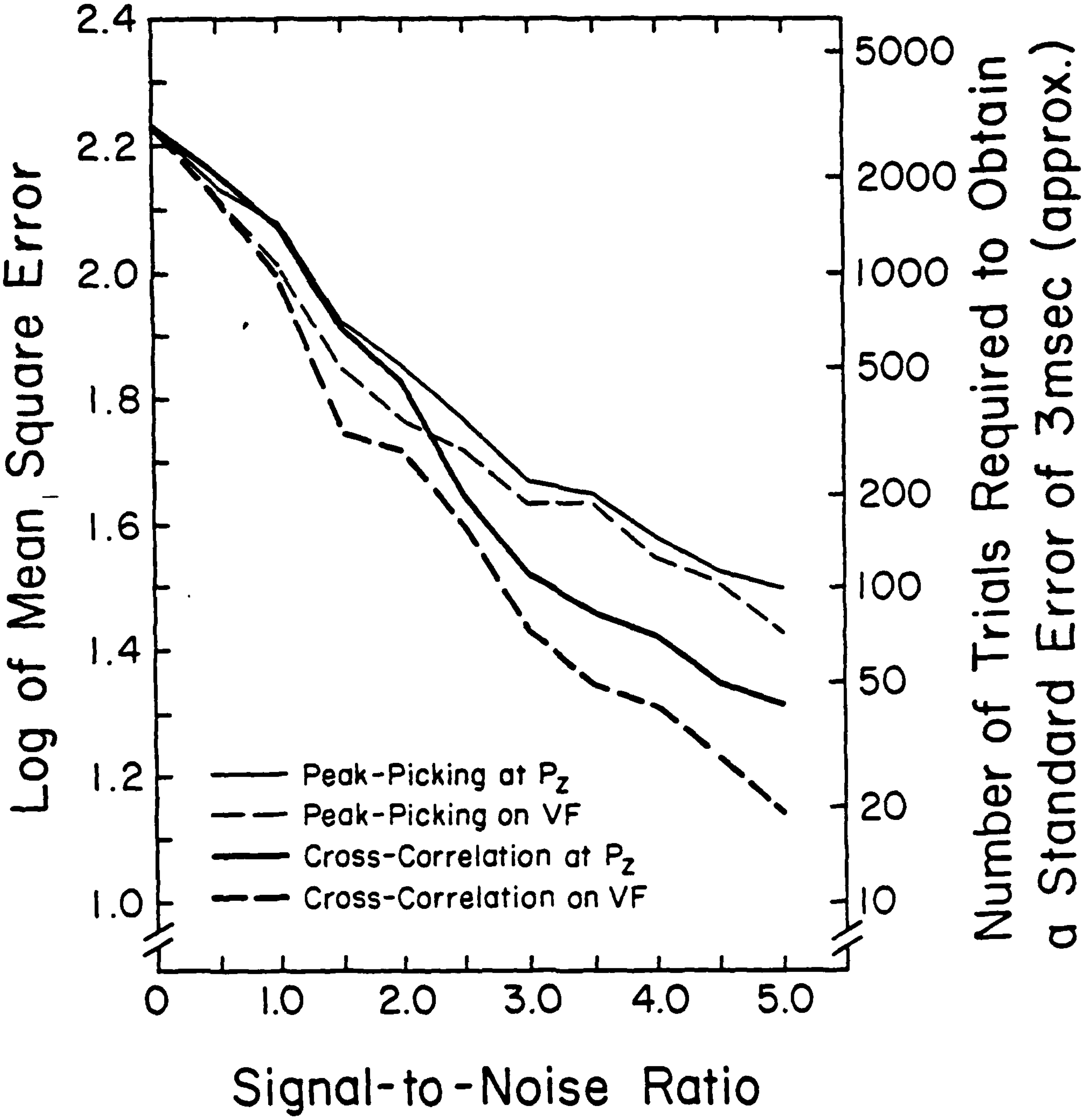
Signal-to-Noise Ratio
0.0
5.0
-200
0
300
800
Cross-Correlation
Woody 2-iterations
Woody 3-iterations
P_z
VF
300
550
800
msec

Fig. 6



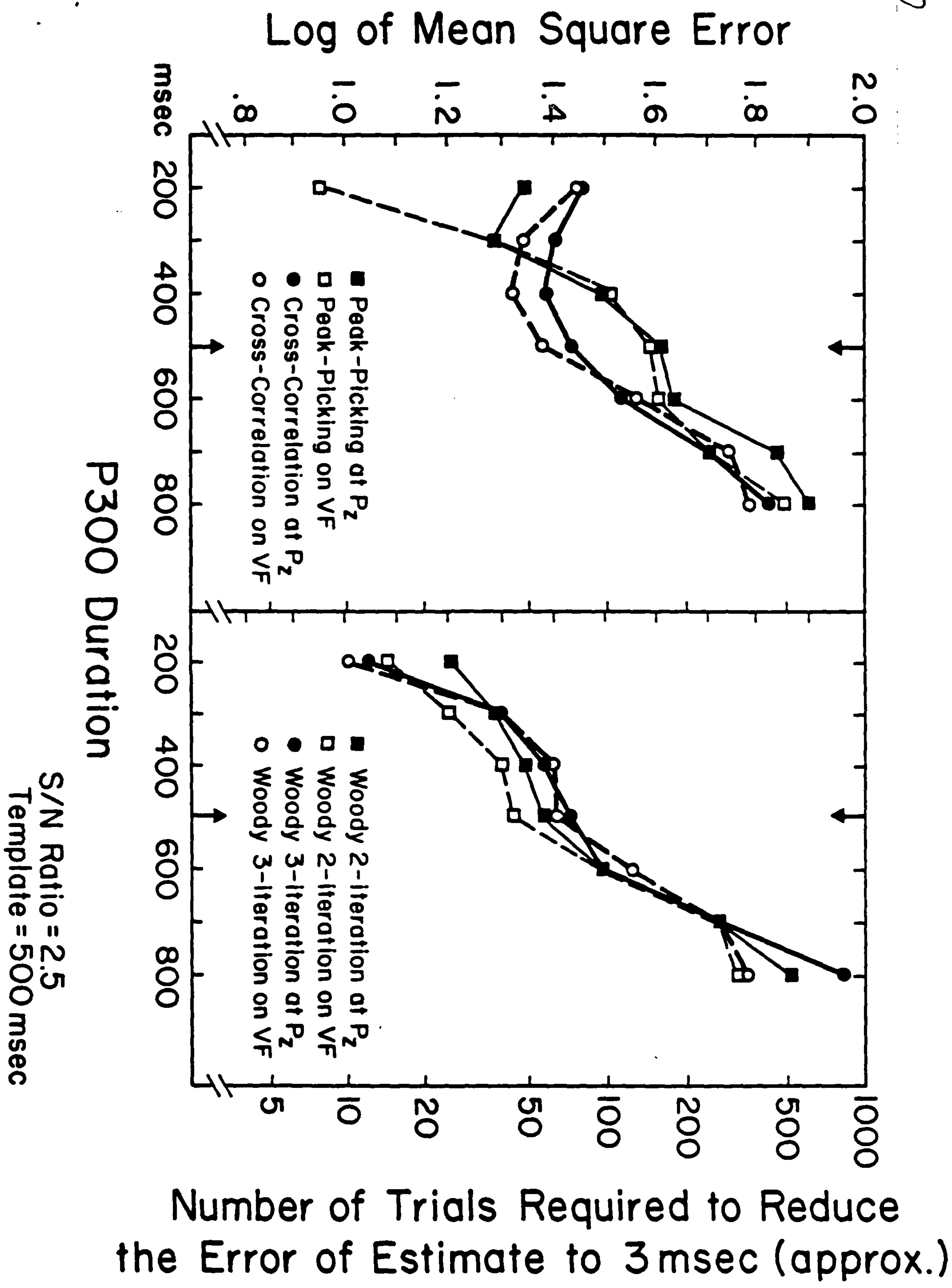
No Component Overlap
Small Component Overlap
Large N200 Overlap
Large Slow Wave Overlap
Log of Mean Square Error
2.5
2.0
1.5
1.0
Number of Trials Required to Obtain a Standard Error of 3msec (approx.)
1000
500
200
100
50
20
10
Peak-Picking at P_z
Peak-Picking on VF
Cross-Correlation at P_z
Cross-Correlation on VF
0 1 2 3 4 5
Signal-to-Noise Ratio

Fig. 7



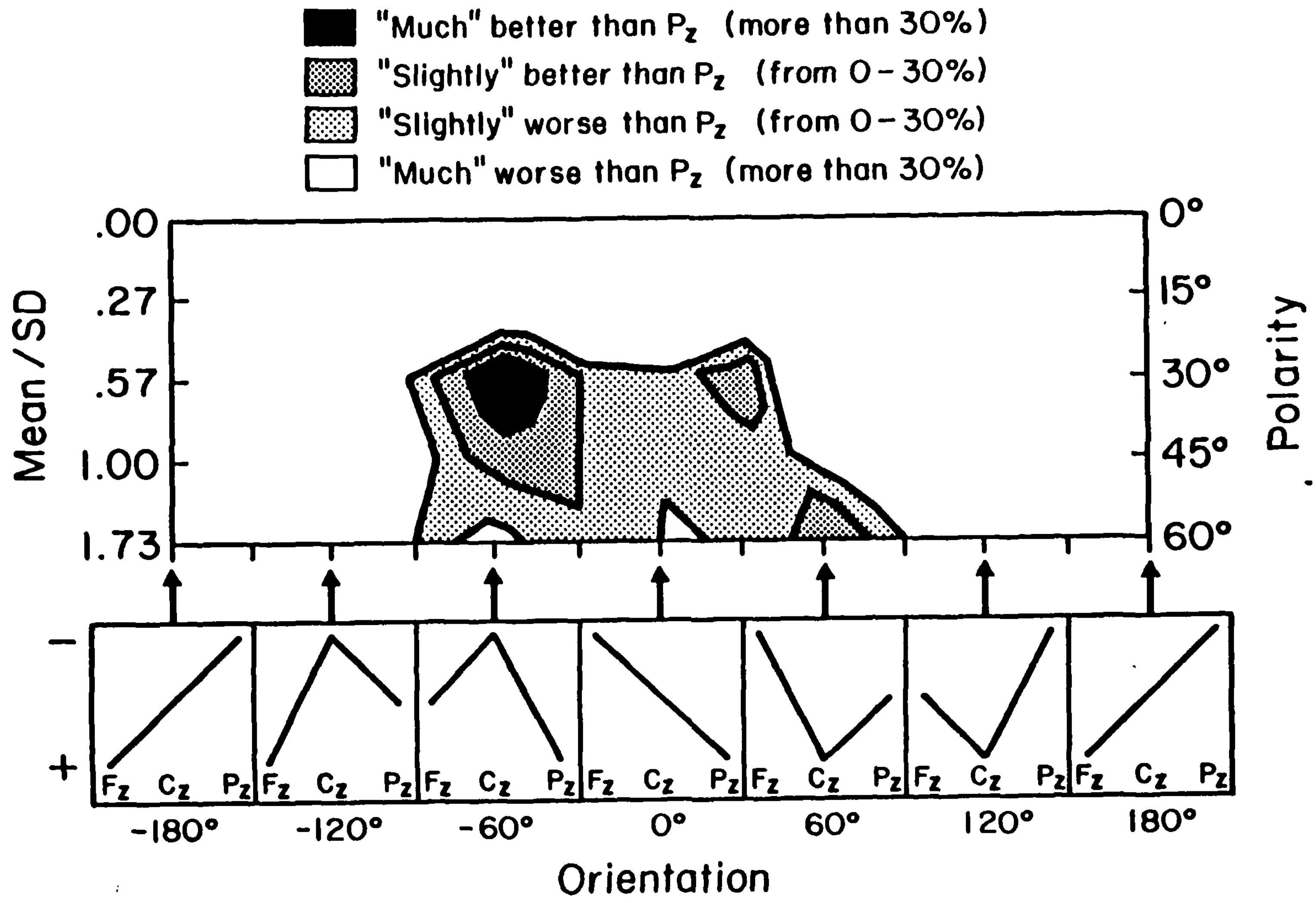
Log of Mean Square Error
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
Number of Trials Required to Obtain a Standard Error of 3msec (approx.)
5000
2000
1000
500
200
100
50
20
10
Peak-Picking at P_z
Peak-Picking on VF
Cross-Correlation at P_z
Cross-Correlation on VF
0
1.0
2.0
3.0
4.0
5.0
Signal-to-Noise Ratio

Fig. 8



Log of Mean Square Error
2.0
1.8
1.6
1.4
1.2
1.0
.8
msec
200
400
600
800
Peak-Picking at Pz
Peak-Picking on VF
Cross-Correlation at Pz
Cross-Correlation on VF
Woody 2-Iteration at Pz
Woody 2-Iteration on VF
Woody 3-Iteration at Pz
Woody 3-Iteration on VF
5
10
20
50
100
200
500
1000
P300 Duration
S/N Ratio = 2.5
Template = 500 msec
Number of Trials Required to Reduce the Error of Estimate to 3 msec (approx.)

Fig. 9


"Much" better than Pz (more than 30%)
"Slightly" better than Pz (from 0-30%)
"Slightly" worse than Pz (from 0-30%)
"Much" worse than Pz (more than 30%)
Mean / SD
.00
.27
.57
1.00
1.73
Polarity
0°
15°
30°
45°
60°
-
+
Fz Cz Pz
-180°
-120°
-60°
0°
60°
120°
180°
Orientation

Appendix B12

PEARL II: Portable Laboratory Computer System for Psychophysiological Assessment Using Event Related Brain Potentials (Notes 1,2,3,4)

Earle Heffley, Brian Foote, Tony Mui, and Emanuel Donchin

Cognitive Psychophysiology Laboratory
Department of Psychology
University of Illinois
603 East Daniel Street
Champaign, Illinois 61820

Published in Neurobehavioral Toxicology and Teratology, 1985, 7, 409-414.

Footnotes

1. Development of PEARL has been supported by the Environmental Protection Agency under agreements EPA-R-605628 and EPA-CR-808974 managed by Dr. David Otto, and by the Air Force Aerospace Medical Research Laboratory through USAF Subcontract SRL Order #25390 and Order #28727 under a program managed by Colonel Robert O'Donnell. Versions of the LABPAK software were developed at the CPL in connection with projects supported by the Air Force Office of Scientific Research through contract F49620-79-C-0233 monitored by Dr. Alfred Fregly, and by the Defense Advanced Research Projects Agency through contract N000-14-76-C-0002 monitored by Dr. Craig Fields.

2. The following members of the CPL Technical Staff also participated in the development of PEARL: Ron Klohr, Wally Meyers, Mike Anderson, and Sara Klohr. Dr. Michael Faiman, Department of Computer Science, University of Illinois, provided major direction in the initial development of PEARL. We would also like to thank the many PEARL users who have made valuable contributions to refinements in the ERP test battery. Dr. Theodore Bashore, Medical College of Pennsylvania, and Dr. Sean O'Connor, University of Connecticut Health Center, have been especially helpful.

3. The versatility of the PEARL system is illustrated by its adaptable acronym. PEARL originally stood for Portable Evoked Average Response Laboratory. EPA knows it as the Portable Environmental Assessment Research Laboratory, while the Air Force calls it the Portable Engineering Assessment Research Laboratory. The CPL, having moved from evoked potentials to event-related potentials, refers to it simply as PEARL.

4. The PEARL Development Project remains an active program. The CPL can, under certain circumstances, produce PEARL systems at cost for interested scientists. Organizations interested in procuring PEARL systems should contact Earle Heffley.

5. LSI-11, Q-Bus, and RT11 are registered trademarks of Digital Equipment Corporation.

6. Organizations that have procured PEARL systems include: the Environmental Protection Agency, the Air Force Aerospace Medical Research Laboratory, the Air Force School of Aerospace Medicine, the University of Connecticut Health Center, Rush-Presbyterian Medical Center, the Medical College of Pennsylvania, the National Institutes of Health, the Technion in Israel, the University of Illinois Aviation Research Laboratory, and the University of Illinois Psychology Clinic. The CPL currently operates ten PEARL systems.

Abstract

Heffley, E., B. Foote, T. Mui, and E. Donchin. PEARL II portable laboratory computer system for psychophysiological assessment using event related brain potentials , NEUROBEHAV. TOXICOL. TERATOL. The PEARL II portable laboratory computer integrates hardware and software to serve as an on-line, real-time, experimental control and data acquisition system. Although the system can be used in many areas of research, PEARL II development has emphasized investigation of physiological responses from human subjects performing complex experimental tasks. PEARL functions as a "turn-key" system which performs standard neurological tests that would be employed, for example, in neurotoxicological assessment. The PEARL system also includes several psychophysiological tests used in human engineering research on performance workload assessment. PEARL can also serve as a tool in basic research on human psychophysiology. The special feature of the PEARL test battery is its suitability for the measurements of event-related brain potentials (ERPs) in these tasks, although other physiological indices such as heart rate may be monitored. In addition, the PEARL system includes a versatile library of laboratory control subroutines that can be used to develop new applications.

Key Words: evoked potentials, event-related potentials, neurotoxicology, BAEP, PREP, SEP, CNV, P300, human field testing, and laboratory computer.

Running Head: Portable computer system for psychophysiological assessment

Background

The PEARL system emerged within the context of the research program conducted at the Cognitive Psychophysiological Laboratory (CPL) at the University of Illinois. The CPL has been engaged since 1969 in research with primary focus on development of the theoretical and empirical basis for the use of Event Related Brain Potentials (ERP) as a tool in the study of cognitive function [1,2]. Laboratory work conducted at the CPL and elsewhere has indicated that ERPs could be used in a wide variety of assessment tasks. In particular, ERPs appeared useful in the measurement of mental workload [3], in mental chronometry [4], in the study of preparatory processes [5], in research on selective attention [6], and in assessment of neurotoxicity [7,8].

While many significant results emerged from ERP laboratories, it was difficult to evaluate the degree to which the procedures could be applied in the working environment of the clinician, the human factor specialist, and the toxicologist. Until a few years ago, the equipment available for research in these environments was costly and bulky. Moreover, users were forced to choose between systems that were preprogrammed for a specific purpose and systems that were quite general yet difficult to program. The problem became particularly acute for the CPL in two of its research programs. The Air Force was interested in evaluating the ERP as a work-load metric within the context of actual, or simulated, aviation. The Environmental Protection Agency was interested in conducting assessments of cognitive and neural function in field sites outside the laboratory. The success of both projects was contingent on the availability of a portable, flexible, replica of the rather sophisticated laboratory developed at the CPL [9]. The PEARL system was developed in response to this need.

In 1977, we began development of the PEARL I laboratory computer system, which was designed to support ERP research both in the laboratory and at field testing sites. The PEARL project became a major hardware and software development effort because the commercial marketplace did not offer a laboratory system that could satisfy the needs of a broad basic research program in psychophysiology. Existing systems could not satisfy two important requirements: First, they did not provide for the execution of a large repertoire of experimental paradigms. In general, commercial systems of the time allowed the user control over relatively few experimental parameters within a very small number of stimulus/response contingencies. Second, commercial equipment lacked adequate capability for the acquisition and retention of massive amounts of data. Most research applications demand that digital records of all raw data be stored for future analysis, which allows for removal of physiological artifacts and for thorough examination of the effects of experimental variables. Thus, PEARL was designed as a programmable, general purpose, laboratory computer system with the power to perform properly managed storage of relatively large volumes of physiological data concurrently with control of complex experimental paradigms.

The development of PEARL I was supported by the Health Effects Research Laboratory (HERL), Environmental Protection Agency, which was interested in using ERPs in field testing situations to assess the consequences of exposure to toxic substances. Scientists from the CPL and from HERL developed specifications for a basic hardware package which was designed to prove the feasibility of developing a portable laboratory computer system for ERP research. The technical staff of the CPL and members of the University of Illinois Computer Science Department collaborated in the design of PEARL I, which was successfully built and tested. One design criterion emphasized in PEARL I was portability and the system could indeed be stored in a large suitcase which could be carried one person. However, this design limited the scope of the system and only relatively simple programs were developed for recording brainstem auditory evoked responses to clicks and for recording P300 potentials to infrequent tones which deviate from regularly presented standard tones (auditory oddball). The success of the initial PEARL project led to a cooperative agreement between the CPL and the EPA to develop an advanced version of the system (PEARL II).

The Air Force Aerospace Medical Research Laboratory (AFAMRL) also became interested in the PEARL project as a means to bring ERPs into their research program on human performance in complex man-machine systems, particularly in aircraft simulators. As noted above, the CPL had been investigating for several years applications of the endogenous ERPs in engineering psychology, and therefore collaboration with AFAMRL provided an excellent opportunity to expand these research efforts into more complex man-machine environments. With the added support of the Air Force, the PEARL project was enlarged to include development of production-quality laboratory interface hardware and to include a major software engineering effort designed to yield an extensive battery of psychophysiological tests. The result of this effort is the PEARL II system for research on ERPs in neurology, neurotoxicology, human engineering, psychopathology, and basic cognitive psychophysiology.

PEARL Project Goals

A set of eight project goals were established based upon discussions with collaborators in the PEARL project and upon consideration of our experience with larger minicomputers in ERP laboratories.

Goal 1: Support for Common ERP Tests. Scientists associated with the PEARL project were interested in a variety of experimental tests, ranging from sensory evoked responses generated by simple stimuli to endogenous event-related potentials recorded in complex manual tracking and monitoring tasks. Because research applications for a given test vary considerably, we sought to offer many variations of each task through parameters that could be easy changed by the user. The need for flexibility in the test battery motivated selection and development of hardware capable of supporting a wide range of data collection rates and processing requirements.

Goal 2: Portability. All research organizations involved in the development of PEARL were interested in a portable system. We sought to design a system that could be packaged in a small number of packing cases, which could be transported in a van or by a commercial air carrier.

Goal 3: Suitability for computer-naive users. Given the variety of tests in the battery and the variations of each particular test, design of a user-friendly system was important. We decided upon a menu-driven battery with a scheme for parameter selection that included levels of protection, legal range specification, and on-line help.

Goal 4: Support for basic research. Basic research in ERPs frequently entails extensive reprocessing of individual data records stored for each stimulus presentation. Many commercially available ERP computer systems do not offer this capability. In order to support off-line data analysis, a database management approach was specified that would allow storage and retrieval of single trial ERP records along with meaningful identification information and other dependent variables, such as subject responses.

Goal 5: System completeness. Most investigators associated with the PEARL project wanted a complete package that would include hardware and software necessary to control experiments, acquire and analyze data and generate numerical and graphic reports. It was vital that scientists be able to execute their experimental plans without having to engage in extensive software or hardware development after delivery of the system. The frustration of watching months, sometime years, pass between the delivery of equipment and the initiation of serious research is painfully common, particularly given the difficulty finding and retaining staff qualified to develop complex real-time programs. Thus, PEARL has been designed as a system that enables an investigator to begin research with basic ERP tests the moment the system is delivered.

Goal 6: Design which facilitates test development. In the hardware domain, specification of laboratory interface devices was guided by the particular needs of ERP research. The experience gained in more than a decade of research at the CPL was crucial in this case, with the result based on extensive practical experience. Our approach was to maintain interaction between scientists, engineers, and programmers during all phases of the project.

Minimizing processor overhead and maintaining precise timing of events during acquisition of a relatively large volume of data were crucial objectives in the design of the system. Many commercially available packages are unsuitable because they are not designed for the relatively unique requirements of the ERP research laboratory. Therefore, we sought to develop a software subroutine library (LABPAK) that would take full advantage of the hardware capabilities in the most efficient manner possible. Thus, development of a specialized, but flexible, programming foundation was a primary goal.

Goal 7: Standardized hardware and software. The core of the PEARL system was to be based upon standardized hardware and software development tools procured from a commercial vendor. This choice was made for two reasons. First, it was necessary to select components that would continue to be supported by the vendor for many years, given the common mortality of product lines in the computer business. For this reason, the Digital Equipment Corporation (DEC) LSI-11 processor and RT11 Operating System were selected as the foundation for the PEARL system (Note 5). Second, if users are to be able to develop their own test battery items, then a standard language processor is a requirement. RT11/FORTRAN was specified as the development system because the compiler generates efficient, reliable code and because most scientists are familiar with FORTRAN. Therefore, development of new testing paradigms was to be supported by a standardized software development environment (RT11/FORTRAN) in conjunction with the LABPAK library of FORTRAN-callable subroutines for flexible control of the specialized laboratory interface devices.

Goal 8: Modularity and Adaptability. Our approach was to develop a system that could grow to meet expanding needs in psychophysiological research by taking advantage of new computer hardware products. We also sought to configure the system so that additional interface modules could be added to increase data collection capabilities. Although PEARL was developed specifically for ERP research, the system is based upon a common microcomputer and it includes general-purpose laboratory interfaces and software.

Organization of PEARL Development Program

The PEARL Development Program is an ongoing research project at the CPL, which has a technical staff dedicated to the project. The initial goals of the project have been satisfied and approximately twenty PEARL II systems are currently in operation. Through interactions with scientists at the several organizations utilizing PEARL systems, new ideas have been generated for ERP applications and for test battery development. In addition, the hardware configuration has evolved to take advantage of new products from the computer industry. Thus, the program continues to be a vital research and development effort.

Organizations that procure PEARL systems agree to become "user-testers" and to furnish feedback to the CPL on the operation of the system. This arrangement is designed to enhance the scientific value of the PEARL software available to participating institutions. Over the past eight years, the CPL has built more than twenty PEARL systems, including many that have been delivered to other laboratories (Note 6).

PEARL Hardware description

PEARL II is a computer system based on a Digital Equipment LSI-11 microprocessor. The hardware configuration includes digital input/output, programmable clocks, an analog-to-digital converter, digital-to-analog converters, a digital magnetic cartridge tape drive, and removable hard disk drives. The processor is connected to memory and peripherals via the standard DEC Q-Bus. Special purpose laboratory peripheral devices, packaging, software, and documentation have all been developed at the CPL. A special feature of the system is the modular front panel which features a flexible scheme for making connections and selecting signal paths (see Figure 1, top). A block diagram of the PEARL II system appears in the lower portion of Figure 1.

The current PEARL II system includes an LSI-11/73 processor, 2 megabytes (MB) of memory, six serial I/O ports, six programmable clocks, parallel input/output (I/O), a 16-channel analog-to-digital (A/D) converter, two dual-channel digital-to-analog (D/A) converters, two removable 10-MB Winchester disk drives, a 1/4-inch cartridge tape drive for data storage, power supplies, packaging, and software. The original PEARL II system utilized an LSI-11/23 processor, 256 kilobytes (KB) of memory, and dual DECtape-II units in place of the Winchester drives.

System Core. The PEARL II system is based on an LSI-11 processor and Q-Bus from Digital Equipment Corporation. The LSI-11 processor was selected because an extensive array of peripheral devices exists, a well-developed real-time operating system is available, and many scientists and programmers are familiar with DEC's PDP-11 series of computers. Another consideration was DEC's apparent commitment to support and expand this product line. This strategy has enable us to take advantage of many new products in the Q-Bus family. For example, the central processor used in PEARL has been upgraded over the years as new versions have become available (from the LSI-11/2, to the LSI-11/23, and recently to the LSI-11/73). Several laboratory interface modules, custom-designed by the CPL, were added to the core system. These modules are described in the following paragraphs.

Programmable Clocks. The PEARL clock module contains six individually programmable clocks capable of supporting the timing of complex testing paradigms. The availability of six hardware clocks permits timing of multiple intervals with a minimum of system overhead.

Analog-to-Digital Converter. The PEARL II A/D system consists of a 16-channel analog multiplexer, a 12-bit A/D converter, and a direct memory access Q-Bus interface. The system samples up to 16 independent channels of electrophysiological signals at rates up to 90K samples per second. The number of channels, total number of points, and sampling rate are all selected under program control. The modular construction of PEARL allows for the inclusion of additional A/D subsystems, which increase the number of channels in groups of 16.

Digital-to-Analog Converters. The D/A system consists of four independent channels which are grouped into two pairs. In most applications, one pair will be devoted to driving an on-line display of waveforms on an oscilloscope for inspection by the system operator. The second pair of D/A channels is available for a variety of purposes, including presentation of auditory stimuli through headphones, generation of simple visual stimuli on a CRT, or driving other external devices that require analog input.

Maxi-Cartridge System. The maxi-cartridge system is one of PEARL's most important features because it allows a large quantity of data to be rapidly stored on a very compact magnetic cartridge tape. The tape system utilizes a 1/4-inch cartridge tape drive and formatter from Digi-Data Corporation. The cartridge system provides laboratory functions similar to industry standard digital tape drives on larger systems. Digitized data are rapidly transferred to tape from computer memory during experimental sessions. The data along with identification codes supplied by the program are recorded serially on tape for later retrieval and analysis. The total unformatted capacity of a single extended-length cartridge tape is approximately 17 million bytes, or 4-MB per each of the four tracks.

Digital Input and Output. A parallel I/O board provides 16 bits of input and 16 bits of output under program control. This interface enables PEARL II to interact with other digital devices and to sense signals from apparatus such as manual response units operated by the subject. The digital I/O unit includes a DEC DRV11 module with additional circuitry for response sensing developed by the CPL.

Other System Components. A complete PEARL laboratory package typically includes a Matrox video display generator, a graphics terminal, and a plotter. The display processor is used to produce visual stimuli such as checkerboard patterns for visual evoked potentials or words for cognitive ERP experiments. The graphics terminal presents on-line subject performance data to the experimenter and is also used for examination of ERP waveforms after each session. Hardcopy records of the waveforms can be made on the plotter. PEARL includes a four-line serial interface unit for system connections to units such as plotters, line printers, modems, voice synthesis modules, and other computer systems. The following devices have been connected to PEARL systems in one or more laboratories: counters for integrating multiple unit activity, a vector display generator, an additional video display generator, an array processor, and nine-track magnetic tape drives.

PEARL Software Overview

A major goal of the PEARL project was development of a package of programs to support ERP tests, with each program allowing selection of the common versions of each test through parameter specification. This approach overcomes limitations in test flexibility and in program maintenance which follow from the manner in which laboratory research software is traditionally developed. In a typical research environment, programs are usually developed for the purpose of conducting some specific study and are rarely generalized beyond the needs of the study at hand. Consequently, recurring software development efforts are required each time the experimental plan is altered. Often, the result is a collection of programs that are difficult to maintain because the software has been rewritten by several authors, with reasons behind coding strategies obscured by successive strata of modification. In application areas where the nature of the research demands that large volumes of complex data be manipulated, the absence of a focused approach to research programming can lead to a geometric explosion of effort over time.

In the PEARL II Project, we have attempted to obviate the reprogramming cycle by raising to the level of preprogrammed parameter selection the manner in which ERP investigators can make substantial changes in a research design. Thus, the investigator specifies a particular version of a test by editing tables of experimental options, labels, and parameters, rather than by programming in a computer language. It is possible, by giving greater initial attention than is customary to the design of a experiment control/data acquisition program, to produce a program that is capable of running a large number of related research paradigms. Many tasks that previously required new programming become mere special cases of the "battery style" programs.

Additional benefits accrue from the battery approach to laboratory software development. The presence of a significant degree of flexibility facilitates exploratory approaches to experimentation, similar to the range of investigation that many statistical packages give in data analysis. With the impediment of having to be concerned with new program development frequently eliminated, the researcher can be in a position to try out research directions that might otherwise have seemed too cumbersome to undertake. Another benefit of the battery approach is that by concentrating effort on a single general purpose program, rather than on an increasingly diffuse collection of programs, greater attention can be given to error checking and debugging. Also, new features added to a general program are instantly available to all the applications that utilize it.

The generalized software battery approach certainly does not satisfy all needs. Often, the best way to realize a new research idea will be to produce a custom computer program to conduct it. Research, by its nature, will always defy attempts to anticipate and to package solutions for the questions it might pose. However, for relatively mature, stable lines of research, the battery scheme can prove quite fruitful. For new directions, the existence of well-conceived programs for similar experiments can provide a valuable model. Further, the subroutine libraries and data management schemes underlying existing battery programs facilitate generation of programs for new experiments. The PEARL II LABPAK programming environment was designed to support development of programs for any sort of experimentation that might be executed with the PEARL hardware.

Within this context, it is useful to distinguish between two modes of operation of the PEARL system. In the applications mode, the user loads standard test battery programs by selecting the appropriate item from a menu of tests. The PEARL software battery includes flexible programs for visual oddball [10], auditory oddball [11], visual Sternberg task [12], visual monitoring [13], dual-task critical (Jex) tracking [14], warned oddball [15], brainstem auditory potentials [16], visual pattern reversal potentials [17], and heart rate monitoring [18]. Each program includes a parameter section that gives the operator considerable control over the manner in which the test is conducted. A brief summary of these tests may be found in Table I.

In the development mode, the user interacts with DEC's RT11 operating system to develop applications programs which are then linked with PEARL device driver modules stored in a standard system library file. The software for the PEARL II system has been designed so that scientists or technicians with modest programming skills can develop applications programs in a high-level programming language. The laboratory support package supplied with PEARL (LABPAK) provides the programmer with a means to control all PEARL II laboratory devices (clocks, A/D converter, digital I/O, D/A converters, and cassette tape system) with standard subroutine calls from a FORTRAN program.

The Battery User Interface

Particular attention has been paid in the PEARL software package to the user/operator interface to the battery items. Such an emphasis seems appropriate for a number of reasons. First, an enhanced user interface can lead to an expansion in the number of investigators utilizing particular tests by reducing the level of computer expertise needed to execute the research. Second, a consistent, simple interaction scheme makes a software package easier for a novice user to learn. Third, as the number of parameters increase, a more powerful user interface is required to deal with the complexity associated with a flexible program.

The battery presents a menu driven user interface to the operator. The initial battery menu gives a list of the items in the battery. The novice user selects items with a cursor controlled by standard arrow keys on the terminal keyboard. After the user becomes familiar with the menus, the cursor-controlled menu selection process may be circumvented with a more efficient parallel, single keystroke synonym mechanism.

Each battery item presents the user with a main menu of its own. While there are item specific variations, the following options will typically be available:

Inspect or Alter Parameters

Run a Block of Trials

Manipulate Tape

Display Data

Exit this Battery Item.

Parameter Specification. The first item on the menu allows the user to inspect or edit the table of option, text, and numeric parameters that configure each battery item to perform the specific task from among the domain of experiments it is capable of executing. This table is called the parameter dictionary. The parameter dictionary is organized into functionally related pages, that in turn are composed of individual parameters. A parameter directory appears as the initial menu in the parameter section, as illustrated in Figure 2a. The directory menu leads to the next level of menus, which contain the individual parameter table entries, as displayed in Figure 2b.

Parameter pages display groups of parameters using symbolic names, together with their current values, legal ranges, units, and access information. Some parameter values are designated as being for inspection only, while other can be changed by the operator. Those variables that are deemed accessible to the operator can be easily modified by first selecting a parameter, then entering a new value. Numeric parameters can be specified as a function of other parameters in combination with numeric constants. A calculator routine built into the battery user interface evaluates such expressions and places the result in the parameter dictionary as the new value for the edited parameter. Another significant feature of the user interface is the parameter description feature. A single keystroke provides additional descriptive information for any parameter to the user, as illustrated at the bottom of the screens depicted in Figure 2. A help screen describing general parameter section facilities is also available.

Parameter tables are stored twice. First, after the user completes editing the table for a particular experiment, the parameter table is written to a disk file. This file can be read for future experimental sessions that will use the same set of parameters. Second, the parameter table, including symbolic names for all parameters, is written to the data storage tape at the beginning of each session. Thus, there is a complete record of all parameters attached to the data file for each session.

Run Section. After parameters have been adjusted according to the plan for the particular experimental session, the operator enters the Run section from the main menu. Initiation of the experiment is preceded by a capsule summary which previews the session, as illustrated in Figure 3a. The Block Preview gives the operator a chance to recheck the experiment specification before data collection begins. If the operator elects to continue, presentation of stimuli and collection of data commence at this point.

During data collection, the operator is presented with extensive information as to the progress of the experiment. With the slow-wave ERP items, for example, a table appears on the operator's terminal giving values for such experimental trial variables as stimulus type, subject response time, estimated P300 peak, and ocular artifact activity, as in Figure 3b. In addition, a running average for any of the recorded leads can be displayed using the D/A system and an oscilloscope. This on-line display is especially useful for monitoring the integrity of the electrophysiological data path.

Display Section. Each battery item contains a Display section which offers waveform displays appropriate to the particular class of ERPs recorded in the experiment. In most cases, two types of waveform displays are offered: Distribution displays overplot waveforms for different electrode recording sites and Measurement displays present waveforms for assessment of basic experimental effects. The Measurement sections allow the experimenter to inspect and score waveform characteristics, including the peak amplitude, peak latency, and area of ERP components. The displays can be dumped to a plotter for hardcopy records (see Figure 4).

PEARL II LABPAK Library

The PEARL II LABPAK Library is a collection of device interface and utility subroutines that allow access to the unique capabilities of the PEARL II hardware. The routines may be called from the FORTRAN level, so that all user program development can be done in a relatively high-level language. The core of the PEARL II LABPAK library has descended directly from the CPL's original LABPAK library, described in Donchin and Heffley [9].

The guiding philosophy behind LABPAK has been to employ fast, carefully coded assembly language subroutines for those functions where real-time or space constraints make using a high-level language impractical, while retaining the convenience of high-level language programming elsewhere. This approach, which is now widespread in the realm of research programming, allows complex application programs to be developed quickly and permits programs to be modified much more readily than would be the case if pure assembly language were used. The use of assembly language in the LABPAK subroutines in tandem with a high-level language processor (FORTRAN) that generates relatively efficient code at the applications level allows the PEARL programming package to fulfill realtime contraints that cannot be satisfied by lab-BASIC processors or pure high-level language implementations.

Program design is further facilitated by the ease with which PEARL II LABPAK routines allow the programmer access to the full power of the PEARL II peripheral devices. The PEARL devices are designed to perform their functions with a minimum of program intervention. For example, the LABPAK A/D routines allow the programmer to schedule with a single call a series of digitizer sweeps, with a specified number of channels, sweep interval, and total number of scans. The PEARL A/D system then conducts data collection in parallel with other system actions, freeing the processor for other functions. When a full bufferload of data have been collected, the program is notified via a designated flag variable. The PEARL II A/D, D/A, programmable clock, and magtape systems may all be operated in this fashion.

LABPAK also contains a number of utility routines. Subroutines have been developed to perform certain operations that might be too slow if they were coded in FORTRAN. Examples of these sorts of operations are block data moves, running average calculation, and display device updates. Recent additions to the PEARL II LABPAK library are subroutines that allow LABPAK programs to access, with performance adequate for realtime applications, memory above the standard 56-KB boundary. These subroutines facilitate development of programs for experiments that require acquisition of large numbers of data points for each trial.

An unusual feature of the PEARL II LABPAK library is its built-in debug trace feature. The programmer may, should the necessity arise, specify that system generate a trace message and optional pause upon the entry to each LABPAK subroutine called from an application program. The programmer need do nothing special to generate programs with this ability; it is included with each program when it is linked with the LABPAK library. The user types a single-line RT11 command to enable the debug trace feature. LABPAK also augments the runtime error checking ability of RT11/FORTRAN by attempting to detect inadvertent changes to pure code or data. This is done using a checksum scheme.

While FORTRAN remains the primary language for the PEARL battery, the project has employed a FORTRAN preprocessor called FLECS [19] for program development. Standard RT11/FORTRAN lacks a number of features that are available in newer programming languages that make development of large applications program much more efficient. FLECS ameliorates many of the shortcomings of FORTRAN by providing features such as modern control structures and variable scope rules. The FLECS preprocessor accepts an input syntax that more closely resembles the C programming language than FORTRAN. The FLECS output is then converted into machine code by the standard FORTRAN processor.

Summary and Evaluation of PEARL Project

The PEARL Project has several aspects. Successes and setbacks have been experienced in each domain, with many lessons learned.

As a Research Tool. The PEARL program may be viewed as an attempt to generate a research tool, specifically a portable experiment control/data acquisition system for ERP research. In this sense, the project has clearly been successful in that more than twenty fully-functional systems have been built and installed in various scientific research facilities, including a mobile van operated by the EPA. The CPL now relies on PEARL exclusively for all its laboratory systems. Although the system has not achieved miniaturization, developments such as removable Winchester disk drives have made the system more truly portable.

