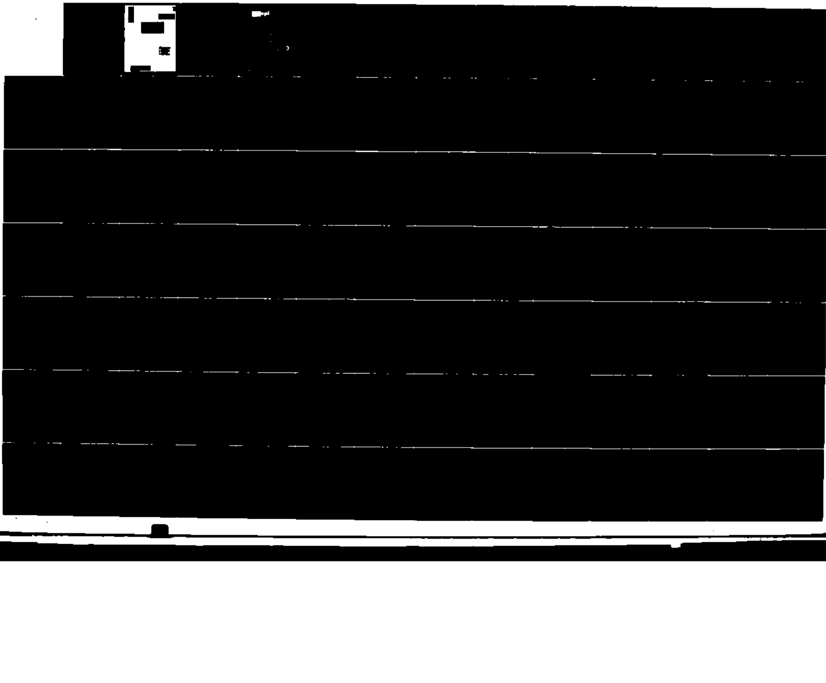
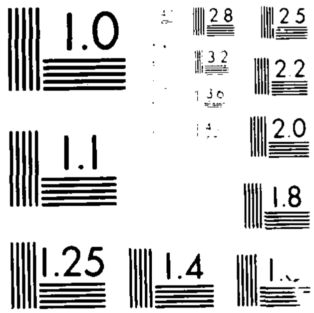
AD-A191 244 THE EVENT-RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING A..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 1/10

UNCLASSIFIED 29 FEB 88 CPL-88-1 AFOSR-TR-88-0316 F/G 5/8 NL




1.0
2.8
2.5
3.2
2.2
2.0
1.1
1.8
1.25
1.4

The Event-Related Brain Potential as an Index of Information Processing and Cognitive Activity: A Program of Basic Research

Prepared by:
The Cognitive Psychophysiology Laboratory

Prepared for:
The Air Force Office of Scientific Research
Life Sciences Directorate

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

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 88-1

5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 88-0316

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

6b. OFFICE SYMBOL (If applicable): NL

7a. NAME OF MONITORING ORGANIZATION: AFOSR

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

7b. ADDRESS (City, State and ZIP Code): Bldg 410, Bolling AFB DC 20332

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

8b. OFFICE SYMBOL (If applicable): NL

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

8c. ADDRESS (City, State and ZIP Code): Bldg 410, Bolling Air Force Base, DC 20332-6448

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2313	A4	

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, Michael Coles, Arthur Kramer

13a. TYPE OF REPORT: Final Technical

13b. TIME COVERED: FROM 87JAN01 TO 87DEC31

14. DATE OF REPORT (Yr., Mo., Day): 88FEB29

15. PAGE COUNT: 11 pgs + appendix

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, motor potentials, mental chronometry, information processing, memory, resource theory, workload, automaticity, communication

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

We review a program of research designed to understand the event-related potential (ERP) so that it can be used as a tool in the study of human information processing and in the assessment of man-machine systems. During the present contract year, we have focused on (a) the use of ERPs in the study of attention and skill acquisition, (b) the use of ERPs in the study of mental chronometry, (c) the use of ERPs in the study of mental resources and workload, (d) the use of ERPs in the study of memory, (e) the development of an animal model of the P300 component, and (f) the use of ERPs as a communication channel.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/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.

10 MAR 1988

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

AFOSR-TR- 88-0316

(2)

Department of Psychology
University of Illinois
Champaign, Illinois 61820

COGNITIVE PSYCHOPHYSIOLOGY LABORATORY

Final Technical Report No. CPL 88-1
F49620-85-C-0041
January 1988

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

A Program of Basic Research

Final Report

Prepared for:

The Air Force Office of Scientific Research
Life Sciences Directorate

TABLE OF CONTENTS

1. INTRODUCTION

This report describes research conducted in the Cognitive Psychophysiology Laboratory during the period 1/1/87-12/31/87 with the support, or partial support, of the present contract.

As in previous years, we have continued to pursue several closely related goals. Our primary mission has been to develop an understanding of the Event-Related Brain Potential (ERP) so that it can be used in the study of human cognitive function and in the assessment of man-machine interactions. To this end, we have conducted research in the following areas:-

a. The use of ERPs in the study of attention and skill acquisition

b. The use of ERPs in the study of mental chronometry

c. The use of ERPs in the study of mental resources and workload

d. The use of ERPs in the study of memory

e. The development of an animal model of the P300 component

f. The use of ERPs as a communication channel

g. Miscellaneous studies

Listed in Section 3 are all chapters, papers, abstracts and presentations that were published, submitted, or "in preparation" in 1987. Since this is the last year of the current contract, we have also listed publications (Sections 4 and 5) that report the results of all research supported by the contract in previous years (1/1/85-12/31/86).

Full reports of the studies for the current year are included in the Appendix. In the following sections, we provide a brief description of the research. Where appropriate, we have referred to the appropriate appendix items.

2. DESCRIPTION OF THE RESEARCH (1/1/87-12/31/87)

Numbers in parentheses refer to full reports of the research listed in Section 3 and provided in the Appendix.

2.1 Attention and Skill Acquisition

The focus of this research has been on the use of ERPs to provide converging evidence for the chronometric and energetical changes that take place during the development of highly skilled behaviors. To this end, we have recorded ERPs along with a number of traditional performance measures such as RMS tracking error, RT and accuracy in paradigms that allow for the development of "automatic" processing (12, 25). Several important results have been obtained from this research. First, we have preliminary evidence which argues that there are a number of dissociable automatic processes that develop at different rates. Second, the results of dual-task manipulations indicate that even tasks which possess "automatic" processes are susceptible to interference in a manner predicted by multiple resource models. Third, the use of P300 amplitude in conjunction with the manipulation of multi-task processing priorities has allowed us to map the attentional requirements of automatic and non-automatic processes.

2.2 Mental Chronometry

The focus of this research has been on the use of ERPs to measure the timecourse of mental processes. A related aim has been to use ERPs to examine the information processing mechanisms responsible for variability in measures of overt behavior. In our most recent research, we have used measures of the readiness potential, recorded from lateral electrode sites above the motor cortices, to assess response preparation. These measures, as well as those of the electromyogram (EMG), suggest that information transmission among processing elements is accomplished continuously - or at least in several steps (1 and 8).

A second line of research has focussed on measures of the readiness potential as predictors of response latency and accuracy. In contrast to the claims of Gevins et al., we have shown that this ERP measure can be used to predict, in advance of an overt response, whether the response will be accurate and how fast it will be (8). This finding clearly reveals the utility of the psychophysiological approach in accounting for variation in overt behavior.

A further line of research examines both the utility of P300 as an index of shifts in subject strategy (6) and as a marker for the duration of stimulus evaluation processes (22).

2.3 Mental Resources and Workload

This research has focused upon the use of ERPs as metrics of mental workload and resource allocation in multi-task environments (for reviews see 9, 15). Previous research conducted in our laboratory has indicated that P300 amplitude behaves in a manner predicted by multiple resource models. Thus, the amplitude of P300 appears to mimic the predicted resource tradeoffs as a function of task difficulty and processing priority. In recent research, we have determined that (a) the reciprocity in P300 amplitudes elicited by two concurrently performed tasks is predictive of single subject performance over a relatively wide range of task variables (20), (b) the P300 effect can be generalized to non-laboratory tasks such as instrument flight (10), (c) and that the visual N190 as well as the P300 components reflect resource tradeoffs both within and across tasks while the N160 is sensitive to only within task resource allocation (11).

2.4 Memory

We demonstrated earlier that there is a relationship between the P300 response to a stimulus and the subsequent memorability of that stimulus. This relationship was observed in subjects who used a rote memorization strategy, but not in those who used elaborative strategies. In a partial replication of this experiment, we demonstrated that the same effects can be observed within the same subjects (16). Furthermore, the data indicate how the P300 measure can be used to evaluate theories of memory that emphasize distinctiveness as a critical attribute in the memorability of events.

2.5 Animal Models

This research, which has been referred to in previous reports, demonstrates the presence of probability-sensitive neuronal activity in rabbits (21). This finding has implications for the search for the neural source of the P300, since probability sensitivity is one of the critical features of the P300.

2.6 ERPs and Communication

We have continued our work on the use of ERPs in communication. We have demonstrated that it is possible to determine, by evaluating of the P300 response to a display, which element in the display the subject has selected. When these elements are letters of the alphabet, the procedure enables subjects to use their ERPs to communicate (17).

2.7 Miscellaneous

In this category, we include studies of methodological issues, theoretical articles, and review papers.

We have continued our efforts to devise a metric for the assessment of topographic information. Without exception, an important defining characteristic of an ERP component is its scalp distribution. Furthermore, it is clear that the distribution of scalp-recorded brain potentials may provide critical clues concerning their neural source. However, until this time, no satisfactory methods have been proposed for the quantification of scalp distribution. A paper describing the current version of our procedure (18) and a validation study (19) are currently under final editorial review. This procedure provides the investigator with the opportunity of dealing with the problem of overlapping components, and of distinguishing statistically between different scalp distributions.

A second methodological contribution has been to extend the method of correction for eye-movement artifact to include both horizontal and vertical movements and to provide a program for the general public to implement the correction procedure (14).

We have published two theoretical articles this year. One deals with general issues concerning the role of theory in cognitive psychophysiology (3), the other examines in detail the context-updating model of the P300 (5). A third article deals with issues in the definition and reliability of measures of the P300 (7). In addition, we have published several chapters dealing with the utility of measures of the ERP in the study of human factors (9 and 15) and psychological issues (2), with the utility of measures of the cardiovascular system in human factors research (4), and with the analysis of aging using the psychophysiological approach (13).

3. PUBLICATIONS FOR THE CURRENT YEAR (1/1/87-12/31/87)

Note that full reports of those entries preceded by a number are provided in the appendix.

3.1 Papers and Chapters

1. Coles, M. G. H., Gratton, G., & Donchin, E. (in press). Detecting early communication: Using measures of movement-related potentials to illuminate human information processing. Biological Psychology.

2. Coles, M. G. H., Gratton, G., & Fabiani, M. (in press). Event-related brain potentials. To appear in Cacioppo, J. T. & Tassinary, L. G., (Eds.) Principles of Psychophysiology: Physical, Social, and Inferential Elements. Cambridge: Cambridge University Press.

3. Coles, M. G. H., Gratton, G., & Gehring, W. J. (1987). Theory in cognitive psychophysiology. Journal of Psychophysiology, 1, 13-16.

4. Coles, M. G. H., & Sirevaag, E. (1987). Heart rate and Sinus Arrhythmia. In A. Gale, & B. Christie (Eds.), Psychophysiology and the electronic workplace. London: John Wiley & Sons.

5. Donchin, E., & Coles, M. G. H. (in press). Is the P300 component a manifestation of context updating? The Behavioral and Brain Sciences.

6. 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.

7. Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (in press). 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., pp. 1-78.

8. Gratton, G., Coles, M. G. H., Sirevaag, E., Eriksen, C. W., & Donchin E. (in press). Pre- and post-stimulus activation of response channels: A psychophysiological analysis. Journal of Experimental Psychology: Human Perception and Performance.

9. Kramer, A.F. Event-related brain potentials. (1987). In A. Gale and B. Christie (Eds.), Psychophysiology and the Electronic Workplace. Sussex, England: John Wiley and Sons.

10. Kramer, A., Sirevaag, E., & Braune, R. (1987). A psychophysiological assessment of operator workload during simulated flight missions. Human Factors, 29(2), p. 145-160.

11. Kramer, A., Sirevaag, E., & Hughes, P. (in press). Effects of foveal task load on visual-spatial event-related brain potentials and performance. Psychophysiology.

12. Kramer, A. F., & Strayer, D. L. (in press). Assessing the development of automatic processing: An application of dual-task and event-related brain potential methodologies. Biological Psychology.

13. Miller, G. A., Bashore, T. R., Farwell, L. A., & Donchin, E. (1987). Geriatric psychophysiology. In K.W. Schaie (Ed.), Annual Review of Gerontology and Geriatrics, Volume 7 (pp. 1-27). New York: Springer-Verlag.

14. Miller, G. A., Gratton, G., & Yee, C. M. (in press). Generalized implementation of an eye movement correction procedure. Psychophysiology.

15. Wickens, C. W. (in press). Application of ERPs to human factors. In J. Rohrbaugh, R. Johnson, and R. Parasuraman (Eds.), Event-related potentials and the brain. New York: Oxford University Press.

3.2 Articles Submitted for Publication

16. Fabiani, M., Karis, D., & Donchin, E. (1987). Effects of strategy manipulation in a von Restorff paradigm. Submitted for publication.

17. Farwell, L. A. & Donchin, E. (submitted). Talking off the top of your head: A mental prosthesis utilizing event-related brain potentials. Electroencephalography and Clinical Neurophysiology.

18. Gratton, G., Coles, M. G. H., & Donchin, E. (1987). A procedure for using multi-electrode information in the analysis of components of event-related potentials: Vector Filter. Psychophysiology. Submitted for publication.

19. Gratton, G., Kramer, A. F., Coles, M. G. H., & Donchin, E. (1987). A simulation study of the latency measures of components of event-related brain potentials. Psychophysiology. Submitted for publication.

20. Sirevaag, E., Kramer, A., Coles, M. G. H., & Donchin. E. (1987). Resource reciprocity: An event-related brain potentials analysis. Acta Psychologica. Submitted for publication.

21. Stolar, N., Sparenborg, S., Donchin, E., & Gabriel, M. (1987). An animal model for the P300 component of the event-related potential in humans. Behavioral Neuroscience. Submitted for publication.

3.3 Abstracts

22. Jenkins, S., Gratton, G., Coles, M. G. H., & Donchin, E. (1987). P300 latency and task requirements (Abstract). Psychophysiology, 24, 594.

23. Sirevaag, E., & Kramer, A. (1987). N100 and P300 tuning effects during an attention switching task (Abstract). Neuroscience, 17, 654.

3.4 Presentations

Coles, M. G. H., Gratton, G., & Donchin, E. (1987). Changing minds: Using measures of movement related potentials to illuminate evaluation processes. Presented at The IV International Conference on Cognitive Neuroscience, Dourdan, France.