Computer Software Engineering. The PEARL Project had its roots in a collaboration between Psychologists and Computer Scientists. It remains a study in the application of ideas from computer science to practical programming problems. The PEARL Battery continues to explore the benefits of attempting to apply ideas such as integrated data management, good user interface design, and more extensive use of graphical presentations to psychophysiological research.

One lesson learned from the project relates to overconfiguring software relative to existing hardware. As the PEARL Battery development progressed, the capacity of existing microcomputer hardware was exceeded by the demands of the elaborate software package deemed necessary to satisfy project goals. Rather than abandon goals that had solid merit, an implicit decision was made to program beyond the abilities of the current hardware. Fortunately, microcomputer hardware and software systems (the LSI-11/ RT11 family in this case) have now grown to the point where the PEARL Battery operates rapidly and efficiently. Had the project goals been set aside temporarily so that the PEARL Battery could be written to operate entirely within the bounds of microcomputer hardware from 5-10 years ago, we would now have a mass of software with an internal structure that would be totally inadequate to the full objectives of the PEARL Project. The practicality of rewriting a major software package is perhaps much less certain than the development of computer systems with faster processors and greater storage capacity.

Conclusion. As scientific research in psychology, physiology, medicine, and human engineering becomes more sophisticated, experimental plans will surely require more complex computer hardware and software for their execution. Further, research objectives in many disciplines call for studies involving greater numbers of subjects, which imposes additional requirements.

The "battery" approach to laboratory systems design offers considerable promise toward meeting the demands of these growing research programs. Good user interfaces and comprehensive data management facilities will be the hallmark of successful laboratory applications systems. In our experience with the PEARL project, we have learned the value of a highly interactive core development group composed of scientists, computer programmers, and engineers. Further, positive and open collaboration with other laboratories is vital in the development of general research tools.

References

1. Donchin, E. Event-related brain potentials: A tool in the study of human information processing. In: H. Begleiter (ed) Evoked Brain Potentials and Behavior, Plenum Press, New York, 13-75, 1979.

2. Donchin, E., W. Ritter and W. C. McCallum. Cognitive psychophysiology: The endogenous components of the ERP. In: Event-Related Brain Potentials in Man, edited by E. Callaway, P. Tueting, and S. H. Koslow, Academic Press, New York, 349-441, 1978.

3. Isreal, J. B., G. L. Chesney, C. D. Wickens and E. Donchin. P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. Psychophysiology, 17: 259-273, 1980.

4. Kutas, M., G. McCarthy and E. Donchin. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 197: 792-795, 1977.

5. Donchin, E., M. G. H. Coles and G. Gratton. Cognitive psychophysiology and preparatory processes: A case study. In: Preparatory States and Processes, edited by S. N. Kornblum and J. Requin, Erlbaum, Hillsdale, New Jersey, 155-178, 1984.

6. Hillyard, S. A., T. W. Picton and D. Regan. Sensation, perception and attention: Analysis using ERPs. In: Event-Related Brain Potentials in Man, edited by E. Callaway, P. Tueting, and S. H. Koslow, Academic Press, New York, 223-347, 1978.

7. Otto, D., S. Baumann and G. Robinson. Application of a portable microprocessor-based system for electrophysiological field testing of neurotoxicity. Neurobehavioral Toxicology and Teratology, 7: 409-414, 1985.

8. Otto, D. (ed) Multidisciplinary Perspectives in Event-Related Brain Potential Research. EPA-600/9-77-043. U.S. Government Printing Office, Washington, DC, 1978.

9. Donchin, E. and E. Heffley. Minicomputers in tne signal-averaging laboratory. American Psychologist, 30: 299-312, 1975.

10. Donchin, E. Surprise!...Surprise? Psychophysiology, 18: 493-513, 1981.

11. Duncan-Johnson, C. C. and E. Donchin. On quantifying surprise: The variation in event-related potentials with subjective probability. Psychophysiology, 14: 456-467, 1977.

12. Sternberg, S. Memory scanning: mental processes revealed by reaction time experiments. American Scientist, 57: 421-457, 1969.

13. Heffley, E. Event-related brain potentials in visual monitoring tasks: Selective attention, display format, and event frequency. Paper presented at the Annual Meeting of the Eastern Psychological Association, 1981.

14. Jex, H. R., J. P. McDonnell and A. V. Phatak. A "Critical" tracking task for manual control research. IEEE Transactions on Human Factors in Electronics, HFE-7: 138-144, 1966.

15. Otto, D., V. Benignus, K. Muller and C. Barton. Effects of age and body lead burden on CNS function in young children: I. Slow cortical potentials. Electroencephalography and Clinical Neurophysiology, 52: 229-239, 1981.

16. Stockard, J. J., J. E. Stockard and F. W. Sharbrough. Brainstem auditory evoked potentials in neurology: Methodology, interpretation, clinical application. In: Electrophysiologic Approaches to Neurologic Diagnosis, edited by M. J. Aminoff, Churchill-Livingstone, New York, 370-413, 1979.

17. Regan, D. Evoked potentials in psychology, sensory physiology and clinical medicine, Chapman & Hall, London, 1972.

18. Singleton, W. T., J. G. Fox and D. Whitfield (eds) Measurement of Man at Work: An Aparaisal of Physiological and Psychological Criteria in Man-machine Systems. Taylor and Francis Ltd., London, 1969.

19. Beyer, T. FLECS User Manual. University of Oregon, Eugene, Oregon, 1975.

Table I: PEARL ERP Test Battery

1. Brainstem Auditory Evoked Potential (BAEP)

   Application: Basic neurological assessment of brainstem auditory pathways

   Major variables: Stimulus intensity, duration, and rate.

2. Pattern Reversal Evoked Potential (PREP)

   Application: Basic neurological assessment of visual pathways

   Major variables: Checkerboard contrast, check size, and reversal rate.

3. Auditory Oddball

   Application: Cognitive abilities to classify auditory stimuli into sets defined by experimenter instructions. Often used as a secondary task to measure mental workload.

   Major variables: Stimulus discriminability, relative probability, interstimulus interval, and type of response.

4. Visual Oddball

   Application: Cognitive abilities to classify visual stimuli into sets defined by experimenter instructions.

   Major variables: Task (word classification, brightness judgment, color discrimination), relative probability, set size, interstimulus interval, and type of response.

5. Visual Monitoring

   Application: Cognitive abilites to process displays of varying levels of complexity.

   Major variabilies: Number of display elements, relative probability of significant events, interstimulus interval, and type of response.

6. Sternberg Memory Task

Application: Short-term memory search task. Sometimes used as a secondary task to assess memory workload.

Major variables: Set size, display size, masking, type of response, and presentation rate.

7. Jex Critical Tracking Task

Application: Test of ability to perform a complex manual tracking task with visual display.

Major variables: Target forcing function characteristics, response dynamics, and other difficulty manipulations.

8. Slow Potentials

Application: Assessment of slow cognitive ERPs and motor potentials.

Major variables: Task (simple reaction time, choice reaction time), foreperiod duration, stimulus probability, type of stimulus classification, and response type.

9. General Purpose Averager

Application: Used in conjunction with triggers to enable ERPs to be recorded in response to externally-generated stimuli. Also, can be used to trigger external devices (somatosensory stimulators, for example).

Major variables: Internal/external trigger, stimulus timing, event classification, and response type.

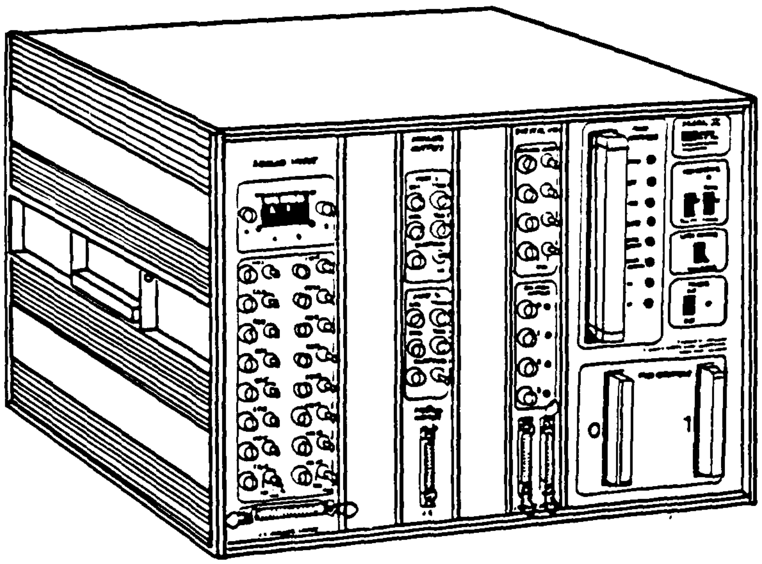
Figure Legends

1. The PEARL II packaging includes a variety of front panel connectors and switches which determine system inputs (top). A schematic representation of the functional components of the system is illustrated (bottom).

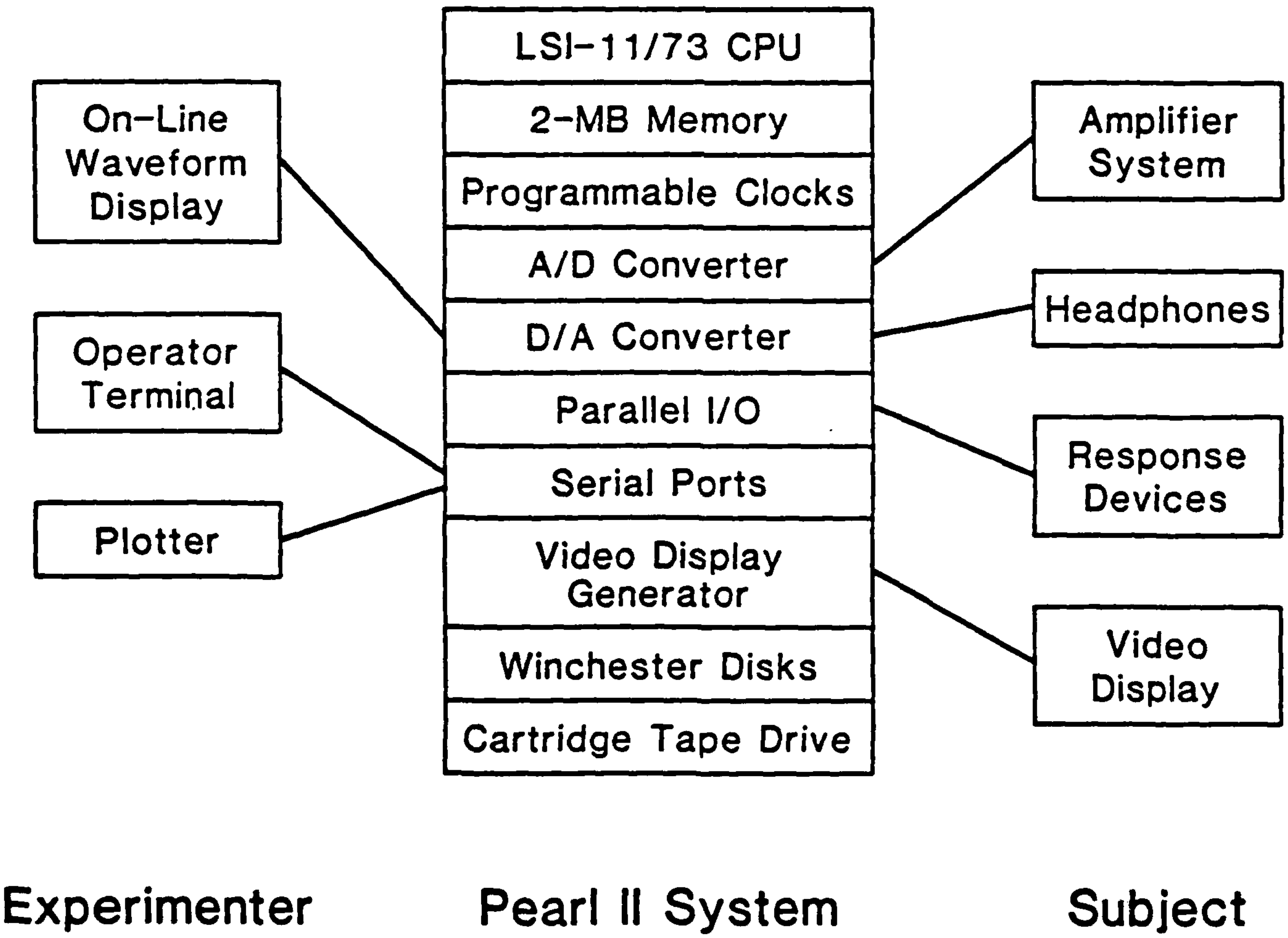
2. Menus are used to select and review parameters for each test. The first page in the parameter section contains a directory of groups of similar parameters (top). The operator can set parameters according to the specific plan for the experiment (bottom). The parameter menus are associated with an underlying data management scheme that provides range checking, protection, and user help.

3. The Block Preview allows the operator to review the characteristics of the experiment before beginning each session (top). During execution of the experiment, the investigator receives an on-line summary for each trial, including: stimulus, response, reaction-time, response code, stimulus onset asynchrony, good/total trials for each category, electro-ocular activity, electro-myogenic activity, and an estimate of the peak latency for the ERP component of interest (bottom).

4. The PEARL II display sections produce hardcopy plots of average waveforms for each stimulus category, along with the number of trials in the average, the electrode channel selected, and ERP component measurements determined by latecy and polarity parameters selected by the operator. Peak amplitude, component area, and peak latency are displayed.



COGNITIVE
PSYCHOPHYSIOLOGY
LABORATORY



V I S U A L   O D D B A L L

Page Name: User ID Directory

Type H for Help, Q to Quit this Section

ID Name	Type	Value	Units	Access	Low	High
-------	Direct	-----	-----	Constant	---	----

Data Parameters -- Basic Data Collection Parameters.
This page contains the IDs controlling basic digitizing parameters such as epoch, number of points, and channel collected, etc.

V I S U A L   O D D B A L L

Page Name: Data Parameters

Calc Epoch =	No	[Y/N]	Epoch	=	2000	MilliSec
Calc Rate =	No	[Y/N]	A/D Rate	=	5000	MicroSec
Calc Pnpts =	Yes	[Y/N]	Points	=	401	#
Start Chan =	1	[1-16]	Channels	=	5	[1-16]
Bytes Used =	4010	Bytes	A/D Gain	=	1	[0-3]
Bytes Free =	21590	Bytes				

Type H for Help, Q to Quit this Section

ID Name	Type	Value	Units	Access	Low	High
Points	Integer	401	#	All	1	12800

Points -- Number of Digitizer Sweeps per Trial.
"Points" is the number of Digitizer Sweeps per Trial the Analog Input System will make. Thus, "Points" Points will be recorded for each active channel.

B L O C K   P R E V I E W   I N F O R M A T I O N

Epoch Length: 2000 Milliseconds. Baseline: 95 Milliseconds.
5 Channels, 401 Points/Chan Digitized every 5000 Microseconds.
4010 out of an available 25600 Bytes/Trial will be Used.

Stimuli are Program Generated.
Trials: 96, 9 Category A Trials (9%), and 87 Category B Trials (91%).
Anticipated Block Duration: 12:48.
SOA will vary between 7200 and 8800 Milliseconds.

Visual Stimuli will be presented on the Matrox Display.

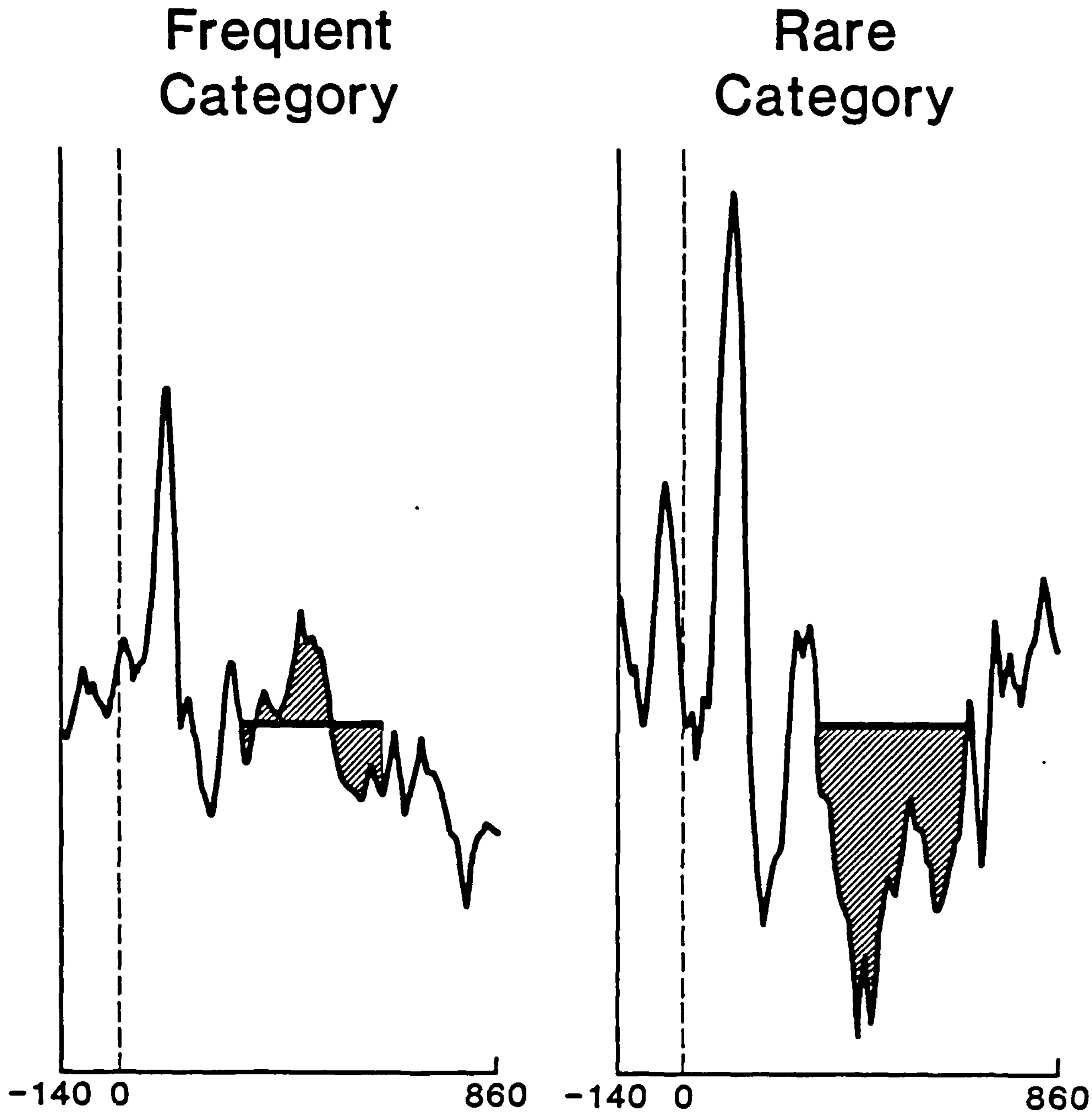
Catogory A Stimuli on Output Bit 0, Category B Stimuli on Bit 1.

Q to Quit, RETURN to Continue

V I S U A L   O D D B A L L

Trial	Stimulus	Response	RT	Resp. Code	SOA	G/T A	G/T B	EOG	EMG	PEAK
1	Potato	Category-B	511	Correct	8729	0/ 0	1/ 1	541	0	595
2	Mouse	Category-A	560	Correct	7716	1/ 1	1/ 1	320	0	630
3	Tomato	Category-B	495	Correct	8183	1/ 1	2/ 2	693	0	520
4	Bear	Category-B	598	Incorrect	7423	1/ 2	2/ 2	250	0	575
5	Horse	Category-A	576	Correct	7969	2/ 3	2/ 2	810	0	460
6	Birch	Category-B	554	Correct	7557	2/ 3	3/ 3	442	0	485
7	Maple	Category-B	521	Correct	8144	2/ 3	4/ 4	568	0	455
8	Radish	Category-B	509	Correct	8293	2/ 3	5/ 5	955	0	470
9	Tulip	Category-B	517	Correct	7625	2/ 3	6/ 6	678	0	510
10	Rose	Category-B	490	Correct	7971	2/ 3	7/ 7	342	0	490
11	Violet	Category-A	588	Incorrect	8658	2/ 3	7/ 8	511	0	585
12	Giraffe	Category-A	567	Correct	8234	3/ 4	7/ 8	280	0	530
13	Monkey	Category-A	539	Correct	7871	4/ 5	7/ 8	535	0	495
14	Daisy	Category-B	543	Correct	7214	4/ 5	8/ 9	489	0	505
15	Lemur	Category-A	592	Correct	8603	5/ 6	8/ 9	306	0	550

PAUSED -- A or B for Sample Stimulus, RETURN to Continue



	Frequent Category	Rare Category
Trials:	20	8
Peak:	5	19
Area:	29	236
Channel:	4	4
Latency:	550	395

PSYCHOPHYSIOLOGY
Vol. 22, No. 3
Printed in U.S.A.

The Interpretation of the Component Structure of Event-Related Brain Potentials: An Analysis of Expert Judgments

ARTHUR F. KRAMER
Department of Psychology, University of Illinois at Urbana-Champaign

ABSTRACT

The analysis of components of the event-related brain potential (ERP) is accomplished through standardized statistical algorithms. However, visual inspection of ERPs is commonly used to guide the selection of the analytic techniques. Thus, the definitional criteria of components employed by the investigators may interact with the choice of analysis procedures. The present study examines the criteria which investigators employ to define components of the ERP, the relative importance assigned to different criteria, and the consistency with which investigators use the definitional criteria. ERPs are simulated so as to vary systematically the amplitude, latency, duration, and electrode distribution of the P300 component. Ten experienced ERP researchers rated the similarity of 153 pairs of simulated ERPs. The ratings were analyzed by a multidimensional scaling procedure. Eight unidimensional judgments on each of the 18 ERPs were also obtained. The results suggest that ERP investigators are capable of accurately recovering the underlying dimensions of a set of simulated ERPs. Furthermore, the degree of experience in analyzing and interpreting ERPs affects the weighting structure of the underlying dimensions. These findings support the commonly held, but previously untested, belief that judgments of ERPs based on visual inspection can be both accurate and reliable.

DESCRIPTORS: Event-related brain potential (ERP), P300, Multidimensional scaling, Component definition, Subjective judgment.

The event-related brain potential (ERP) is a transient series of voltage oscillations that can be recorded from the scalp in response to the occurrence of a discrete event (Donchin, 1975; Regan, 1972). The ERP is viewed as a sequence of separate but sometimes temporally overlapping components which are influenced by changes in the physical parameters of stimuli (exogenous components) or psychological constructs such as expectancy, memory, and strategy (endogenous components). Components are typically labeled with either an "N" or "P" denoting negative or positive polarity and a number indicating their minimal latency measured from the onset of an eliciting event (e.g. P300 is a positive going component which occurs at least 300 ms after a stimulus).

I gratefully acknowledge the helpful comments of Michael Coles, Monica Fabiani, Gabriele Gratton, and Larry Jones. This research was supported by the Air Force Office of Scientific Research under contract No. F49620-83-0144 with Dr. Alfred Fregly as technical monitor.

Address requests for reprints to: Arthur F. Kramer, Department of Psychology, University of Illinois at Urbana-Champaign, 603 East Daniel Street, Champaign, Illinois 61820.

The attributes of an ERP that have served as definitional criteria for "components" include: electrode distribution, polarity, amplitude, and latency (Donchin, Ritter, & McCallum, 1978; Sutton & Ruchkin, 1984). Variance in the amplitude and latency of the endogenous ERP components is normally accounted for by the tasks assigned to the subjects. Thus, the sensitivity of the ERP components to experimental manipulations represents another criterion. The electrode distribution refers to the relative amplitude and polarity of a component across the scalp for a fixed temporal interval. Thus, one component may be positive at a parietal location and negative at a frontal site at time t(n) while another component might possess the opposite polarity-location relationship at time t(n). The latency criterion is applied with respect to its changes as a function of experimental manipulations. Thus, particular experimental manipulations are expected to result in systematic changes in the latency of a component. Furthermore, each component is assumed to operate in a specific latency range. The amplitude of a component is also useful as a definitional criterion with respect to its modulation as a function of experimental manipulations.

Another criterion employed by some investigators is the neuroanatomical source of the scalp recorded potential (Goff, Allison, & Vaughan, 1978). Although the neuroanatomical source provides a means by which to discriminate between scalp recorded components, it cannot be derived directly from the morphological characteristics of the waveform as are the other criteria. It is also difficult to realize this criterion in intact humans. Due to these differences, this criterion will not be dealt with further in the present paper.

Although the definitional criteria may appear to lead to a straightforward method of differentiating among components, in practice the definitional process is not quite so simple. One obvious issue is the manner in which investigators weight different criteria. Do investigators apply equal weighting coefficients to each of the dimensions (electrode distribution, latency, and amplitude) or is there a differential weighting of the dimensions? A second question involves the evaluation of the consistency with which these weights are applied. The answers to these questions depend on at least three factors: 1) the laboratory or laboratories conducting the study in which the ERPs are obtained, 2) the reliability of the experimental paradigm employed, and 3) the signal/noise ratio of the components.

Laboratories differ in the magnitude of the latency range they will accept for a given component. Some laboratories base their decisions on the specific experimental manipulations. For example, it has been suggested that P300 latency is related to the duration of stimulus evaluation required for a task relevant event (Donchin, 1981; McCarthy & Donchin, 1981). If the relationship between P300 latency and stimulus evaluation is valid, one would expect that the latency of P300 would vary with the processing requirements of the task. Thus, in this case the latency range of the P300 would be potentially wide. However, other laboratories tend to define new components if any large differences in latency are observed. In this case the latency range of a component is relatively small. The range of variance across laboratories in dealing with electrode distribution is somewhat smaller than with latency although some laboratories allow some degree of intersubject variability while others do not (Courchesne, 1978). The interlaboratory variability in dealing with the polarity of the component is relatively small.

Thus far we have discussed differences among laboratories in assigning weights to the definitional criteria. What about differences among different investigators within a single laboratory? Although most of the pattern recognition and signal extraction procedures employed in ERP research are performed by standardized statistical alogrithms (see Donchin & Heffley, 1978; Coles, Gratton, Kramer, & Miller, in press), visual inspection of the waveforms is commonly used to guide the selection of the analytic techniques. Thus, the investigator interacts with the data set and derives expectations concerning statistical outcomes based on the visual characteristics of the waveforms. It would seem, therefore, that inter-investigator variability in the selection of weights for the definitional criteria would contribute to variance in the later decisions about the types of analytic techniques to be applied to the data as well as to variance in the interpretation of the waveforms within their experimental context.

Since the amplitude and latency of components provide definitional criteria only within an experimental framework, the reliability of the relationship between these ERP attributes and experimental manipulations is of concern. More faith is placed in these criteria in well established paradigms than in newly tested experimental situations. For example, the relationship between P300 amplitude and subjective probability, larger amplitude P300s elicited by rare events, has been demonstrated numerous times and is very robust (Pritchard, 1981). Thus, obtaining the probability effect is taken as strong evidence for the existence of the P300 component in a data set.

The identification and measurement of ERP components is complicated by a relatively low signal to noise ratio in the single trial data (Coles et al., in press; John, Ruchkin, & Vidal, 1978). The interaction of the noise with the definitional criteria of the component makes the assignment of any set of weights difficult. Thus, in terms of the signal to noise ratio, the issue becomes not only one of the differential weighting of the different criteria but also the detection and identification of the attributes of the component in the waveform. The ERP is typically defined as that portion of the waveform which is time locked to a stimulus or response (Donchin, Kramer, & Wickens, 1982). Noise is defined as everything else (e.g. ongoing EEG). The noise distribution possesses an amplitude of between 50 and 100 microvolts while the ERP amplitude ranges from 1 to 25 microvolts. Clearly, it would be difficult to extract an ERP component with this signal to noise ratio. Several methods have been proposed to deal with this problem. In one such procedure, signal averaging, the degree of attenuation of the noise is inversely proportional to the square root of the number of single trial waveforms in the sample. Although the signal averaging procedure serves as an efficient method of reducing the noise and thereby making the extraction of the component easier, there are several assumptions which must be met prior to its use. These assumptions include: 1) the component must be temporally invariant over repeated presentations of the stimulus, 2) the morphological characteristics of the component must be invariant over trials, and 3) the noise must not be systematically related to the component. However, if the temporal invariance assumption is not met, iterative autocorrelation procedures can be employed to temporally align the single trial waveforms prior to averaging (Nahvi, Woody, Ungar, & Sharafat, 1975; Woody, 1967). Other procedures such as filtering also serve to reduce the amplitude of the noise relative to the amplitude of the component, provided that the two distributions are not highly correlated.

The problems of component identification and measurement discussed above illustrate some of the difficulties in the analysis and interpretation of ERPs. Two of these issues, the inter-investigator variability in assigning weights to the definitional criteria and the detection and identification of components in waveforms, can be thought of in terms of questions of consistency and validity, respectively. To begin with, investigators must be able to recover the underlying dimensions of the data. These dimensions can be thought of as the definitional attributes of the components. As has been mentioned previously, visual inspection of waveforms usually guides the selection of signal extraction and pattern recognition techniques to be applied to the data set. Thus, investigators must be capable of accurately perceiving the important characteristics of the waveform. This can be translated into a question of validity. Given a known set of inputs (dimensions) can the ERP investigator reliably recover them? This question is addressed in the present study by simulating changes in three P300 dimensions (latency, amplitude, and electrode distribution) and asking investigators from one laboratory to make similarity judgments on a pair by pair basis.

The dimensions employed by the investigators in making the similarity judgments will be evaluated with a multidimensional scaling (MDS) solution. MDS is a class of data reduction techniques designed to derive the underlying structure of a data set. It is particularly well-suited for the situation in which the dimensions employed by subjects to differentiate complex stimuli are unknown. In the case of ERPs, the data are most certainly complex and the dimensions used by investigators to distinguish between components are uncertain. The input to the MDS procedure is a matrix of numbers that express the similarity or dissimilarity of pairs of stimuli. In the present case, these are the similarities of simulated ERPs which vary systematically along three dimensions: latency, amplitude, and electrode distribution of P300. The output of the MDS procedure is a set of two-dimensional maps which present a geometric configuration of the stimuli. Stimuli judged to be highly similar are represented as points close to each other in the spatial maps. Stimuli judged to be dissimilar are represented as points distant from one another.

Once it has been established that the ERP investigators can in fact recover the underlying component dimensions, the question of inter-investigator consistency in selecting dimensional weights can be considered. This issue will be addressed by examining the single subject weights assigned to the dimensions. The INDSCAL individual differences model for MDS (Carroll & Chang, 1970; Carroll & Wish, 1974) will be employed in the present study. The SINDSCAL program that implements this technique provides several useful outputs (Pruzansky, 1975). One output is an aggregate solution of the underlying dimensions in the data set. The aggregate solution provides information concerning the dimensions employed by the subjects in making the pairwise comparisons. Another output is a set of single subject weights on the dimensions. The INDSCAL model assumes that a single set of dimensions is adequate to represent the underlying structure of the data. Although all of the subjects are assumed to employ the same dimensions in making their pairwise comparisons, single subject preferences are accommodated by permitting subject weights on each of the dimensions to vary across subjects. Thus, the dimensional structure of the data is fixed for a given solution while the subject weights are free to vary.

Method

Subjects

Eight graduate students and two research associates served as subjects. All of the participants held a research position at the Cognitive Psychophysiology Laboratory, University of Illinois, at the time of the study. Each of the subjects was familiar with the field of electrophysiology, including the design of ERP experiments and the analysis and interpretation of ERPs. Experience ranged from 9 months to 10 years. None of the subjects had seen the specific stimuli (ERPs) prior to the experiment.

Stimuli

The stimuli consisted of 18 plots (16 × 16cm) of three simulated ERPs. Each set of waveforms included x and y axes which provided the subjects with temporal (x axis) and voltage (y axis) information, a stimulus marker, and three electrodes (defined as F_Z - frontal, C_Z - central, and P_Z - parietal). The sets of ERPs were composed of four components—N100, P200, N200, and P300—as well as four bands of random noise—.25 to 5.00 Hz, 5.25 to 10.00 Hz, 10.25 to 15.00 Hz, and 15.25 to 20.00 Hz. The amplitude of the noise bands was set at six units. Three of the four components did not change over the 18 ERP plots. N100 occurred at 100 ms post-stimulus, with a negative polarity, an amplitude of 20 units, and a duration of 110 ms. Weighting factors for the amplitude of each of the components were employed so that the electrode distributions of the components would be consistent with those in the published literature. The amplitude weighting factors for the N100 components were −.592, −1.816, and −.592. These weighting factors corresponded to the three electrodes: F_Z, C_Z, and P_Z, respectively. Thus in the case of N100, the component was most negative at C_Z and equal in amplitude at F_Z and P_Z. P200 occurred at 180 ms post-stimulus, with a positive polarity, an amplitude of 25 units, and a duration of 110 ms. The amplitude weighting factors employed for the P200 component were 1.0, 1.[illegible], and .5. N200 occurred at 220 ms post-stimulus, with a negative polarity, an amplitude of 25 units, and a duration of 200 ms. The amplitude weighting factors employed for the N200 com[illegible]ere −1.2, .[illegible], and .4.

Since the P300 com[illegible] of particular interest in the present study, the parameters used in its construction constituted levels of three independent variables. The amplitude of the P300 was either 60 or 120 units. The peak of the P300 occurred at either 280, 380, or 680 ms post-stimulus. Duration of the P300 was perfectly correlated with its latency. The three durations which corresponded to the latency values were 320, 350, and 440 ms. Electrode distribution was also an independent variable and therefore influenced the choice of the amplitude weighting factors. One set of weighting factors, hereafter to be referred to as the P_Z distribution, was .4, .8, and 1.2. A second set of weighting factors, hereafter to be referred to as the C_Z distribution, was .4, 1.2, and .8. The third set of weighting factors, hereafter to be referred to as the Flat distribution, was .4, .4, and 1.2. Polarity of the P300 component was always positive. Thus the three independent variables (2 amplitudes, 3 latencies, and 3 electrode distributions) combined to form 18 sets of ERPs. Each of the waveforms was digitally filtered (−3dB at 6.29 Hz, 0dB at 14.29 Hz) prior to reproduction. The filter parameters employed were consistent with those used in the analysis of endogenous components of the ERP (e.g. N200, P300). The ERPs employed in the study are presented in Figures 1, 2, and 3. Figure 1 presents small and large amplitude P300s at 280 ms post-stimulus for each of the three electrode distributions. Figures 2 and 3 present the same sequence of ERPs for 380 and 680 ms post-stimulus, respectively.

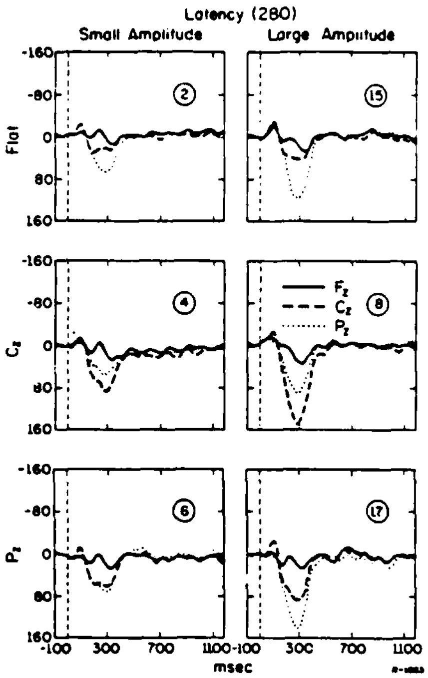


Figure 1. Simulated ERPs with large and small amplitude P300s at 280 ms post-stimulus for each of the three electrode distributions. Each plot contains three waveforms (F_Z, C_Z, P_Z). The stimulus number is located in the circle.

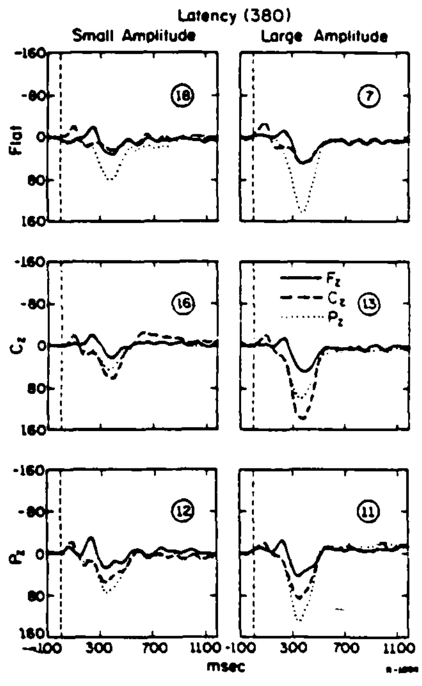


Figure 2. Simulated ERPs with large and small amplitude P300s at 380 ms post-stimulus for each of the three electrode distributions. Each plot contains three waveforms (F_Z, C_Z, P_Z). The stimulus number is located in the circle.

Procedure

Subjects were initially presented with the 18 sets of ERPs to familiarize them with the range of variability in the set. As the subjects were viewing the ERPs the experimenter read a set of instructions. Subjects were free to ask questions concerning the procedure or ERPs at any time. The instructions were as follows:

> You will be making pairwise comparisons of each of the 18 sets of ERPs with all of the other ERP sets. The total number of pairwise comparisons is 153. You will be judging the similarity of the ERPs on a one to seven scale;

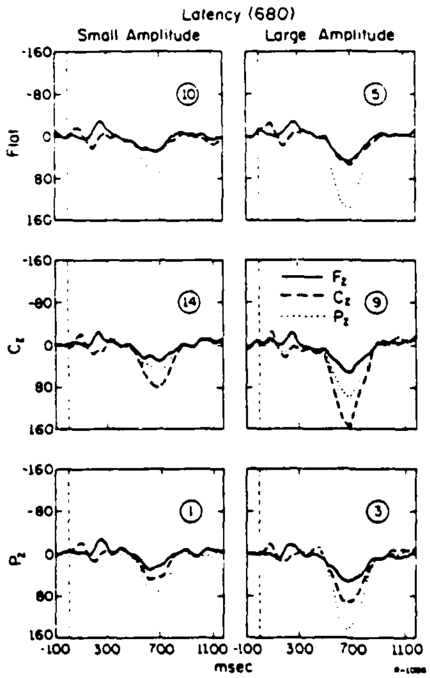


Figure 3. Simulated ERPs with large and small amplitude P300s at 680 ms post-stimulus for each of the three electrode distributions. Each plot contains three waveforms (F_Z, C_Z, P_Z). The stimulus number is located in the circle.

one representing very similar ERPs and seven representing very dissimilar ERPs. Each of the sets of ERPs consists of an x axis which indicates the temporal scale of the recording epoch (1.28 s), a y axis which provides voltage information, and a stimulus marker (dashed, vertical line). The solid line indicates the ERP recorded from the frontal electrode (F_Z), the dashed line represents the ERP recorded at the central electrode (C_Z), and the dotted line shows the ERP recorded at the parietal electrode (P_Z). Eighty units on the y axis correspond to 10 microvolts. Positive polarity is plotted down, negative polarity is plotted up.

You can think of these waveforms as being generated in a visual discrimination experiment in which the subject must detect and identify one of two stimuli. These stimuli are presented in a .20/.80 Bernoulli series. The subjects are instructed to count the total number of occurrences of either the low or high probability stimuli. The stimuli are either two names (e.g. Nancy and David) or exemplars of two categories (e.g. birds and furniture). The waveforms that you will judge can be thought of as single subject averages. Think about how each of these manipulations would affect the ERP components.

In actuality these waveforms have been simulated. The experimental description is provided so that you have a context in which the 18 waveforms could have been generated. Do you have any questions?

At this point subjects began to make the pairwise comparisons. The order of the presentation of the ERPs was random across subjects and ERP pairs. Eight of the pairwise comparisons were repeated so as to provide a reliability check on subjects' rating performance. Subjects were permitted to take a maximum of 30 seconds per pairwise comparison. Short breaks were permitted at any time during the rating procedure. The ERPs were printed on semi-transparent paper so that the subjects had the option of viewing them either side by side or by superimposing one on top of the other.