24. Gehring, W. J., Strayer, D. L., Kramer, A., Donchin, E., & Miller, G. (1987). An evaluation of age differences in the development of automaticity. Proceedings of The IV International Conference on Cognitive Neuroscience, Paris-Dourdan, France.

Gratton, G. (1987). The use of scalp distribution to separate components of the Event-Related Brain Potential. Presented at The IV International Conference on Cognitive Neuroscience, Dourdan, France.

Gratton, G. & Coles, M. G. H. (1987). Detecting early communication: Can response preparation begin before stimulus evaluation ends? Presented at the Annual Hoosier Mental Life Meeting, Monticello, IL.

Gratton, G., & Coles, M. G. H. (1987). Generalization and evaluation of eye-movement correction procedures. Presented at the Workshop on Removing Eye Movement Artefacts from the EEG, Tilburg, Holland.

25. Kramer, A. F., & Strayer, D. L. (1987). P300 operating characteristics: Performance/ERP analysis of dual-task demands and automaticity. Proceedings of the 4th International Conference of Neurosciences, Paris-Dourdan, France.

26. Strayer, D.L., Gehring, W.J., Kramer, A.F., & Miller, G.A. (1988, April). Adult age differences in the development of automaticity: A psychophysiological assessment. Paper to be presented at the Eleventh Psychology in the Department of Defense Symposium, Colorado Springs.

4. PUBLICATIONS FOR 1/1/85-12/31/85

4.1 Papers 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. Donchin, E., Miller, G. A., & Farwell, L. A. (1986). The endogenous components of the event-related potential -- A diagnostic tool? In E. Fliers (Ed.), Progress in Brain Research. Amsterdam: Elsevier.

4. 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.

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

6. 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.

4.2 Abstracts and Conference Presentations

7. 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.

8. 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.

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

5. PUBLICATIONS FOR 1/1/86-12/31/86

5.1 Papers and Chapters

1. Coles, M. G. H., & Gratton, G. (1986). 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, 409-425.

2. Coles, M. G. H., Gratton, G., Kramer, A., & Miller, G. A. (1986). Principles of signal acquisition and analysis. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), Psychophysiology: Systems, Processes, and Applications. Volume I : Systems (pp. 183-221). New York: Guilford Press.

3. Donchin, E., Karis, D., Bashore, T. R., Coles, M. G. H., & Gratton, G. (1986). 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 (pp. 244-267). New York: Guilford Press.

4. Donchin, E., Kramer, A. F., and Wickens, C. (1986). Applications of brain event-related potentials to problems in engineering psychology. In M. G. H. Coles, E. Donchin, and S. W. Porges (Eds.), Psychophysiology: Systems, Processes, and Applications. New York: Guilford Press.

5. Fabiani, M., Karis, D., & Donchin, E. (1986). P300 and recall in an incidental memory paradigm. Psychophysiology, 23, 298-308.

6. Fabiani, M., Karis, D., & Donchin, E. (1986). P300 and memory. In W. C. McCallum, R. Zappoli, & F. Denoth (Eds.) Psychophysiology: Studies in Event-Related Potentials. Supplement 38 to Electroencephalography and Clinical Neurophysiology, 63-69.

7. Gratton, G., Coles, M. G. H., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1986). An ERP/EMG/RT approach to the continuous flow model of cognitive processes. In W. C. McCallum, R. Zappoli, & F. Denoth (Ed.), Cerebral Psychophysiology: Studies in Event-Related Potentials, Suppl. 38 to Electrophysiology and Clinical Neurophysiology, 120-122.

8. Hockey, G. R. J., Coles, M. G. H. & Gaillard, A. W. K. (1986). Energetical issues in research on human information processing. In G. R. J. Hockey, A. W. K. Gaillard, & M. G. H. Coles (Eds.), Energetics and Human Information Processing. Dordrecht, The Netherlands: Martinus Nijhof, pp. 3-21.

9. Kramer, A. F., and Donchin, E. (1986). Brain potentials as indices of orthographic and phonological interaction during word matching. Journal of Experimental Psychology: Learning, Memory and Cognition, 13, 76-86.

10. Kramer, A. F., Schneider, W., Fisk, A., and Donchin, E. (1986). The effects of practice and task structure on components of the event-related potential. Psychophysiology, 23, 33-47.

5.2 Abstracts and Conference Presentations

11. Bosco, C. M., Gratton, G., Kramer, A. F., Coles, M. G. H., Wickens, C., & Donchin, E. (1986). Partial information and components of the Event-Related Brain Potential (Abstract). Psychophysiology, 23, 426.

12. Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1986). The reliability of measurement of the P300 component of the ERP. (Abstract). Psychophysiology, 23, 434.

13. Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1986, abstract). The reliability of measurement of the P300 component of the event-related brain potential. Psychophysiology, 23, 434.

14. Farwell, L.A., and Donchin, E. (1986). The "Brain Detector:" P300 in the detection of deception. Psychophysiology, 23, 434.

15. Farwell, L.A., Donchin, E., and Kramer, A.F. (1986). Talking heads: a mental prosthesis for communicating with event-related potentials of the EEG. Psychophysiology, 23: 434.

16. Gehring, W., Gratton, G., Coles, M. G. H., & Donchin, E. (1986). Response priming and components of the Event-Related Brain Potential (Abstract). Psychophysiology, 23, 437-438.

17. Gratton, G., & Coles, M. G. H. (1986). Lateralized brain potentials and priming (Abstract). Psychophysiology, 23, 416-417.

18. Kramer, A. F. (1986). Interaction between workload and Training: Converging Evidence from Psychophysiology and Performance Measurement. Proceedings of the Human Factors Society, 30, 1137-1141.

19. Strayer, D. L, Coles, M. G. H., Buckley, J., & Donchin, E. (1986). Stimulus repetition effects: Evidence for processes in cascade. Psychophysiology, 23, 466.

Proceedings of the Fourth International Conference on Cognitive Neuroscience (Biological Psychology), in press.

Detecting early communication: Using measures of movement-related potentials to illuminate human information processing

Michael G. H. Coles*, Gabriele Gratton, & Emanuel Donchin

Cognitive Psychophysiology Laboratory

University of Illinois

Champaign, Illinois 61820

USA

Running Title: Early communication

* Requests for reprints should be sent to Michael G.H. Coles, University of Illinois, Psychology Department, 603 E. Daniel St., Champaign, Illinois, 61820, U.S.A.

Abstract

In this paper we review evidence that suggests that the stimulus evaluation system can pass information to the response activation system before evaluation is completed ("early communication"). This evidence is derived from measures of the eralized readiness potential, which have been related in previous research to the preparation for movement. Early communication is evident in conflict and congruence paradigms. In both paradigms, a single stimulus, or two different stimuli, deliver two aspects of information. In the conflict paradigm, the first aspect of information (derived from preliminary evaluation) primes the incorrect response, while the second primes the correct response. In the congruence paradigm, information derived from preliminary and complete evaluation is congruent. In both paradigms, lateralized readiness potential measures suggest that preliminary evaluation is able to prime the response system, although the overt motor response may not be released until evaluation is completed. This demonstration of early communication has both theoretical and practical implications. First, it does not support single-decision models of information processing. Second, it suggests that the lateralized readiness potential, a continuous, analog measure of the activity of the response system, can be used to make inferences about the nature of the evaluation process, and to localize the effects of various manipulations on the information processing system.

Detecting early communication: Using measures of movement-related potentials to illuminate human information processing

Michael G. H. Coles, Gabriele Gratton, & Emanuel Donchin

1. Introduction

A central issue in contemporary research on human information processing concerns the nature of the transmission of information between elementary information processing activities (e.g., Meyer, Yantis, Osman, & Smith, 1986, and Meyer, Osman, Irwin, & Yantis, this volume). Two classes of models can be identified: discrete models, that assume that information is transmitted discretely, only when processing at a particular level of the system (or stage) is completed (e.g., Sanders, 1980; Sternberg, 1969), and continuous models, that assume that information is transmitted continuously, as soon as it is available (e.g., Eriksen & Shultz, 1979; Grice, Nullmeyer, & Spiker, 1982; McClelland, 1979). A hybrid model has been proposed by Miller (1982), the asynchronous discrete coding model. Like discrete models, this model assumes that information is transmitted discretely, but like continuous models, it assumes that information may be transmitted before processing at a particular level is completed. Thus, information is transmitted in "chunks," whose size and number may vary as a function of the nature of the information.

The psychophysiological evidence we review in this paper suggests that communication between stimulus evaluation and response activation systems can take place continuously, or at least in chunks. In particular, the activity of the motor system can be influenced by preliminary phases of stimulus evaluation.

1.1 Early Communication

To address the question of the nature of communication, it is assumed that the information processing system can be represented by two major subsystems: a stimulus evaluation system responsible for the identification of stimulus information, and a response activation system that is directly responsible for the generation of overt behavioral responses. This assumption, which is prevalent in contemporary theorizing about information processing (see, for example, Grice et al., 1982; Posner, 1978), can be traced directly to the work of Donders (1868/1969) and his followers (e.g. Sanders, 1980; Sternberg, 1969). A further assumption is that each of these systems can pass through intermediate, states, or levels of activation, before it achieves a threshold level (Eriksen & Schultz, 1979; McClelland, 1979; Meyer et al., 1986; Miller, 1982). For the stimulus evaluation system, this threshold level corresponds to "complete" evaluation of the stimulus (by which we mean that sufficient information has been extracted from the stimulus to enable the subject to produce the correct response). For the response system, the threshold level corresponds to the initiation of the response. These ideas are represented by the activation by time functions in Figure 1.

Insert Figure 1 About Here

Given these assumptions, the question of communication between stimulus evaluation and response systems can be rephrased as follows: can the stimulus evaluation system transmit information to the response system before evaluation is completed? If so, then the response system should be influenced by preliminary phases of evaluation. This type of influence can be referred to as "early communication." Note that, in Figure 1, the presence of early communication is associated with a subthreshold increase in the level of response activation.

2. Measuring Partial Response Activation

The response activation function depicted in Figure 1 is, of course, hypothetical. To explore this function and detect early communication we must identify procedures that can be used to reveal the presence of subthreshold response activation. One approach to this problem is to use experimental manipulations which allow one to infer the presence of "early communication" effects from measures of overt behavior (e.g., Miller, 1982; Proctor & Reeve, 1985; Kounios, Osman, & Meyer, 1987). This approach may also incorporate the analysis of reaction time distributions (e.g. Logan & Cowan, 1984; Meyer et al., 1986). Another approach involves the use of measures of peripheral subthreshold motor activity (the electromyogram) to infer the presence of partial response activation (Coles, Gratton, Bashore, Eriksen, & Donchin, 1985; Eriksen, Coles, Morris, & O'Hara, 1985).

In this paper, however, we focus on event-related brain potential measures, and in particular on the lateralized negative potentials that have been found by many investigators to precede left- and right-hand overt responses. As we will discuss later, these measures satisfy several criteria for measures of response activation. First, they are intimately related to the the activity of the motor system. Second, a fixed relationship can be observed between these measures and the emission of the overt response. Third, they are affected by experimental variables that are assumed to produce response activation (e.g., priming, bias, etc.). Fourth, they are related to the parameters of the subsequent overt response (accuracy and latency).

These lateralized negativities were first observed in studies of voluntary movements (Kornhuber & Deeke, 1965). When subjects anticipate making a response with a particular hand, an increase in negativity occurs that is larger at scalp sites contralateral to the responding hand (Kutas & Donchin, 1974; Vaughan, Costa, & Ritter, 1972).

The scalp sites at which these potentials are maximal are those over lateral central areas (C3 and C4) or over adjacent sites (frequently described as C3' and C4'). Note that these sites are in close proximity to those areas of the brain that are presumed to control movement (Penfield & Jasper, 1954). Of course, proximity to a particular brain area is no guarantee that an electrode is sensitive to activity of that brain area. However, in this case, there is evidence from a variety of studies, some of which we review below, to relate the lateralized negative scalp potentials to activity in motor brain areas.

We refer to these negativities as the "lateralized readiness potential" and will focus in this paper on the difference in potential between scalp sites contralateral and ipsilateral to particular responses.

The claim that this potential is related to the activity of the motor system is supported by functional and neurophysiological evidence. First, as we noted previously, this potential precedes movements. Second, neurophysiological data suggest that the source of the potential may lie, at least in part, in motor regions of the cortex (e.g., Arezzo & Vaughan, 1975; Okada, Williamson, & Kaufman, 1982). There are also strong similarities between the activity of units in the motor cortex and the potential (Requin, 1985). For these reasons, it is clear that the potential is related to motor activity.

The *fixed* relationship between the lateralization of the readiness potential and the production of the motor response is evident in data from Gratton, Coles, Sirevaag, Eriksen, and Donchin (in press). In this study, a fixed level of lateralization was observed at the moment of the onset of EMG activity, regardless of response accuracy or latency.

A variety of studies have demonstrated that the measure is influenced by those variables that are assumed to produce partial response activation. In one class of studies, information about a future response is provided by a valid precue. Such information should be associated with a sub-threshold increase in activation of the predicted response. In the most extreme case, the precue provides complete information about the response to be given, and the imperative stimulus only provides temporal information. For example, Kutas and Donchin (1980) studied the lateralized readiness potential in the context of a two-choice reaction time task, in which the two responses were assigned to left and right hands. In one condition the precue (warning stimulus) perfectly predicted the imperative stimulus that would occur one second later. In another condition, the precue did not provide advance information. In the precued condition, the lateralized readiness potential developed during the foreperiod with greater negativity at scalp sites contralateral to the response associated with the predicted stimulus. In the non-precued condition, on the average, lateralization was not evident until after the response relevant information was delivered by the imperative stimulus. These data are presented in Figure 2. Rohrbaugh, Syndulko, and Lindsley (1976) and Gaillard (1978) observed a similar lateralization of the readiness potential during the foreperiod of simple warned reaction time tasks.

Insert Figure 2 About Here

Lateralization effects can also be observed when the precue provides only probabilistic information about the upcoming imperative stimulus (and future response). In fact, when the validity of the precue is .8, lateralization effects are still present (Bosco et al., 1986; Gehring, Gratton, Coles, & Donchin, 1986).

Another line of evidence comes from studies of the relationship between lateralization in the foreperiod of reaction time tasks and the accuracy and latency of responses. In particular, the probability that a particular fast-guess response will be given depends on the degree of lateralization during the foreperiod (Gehring et al., 1986; Gratton et al., in press). Additional evidence indicates that, in a simple reaction time task, fast responses are associated with greater lateralization than slow responses (Rohrbaugh et al., 1976).

Evidence for the role of lateralized brain electrical activity in the preparation for movement is also provided by studies of the effects of externally applied direct currents on reaction time. Birbaumer and his colleagues (e.g., Jaeger, Elbert, Lutzenberger, & Birbaumer, in press) applied currents of different directions at C3 and C4. This procedure is believed to result in lateralized polarization of cortex. These investigators report a small, but significant effect on reaction time as a function of the hand used to respond and direction of current flow.