Following the 153 pairwise comparisons subjects were asked to rate each of the 18 sets of ERPs on eight univariate scales. These scales included ratings of P300 area, N200 amplitude, P300 electrode distribution, Classic P300, P300 latency, P300 amplitude, noise, and P300 duration. Each of the univariate ratings was performed on a seven-point scale. Experimental sessions lasted between 1½ and 2 hrs.

Results and Discussion

SINDSCAL Solution

The 10 single subject dissimilarity matrices obtained from the pairwise judgment of 18 ERP plots were submitted to the SINDSCAL individual differences procedure. Solutions were derived in one through five dimensions. The variance accounted for in each solution from one to five dimensions is shown in Figure 4. As can be seen from the figure, there is a relatively large increment in the variance accounted for from one to two dimensions (.24), somewhat less of an increase from two to three dimensions (.12), and a fairly small increase from three to four and four to five dimensions (.03 each). Visual inspection of the graphic representation of the fit statistic (R square) would suggest that three dimensions provide a reasonable fit to the stimulus space data (Kruskal & Wish, 1978). This is evidenced by the knee in the curve subsequent to three dimensions.

The three-dimensional SINDSCAL solution will be interpreted in two ways. First, the simulated ERPs will be represented spatially in two-space MDS plots and an intuitive interpretation of the dimensions will be offered. A second method of interpreting the SINDSCAL solution will involve regressing the three SINDSCAL dimensions on each of the univariate scales. For a univariate rating scale to provide a useful interpretation of a SINDSCAL dimension, two conditions must be fulfilled: 1) the multiple correlation for the scale must be high and statistically significant, and 2) the scale must have a high regression weight on that dimension. Satisfaction of these conditions indicates that the univariate scale closely corresponds to the dimension. The two methods of interpretation will be used to converge on a single explanation of the derived three dimensional SINDSCAL space.

Figure 5 provides a graphic representation of the 18 sets of ERPs in the 1-2 plane. The x axis represents the first dimension, the y axis represents the

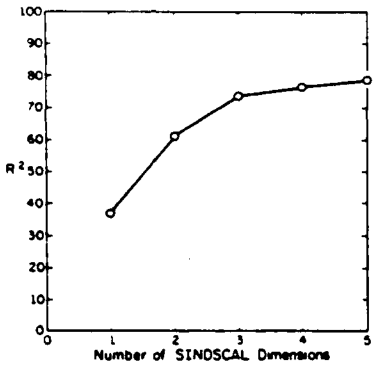


Figure 4. Variance accounted for by each of the dimensions in a five-dimensional SINDSCAL MDS solution.

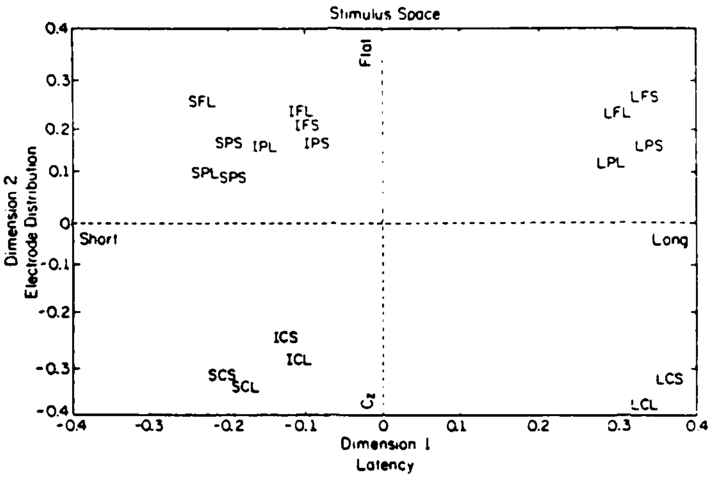


Figure 5. Stimulus space for MDS dimensions 1 (x axis) and 2 (y axis). The three-letter codes represent the levels of the independent variables. The first letter indicates the latency of the P300 component (S = short, I = intermediate, L = long). The second letter denotes the electrode distribution (P = P_z distribution, F = Flat distribution, C = C_z distribution). The third letter indicates the amplitude of the P300 component (S = small, L = large).

second dimension. Each of the 18 ERPs is indicated by a three-letter code which corresponds to levels of the independent variables. The first letter represents the latency of the P300 peak (S = short, I = intermediate, L = long). The second letter indicates the electrode distribution of the P300 (F = Flat distribution, C = C_z distribution, P = P_z distribution). The third letter represents the amplitude of the P300 component (S = small, L = large). Visual inspection of the 1-2 stimulus space suggests that dimension 1 represents the latency of the P300 component while dimension 2 depicts the electrode distribution of the P300. The geometric configuration of the ERPs on the x axis (P300 latency) closely corresponds to the actual latency manipulation. The short (280 ms) and intermediate (380 ms) latency P300s form a tight cluster while the long latency P300s (680 ms) are located at the other end of the dimension. Thus the relative distances between the simulated latencies are well represented by the subjects' judgments of them.

The rank ordering of the ERPs on the second dimension (P300 electrode distribution) also provides some interesting information. In this case the Flat and P_z distributions form a relatively tight cluster while the C_z distribution ERPs are located at the other end of the dimension. This geometric configuration corresponds well with investigators' perceptions of the important features of the electrode distribution in defining the P300 component. In both the P_z and Flat distributions the P300 component is maximal at the P_z electrode site whereas in the C_z distribution the P300 is maximal at the C_z site. P300 has typically been associated with a P_z maximal distribution, although C_z is usually found to possess a larger positive amplitude than F_z (Donchin et al., 1978). The present results suggest that the investigators in the sample placed the most importance on the maximally positive electrode site when making their pairwise judgments. The other two electrode sites (F_z and C_z in the Flat distribution and F_z and P_z in the C_z distribution) appeared to be less salient features of the electrode distribution.

Figure 6 presents a graphic representation of the 18 sets of ERPs in the 1-3 plane. As mentioned previously, the first dimension appears to represent the latency of the P300 component. The third dimension (y axis) seems to correspond to the amplitude of the P300 component. Visual inspection of the stimulus space suggests two clusters of ERPs: one with the large amplitude P300s and the other with the small amplitude P300 components.

The reliability of each subject's rating strategy over time was assessed by repeating eight of the 153 pairwise comparisons. The single subject Pearson correlation coefficients ranged from .80 to .96, with an average correlation across the 10 subjects of .88. These relatively high correlations suggest that subjects were uniformly consistent in their rating strategies over the 153 judgments.

The three-dimensional solution obtained from the SINDSCAL procedure corresponds well with the physical characteristics of the ERPs. Each of the derived dimensions appears to represent one of the dimensions of the simulated ERPs. Furthermore, the relative differences among the physical dimensions also appear to be well represented. This is

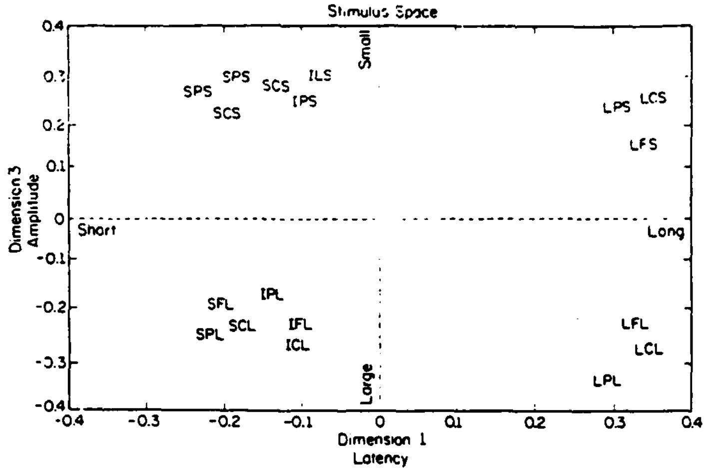


Figure 6. Stimulus space for MDS dimensions 1 (x axis) and 3 (y axis). The three-letter codes represent the levels of the independent variables. The first letter indicates the latency of the P300 component (S = short, I = intermediate, L = long). The second letter denotes the electrode distribution (P = P_z distribution, F = Flat distribution, C = C_z distribution). The third letter indicates the amplitude of the P300 component (S = small, L = large).

evidenced by the clusters formed on the latency dimension. In addition to accurately representing the physical characteristics of the stimuli, the SINDSCAL solution has also provided information concerning the important subjective aspects of the electrode distribution. The electrode site at which P300 is maximally positive appears to be perceived differently from the other two electrode sites.

Univariate Ratings

In addition to making the 153 pairwise comparisons, subjects also rated each of the 18 sets of ERPs on eight univariate scales. These scales included: P300 area, N200 amplitude, P300 electrode distribution, Classic P300, P300 latency, P300 amplitude, noise, and P300 duration. The average ratings for the 10 subjects can be found in Table 1.

Table 1

Average ratings on each of eight univariate scales

	Average Ratings[b]							
Stimuli[a]	P300 Area	N200 Amplitude	Electrode Distribution	Classic P300	P300 Latency	P300 Amplitude	Noise	P300 Duration
1	3.36	4.82	5.64	4.91	5.55	3.18	3.18	4.09
2	2.73	3.45	6.00	5.55	1.45	2.91	3.55	3.45
3	5.91	4.18	6.27	5.36	5.45	6.27	3.09	4.73
4	3.09	2.55	2.64	2.55	1.27	3.00	2.36	3.91
5	5.73	4.18	4.64	3.91	5.18	5.27	3.09	4.82
6	2.64	4.36	5.36	5.55	1.55	3.09	4.09	2.73
7	4.82	4.91	4.55	4.36	2.91	5.36	3.55	3.45
8	4.55	1.09	2.36	2.73	1.64	5.00	2.64	3.45
9	5.09	4.55	2.55	2.00	5.36	5.36	2.91	5.55
10	3.00	3.09	5.18	3.91	5.09	2.73	3.64	3.55
11	4.82	4.91	6.27	6.00	3.09	5.36	2.73	3.45
12	3.18	5.45	6.09	5.55	3.18	3.18	2.64	3.55
13	4.64	5.18	2.45	2.64	3.18	5.36	3.64	3.73
14	2.18	4.82	1.82	1.73	5.18	2.55	4.09	3.82
15	4.00	3.36	4.27	4.18	1.36	5.91	3.45	2.82
16	2.36	5.09	3.00	3.00	2.91	2.18	3.09	3.09
17	5.00	1.45	6.18	6.09	1.45	5.82	2.45	3.00
18	3.09	5.09	4.55	4.18	3.27	3.27	2.64	2.91

[a]Stimulus numbers refer to the ERP numbers in Figures 1, 2, and 3.
[b]P300 amplitude, P300 area, N200 amplitude, and noise were rated on a scale of small (1) to large (7). P300 latency and P300 duration were rated on a scale of short (1) to long (7). P300 electrode distribution and Classic P300 were rated on a scale of no (1) to yes (7).

The univariate scales were used for two purposes. First, the scales were correlated with the physical changes in the ERPs. This served to provide a measure of the correspondence between the physical changes in the ERPs and the subjects' perception of these changes. Pearson correlation coefficients were computed between the simulated attributes of the waveforms (P300 amplitude, P300 latency, and P300 electrode distribution) and the eight univariate scales.

The physical amplitude of the P300 component was significantly correlated with ratings of both P300 amplitude (.97) and P300 area (.91). These high correlations indicate that subjects were successful at detecting and systematically representing the physically simulated changes in the amplitude of the P300 component. Furthermore, the high correlation between P300 area and physical amplitude suggests that subjects realized that area and amplitude covaried across stimuli.

The physical latency of the P300 component was significantly correlated with ratings of P300 latency (.99), P300 duration (.68), and N200 amplitude (.51). Although both P300 duration and P300 latency were significantly correlated with the physical latency of the P300, the relatively higher P300 latency correlation suggests that latency represents a more salient cue for this group of subjects. The significant correlation between N200 amplitude and the physical latency of the P300 component is interesting since N200 amplitude was not varied across stimuli. However, comparison of N200 amplitudes across the three latencies (see Figures 1, 2, and 3) reveals that N200 amplitude does in fact appear to increase with increasing P300 latency. Although the input values of N200 amplitude were constant across ERPs, the overlap of P300 with N200 at the shorter P300 latencies produced an attenuation of N200 amplitude. Thus in this case subjects have accurately perceived an artifactual change in the ERPs.

Since electrode distribution is represented by a rank ordering of the three electrodes (F_Z, C_Z, and P_Z) both Pearson and Spearman correlation coefficients were computed. Both of these techniques yielded approximately the same results, so only the Pearson correlation coefficients will be presented. The physical differences in electrode distribution were significantly correlated with ratings of P300 electrode distribution (.94) and Classic P300 (.93). The high correlation between Classic P300 and the physically simulated electrode distributions suggests that subjects in the present study place a great deal of importance on electrode distribution in defining ERP components.

In addition to providing information about the correspondence between the physical changes in the simulated ERPs and the subjects' perceptions of them, the ratings obtained from the univariate scales can also be used to help interpret the MDS dimensions. This is achieved by regressing the MDS coordinates (the weights for the 18 ERPs on the three dimensions) on each of the univariate scales. Thus, in the present case, eight separate multiple regression equations were calculated. The three standardized weights (corresponding to the three MDS dimensions) and the multiple correlation for each scale can be used in the interpretation of the MDS dimensions (see Table 2). For example, the two highest regression weights on dimension 1 correspond to the P300 latency and P300 duration scales (.95 and .72, respectively). Dimension 1 was defined as P300 latency in the interpretation of the SINDSCAL solution. Given that the multiple correlation for the scale is high and statistically significant, the high regression weight indicates a close correspondence between the univariate scale and the MDS dimension. Thus, the regression of the MDS dimensions on the univariate scales provides converging evidence for the interpretation of dimension 1 as a P300 latency dimension.

The two highest weights on dimension 2 occur for the P300 electrode distribution and Classic P300 scales (.85 and .75, respectively). Dimension 2 was defined as the P300 electrode distribution. The relatively high weight on the Classic P300 scale suggests that the subjects relied heavily on electrode distribution in defining an ideal P300 component. For dimension 3, the two highest weights are from the P300 amplitude and P300 area scales (−.97 and −.93, respectively). The high weights for both P300 amplitude and P300 area are consistent with the correlation between these properties in the simulated ERPs.

The correspondence between the univariate scales and the MDS dimensions may also be rep-

Table 2

Regression weights for predicting the mean ratings on the eight univariate scales from the three SINDSCAL dimensions

Univariate Scale	Regression Weights: Dimension 1	Dimension 2	Dimension 3	Multiple Correlation
P300 Area	.11	.13	−.93	.95*
N200 Amplitude	.27	.12	.23	.36
P300 Amplitude	−.08	.12	−.97	.98*
Noise	.26	.08	.19	.33
P300 Latency	.95	.05	−.04	.95*
P300 Duration	.72	−.25	−.32	.85*
Classic P300	−.29	.75	−.01	.84*
P300 Electrode Distribution	−.08	.85	.01	.85*

*$F(3/14)$ significant at less than .001 level.

resented geometrically by using the regression weights to plot univariate property vectors in the MDS space. The length of the property vector represents a measure of the importance of the univariate scale in the MDS space, while the direction of the property vector corresponds to the relationship between the MDS and univariate scale values. Figure 7 illustrates the correspondence between the univariate scales and the MDS dimensions. The P300 latency property vector lies close to dimension 1, the P300 latency dimension. The P300 electrode distribution property vector corresponds closely to dimension 2, the P300 electrode distribution dimension.

The regression of the three MDS dimensions on each of the eight univariate scales, as well as the correlation of the univariate scales with the physical changes in the simulated ERPs, have provided converging support for the graphic interpretation of the MDS dimensions. The results suggest that subjects in the present study were in fact capable of recovering the actual physical dimensions of the stimuli. Furthermore, subjects were aware of the covariation among the physically manipulated aspects of the waveforms and other attributes (e.g. P300 amplitude and P300 area, P300 latency and N200 amplitude). Therefore, the answer to our first question is an unequivocal yes. The ERP researchers employed as subjects are capable of detecting and identifying systematic variations in simulated waveforms.

Although the ERP researchers were able to accurately and reliably recover the underlying dimensions of the set of simulated ERPs, it would be unwise to overgeneralize these results. The ERPs employed in this study were is some sense ideal because they possessed a relatively high signal to noise ratio and a small amount of component overlap. Furthermore, the component structure of the ERPs is not always as simple as that represented in the present study. For example, the P300 component is sometimes composed of a series of subcomponents which are difficult to distinguish (e.g. P3a, P3b, and slow wave). Additional studies are necessary to evaluate the accuracy and reliability of investigators' judgments with more complex waveforms.

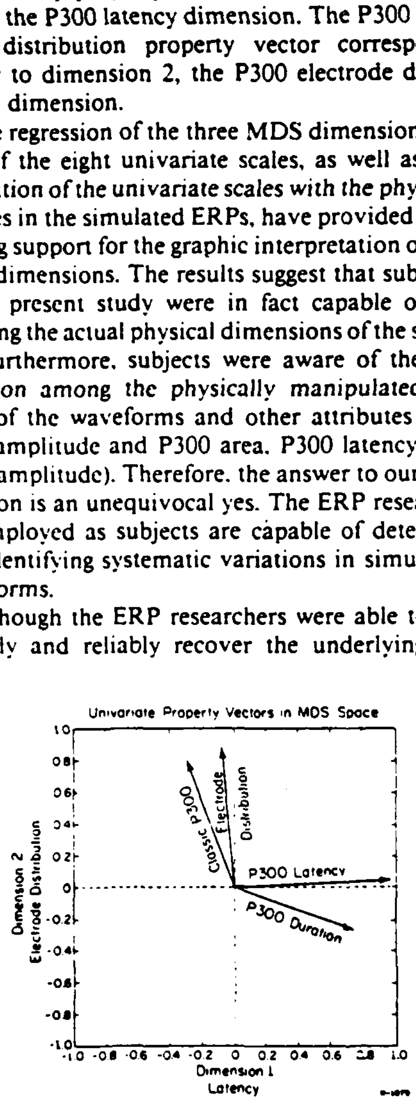


Figure 7. A graphic illustration of the univariate property vectors plotted in two-dimensionsal MDS space. The x axis represents the first dimension while the y axis shows the second dimension. The length of each vector represents the goodness of fit in the space; longer vectors indicate a good fit. The direction of each vector corresponds to the relationship between the univariate and MDS ratings.

Subject Weights

The SINDSCAL procedure yields both an aggregate solution which identifies the dimensional structure of the similarity judgments and a set of single subject weights for each of the dimensions. The magnitude of the weights provide information concerning individual subjects' emphasis of specific dimensions. A high weight indicates a salient dimension while a low weight suggests that a particular dimension did not contribute substantially to the subject's judgments.

As can be seen from Table 3, the weights assigned to the three dimensions vary across subjects. This suggests a great deal of variability across subjects in the use of the three dimensions. Aspects of subjects' weights which can be noted by visual inspection of the table include: (1) The most salient dimension appears to be dimension 1 (P300 latency). Five subjects have their highest weight on this dimension. (2) Electrode distribution (dimension 2) seems to be the next most important dimension with three subjects having their highest weight on this dimension. Only one subject placed the most

Table 3

Subject weight matrix, difference measures, and years of ERP experience

Subject Weights				
Dimension One	Dimension Two	Dimension Three	Difference Measures	Years of ERP Experience
.257	.725	.302	.936	2
.635	.479	.311	.648	3
.463	.386	.584	.396	4
.586	.605	.394	.422	2
.552	.444	.278	.548	3
.836	.103	.147	1.466	1
.509	.681	.210	.942	1
.797	.276	.348	1.047	1
.723	.271	.252	.942	1
.515	.491	.430	.170	10

importance on the third dimension (P300 amplitude). The rank ordering of subjects' highest weight is consistent with the rank order of the SINDSCAL dimensions. (3) The magnitude of the differences among the three weights varies across the 10 subjects. The difference measure was calculated by subtracting subject weights 1−2, 1−3, and 2−3, and summing the differences. As can be seen from the table, the magnitude of differences among the weighting coefficients for each of the dimensions appears to decrease with increased ERP experience. The Pearson correlation coefficient between the difference measure and ERP experience was .74, indicating that the experience in analysis and interpretation of ERPs accounts for 55% of the variance in the differential weights of the dimensions. This suggests that, with experience, ERP investigators tend to weight equally each of the important aspects of ERP components. Thus each of the dimensions becomes equally important in defining an ERP component.

As mentioned previously, different attributes of dimensions seem to be differentially weighted. For example, the most positive electrode in the P300 electrode distribution appears to be more important to the subjects than the other two electrodes. Thus, although the attributes or levels of a dimension may be differentially weighted, the different dimensions appear to receive equal weight in experienced subjects' similarity judgments.

Summary

The present study has demonstrated that ERP investigators are capable of accurately recovering the underlying dimensions of a set of simulated ERPs. Furthermore, experience in analyzing and interpreting ERPs appears to lead to an equal weighting of each of the underlying dimensions. These findings support the commonly held, but previously untested, belief that visual inspection of ERP waveforms can be performed accurately and reliably. In addition, the findings suggest that with experience, ERP researchers are able to attend equally well to several dimensions of a "component." However, it is important to note that the simulated ERPs in the present study are not representative of the entire set of waveforms that an ERP investigator normally encounters. Additional studies are needed to assess investigators' ability to recover underlying dimensions of more structurally complex ERPs.

REFERENCES

Carroll, J.D., & Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckhart-Young Decomposition." *Psychometrika, 35*, 283–319.

Carroll, J.D., & Wish, M. (1974). Models and methods for three-way multidimensional scaling. In D.H. Kranz, R.C. Atkinson, R.D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (pp. 57–105). San Francisco: W.H. Freeman.

Coles, M.G.H., Gratton, G., Kramer, A., & Miller, G. (in press). Principles of signal acquisition and analysis. In M.G.H. Coles, E. Donchin, & S.W. Porges (Eds.), *Psychophysiology: Systems, processes and applications.* New York: Guilford Press.

Courchesne, E. (1978). Changes in P3 waves with event repetition: Long term effects on scalp distribution and amplitude. *Electroencephalography & Clinical Neurophysiology, 45*, 754–766.

Donchin, E. (1975). Brain electrical correlates of pattern recognition. In G.F. Inbar (Ed.), *Signal analysis and pattern recognition in biomedical engineering* (pp. 199–218). New York: John Wiley.

Donchin, E. (1981). Surprise! . . . Surprise? *Psychophysiology, 18*, 493–513.

Donchin, E., & Heffley, E. (1978). Multivariate analysis of event related potential data: A tutorial. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related potential research* (pp. 555–572). EPA 600/9-77-043, Washington, DC: U.S. Government Printing Office.

Donchin, E., Kramer, A., & Wickens, C. (1982). Probing the cognitive infrastructure with event-related brain potentials. *Proceedings of the AIAA Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics* (pp. 371–387). Washington, DC: U.S. Government Printing Office.

Donchin, E., Ritter, W., & McCallum, W.C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. Koslow (Eds.), *Event related brain potentials in man* (pp. 349–412). New York: Academic Press.

Goff, W.R., Allison, T., & Vaughan, H.G. (1978). The functional neuroanatomy of event related potentials. In E. Callaway, P. Tueting, & S. Koslow (Eds.), *Event related brain potentials in man* (pp. 1–80). New York: Academic Press.

John, E.R., Ruchkin, D.S., & Vidal, J.J. (1978). Measurement of event related potentials. In E. Callaway, P. Tueting, & S. Koslow (Eds.), *Event related brain potentials in man* (pp. 93–138). New York: Academic Press.

Kruskal, J.B., & Wish, M. (1978). *Multidimensional scaling.* Beverly Hills, CA: Sage Publications.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science, 211*, 77–80.

Nahvi, M.J., Woody, C.D., Ungar, R., & Sharafat, A.R. (1975). Detection of neuroelectric signals for multiple data channels by optimum linear filter method. *Electroencephalography & Clinical Neurophysiology, 38*, 191–198.

Pritchard, W.S. (1981). Psychophysiology of P300. *Psychological Bulletin, 89*, 506–540.

Pruzansky, S. (1975). *SINDSCAL: A computer program for individual differences in multidimensional scaling.* Unpublished manuscript, Bell Laboratories.

Regan, D. (1972). *Evoked potentials in psychology, sensory physiology and clinical medicine.* London: Chapman Hall.

Sutton, S., & Ruchkin, D.S. (1984). The late positive complex: Advances and new problems. In R. Karrer, J. Cohen, & P. Tueting (Eds.), *Brain and information: Event-related potentials* (pp. 1–23). New York: New York Academy of Sciences.

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

(Manuscript received October 3, 1984; accepted for publication January 18, 1985)

In press,
Journal of Experimental Psychology: Learning, Memory and Cognition

Brain Potentials as Indices of Orthographic and Phonological Interaction during Word Matching

Arthur F. Kramer and Emanuel Donchin

Cognitive Psychophysiology Laboratory
University of Illinois
Champaign, Illinois

Acknowledgment

The experiment was supported by AFOSR contract number F49620-79-C-0233 monitored by Dr. Al Fregly. We gratefully acknowledge the helpful comments of Monica Fabiani, Amir Mane, John Polich and Mick Rugg. Requests for reprints should be sent to Arthur F. Kramer, Department of Psychology, 603 East Daniel Street, University of Illinois, Champaign, Illinois, 61820.

Running Head: ERPs and Word Matching

Abstract

The interaction between orthographic and phonological codes in a same-different judgment task was studied by requiring subjects to decide if two visually presented words either looked alike or rhymed. Word pairs were selected from four different lists. Words either rhymed and looked alike, rhymed but did not look alike, looked alike but did not rhyme, or neither looked alike or rhymed. Reaction time and percent error increased whenever there was a conflict between the orthography and phonology of the words. The N200 component of the Event-Related Brain potential (ERP) indicated that subjects' were capable of detecting phonological differences between words within 160 msec from the presentation of a word pair. The amplitude of the N200s also varied with the degree of mismatch between words. N200s were largest when both the orthography and phonology mismatched, of intermediate amplitude when either orthography or phonology mismatched, and smallest when both orthography and phonology matched. P300 latency was consistent with reaction time, increasing whenever there was a conflict between the two codes. Taken together, behavioral measures and the ERP data suggest that the extraction of the orthographic and phonological aspects of words occurs early in the information processing sequence.

Descriptors: Event-related brain potentials (ERP), P300, N200, orthography, phonology, word matching, code interaction

Brain Potentials as Indices of Orthographic and Phonological Interaction during Word Matching

Arthur F. Kramer and Emanuel Donchin

When different aspects of a visual display call for conflicting responses, reaction times (RTs) are usually increased relative to arrangements in which response conflicts do not occur. For example, in the standard Stroop paradigm subjects are instructed to name the ink color in which a word is printed (Stroop, 1935). Three different categories are employed. In the congruent condition both the word and the ink color in which the word is printed are the same (the word "blue" is displayed in blue ink). In the incongruent condition the ink color and the word do not match (the word "red" is printed in yellow) while in the neutral condition a non-color word is presented (the word "house" is shown in green ink). More time is required to name the ink color when the word and the ink color conflict. Thus, although subjects are instructed to respond to the ink color and ignore the meaning of the word, it appears that codes for both properties are activated. Eriksen and colleagues (Eriksen & Eriksen, 1974; Eriksen & Shultz, 1979) have observed a similar interference effect in a paradigm in which subjects were required to respond to a target letter in an array of non-target letters. When the flanking non-target letters called for a conflicting response, RT to the target was lengthened.

In the present study we examine the temporal characteristics of another interference effect which is observed when subjects are instructed to selectively process either the orthographic or phonological aspects of words (LaBerge, 1972; LaBerge & Sammuels, 1974; Seidenberg & Tanenhaus, 1979). In a study which examined the interaction between orthographic and phonological codes, subjects were presented with a target word which was followed by three semantically unrelated monosyllable words (Nolan, Tanenhaus & Seidenberg, 1981). The subjects task was to detect the word which rhymed with the target. Target stimuli were presented in both the auditory and visual modalities, while the words to be monitored were spoken. The results indicated that for words which did not rhyme, RTs to orthographically similar pairs were significantly longer than the RT to the orthographically dissimilar pairs. Thus, although the orthographic information was not required to perform the rhyme monitoring task, subjects were unable to ignore it. The investigators suggested that the activation of one code also served to activate the other code in a manner similar to that proposed by Morton's (1969) logogen model or the spreading activation hypothesis (Collins & Loftus, 1975).

Polich, McCarthy, Wang and Donchin (1983) obtained similar results employing a visual matching task. The subjects' task was to decide if two words either looked alike (orthographic match) or rhymed (phonological match). The stimulus lists were composed of words which rhymed and looked alike, rhymed but did not look alike, looked alike but did not rhyme and neither looked alike or rhymed. When subjects were to decide if the words rhymed, the RT was longest whenever there was a conflict between the orthographic and phonological codes. The orthographically similar but phonologically dissimilar words (COUGH-DOUGH) produced the longest RT. In the visual matching conditions, the orthographically dissimilar but phonologically similar word pairs yielded the longest RT (MOOSE-JUICE). Thus, both the orthographic and phonological information affected subjects' performance, although one of the two codes was irrelevant in each of the matching tasks. The results support the proposition that multiple codes may be invoked and processed, perhaps to the action level in parallel. The data also indicate that task demands influence how the two codes interact to affect performance.

Although each of the interference effects described above have proven to be very robust, the determination of the locus at which the relevant and irrelevant codes interact has been elusive. For instance, the prolonged RT in the incongruent Stroop condition has been attributed to interference between codes during both stimulus encoding (Hock & Egeth, 1970; Teece & Dimartino, 1965) and response selection and execution (Keele, 1972; Posner & Synder, 1975). Similarily, the interference between the targets and non-targets in the Eriksen paradigm has been attributed to competition at the recognition stage (Proctor, 1981) as well as to interactions at the response level (Eriksen, Coles, Morris & O'Hara, 1985).

An obstacle in discovering the locus of these interactions has been the difficulty of decomposing the human information processing sequence. Whether one adopts a serial, or a parallel model of the information processing system there is no doubt that many structures are involved, and many different processes are activated between the instant a stimulus is presented and the emission of a response. Yet, a mental chronometry which is based on RTs is locked into an analysis of the final outcome of this multiplicity of interacting processes. Much of contemporary mental chronometry consists of attempts to overcome this limitation by developing indirect methods for decomposition (Posner, 1978; Sternberg, 1969). Clearly, it would be useful if a metric could be developed which is sensitive to the duration of a specific subset of the processes reflected by our traditional measures of motor output. One class of techniques which have proven to be increasingly useful in the measurement of mental processes is the event-related brain potential (ERP). In the present study we employ components of the ERP in conjunction with RT and accuracy measures to increase the resolution with which we can measure the interaction between orthographic and phonological codes in a word matching task.

Event Related Brain Potentials and Information Processing

The ERP is a transient series of voltage oscilliations in the brain that can be recorded from the scalp in response to the occurrence of a discrete event (Donchin, 1975; Regan, 1972). The ERPs are viewed as a sequence of separate commonents commonly labeled with an "N" or "P" denoting polarity, and a number which indicates the minimal latency measured from the onset of the eliciting event (e.g. N200 is a negative going component which occurs at least 200 msec after a stimulus). The components of the ERP are customarily divided into two categories. Exogenous components represent an obligatory response of the brain to the presentation of a stimulus. These components are primarily sensitive to physical attributes of the stimuli such as intensity, modality and rate of presentation. The latency and amplitude of the Endogenous components are influenced by the processing demands of a task rather than the physical characteristics of the stimuli. The use of these ERP components in conjunction with more traditional measures such as RT and percent correct has lead to a more complete understanding of several cognitive processes (Donchin et al, 1978; 1983; Duncan-Johnson, 1981; Picton et al, 1978; Pritchard, 1981). Two endogenous components, the N200 and P300, will be used in the present study to augment the chronometric information provided by measures of RT.

The peak latency of the P300 component is influenced by the duration of stimulus evaluation processes while it is relatively unaffected by the duration of the processes of response selection and execution (Maligero, Bashore, Coles & Donchin, 1984; Squires, Donchin, Herning & McCarthy, 1977).

McCarthy and Donchin (1981) orthogonally manipulated two independent variables in an additive factors design. One factor, stimulus discriminability, has been shown to affect an early encoding stage of processing while the second factor, stimulus-response compatibility, influences the later stages of response selection and execution. The subjects' task was to decide which of two target stimuli, the words RIGHT or LEFT, were embedded in a matrix of characters on a CRT. Stimulus discriminability was manipulated by varying the amount of noise in the matrix. Stimulus-response compatibility was manipulated by requiring subjects to respond to the target stimulus with the compatible or incompatible hand (i.e. compatible - respond to the word RIGHT with the right hand). The RT increased additively when the target word was embedded in noise and when the response was incompatible with the stimulus. P300 latency was prolonged by the addition of the noise to the target matrix, but was not affected by the incompatibility between the stimulus and response. Thus, the results support the conclusion that P300 latency is affected by a subset of processes which affect RT. Additional evidence of P300's sensitivity to stimulus evaluation processes has been obtained in varied mapping visual search paradigms. In these studies, both RT and P300 latency increase montonically with increasing memory load (Adam & Collins, 1978; Brookhuis, Mulder, Mulder & Gloerich, 1983; Ford, Mohs, Pfefferbaum & Kopell, 1979; Ford, Pfefferbaum & Kopell, 1980; Gomer, Spicuzza & O'Donnell, 1976).

Another component of the ERP, the N200, has been found to be a reliable indicator of stimulus mismatch (Naatanen, Simpson & Loveless, 1982; Naatanen & Gaillard, 1983). The N200 appears to be an automatic response to stimulus mismatch since it occurs regardless of the focus of subjects attention

(Naatanen, Gaillard & Varey, 1981; Squires et al., 1977). N200's have been elicited by mismatches in visual, auditory and somatosensory modalities in a diverse set of experimental paradigms including selective attention tasks, omitted stimulus paradigms, lexical decision tasks and word matching paradigms (Ford, Pfefferbaum & Kopell, 1982; Ford, Roth & Kopell, 1976; Klinke, Fruhstorter & Finkenzeller, 1968; Kramer, Ross & Donchin, 1982). Although the N200 has been described as an index of physical stimulus mismatch, N200's have also been elicited by orthographic, phonological and semantic mismatches (Sandquist et al., 1980). Furthermore, the amplitude of the N200 appears to be systematically related to the magnitude of stimulus mismatch, with larger amplitude N200's being elicited by more deviant stimuli (Ford et al., 1976; Kramer et al., 1982).

Several recent studies have illustrated the advantages of the joint use of traditional performance measures and ERPs in the study of the mental chronometry of code interactions. In one such study, subjects were instructed to name the color of the ink in which a word was printed (Duncan-Johnson & Kopell, 1981). However, unlike most Stroop tasks the RT measure was augmented by measuring the latency of components of the ERP. The investigators reasoned that since the latency of the P300 component is sensitive to factors which influence stimulus evaluation time but is relatively unaffected by response selection and execution processes the pattern of P300 latencies and RT could be used to discriminate between stimulus encoding and response selection interpretations of the Stroop effect. For instance, a pattern of results in which P300s elicited in the incongruent condition were prolonged relative to the congruent condition would suggest that at least some portion of the code interaction took place prior to response selection and execution. On the other hand, if P300 latencies did not differ in the two conditions while RT was lengthened in the incongruent condition support would be provided for the response selection hypothesis. The results were consistent with the former interpretation. The RTs were prolonged in the incongruent condition relative to the congruent condition while P300 latency was unaffected by the relationship between the word and the color of the ink in which it was printed.

A similar approach was pursued by Coles et al. (1985) to study the interaction between target and non-target stimuli in the Eriksen paradigm. In this study measures of RT and accuracy were augmented by measures of the electromyogram (EMG) as well as by the latency of two components of the ERP, the P300 and the readiness potential (RP). The results obtained from this multivariate approach to the study of code interactions provided support for the proposal that the noise/compatibility manipulation influenced both stimulus evaluation and response competition processes. Furthermore, the finding of a gradual build-up of response competition was consistent with a continuous flow model of human information processing (Eriksen & Schultz, 1979). The results of the Polich et al. (1983) investigation of orthographic and phonological code interactions in a word matching task also demonstrated the usefulness of the joint use of behavioral and psychophysiological measures. However, while the Polich et al. study provided initial insights, there appeared considerable differences between individual subjects on the pattern of interactions between tasks, as reflected by RT and ERP measures.

The present study was designed as a comprehensive investigation of the effects that interaction between orthographic and phonological codes may have on both overt responses and ERPs. It was our assumption that the N200 component would provide an indication of the earilest time at which mismatches were detected between orthographic and phonological codes. Furthermore, we predicted that the amplitude of the N200 would be sensitive to the degree of mismatch as well as the task relevance of the two codes. We also expected that the P300 component, which is a reliable measure of the duration of stimulus evaluation processes, should aid in defining the locus of the orthographic and phonological code interaction. A pattern of increases in the latency P300 whenever there is a conflict between the orthography and phonology of the words would lend support to the proposal that the interactions between the codes take effect early in the processing sequence.

Method

Subjects

Forty subjects (20 male) participated in the study. They were selected from a population of 400 undergraduate students on the basis of their scores on the Nelson-Denny Reading test (reading comprehension and vocabulary). Half the subjects were selected from the upper 20th percentile and the other half from the lower 50th percentile. The effect of reading ability on performance in the match task has been reported elsewhere (Kramer, Ross & Donchin, 1982) and will not be dealt with further in the current paper. All subjects were right handed, native speakers of English. All subjects had either normal or corrected-to-normal vision.

Stimuli and Apparatus

The pairs of words employed as experimental stimuli were those used by Polich et al. (1983) and are presented in Appendix A. Each word pair is abbreviated such that the phonological and/or orthographic similarity of the constituent stimulus is indicated.

RO: Rhymed and looked alike (MATCH-PATCH)

RX: Rhymed but did not look alike (BLARE-STAIR)

OX: Looked alike but did not rhyme (CATCH - WATCH)

XX: Neither looked alike or rhymed (SHIRT - WITCH)

We label these pairs with an R to indicate a rhyme match and an O to indicate an orthographic match. The absence of a rhyme or orthographic match is indicated by an X. Each of the words was a single syllable of either four or five letters. Words were considered to be orthographically similar (RO & OX lists) if all letters but the first matched. Word frequency was highly variable within word lists ranging from 1 to 10,000 occurrences per million (Kucera & Francis, 1967). The mean word frequency (standard deviations in parentheses) for each of the lists was: RO - 196 (332), RX - 68 (137), OX - 313 (662), XX - 313 (1037). Although the difference in word frequency among the four stimulus lists was significant, Polich et al. found that manual RT in discriminating the stimulus words from a set of unrelated pseudowords as well as vocal RT for pronouncing the words did not differ significantly among the lists. Stimulus words were displayed on a Hewlett Packard CRT which was positioned approximately 70 cm from the subjects. Words subtended a visual angle of between 1.9 and 2.3 degrees in length and 0.5 degrees in height.

ERP Recording

The EEG was recorded from three midline sites (Fz, Cz, Pz according to the International 10-20 system; Jasper, 1958) and referred to linked mastoids. Two ground electrodes were positioned on the left side of the forehead. Burden Ag-AgCl electrodes, affixed with collodion, were used for scalp and mastoid recording. Beckman Bipotential electrodes were affixed with adhesive collars and placed laterally and supra-orbitally to the right eye to record EOG and for the ground recording. Electrode impedances did not exceed 5 kohms.

The EEG and EOG were amplified with Van Gogh Model 50000 amplifers (time constant 10 sec and upper half amplitude of 35 Hz). Both EEG and EOG were sampled for 2560 msec, begining 100 msec prior to stimulus onset (presentation of the first word). The data were digitized every 10 msec. ERPs were digitally filitered off-line (-3dB at 8.8 Hz; 0 dB at 20 Hz) prior to statistical analysis.

Stimulus Generation and Data Collection

Stimulus presentation and data acquisition were governed by a PDP 11/40 computer interfaced with an Imlac graphics processor (see Donchin & Heffley, 1975). Single trial and average EEG and EOG were monitored on-line using a GT-44 display. Digitized single trial data were stored on magnetic tape for later analysis. Evaluation of each EOG record for saccades and blinks was conducted off-line by calculating its variance and comparing this to a preset criterion for acceptance. Single trial EEG containing unacceptable EOG was discarded prior to statistical analysis.