In summary, data from a variety of sources converge in suggesting that the lateralized readiness potential can be used as a measure of the relative activation of responses (when the response are assigned to different hands). The potential (a) appears to be generated (at least in part) in motor regions of the brain, (b) has a fixed level at the moment of response emission, (c) is influenced by those variables that affect response activation, and (d) is related to the accuracy and latency of overt behavioral responses.

In measuring the lateralized readiness potential it is important that steps are taken to ensure that only movement-related potentials are represented in the lateralization waveforms. First, experimental conditions are established so that parameters of left- and right-hand responses (such as response latency) are comparable. Second, averages for left- and right-hand responses are computed separately. Third, these averages are then combined with equal weight. If these steps are followed, asymmetrical activity unrelated to the side of movement should average to zero leaving only movement-related activity (see also De Jong, Wierda, Mulder, & Mulder, in press; Gratton et al., in press).

3. Early Communication Paradigms

Having found a suitable measure of partial response activation, we now need to consider paradigms in which we can use the measure to investigate the question of early communication between stimulus and response systems. Recall that we are trying to determine whether preliminary stimulus evaluation can lead to partial response activation. Thus, a suitable paradigm must provide the opportunity for the stimulus evaluation system to transmit preliminary information to the response system that is in some way distinguishable from that derived from complete evaluation. By "complete evaluation" we mean evaluation of all the information required to give the correct response. In the present paper we consider two paradigms that appear to provide this opportunity, the conflict paradigm and the congruence paradigm.

3.1 The Conflict Paradigm

The conflict paradigm is based on the idea that two aspects of the stimulus information prime different responses. Furthermore, the manipulations are arranged such that the two aspects are analyzed (and are therefore available to the response system) at different moments in time. Thus, the effects of preliminary and complete evaluation are in conflict. In particular, experimental contingencies in a two-choice reaction time task are arranged such that preliminary evaluation should lead to one response, while complete evaluation should lead to the other response. If the responses are assigned to different hands, brain potentials associated with one response should be distinguishable from those associated with the other response. Since preliminary evaluation should lead to the activation of the incorrect response, while complete evaluation should lead to activation of the correct response, a reversal in lateralization when conflicting information is presented provides evidence for early communication. These ideas are represented diagrammatically in Figure 3.

Insert Figure 3 About Here

We describe here three different instances of the conflict paradigm. They involve different degrees of separability of the two aspects of the information.

In the first instance (Precueing Task, e.g., Bosco et al., 1986; Gehring et al., 1986), preliminary information is delivered by one stimulus, while complete information is delivered by a second stimulus presented sometime later. Thus, information 1 and 2 (see Figure 3) are delivered at different times.

In the second instance (Noise-Compatibility Task, e.g., Gratton et al., in press; Smid, Mulder, & Mulder, 1988), all the information is presented in a single visual array. However, the conflicting response-relevant dimensions of the array are spatially separated. A target letter, to which the subject must respond, is surrounded by noise letters that call for the incorrect response.

In the third instance ("Stroop" Task, Buckley, Gratton, Kramer, Coles, & Donchin, 1988), the competing response-relevant dimensions are provided by the same stimulus. The stimulus is a word whose orthographic attributes call for the incorrect response and whose phonologic characteristics call for the correct response.

Thus, the three examples represent different degrees of separability of the two response-relevant stimulus dimensions. In the precueing task, they are temporally separated. In noise-compatibility and "Stroop" tasks, they are different attributes of the same stimulus. In the noise-compatibility task, the attributes are also spatially distinguishable.

3.1.1 The Precueing Task. In this task (Bosco et al., 1986; Gehring et al., 1986) one of two warning stimuli informs the subject that the same, or the other, stimulus will appear after a second or so with a particular probability (e.g., .8). Subjects must execute a response to the second stimulus (i.e., imperative stimulus) with one hand or the other. In this situation, reaction time data suggest that subjects generally use the warning information to prime the appropriate response.

As can be seen in the upper panel of Figure 4, the lateralized

readiness potential data suggest that subjects prime the appropriate response during the foreperiod, when the precue provides useful information. As the lower panel of Figure 4 indicates, when the precue has no predictive value, there is, on the average, no lateralization during the foreperiod. The lateralization only develops after the imperative stimulus.1
In the predictive condition (upper panel), if the imperative stimulus

---

Insert Figure 4 About Here

---

matches the warning stimulus, the degree of lateralization increases in the direction of the predicted (same) response. However, if an unpredicted imperative stimulus is presented, as happens on 20% of the trials, then the lateralization reverses in the direction of the (different) response. These data suggest that, when conflicting information is delivered at different points in time, subjects begin by priming one response and then reverse their priming and execute the other response.

The next two experiments address a more challenging question: If information calling for conflicting responses is carried by the same stimulus, will we see the same pattern of reversal in lateralization? That is, will the changes in lateralized readiness potential evident in Figure 4 be compressed into the post-stimulus epoch? If the answer to this question is "yes," then we will have evidence that preliminary phases of evaluation can lead to the activation of the response system.

3.1.2 The Noise-Compatibility Task. In this task (Gratton et al, in press; Smid et al., 1988), an array of letters is presented on a screen, and subjects are instructed to respond with their left or right hands depending on the "target" letter appearing on the center of the array. Letters that flank the target letter (noise letters) can be the same as the target (compatible-noise trials), or those that call for the other response (incompatible-noise trials).

Coles et al. (1985) and Gratton et al. (in press) have shown that information about the letters in the array is available during preliminary phases of stimulus evaluation, but that subjects cannot locate the target letter until later (cf. Treisman & Gelade, 1980). Therefore, we should see the reversal in the lateralized brain potential when the target letter and noise letters conflict - if early communication occurs.

Data from the Gratton et al. (in press) study are presented in Figure 5. In the top panel, the overt behavioral data are presented in the form of speed-accuracy trade-off function that Luce (1986) calls "conditional accuracy functions." The function for incompatible noise stimuli reveals that there is a response latency (between 150 and 250 ms) for which accuracy is less than chance. This is not true for compatible stimuli. Accuracy is above chance for responses to compatible noise trials with a latency between 150 and 250 ms. This indicates that, when subjects respond at this latency, the accuracy of their responses is more dependent on the type of noise letters than on the target letter. It appears that, at this latency, the subject has identified what the letters are, but does not know where they are.

Insert Figure 5 About Here

For responses of longer latency, response accuracy is high regardless of the compatibility of the noise. This suggests that the accuracy of these responses is the result of correct identification and localization of the target letter.

Thus, the conditional accuracy functions support the idea that preliminary stimulus evaluation results in identification of the letters in the array, and that later, "complete" evaluation results in the location of the target letter.

Lateralized readiness potential data are shown in the middle panel of Figure 5. Note that these functions, like the conditional accuracy functions suggest that there is a time following stimulus presentation when the incorrect response is preferentially primed if the array contains incompatible noise.

Although suggestive, these data do not necessarily support the idea of early communication. In fact, both lateralization waveforms and conditional accuracy functions are the result of aggregating data over trials - and, for this reason, they may not present an accurate picture of what is going on during individual trials. The lateralization data are based on averages of all trials regardless of the accuracy of the response. We know that the readiness potential lateralizes just prior to the execution of an overt response. For incorrect responses, the lateralization would be associated with a downward deflection of the trace. Thus, the reversal seen in the average lateralization waveform for incompatible trials may be due to the fact that, at intermediate latencies, the average lateralization values reflect the contribution of the larger number of trials with incorrect responses at these latencies. The evidence we need to demonstrate that early communication can occur is a reversal in the lateralization on individual trials. Obviously, this is a problem because one must aggregate data in order to enhance the signal/noise ratio of the lateralized readiness potential.

We reasoned that a solution to this problem was to select a subset of trials for which only the correct overt response was given and for which overt response latency was relatively constant (300-349 msec). We have evidence that responses of this latency are not likely to be the result of fast guesses (Gratton, et al., in press) and that the subjects were responding mainly on the basis of a complete evaluation of the stimulus. Thus, any reversal in the lateralized readiness potential seen for this subset of trials cannot be attributed to the inclusion of incorrect response trials. Rather, it should be attributed to partial incorrect response activation as a result of early communication on these trials.

Averages for these trials are shown in the lower panel of Figure 5. Note that, on incompatible response trials the reversal in lateralization is clearly visible and significant.2 These data support the inference that early communication is possible and that, at least on some trials in this task, preliminary information was transmitted to the response system before the stimulus was completely evaluated.

Similar data have been obtained by Smid et al. (1988). Their data are shown in Figure 6. More conditions were included in this experiment; however, the reversal in lateralization is clearly evident in the data for incompatible noise. These averages are based on correct response trials only - but the latency of the overt response was not restricted.

Insert Figure 6 About Here

3.1.3 The "Stroop" Task. In this task, conflicting information is carried by different attributes of the same stimulus. In the classic Stroop paradigm, names of colors are written in different colored inks. When the color name and ink color are different, and the subjects must name the ink color, conflict is evident in reaction time and accuracy measures.

Buckley et al. (1988) studied the lateralized readiness potential in a variant of the Stroop task. In this experiment, information derived from an orthographic analysis of a visually presented word conflicted with information derived from a phonological analysis (cf. Polich, McCarthy, Wang, & Donchin, 1983; Kramer & Donchin, 1987). Two words, separated by 1000 ms, were presented to the subject. Four types of word pairs were used, in which:

a. Words rhymed and looked alike (e.g. COOK - BOOK);

b. Words rhymed and did not look alike (e.g. DOUGH - FLOW);

c. Words did not rhyme but looked alike (e.g. DOVE - MOVE);

d. Words neither rhymed nor looked alike (e.g. TABLE - SOAP).

In one session, the subjects' task was to judge whether the words rhymed, in another, subjects had to judge whether the words looked alike. Note that conflict is present in two types of word pairs: DOUGH - FLOW and DOVE - MOVE. For half the subjects, "yes" judgements were indicated with the right hand, and "no" judgements with the left. The converse was the case for the other subjects.

Reaction time data revealed that the effects of conflict were much more potent in the rhyme task than in the visual task. A detailed analysis of conditional accuracy functions suggested that information about the orthography was available before information about phonology.

The question is, therefore, does the reversal in lateralization occur for conflict words in the rhyme task? No reversal would be expected in the visual task because the correct response can be made on the basis of the orthographic information, which is available first. The relevant data are shown in the upper panel of Figure 7. In the conflict conditions, when the words look alike but do not rhyme, or rhyme but do not look alike, there is a reversal in the lateralized brain potential.3 When there is no conflict, there is no reversal. The lateralization data for the visual task are presented in the lower panel of Figure 7. When the subject's task is to judge whether the words look alike (the visual task), there is no reversal in lateralization when phonology conflicts with orthography.

Insert Figure 7 About Here

3.1.4 Summary. The data from the different instances of the conflict paradigm converge in suggesting that early communication between the stimulus evaluation and response systems does indeed occur. Whether conflicting information is presented at different moments in time, or by different attributes of the same physical stimulus, subjects appear to activate the incorrect response based on preliminary evaluation. The correct response may be activated later, when more information about the stimulus is processed.

3.2 Congruence Paradigm

In the previous section, we have seen how the lateralized readiness potential provides evidence for early communication when the effects of preliminary evaluation conflict with those of "complete" evaluation. As with the conflict paradigm, the congruence paradigm focuses on the idea that a particular response is somehow influenced by two distinct aspects of stimulus information. However, while in the conflict paradigm the information is in conflict, in the congruence paradigm it is congruent. The two aspects of information may be delivered by two different stimuli presented at different times, or by the same stimulus. In the latter case, the experimental contingencies are such that one aspect of the information is available before the other. The congruence paradigm is illustrated in Figure 8.

Insert Figure 8 About Here

The Bosco et al. (1986) and the Gehring et al. (1986) experiments, described earlier, provide examples of a situation in which a warning stimulus delivered information that was congruent with a subsequent imperative stimulus 80% of the time. As can be seen in Figure 4, under these circumstances lateralization develops in the foreperiod in the direction of the expected response, and when a congruent imperative stimulus is presented, the lateralization further develops in the same direction. Can we see a similar influence of partial priming when the two aspects of the stimulus information are presented simultaneously? In the Gratton et al. (in press) study, this situation is realized when target and noise letter are the same (the compatible-noise condition). As the bottom panel of Figure 5 shows, this condition is associated with an early lateralization in the direction of the correct response, which is quite distinguishable from the later lateralization, and which is symmetrical to the lateralization in the direction of the incorrect response in the incompatible-noise condition.

In the preceding examples, the number of alternative responses was always two, and the target stimulus could vary along one dimension. Furthermore, the partial information (the precue or the noise letters) was not always valid. In contrast, in the Miller paradigm (Miller, 1982), used by De Jong et al. (in press), the partial information is always valid, but it is incomplete. In this paradigm, subjects perform a four-choice reaction time task, in which responses are assigned to the index and middle fingers of each hand. Imperative stimuli vary along two dimensions (e.g., size and letter name), and each response is assigned to a unique combination of the two dimensions. In one condition, a precue (occurring .5 sec before an imperative stimulus) provided information that the forthcoming imperative stimulus would require a response by one of two fingers of a particular hand. In this case, a lateralization developed in the foreperiod in the direction associated with the cued hand. Following the imperative stimulus which specified the finger to be used in the response, the lateralization continued to develop in the same direction. Lateralization in the foreperiod was not observed when the precue specified the finger (on different hands) that would have to be used.

In another condition, De Jong et al. (in press) sought to determine whether early lateralization would occur when the partial information about hand and complete information about finger were delivered by the same stimulus. To insure the temporal dissociation between the two aspects of the information, the discriminability of the two stimulus dimensions was made different. It was assumed that information concerning the easy-to-discriminate dimension would be processed faster than the information concerning the difficult-to-discriminate dimension. Thus, if early communication exists, there should be partial response priming on the basis of the easy dimension. Three basic stimulus/response assignments were used. In the first, analysis of the easy dimension should permit the subject to prime the response hand (PREPARE HANDS), in the second, the same finger on each hand (PREPARE FINGERS), and, in the third, a different finger on each hand (PREPARE NEITHER).

De Jong et al. (in press) predicted that partial response priming should be revealed by earlier lateralization of the readiness potential in the PREPARE HANDS condition, than in the other two conditions. This is because information provided by the easy-to-discriminate dimension should permit the subject to start preparing the response hand before selecting the appropriate finger if information about the hand was available early in the evaluation process. Note that early communication might occur in the other two conditions, but it would not be detectable since it would result in partial priming of individual fingers on both hands.

In fact, De Jong et al. (in press) found no evidence for earlier lateralization in the PREPARE HANDS condition. The onset of the lateralization did not differ among the three conditions. These data suggest, therefore, that response priming on the basis of preliminary stimulus evaluation does not occur in this case.