Procedure

The subjects performed a word comparison task. In one condition they had to decide whether two visually presented words looked alike (orthographic match). In the other condition they had to decide if the words rhymed (phonological match). Subjects depressed one response button with the thumb of one hand if the words were the same and another with the thumb of the other hand if they were different. The buttons assigned to the different responses were counterbalanced across subjects. The subjects were given a maximum of 2000 msec to indicate their response. Instructions emphasized both speed and accuracy.

Each subject received all of the 200 word pairs in each condition (50 per list -- see Appendix A). The order of the word pairs was reversed between match conditions with the initial presentation order counterbalanced across subjects. Subjects were unable to predict which items would be presented on any trial since word pairs occured randomly and with equal probability. Prior to the presentation of the experimental blocks, subjects received 40 practice trials in each of the match conditions. The word lists employed during the practice trials were composed of word pairs not included in the experimental stimulus sets (ten words per list).

The temporal sequence of each trial was as follows: A fixation dot appeared on the CRT and was displayed for 1000 msec. The first word of the pair followed the presentation of the fixation dot. The word was displayed for 200 msec and was followed by a blank screen for 600 msec. The second word of the pair was then displayed for 200 msec. The next trial followed in four to six secs. Each subject was presented with the orthographic and phonological matching conditions and all four word pair lists (RO, RX, OX, XX).

Results

Reaction Time and Error Data

RT was defined as the interval between the onset of the second word in the pair and the subject's key-press indicating whether this stimulus matched the first stimulus. In the Rhyme Match condition the answer was "yes" for the RO and RX lists; in the Visual Match condition the answer was "yes" for the RO and OX stimuli. The average RT for correct responses and the average error rate, taken over all subjects for each condition are presented in Figure 1. A numerical comparison of the mean RT and error rate is presented in Table 1.

Insert Table 1 and Figure 1 About Here

The subjects responded faster when they were assessing the visual similarity of the word pairs than they did when deciding if the words rhymed (680 vs. 824 msec; $F(1,36)=116.1$, MSE=13,750, $p<.001$). The RTs also differed significantly across lists ($F(3,108)=123.5$, MSE=3,234, $p<.001$). More important, the pattern of differences among lists varied with the type of judgment task such that the interaction between matching condition and list types was significant ($F(3,108)=110.3$, MSE=2,080, $p<.001$).

The pattern of RT effects across word lists within each of the match conditions was assessed with planned comparisons. In the rhyme match condition, subjects responded "yes" more quickly than they responded "no" (774 vs. 874 msec; $F(1,36)=8.9$, $p<.01$). This is consistent with the general finding that "same" judgments require less time than "different" judgments (Nickerson, 1978; Proctor, 1981). Furthermore, to reject the OX words as rhyming took substantially longer: RT to the OX pair was longer than the RT to the other three word pairs ($p<.01$ for all comparisons). Items in the RX list were responded to more slowly then either the RO or XX pairs ($F(1,36)=14.2$, $p<.01$). The time required to make a rhyme match decision with words which were either both phonologically and orthographically similar (RO) or dissimilar (XX) was not significantly different. Thus, when the task was to detect rhymes, word pairs in which the orthographic and the phonological codes conflicted produced longer RTs than word pairs in which both codes either matched or mismatched. However, when the words matched orthographically but mismatched phonologically the RT was even further increased. Presumably the orthographic code interacts with the phonological decision even though the orthography of the word is irrelevant to the rhyme match task.

In the visual match condition, subjects judged the visual similarity of word pairs faster when the words matched than when the words mismatched ($F(1,36)=11.3$, $p<.01$). The mean RT for the "yes" response (RO & OX) was 657 msec while the "no" response was 703 msec (RX & XX). Unlike the rhyme condition, the mean RT for the RO list was significantly shorter than that for the XX list ($F(1,36)=4.7$, $p<.05$). Although the difference between the RO and XX lists was approximately the same in the rhyme and visual match conditions (32 & 40 msec, respectively), the within subject variability was substantially smaller in the latter case. There were no significant difference between the RTs elicited by the OX and XX pairs during visual matching (674 & 679 msec, respectively). However, the mean RT for the OX list was significantly longer than that for the RO list ($F(1,36)=4.2$, $p<.05$). The longest mean RT was elicited by the RX word pairs ($p<.05$ for all comparisons). Thus, the conflict between the orthography and phonology appears to retard the response in the visual match task, as it did in the rhyme task.

Figure 1(b) displays the error rates for the four word lists in the two match conditions. The mean error rate across the four word lists and two match conditions was 4.6%. Subjects made significantly more errors in the rhyme (6.8 vs 2.4%) than in the visual condition ($F(1,36)=56.2$, MSE=56.2, $p<.001$). The word lists also differed significantly in error rates

($\underline{F}$(3,108)=44.6, MSE=31.2, $\underline{p}$<.001). A significant interaction between match condition and word list was also obtained such that higher error rates were associated with lists that elicited longer RTs ($\underline{F}$(3,108)=56.3, MSE=29.4, $\underline{p}$<.001). These data suggest that subjects were not trading speed for accuracy when performing the tasks.

It is clear that there is a task dependent increase in RT whenever the orthography of the words call for one response and the phonology calls for another. Moreover, the extra time taken for the overt responses to the word pairs with mismatching codes is not <u>necessary</u> for reading and comparing the words since this is exactly what the subject accomplishes for the RO and XX word pairs in some 800 msec. Why then is the response delayed? Where is the time spent? How early is the information about the mismatch available to the subject? Are any other aspects of the processing delayed by the conflict? It is for answers to these questions that we turn to an examination of the ERPs.

<u>Event-Related Potentials</u>

Figure 2 presents the average parietal ERPs elicited by the word pairs for each of the lists in both of the match conditions. The average ERPs were composed of single trial ERPs elicited by word pairs which were responded to correctly.

Insert Figure 2 About Here

It is clear that the course of the ERPs elicited by the first word in a pair is the same regardless of the word pair or the task. Indeed, this segment of the ERPs confirms one of the main assumptions of our experimental design, namely that the subjects could not know which class of pairs appeared on a given trial until the second word appeared. 1 The consistency of the ERP pattern following S1 gives way to a diversity of patterns across the four lists in the part of the epoch following S2. The four curves diverge mostly in two segments of the epoch. It is quite evident that a negative going peak with a latency of about 350 msec is elicited with different amplitudes by the different word pairs. This pattern also varies with the subjects task. The negative variation is followed by the obvious variance in a positive going component whose latency ranges from 450 to 700 msec across the different word pairs. These ERP patterns can be used to support the construction of a more detailed model of the information processing activities elicited by the stimuli than is available from just the RT results (for a review of the assumptions that underly our use of ERPs in interpreting human information processing see Donchin, 1981; Donchin, Kramer & Wickens, 1982; Pritchard, 1981). The effects of the experimental variables on the ERP components was assessed with a Principal Components Analysis procedure (for a review of the use of PCA in studies of ERPs, see Donchin & Heffley, 1979).

Two separate PCAs were performed on the data. In the first procedure the components of the ERP were derived and analyzed by applying the PCA technique to the averaged raw data. The data base submitted to the PCA consisted of 960 ERPs (40 subjects x 2 match conditions x 4 lists x 3 electrodes), each composed of 256 time points. The PCA was performed on the covariance matrix obtained by computing the covariance between the voltages recorded at each pair of time points over the entire data set. The component scores computed from a linear combination of time points x loading coefficients were then subjected to a repeated measures ANOVA.

It is evident however, even in the data of Figure 2, that there is a high degree of latency variability in the P300 component across experimental conditions. It is difficult to interpret the results of a PCA when the data show much latency variability. A latency adjustment procedure (Woody, 1967) was therefore employed to create a data set in which P300 latency is less variable. The latency adjustment procedure uses a lagged cross correlation algorithm to detect and align the desired component so as to eliminate the latency jitter across single trials. This procedure helps reduce the confounding of changes in amplitude with variability in the latency of a specific ERP component. The latency adjustment procedure yields latency adjusted single-trial waveforms which can be averaged according to experimental conditions and submitted to the PCA. The latency adjustment procedure also determines component peak latencies for each subject in each condition. These latency values can then be analyzed much in the same manner as RT.

Thus, two separate PCA's were performed: one on unadjusted, the other on latency adjusted ERPs. The component scores obtained from the first PCA were used to assess the effect of experimental manipulations on ERP components other than the P300. The second PCA was employed to evaluate the effect of experimental factors on the P300.

Insert Figure 3 About Here

The component scores obtained from the PCAs were submitted to a repeated measures analysis of variance.

<u>Principal Components Analysis</u> The PCAs revealed the existence of several components in the data. Figure 3 presents the first four components derived from the PCA of the unadjusted average waveforms.2 Of primary interest are the components labeled "N200" and "P300" in Figure 3. These aspects of the variance were identified as N200 and P300 respectively because of their temporal relation to the stimulus as well as their scalp distribution. In reviewing the results it is useful to note that we were assessing the degree to which the subject's matching task affected the ERP, the degree to which different word pairs elicited different ERPs, and the extent to which these two factors interacted.

<u>N200 Amplitude</u> As portrayed both in the average waveforms in Figure 2, a negativity with a latency of about 200 msec is elicited prominently by the XX items in both the visual and rhyme matching task. A somewhat smaller N200 is elicited by the RX pairs in both conditions. Similarly, the OX items elicit a slightly smaller N200 in the rhyme match condition, though not in the visual match condition. These impressions are corroborated by the analysis of the component scores which are illustrated in figure 4. The N200 amplitude difference between lists is significant ($F(3,108)=44.3$, $MSE=.48$, $p<.001$). Moreover, there is a significant interaction between word list and match condition ($F(3,108)=7.8$, $MSE=.29$, $p<.001$).

Planned comparisons were performed on the N200 elicited by the different word lists in both the rhyme and visual match conditions. In the rhyme match condition, the XX word pairs list elicited a larger N200 than did words in the other three lists ($p<.05$ for all comparisons; see Figure 4). Significant differences were also obtained between the N200 elicited by the RO and RX pairs ($F(1,36)=6.6$, $p<.05$) and the RO and OX pairs ($F(1,36)=6.9$, $p<.05$). In the visual match condition, XX word pairs produced a larger N200 than words in the other three lists ($p<.01$ for all comparisons). However, the N200 amplitude did not differ significantly between the RO and OX lists but did differ between the RO and RX ($F(1,36)=6.4$, $p<.05$) and RX and XX ($F(1,36)=5.0$, $p<.05$) comparisons.

The pattern of N200 results suggests that when the words in a pair did not match on both the orthographic and the phonological dimensions (XX pairs) or when the words mismatched orthographically (RX pairs), an N200 was elicited regardless of the subject's task. The dual mismatch elicited a larger N200 than the single mismatch. The N200 elicited by the phonologically mismatching OX pairs appears to be task specific since it occurs only when subjects are performing the rhyme match task. Perhaps the most important aspect of these data is the finding that within 160 msec of the presentation of the stimulus pair, subjects have processed the orthographic and phonological characteristics of the words.

P300 Amplitude In contrast to the N200, the amplitude of the P300 component demonstrates a somewhat different pattern. The amplitude of P300 varied significantly with the word list ($F(3,108)=20.3$, $MSE=.42$, $p<.001$). As can also be observed in the waveforms of Figure 2, the largest P300 is elicited by the RX words. In addition larger P300s are elicited in the rhyme match condition relative to the visual match condition ($F(1,36)=5.8$, $MSE=1.78$, $p<.05$). These two effects interact significantly ($F(3,108)=3.2$, $MSE=.26$, $p<.05$) and primarily reflect the fact that the OX pairs elicit a relatively larger P300 during the rhyme match condition than they do in the visual match condition. This pattern of P300 component scores is illustrated in Figure 5.

Insert Figures 4 & 5 About Here

P300 Latency The pattern of the latencies of the P300 was quite similar to the pattern displayed by the RTs. Table 1 presents the mean P300 latencies with their respective standard deviations for the four word lists in both the rhyme and visual match conditions. The P300 was significantly longer in latency in the rhyme (586 vs 510 msec) than in the visual condition ($F(1,36)=67.5$, $MSE=54.8$, $p<.001$). The difference in average latency of the P300 between the rhyme and visual conditions is approximately one-half that of the difference between the corresponding RTs (76 vs 149 msec). It is generally the case that RT differences are twice the magnitude of the differences in P300 latency (Duncan-Johnson, 1981; Ford et al., 1981; Kutas et al., 1977; McCarthy & Donchin, 1981). The latency of the P300 appears to depend on the duration of a subset of the processes that determine RT such that processes which effect P300 latency do not include the processes involved in response selection and execution. Additional evidence indicates that P300 invocation must await the completion of stimulus categorization and identification. Within this framework it is important to note that the P300 latencies demonstrated the same pattern of results as provided by RT. Thus, the interaction between match condition and word list was significant ($F(3,108)=27.5$, $MSE=37.3$, $p<.001$).

Planned comparisons were performed on the mean P300 latency between conditions. As was observed for the RT data, subjects judged the phonological similarity of words more quickly for matches than they did for mismatches ($F(1,36)=7.6$, $p<.01$). The mean P300 latency for the "yes" response (RO & RX) was 549 msec while the mean laency for the no's (OX & XX) was 627 msec. For the rhyme match condition, P300 latency did not differ significantly between the RO and RX, RO and XX, or the RX and XX word lists but did differ between the RO and OX ($\underline{F}(1,36)=6.9$, $\underline{p}<.05$), RX and OX ($\underline{F}(1,36)=5.1$, $\underline{p}<.05$) and OX and XX lists ($\underline{F}(1,36)=4.6$, $\underline{p}<.05$). If one accepts the assertion that P300 latency is a measure of stimulus categorization time, then the relatively long latency of the P300 component elicited by the OX word pairs indicates that the phonological and orthographic codes interact and have an effect on the duration of stimulus categorization prior to the response selection and execution stage. Thus the P300 latency results obtained in the rhyme match condition provide additional information concerning the pattern of interactions of phonological and orthographic codes.

In the visual match condition, P300 latency was shorter for the "yes" than the "no" responses (481 vs. 540 msec; $\underline{F}(1,36)=5.3$, $\underline{p}<.05$). Statistically significant differences were obtained for the RO and RX ($\underline{F}(1,36)=4.7$, $\underline{p}<.05$) as well as the RX and OX comparisons ($\underline{F}(1,36)=4.8$, $\underline{p}<.05$). Although the P300 latency differences across lists was smaller in the visual than in the rhyme match condition (76 vs 147 msec), the results obtained in the visual condition also provide support for an interaction of phonological and orthographic codes prior to response selection.

Discussion

The analysis of the ERPs was undertaken with the hope that it would provide a clue to the substantial delay in the subject's response to word pairs in which the orthographic and phonologic codes call for conflicting responses. At least two alternate explanations are possible for the delay:

(1) It may be attributed to a tendency to withhold the response when such a conflict is discovered; (2) The processing of the items may be delayed by the conflicting codes. While the RT data does not allow discriminations between these positions the ERP results suggest that the code conflicts play a role at a fairly early stage, even before the N200 component is elicited. The clearest evidence for this assertion comes from a comparison of the RO and the OX patterns in the two conditions. Note that the RO words are pairs in which both the orthography and phonology match and that the subject responds affirmatively to these pairs in both conditions. The wave pattern of the ERP in both conditions is virtually identical and the latency is short in both conditions with but a trace of an N200.

When the subject is required to judge orthographic similarity in the visual match condition, the ERP elicited by the OX pairs is virtually identical to the ERP elicited by the RO pairs. Hence, the distinctly different ERP pattern elicited by the OX items in the rhyme conditions must be due to the additional processing requirement imposed by the task. The OX ERP for the rhyme condition differs from the RO ERP in two respects -- it is characterized by an N200 that is larger in amplitude than the N200 in the RO ERP. Moreover, the latency of the P300 is increased in the OX condition. If the N200 is a consequence of the detection of a mismatch (Naatanen, 1982), and if the latency of P300 is indicative of additional categorization time, then the special nature of the OX pair must be detected very early in the processing sequence. While there may well be additional delays before the overt response is emitted, there is clearly a very early perturbation of the processing activities by the conflict.

The difference between the RO and the OX as early as the N200 is especially remarkable considering the fact that it reflects the mismatch between the sounds associated with such character strings as "OUGH" when they appear in words such as COUGH and DOUGH. The remarkable feature is that in order to "know" that the sounds are different the system must have identified the two words including presumably their lexical sense. After all, there is no cue in the character string "ough" itself to the phonological mismatch. The cue is in the additional character ("C" and "D", making the COUGH and DOUGH, respectively). Thus, within 160 msec or so after the presentation of the second stimulus, by the point in time at which these ERP waveforms for the OX and RO diverge, the words as lexical entities have been adequately processed to determine the differential phonology.

The pattern of RT results obtained in the present study appears to be consistent with the dual-encoding hypothesis: regardless of the task instructions, subjects were unable to ignore the irrelevant information (LaBerge, 1972; Nolan et al., 1981; Polich et al., 1983; Seidenberg & Tanenhaus, 1979; Tanenhaus et al., 1980). Information conflict served to slow RT whenever there was a interaction between the orthography and phonology of the words. In the rhyme match condition words which were orthographically similar but phononologically dissimilar retarded RT by 240 msec when compared with words which both possessed the same codes. In the visual match condition words which were phonologically similar but orthographically dissimilar delayed RT by 87 msec when compared with the RO word pairs.

Similar to the pattern of RTs, P300 latency was increased whenever there was a conflict between the two codes. Since P300 latency has been shown to be sensitive to manipulations of factors which affect stimulus evaluation and categorization while being relatively insensitive to the manipulation of response selection and execution factors, it appears reasonable to assume that the orthographic and phonological code interaction occurs at least as early as the stimulus categorization stage of processing (Ford et al., 1980; Kutas et al., 1977; McCarthy & Donchin, 1981; Squires et al., 1977). Changes in the latency of the P300 component accounted for approximately 50% of the change in RT over the word lists. This is consistent in both the rhyme and visual match conditions. The difference between P300 latency and RT suggests that the interaction of phonological and orthographic codes began during the stimulus evaluation stage but that the interaction was magnified during later stages of processing perhaps reflecting the effects of a cascading process (McClelland, 1979).

An analysis of the N200 amplitude effects indicated that the orthographic mismatches produced N200s regardless of the task while phonological mismatches elicited N200s only when the phonology of the word was task relevant. The word lists in which both of the codes mismatched produced the largest N200s regardless of the match condition. The comparison among the N200s elicited by the XX, RX and RO word pairs is consistent with previous research which has obtained a high positive relationship between the degree of mismatch and the amplitude of the N200 component (Naatanen et al., 1980). This effect is evidenced by the large amplitude N200 elicited by words which both orthographically and phonologically mismatch, an intermediate size N200 elicited by words which orthographically mismatch and the absence of an N200 when words were both orthographically and phonologically similar.

A comparison of the N200 amplitudes elicited by the RX and OX word pairs in the rhyme and visual match conditions is noteworthy. The RX word pairs elicit an N200 regardless of match conditions while the OX word pairs elicit an N200 only in the rhyme match condition. It appears that the N200 elicited by orthographic mismatches is in some sense "automatic" since it does not depend on the subjects task. On the other hand, a phonological mismatch produces an N200 only when the phonology of the word pair is task relevant. Thus, the nature of the code appears to interact with its task relevance in eliciting the N200 component.

Figure Captions

Figure 1. Mean correct reaction time and percent error for the rhyme and visual match conditions and the four word lists. Each of the means represents an average of 40 subjects.

Figure 2. Grand average parietal ERPs for the four word lists and the two match conditions.

Figure 3. Component loadings for the first four components extracted in the Principal Components Analysis of the unadjusted ERP data.

Figure 4. N200 component scores for each of four word lists for both rhyme and visual match conditions.

Figure 5. P300 component scores for each of four word lists for both rhyme and visual match conditions.

Footnotes

1 Further support for our assumption that subjects could not predict the list type on the basis of the first word was obtained in a study in which we asked a group of subjects to predict the likelihood of another word either rhyming or looking like words selected from the four lists. Each of eight subjects rated 20 words selected randomly from each list. Their predictions of orthographic and phonological matches did not differ significantly among the four word lists ($p > .45$).

2 It is important for the reader to realize that in Figure 2 we plot voltages as a function of time, and therefore the components can be either positive or negative. In figure 3 we plot component loadings which reflect the contribution of a component to the variance of the recorded voltages. Thus, the upward plot of N200 and P300 in figure 3 should not be confused with the bidirectional plot of these components in figure 2.

References

Adam, N. & Collins, G.I. (1978). Late components of the visual evoked potential to search in short term memory. Electroencephalography and Clinical Neurophysiology, 44, 147-156.

Brookhuis, K.A., Mulder, G., Mulder, L.J.M. & Gloerich, A.B.M. (1983). The P300 complex as an index of information processing: The effects of response probability. Biological Psychology, 17, 277-296.

Coles, M.G.H., Gratton, G., Bashore, T.R., Eriksen, C.W. & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. Journal of Experimental Psychology: Human Perception and Performance, 11, 529-553.

Collins, A.M. & Loftus, E.F. (1975). A spreading activation theory of semantic processing. Psychological Review, 82, 407-428.

Donchin, E. (1981). Surprise! ... Surprise? Psychophysiology, 18, 493-513.

Donchin, E. & Heffley, E.F. (1975). Minicomputers in the signal averaging laboratory. American Psychologist, 30, 299-312.

Donchin, E. & Heffley, E. (1979). Multivariate analysis of event related potential data: A tutorial review. In D. Otto (Ed.), Multidisciplinary Perspectives in Event Related Potential Research. EPA 600/9-77-043, Washington, D.C.: U.S. Government Printing Office.

Donchin, E., Kramer, A. & Wickens, C.D. (1982). Probing the cognitive infrastructure with event-related brain potentials. Proceedings of the AIAA Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics. Edwards Air Force Base, California.

Donchin, E., Ritter, W. & McCallum, C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting & S. Koslow (Eds.), Brain Event-Related Potentials in Man, New York: Academic Press.

Duncan-Johnson, C.C. (1981). P300 latency: A new metric for information processing. Psychophysiology, 18, 207-215.

Duncan-Johnson, C.C. & Kopell, B.S. (1981). The stroop effect: Brain potentials localize the source of interference. Science, 214, 938-940.

Eriksen, B.A. & Eriksen, C.W. (1974). Effects of noise letters on the identification of target letter in a nonsearch task. Perception and Psychophysics, 16, 143-149.

Eriksen, C.W., Coles, M.G.H., Morris, L.R., O'Hara, W.R. (1985). An electromyographic examination of response competition. Bulletin of the Psychonomic Society, 23, 165-168.

Eriksen, C.W. & Schultz, D.W. (1979). Rate of information processing in visual search: A continuous flow conception and experimental result. Perception and Psychophysics, 25, 249-263.

Ford, J.M., Mohs, R.C., Pfefferbaum, A. & Kopell, B.S. (1980). On the utility of P300 latency and reaction time for studying cognitive processes. In H.H. Kornhuber and L. Deecke (Eds.), Motivation, Motor and Sensory Processes of the Brain: Electrical Potentials, Behavioral and Clinical Use. Netherlands: North Holland Biomedical Press.

Ford, J.M., Pfefferbaum, A. & Kopell, B.S. (1982). Event-related potentials to a change of pace in a visual sequence. Psychophysiology, 19, 173-177.

Ford, J.M., Roth, W.T. & Kopell, B.S. (1976). Auditory evoked potentials to unpredictable shifts in pitch. Psychophysiology, 13, 32-39.

Gomer, F.E., Spicuzza, R.J. & O'Donnell, R.D. (1976). Evoked correlates of item recognition during memory scanning tasks. Physiological Psychology, 4, 61-65.

Hock, H.S. & Egeth, H. (1970). Verbal interference with encoding in a perceptual classification task. Journal of Experimental Psychology, 83, 299-303.

Jasper, H.H. (1958). The ten twenty electrode system of the international federation. Electroencephalography and Clinical Neurophysiology, 10, 371-375.

Keele, S.W. (1972). Attention demands of memory retrieval. Journal of Experimental Psychology, 93, 245-248.

Klinke, R., Fruhstorfer, H. & Finkenzeller, P. (1968). Evoked resposes as a function of external and stored information. Electroencephalography and Clinical Neurophysiology, 25, 119-122.

Kramer, A.F., Ross, W. & Donchin, E. (1982). Individual differences in human information processing: A chronometric analysis of the role of orthographic and phonological cues in a non-lexical decision task. Psychophysiology, 19, 330-331.

Kutas, M., McCarthy, G. & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 197, 792-795.

LaBerge, D. (1972). Beyond auditory coding. In J.F. Kavanagh & I.G. Mattingly (Eds.), Language by ear and by eye. Cambridge Mass.: MIT Press.

LaBerge, D. & Samuels, S.J. (1974). Toward a theory of automatic information processing in reading. Cognitive Psychology, 6, 293-323.

Magliero, A., Bashore, T.R., Coles, M.G.H. & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. Psychophysiology, 21, 171-186.

McCarthy, G. & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. Science, 1981, 211, 77-80.

McClelland, J.L. (1979). On time relations of mental processes: A framework for analyzing processes in cascade. Psychological Review, 86, 287-330.

Morton, J. (1969). Interaction of information in word recognition. Psychological Review, 76, 163-178.

Naatanen, R. (1982). Processing negativity: An evoked-potential reflection of selective attention. Psychological Bulletin, 92, 605-640.

Naatanen, R. & Gaillard, A.W.K. (1983). The orienting reflex and the N2 deflection of the ERP. In A.W.K. Gaillard & W. Ritter (Eds.), Tutorials in Event Related Potential Research: Endogenous Components. Amsterdam: North Holland Publishing Company.

Naatanen, R., Gaillard, A.W.K. & Varey, C.A. (1981). Attention effects on auditory EPs as a function of inter-stimulus interval. Biological Psychology, 13, 173-187.

Naatanen, R., Hukkanen, S. & Jarilehto, T. (1980). Magnitude of stimulus deviance and brain potentials. In H.H. Kornhuber & L. Deecke (Eds.), Motivation, Motor and Sensory Processes of the Brain Electrical Potentials, Behavioral and Clinical Use. Netherlands: North Holland Biomedical Press.

Naatanen, R., Simpson, M. & Loveless, N.E. (1982). Stimulus deviance and evoked potentials. Biological Psychology, 14, 53-98.

Nelson, M.J. & Denny, E.C. (1960). The Nelson-Denny reading test. Houghton Mifflin Company.

Nickerson, R.S. (1978). On the time it takes to tell things apart. In J.Requin (Ed.), Attention and Performance VII. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Nolan, S.D., Tanenhaus, M.K. & Seidenberg, M.S. (1981). Multiple code activation in word recognition: Evidence from rhyme monitoring. Center for the Study of Reading, University of Illinois, Technical Report No. 204.

Picton, T.W., Campbell, K.B., Baribeau-Braun, J. & Proulx, G.B. (1978). The neurophysiology of human attention: A tutorial review. In J. Requin (Ed.), Attention and Performance VII. Hillsdale, New Jersey; Lawrence Erlbaum Associates.

Polich, J., McCarthy, G., Wang, W.S. & Donchin, E. (1983). When words collide: Orthographic and phonological interference during word processing. Biological Psychology, 16, 155-180.

Posner, M.R. (1978). Chronometric Explorations of Mind. Hillsdale, New Jersey: Lawrence Erlbaum.

Posner, M.R. & Synder, C.R.R. (1975). Attention and cognitive control. In R.L. Solso (Ed.), Information Processing and Cognition: The Loyola Symposium. Hillsdale, New Jersey: Lawrence Erlbaum.

Pritchard, W.S. (1981). Psychophysiology of P300. Psychological Bulletin, 89, 506-540.

Proctor, R.W. (1981). A unified theory for matching task phenomena. Psychology Review, 88, 291-326.

Regan, D. (1972). Evoked Potentials in Psychology, Sensory Physiology and Clinical Medicine. London: Chapman Hall.

Sanquist, T.F., Rohrbaugh, J., Syndulko, K. & Lindsley, D.B. (1980). An event-related potential analysis of coding processes in human memory. In H.H. Kornhuber & L. Deecke (Eds.), Motivation, Motor and Sensory Processes of the Brain: Electrical Potentials, Behavioral and Clinical Use. Netherlands: North Holland Biomedical Press.

Seidenberg, M.S. & Tanenhaus, M.K. (1979). Orthographic effects on rhyme monitoring. Journal of Experimental Psychology: Human Learning and Memory, 5, 546-554.

Squires, K.C., Donchin, E., Herning, R.I. & McCarthy, G. (1977). On the influence of task relevance and stimulus probability on event-related potential components. Electroencephalography and Clinical Neurophysiology, 42, 1-14.

Squires, N.K., Donchin, E., Squires, K.C. & Grossberg, S. (1977). Bisensory stimulation: Infering decision related processes from the P300 component. Journal of Experimental Psychology: Human Perception and Performance, 3, 299-315.

Sternberg, S. (1969). The discovery of processing stages: Extension of Donders method. Acta Psychologica, 30, 276-315.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18, 643-662.

Tanenhaus, M.K., Flanigan, H. & Seidenberg, M.S. (1980). Orthographic and phonological activation in auditory and visual word recognition. Center for the Study of Reading, University of Illinois, Technical Report No. 178.

Teece, J.J. & Dimartino, M. (1965). Effects of heightened drive (shock) on performance in a tachistoscopic color-word interference task. Psychological Reports, 16, 93-94.

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

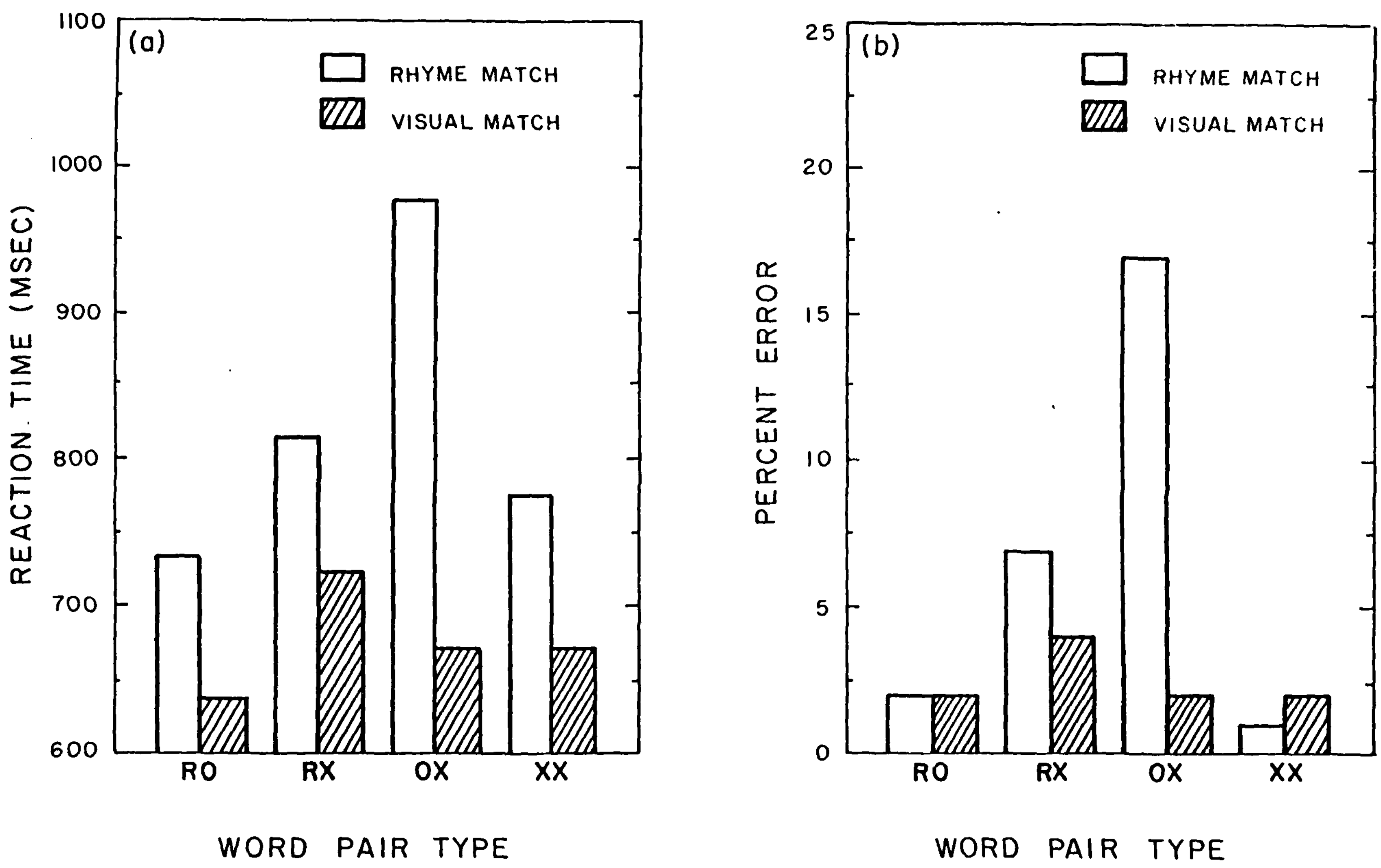
Table 1

Means (msec), standard deviations (SD) and error rates (%) for the four word lists in the Rhyme and Visual match conditions.

	Rhyme Match				Visual Match			
	RO	RX	OX	XX	RO	RX	OX	XX
RT	738	809	978	770	639	726	674	679
SD	61	68	84	73	58	64	68	61
P300 Latency	530	568	676	578	482	556	479	524
SD	43	38	33	40	40	32	35	37
Error Rate (%)	2	7	17	1	2	4	2	2

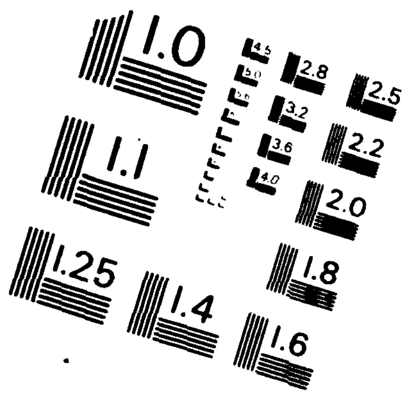
Appendix A

RO		RX		OX		XX	
PLEA	FLEA	MAKE	ACHE	SAID	PAID	MIND	WALL
TRIM	GRIM	NONE	STUN	HOME	SOME	HURT	POST
LAST	PAST	DUNE	MOON	HAVE	CAVE	DORM	PIPE
DEAL	SEAL	FOOL	DUAL	BOMB	TOMB	WOOD	TAPE
TILT	WILT	HAIL	SALE	DULL	PULL	MEAT	DIRT
CLOT	BLOT	HAIR	CARE	WARD	CARD	TEST	SOCK
MARK	DARK	TOOL	RULE	LUST	MUST	SWIM	EAST
BORN	WORN	THOU	PLOW	YOUR	HOUR	LEAF	SOON
RICE	NICE	LOAN	BONE	MINT	PINT	JAZZ	FILL
FILE	TILE	FEUD	MOOD	DOLL	TOLL	FERN	BACK
SOFT	LOFT	BLUE	FLEW	FOUL	SOUL	BANK	SURE
NUMB	DUMB	GAIT	HATE	HUSH	BUSH	DOOR	SING
COIL	BOIL	DIAL	MILE	CROW	BROW	BOOK	NOUN
MINK	PINK	NAIL	PALE	GASP	WASP	SICK	DULL
RUNG	SUNG	HOPE	SOAP	BONE	GONE	MEAN	HIGH
HILL	WILL	NEWS	LOSE	NOSE	LOSE	REEL	EACH
MUCH	SUCH	NUDE	LEWD	GOLF	WOLF	DIME	BOAT
MADE	FADE	HAIL	MALE	FOOD	HOOD	WHAT	SIDE
COLD	TOLD	COMB	FOAM	SOUR	FOUR	THAT	HAND
PAIR	FAIR	BEAR	FARE	BOTH	MOTH	PARK	ROAD
LIKE	BIKE	ROOM	FUME	HARM	WARM	TURN	FORM
HAND	LAND	FATE	BAIT	HAVE	GAVE	FROM	GIFT
DAMP	LAMP	ROAM	DOME	BOOT	FOOT	HALF	TALK
HARM	FARM	WEAR	RARE	BEAR	FEAR	GIRL	JUMP
LOAD	TOAD	SHOE	CHEW	WERE	HERE	WALK	LEFT
PITCH	DITCH	JEWEL	SPOOL	WASP	RASP	STACK	BRAIN
MIGHT	TIGHT	STOOL	CRUEL	GIVE	HIVE	SHELF	FIGHT
POINT	JOINT	REIGN	TRAIN	WAND	HAND	CLEAR	GROUP
CRANK	DRANK	CROAK	SPOKE	LOVE	MOVE	ROUND	GRIPE
POISE	NOISE	PHONE	KNOWN	SEAR	BEAR	GLASS	DRIVE
BRACE	TRACE	FLOAT	QUOTE	WORD	LORD	STORE	WHOLE
GUILT	BUILT	BLOOM	SPUME	WEAR	HEAR	SHIRT	WITCH
HOUSE	MOUSE	SPOON	PRUNE	DONE	HONE	LEARN	CLEAR
PORCH	TORCH	FRUIT	SHOOT	WERE	HERE	THING	STORM
CREAM	DREAM	GREAT	STATE	BOOT	FOOT	FORCE	BRAKE
BLAME	FLAME	TOUGH	STUFF	CATCH	WATCH	SHEET	THREE
BRIBE	TRIBE	WRITE	LIGHT	ROUGH	DOUGH	LEAVE	TWICE
FENCE	HENCE	BRAIL	WHALE	YOUTH	SOUTH	PLATE	STEAM
YIELD	FIELD	MOOSE	JUICE	GROWN	CROWN	BLESS	STAIR
BARGE	LARGE	WAIST	BASTE	DROVE	PROVE	STICK	TEACH
VAULT	FAULT	FRAIL	SCALE	BEARD	HEARD	GRAPE	SKIRT
HINGE	BINGE	STAIN	CRANE	FLOWN	CLOWN	BLAME	CLOCK
MATCH	PATCH	BLOWN	STONE	CLOVE	GLOVE	CHAIR	START
SIGHT	LIGHT	STARE	FLAIR	FREAK	BREAK	CHAIN	THINK
BRUSH	CRUSH	GRAIN	BLAME	COUCH	TOUCH	BROWN	MARCH
GLAND	BLAND	STOLE	DROLL	HORSE	WORSE	PLANT	THESE
BATCH	LATCH	PLANE	TRAIN	COUTH	MOUTH	GRANT	ROAST
GROVE	DROVE	BLARE	STAIR	TASTE	CASTE	EIGHT	WHARF
SOUND	POUND	GROAN	FLOWN	COULD	MOULD	PEACH	DRESS
CLOCK	FLOCK	BREAK	FLAKE	BOUGH	TOUGH	SHIFT	TORCH