3.2.1 Summary. In contrast to the conflict paradigm, the congruence paradigm does not provide consistent support for the idea of early communication. When the congruent information is delivered at different moments in time by two stimuli, the readiness potential clearly suggests that the subjects prime their responses on the basis of the preliminary, partial information. When congruent information is delivered by the same stimulus, the effect of early communication is not always apparent. It is evident when, as in the compatible condition of the noise-compatible task, the preliminary information is sufficient to lead to the preparation of one, and only one response. It is not evident, however, in the Miller task, used by De Jong et al. (in press), where preliminary information is compatible with more than one response. In the PREPARE-HANDS condition, such information identifies the hand to be used in responding, but does not specify the finger.

This difference between these two kinds of congruence tasks may be associated with different strategies. Since preliminary evaluation in the Miller task is insufficient to define the required response, subjects may prime that response to a lesser degree than when the response is completely specified. Thus, in the Miller task the lateralization associated with partial evaluation will be smaller, and therefore less detectable, than in the noise-compatibility task. Two other possibilities are suggested by De Jong et al. (in press). First, subjects may not use the results of preliminary evaluation in the Miller task to prime their responses. Instead, they may wait until the stimulus has been completely evaluated and then select and prepare the required response in one step, rather than selecting the response hierarchically (e.g., hand first, then finger). In this case, of course, no effects on the lateralized readiness potential would be expected. Second, while subjects may select their responses hierarchically, such a strategy may not be driven by the relative time at which information about the two dimensions is available, but by the translation rules that map stimuli to responses. In terms of the model we presented in Figure 1, these strategic effects can be conceptualized in terms of adjustments in the way in which stimulus evaluation and response systems communicate (Logan, 1980).

4. Discussion

The data reviewed in the preceding sections suggest that, under at least some conditions, the readiness potential measures detect the presence of early communication. Different aspects of the stimulus information produce distinguishable effects on the response system. In the conflict paradigm, the incorrect response can be activated first at a subthreshold level as a result of preliminary stimulus evaluation. Later on, the correct response is activated at a threshold level as a result of "complete" evaluation. In the congruence paradigm, the correct response can be activated first at a subthreshold level by partial evaluation, and later at a threshold level by complete evaluation.

Early communication is not predicted by models that postulate a single decision at the interface between the stimulus evaluation and the response systems (Sternberg, 1969). In fact, such models assume that information is transmitted between different elements of the information processing system in a discrete, all-or-none fashion. In some of the experiments we have described, information is clearly not transmitted in a single chunk. In this sense, the data are compatible with continuous flow models (Eriksen & Schultz, 1979; Grice et al., 1982), but they are also compatible with the asynchronous discrete model of Miller (1982). Although the form of the lateralized readiness potential suggests a continuous modulation of the level of response activation, this continuity maybe an artifact of the averaging process. A small number of discrete "quanta" of activations of the correct and incorrect response may produce the pattern of lateralization data we observe in the conflict and congruence paradigms. In fact, at the present time, we have no way to distinguish between models that propose continuous communication and those that propose communication in multiple chunks.

The possibility of early communication means that the response system can be sensitive to intermediate phases of the stimulus evaluation process. Therefore, measures of partial response activation provided by the lateralized readiness potential can be used to investigate how stimulus information is analyzed. For example, the latency of the lateralization of the readiness potential could be used as an index of the timing of intermediate evaluation phases (see Buckley et al., 1988; De Jong et al., in press; Gratton et al., in press). When lateralization is observed, following stimulus presentation, the subject presumably has access to a particular aspect of the stimulus information.

Even in those cases in which the lateralization of the readiness potential does not reveal the presence of early communication, this measure can be used to make inferences about the strategies used by the subject to perform the task. We have seen an example of this in the De Jong et al. (in press) study. Note that to use the lateralization data in this way it is essential to demonstrate that the different dimensions of stimulus information manipulated (a) can be processed independently, and (b) activate responses at different times. Only when it is possible to demonstrate that the subject might have performed the task differently, can it be concluded that strategic choices, rather than cognitive limitations, are responsible for the failure to detect early communication.

The apparent sensitivity of the lateralized readiness potential to subthreshold response activation suggests a number of other situations in which this measure can be used to understand the information processing system. For example, it could be used to determine whether a particular interference effect involves subthreshold activation of inappropriate responses, or is instead localized at more central levels of the information processing system. Questions about the locus of different stimulus-response compatibility effects (e.g., Magliero, Bashore, Coles, & Donchin, 1984; Ragot, 1984) may be addressed in this way. Furthermore, the onset of the lateralization of the readiness potential can be used as a marker for the occurrence of a particular internal event. In fact, the information processing system must have distinguished between two alternative hypotheses about the stimulus by the time the lateralization appears.

In general, then, the lateralization measure can be used, in conjunction with other psychophysiological measures (EMG onset, P300 latency, etc.) and measures of overt behavior (reaction time, accuracy, etc.) to explore the complexities of human information processing. In this paper, we have described one way in which the lateralization of the readiness potential can be used, namely in the detection of "early communication."

Footnotes

1. When the warning stimulus predicted the imperative stimulus with a probability of .8, the average value of lateralization for the last 300 ms of the foreperiod (-0.68 microvolts) was significantly less than 0, $t(5)=-3.38$, $p<.05$. When the warning stimulus had no predictive value, the average lateralization (-0.18 microvolts) was not significantly different from 0, $t(5)=-1.30$.

2. The average activity in the interval between 170 and 210 ms after array presentation was significantly positive, $t(5)=2.78$, $p<.05$.

3. The average activity in the interval between 310 and 350 ms after the presentation of the second word was significantly more negative in the *no-conflict than in the conflict condition*, -0.31 vs. 0.09 *microvolts*, $t(9)=-2.78$, $p<.05$. However, the average activity for the conflict condition (the "dip") visible in Figure 7 was not significantly different from 0, $t(9)=0.71$. Note that in this task, as compared to the noise-compatibility task, there was considerably more variability in reaction time. This implies that separate averages should be computed for trials with different response latencies (as we did in Gratton et al., in press). However, this was not possible, since only a limited number of word pairs could be obtained for the "conflict" condition (a total of 110). Therefore, the power of this experiment was considerably less than that of the other experiments described in this paper, and the lack of a significant result is certainly not surprising. Note, however, the consistency of the waveforms obtained for the "conflict" and "no conflict" conditions with those obtained in comparable conditions of the noise-compatibility task.

Acknowledgements

The preparation of this manuscript was made possible, in part, by a grant from the National Institute of Mental Health, MH41445. We wish to thank Ritske De Jong, Monica Fabiani, Bill Gehring, Gordon Logan, Richard Ragot, Bernard Renault, and Dave Strayer for their comments on earlier versions of this manuscript.

References

Arezzo , J., & Vaughan, H.G. Jr. (1975). Cortical potentials associated with voluntary movements in monkey. Brain Research, 88, 99-104.

Bosco, C. M., Gratton, G., Kramer, A. F., Wickens, C. D., Coles, M. G. H., & Donchin, E. (1986). Information extraction and components of the event-related brain potential. Psychophysiology, 23, 426 (Abstract).

Buckley, J., Gratton, G., Kramer, A.F., Coles, M.G.H., & Donchin, E. (1988). A psychophysiological analysis of the priming of responses in orthographic and phonological tasks. In preparation.

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.

De Jong, R., Wierda, M., Mulder, G., & Mulder, L. J. M. (in press). The use of partial information in response processing. Journal of Experimental Psychology: Human Perception and Performance.

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

Eriksen, C. W., Coles, M. G. H., Morris, L. R., & O'Hara, W. P. (1985). An electromyographic examination of response competition. Bulletin of the Psychonomic Society, 23, 165-168

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

Gaillard, A. (1978). Slow brain potentials preceding task performance. Doctoral Dissertation. Amsterdam: Academische Pers B. V.

Gehring, W., Gratton, G., Coles, M. G. H., & Donchin, E. (1986). Response priming and components of the event-related brain potential. Psychophysiology, 23, 437 (Abstract).

Gratton, G., Coles, M. G. H., Sirevaag, E., Eriksen, C. W., & Donchin E. (in press). Pre- and post-stimulus activation of response channels: A psychophysiological analysis. Journal of Experimental Psychology: Human Perception and Performance.

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

Jaeger, D., Elbert, T., Lutzenberger, W., & Birbaumer, N. (in press). The effects of externally applied transcephalic weak direct currents on lateralization in choice reaction tasks. _Journal of Psychophysiology_.

Kornhuber, H. H., & Deecke, L. (1965). Hirnpotentialanderungen bei Wilkurbewegungen und passiven Bewegungen des Menschen: Bereitschaftpotential und reafferente Potentiale. _Pflugers Archives fur die gesammte Physiologie_, _248_, 1-17.

Kounios, J., Osman, A. M., & Meyer, D. E. (1987). Structure and process in semantic memory: New evidence based on speed-accuracy decomposition. _Journal of Experimental Psychology: General_, _116_, 3-25.

Kramer, A. F., & Donchin, E. (1987). Brain potentials as indices of orthographic and phonological interaction during word matching. _Journal of Experimental Psychology: Learning, Memory, and Cognition_, _13_, 76-86.

Kutas, M., & Donchin, E. (1974). Studies of squeezing: The effect of handedness, the responding hand and response force on the contralateral dominance of the readiness potential. _Science_, _186_, 545-548.

Kutas, M., & Donchin, E. (1980). Preparation to respond as manifested by movement-related brain potential. _Brain Research_, _202_, 95-115.

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. _Cognitive Psychology_, _12_, 523-553.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. Psychological Review, 91, 295-327.

Luce, R. D. (1986). Response times: Their role in inferring elementary mental organization. New York: Oxford.

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.

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., Osman, A. M., Irwin, D. E., & Yantis, S. G. (in press). Modern mental chronometry. Biological Psychology.

Meyer, D. E., Yantis, S. G., Osman, A. M., & Smith, J. K. (1986). Temporal properties of human information processing: Tests of discrete versus continuous models. Cognitive Psychology, 17, 445-518.

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.

Okada, Y. C., Williamson, S. J., & Kaufman, L. (1982). Magnetic fields of the human sensorimotor cortex. International Journal of Neurophysiology, 17, 33-38.

Penfield, W., & Jasper, H. (1954). Epilepsy and the functional anatomy of the human brain. Boston, MA: Little Brown.

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. I. (1978). Chronometric explorations of mind. Hillsdale, N.J.: Erlbaum.

Proctor, R. W., & Reeve, T. G. (1985). Compatibility effects in the assignment of symbolic stimuli to discrete finger responses. Journal of Experimental Psychology: Human Perception and Performance, 11, 623-639.

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

Requin, J. (1985). Looking forward to moving soon: Ante factum selective processes in motor control. In M.I. Posner & O.S.M. Marin (Eds.), Attention and Performance X (pp. 147-167). Hillsdale, NJ: Erlbaum.

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

Sanders, A. F. (1980). Stage analysis of reaction processes. In G. E. Stelmack & J. Requin (Eds.), Tutorials in motor behavior. Advances in Psychology (Vol. 1, pp. 331-354). Amsterdam: North Holland.

Smid, H. G. O. M., Mulder, G., & Mulder, L. J. M. (1988). The continuous flow model: A psychophysiological investigation of perceptual and motor aspects. In preparation.

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.

Vaughan, H. G., Costa, L. D., & Ritter, W. (1972). Topography of the human motor potential. Electroencephalography and Clinical Neurophysiology, 25, 1-10.

Figure Legends

Figure 1. Schematic representation of stimulus evaluation and response activation systems and the concept of "early communication." Each system may pass through intermediate states (levels of activation). The evaluation system may or may not pass information to the response activation system before stimulus evaluation is completed.

Figure 2. Lateralized readiness potentials time-locked to the motor response, or to the respond stimulus, for three different conditions. Voluntary: the subject produced self-paced squeezes. Warned: the subject responded to the respond stimulus (RS) in a choice reaction time when the warning stimulus (WS) provided information about the hand to be used. Choice Warned: the subject responded to the respond stimulus (RS) in a choice reaction time task when the warning stimulus (WS) did not provide information about the hand to be used. (Copyright 1980, Elsevier Science Publishers. Adapted with permission from the author and publisher from Kutas & Donchin, 1980).

Figure 3. The Conflict Paradigm. The subject receives conflicting information (information 1 and information 2 are associated with different responses). Preliminary evaluation (based on information 1) provides initial evidence for the incorrect response. Complete evaluation (based on information 2) provides evidence for the correct response. Correct and incorrect responses are assigned to different hands. Thus, a reversal in lateralization of the readiness potential will reveal the presence of early communication.

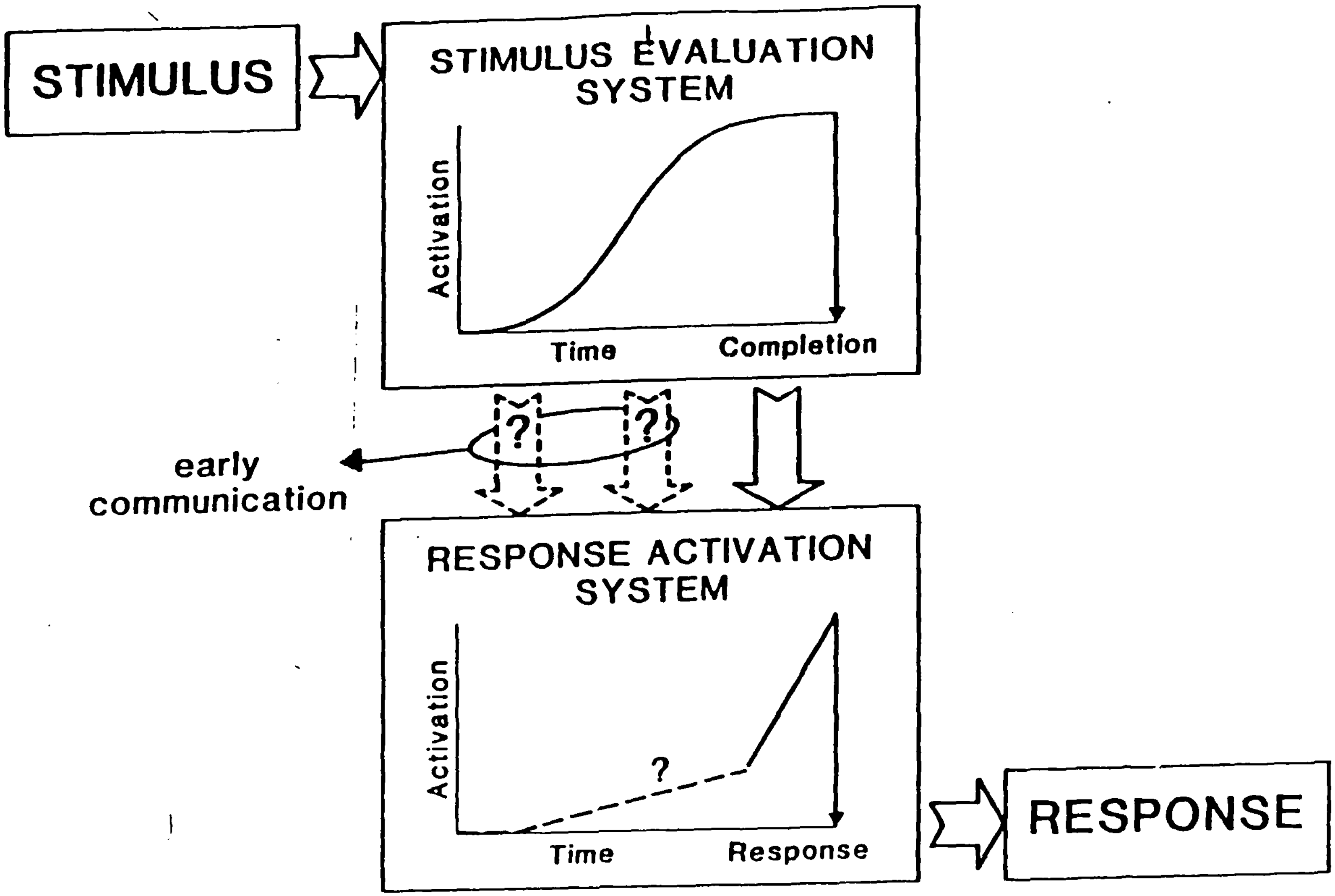
Figure 4. S1 indicates the time of presentation of the warning stimulus; S2 indicates when the imperative stimulus was presented. Upward deflections indicate greater negativity at the scalp site contralateral to the same response - and can be considered to reflect activation of the "same" response. _Panel A_ presents the lateralization data for the condition in which the precue predicted the imperative stimulus with a probability of .8. The solid line refers to trials in which the predicted imperative stimulus ("same") actually occurred, while the dashed lines refers to trials in which the unpredicted imperative stimulus ("different") was presented. _Panel B_ presents comparable data for the condition in which the precue did not have any predictive value. (From Bosco et al., 1986.)