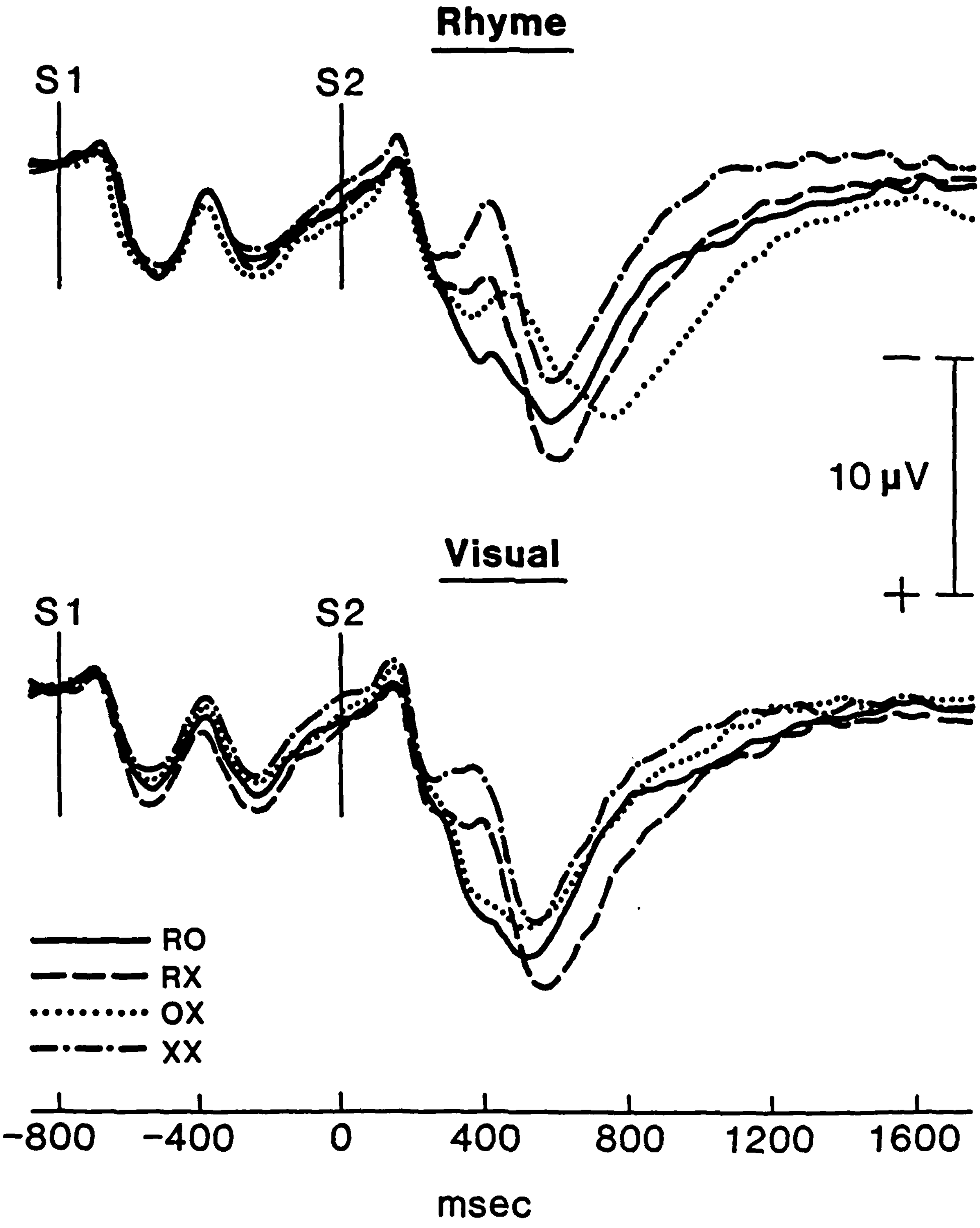


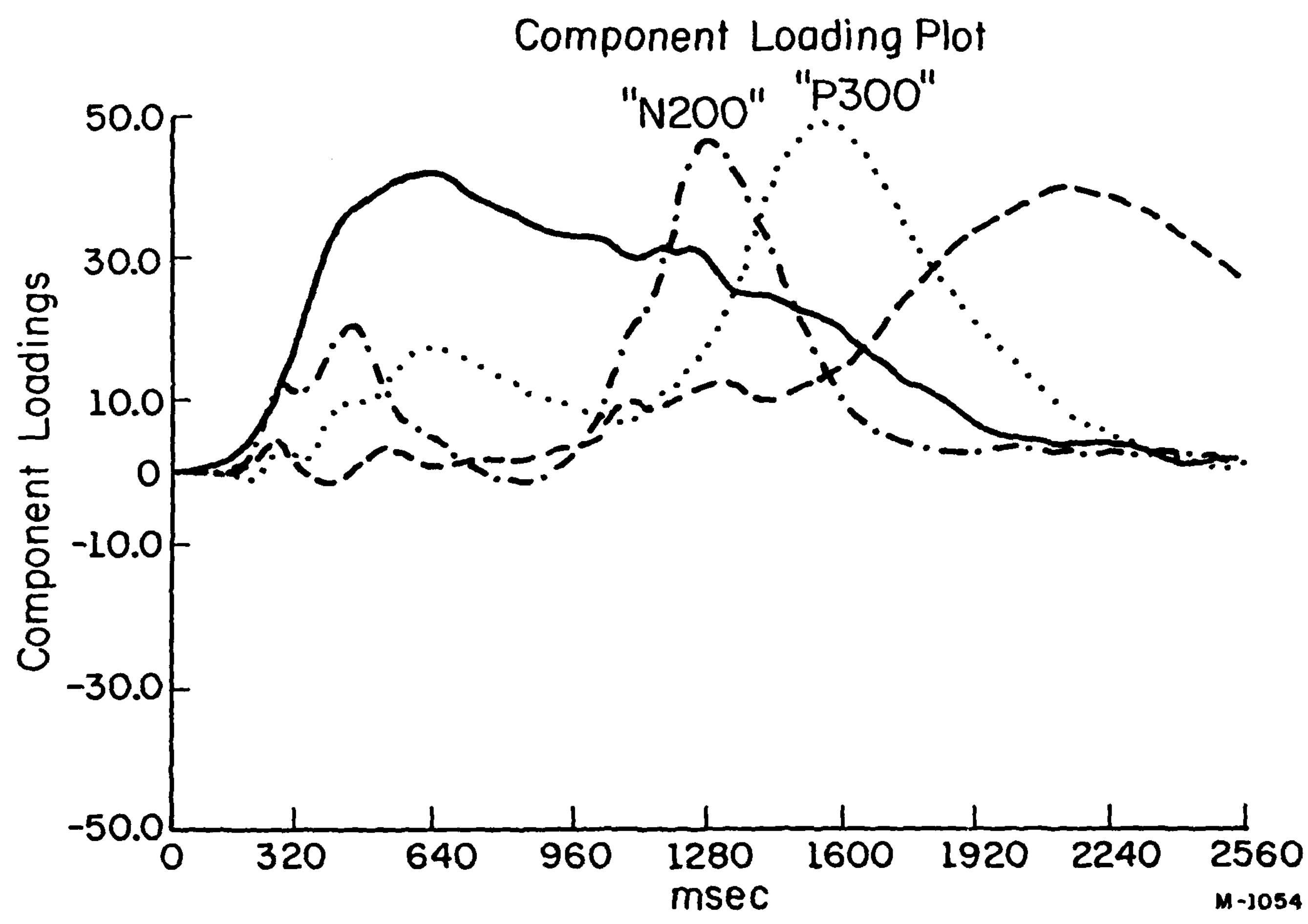
(a)
RHYME MATCH
VISUAL MATCH
REACTION. TIME (MSEC)
1100
1000
900
800
700
600
RO
RX
OX
XX
WORD PAIR TYPE
(b)
RHYME MATCH
VISUAL MATCH
PERCENT ERROR
25
20
15
10
5
0
RO
RX
OX
XX
WORD PAIR TYPE

AD-A170 794 THE EVENT-RELATED BRAIN POTENTIAL AS AN INDEX OF 8/8
INFORMATION PROCESSING A..(U) ILLINOIS UNIV CHAMPAIGN
COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL.
UNCLASSIFIED 20 FEB 86 CPL-86-1 F49620-85-C-0041 F/G 5/10 NL

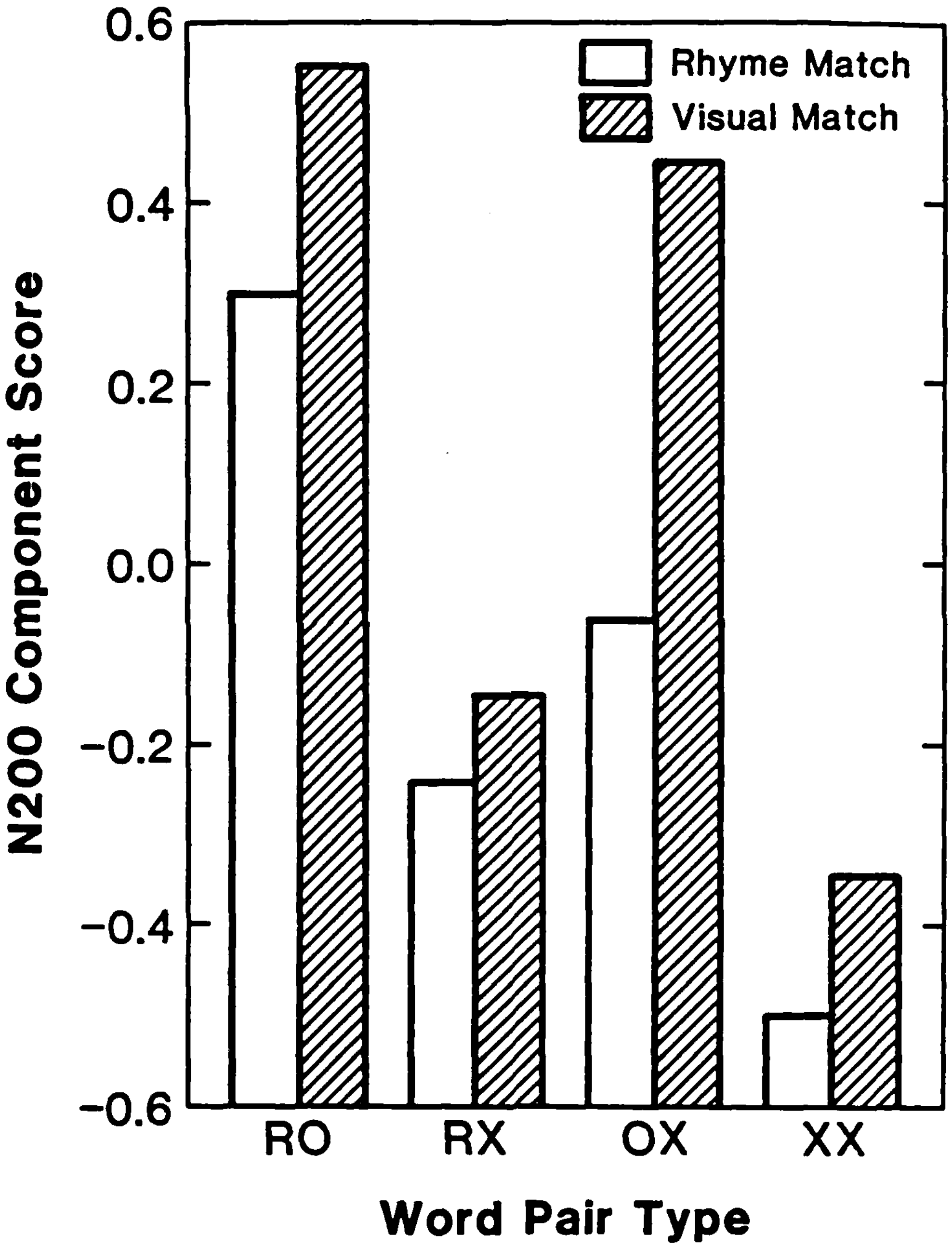


MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963

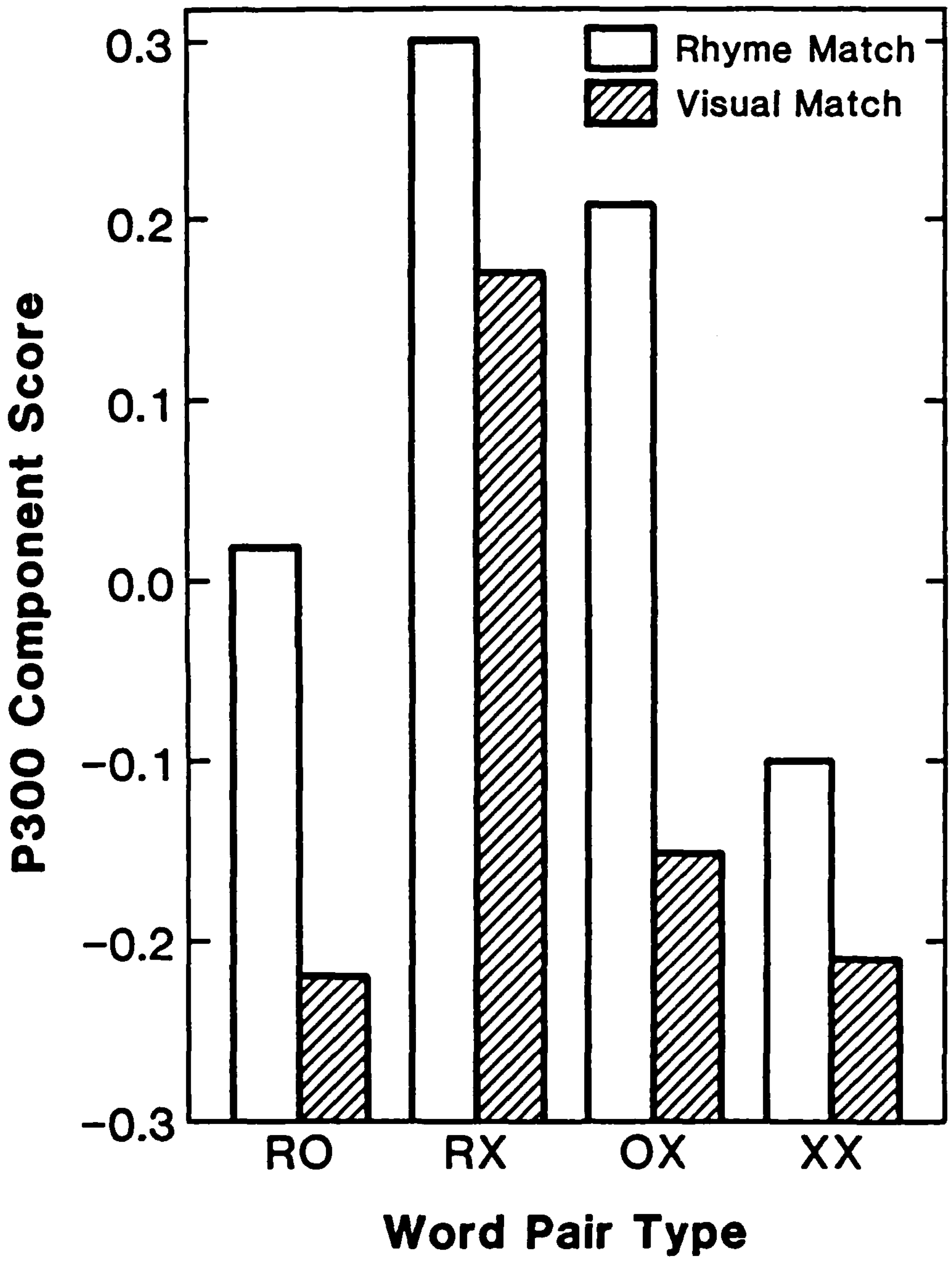

Rhyme
S1
S2
10 µV
Visual
S1
S2
RO
RX
OX
XX
-800
-400
0
400
800
1200
1600
msec



Component Loading Plot
"N200"
"P300"
Component Loadings
50.0
30.0
10.0
0
-10.0
-30.0
-50.0
0
320
640
960
1280
1600
1920
2240
2560
msec
M-1054



0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
N200 Component Score
Rhyme Match
Visual Match
RO
RX
OX
XX
Word Pair Type



P300 Component Score
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
Rhyme Match
Visual Match
RO
RX
OX
XX
Word Pair Type

Journal of Experimental Psychology:
Human Perception and Performance

VOL. 11, NO. 4 AUGUST 1985

Processing of Stimulus Properties: Evidence for Dual-Task Integrality

Arthur F. Kramer, Christopher D. Wickens, and Emanuel Donchin
University of Illinois

The conditions under which dual-task integrality can be fostered were assessed in a study in which we manipulated four factors likely to influence the integrality between tasks: intertask redundancy, the spatial proximity of primary and secondary task displays, the degree to which primary and secondary task displays constitute a single object, and the resource demands of the two tasks. The resource allocation policy is inferred from changes in the amplitude of the P300 component of the event-related brain potential. Twelve subjects participated in three experimental sessions in which they performed both single and dual tasks. The primary task was a pursuit step tracking task. The secondary tasks required subjects to discriminate between different intensities or different spatial positions of a stimulus. Task pairs that required the processing of different properties of the same object resulted in better performance than task pairs that required the processing of different objects. Furthermore, these same object task pairs led to a positive relation between primary task difficulty and the resources allocated to secondary task stimuli. Intertask redundancy and the physical proximity of task displays produced similar effects of reduced magnitude.

The concurrent processing of information relevant to several tasks has been addressed in the psychological literature from the early writings of James (1890) to contemporary investigations of dual-task performance in complex, operational environments. Substantial theoretical and empirical effort has been expended in mapping the conditions under which the demands imposed by tasks performed concurrently interact so that performance on one or both tasks degrades. Navon and Gopher (1979) proposed that these conditions are related to the degree to which two tasks compete for "processing resources." It is critical to note that at present there is no objective, analytic way to specify the resources required to perform a task. Rather, the concept of resources serves as a convenient model that allows us to deal with the competition between tasks using the tools of economics.

In principle, tasks that are more successfully time shared are presumed to compete less for common resources than tasks that cannot be performed concurrently with ease. In much of the early work in this domain, resources were treated as undifferentiated. However, it has become increasingly evident that the data are more consistent with models that assume the existence of specific resources. Wickens (1980) proposed a multiple resource

This article is based on a doctoral dissertation submitted to the University of Illinois by the first author. The research was supported by the Air Force Office of Scientific Research under Contract F49620-83-0144, Alfred Fregly, technical monitor. We gratefully acknowledge the helpful comments of Michael Coles, Gabriele Gratton, and Demetrios Karis.

Requests for reprints should be sent to Arthur F. Kramer, Department of Psychology, University of Illinois, 603 East Daniel, Champaign, Illinois 61820.

model according to which processing resources may be represented by three dimensions: stages of processing, modalities of processing, and codes of processing. This theoretical conceptualization of the processing structure of dual-task performance has received considerable empirical support. Attempts to perform concurrently tasks that require the use of the same modalities, the same codes, or the same stages of processing generally show a large effect of competition. This is not the case when concurrently performed tasks require resources from different structures (Alwitt, 1981; Isreal, Chesney, Wickens, & Donchin, 1980; North, 1977; Trumbo, Noble, & Swink, 1967; Wickens & Kessel, 1979; Wickens, Sandry, & Vidulich, 1983).

The present article is concerned with the availability of cooperative processing strategies that make it possible for the processing routines necessary for one task to be used in the processing of another task. This emphasis on cooperative processing strategies is in marked contrast with the common approach to the analysis of dual-task performance. In general, investigators have assumed that the processing necessary to meet the performance criteria for one of the tasks is either independent of or at odds with the processing required for meeting the criteria for the other task. That is, the emphasis has always been on the competition for resources between tasks. In this article, we focus on the cooperation between tasks. Specifically, we propose that the stimulus ensembles associated with different tasks can, on occasion, be manipulated so that a cooperative concurrence benefit can be realized.

The relation between performance and processing resources forms the basis for dual-task performance models (Kahneman, 1973; Navon & Gopher, 1979; Wickens, 1980). Therefore, it is useful to describe briefly the types of relations that have been proposed. Figure 1 illustrates the different relations among resource functions that are presumed to underlie performance in different dual-task situations. Three cases are examined: resource reciprocity, separate resources, and dual-task integrality. The left side of the figure depicts the change in performance of a primary and a secondary task as primary task difficulty increases. These performance functions correspond to the functions shown on the right side of the figure, which map the hypothetical allocation of resources between the two tasks on primary task difficulty. In dual-task paradigms in which tasks are assigned different priorities, it is assumed that primary task performance is protected at the expense of performance on the secondary task (Rolfe, 1971; Wickens, 1979). Thus, in the performance–difficulty functions shown in the left column of Figure 1, there is no decrement in primary task performance with increases in task difficulty.

In the resource reciprocity case, increasing the difficulty of the primary task results in a decrease in the performance on the secondary task. The corresponding resource–difficulty function displays a trade-off between the two tasks. Increasing primary task difficulty leads to an increased demand for resources by the primary task. This results in a decreasing supply of resources for the secondary task, and hence, its performance deteriorates. In the separate resource case, increasing the difficulty of the primary task has no effect on the performance of the secondary task. This suggests a resource–difficulty function that reflects the insensitivity of the secondary task to the withdrawal of resources. That is, we infer the separateness of the resources from the failure of the two tasks to compete.

The two relations described thus far are those on which most of the modern analyses of dual-task performance have concentrated. In this report, however, we focus on the third relation, which we have labeled *dual-task integrality*. In this case, secondary task performance improves with the increasing difficulty of the primary task. Thus, it is assumed that the secondary task can benefit from the additional resources supplied to cope with the increase in primary task demands. The corresponding resource–difficulty function displays a single function that represents the resources allocated to both tasks.

The notion of dual-task integrality derives from Garner's concept of integral dimensions. Two dimensions are said to be integral if there must be a level specified on one dimension in order for a level on the other dimension to be realized (Garner, 1969). That is, dimensional integrality is observed when the dimensions are in some sense close to each other. In a series of experiments, Garner and co-workers found that a pair of dimensions that are integral normally show cooperative performance benefits (Garner, 1970; Garner & Felfoldy, 1970; Garner & Flowers, 1969). These occur in stimulus classification tasks when the levels along the two dimensions are correlated. In the present article, we argue that dual-task integrality also occurs when two tasks are in some sense close to each other. *Closeness* in this case is defined by features of the two tasks such as a correlation of their stimulus elements or the fact that the tasks depend on a common spatial location or on the dimensions of a common object. The consequence of Garner's dimensional integrality was a facilitation in the classification of redundant dimensions. The implications of dual-task integrality are proposed to be the cooperative sharing of mobilized resources, as shown in the lower panel of Figure 1.

Several factors are presumed to influence the degree of integrality between tasks. One factor, the redundancy between components of the tasks, can be illustrated with reference to a hypothetical experiment. For example, subjects may be instructed to perform two separate tasks concurrently. One task requires tracking a target with a cursor along a single axis on a cathode ray tube (CRT), and the other task calls for a discrimination between flashes that differ in brightness. What if an event embedded in one of the tasks predicts with some degree of certainty the appearance

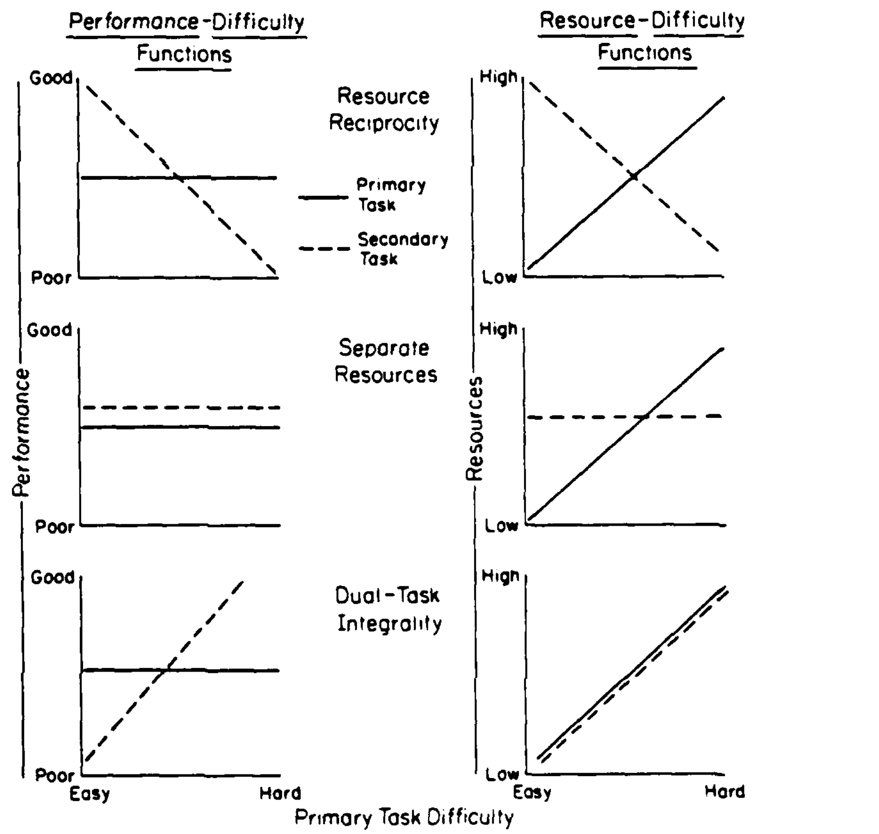


Figure 1. Relations between resource and performance functions. (Left panel presents the performance-difficulty functions; right panel presents the corresponding resource-difficulty functions. Primary task difficulty is represented on the abscissa on both panels. The primary task is indicated by the solid line; the secondary task is represented by the dashed line.)

of a change in events associated with the other task? For example, the spatial position of the tracking target may predict (i.e., may be correlated with) the brightness of the secondary task stimulus. Thus, although the assignments to the subject when specified in terms of the proper responses to the two stimulus sets have not changed, an overlap has been introduced between the two stimulus sets. The correlation between the target position and flash brightness incorporates the tracking stimuli into the brightness discrimination task. Such overlap between tasks may sometimes facilitate, and sometimes hinder, performance in a dual-task situation. In this study we examine some of the conditions that allow operators to benefit from such concurrency (see Navon & Gopher, 1979).

Recent models of attention have emphasized the influence of the spatial location of stimulus properties on the efficiency of processing. Treisman (1977), in her feature integration model of attention, has argued that features occurring within the same central fixation of attention are processed in parallel and combined to form a single object. Once the object has been formed, it is perceived and stored in memory as such. The importance of the spatial location of stimulus properties has also been emphasized in the research on integral and separable dimensions. By definition, integral dimensions must occur in the same space and time (Lockhead, 1966). In the context of the present study, it is of interest to determine whether the spatial proximity of different tasks influences dual-task performance, and if so, how spatial proximity interacts with other factors that affect the processing of dual tasks.

Investigators have reported that the processing of several stimulus dimensions is more efficient if the dimensions are incorporated in a single object rather than in several objects. In one such study, subjects were to detect a different dimension on each of three separate objects, the same dimensions on three separate objects, or three different dimensions on the same object (Lappin, 1967). Identification of the dimensions was best when the three dimensions were located on the same object. Kahneman and Treisman (1984) have investigated further the advantages of incorporating different dimensions within a single object. In their studies, superior performance resulted from a change in the relations among the properties of the stimuli, though the type of processing required by the tasks remained the same. When the entities to be processed appeared as properties of a single object, or context, performance was enhanced.

Kahneman and co-workers (Kahneman & Henik, 1981; Kahneman & Treisman, 1984; Kahneman, Treisman, & Burkell, 1983) and Duncan (1984) have underscored the importance of the object in attention by suggesting that attentional competition arises between, not within objects. This argument implies that tasks requiring the processing of different dimensions of the same object will be processed with the same resource framework. Tasks that necessitate the processing of separate objects will compete for processing resources. Thus, the degree to which two separate tasks can be integrated into a single object will presumably determine the resource competition between the tasks. The hypothesis of competition for resources between objects and not among different dimensions of the same object will be investigated further in the present study by manipulating the stimulus relations among dual-task pairs.

The studies described thus far have shown a processing advantage for stimulus dimensions that are integrated into a single object. Other investigators have shown that it is difficult to selectively attend to one dimension of an object while ignoring other dimensions (Kahneman & Chajczyk, 1983; Polich, McCarthy, Wang, & Donchin, 1983; Seidenberg & Tanenhaus, 1979; Stroop, 1935). In these studies it appears that the processing of irrelevant dimensions will be carried out regardless of their effect on the relevant task. Thus, in some cases the additional processing will have facilitating effects on performance, whereas in other cases interference between the relevant and irrelevant dimensions will be produced. Indeed, there is ample evidence for patterns of both facilitation and interference (Garner, 1974; Pomerantz & Garner, 1973; Reicher, 1977; Weisstein & Harris, 1974). In these studies, the stimuli are generally presented briefly and at low levels of illumination. Studies that have presented suprathreshold stimuli at durations exceeding

200 ms have found that subjects are capable of selectively attending to particular properties of objects (Donchin & Cohen, 1967; Kramer, Wickens, & Donchin, 1983). In the present study we predict that selective attention can be directed to specific dimensions of objects because the objects are readily perceptible and the subjects are not stressed for a speeded response.

Event-Related Brain Potentials and Processing Resources

One difficulty in resolving issues regarding dual-task integrality is finding a way of assessing resource allocation that is independent of the criterion responses. In this study we complimented the standard tools of dual-task analysis by using a psychophysiological index of resource allocation, the amplitude of the P300 component of the event-related brain potential (ERP).

The ERP is a transient series of voltage oscillations in the brain that can be recorded on the scalp following a discrete event (Donchin, 1975; Regan, 1972). The ERP has traditionally been partitioned into a number of separate components. In most cases component labels indicate both the polarity and approximate latency of the peak (e.g., *N100* is a negative peak occurring approximately 100 ms after stimulus onset). The amplitude and latency of the early components, those occurring within the first 100 ms, have been shown to be influenced by the physical attributes of stimuli (e.g., intensity, modality, presentation rate). These components have been labeled *exogenous*. Later components such as N200 and P300 are nonobligatory responses to stimuli. These *endogenous* components vary in amplitude, latency, and scalp distribution with the strategies, expectancies, and other mental activities triggered by the event eliciting the ERP. These components are largely uninfluenced by the physical attributes of the stimuli. The amplitude of one such endogenous component, the P300, is one of the dependent variables in the present study.

The P300 component of the ERP has been found useful in providing information concerning the allocation of resources to concurrently performed tasks. It has been shown that P300s elicited by discrete secondary task events decrease in amplitude with increases in the perceptual/cognitive difficulty of the primary task (Isreal, Chesney, Wickens, & Donchin, 1980; Kramer et al., 1983) but not with increases in purely motor demands (Isreal, Wickens, Chesney, & Donchin, 1980). The P300 amplitude, when plotted against primary task difficulty, can serve as a representation of the resource–difficulty functions described in Figure 1. If this is indeed the case, then we would predict that P300s elicited by primary task events would increase in amplitude with increases in the difficulty of the primary task. This hypothesis was confirmed in a series of studies in which P300s were elicited by discrete spatial changes in the position of a target in a tracking task (Wickens, Kramer, Vanasse, & Donchin, 1983; Sirevaag, Kramer, Coles, & Donchin, 1984). Increasing the difficulty of the tracking task by decreasing the stability of the control dynamics resulted in a systematic increase in primary task P300 amplitude, a finding similar to that depicted in the top right of Figure 1. The same increase in difficulty produced a decrease in secondary task P300s.

The reciprocal relation between P300s elicited by primary and secondary task stimuli as a function of primary task difficulty is identical to the resource trade-offs presumed to underlie dual-task performance decrements. Thus, the hypothetical resource functions illustrated on the right side of Figure 1 might be inferred from changes in the amplitude of P300 as a function of task difficulty. In the present experiment, P300s will be used to provide information concerning the resource framework of dual-task combinations as various perceptual characteristics are manipulated that are presumed to affect the degree of dual-task integrality. Specifically we will manipulate the spatial proximity of the primary and secondary task displays, the degree to which the primary and secondary task displays constitute a common object, the similarity of resource demands of primary and secondary tasks, and the degree of correlation between primary and secondary task events. The first three factors will be varied in a factorial design, and the correlation manipulation will be varied at a fixed level of position and stimulus resource demand variables.

Method

Subjects

Twelve right-handed persons (6 men and 6 women) were recruited from the student population at the University of Illinois and paid for their participation in the study. None of the students had any prior experience with the pursuit step tracking task. All of the subjects had normal or corrected to normal vision.

Step Tracking and Discrimination Tasks

The single axis pursuit step tracking task is illustrated in Figure 2. The tracking display, which consisted of the computer driven target and the subject controlled cursor, was presented on a Hewlett Packard CRT positioned approximately 70 cm from the subjects. The target and cursor were 1.2 cm × 1.2 cm in size and subtended a visual angle of 1.0°. The target changed its position along the horizontal axis once every 3.6 s to 4 s, and the subjects' task was to nullify the position error between the target and cursor. The cursor was controlled by manipulating a joystick with the right hand. Pursuit step tracking was defined as the primary task. The dynamics for the tracking stick were composed of a linear combination of first order (velocity) and second order (acceleration) components. That is, the system output, $X(t)$, is represented by the following equation: $Xt = [(1 - a) \int u dt] + [a \iint u dt]$, where u = stick position, t = time, a = difficulty level.

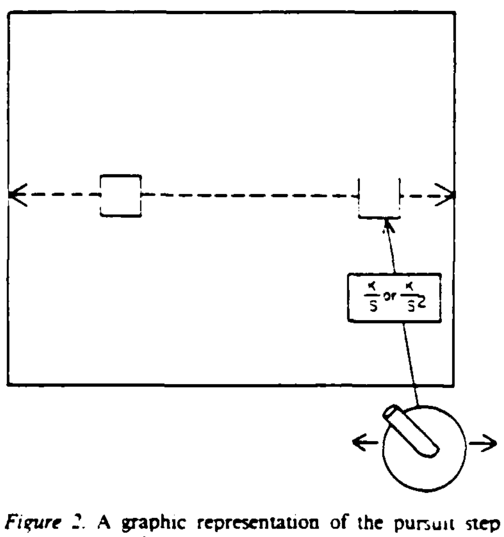


Figure 2. A graphic representation of the pursuit step tracking task. (Subjects' task was to track the computer controlled target with the cursor along the horizontal axis; difficulty of the tracking task was manipulated by changing the control dynamics from a first order [K/S] to a second order [K/S²] system.)

The task was conducted at three different levels of the system order manipulation: First, in the relatively easy condition, a was set to 0, a pure first order (velocity) system. Second, in the moderate difficulty condition, a was set to 0.5, a 50/50 combination of first and second order dynamics. Third, in the difficult tracking condition, a was set to 1.0, a pure second order (acceleration) system. Numerous investigators have validated the increasing difficulty associated with higher order control (Kramer et al., 1983; North, 1977; Trumbo et al., 1967; Vidulich & Wickens, 1981). Converging evidence using Sternberg's additive factors paradigm indicates that the demands of higher order tracking are in fact perceptual/cognitive in nature, given the requirement to process higher derivatives of the error signal to maintain stable control (Wickens, Derrick, Micallizi, & Beringer, 1980).

The subjects' secondary task required counting covertly the total number of occurrences of a relevant probe. Probes were presented in a Bernoulli series. The probability that either of the two stimuli would occur on any one trial was .50. In different experimental blocks, subjects counted bright flashes of a horizontal bar, bright flashes of a cursor, translational changes of a cursor upward, or translational changes of a horizontal bar downward (see Figures 3 and 4). Secondary task probes occurred either on the same horizontal axis as the tracking task or 2 cm (1.5° of visual angle) below it. A probe was presented every 3.6 s to 4 s. The presentation of the probe was temporally constrained so that it occurred 1.8 s to 2 s subsequent to a step change of the target in the tracking task. Thus, the temporal sequence of the presentation of the probes (secondary task stimuli) and of changes in the spatial position of the tracking target was fixed, and the temporal interval between these stimuli was variable. Although changes in the spatial position of the target were discrete events, the tracking task was performed continuously because the subjects were required to constantly manipulate the joystick to nullify the position error between the target and the cursor.

In the dual-task blocks, subjects concurrently performed the tracking task and the count task. At the conclusion of each block of trials, subjects reported their total count. At this time subjects also rated the subjective difficulty of the block on a bipolar scale from 1 (*easy*) to 7 (*difficult*). Following each block the subjects were informed of their count accuracy and root mean square (RMS) tracking error.

Recording System

Electroencephalographic activity (EEG) was recorded from three midline sites (Fz, Cz, and Pz, according to the International 10-20 system; Jasper, 1958) and referred to linked mastoids. Two ground electrodes were positioned on the left side of the forehead. Burden Ag-AgCl electrodes affixed with collodion were used for scalp and mastoid recording. Beckman Biopotential electrodes, affixed with adhesive collars, were placed below and supraorbitally to the right eye to record electro-oculogram (EOG), and this type of electrode was also used for ground recording. Electrode impedances did not exceed 5 kohms/cm.

The EEG and EOG were amplified with Van Gogh model 50000 amplifiers (time constant 10 s and upper half amplitude of 35 Hz, 3-dB [SPL] octave roll-off). Both EEG and EOG were sampled for 1,280 ms, beginning 100 ms prior to stimulus onset. The data were digitized every 10 ms. The ERPs were filtered off-line (−3 dB at 6.29 Hz, 0 dB at 14.29 Hz) prior to statistical analysis. Evaluation of each EOG record for eye movements and blinks was conducted off-line. The EOG contamination of EEG traces was compensated for through the use of an eye movement correction procedure (Gratton, Coles, & Donchin, 1982).

Design

A repeated measures, four-way factorial design was used. The factors were primary task difficulty (count only, first order, first/second order, and second order control dynamics); the relation between primary and secondary task stimulus objects (same or different objects); the spatial position of the primary and secondary task displays (same or different); and the type of secondary task (intensity or translational discriminations).[1]

The degree of correlation between the primary and secondary tasks was also manipulated, although this manipulation was not orthogonal to the other four factors. Subjects performed the dual tasks with either low or high (0 or .85, respectively) correlation at each level of difficulty in the same-object–same-position and different-object–same-position conditions with the intensity discrimination secondary task. In the correlated dual-task conditions, a change in the intensity of the secondary task stimulus

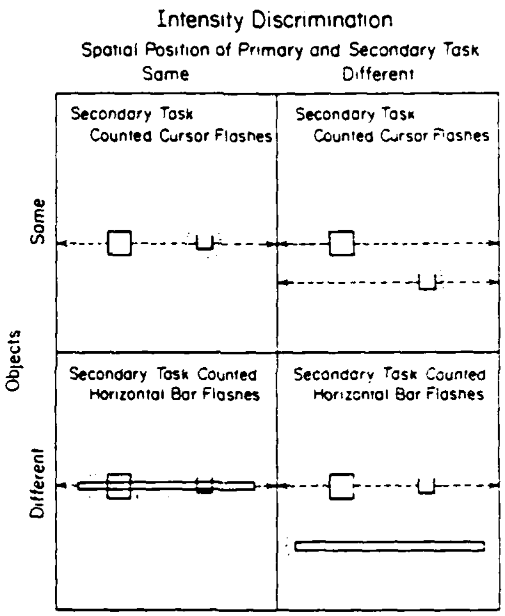


Figure 4. A graphic illustration of the intensity discrimination secondary tasks and their relation to experimental manipulations.

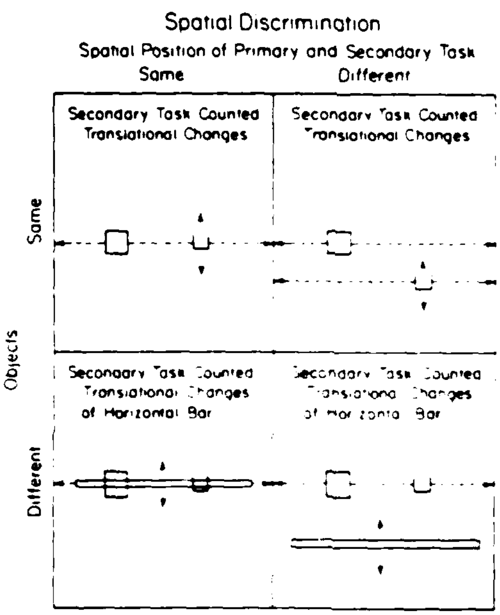


Figure 3. A graphic illustration of the spatial discrimination secondary tasks and the tracking task. (Relation of the task configurations to the experimental manipulations is represented along the abscissa [task display position] and the ordinate [relevant objects]. In the same-object condition, one of the tracking elements is also used for the secondary task discrimination.)

(cursor or horizontal bar) correctly predicted the direction of the subsequent change in the spatial position of the target in the tracking task 85% of the time. Thus, a bright flash was highly predictive of a movement of the target to the right; a dim flash would indicate with a high degree of certainty a jump to the left.

Procedure

Each of the 12 subjects participated in all of the experimental conditions. One practice and two experimental sessions, run on successive days, were required to complete the experiment. The practice session included 24 blocks of tracking and 6 secondary task count blocks. Each of the tracking blocks lasted 4 min. Subjects performed 8 blocks of tracking at each of the three levels of system order. Secondary task blocks lasted approximately 6 min. *Although subjects did not count the probes* in the tracking blocks, ERPs were digitized from both step changes of the target and presentations of the probes. Thus, these blocks served as practice as well as an indication of subjects' allocation of processing resources between the tracking task and the irrelevant probe stimuli.

The experimental sessions began with 3 tracking blocks each lasting approximately 3 min. Following the tracking blocks, subjects performed 15 dual-task blocks. The [illegible] dual-task blocks divided between Sessions 2 and 3 consisted of 24 blocks from the 3 (tracking difficulty levels) ×

[1] The cu[illegible] different-position condition is actually quite similar to the cursor–same-position condition.

2 (type of stimuli) × 2 (task position) × 2 (secondary task) factorial design and 6 blocks in which dual tasks in the same-object–same-position and different-object–same-position conditions were highly correlated (.85). Each of the dual-task blocks lasted approximately 6 min. Subsequent to the dual-task blocks, subjects again performed 3 single-task tracking blocks. The ERPs, subjective ratings, and RMS tracking error were recorded during the experimental sessions. The order of the experimental blocks was counterbalanced across subjects.

Results

Measures of Tracking Performance

Table 1 presents the mean RMS tracking error for each level of system order during both single- and dual-task performance. The RMS error data have been collapsed across the type of secondary task (intensity or translational discrimination) because this manipulation did not significantly affect performance. As suggested in the table, higher levels of system order result in increases in subjects' tracking error, $F(2, 22) = 40.84$, $p < .001$. Planned comparisons indicated that subjects performed significantly better with first order than they did with first/second order tracking, $F(1, 11) = 5.64$, $p < .05$. Performance was also better in the first/second order condition than it was during second order tracking, $F(1, 11) = 8.58$, $p < .05$. Thus, our manipulation of tracking difficulty was successful in influencing subjects' performance in single- and dual-task conditions.

A significant main effect was obtained for tracking condition, $F(6, 66) = 35.44$, $p < .001$. Furthermore, the Tracking Condition × System Order interaction was also significant, suggesting greater differences between tracking conditions at higher levels of system order, $F(6, 66) = 4.24$, $p < .01$. Planned comparisons indicated that tracking error was significantly lower in the uncorrelated dual-task conditions when the secondary task involved counting changes of the cursor than when subjects were required to perform the secondary task by counting changes in the horizontal bar, $F(1, 11) = 7.1$, $p < .05$.

The differential effect of the type of secondary task object on tracking performance may be due to the relation of the objects to the primary task and is consistent with the task integration hypothesis as set forth in Figure 1. The cursor is clearly a necessary component of the tracking task, whereas the horizontal bar is not necessary for primary task performance. Therefore, subjects may find it more difficult to track and count probes if the probes are extraneous to the tracking task than if the probes occur within the primary task stimuli. If this interpretation is correct, we would expect that integration of the two tasks, achieved by correlating events in the secondary task with events in the primary task, would reduce the differences in RMS error between the two conditions. A comparison of the correlated and uncorrelated dual-task pairs supports this interpretation. The difference in RMS error between the horizontal bar and cursor conditions was eliminated when the primary and secondary

Table 1
Mean RMS Tracking Error Averaged Across Subjects for Single-Task and Dual-Task Tracking Conditions

	System order					
	First		First/second		Second	
Tracking condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Single-task	159	8	175	14	209	21
Dual-task (uncorrelated)						
Cursor–same position	154	9	166	12	191	15
Cursor–different position	155	8	164	9	192	19
Bar–same position	169	11	187	14	223	24
Bar–different position	166	8	184	16	221	20
Dual-task (correlated)						
Cursor–same position	155	7	163	10	194	18
Bar–same position	157	9	164	13	195	16

tasks were correlated ($p > .60$). The RMS error scores obtained in the single-task tracking conditions were not significantly different from those in the dual-task conditions ($p > .30$).

Subjective Measures of Task Difficulty

Subjects rated the subjective difficulty of each block of trials on a bipolar scale from 1 (*easy*) to 7 (*difficult*). The average ratings of difficulty for each level of system order in the single- and dual-task conditions are presented in Table 2. The subjects' perception of difficulty increased with increases in the system order of the tracking task for all of the tracking conditions, $F(3, 33) = 44.39$, $p < .001$. Subjects rated the difficulty of the uncorrelated dual tasks higher when performing the secondary task with the horizontal bar than they did when counting the intensity or translational changes of the cursor, $F(1, 11) = 13.84$, $p < .01$. Ratings of difficulty in the dual-task horizontal bar conditions were also significantly higher than those obtained for the single-task tracking conditions, $F(1, 11) = 9.31$, $p < .01$. The cursor and bar conditions were judged to be equally difficult in the correlated dual-task blocks ($p > .40$). Thus, subjects' ratings of tracking difficulty are consistent with their overt performance, as measured by RMS tracking error.

The accuracy with which subjects counted the secondary task probes was not significantly affected by any of the experimental manipulations. Subjects' counting accuracy exceeded 97% in all of the experimental conditions.

Table 2
Mean Subjective Difficulty Ratings for Single-Task Tracking and Dual-Task Conditions

	System order		
Tracking condition	First	First/second	Second
Single-task	2.2	3.7	4.6
Dual-task (uncorrelated)			
Cursor–same position	2.4	3.8	4.7
Cursor–different position	2.3	3.6	4.8
Bar–same position	3.3	4.6	6.0
Bar–different position	3.2	4.6	5.9
Dual-task (correlated)			
Cursor–same position	2.6	3.7	4.8
Bar–same position	2.5	3.8	4.8

Note. 1 = easy; 7 = difficult.

Event-Related Brain Potentials

The treatment of the ERP data is divided into two sections. The first section examines the ERPs elicited by changes in the spatial position of the tracking target. The second section is concerned with the effects of the experimental manipulations on the ERPs elicited by the secondary task probes in the correlated and uncorrelated dual-task conditions.

Primary task events. Figure 5 presents the ERPs elicited by changes in the spatial position of the tracking target in the dual-task conditions for the parietal recording site. It is evident that the ERPs differ in the amplitude of the positive component as the difficulty of the primary task is varied. This amplitude difference appears as early as 350 ms after the stimulus and continues to the end of the recording epoch. Across all conditions, it appears that the largest positivity is elicited when tracking is the most difficult, a trend that replicates the basic finding of Wickens, Kramer, Vanasse, and Donchin (1983).

The single-trial ERPs acquired during dual-task performance were averaged separately for each of the experimental conditions. Each of the average ERPs was composed of 50 to 60 single trials. The amplitude of the positive component was expressed as the difference between the maximum positive deflection occurring between 350 ms and 550 ms after the presentation of the relevant stimulus and the baseline, which was defined as the average voltage recorded over the 100-ms epoch just preceding the stimulus (Donchin & Heffley, 1979).

It is customary to define ERP components in terms of their latency relative to a stimulus or response, scalp distribution, and sensitivity to experimental manipulations (Donchin, Ritter, & McCallum, 1978). The deflection in the waveform becomes increasingly positive from the Fz to the Pz electrode, $F(2, 22) = 115.08$, $p < .001$, and the base to peak measures were maximal in the epoch associated with P300 (350 ms to 550 ms). Based on these criteria the positive deflection can be identified as the P300 (Donchin, Kramer, & Wickens, 1982; Sutton & Ruchkin, 1984).

A repeated measures analysis of variance of these amplitudes indicated that the P300 was influenced by the system order of the tracking task. Increases in system order were associated with increases in the amplitude of the P300 component, $F(2, 22) = 39.14$, $p < .001$. Planned comparisons indicated that the P300s elicited in the second order tracking conditions were significantly larger than those obtained in the first/second order conditions, $F(1, 11) = 6.96$, $p < .05$. The P300s elicited in the first/second order conditions were larger than those obtained during first order tracking, $F(1, 11) = 5.14$, $p < .05$. Thus, consistent with previous research, the amplitude of the P300s elicited by discrete changes in a primary task increases with increases in the difficulty of that task. None of the other main effects or interactions attained statistical significance for the primary task P300s. Because the P300 component of the ERP represents the major focus of the experimental hypotheses, other components will not be discussed in the present article (see Kramer, 1984, for details).

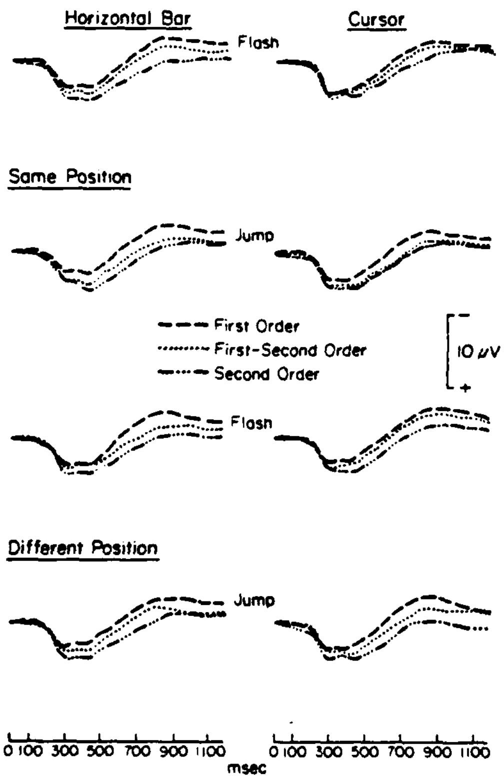


Figure 5. Grand average parietal event-related brain potentials elicited by changes in the spatial position of the tracking target in the dual-task blocks.

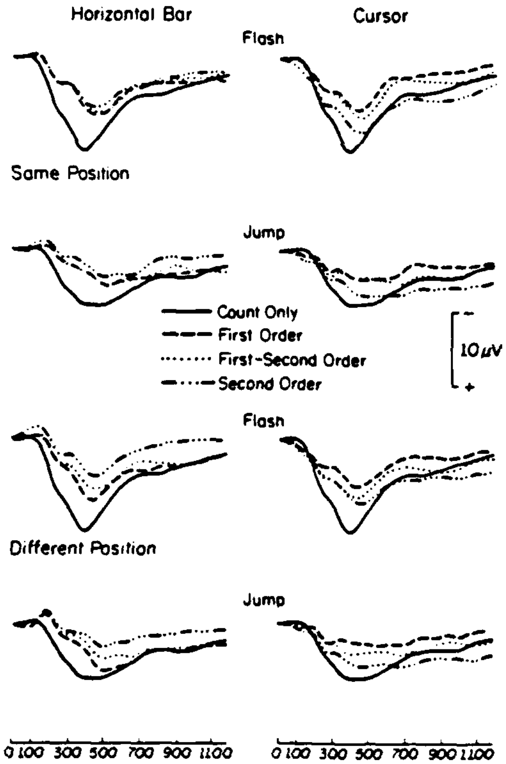


Figure 6. Grand average parietal event-related brain potentials elicited by the secondary task probes during the performance of the pursuit step tracking task.

Secondary task probes: Uncorrelated dual tasks. Figure 6 presents the average parietal ERPs elicited by the secondary task probes during the performance of the step tracking task. Several aspects of the waveforms are noteworthy. In all of the experimental conditions, the single-task count-only block yields an ERP characterized by a large positive deflection with a latency of 400 ms. This positive deflection has been identified as the P300 component according to the criteria enumerated above.[2] The three levels of system

[2] It may be confusing to those not familiar with the ERP literature that the label *P300* is attached to deflections whose latency is often quite a bit longer than 300 ms. This seems to belie the implication of the label that the

order elicit varying degrees of positivity that appear to depend on the particular experimental condition. For example, for all experimental conditions in which the secondary task probe is the cursor, the positive deflection is largest for the second order condition, of intermediate amplitude in the first/second order condition, and smallest in amplitude in the first order condition, $F(2, 22) = 28.1$, $p < .001$, a trend similar to that observed for the P300s elicited by the primary task probes, as illustrated in Figure 5. This ordering of tracking difficulty conditions does not appear to be influenced by the position of the secondary task probe relative to the tracking task or by the type of discrimination required of the subject.

In the two conditions in the lower left of Figure 6 in which the horizontal bar is counted and is located below the tracking task, the sequence of the ERPs elicited by different levels of system order is clear and consistent. However, the order is the inverse of that obtained in the cursor conditions. The first order tracking condition elicits the largest positivity, the first/second order condition elicits an intermediate level of positivity, and the second order condition produces the smallest amplitude, $F(2, 22) = 24.2$, $p < .001$. This trend in P300 amplitude is typical of secondary task probe stimuli (Isreal, Wickens, Chesney, & Donchin, 1980; Natani & Gomer, 1981). In the two conditions in which the horizontal bar is superimposed on the tracking task, the ERPs elicited by different levels of system order are not significantly different from each other. Finally, the pattern of effects shown by the flashes were in all respects identical to those shown by the jumps.

The opposite effect of system order on P300 amplitude was predicted for the cursor and horizontal bar conditions. This prediction derives from the resource structure inferred from Figure 1. We hypothesized that if two tasks required that the subjects process different properties of the same object, then the resource structure of the two tasks would be similar. The direct relation between P300 amplitude and system order for the primary task events and cursor probes is consistent with this hypothesis. It was also argued that if two tasks required the processing of different objects and these tasks overlapped in their resource demands as defined by the multiple resource model, then the relation between P300 amplitude and system order would be reciprocal between primary and secondary tasks. This hypothesis was confirmed with the dual-task combination of the tracking task and horizontal bar at different spatial locations. Thus, the results obtained in the present study are consistent with both hypotheses concerning the resource structure of dual tasks. When two tasks require the processing of different properties of the same object, then the amplitude of the P300s elicited by stimuli associated with each task will change in the same direction with changes in system order. If, on the other hand, the two tasks require the processing of different objects, then as the amplitude associated with one increases, the amplitude associated with the other will decrease. That is, we will obtain a reciprocity in P300 amplitudes for concurrent tasks that require the processing of different objects.

The ERPs elicited by the probes used during Session 1 served as a baseline for the secondary task probes used in the later experimental sessions. In the first session, subjects were instructed to ignore the probes and concentrate on performing the tracking task. Thus, the P300s elicited by the probes in Session 1 provide an index of subjects' ability to ignore extraneous stimuli while performing a task. Figure 7 presents the average parietal ERPs elicited by the probes during the performance of the tracking task in the practice session. It is clear from a comparison of the waveforms shown in Figures 6 and 7 that uncounted probes elicit P300s of a relatively small size. Indeed, the waveforms presented in Figure 7 present no evidence of a P300. Furthermore, the waveforms elicited by the uncounted probes did not discriminate among levels of tracking difficulty in any of the experimental conditions.

This result confirms our prediction that the ignored stimulus properties will not elicit a P300. This effect is obtained regardless of the relation of the probe stimuli to the primary task objects. Therefore, based on the

peak occurs at 300 ms. The problem arises because the component seems to reflect activity triggered at the completion of certain processing activities. Thus, the triggering event is internal and results in a variable latency to the external, eliciting event.

P300 amplitude measure, it appears that subjects are capable of directing their attention to one property of an object while ignoring another property of the same object (for additional evidence see Donchin & Cohen, 1967; Heffley, Wickens, & Donchin, 1978; Kramer et al., 1983).

Secondary task probes: Correlated dual tasks. The analysis and discussion of the ERPs elicited by the secondary task probes have thus far been concerned with dual tasks that are uncorrelated. Can we expect the relations observed with uncorrelated dual tasks to generalize to situations in which the events in one task predict the events in the other task with some degree of certainty? What effects will intertask correlation have on the resource structure of the two tasks? These questions are examined in the present section by analyzing the effects of intertask correlation on the relation between P300 amplitude and system order.

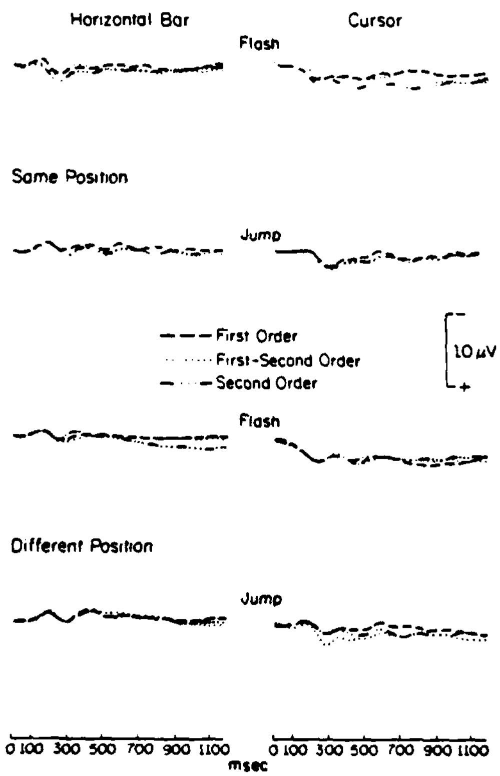


Figure 7. Grand average parietal event-related brain potentials elicited by the probes during the performance of the pursuit step tracking task in the first session. (Subjects were instructed to ignore the probes in this session.)

Figure 8 presents the average parietal ERPs elicited by the correlated and uncorrelated dual-task conditions when the cursor and bar are flashed. There are several interesting aspects of these waveforms. A comparison of the ERPs elicited in the correlated and uncorrelated cursor probe conditions suggests that system order has the same effect on the ERPs in both conditions. The ERPs elicited by the cursor probes during second order tracking are characterized by a large P300. In the first/second order condition, the cursor probes elicited a P300 of an intermediate amplitude, and in the first order condition the P300s were smallest in amplitude, $F(2, 22) = 10.9$, $p < .001$.

An examination of the waveforms elicited by the horizontal bar probes presents a different picture. As noted previously, the effect of system order is not significant in the uncorrelated horizontal bar condition. However, the ERPs elicited in the correlated horizontal bar condition increase in positivity with increases in system order, $F(2, 22) = 12.3$, $p < .001$. Thus, it appears that the effect of system order on the ERPs is the same across the two cursor conditions and the correlated horizontal bar condition. Correlating the tracking and probe events performs the same integrating function on the processing as that accomplished by combining them into a common object.

These results indicate that when two tasks require the processing of the same object, as was the case for the uncorrelated dual-task cursor condition, an increase in the correlation between the tasks does not have a large effect on the resources allocated to the tasks. The relation between P300 amplitude and system order was not significantly different in the correlated and uncorrelated dual-task conditions. Thus, when the two tasks require the processing of different properties on the same object, the processing of the tasks is in some sense integrated and intertask correlation does not enhance this integrality further. However, when two concurrently performed tasks require the processing of separate objects, as was the case in the horizontal bar conditions, the presence of intertask correlation does appear to enhance the integrality between tasks. This increase in dual-task

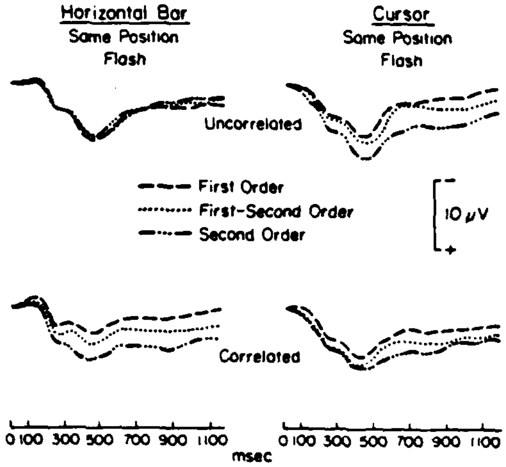


Figure 8. Grand average parietal event-related brain potentials elicited by the secondary task probes in the correlated and uncorrelated dual-task conditions.

integrality is inferred from the change in the relation between P300 amplitude and system order in the correlated and uncorrelated horizontal bar conditions. The P300 amplitude changes with system order in the correlated condition in the same manner that it does when P300 is elicited by primary task events, suggesting an overlap in the resource structure between tasks.

Discussion

In most dual-task combinations, increasing the difficulty of one task is assumed to consume resources that normally would have been used in the processing of the other task. The resources shared by these two tasks are presumed to be reciprocal in nature. Thus, increasing the difficulty of one task leads to a decrement in performance on one or both of the concurrently performed tasks. In other cases, the two tasks require different processing resources and therefore do not result in resource trade-offs. These tasks are generally performed as well together as they are alone (Navon & Gopher, 1979; Wickens, 1984). Under conditions of dual-task integrality, one task increases processing demands within the domain of the other task. Therefore, in the case of dual-task integrality, resource reciprocity is not obtained, although both tasks require the same resources. Dual-task integrality results in a facilitation in performance in one or both tasks when performed concurrently. Facilitation is relative to conditions in which the two tasks are performed separately or when the stimulus relations but not the processing requirements change between dual-task pairs.

Dual-task integrality is operationally defined in the current context as occurring when the amplitude of the P300s elicited by secondary task probes increases with increases in the difficulty of the primary task. Four experimental variables were manipulated, each intended to foster increasing degrees of integrality between the primary and secondary tasks. The pattern of data allows the ordering of these variables in terms of the degree to which they fostered integrality. In discussing the data, reference is made to Figure 9, in which P300 amplitude in each condition is shown as a function of the system order of the tracking task.

First and most consistent are the effects of the object properties on dual-task integrality as inferred from changes in the amplitude of P300. When the relevant stimuli associated with both tasks were incorporated in the primary task object, integrality was observed at its maximum value. The P300s elicited by secondary task events increased in amplitude with increases in the difficulty of the primary task. Given that the two tasks required the processing of different properties of the same object, neither a change in the specific properties (spatial or intensity) nor a change in the correlation between tasks could alter the high degree of integrality. Furthermore, the object-derived benefit was also reflected by the RMS error data. Tracking performance was superior when the two tasks required the processing of different properties of a single object as compared with the processing of separate objects. These results are consistent with previous findings suggesting that different properties of an object tend to be processed in parallel (Kahneman & Henik, 1981; Lappin, 1967; Treisman, 1977). The important knowledge added by the present study concerns the direct measure of resource investment and the characteristic that reciprocity is defined here in terms of a resource-demand manipulation and not just an absolute performance level.

Second, and equally strong, is the effect of correlation on dual-task integrality. When the two tasks are correlated, integrality is shown. The P300s elicited by secondary task events that are correlated with events in the primary task increase in amplitude with increases in the difficulty of the primary task. When events in the two tasks are not correlated and the tasks require the processing of different objects, integrality is lost and reciprocity is sometimes shown. There are several reasons why correlation may produce integrality. Again, the concept of an object may underlie this effect. Different properties of a single object are typically correlated as we experience them in the real world. So, turning this around, the correlation of stimuli may foster object perception and, hence, dual-task integrality. Garner and co-workers found that the processing of integral stimulus dimensions is enhanced when the dimensions are correlated, presumably because the integral dimensions function as a single unit (Garner, 1969; Garner & Felfoldy, 1970). In the present study it appears that two tasks that are correlated also seem to function as a unit and therefore benefit from the redundancy.

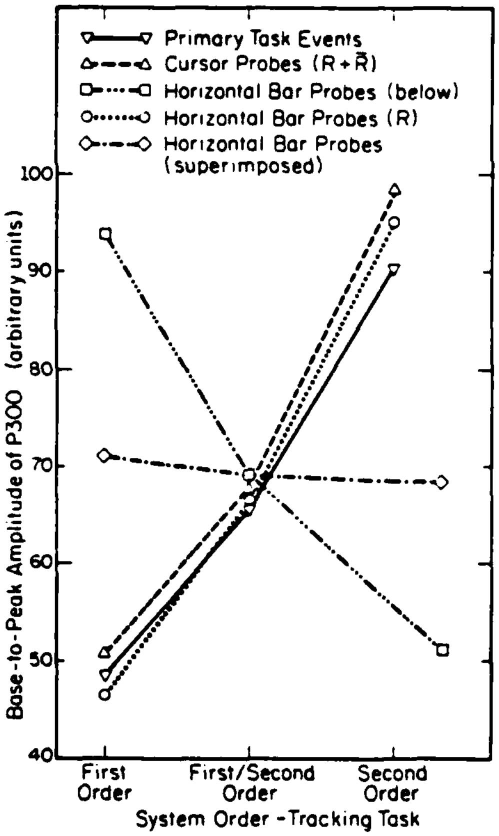


Figure 9. A graphic summary of the P300 results. (Amplitude of the P300s as a function of primary task difficulty is reported for each of the experimental manipulations. R = the correlated conditions.)

Third, spatial location fosters integrality, although to a lesser extent than the properties of an object or the correlation between tasks. Of course if the two tasks require the processing of different properties of a single object, this guarantees a common spatial location. However, even when there were different primary and secondary task objects (horizontal bar probes), we found that locating them together in space, though it did not produce integrality, still reduced the level of reciprocity so that the P300 function was flat. Again, returning to real world experience, it is true that the properties of an object are typically close together in space; however, proximity does not guarantee integrality. The ease with which subjects can focus on some information at a location in space while completely ignoring other information at the same location has been demonstrated in several experiments (Donchin & Cohen, 1967; Fischer, Haines, & Price, 1980; Neisser & Becklen, 1975).

Thus far we have argued that when two tasks require the processing of different properties of a single object, integrality is observed. Other investigators have also found that it is difficult to selectively attend to one property of an object while ignoring other properties (Kahneman & Chajczyk, 1983; Stroop, 1935). This seems to be especially true if the two properties are integral, in the sense that for one property to be realized, there must be a level specified on the other property (Garner, 1970). In the first session of the present study, subjects were instructed to perform the tracking task and ignore the extraneous probes. The probes were changes in properties of the primary task objects that were not necessary for tracking performance. These probes became the secondary task stimuli in the later experimental sessions. These ignored probes did not elicit a P300 component. However, P300s were elicited by the probes when they represented a secondary task. Although the presence or absence of the P300 does not in and of itself indicate the success or failure of selective attention, it does provide information concerning the amount of task-related processing (Hillyard & Kutas, 1983; Picton, Campbell, Baribeau-Braun, & Proulx, 1978). The P300 results suggest that the task-relevant properties of the primary task objects were processed to a greater extent than the irrelevant properties. Thus, it appears that when the properties of an object are highly discriminable, subjects are capable of selectively attending to specific properties of an object while ignoring others.

The resource framework inferred from the P300 provides a theoretical account of the effect of several factors on the phenomenon of dual-task integrality. The results also have practical implications. The P300 component has been used as a measure of cognitive workload. The P300s elicited by secondary task stimuli decrease in amplitude with increases in primary task difficulty; P300s elicited by discrete primary task events increase in amplitude with increases in the difficulty of the primary task. The resources allocated to tasks have been inferred from changes in P300 amplitude. The results obtained in the present study suggest that the reciprocal relation between the primary and secondary tasks depends on the structure of the dual task. For example, the relation between P300 amplitude and task difficulty changes from the case in which the two tasks require the processing of different properties of the same object to the situation in which the two tasks necessitate the processing of different objects. Furthermore, intertask correlation and the physical proximity of task displays also have a significant effect on the resource structure of the dual-task pair. These findings suggest that a reliable analysis of the processing demands of a task can only take place within a theoretical framework. The concept of dual-task integrality offers one such framework.

References

Alwitt, L. F. (1981). Two neural mechanisms related to modes of selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 324–332.

Donchin, E. (1975). Brain electrical correlates of pattern recognition. In G. F. Inbar (Ed.), *Signal analysis and pattern recognition in biomedical engineering* (pp. 199–218). New York: Wiley.

Donchin, E., & Cohen, L. (1967). Average evoked potentials and intramodality selective attention. *Electroencephalography and Clinical Neurophysiology, 22*, 537–546.

Donchin, E., & Heffley, E. (1979). Multivariate analysis of event related potential data: A tutorial. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related potential research* (Report No. EPA 600/9-77-043, pp. 555–572). Washington, DC: U.S. Government Printing Office.

Donchin, E., Kramer, A., & Wickens, C. (1982). Probing the cognitive infrastructure with event-related brain potentials. In M. L. Frazier & R. C. Crombie (Eds.), *Proceedings of the workshop on flight testing to identify pilot workload and pilot dynamics* (Report No. AFFTC-TR-82-5, pp. 371–387). Edwards Air Force Base, CA: U.S. Air Force.

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. 349–412). New York: Academic Press.

Duncan, J. (1984). Selective attention and the organization of visual attention. *Journal of Experimental Psychology: General, 113*, 501–517.

Fischer, E., Haines, R., & Price, T. (1980). *Cognitive issues in head-up displays* (Technical Paper No. 1711). Washington, DC: National Aeronautics and Space Administration.

Garner, W. R. (1969). Speed of discrimination with redundant stimulus attributes. *Perception and Psychophysics, 6*, 221–224.

Garner, W. R. (1970). The stimulus in information processing. *American Psychologist, 25*, 350–358.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing tasks. *Cognitive Psychology, 1*, 225–241.

Garner, W. R., & Flowers, J. H. (1969). The effect of redundant stimulus elements on visual discriminations as a function of element heterogeneity, equal discriminability, and position uncertainty. *Perception and Psychophysics, 6*, 216–220.

Gratton, G., Coles, M. G. H., & Donchin, E. (1982). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55*, 468–484.

Heffley, E., Wickens, C. D., & Donchin, E. (1978). Intramodality selective attention and P300: A reexamination in a visual monitoring task. *Psychophysiology, 15*, 269–270.

Hillyard, S. A., & Kutas, M. (1983). Electrophysiology of cognitive processing. *Annual Review of Psychology, 34*, 33–61.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. *Psychophysiology, 17*, 259–273.

Isreal, J. B., Wickens, C. D., Chesney, G. L., & Donchin, E. (1980). The event-related brain potential as an index of display monitoring workload. *Human Factors, 22*, 211–224.

James, W. (1890). *Principles of psychology* (Vol. 1). New York: Henry Holt.

Jasper, H. H. (1958). The ten twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10*, 371–375.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D., & Chajczyk, D. (1983). Tests of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 497–509.

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 181–209). Hillsdale, NJ: Erlbaum.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 29–62). New York: Academic Press.

Kahneman, D., Treisman, A., & Burkell, J. (1983). The cost of visual filtering: A new interference effect. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 510–522.

Kramer, A. F. (1984). *The processing of stimulus attributes: Evidence for dual-task integrality.* Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Kramer, A., Wickens, C. D., & Donchin, E. (1983). An analysis of the processing demands of a complex perceptual-motor task. *Human Factors, 25*, 597–622.

Lappin, J. S. (1967). Attention in the identification of stimuli in complex visual displays. *Journal of Experimental Psychology, 75*, 321–328.

Lockhead, G. H. (1966). Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psychology, 72*, 95–104.

Natani, K., & Gomer, F. E. (1981). *Electrocortical activity and operator workload: A comparison of changes in the electroencephalogram and in event-related potentials* (Tech. Rep. No. MDC E2427). St. Louis, MO: McDonnell Douglas Astronautics Company.

Navon, D., & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review, 86*, 214–255.

Neisser, U., & Becklen, R. (1975). Selective looking: Attention to visually specified events. *Cognitive Psychology, 7*, 480–494.

North, R. A. (1977). *Task components and demands as factors in dual-task performance* (Rep. No. ARL-77-2/AFOSE-77-2). Unpublished report, University of Illinois at Urbana-Champaign, Aviation Research Laboratory.

Picton, T. W., Campbell, K. B., Baribeau-Braun, J., & Proulx, G. B. (1978). The neurophysiology of human attention: A tutorial review. In J. Requin (Ed.), *Attention and performance VII* (pp. 429–468). Hillsdale, NJ: Erlbaum.

Polich, J., McCarthy, G., Wang, W. S., & Donchin, E. (1983). When words collide: Orthographic and phonological interference during word processing. *Biological Psychology, 16*, 155–180.

Pomerantz, J. R., & Garner, W. R. (1973). Stimulus configuration in selective attention tasks. *Perception and Psychophysics, 18*, 355–361.

Regan, D. (1972). *Evoked potentials in psychology, sensory physiology and clinical medicine.* London: Chapman & Hall.

Reicher, G. M. (1977). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81*, 275–280.

Rolfe, J. M. (1971). The secondary task as a measure of mental load. In W. T. Singleton, J. G. Fox, & P. Witfield (Eds.), *Measurement of man at work.* London: Taylor & Francis.

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme matching. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 546–554.

Sirevaag, E., Kramer, A. F., Coles, M. G. H., & Donchin, E. (1984). P300 amplitude and resource allocation. *Psychophysiology, 21*, 598–599.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.

Sutton, S., & Ruchkin, D. S. (1984). The late positive component: Advances and new problems. In R. Karrer, J. Cohen, & P. Tueting (Eds.), *Brain and information: Event-related potentials* (pp. 1–23). New York: New York Academy of Sciences.

Treisman, A. (1977). Focused attention in the perception and retrieval of multidimensional stimuli. *Perception and Psychophysics, 22*, 1–11.

Trumbo, D., Noble, M., & Swink, J. (1967). Secondary task interference in the performance of tracking tasks. *Journal of Experimental Psychology, 73*, 232–240.

Vidulich, M., & Wickens, C. D. (1981). *Time-sharing manual control and memory search: The joint effects of input and output modality competition, priorities and control order* (Report No. ERL-81-4/ONR-81-4). Unpublished report, University of Illinois at Urbana-Champaign, Engineering Research Laboratory.

Weisstein, N., & Harris, C. S. (1974). Visual detection of line segments: An object superiority effect. *Science, 186*, 752–755.

Wickens, C. D. (1979). Human workload measurement. In N. Moray (Ed.), *Mental workload: Its theory and measurement* (pp. 79–100). New York: Plenum Press.

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 239–258). Hillsdale, NJ: Erlbaum.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 63–102). New York: Academic Press.

Wickens, C. D., Derrick, W. D., Micallizi, J., & Beringer, D. (1980). The structure of processing resources. In *Proceedings of the 24th Annual Meeting of the Human Factors Society* (pp. 253–256). Santa Monica, CA: Human Factors Society.

Wickens, C. D., & Kessel, C. (1979). The effects of participatory mode and task workload on the detection of dynamic system failures. *IEEE Transactions on Systems, Man and Cybernetics, SMC-13*, 21–31.

Wickens, C. D., Kramer, A., Vanasse, L., & Donchin, E. (1983). The performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information processing resource. *Science, 221*, 1080–1082.

Wickens, C. D., Sandry, D. L., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors, 25*, 227–248.

Received December 15, 1984
Revision received March 5, 1985 ■

Resource Reciprocity: An Event-Related Brain Potentials Analysis

Erik Sirevaag, Arthur Kramer, Michael G.H. Coles, Emanuel Donchin
University of Illinois
Department of Psychology
Champaign, Illinois

Abstract

The amplitude of the P300 component of the Event-Related Potential (ERP) has proven useful in identifying the resource requirements of complex perceptual-motor tasks. In dual-task conditions, increases in primary task difficulty result in decreases in the amplitude of P300s elicited by secondary tasks. Furthermore, P300s elicited by discrete primary task events increase in amplitude with increases in the difficulty of the primary task. The reciprocity in P300 amplitudes has been used to infer the processing tradeoffs that occur during dual-task performance. The present study was designed to investigate further the P300 amplitude reciprocity effect under conditions where primary and secondary task ERPs could be concurrently recorded within the same experimental situation. Forty subjects participated in the study. Measures of P300 amplitude and performance were obtained within the context of a pursuit step tracking task performed alone and with a concurrent auditory discrimination task. Task difficulty was manipulated by varying both the number of dimensions to be tracked (from one to two), and the control dynamics of the system (velocity or acceleration). ERPs were obtained from both secondary task tones and primary task step changes. Average root-mean-square (RMS) error estimates were also obtained for each tracking condition. The data indicated that increased primary task difficulty, reflected in increased RMS error scores, was also associated with decreased secondary task P300 amplitudes and increased primary task P300 amplitudes. Since the increases in primary task P300 amplitudes were complementary to the decrements obtained for the secondary task, the hypothesis of reciprocity between primary and secondary task P300 amplitudes was supported across several different manipulations of primary task difficulty.

Additional research is currently under way to examine a variety of additional issues. If resources are a limited commodity which can be flexibly allocated to the performance of a given task, it should be possible to manipulate policies of resource allocation during dual task performance as a function of instructional set as well as task difficulty. Thus, if a subject is instructed to pay more attention to one task than another, P300 amplitude should reflect an increased allocation of resources to the attended task. Another area of interest concerns the question of dual task integrality as opposed to reciprocity. In other words, under what conditions will two tasks be combined to produce a new "single" task (integrality) instead of competing for resources as in the present study (reciprocity).

KEYWORDS: EVENT-RELATED BRAIN POTENTIALS, P300, RESOURCES, MENTAL WORKLOAD, DUAL-TASK PERFORMANCE

Introduction

This study is concerned with the measurement of mental workload. Two different measures will be utilized to assess policies of resource allocation between two concurrently performed tasks. These measures, electrophysiological concomitants of resource allocation, and performance on an assigned task, were obtained during task conditions differing in difficulty. The workload related to various manipulations of difficulty was assessed by dual-task techniques (e.g. Knowles, 1963; Rolfe, 1971; Brown, 1978).

Modern theories of workload assume that attention can be likened to the limited processing capacity of a general purpose computer (Moray, 1967; Kahneman, 1973; Norman and Bobrow, 1975). The allocation of this limited processing commodity to the performance of a given task is determined by the motivation of the operator and the demand characteristics of the task. The latter can, in turn, be manipulated either by changing the nature of the task (e.g. reducing stimulus discriminability, changing the pacing of the task), or by varying the level of performance required from subjects engaged in the task. Thus, workload is a hypothetical construct that refers to the interaction between task, subject, and circumstances (Gopher and Donchin, in press). Measures of workload must be sensitive to these interactions.

Measures of task performance are not adequate workload metrics for two reasons. (1) They cannot be generalized across tasks that require different performance measures. (2) Operators may change their strategies so that they maintain a constant performance level even though the workload is increased. Several alternative techniques have therefore been proposed for workload assessment (for reviews see Wickens, 1984; Gopher and Donchin, in press; Williges and Wierwille, 1979).

A common approach to workload measurement is the dual-task technique (Brown, 1978; Knowles, 1963; Rolfe, 1971). This technique views the measurement of workload as the estimation of the amount of resources drawn from a limited resource pool by the task whose workload is being assessed. If the limited resources are allocated according to task demands and performance requirements, then two tasks performed concurrently must compete for these limited resources. Furthermore, if subjects are instructed to optimize their performance on one task, fewer resources will be available for the other task. Increases in primary task workload entail the allocation of a larger share of the resources to the primary task. Inevitably, these resources are no longer available to the secondary task and performance on the secondary task deteriorates. The use of the dual-task paradigm in workload assessment has been validated and refined in a variety of different conditions (Allport, Antonis, and Reynolds, 1972; Kahneman, 1973; Navon and Gopher, 1979; Norman and Bobrow, 1975; North, 1977; Schaffer, 1975; Schneider and Fisk, 1981; Sperling and Melchner, 1978; Wickens, 1976).

Yet it became evident, as research progressed, that the concept of an undifferentiated resource pool (Kahneman, 1973) cannot account for the complex of patterns and results. To account for fact that different combinations of tasks yield different estimates of workload, Wickens (1980) proposed a multiple resource model in which resources are defined in terms of three dichotomus dimensions: stages of processing (perceptual/cognitive vs. response selection processes); codes of processing (spatial and verbal); and modalities of processing (auditory vs. visual). However, even by utilizing a multiple resource theory for task design and analysis, there are several practical problems in implementing a dual-task study (Brown, 1978; Ogden, Levine, and Eisner 1979). In particular, it has proven difficult to design secondary tasks that are sensitive to the resource demands of the primary task, yet do not intrude upon its performance. This is especially difficult when the secondary task requires an overt motor response.

As a consequence, several investigators suggested the use of secondary tasks which allow the use of psychophysiological measures and which do not require the subject to respond overtly to a secondary task. Psychophysiological measures that index arousal have been used in workload assessment for some time (Berlyne, 1960; Howitt, 1968; Roscoe, 1978; Wierwille, 1979). However, it is also possible to use psychophysiological measures that are affected by cognitive rather than affective factors. Variance in these measures can play a role identical to that played by overt-response measures in the dual-task paradigm. These procedures rely on the endogenous components of the Event-Related Brain Potential (ERP), which is sensitive to specific aspects of the information processing sequence, within the framework of the secondary task paradigm (Isreal, Wickens, Chesney, and Donchin, 1980; Isreal, Chesney, Wickens, and Donchin, 1980; Kramer, Wickens, and Donchin, 1983). It is this approach which is employed in the present study.