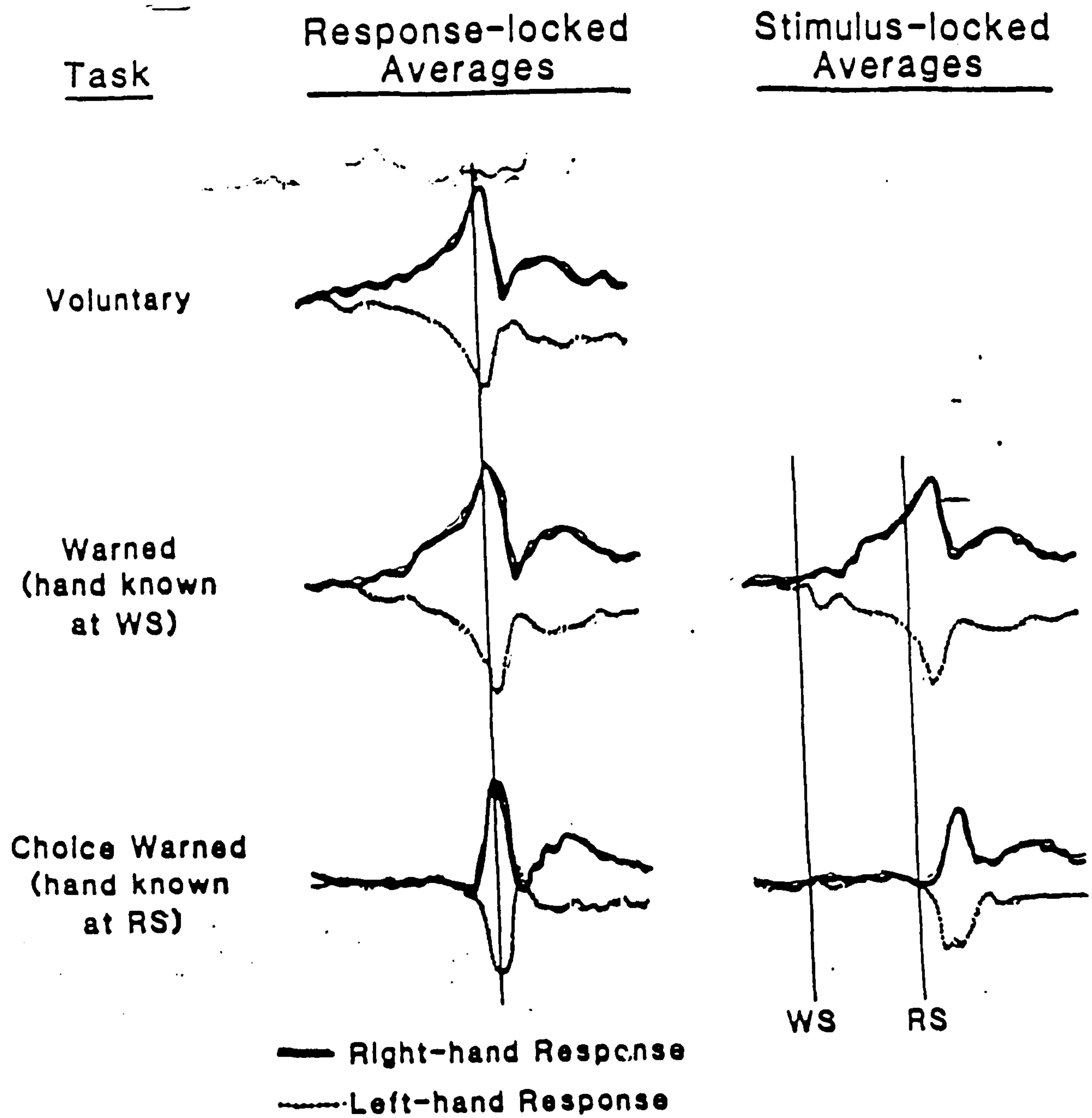
Figure 5. _Top panel_. EMG response accuracy is plotted as a function of EMG response latency for compatible and incompatible conditions. _Middle panel_. Lateralization values for the two compatibility conditions. Upward deflections are given by greater negativity at the electrode site contralateral to the correct response and thus indicate correct response activation. Downward deflections indicate incorrect response activation. _Bottom panel_. Lateralization values for a subset of trials on which (a) EMG latency was 300-349 msec, and (b) the first discernible EMG response was with the correct hand. (Copyright 1988, American Psychological Association. Reprinted with permission from the author and publisher from Gratton et al., in press).

Figure 6. Lateralization of the readiness potential for different conditions of the noise/compatibility experiment. In the No Noise condition, target letters were presented alone. In the Neutral Noise condition, noise letters were not associated with any experimentally defined response. For Compatible and Incompatible Noise conditions, noise letters were those that called for the same or the opposite response as the target. (Adapted with permission of the author from Smid et al., in preparation.)

Figure 7. Lateralization data averaged over "yes" and "no" responses, when conflict was or was not present for the rhyme judgment task. Upward deflections correspond to greater negativity at the scalp site contralateral to the correct response. Averages are based on all correct response trials regardless of response latency. There were insufficient data to segregate trials as a function of reaction time. *Top Panel*. Data from the Rhyme Task. *Bottom Panel*. Data from the Visual Task.

Figure 8. The Congruence Paradigm. The subject receives congruent information (information 1 and information 2 are associated with the same response). Preliminary evaluation (based on information 1) provides partial evidence for the correct response. Complete evaluation (based on information 2) provides complete evidence for the correct response. A "two-step" increase in lateralization of the readiness potential will reveal the presence of early communication.


STIMULUS
STIMULUS EVALUATION SYSTEM
Activation
Time
Completion
?
?
early communication
RESPONSE ACTIVATION SYSTEM
Activation
?
Time
Response
RESPONSE


Task
Response-locked Averages
Stimulus-locked Averages
Voluntary
Warned (hand known at WS)
Choice Warned (hand known at RS)
WS
RS
Right-hand Response
Left-hand Response

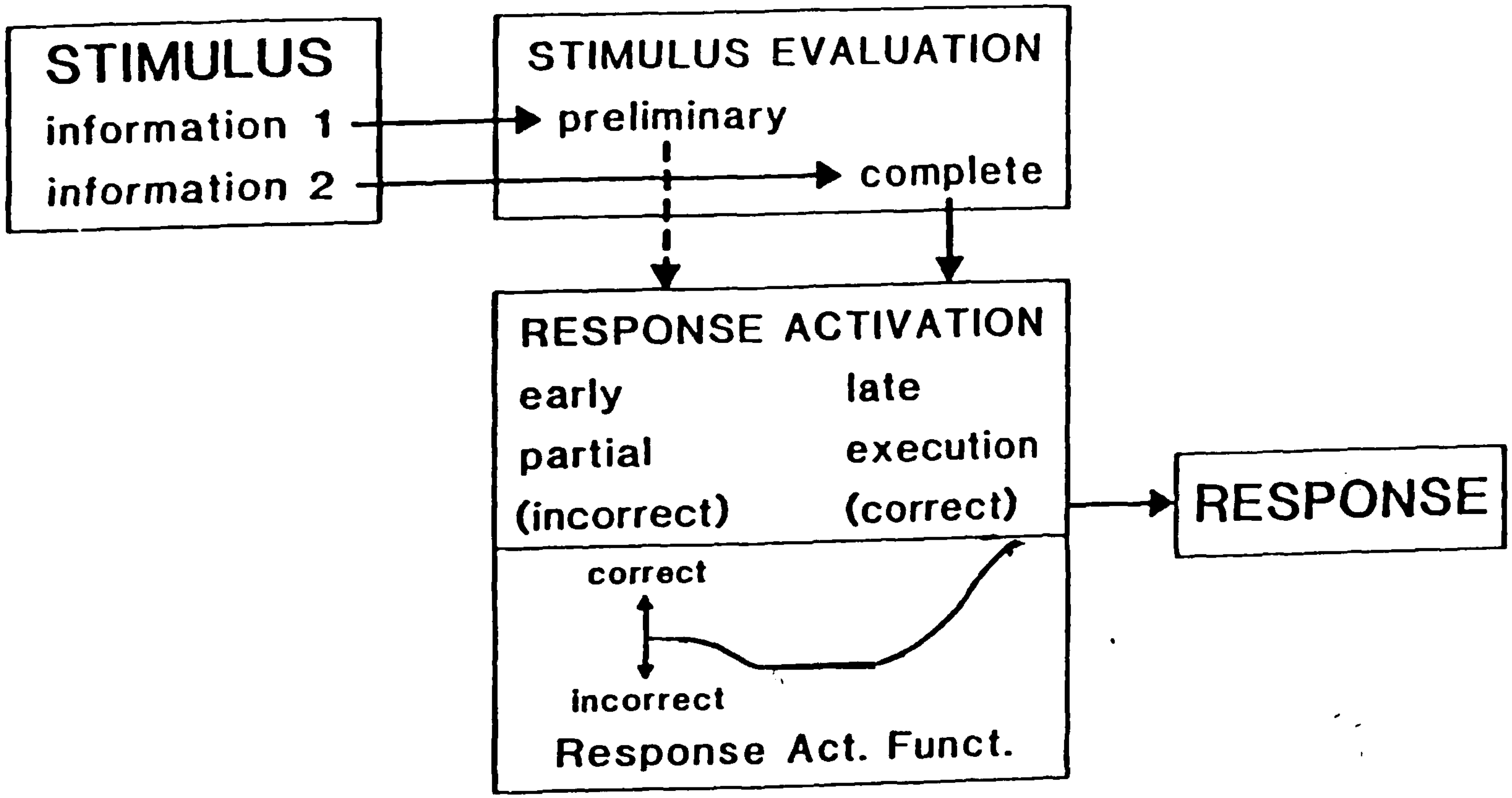
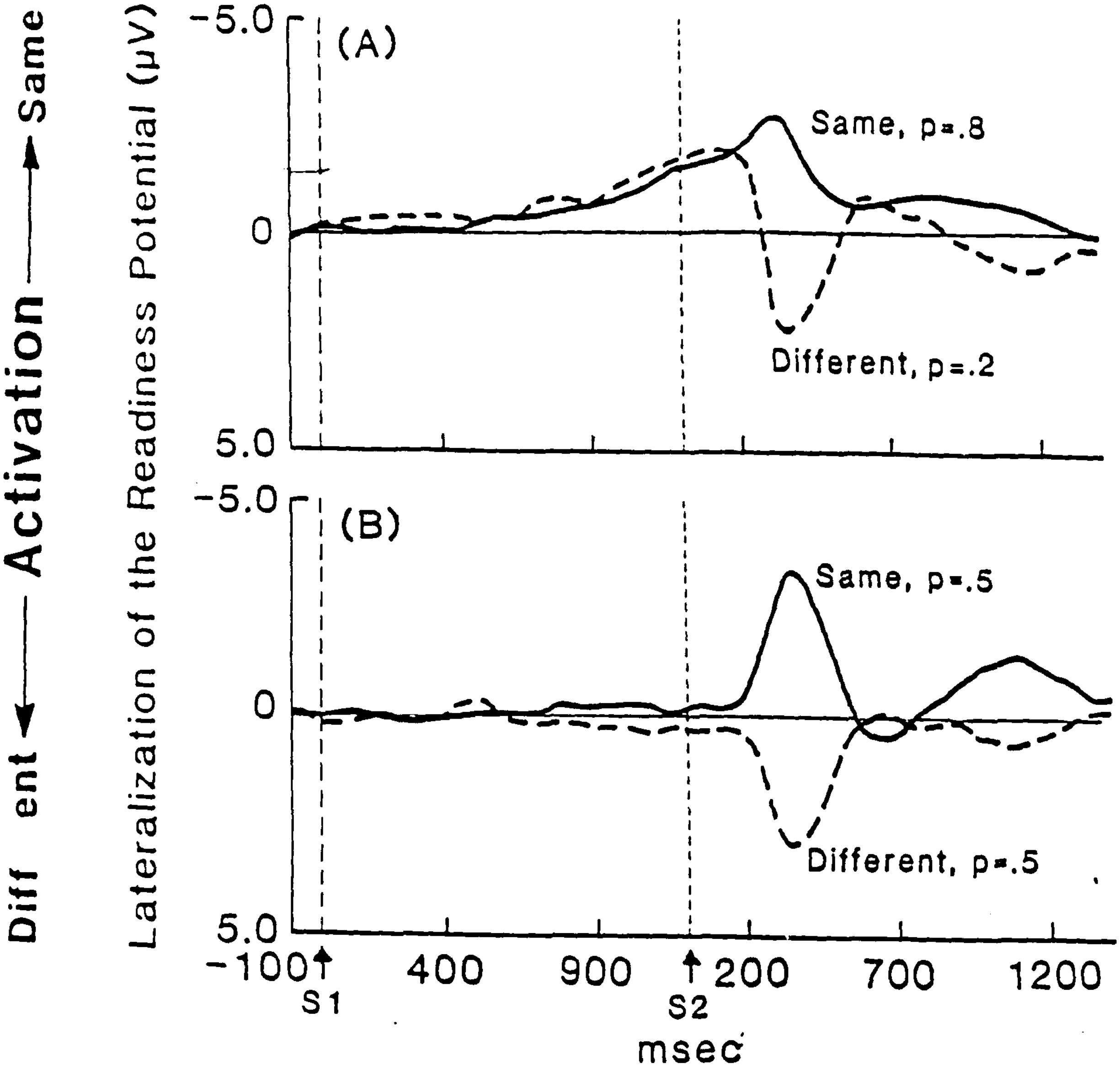

STIMULUS
information 1
information 2
STIMULUS EVALUATION
preliminary
complete
RESPONSE ACTIVATION
early
partial
(incorrect)
late
execution
(correct)
correct
incorrect
Response Act. Funct.
RESPONSE

Fig 4



Same: H → H

S → S

Different: H → S

S → H

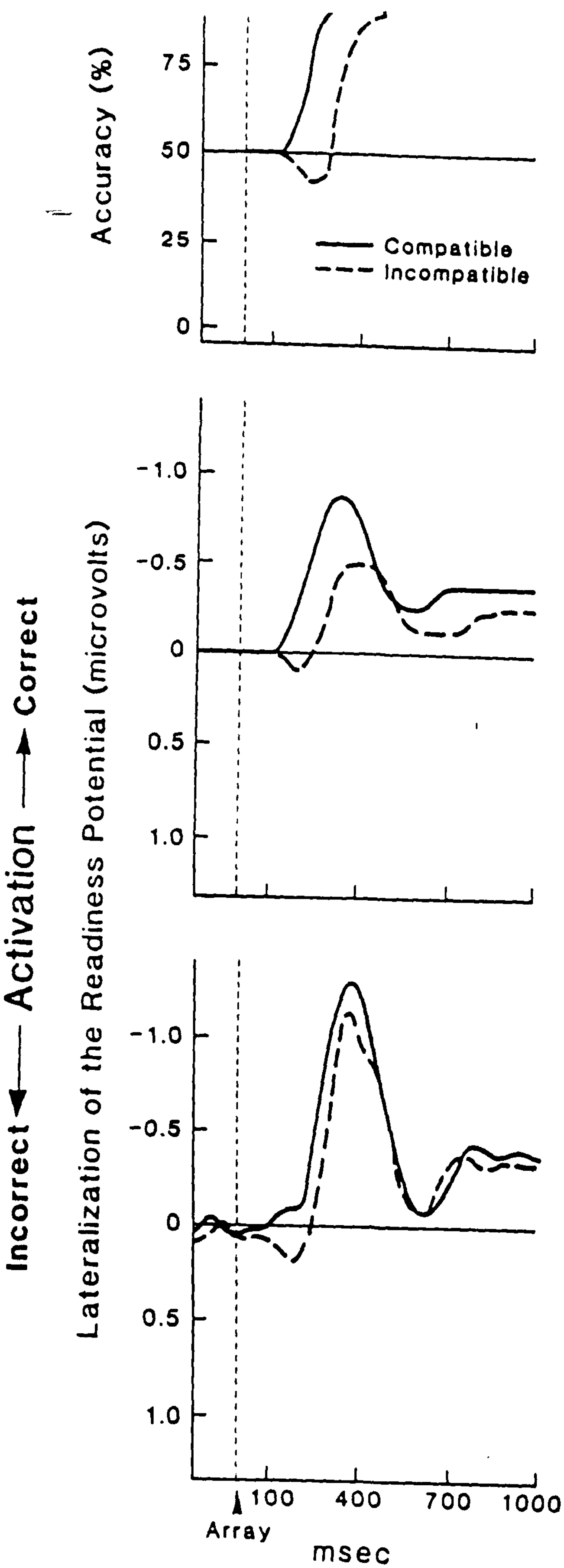
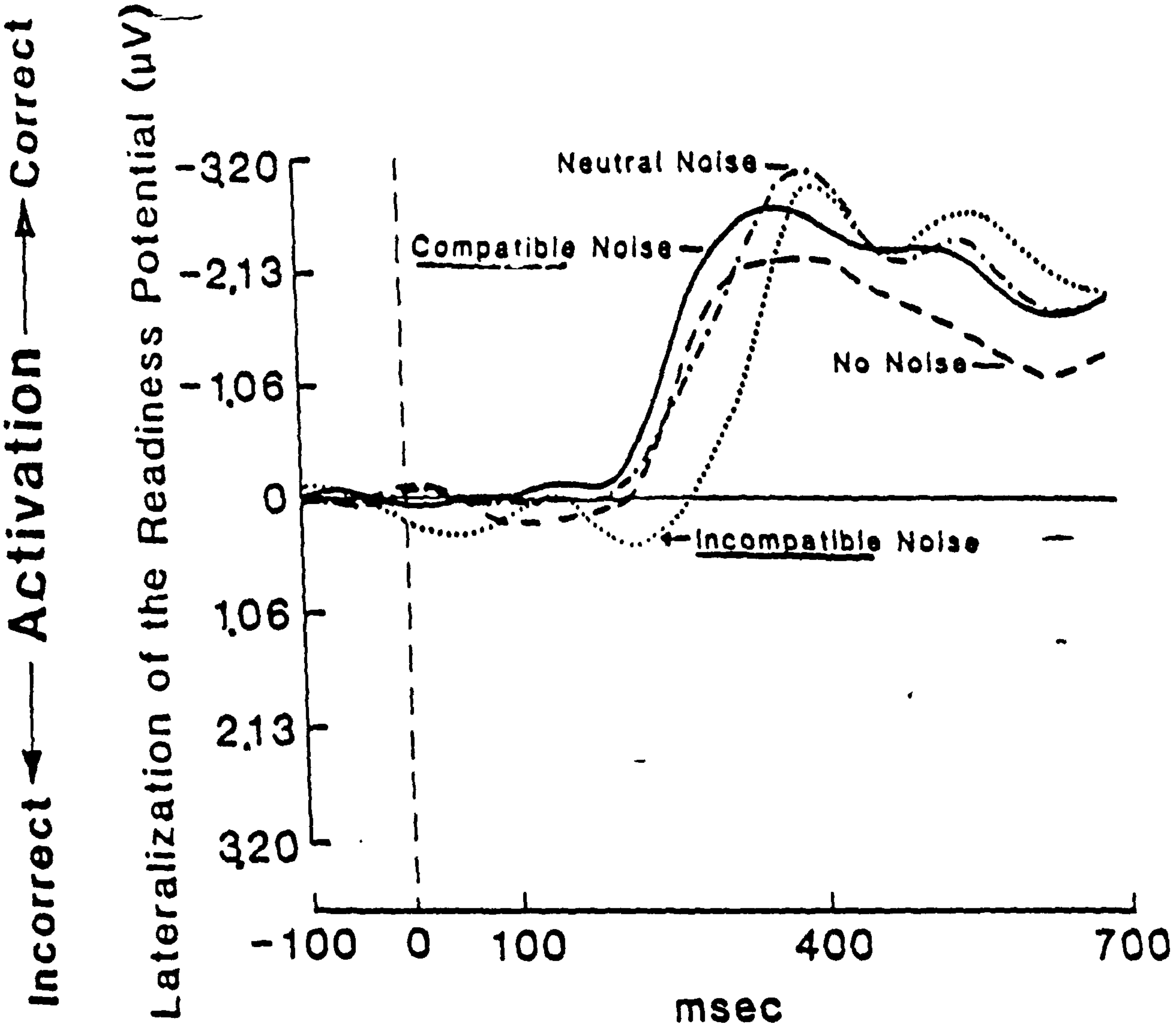
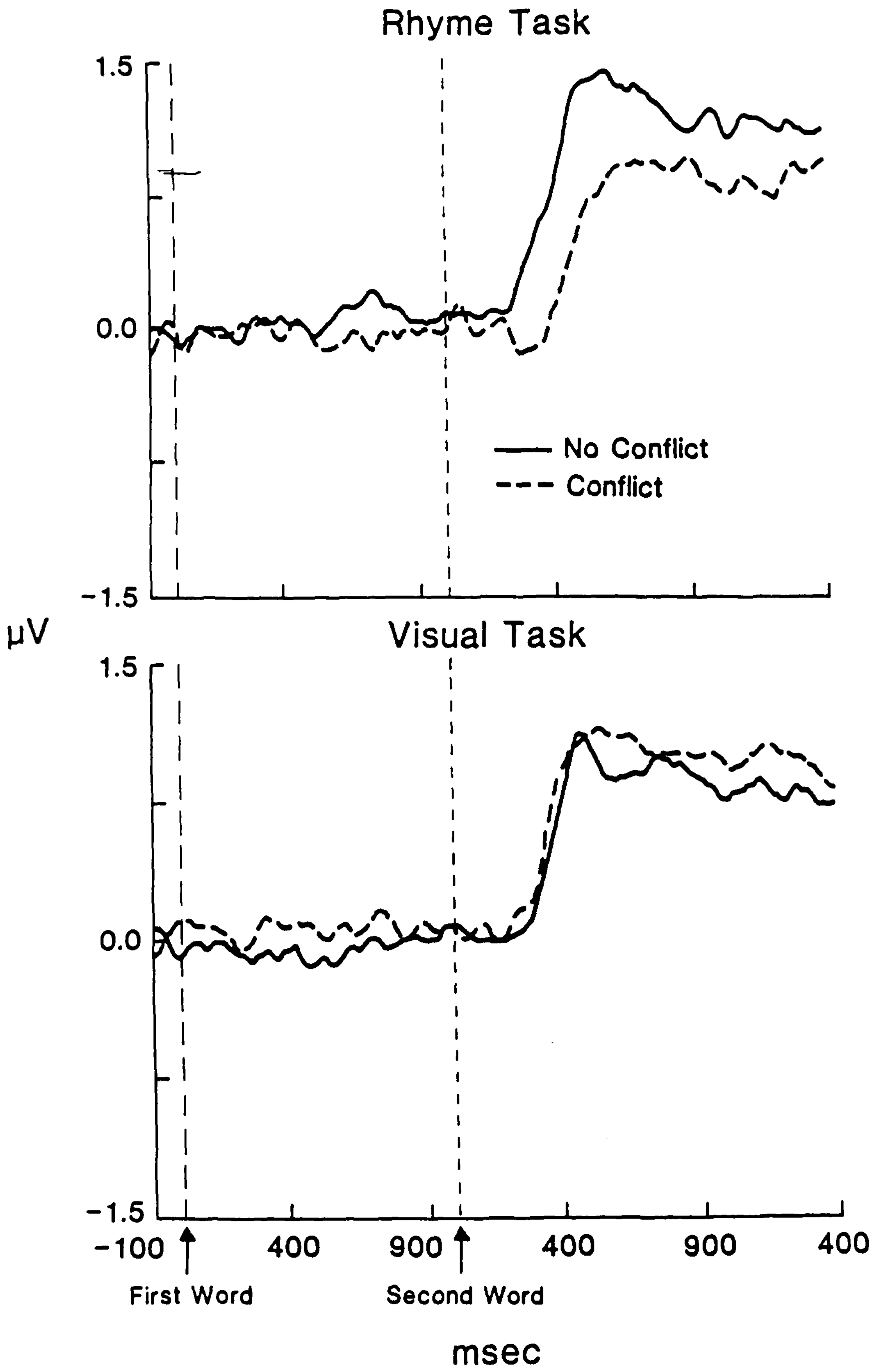
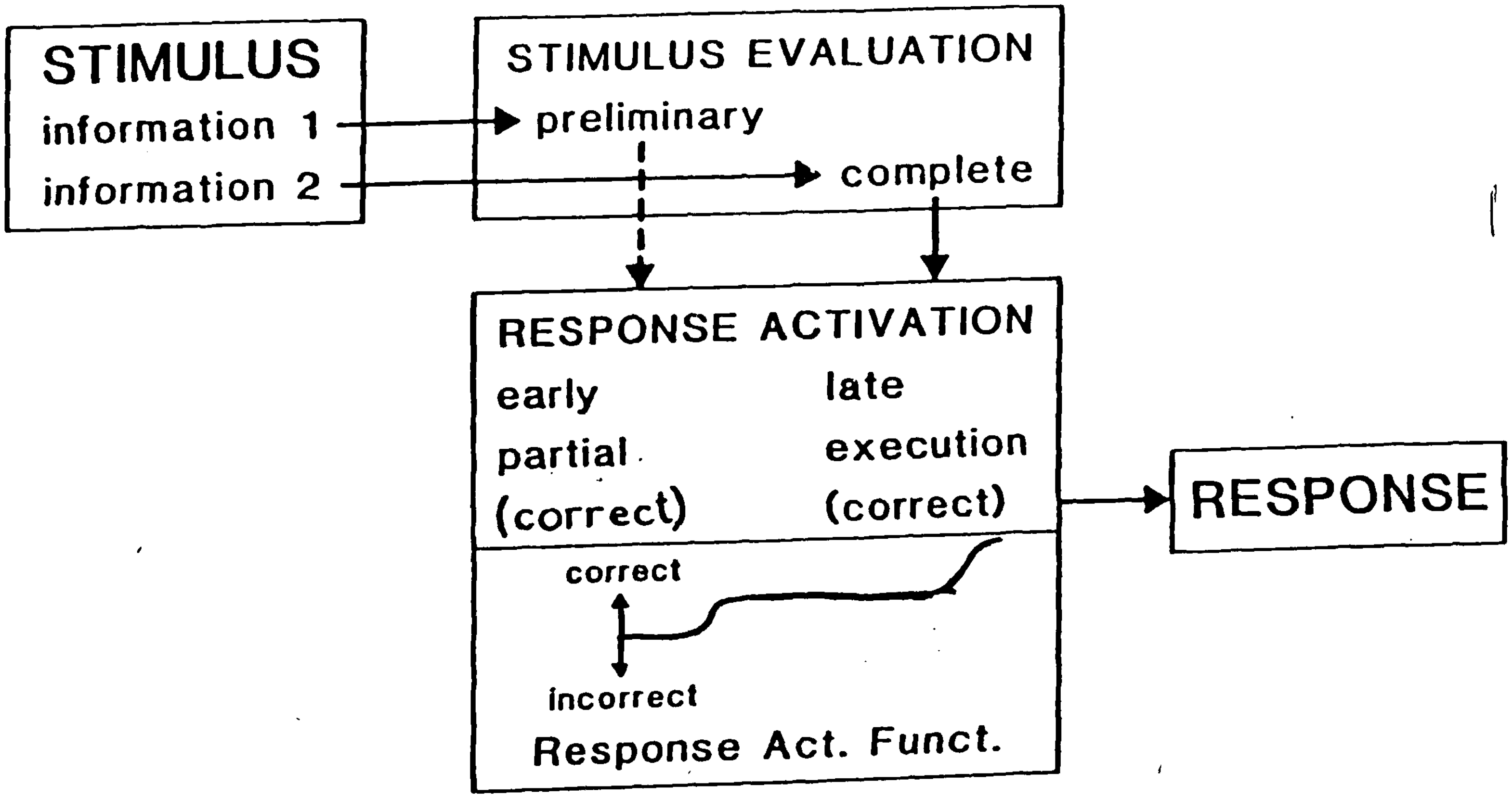

Accuracy (%)
75
50
25
0
Compatible
Incompatible
Incorrect ← Activation → Correct
Lateralization of the Readiness Potential (microvolts)
-1.0
-0.5
0
0.5
1.0
100
400
700
1000
Array
msec

Fig. 6


Incorrect ← Activation → Correct
Lateralization of the Readiness Potential (µV)
-3,20
-2,13
-1,06
0
1,06
2,13
3,20
Neutral Noise
Compatible Noise
No Noise
Incompatible Noise
-100
0
100
400
700
msec


Rhyme Task
1.5
0.0
-1.5
No Conflict
Conflict
μV
Visual Task
1.5
0.0
-1.5
-100
400
900
400
900
400
First Word
Second Word
msec


STIMULUS
information 1
information 2
STIMULUS EVALUATION
preliminary
complete
RESPONSE ACTIVATION
early
partial
(correct)
late
execution
(correct)
correct
incorrect
Response Act. Funct.
RESPONSE

Chapter 15

Event-Related Brain Potentials

Michael G. H. Coles, Gabriele Gratton, & Monica Fabiani

Cognitive Psychophysiology Laboratory
University of Illinois at Urbana-Champaign

Acknowledgements

Preparation of this chapter was supported in part by a grant from NIMH (MH41445). Bill Gehring, Greg Miller, John Cacioppo and Lou Tassinary, made helpful comments on preliminary drafts.