Event-Related Brain Potentials and Information Processing The ERP is a component of electroencephalographic (EEG) activity which is time-locked to an event (Donchin, Ritter, and McCallum, 1978). By averaging over records which follow repetitions of the event, background activity unrelated to the processing of the event diminishes while the time-locked activity is enhanced. The P300 component of the ERP is represented by a positive voltage deflection maximal over the parietal scalp with a minimal latency of 300 msec. (Sutton, Braren, Zubin, and John, 1965).

Given that the amplitude of the P300 is proportional to the extent to which a subject utilizes the information provided by a stimulus (Johnson and Donchin, 1978, 1982; Duncan-Johnson and Donchin, 1977, 1978; Donchin, Kubovy, Kutas, Johnson, and Herning, 1973), it seemed reasonable to propose that this component may serve as an index of task relevance and can therefore be used to infer the tradeoffs of processing resources presumed to underlie the performance of dual-tasks.

A series of dual-task studies (reviewed by Donchin, Kramer, and Wickens, 1982) provided evidence that the amplitude of the P300 varies as a function of cognitive workload. In these studies, P300s elicited by tones associated with a secondary task requiring auditory discriminations decreased in amplitude with increases in the perceptual/cognitive difficulty of concurrently performed primary tasks. In effect, this series of studies demonstrated that the amplitude of P300 elicited by secondary task stimuli is decreased as the demands placed on subjects by the primary tasks are increased. While these results are consistant with the model, there was one implication of the model that had not been tested. We refer to the fate of the P300 associated with the primary task.

If P300 amplitude does, in fact, reflect the resource tradeoffs that occur during dual-task performance, then P300s elicited by primary task events should increase in amplitude with increases in the workload of the primary task. An even stronger prediction is that in dual-task studies in which ERPs can be recorded in response to discrete primary <u>and</u> secondary task events, there should be a reciprocal relationship between primary and secondary task P300 amplitudes. This prediction is implied by the limited capacity model which assumes that a fixed pool of resources is fully allocated between the two tasks. Note that because resources are not directly observable, we cannot directly confirm this "resource reciprocity" hypothesis. However, if it is assumed that P300 amplitude is proportional to the amount of resources allocated to the tasks we predict a reciprocity in primary and secondary task P300s. Thus, as additional resources are allocated to the primary task, secondary task P300s should decline and primary task P300s should increase in amplitude.

The amplitude reciprocity hypothesis was tested by Wickens, Kramer, Vanasse, and Donchin (1983) who required subjects to track a discretely changing target with a cursor . The ERPs elicited by the step changes of the primary task were digitized and recorded in one experimental run; while those elicited by the tones counted during the secondary task were recorded in a separate session. Task demands were varied by manipulating the number of time integrations between the joystick output and the movements of the cursor on the screen (so that the dynamics of the system changed from "velocity" to "acceleration" control). The data confirmed that P300s associated with the step changes increased in amplitude with increased primary task difficulty; while secondary task P300 amplitude decreased in a complementary manner.

While these data did confirm resource reciprocity, we considered it necessary to test the resource-reciprocity hypothesis with ERPs that were elicited by primary and secondary task events recorded within the <u>same</u> block of trials. In other words, the case for amplitude reciprocity would be stronger if a reciprocal relationship between <u>concurrently</u> recorded primary and secondary task ERPs is found. We report here a study that permits this comparison.

Previous research has indicated that while P300 amplitude is sensitive to increases in the system order of a tracking task (Kramer, Wickens, and Donchin, 1983), manipulations of the number of dimensions in which a subject is required to track produce no changes in secondary task P300 amplitude (Wickens, Isreal, and Donchin, 1977). Therefore, an orthogonal manipulation of dimensionality and system order should provide conditions with varying degrees of primary and secondary task resource competition.

A step tracking task was developed in which subjects performed in four conditions (2 system orders x 2 dimensions) within the context of both single and dual-task instructions (i.e. the presence or absence of a concurrent auditory discrimination task). This design provides a unique opportunity to examine whether the reciprocity of P300 amplitude can be demonstrated in a study involving orthogonal combinations of dependent variables and concurrently recorded primary and secondary task ERPs.

Method

Subjects Forty dextral males between the ages of 18 and 25 were paid for their participation in the study. None of the subjects had any previous experience with the step tracking task. All subjects had normal hearing and normal or corrected to normal vision.

Tasks As a primary task, subjects tracked a target by moving a cursor on the display screen. The target and cursor were both square (.5 x .5 cm) and were displayed with equal intensity on a Cathode Ray Tube (CRT) 1.2 m from the subject. Movements of the target square were under computer control. The targets moved in discrete jumps to random positions on the CRT with an average inter-move interval of 3.8 sec. The jumps could occur either solely in the horizontal or in both the horizontal and the vertical dimensions depending on the experimental condition. The sequence of jumps was constrained so that an equal number of jumps to the left and right, as well as up and down, were executed in a given block. While changes in the spatial position of the target occured in discrete steps, the subject was required to exercise continuous control over the joystick to cancel the error between the tracking elements. In each dual-task block the target changed position approximately 60 times.

Subjects controlled the position of the cursor on the screen by manipulating a joystick with their right hand. The dynamics of the system response to movements of the joystick were determined by the following equation:

$$X(T)=[(1-A)\int U(T)\,dt]+[(A)\iint U(T)\,dt]$$

where, U=stick position;
T=time;
A=contribution of the second order component.

The value of "A" was varied across experimental conditions. When A=0, the system is a pure first-order system in which movements of the stick increase or decrease the velocity with which the cursor moves. This will be referred to as the velocity condition. When A=95, the system includes a second-order component. That is, the joystick controlled the acceleration of the cursor. In this "acceleration" condition, it is considerably more difficult to achieve control over the cursor's movements.

For the secondary task, the subjects monitored a Bernoulli sequence of auditory stimuli presented binaurally through TDH-39 headphones. The tones, each of which could either be low pitched (1200 Hz) or high pitched (1400 Hz) were selected at random; (p [high] = p [low] = .5 across blocks).

Within a given block, from 26 to 35 target tones were presented and the interval between individual tones averaged 3.8 sec. The duration of both tones was 60 msec (including a 10 msec rise/fall time). Since we were interested in ERPs to both step and tone stimuli, the presentation schedule for the tones was constrained so that the recording epochs of the tones and the step changes did not overlap.

Recording system Electroencephalographic (EEG) activity was recorded from three midline sites (Fz, Cz, and Pz according to the International 10-20 system: Jasper, 1958) referenced to linked mastoids. Two ground electrodes were attached to the forehead. The scalp and mastoid electrodes were Burden Ag-AgCl electrodes affixed with collodion. The vertical electro-oculogram (EOG) was recorded from Beckman Biopotential electrodes affixed with adhesive collars above and below the subject's right eye. Beckman electrodes were also used for the grounds. All electrode impedances were below 10 kohms/cm.

The EEG and EOG were amplified by Van Gogh model 50000 amplifiers with a 10 sec time constant and an upper half amplitude of 35 Hz, 3 db/octave rolloff. The recording epoch for both the EEG and EOG was 1280 msec and began 100 msec prior to either the primary task step changes or the secondary task tones. The data channels were digitized every 10 msec and were also filtered off-line (-3 db at 6.29 Hz) prior to further analysis.

Stimulus generation and data collection Presentation of the stimuli and collection of the data were under the control of a PDP 11/40 computer (see Donchin and Heffley, 1975). On line monitoring of both average and single trial EEG and EOG was accomplished by a GT-44 display. Contributions of the EOG to the EEG waveforms were evaluated and eliminated off-line by submitting the data to an eye-movement correction algorithm developed by Gratton, Coles, and Donchin (1983).

Procedure Two aspects of the tracking task could vary. The target could move in one or two dimensions. The control system could be either a "velocity" or an "acceleration" system. Thus, four different formats of the tracking task were obtained by crossing target-movement with control dynamics. These tasks could be performed either alone (single task conditions), or concurrently with the auditory discrimination task (dual-task conditions). Each of the forty subjects participated in all of the experimental conditions. Each condition lasted approximately 15 minutes and was followed by a short (2 min) break.

Following electrode placement, subjects were told that they were about to participate in a study to assess the effects of increased task difficulty under both single and dual-task conditions. Before receiving practice in the step tracking task, all subjects performed three blocks of the auditory discrimination task to familiarize themselves with the stimuli. Subjects then performed three single task tracking blocks as practice. The practice trials consisted of one block of velocity tracking in two dimensions, and two blocks of acceleration tracking (one in one dimension and one in two dimensions). After completing the practice blocks, subjects were instructed to assign the tracking task top priority. Thus, tracking was defined as the primary task. The subjects were told that while they should try to count tones accurately, their goal was to perform the tracking task as well as possible. The experimental blocks were then presented in the fixed order displayed in Table 1.

<Insert Table one about here>

Note that single task blocks always preceded dual-task blocks, and easier tracking conditions preceded difficult tracking conditions. The order displayed in Table 1 was chosen so that any learning due to practice should improve performance during the more difficult conditions. Thus, this conservative design will tend to diminish differences between easy and difficult tracking conditions rather than amplify them. On every single and dual-task block, data concerning tracking accuracy was collected for each trial by recording every 50 msec the distance between the subject controlled cursor and the target square and then computing the root-mean-square (RMS) error defined by these values. Errors in counting of the tones were also recorded during all dual-task conditions.

The ERPs were recorded in all conditions. Frontal, central, and parietal (Fz,Cz, and Pz) electrode outputs were digitized and recorded to the step changes during all single task tracking, and dual-task tracking conditions. ERPs associated with the tones in the auditory discrimination task were recorded during all dual-task tracking conditions.

Results

The primary and secondary task overt performance data will first be examined to assess the extent to which the variations in dimensionality and system order modulated the performance of the primary task, as well as to determine whether subjects did, in fact, protect the level of performance on the primary task even as the task demands increased. This is a critical observation if the dual-task methodology is to be applied. Secondary task performance decrements cannot be properly interpreted if subjects do not perform the primary task as well during dual-task conditions as they did during single task conditions. The RMS error measures will be used to define the difficulty level of a given block of trials. According to this operational definition, the claim that a manipulation has increased the difficulty of the primary task will be made if, and only if, an increase in RMS error was produced.

With these observations established, the ERPs elicited during the dual-tasks can be analyzed to assess the effects of increased primary task difficulty on the amplitude of P300s associated with the tone stimuli of the secondary task. Recall that we predicted that decreased secondary task P300 amplitudes would be associated with increases in the difficulty of the tracking task. Confirmation of this prediction is consistent with the existence of resource reciprocity. Finally, the P300s associated with the primary task step changes will be evaluated to determine whether increased primary task P300 amplitudes were associated with decreased secondary task P300s. Primary and secondary task P300 amplitude reciprocity will be evaluated both across and within individual subjects.

Overt response data The average root mean square (RMS) error for a given block reflects the average distance between the cursor and the target square. Low values of the RMS error scores, therefore, reflect increased tracking accuracy. Because the primary and secondary task P300 amplitude data were range-corrected to facilitate direct comparison, the RMS data were also corrected according to the following transformation:

$$X(T) = 100 * \frac{X(I) - X(MIN)}{X(RNG)}$$

where, X(T)= transformed score;
X(I)= single block score;
X(MIN)= minimum score for a given subject;
X(RNG)= range of scores for a given subject.

In Table 2 we present the values of the mean, range-corrected, RMS error scores for all single and dual-task tracking conditions.

<Insert Table 2 about here>

It is evident that the introduction of the secondary task did not impair performance of the primary task. The RMS error scores during dual-tasks were not statistically different [$F(1,39)=3.89$, $p>.05$] than the RMS errors during single task tracking.

Because the RMS error data were compared with ERP data collected only during dual-task conditions, the RMS data from the dual-task conditions alone were submitted to range correction and the effects of the dimensionality and order manipulations were examined with respect to this data base. Tracking accuracy declined, that is error rate increased, as dimensionality increased [$F(1,39)=321.81$; $p<.0001$], and as the control order was increased from a velocity to an acceleration system [$F(1,39)=2246.03$; $p<.0001$].

The effect of system order was consistently larger than the effect of dimensionality. It is noteworthy that dimensionality affects the accuracy of tracking largely for the second order systems. This can be seen from the interaction between order and dimensionality [$F(1,39)=132.17$; $p<.0001$]. Tukey tests (Tukey, 1977) performed on specific pairwise comparisons indicate that order significantly affected performance in both one [$F(1,39)=202.21$; $p<.0001$] and two dimensions [$F(1,39)=4196.28$; $p<.0001$]. Similarly, the effect of dimensionality was significant for both velocity [$F(1,39)=9.37$; $p<.01$], and acceleration [$F(1,39)=247.18$; $p<.0001$] control systems.

Also presented in Table 2 are the mean number of counting errors during the auditory discrimination task. While analyses of these data indicated that counting performance was significantly impaired by the manipulation of system order [$F(1,39)=9.32$; $p<.01$], the magnitude of this effect is small. Because the average number of errors in even the most difficult condition was less than two, we assume that changes in the ERPs associated with the secondary task as a function of increased primary task difficulty reflect a reallocation of resources to the primary task rather than an unacceptable level of secondary task performance.

In summary, the RMS error data indicate that the manipulations of control order and dimensionality successfully produced a range of tracking performance suitable for the analysis of P300 amplitude reciprocity under varying levels of primary and secondary task competition for processing resources. Furthermore, the RMS error data confirm that subjects protected their performance on the primary task, for there was no significant increase in RMS error scores due to the imposition of the secondary task. Finally, the small number of counting errors in all dual-task blocks provides support for the claim that changes in the secondary task waveforms cannot be explained simply as the result of inadequate performance of the auditory discrimination task.

Secondary task ERP data The secondary task ERP data will now be examined to determine the extent to which variations in primary task workload (as reflected in the RMS error scores) are manifested in P300 amplitude variability associated with different secondary task conditions.

The grand average ERPs for targets which were associated with the dual-task tone count conditions are displayed in Figure 1. As predicted, the one-dimensional velocity condition was associated with the largest secondary task P300, and the smallest secondary task P300 was produced by the most difficult two-dimensional acceleration condition.

<Insert Fig. 1 about here>

A Principal Components Analysis (PCA) was conducted on the secondary task ERP data to quantify the effects of manipulations of system order and dimensionality upon the P300 (see Donchin and Heffley, 1978; Coles, Gratton, Kramer, and Miller, in press, for a discussion of this procedure). The main advantage of this technique is its ability to separate temporally overlapping components. A data matrix consisting of 960 trials (40 Subjects x 2 Dimensions x 2 Control Orders x 2 Stimulus Categories x 3 Electrodes) was submitted to the PCA. The grand mean waveform and the component loadings derived from the PCA are displayed in Figure 2.

<Insert Fig. 2 about here>

It has been suggested (Donchin, Ritter, and McCallum, 1978; Donchin, Kramer, and Wickens, 1982) that ERP components be identified according to three criteria: their latency relative to a stimulus or a response; their amplitude distribution across different electrode sites; and their sensitivity to task manipulations. Utilizing these criteria component 5 can be identified as the secondary task P300 component. This component was active in the appropriate latency range, and displayed the Pz maximal scalp distribution [$F(2,78)=12.91$; $p<.01$) traditionally associated with the P300. By evaluating the PCA derived component with the P300 we are not claiming that the two are synonymous, only that the experimental variance represented by the PCA presents a pattern of results consistant with the P300 component.

Having established component 5 as the P300 we will now examine the effects of the experimental manipulations upon the amplitude of this component. The requirement to track in two dimensions significantly reduced the amplitude of the P300s associated with the concurrent auditory discrimination task [F(1,39)=30.84; p<.0001] when compared with one dimensional tracking conditions. Similar results were obtained for the system order manipulation. Thus, smaller secondary task P300s were produced during velocity tracking conditions than when the tracking task required acceleration control [F(1,39)=21.49; p<.0001).

Because the above analyses indicated that component 5 displayed the traditional latency, scalp distribution, and sensitivity to experimental manipulations, the component scores for this factor at Cz were used as numerical estimates of P300 amplitude for each subject in all of the dual-task tracking conditions. These amplitude estimates were obtained so that variations in secondary task P300 amplitude could be directly compared with primary task P300 amplitude variability both within and across subjects. Since the results obtained from separate PCAs cannot be directly compared, the component scores were submitted to the range correction algorithm outlined above for the RMS error scores.

The mean range-corrected P300 component scores for the secondary task are presented in Table 3. Note that higher component scores reflect increased P300 amplitude.

<Insert Table 3 about here>

The analysis of variance conducted upon the range-corrected scores confirms that the manipulations of dimensionality [F(1,39)=17.24; p<.0002] and system order [F(1,39)=41.30; p<.0001] significantly reduced secondary task P300 amplitude with no significant interaction [F(1,39)=0.97; p=0.33]. Thus, if increased workload is defined as resulting from any manipulation which significantly increases RMS error, these data confirm that increased primary task workload, in every case, was associated with a reduction in the amplitude of P300s generated by the concurrently performed auditory discrimination task, confirming the previous studies cited above.

Primary Task ERPs The ERPs elicited by the step changes in the dual-task tracking conditions are displayed in Figure 3.

<Insert Fig. 3 about here>

The ERP pattern in this condition is quite different than that recorded in response to the secondary task. These waveforms are dominated by a large deflection that is maximal at the central electrode (Cz). It is also evident that the scalp distribution and the early latency of the peak suggest that this positive deflection is not a P300. We conclude, rather, that P300 in these waveforms overlaps with this component producing differential returns to baseline for the different conditions. It is evident that the amplitude of the P300 component in all of the conditions was quite small. This is not surprising given that the primary task step changes were all equiprobable.

Figure 4 displays the effect of system order upon the primary task waveforms in both one and two dimensions.

<Insert Fig. 4 about here>

The cross-hatched areas indicate regions of increased positivity associated with increased system order. However, the differences evident in the superaverages are small, presumably due to overlap with the earlier Cz maximal component. A procedure similar to the one outlined for the analysis of secondary task P300 amplitude was followed for the analysis of the primary task ERP data.

A single PCA was performed on the waveforms associated with both single and dual-task step changes. The data matrix submitted to the PCA consisted of 960 trials (40 Subjects x 2 Task levels x 2 Dimensions x 2 Control Orders x 3 electrodes), and four of the components extracted were Varimax rotated. The component structure extracted by this PCA is displayed in Figure 5.

<Insert Fig. 5 about here>

Component 1 is active in the appropriate latency range, and with the correct parietal maximal scalp distribution [$F(2,78)=219.07$; $p<.0001$] to enable its identification as the component corresponding to P300.

Overall, primary task P300 amplitude increased both as a function of increasing the number of dimensions [$F(1,39)=6.20$; $p<.05$] as well increasing the control order [$F(1,39)=33.32$; $p<.001$] of the tracking task with no significant interaction. Furthermore, both the dimension and order effects interacted with electrode site such that modulation of the component was greater at Cz [$F(2,78)= 7.13$; $p<.01$; and $F(2,78)=28.13$; $p<.01$, respectively] even though the component loaded maximally on the Pz electrode as noted above.

Because in both single and dual-task conditions the amplitude differences were significantly greater at Cz, numerical estimates of primary task P300 amplitude were obtained by extracting, for every subject, the component scores at Cz output by the PCA outlined above for the different dual-task tracking conditions. The mean component scores are presented in Table 3. Once again, to facilitate amplitude comparisons between the primary and secondary task P300s, the measures were corrected for range. An examination of Table 3 reveals that as the difficulty of the primary task increased (as reflected in increased RMS error scores and decreased secondary task P300 amplitudes), the amplitude of the P300s associated with primary task events also increased. Analysis of variance conducted on the range-corrected estimates of primary task P300 amplitude confirms that larger P300s were associated with two dimensional tracking conditions

[$F(1/39)=5.45$; $p<.02$], as well as with conditions requiring acceleration control [$F(1/39)=28.57$; $p<.0001$]. The dimension by order interaction was not significant [$F(1/39)=1.62$; $p>0.20$]. Thus, decreases in the amplitude of secondary task P300s were accompanied in every case by increased primary task P300 amplitudes.

Combined Analysis The preceding analysis can be criticized on the grounds that the P300 amplitude estimates for the primary and secondary tasks were assessed using different PCAs. Therefore, an additional analysis was conducted in which a single PCA was performed on the waveforms from the concurrently performed primary visual and secondary auditory tasks. Thus, a data matrix consisting of 960 trials (40 Subjects x 2 Task Levels x 2 Dimensions x 2 Control Orders), was submitted to a PCA in which 5 factors were extracted and Varimax rotated. The component structure output by this analysis is presented in Figure 6. Interpretation of the Anova conducted on the output of this PCA was complicated by the fact that the primary and secondary tasks required subjects to process stimulus information presented in different modalities. As a result the waveforms associated with the visual primary task differed considerably from the waveforms of the auditory secondary task. For example, the P300 component of the primary task was evident at a longer latency than the P300 component of the secondary task. Additionally, the structure of the components surrounding the P300 was different for the two tasks.

Because of the differences in latency and component structure, the P300s for the primary and secondary task emerged as two separate components in this analysis. Since component 2 has a parietal maximum scalp distribution and is active in the appropriate latency range we identify it as the P300 component for the primary task. Component 5 meets these criteria for the secondary task, and is therefore identified as the P300 for the auditory task. The same procedure described above was applied to obtain numerical estimates of P300 amplitude. Thus, the relevent component scores at Cz for each condition were extracted and range corrected for every subject. The mean estimates of P300 amplitude averaged across subjects are presented in Table 4.

<Insert Table 4 about here>

It is apparent that the general trends evident in Table 3 are also present in Table 4, for the primary and secondary task P300s are ordered in precisely the same way.

The estimates of P300 amplitude were then submitted to a repeated measures analysis of variance. For the primary task, P300 amplitude was greater during conditions requiring acceleration control than velocity control [$F(1,39)=35.58$; $p<.0001$]. The overall effect of increasing the number of dimensions to be tracked was marginally significant [$F(1,39)=3.88$; $p=.056$]. Analysis of the simple main effects indicated that P300 amplitude did not increase as a function of increased dimensionality for velocity systems [$F(1,39)=0.38$; $p>.50$] but did increase as dimensionality increased for acceleration systems [$F(1,39)=5.52$; $p<.025$]. The interaction between dimensionality and system order was not significant [$F(1,39)=1.56$; $p>.20$]. These results closely parallel those obtained when the waveforms associated with the step changes were analyzed separately.

Secondary task P300s revealed a reciprocal pattern of results with respect to the primary task changes. Secondary task P300s were larger during velocity tracking conditions [$F(1,39)=13.48$; $p<.001$]; as well as those involving one dimensional tracking [$F(1,39)=26.39$; $p<.0001$]. The dimension by system order interaction was not significant [$F(1,39)=0.33$;

p=.57]. Again, these results are identical to those obtained when secondary task P300 amplitudes were analyzed in isolation from the ERPs associated with the step changes.

Reciprocity The purpose of this study was to assess the degree to which P300 amplitudes associated with the two tasks would be reciprocal. In Figure 7 the amplitude of the P300 components in the primary and secondary tasks extracted from Table 3 are plotted as a function of the RMS error scores for each of the dual-task conditions. Inspection of the amplitude estimates in Table 4 reveals that the conclusions would remain the same had the results of the combined analysis been used instead. The line at the top of Figure 7 represents the sum of the primary and secondary task P300 amplitudes. Perfect amplitude reciprocity would generate a function with a slope of zero and an intercept value of 100. As can readily be seen by examining the obtained function, the evidence for amplitude reciprocity is quite good. Difficult tracking conditions produced a demand for perceptual resources resulting in increased primary task P300s and decreased secondary task P300s. Furthermore, the greater the increase in primary task P300 the greater the decrease in secondary task P300. This experiment, therefore, provides the first evidence for amplitude reciprocity obtained from concurrently recorded primary and secondary tasks of different modalities.

To determine the extent to which this pattern of reciprocity held true within subjects, separate reciprocity functions were obtained for each subject and the regression lines for these functions were computed. If the single subjects also demonstrated significant reciprocity the mean slope of these derived functions should equal zero and the mean intercept should equal 100. These data are presented in Table 5. Although there was significant variability across the subjects (indicating the presence of instances of both under and over reciprocity) the obtained value of 0.04 for :he mean slope did not differ significantly from the predicted value of 0 $\underline{t}$=0.44, p>0.10); and the mean intercept value of 92.99 did not differ ;ignificantly from the predicted value of 100 ($\underline{t}$=1.57, p>0.05). Thus, evidence in support of the reciprocity theory was obtained both across and vithin subjects.

<Insert Table 5 about here>

Conclusions

This experiment confirms the existence of a reciprocal relationship)etween the amplitudes of the P300s associated with two concurrently)erformed tasks. The predicition of reciprocity derives from a large body)f evidence which has indicated that variations in P300 amplitude are ;ensitive to the manner in which subjects allocate processing resources)etween two tasks under dual-task conditions. In other words, P300 ımplitude has emerged as a psychophysiological metric of the resource trade-)ffs that are presumed to underly the concept of mental workload (Kahneman, .973; Navon and Gopher, 1979; Wickens, 1980, 1981; Gopher and Donchin, in)ress).

The RMS error data confirm that the orthogonal manipulation of system ırder and dimensionality employed in this study successfully produced a wide ariability in performance within which to assess the reciprocity of primary nd secondary task P300 amplitudes. Because the difficulty of the secondary ask was held constant during all the step-tracking conditions, the model of esource reciprocity upon which this experiment is based predicts that as he tracking task is made more difficult, primary task P300 amplitudes should become larger, due to the allocation of additional processing resources; and secondary task P300 amplitudes should decline as a result of the drain upon this limited commodity.

The data collected during this experiment confirm this assertion. As the difficulty of the primary task increased, the RMS error measures also increased. Furthermore, the amplitude of the P300s associated with primary task step changes increased, while the amplitude of the secondary task P300s elicited by the auditory stimuli decreased in the predicted fashion. This result was obtained when amplitude measures were derived from both individual and combined PCAs. In all conditions, the increase in primary task P300 amplitude was proportional to the decrease in secondary task P300 amplitude. An examination of Figure 7 confirms that the summation of primary and secondary task P300 amplitudes yields an approximately constant value.

The validation of P300 amplitude as a metric of a particular aspect of the workload demands of a task has a number of theoretical and applied implications. As mentioned earlier, the auditory discrimination task is an attractive secondary task for a number of reasons. The most important of these reasons is that such a task can be applied in a relatively non-obtrusive fashion in many different situations since there is no need for an overt response. Thus, because subjects can count the stimuli rather than respond overtly to them, competition for response related processing resources is reduced. The RMS error data from this experiment confirm that, indeed, a secondary auditory discrimination task can be imposed in a dual-task setting with no evident cost to the performance of the primary task.

Another advantage of the discrimination task is that stimuli of different modalities can be used to elicit P300s. The modality of the secondary task can, therefore, be chosen to eliminate competition for modality specific processing resources. In this experiment, an auditory secondary task was chosen because the step-tracking task required visual stimulus processing. Had the primary task relied more upon auditory processing, a visual secondary task could have been employed. Although some preliminary work investigating the effects of overlapping primary and secondary task stimulus integrality has been carried out (Kramer, Wickens, and Donchin, 1985), further research is needed to determine the extent to which P300 amplitude decrements associated with secondary tasks of different modalities can be compared to assess workload.

For these reasons, P300 amplitude measured under dual-task conditions can be used in the analysis of demands placed upon operators in complex man-machine systems. The P300 is a relatively unobtrusive measure sensitive to graded changes in task difficulty. Furthermore, the P300 is diagnostic of perceptual/cognitive resource demands as opposed to response-related processing (Isreal et al., 1980). Finally, it is conceivable that with further refinements, such as the application of step-wise discriminant analysis techniques (Donchin and Herning, 1975), the bandwidth and reliability of the P300 may be of sufficient quality to permit the analysis of workload on a moment by moment basis.

In addition to validating the prediction of P300 amplitude reciprocity, this experiment produced a number of ancillary findings. In particular, previous findings concerning the nature of the manipulation of system order obtained in this laboratory have been both replicated and extended. Thus, the conclusion by Wickens et al. (1983) that the manipulation of system order during a one-dimensional tracking task produces a salient drain on perceptual/cognitive processing resources has been confirmed and extended to the two dimensional case. Secondary task P300s declined and primary task P300s increased in amplitude as a function of increased system order in both one and two dimension.

However, Wickens et al. (1977) reported that changing the dimensionality of tracking has no effect on P300 amplitude in velocity systems. This finding was not replicated in our study. Primary and secondary task P300 amplitude did significantly vary as a function of this manipulation when subjects were tracking with a velocity control system. However, Wickens et. al used a different paradigm where the primary task consisted of a compensatory tracking task, while the present study utilized a pursuit step tracking task. Thus, the nature of the dimensionality manipulation was different in the two paradigms. This difference may account for our failure to replicate the earlier findings. Secondly, it should be recalled that the magnitude of the dimensionality manipulation in velocity systems produced the smallest changes in RMS error and P300 amplitude of any of the manipulations invoked. Thus, although this manipulation produced a significant change in workload, the magnitude of this change was quite small. It is conceivable that this effect only attained significance due to the power of a design involving 40 subjects.

In conclusion, this experiment demonstrates using concurrently recorded ERPs associated with primary and secondary tasks, that there is a reciprocal relationship between the amplitudes of the P300s associated with the two tasks. Furthermore, this relationship was investigated and confirmed under a variety of levels of primary and secondary task competition for processing resources. The results can be intrepretated within a model of dual-task performance in which the allocation of processing resources to the two tasks was presumed to determine primary and secondary task P300 amplitude. Thus, because a reciprocal relationship between the allocation of processing resources to the two tasks was presumed to exist, a reciprocal relationship between primary and secondary task P300 amplitudes was predicted. This prediction of primary and secondary task P300 amplitude reciprocity was confirmed in all of the conditions in which it was tested. Additionally, the zero slope of the derived reciprocity function is evidence that the total supply of resources available for allocation to the primary or the secondary task remained relatively constant for all the tracking conditions employed in this experiment. Thus, this experiment further illustrates the utility of the P300 as a tool to aid in the analysis of mental workload.

References

Allport, D.A., Antonis, B., and Reynolds, P. (1972) On the division of attention: A disproof of the single channel hypothesis. Quarterly Journal of Experimental Psychology, 24, 255-265.

Berlyne, D.E. (1960) Conflict, Arousal, and Curiosity. New York: McGraw-Hill.

Brown, I.D. (1978) Dual-task methods of assessing workload. Ergonomics, 21, 221-224.

Coles, M.G.H., Gratton, G., Kramer, A., and Miller, G. (in press) Principles of signal acquisition and analysis. In Coles, M.G.H., Donchin, E., and Porges, S. (Eds.) Psychophysiology: Systems, Processes, and Applications. Vol. I: Systems. New York: Guilford Press.

Donchin, E. & Heffley, E. (1975) Minicomputers in the signal averaging laboratory. American Psychologist, 30, 299-312.

Donchin, E. & Herning, R.I. (1975) A simulation study of the efficacy of stepwise discriminant analysis in the detection and comparison of event-related potentials. Electroencephalography and Clinical Neurophysiology, 38, 51-68.

Donchin, E., Kramer, A., & Wickens, C.D. (1982) Probing the cognitive infrastructure with event-related brain potentials. Proceedings of the AIAA Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics. Edwards Air Force Base, California, pp. 371-387.

Donchin, E., Kubovy, M., Kutas, M., Johnson, R. jr., & Herning, R.I. (1973) Graded changes in evoked response (P300) amplitude as a function of cognitive activity. Perception and Psychophysics, 14, 319-324.

Donchin, E., Ritter, W., & McCallum C. (1978) Cognitive psychophysiology; The endogenous components of the ERP. in E. Callaway, P. Tueting, & S. Koslow (Eds.), Brain event-related potentials in man. New York: Academic Press, pp. 349-441.

Duncan-Johnson, C.C. & Donchin, E. (1977) On quantifying surprise: The variation in event-related potentials with subjective probability. Psychophysiology, 14, 456-467.

Duncan-Johnson, C.C. & Donchin, E. (1978) Series-based vs. trial-based determinants of expectancy and P300 amplitude. Psychophysiology, 15, 262.

Gopher, D. & Donchin, E. (in press), What is workload -- An introduction, in L. Kaufman and K. Boff (Eds.), Handbook of perception and human performance. New York: John Wiley.

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.

Howitt, J.S. (1968) Flight deck workload studies in civil transport aircraft. Nato AGARD Conference Proceedings, 56, 1-7.

Isreal, J.B., Chesney, G.L., Wickens, C.D., & Donchin, E. (1980) P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. Psychophysiology, 17, 259-273.

Isreal, J.B., Wickens, C.D., Chesney, G.L. & Donchin, E. (1980) The event related brain potential as an index of display monitoring workload. Human Factors, 22, 211-224.

Jasper, H.H. (1958) The ten twenty electrode system of the International Federation. Electroencephalography and Clinical Neurophysiology, 10, 371-375.

Johnson, R.E. jr., and Donchin, E. (1978) On how P300 amplitude varies with the utility of the eliciting stimuli. Electroencephalography and Clinical Neurophysiology, 44, 424-437.

Johnson, R.E. jr., and Donchin, E. (1982) Sequential expectancies and decision making in a changing environment: An electrophysiological approach. Psychophysiology, 19, 183-200.

Kahneman, D. (1973) Attention and Effort. Englewood Cliffs, N.J.: Prentice Hall.

Knowles, W.B. (1963) Operator loading tasks. Human Factors, 5, 155-161.

Kramer, A.F., Wickens, C.D. & Donchin, E. (1983) An analysis of the processing requirements of a complex perceptual-motor task. Human Factors, 25, 597-622.

Kramer, A.F., Wickens, C.D. & Donchin, E. (1985) The processing of stimulus properties: Evidence for dual-task integrality. Journal of Experimental Psychology: Human Perception and Performance, 12, 389-407.

Moray, N. (1967) Where is attention limited? A survey and a model. Acta Psychologica, 27, 84-92.

Navon, D. & Gopher, D. (1979) On the economy of the human processing system. Psychological Review, 86, 214-255.

Norman, D., & Bobrow, D. (1975) On data-limited and resource-limited processes. Cognitive Psychology, 7, 44-64.

North, R.A. (1977) Task functional demands as factors in dual-task performance. Proceedings of the Twenty-first Annual Meeting of the Human Factors Society, San Fransisco: October.

Ogden, G.D., Levine, J.W., & Eisner, E.J. (1979) Measurement of workload by secondary tasks. Human Factors, 21, 529-548.

Rolfe, J.M. (1971) The secondary task as a measure of mental load. In W.T. Singleton, J.G. Fox, & D. Whitfield (Eds.), Measurement of Man at Work. London: Taylor & Francis, pp. 135-148.

Roscoe, A.H. (1978) Assessing Pilot Workload. AGARD-A6-233.

Schneider, W. & Fisk, A.D. (1982) Concurrent automatic and controlled visual search: Can processing occur without cost? Journal of Experimental Psychology: Learning, Memory, & Cognition, 8, 261-278.

Shaffer, H. Multiple attention in continuous verbal tasks. (1975) In S. Dornick & P.M. Rabbitt (Eds.) Attention and Performance V. New York: Academic Press.

Sperling, G. & Melchner, M. (1978) Visual search, visual attention and the attention operating characteristic. In J. Requin (Ed.), Attention and Performance VIII. Hillsdale, N.J.: Erlbaum, pp. 675-686.

Sutton, S., Braren, M., Zubin, J. & John, E.R. (1965) Evoked-potential-correlates of stimulus uncertainty. Science, 150, 1187-1188.

Tukey, J.W. Exploratory Data Analysis. (1977) Reading, Mass.: Addison Wesley.

Wickens, C.D. (1976) The effects of divided attention in information processing and tracking. Journal of Experimental Psychology: Human Perception and Performance, 2, 1-13.

Wickens, C.D. (1980) The structure of processing resources. In R.S. Nickerson (Ed.) Attention and Performance VIII. Hillsdale, N.J.: Lawrence Erlbaum, pp. 239-254.

Wickens, C.D. (1981) Workload: In defense of the secondary task. Personnel Training and Selection Bulletin, 2, 119-123.

Wickens, C.D. (1983) Processing resources in Attention. In R. Parasuraman, J. Beatty, & R. Davies (Eds.), Varieties of Attention. New York: Academic Press, pp. 63-98.

Wickens, C.D. (1984) Engineering Psychology and Human Performance. Columbus, Ohio: Charles Merrill Publishing Company.

Wickens, C.D., Isreal, J.B., Donchin, E. (1977) The event-related cortical potential as an index of task workload. In A.S. Neal & R.F. Palasek (Eds.), Proceedings of the Human Factors Society 21st Annual Meeting, Santa Monica, California.

Wickens, C., Kramer, A., Vanasse, L., and Donchin, E. (1983) Performance of concurrent tasks: A Psychological analysis of the reciprocity of information processing resources. Science, 221, 1080- 1082.

Wierwille, W.W. (1979) Physiological measures of aircrew mental workload. Human Factors, 21, 575-594.

Hilliges, R.C. & Wierwille, W.W. (1979) Behavioral measures of aircrew mental workload. Human Factors, 21, 549-574.

TABLE 1

ORDER OF TASK PRESENTATION

	VELCOCITY IN ONE DIMENSION	ACCELERATION IN ONE DIMENSION	VELOCITY IN TWO DIMENSIONS	ACCELERATION IN TWO DIMENSIONS
SINGLE	1			
DUAL	2			
SINGLE			3	
DUAL			4	
SINGLE		5		
DUAL		6		
SINGLE				7
DUAL				8

TABLE 2

MEAN OVERT PERFORMANCE DATA
(Standard Deviations are enclosed in parentheses)

	VELOCITY IN ONE DIMENSION	ACCELERATION IN ONE DIMENSION	VELOCITY IN TWO DIMENSIONS	ACCELERATION IN TWO DIMENSIONS
SINGLE TASK RMS ERROR	4.30 (4.09)	41.90 (15.23)	5.70 (7.01)	91.05 (11.32)
DUAL-TASK RMS ERROR	2.25 (4.32)	46.33 (18.81)	6.75 (7.64)	95.92 (6.22)
COUNTING ERRORS	0.85 (0.85)	1.73 (1.84)	1.10 (1.61)	1.58 (1.67)

TABLE 3

MEAN RANGE-CORRECTED P300 AMPLITUDE
(Standard Deviations are enclosed in parentheses)

	VELOCITY IN ONE DIMENSION	ACCELERATION IN ONE DIMENSION	VELOCITY IN TWO DIMENSIONS	ACCELERATION IN TWO DIMENSIONS
SECONDARY TASK	69.59 (39.13)	44.08 (34.43)	53.02 (38.53)	19.19 (31.54)
PRIMARY TASK	27.39 (34.87)	53.17 (35.58)	35.06 (39.19)	76.65 (32.18)

TABLE 4

MEAN RANGE-CORRECTED P300 AMPLITUDE
COMBINED ANALYSIS
(Standard Deviations are enclosed in parentheses)

	VELOCITY IN ONE DIMENSION	ACCELERATION IN ONE DIMENSION	VELOCITY IN TWO DIMENSIONS	ACCELERATION IN TWO DIMENSIONS
SECONDARY TASK	72.44 (34.50)	44.01 (33.79)	55.47 (39.18)	19.29 (33.77)
PRIMARY TASK	28.07 (32.32)	57.37 (35.56)	33.67 (39.14)	77.80 (30.98)

TABLE 5

MEAN SLOPE AND INTERCEPT VALUES FOR INDIVIDUAL RECIPROCITY FUNCTIONS

SUB.	SLOPE	INTERCEPT	SUB.	SLOPE	INTERCEPT
1	0.03	97.35	21	0.11	90.02
2	0.16	102.49	22	0.61	78.47
3	0.43	87.76	23	0.32	95.09
4	0.11	81.95	24	-1.32	164.21
5	0.56	62.72	25	-0.79	94.91
6	-0.22	126.22	26	0.22	86.06
7	0.20	75.72	27	-0.57	109.80
8	0.39	73.00	28	0.00	108.03
9	0.41	70.31	29	-0.45	137.51
10	-1.10	136.28	30	-0.28	115.70
11	0.08	86.48	31	0.11	84.17
12	0.88	61.95	32	-0.23	92.97
13	-0.45	116.39	33	0.06	71.55
14	-0.21	105.22	34	1.17	39.32
15	-1.21	124.00	35	0.24	91.04
16	-0.41	114.11	36	0.44	84.86
17	-0.13	73.38	37	-0.42	109.49
18	0.08	93.47	38	0.47	99.51
19	0.94	35.61	39	1.32	23.11
20	0.24	92.36	40	-0.23	126.94

Figure Legends

Figure 1 The scalp distribution of the grand average ERPs elicited by secondary task target tones during the dual-task tracking conditions.