1. Introduction

Ever since Berger (1929) demonstrated that it is possible to record the electrical activity of the brain by placing electrodes on the surface of the scalp, there has been considerable interest in the relationship between these recordings and psychological processes. While Berger and his followers focussed their attention on spontaneous rhythmic oscillations in voltage - that is, on the electroencephalogram or EEG (see chapter 14) - more recent research has concentrated on those aspects of the electrical potential that are specifically time-locked to events - that is, on event-related potentials or ERPs. The ERPs are regarded as manifestations of brain activities that occur in preparation for or in response to discrete events, be they internal or external to the subject. Conceptually, the ERPs are regarded as manifestations of specific psychological processes. Later in this chapter, we shall review what is known about their underlying sources and their relationship to physiological function (Section 2). However, we shall focus for the most part on the relationship between the potentials and psychological function (Sections 3, 4, and 5).

1.1 Deriving event-related potentials

The procedures used to derive ERPs begin with the same amplifiers and filters used to obtain EEG (see Figure 1). Electrodes are attached to the scalp at various locations and connected to amplifiers. The locations are usually chosen according to the International 10-20 system (Jasper, 1958), such that between-laboratory and between-experiment comparisons are possible. The outputs of the amplifiers are converted to numbers by a device for measuring electrical potentials, an analog-digital converter. The potentials are sampled at a frequency ranging from 100 to 10000 Hz (cycles per second) and may be stored for subsequent analysis.

The ERP is small (a few microvolts) in comparison to the EEG (about 50 microvolts). Thus, the analysis generally begins with a procedure to increase the discrimination of the "signal" (the ERP) from the "noise" (background EEG). The most common procedure involves averaging samples of the EEG that are time-locked to repeated occurrences of a particular event. The number of samples used in the average will depend on the signal-to-noise ratio. However, in all cases the samples are selected so as to bear a constant temporal relationship to an event. Since all those aspects of the EEG that are not time-locked to the event are assumed to vary randomly from sample to sample, the averaging procedure should result in a reduction of these potentials leaving the event-related potentials visible.[1] The resulting

---

[1] Note that this assumption may not always be valid, as, for example, in the case of variability in the latency and other characteristics of the ERP from sample to sample (see 3.2.2). Furthermore, the ERP derived by averaging may include potentials that do not originate in the brain but are time-locked to the event (see 3.2.1).

voltage x time function (see Figure 2) contains a number of positive and negative peaks which are then subjected to a variety of measurement operations (see 3.2). These peaks are generally described in terms of their characteristic polarity and latency. Thus, P300 refers to a positive peak with a modal latency of 300 ms. Other descriptors can include reference to the psychological or experimental conditions that control the potential (e.g. readiness potential, mismatch negativity) and to the scalp location at which the potential is maximal (e.g. frontal P300). Note that each peak in the ERP waveform is usually associated with a particular distribution across the scalp. Thus, spatial (topographic) distribution is regarded as an important discriminative characteristic of the ERP (Donchin, 1978; Sutton & Ruchkin, 1984). The relationship between spatial distribution and underlying brain activity will be discussed in Section 2, while Section 3 provides a more thorough description of the methods used to extract and measure ERPs. Section 4 reviews some of the more commonly measured components.

1.2 The endogenous versus exogenous distinction

From a psychological point of view, it is convenient to distinguish between different types of ERPs. First we can identify those ERPs whose characteristics are mostly controlled by the physical properties of an external eliciting event. Such potentials are considered to be obligatory and are referred to as "sensory" or "exogenous." Second, we can identify ERPs whose characteristics are determined more by the nature of the interaction between the subject and the event. For example, some ERPs vary as a function of the information processing activities required of the subject; others can be elicited in the absence of an external eliciting event. These potentials are referred to as "endogenous". Naturally, it is the endogenous potentials that are the focus of those researchers interested in cognitive function. (For a discussion of the distinction between exogenous and endogenous potentials, see Donchin, Ritter, and McCallum, 1978).

2. Physiological Basis of ERPs

In this section, we review evidence that relates the scalp-recorded electrical activity to its underlying anatomical and physiological basis. For a detailed review of this relationship, see Nunez (1981), or Allison, Wood, and McCarthy (1986).

2.1 From the brain to the scalp: the generation of ERPs

It is generally assumed that ERPs are distant manifestations of the activity of populations of neurons within the brain. This activity can be recorded on the surface of the scalp because the tissue that lies between the source and the scalp acts as a volume-conductor. Since the electrical activity associated with any particular neuron is small, it is only possible to record at the scalp the integrated activity of a large number of neurons. Two requirements must be met for this integration to occur: (a) the neurons must be active synchronously, and (b) the electric fields generated by each particular neuron must be oriented in such a way that their effects at the scalp cumulate. As a consequence, only a subset of the entire brain electrical activity can be recorded from scalp electrodes.

Two considerations further restrict the likely sources of the ERP. First, since the ERP represents the synchronous activity of a large number of neurons, it is probably not due to the summation of pre-synaptic potentials (spikes), because these potentials have a very high frequency and short duration. On the other hand, post-synaptic potentials, having a relatively slower time course, are more likely to be synchronous, and therefore to summate to produce scalp potentials. Thus, it is commonly believed that most scalp ERPs are the summation of the post-synaptic potentials of a large number of neurons that are activated (or inhibited) synchronously (see Allison, Wood, & McCarthy, 1986).

A second consideration concerns the orientation of neuronal fields. Since the electric fields associated with the activity of each individual neuron involved must be oriented in such a way as to cumulate at the scalp, only neural structures with a specific spatial organization may generate scalp ERPs. Lorente de No (1947) specified the spatial organizations that are required for the distant recording of the electrical activity of a neural structure. He distinguished between two types, "open fields" and "closed fields" (see Figure 3).

A structure having an open field organization is characterized by neurons that are ordered so that their dendritic trees are all oriented on one side of the structures, while their axons all depart from the other side. In this case, the electric fields generated by the activity of each neuron will all be oriented in the same direction and summate. Only structures with some degree of open field organization generate potentials that can be recorded at the scalp. Open fields are obtained whenever neurons are organized in layers, as in most of the cortex, parts of the thalamus, the cerebellum, and other structures.

A structure with a closed field organization is characterized by neurons that are concentrically or randomly organized. In both cases, the electric fields generated by each neuron will be oriented in very different, sometimes opposite, directions, and therefore will cancel each other. Examples of closed field organization are given by some midbrain nuclei.

From this analysis it is clear that ERPs represent just a sample of the brain electrical activity associated with a certain event. Thus, it is entirely possible that a sizeable portion of the information processing transactions that occur after (or before) the anchor event are silent as far as ERPs are concerned. For this reason, some caution should be used in the interpretation of ERP data. For instance, if an experimental manipulation has no effect on the ERP, we cannot conclude that it does not influence brain processes.

2.2 From the scalp to the brain: inferring the sources of ERPs

So far we have examined how particular properties of neuronal phenomena may determine whether they will be recorded at the scalp. We have approached the problem of ERP generation in a direct fashion, from properties of the generators to predictable scalp observations. In most cases, however, we have only limited information about the neural structure(s) responsible for a specific aspect of the ERP. Our database consists of observations of voltage differences between scalp electrodes or between scalp electrodes and a reference electrode. To determine which neural structures are responsible for the scalp potential - that is, to identify the neural generators of ERPs - we must solve the inverse problem. We have to infer the unique combination of neural generators whose activity results in the potential observed at the scalp.

In solving this problem, we are confronted with an indefinite number of unknown parameters. In fact, an indefinite number of neural generators may be active simultaneously, and each of them may vary in amplitude, orientation of the electric field, and location inside the head. Since a limited number of observations (the voltage values recorded at different scalp electrodes) is used to estimate an indefinite number of parameters, it is clear that the inverse problem does not have a unique solution. A further complication is that the head is not a homogeneous medium. Therefore the electric field generated by the activity of particular structures is difficult to compute. A particularly important distortion of the electric fields is caused by the skull - a very low conductance medium that reduces and smears electric fields. For all these reasons, we cannot determine unequivocally which structures are responsible for the ERP observed at any point in time, when the only information available is given by the potentials recorded at scalp electrodes.

In spite of these problems, investigators have tried to identify the neural sources of the scalp ERP using a variety of approaches, involving both non-invasive and invasive techniques. Non-invasive techniques are based on scalp recordings. They involve complex mathematical procedures and depend on a number of restrictive assumptions. Invasive techniques are based on recordings from indwelling electrodes (in humans or in animals) or on lesion data.

A non-invasive approach to the inverse problem is to generate several alternative hypotheses about neural structures that may be active at a particular point in time, and that may be responsible for an observed scalp ERP. The distribution of potentials across the scalp that would be generated by each of these structures can then be computed using a direct approach. Finally, the structure whose activity best accounts for the observed scalp distribution can be identified. This approach requires that specific neurophysiological knowledge exists about candidate underlying structures. First, we must have some reason to restrict the number of candidate neural generators to a manageable number. Second, we must have sufficient knowledge of the anatomy and physiology of each of these candidate neural generators to be able to compute the distributions of the scalp potentials associated with their activity. At present, such knowledge exists only in a very limited number of cases (see, for example, Scherg & von Cramon, 1985).

A more sophisticated and promising non-invasive procedure involves the use of magnetic field recordings. Magnetic fields generated by brain activity are extremely small in relation to magnetic fields generated by environmental and other bodily sources. Therefore, their measurement is both difficult and expensive. The advantage of measuring the magnetic field is that it is practically insensitive to variations of the conductive media (such as those due to the presence of the skull). It is therefore much easier to compute the source of a particular field. An in-depth discussion of the problems and peculiarities of the magnetic technique is beyond the scope of this chapter and can be found elsewhere (Beatty, Barth, Richer, & Johnson, 1986). We will only note here that using the magnetic technique to determine the source of neural components still requires assumptions about the number of neural structures active at a particular moment in time.

Invasive techniques for the identification of the sources of ERP components are based on the implantation of electrodes within the brain of humans or animals. Research on humans has been made possible by the need for recording EEG activity in deep regions of the brain for diagnostic purposes (Halgren et al., 1980; Wood et al., 1984). A problem with the human research is that the indwelling electrodes are located according to clinical, rather than scientific criteria, and therefore may fail to map the regions involved in the generation of scalp ERPs. This may be solved by research on animals (Buchwald & Squires, 1983; Csepe, Karmos, & Molnar, 1987; Starr & Farley, 1983). However, animal research is problematic because it is difficult to determine whether the ERP observed in animals corresponds to that observed in humans. This is because of fundamental differences in the anatomy of animal and human brains. Finally, a general problem with depth recording is that it is difficult to know the extent to which the scalp recorded ERP is due to the activity of the structures that have been identified by the indwelling electrodes. This problem can be resolved in animal research if lesions in the structure identified as the candidate generator result in elimination of the scalp potential.

In summary, although several techniques have been used for identifying the source of ERP components, none of them appears to be likely to give definitive answers in all cases. However, the convergence of several techniques may provide useful information about the neural structures whose activity is manifested at the scalp by the ERP.

3. The Inferential Context

In this section, we review the process through which we come to make inferences about psychological processes and states from the measurement of ERPs. To begin with, however, we need to address a number of assumptions about the "meaning" of the ERP and a variety of measurement issues.

3.1 The concept of components

As we noted above, the ERP can be described as a voltage x time function. We assume that the various voltage fluctuations represented by this function reflect the activities of neuronal populations and that, in turn, these neuronal populations are responsible for the execution of some psychological process. In practice, the tendency in cognitive psychophysiology has been to focus on processes identified by cognitive psychologists as candidate psychological processes.

The total ERP is assumed to be a manifestation of the aggregate of a number of ERP "components". The components can be defined in three different ways (Fabiani, Gratton, Karis, & Donchin, in press; Naatanen & Picton, 1987). First, components can be defined in terms of the peaks and troughs (maxima and minima) that are observed in the ERP trace. Second, components can be defined as aspects of the ERP waveforms that are functionally associated, that is, they covary across subjects or conditions or location on the scalp. Third, components can be defined in terms of those neural structures that generate them. These definitions may converge in some circumstances. However, as Naatanen and Picton (1987) have indicated, a peak in the ERP waveform (for example, the N1) may represent the summation of several functionally and structurally distinct components. Thus, the adoption of one or another of these definitions will have important consequences for the interpretation of the component structure of the ERP waveform. A corollary of this is that different measurement procedures will be required depending on the type of component definition that is adopted. These procedures will be reviewed in subsequent sections after a brief discussion of general measurement issues.

3.2 General measurement issues

3.2.1 Artifacts. The potential recorded at the scalp can be influenced by sources of electrical activity that do not arise from the brain. Examples of these sources of artifacts include the movement of eyeballs and eyelids, tension of the muscles in the head and neck, and the electrical activity generated by the heart. These artifacts can be dealt with in the following ways. First, one can set up the recording situation so that artifacts are minimized. This can be accomplished by suitable choice of electrode locations and of the subject's task and environment. Second, one can simply discard records which contain artifacts. Unfortunately, this procedure may lead to a bias in the selection of the observations and/or subjects. Third, one can use filters (see below) to attenuate artifactual activity. This procedure is useful when the frequency of the artifactual activity is outside the frequency range of the ERP signal of interest. For example, the frequency of electromyographic activity is higher than that of most endogenous ERP components. Fourth, one can attempt to measure the extent of the artifact and then remove it from the data. This procedure has been used most frequently in the case of ocular artifacts (e.g. Gratton, Coles, & Donchin, 1983).