Figure 2 Panel A) Grand average waveform with +/- one standard deviation unit for secondary task ERPs. Panel B) Component loadings for the first five components extracted from a PCA of the secondary task ERPs.

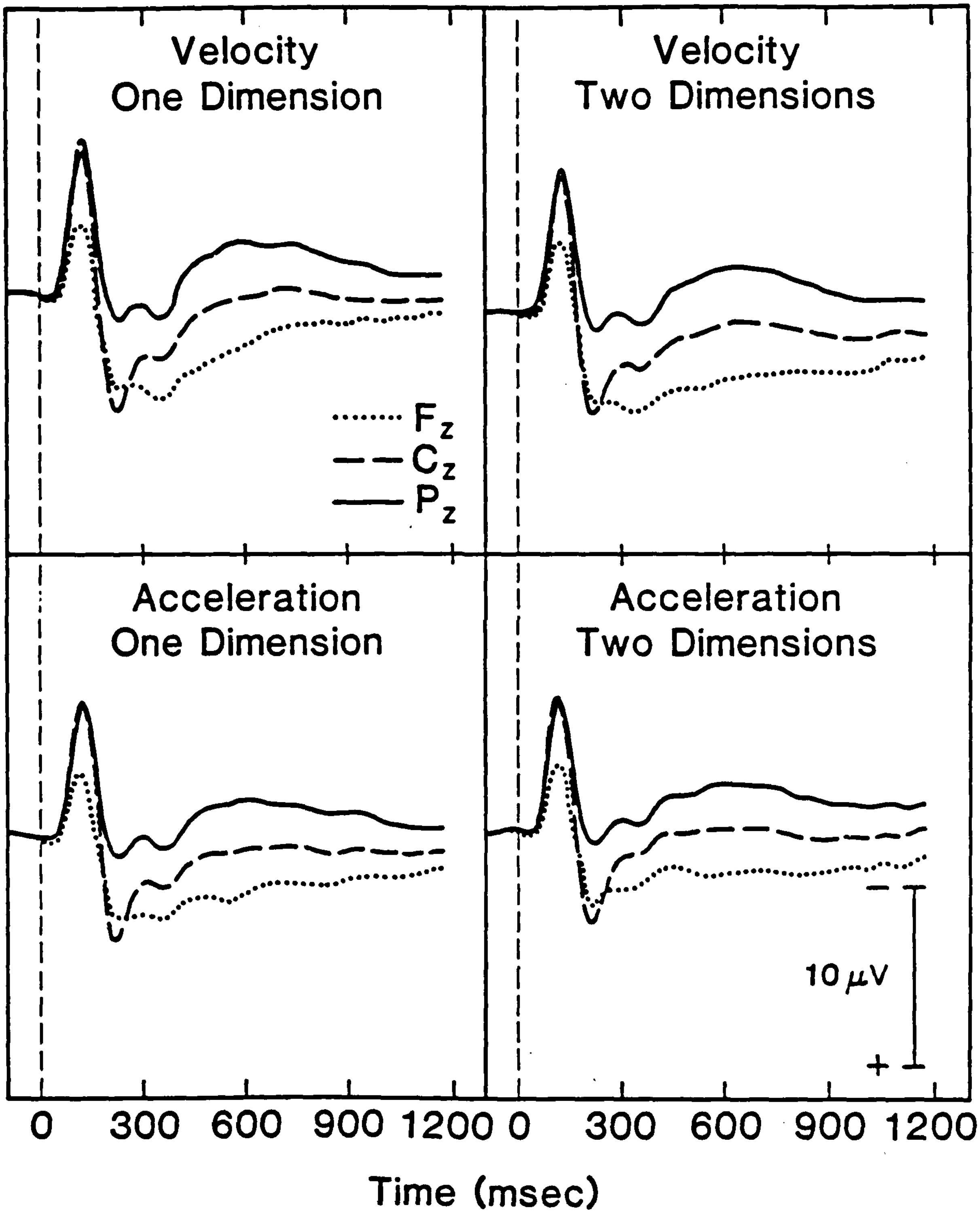
Figure 3 The scalp distribution of the grand average ERPs elicited by the primary task step changes for the different step tracking conditions.

Figure 4 The effect of increased system order upon the primary task parietal waveforms is shown. The cross-hatched areas indicate increased positivity as a function of increased system order.

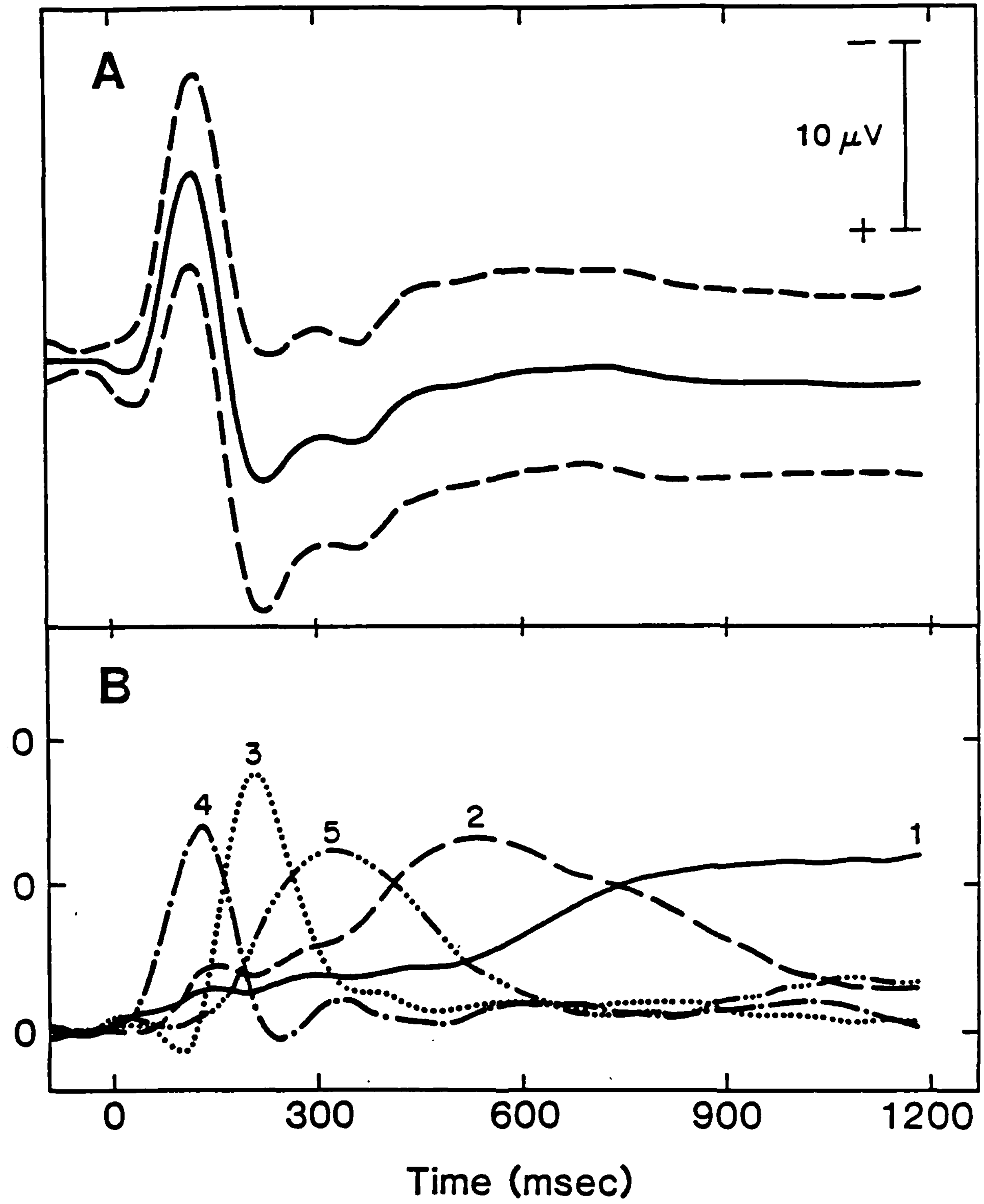
Figure 5 Panel A) Grand average waveform with +/- one standard deviation for the primary task waveforms. Panel B) Component loadings for the first 4 components extracted from a PCA of the primary task ERPs.

Figure 6 Panel A) Grand average waveform with +/- one standard deviation for the combined primary and secondary task ERPs. Panel B) Component loadings for the first five components extracted from a PCA of combined primary and secondary task ERPs.

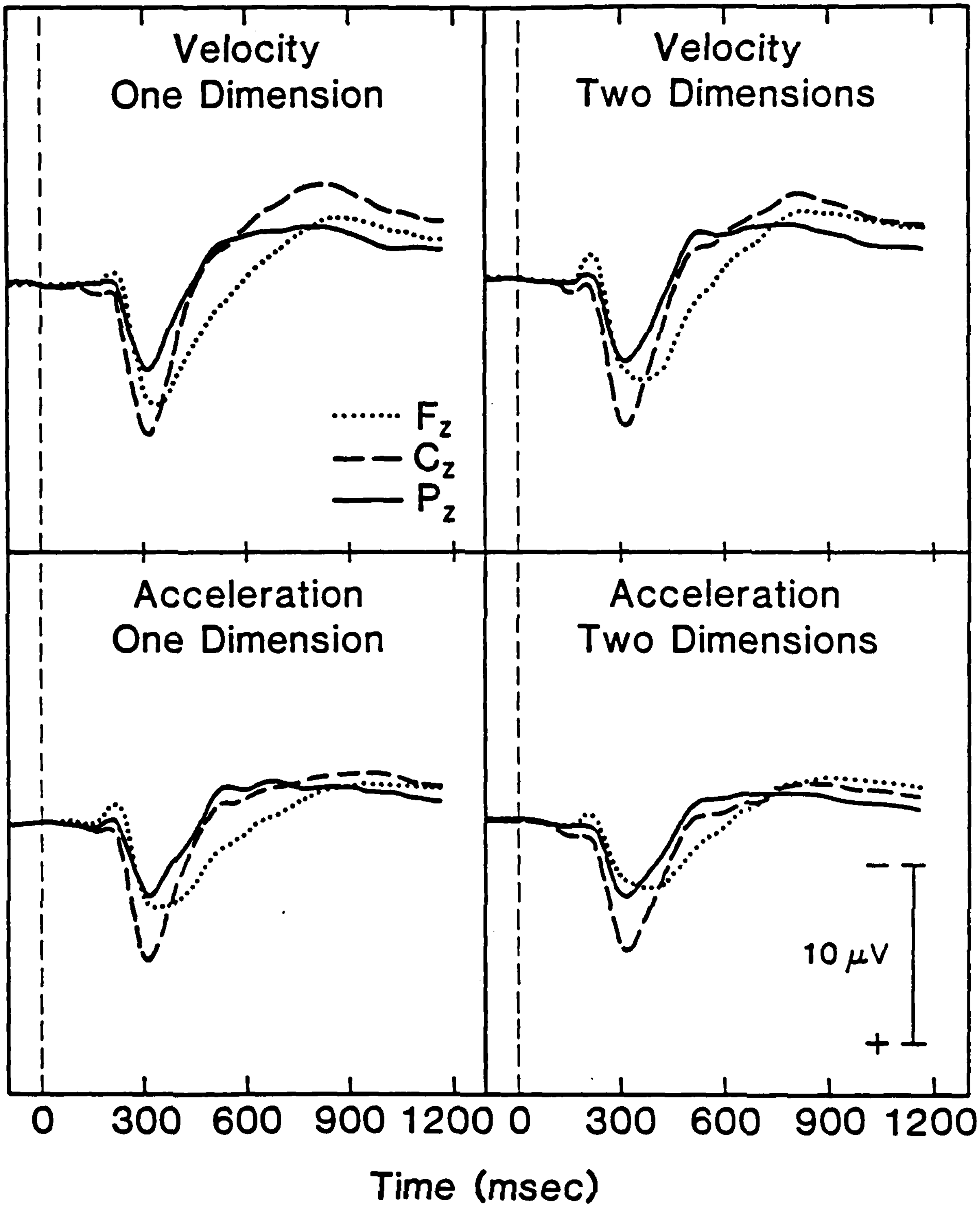
Figure 7 Mean range corrected primary and secondary task P300 amplitudes for all of the dual-task tracking conditions plotted as a function of the associated RMS error score. The reciprocity function represents the sum of the primary and secondary task amplitude measures.

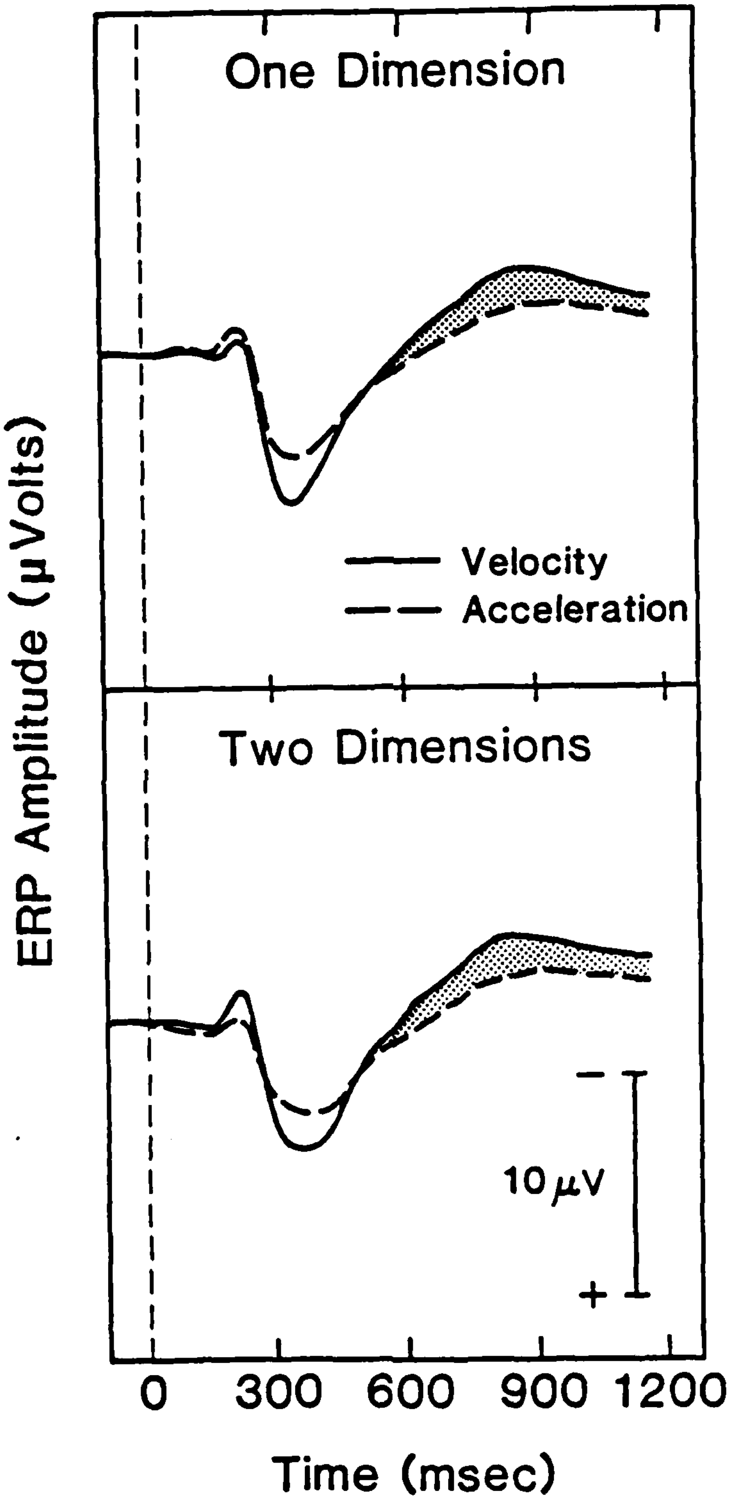


Velocity
One Dimension
Velocity
Two Dimensions
Fz
Cz
Pz
Acceleration
One Dimension
Acceleration
Two Dimensions
10 μV
+
0
300
600
900
1200
0
300
600
900
1200
Time (msec)

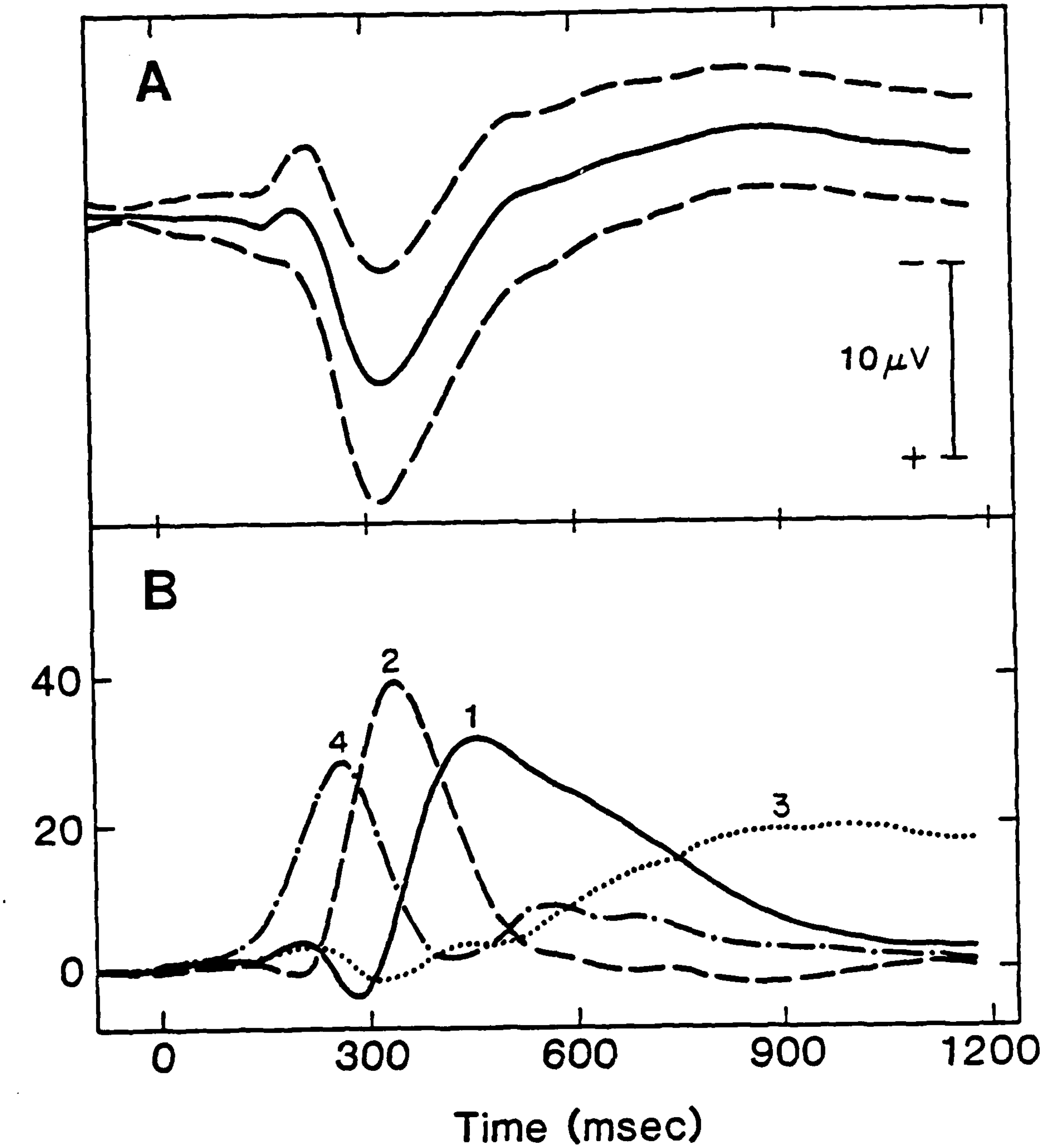


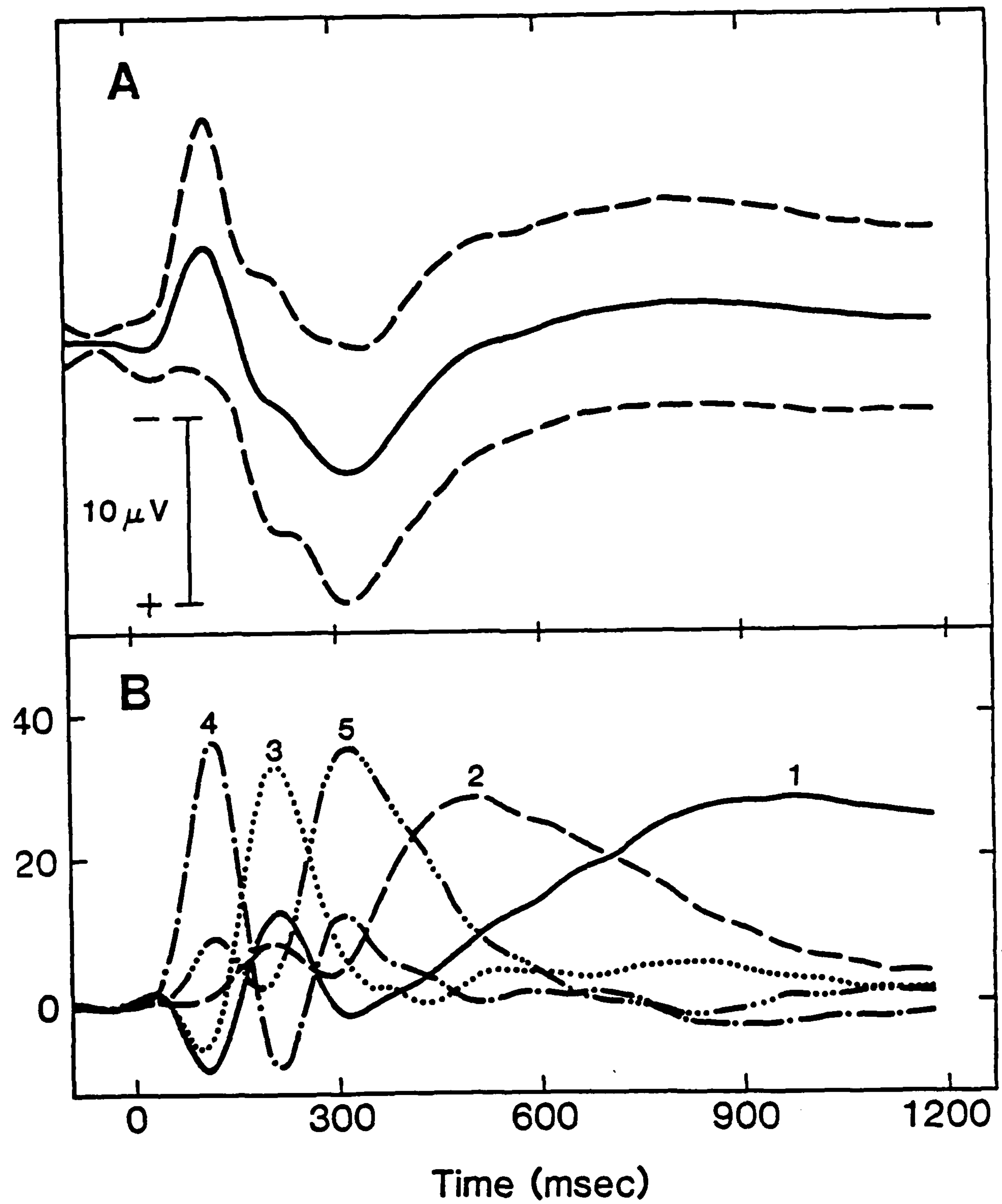
A
10 μV
−
+
B
4
3
5
2
1
0
0
0
0
300
600
900
1200
Time (msec)


Velocity
One Dimension
Velocity
Two Dimensions
Fz
Cz
Pz
Acceleration
One Dimension
Acceleration
Two Dimensions
10 μV
0
300
600
900
1200
Time (msec)

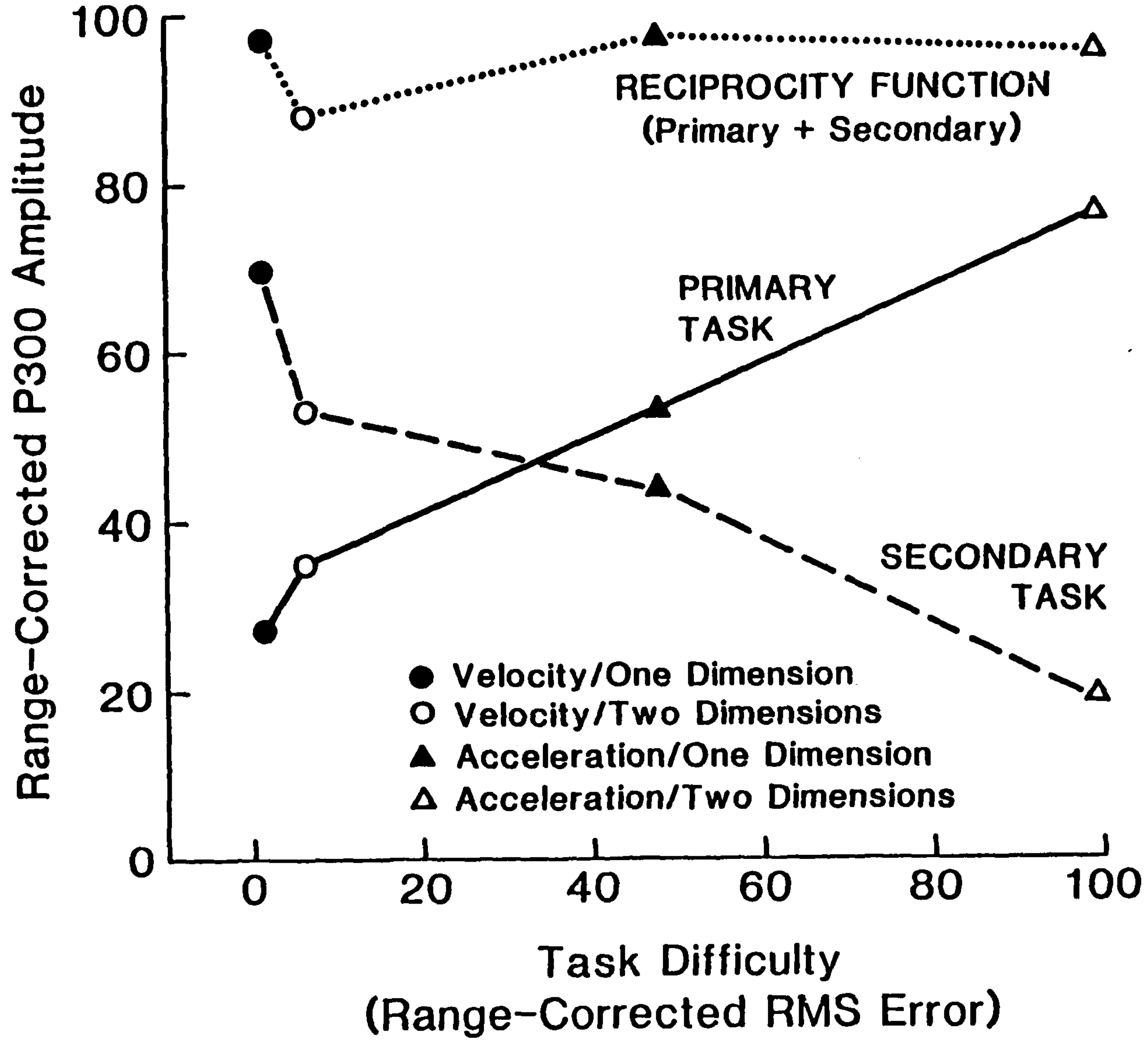


One Dimension
Velocity
Acceleration
Two Dimensions
ERP Amplitude (μVolts)
10 μV
−
+
0
300
600
900
1200
Time (msec)


A
10 μV
−
+
B
40
20
0
4
2
1
3
0
300
600
900
1200
Time (msec)



A
10 μV
−
+
B
40
20
0
4
3
5
2
1
0
300
600
900
1200
Time (msec)



100
80
60
40
20
0
Range-Corrected P300 Amplitude
RECIPROCITY FUNCTION
(Primary + Secondary)
PRIMARY
TASK
SECONDARY
TASK
Velocity/One Dimension
Velocity/Two Dimensions
Acceleration/One Dimension
Acceleration/Two Dimensions
0
20
40
60
80
100
Task Difficulty
(Range-Corrected RMS Error)

Appendix B17

Abstract presented at the 25th Annual Meeting of the Society for Psychophysiological Research, Houston, 1985. Psychophysiology, 22, 588-589.

Effects of Strategy Manipulation on P300 Amplitude in a von Restorff Paradigm

Monica Fabiani, Demetrios Karis and Emanuel Donchin
Cognitive Psychophysiology Laboratory
University of Illinois at Urbana-Champaign

Donchin (1981) proposed that P300 manifests the updating of schemas in working memory ("context updating"). This theory led to the hypothesis that the larger the P300 elicited by a word, the more likely its subsequent recall. In a previous study we tested this hypothesis using a von Restorff paradigm (Karis, Fabiani & Donchin, 1984). We recorded ERPs to words in a series that contained a deviant word (an "isolate"). The isolation was achieved by changing the size of the word. In general, isolated items are better recalled than comparable non-deviant items (the von Restorff effect). We found that subjects who displayed the largest von Restorff effect reported using rote mnemonic strategies, and had a low recall performance. For these subjects, isolates later recalled elicited a larger P300 on initial presentation, than isolates that were not recalled. Subjects displaying the lowest von Restorff effect had the best recall performance and reported using elaborative strategies. For these subjects, P300 amplitude did not predict subsequent recall. Therefore, we concluded that the relationship between P300 amplitude and recall emerges only when it is not overshadowed by subsequent elaborative processing.

In the present study we manipulated strategies by instructions to determine whether the relationship between recall and P300 amplitude depends indeed on the subject's mnemonic strategies. Instructions to use "rote" strategies required the subject to repeat each word as it was presented, while "elaborative" instructions required the subject to combine words into images, sentences, or stories.

Nine subjects were run for three sessions in a von Restorff paradigm. The first session was devised to assess the subjects' natural strategies, and their ability to change their strategy according to the instructions. In both the second and third session ERPs were recorded, and subjects were instructed to use one strategy during the first half of the session, and the other during the second half. The order of presentation was counterbalanced across subjects.

Strategy instructions were effective in manipulating the performance of the subjects. When instructed to use rote strategies, subjects recalled significantly fewer words, and displayed a significantly higher von Restorff effect, then when they used elaborative strategies.

ERP analyses on the nine subjects also supported our predictions. When subjects were instructed to use rote strategies, the P300s elicited by words later recalled were significantly larger than those elicited by words later not recalled. However, when subjects were instructed to use elaborative strategies, there was no memory effect on P300 amplitude, but there was a memory effect on a frontal-positive Slow Wave. This component was also observed in the "elaborators" of the previous experiment.

Thus, these data support the theory that P300 manifests the updating of schemas in working memory. The relationship between P300 amplitude and recall is most evident when the subjects base their recall on their first encoding of the word (rote strategies), while it is overshadowed when subjects continue their processing well beyond the P300 time-range (elaborative strategies). There also seems to be a relationship between the amplitude of a later Slow Wave and recall when the subjects use elaborative strategies.

Abstract presented at the 25th Annual Meeting of the Society for Psychophysiological Research, Houston, 1985. Psychophysiology, 22, 589.

A SPECIFIC MEMORY DEFICIT IN ELDERLY SUBJECTS WHO LACK A P300

Lawrence A. Farwell, Ron D. Chambers, Gregory A. Miller,
Michael G.H. Coles, & Emanuel Donchin
Cognitive Psychophysiology Laboratory
University of Illinois at Urbana-Champaign

In a previous study of 53 healthy community residents aged 60-82 who completed four variants of the Oddball task, we found that the amplitude of the P300, which varied considerably across subjects, was stable within subjects (Marshall et al., 1983). Strikingly, about 20% of the subjects showed a very small P300. The reliability of this observation was confirmed by retesting 27 subjects -- 9 subjects with the largest and 9 with the smallest P300s, plus another 9 randomly selected from the entire range. No psychometric or demographic attribute differentiated subjects with high and low P300.

In the present study, the hypothesis was tested that small-P300 subjects cannot adequately maintain their Working Memory. This hypothesis is derived from the suggestion that P300 is a manifestation of the updating of Working Memory (Donchin, 1981).

Two years after initial testing, we recalled 9 subjects with a high P300 ("Highs") and 7 subjects wth a low P300 ("Lows") and challenged them with a digit-pair memory task derived from Milner's procedure for assessing temporal lobe damage. On each trial, the subject was presented with a digit followed by a second digit 1000 msec later. The subject was instructed to indicate whether the pair had appeared previously in that trial block. In each block, one specific pair of digits was repeated approximately every 3 trials. All other pairs were different. Reaction time and error rate served as dependent variables.

In general, all subjects correctly identified repeating and nonrepeating pairs. Furthermore, the two groups were equally fast in their response to repeating digit pairs. However, Lows were consistently slower than Highs in responding to the nonrepeating pairs. The correlation between RT for the nonrepeating pairs in the digit task and P300 amplitude in the original Oddball battery was significantly positive. Importantly, Highs had shown neither larger N100 amplitude nor faster choice RT during the oddball screening, ruling out gross attentional, motivational, or motor deficits as a basis for the group differences.

These data suggest that subjects with a low P300 require more time to determine, or to report, that a digit pair is not one that they have already seen. Thus, in Low subjects, the "recognition" required with the repeating pairs is not impaired, but "recall" -- that is, a search of memory unaided by external cues -- does show impairment.

To confirm this interaction of P300 amplitude with memory search demands, a new sample (N=32) was recruited, and the basic findings of the first study were replicated. Thus, exceptionally low amplitude of the P300 characteristic of a subject reflects a deficit specific to recall from working memory, rather than a generalized deficit so often seen in special populations. These data support the interpretation of P300 as a manifestation of the updating of working memory and suggest that it is useful to examine the deterioration of memory performance in the elderly in terms of specific properties such as the maintenance of working memory.

Abstract presented at the 25th Annual Meeting of the Society for Psychophysiological Research, Houston, 1985. Psychophysiology, 22, 592.

Examining Stimulus Evaluation and Response Preparation with Psychophysiological Measures

Gabriele Gratton, Michael G.H. Coles, Erik Sirevaag, Charles W. Eriksen, and Emanuel Donchin

Cognitive Psychophysiology Laboratory - Department of Psychology
University of Illinois at Urbana-Champaign

In a previous experiment (Coles, Gratton, Bashore, Eriksen, & Donchin, in press), we examined the responses of subjects to visual displays containing both target and noise information. We found that response accuracy was determined by three mechanisms: (a) a mechanism dependent on the processing of all the features present in the display ("feature analysis"); (b) a mechanism dependent only on the information presented at the target location ("location analysis"); and (c) a response priming or bias mechanism.

In the present study, we replicated and extended the first study. One of four stimulus arrays (HHHHH, SSHSS, SSSSS, and HHSHH) was presented for 100 ms every 4.5 to 6.5 s. Six subjects were instructed to respond with one hand when the central letter of the array was an H, and with the other when it was an S. The array was preceded by 1000 ms by a warning tone. We obtained measures of event-related brain potentials, electromyogram, and analog representations of the overt behavioral response (a squeeze) on every trial.

We derived descriptions of response accuracy as *a function of response* latency (speed-accuracy trade-off functions). These functions were characterized by three phases: (a) an early phase for which response accuracy was at a chance level for any kind of display; (b) an intermediate phase in which response accuracy was above chance for displays with no conflict and below chance for displays with conflict; and (c) a late phase in which response accuracy was high for any kind of display. These results confirmed our previous conclusions that response accuracy at each latency is determined by the relative weight of three mechanisms: bias, feature analysis, and location analysis.

A decomposition of the speed-accuracy functions suggested that the contributions of feature and location analysis to response accuracy, although peaking at different moments in time, may start simultaneously. However, this conclusion may be inaccurate because of trial-to-trial variation in the speed of evaluation processes. In fact, when the speed of evaluation processes is taken into account by using the ratio between RT and P300 latency (instead of RT), location analysis appears to contribute to response accuracy only when the contribution of the feature analysis has reached its peak.

We hypothesized that the priming, or bias mechanism could be attributed to an activation of one of the response independent of the stimulus ("aspecific priming"). To study this mechanism, we analyzed the scalp negativity preceding the presentation of the array. We hypothesized that the negativity indicates motor preparation, and that the accuracy of guesses can be predicted by comparing the side of the scalp at which the negativity was larger with the side of the correct response. Indeed, the accuracy of fast guesses was above chance for trials with negativity maximum on the side contralateral to the correct response and below chance for trials with negativity maximum on the side ipsilateral to that response. This result indicates that response accuracy is at least in part attributable to "aspecific priming."

Abstract presented at the 25th Annual Meeting of the Society for Psychophysiological Research, Houston, 1985. Psychophysiology, 22, 592. Appendix B20

Dual-Task Processing and Visual Selective Attention: An Event-Related Brain Potentials Analysis

Arthur F. Kramer and Erik Sirevaag
Department of Psychology
University of Illinois

The study of selective attention can be characterized in terms of the models proposed to account for the phenomenon, as well by the paradigms used to experimentally manipulate attentional processes. Although a multitude of attentional models exist, they can be categorized into two general groups: structural and capacity models. Structural models suggest that attention is allocated on the basis of channels of information. A channel is defined as a physical or semantic characteristic of a stimulus that can be selectively processed. The major disagreement among structural models is the level at which selection occurs. ERPs have been useful in addressing this issue. Several early components have been shown to discriminate between attended and unattended channels (e.g. left and right visual fields) while the P300 is sensitive to target stimuli within an attended channel. This hierarchical relationship among ERP components suggests that selection occurs at different stages within the information processing sequence. Capacity models of attention, which postulate a hypothetical resource structure underlying performance, have been concerned with delineating the antecedant conditions which lead to dual-task decrements. P300s have been shown to mimic the resource tradeoffs presumed to underlie dual-task performance. When two tasks are performed concurrently, P300s elicited by primary task events increase in amplitude while P300s recorded from secondary task stimuli decrease in amplitude with increases in the difficulty of the primary task.

Structural approaches are concerned with assessing the allocation of attention across stimuli within a task while capacity approaches assess resource tradeoffs between tasks. The current study examined the interaction between these two types of attentional processes by investigating the effects of the cognitive difficulty of a foveally presented task on the processing of parafoveal stimuli. ERPs, response accuracy and reaction time were employed as measures of the selective processing of task relevant events.

The foveal task required subjects to monitor a simulated process control plant for system failures. The multiattribute process was represented by a triangle which changed shape as a function of the values of three system variables. The difficulty of the monitoring task was manipulated by varying the complexity of the rule for detecting a system failure. Once a failure had been detected subjects pressed a response button to reset the system. Failures occurred with an ISI of between 5 and 50 sec. The parafoveal task required the discrimination between vertical bars of different lengths. The bars were either long (.80 - targets) or short (.20 - standards) and were presented in the left and right visual fields. During a block of trials, subjects attended to one field and depressed a response button whenever a short bar occurred. Bars were presented every 1.2 to 1.5 sec at either two or seven degrees of visual angle from the monitoring task. ERPs were elicited by the bars. Subjects performed the foveal and parafoveal tasks both separately and together.

RT and accuracy measures indicated that the difficulty of the foveal task influenced subjects performance on the peripheral task. Three ERP components (N170, P230 & N310) discriminated between stimuli in the attended and unattended visual fields. The N170 decreased in amplitude with increases in the difficulty of the monitoring task. The N310 was larger for the targets than the standards. However, this difference diminished with increasing foveal task difficulty. P300s were larger for the attended than the unattended targets. The amplitude of the P300s decreased with increased difficulty of the monitoring task. The implications of these results for models of attention will be discussed.

END

DTIC

9 — 86