3.2.2 Signal-to-noise ratio. The ERP consists of a series of fluctuations in voltage that are time-locked to an event. These voltage changes are typically small (a few microvolts) in relation to the background EEG (about 50 microvolts) in which they are imbedded. Thus, a major measurement problem concerns the extraction of the ERP signal from the background noise.

Several procedures have been advocated to increase the signal-to-noise ratio, including filtering, averaging, and pattern recognition (see Coles, Gratton, Kramer, & Miller, 1986, for a more detailed discussion).

Filtering involves the attenuation of noise, whose frequency is different from that of the signal. For example, most endogenous components have frequencies of between 0.5 and 20 Hz. Thus, at the time of recording or later at the time of analysis, analog or digital filters can be used to attenuate activity outside this frequency range. Great care should be taken in the selection of filters. The amplitude and latency of an ERP component (as well as the general ERP waveform) will be distorted if the bandpass of the filter excludes frequencies of interest (see Figure 4).

Averaging involves the summation of a series of EEG epochs (or trials), each of which is time-locked to the event of interest. These EEG epochs are assumed to be given by two sources: first, the ERP, and second, other voltage fluctuations that are not time-locked to the event. Since, by definition, these other fluctuations are random with respect to the event, they should average to zero, leaving the time-locked ERP both visible and measurable. If it is the case that (a) the ERP "signals" are constant over trials, (b) the noise is random across trials, and (c) that the ERP "signals" are independent of the background noise, then the signal-to-noise ratio will be increased by the square root of the number of trials included in the average. Note that the utility of the averaging procedure also depends on the fact there are no correlated signals (such as the EOG - see 3.2.1) that are also time-locked to the event.

One of the problems with the averaging procedure is that the three assumptions described in the previous paragraph may not always be satisfied in the typical experiment. In particular, if the latency of the ERP varies from trial to trial (latency jitter), the average ERP waveform will not be representative of the actual ERP of any individual trial. A related problem is that the investigator may be interested in measures of the ERP on individual trials. Thus, a major thrust in ERP methodology has been to derive procedures for single-trial analysis.

Pattern recognition techniques allow the investigator to identify segments of the EEG epoch that contain specific features, such as a particular pattern of peak and troughs characteristic of an ERP component. The advantage of these techniques is that they allow the investigator to identify and measure components on individual trials. Examples of pattern recognition techniques are cross-correlation, Woody filter (Woody, 1967), and step-wise discriminant analysis (Donchin & Herning, 1975; Horst & Donchin, 1980; Squires & Donchin, 1976). In the case of cross-correlation and Woody filter, the individual trial epoch is scanned to determine the segment that best resembles an "ideal" template corresponding to the component of interest. In the case of discriminant function, the procedure begins with the selection of two sets of waveforms, that are presumed to differ in terms of a specific component. Then, features are identified (in the form of numerical weights) that best discriminate between the two sets of waveforms. These features are considered to represent the defining characteristic of the component and individual trials can be examined to determine the extent to which the features are present. Note that the cross-correlation and Woody filter procedures can yield both amplitude and latency estimates for individual trials, while discriminant function procedures yield only amplitude estimates. (For a general discussion of pattern recognition techniques, see Glaser & Ruchkin, 1976.)

3.3 Component quantification

In this section, we describe procedures that have been used to quantify ERP components. As mentioned earlier, these measurement operations will depend on the way in which ERP components are defined.

3.3.1 Peak measurement. As indicated above, components can be defined in terms of peaks or troughs having characteristic polarities and latency ranges. Thus, the measurement operation involves the assessment of either amplitude of the peak in microvolts, or the assessment of its latency in msec (see Figure 5). Amplitude is usually referred either to the baseline, pre-event, voltage level (base-to-peak amplitude) or to some other peak in the ERP waveform (peak-to-peak amplitude). Latency is referred to the time of occurrence of the event.

When the component under analysis does not have a definite peak, it is customary to measure the integrated activity across a particular latency range (area measure).

3.3.2 Covariation measures. As mentioned above, components can also be defined in terms of segments of the ERP waveform that exhibit covariation across subjects, conditions, or scalp locations. As a consequence, procedures are needed to identify and measure these segments.

One the most commonly used procedures is Principal Component Analysis (PCA). This procedure can be applied to describe the structure of the variance associated with a set of ERP waveforms. This structure is inferred from the pattern of covariance between the voltage values for each timepoint in the waveform. The outcome of PCA is a set of components each characterized by a particular pattern of weights (loadings, or eigenvectors) for each timepoint (see Figure 6). Then, the weights for each component can be used as a linear filter to derive measures of amplitude of each component in a particular waveform. Each point in the waveform is multiplied by the weight for that point. The resulting values are summed to yield factor scores (or measures of component amplitude). A more detailed description of the application of PCA to ERPs can be found in Donchin and Heffley (1978).

3.3.3 Source activity measures. A third way of defining components is in terms of underlying sources. According to this definition, we should quantify the activity of these sources to provide latency and amplitude measures of the different components. As we noted earlier (Section 2), the relationship between scalp electrical activity and source activity is difficult, if not impossible, to describe. Thus, for the time-being, this type of component measurement, although theoretically important, is practically unfeasible.

3.3.4 Problems in component measurement. In this section we discuss two specific problems that arise during component measurement.

The first problem concerns the commensurability of the measurements of different waveforms. Is a particular component, recorded under a particular set of circumstances, the same as that recorded in another situation? This is a particular problem when we define components as a peak observed at a particular latency. For example, if the latency of the peak differs between two experimental conditions, we would be led to conclude that different components are present in the two sets of data. How can we be sure that the same component varies in latency between the two conditions, rather than that two different components are present in the two different conditions? The problem is not resolved by using PCA. In this case, how can we know whether two components extracted by separate PCAs conducted on different sets of data reflect the same activity? A solution to this problem can only be derived from a careful examination of the pattern of results obtained, and from a comparison of these results with what we already know about different ERP components. Of course, this means that we are including a large number of empirical and theoretical arguments in the definition of each ERP component, which, of course, may differ from one component to another, and from time to time. Thus, the definition of a component may include not only polarity and latency, but also distribution across the scalp and sensitivity to experimental manipulations (see, for example, Fabiani et al., in press). Thus, it is clear that a correct interpretation of ERP component structure requires some background information about the components themselves (see section 4).

A second problem in component measurement is that of component overlap. Usually, ERP components do not appear in isolation, but several of them may be active at the same moment in time. This reflects the parallel nature of brain processes. When this occurs, it is difficult to attribute a particular portion of scalp activity to a particular component. Peak and area measures are particularly susceptible to this problem, but even PCA can, in some cases, misallocate variance across different components (see Wood & McCarthy, 1984). As a result, we may attribute a difference obtained between two particular experimental conditions to the wrong component.

Several procedures have been proposed to solve the problem of component overlap, but none of them seems to have universal validity. In some cases, it can be assumed that only one component varies between two experimental conditions. In this case, the variation of this component can be isolated by subtracting two sets of waveforms. Unfortunately, very rarely can we assume that the effect of an experimental variable is so selective. Furthermore, the subtraction procedure implies that only amplitude, and not latency, varies across experimental conditions.

A procedure that may help solve both these problems is vector filtering (Gratton, Coles, & Donchin, submitted). This procedure begins with the assumption that components can be defined in terms of scalp distribution. Any component can be represented by a specific profile of amplitude values at different electrode locations. This profile is then used as a "distributional filter" to determine the extent to which the component is present in a particular epoch (see Figure 7). This procedure can be applied to several components simultaneously, and, thus, it is possible to separate the contribution of several, overlapping, components to an observed waveform. Note that, in contrast to PCA, neither the latency nor the time-course of the component (but only the scalp distribution) need be constant over trials. The assumption of constancy of distribution may be more valid than assumptions of constancy of latency and time-course.

Note that distributional filters perform the same kind of operations in the spatial domain that frequency filters perform in the frequency domain. While frequency filters apply different weights to activity in different frequency bands, distributional filters apply different weights to activity from different spatial locations.

3.4 Experimental Logic

In this section, we review the experimental procedures that lead to a specification of the functional significance of the components. Some theory about the functional significance of a component is essential to the understanding of the changes that this component will exhibit as a function of specific contexts. The development of a theory about the functional significance of a component is a complex process that involves studies of the component's antecedents and consequences, as well as speculations about the psychological function it manifests.

To illustrate this process, we shall focus on the approach adopted in our laboratory (see Coles, Gratton, & Gehring, 1987; Donchin, 1981; Donchin & Coles, in press; Donchin, Karis, Bashore, Coles, & Gratton, 1986). While this approach is oriented towards cognitive ERP research, and to the P300 component in particular, there is no reason, in principle, why it could not be applied to other approaches to ERPs and to other components.

3.4.1 Antecedent conditions. In the case of all ERP components, the initial phase in the process begins with the "discovery" of a component. For most endogenous ERP components, this discovery occurred during the 1960s and 1970s; for the P300, this occurred in 1965 (Sutton, Braren, Zubin, & John, 1965). After this report, there followed a period in which attempts were made to map out the antecedent conditions that influence the amplitude and latency of the component.

For amplitude, a large series of studies focussed on the "oddball" paradigm in which subjects are presented with two stimuli, or classes of stimuli, that occur in a Bernouilli sequence. The probability of one stimulus is generally less than that for the other and the subject's task is to count the rarer of the two stimuli. The basic conclusion of these kinds of studies is that the amplitude of the P300 is sensitive to the probability of task relevant events (see, for example, Duncan-Johnson & Donchin, 1977, and Figure 8; see also Donchin et al., 1986a). Further research indicated that it is subjective, rather than objective, probability that controls the amplitude of P300 (Squires, Wickens, Squires, & Donchin, 1976). In addition, the stimuli or stimulus classes can be as diverse as male versus female names (Kutas, McCarthy, & Donchin, 1977) or pictures of politicians versus pictures of others (Towle, Heuer, & Donchin, 1980), or even the absence of the stimulus (Sutton, Tueting, Zubin, & John, 1967). Furthermore, the scalp distribution is independent of the modality of the stimulus (Simson, Vaughan, & Ritter, 1976).

A second factor controlling the amplitude of P300 is the task-relevance of the eliciting event. Thus, P300s are only elicited if the subject must use the stimuli to perform the assigned task. If the events occur while the subject is performing another task (such as a word puzzle) then even rare events do not elicit the P300 (see Figure 8: Duncan-Johnson & Donchin, 1977). Furthermore, the P300 to an event is directly related to the event's utility in terms of the subject's task (Johnson & Donchin, 1978; Bosco et al., 1986).

A final series of studies (see Donchin, Kramer, & Wickens, 1986) has demonstrated that the P300 is related to the processing resources demanded by a particular task. In a dual-task situation, P300 amplitude to primary-task events increases with the perceptual/cognitive resource demands, while the P300 response to the concurrent secondary-task decreases (Sirevaag, Kramer, Coles, & Donchin, 1984).

As far as P300 latency is concerned, the research has focussed on the identification of those processes that have elapsed prior to the elicitation of P300. As an initial observation, it can be noted that, if P300 amplitude is sensitive to probability, then, at the very least, processes required to establish the rareness of an event must occur prior to the P300 process. On the basis of this observation, Donchin (1979) proposed that P300 latency may reflect stimulus evaluation or categorization time. This idea was supported by the observation that the correlation between P300 latency and reaction time is higher when subjects are given accuracy rather than speed instructions. Furthermore, as categorization becomes more difficult, so P300 latency becomes longer (see Figure 9: Kutas, McCarthy, & Donchin, 1977).

If P300 latency reflects stimulus evaluation time, then it should not be affected by factors that influence response-related processes. Several studies (McCarthy & Donchin, 1981; Magliero, Bashore, Coles, & Donchin, 1984; Ragot, 1984) demonstrate that manipulations that should affect the duration of response-related processes (i.e. stimulus-response compatibility) have little or no effect on P300 latency, while manipulations of stimulus complexity have a large effect. (Both these manipulations have a substantial effect on reaction time.) Thus, it appears that the P300 is more dependent on the completion of processes of stimulus evaluation and categorization than on those related to the current overt response.

3.4.2 Formation of theory. Taken together, these data suggest that the P300 is dependent on (a) the subjective probability of task relevant effects, (b) the value or meaning of the event in the context of the task, and (c) the psychological resources allocated to the processing of the event. Furthermore, the P300 is not emitted until the event has been categorized. If responses are based on incomplete stimulus evaluation, the P300 will occur after the overt response. (See also Johnson, 1986, for a discussion of the determinants of P300, and Donchin et al., 1984, for a discussion of the relationship between P300 and the orienting reflex.)

With these observations in mind, one can proceed to speculate about the functional significance of the P300. Thus, Donchin (1981; Donchin & Coles, in press) argued that it is a manifestation of a process related to the updating of models of the environment or context in working memory. Such an updating will depend on the processing of the current event (hence, the P300 latency/stimulus evaluation time relationship) but will have implications for the processing of and response to future events.[2]

3.4.3 Consequences. These kinds of statements relating to functional significance can be tested by an examination of the predicted consequences of variation in the latency or amplitude of the P300 for the outcome of the interaction between the subject and the environment.

For example, as far as P300 latency is concerned, if the P300 occurs after the stimulus has been evaluated, then the quality of the subject's response to that event should depend on the timing of that response relative to the occurrence of the P300. Thus, Coles, Gratton, Bashore, Eriksen, and Donchin (1985) showed that, for a given response latency, response accuracy was

---

[2] Note that other theories of P300 have been offered by, for example, Desmedt (1981), Rosler (1983), and Verleger (in press). Both Desmedt and Verleger propose that the P300 is related to the termination or "closure" of processing periods, while Rosler proposed that P300 reflects controlled processing. As we have argued elsewhere (Donchin & Coles, in press), these theories do not account for the richness of the results concerning the consequences of the P300 that have accrued in the last decade.

higher the shorter the P300 latency. This is illustrated in Figure 10.

For P300 amplitude, we expect that, to the extent that the subject's future behavior depends on the degree to which an event leads to a change in their model of the environment, that behavior will be related to P300 amplitude. Thus, there have been several studies that have demonstrated a relationship between the memorability of an event, assessed at some future time, and the amplitude of the P300 response to the event at the time of initial presentation (see section 4.4.1 for a detailed discussion).

As another example, we have demonstrated that the subject's future strategy as revealed in overt behavior can be predicted from the P300 response to current events (Donchin, Gratton, Dupree, & Coles, in press). In particular, in a speeded choice reaction time task, the amplitude of the P300 following an error was related to the latency and accuracy of overt responses on subsequent trials.

These kinds of experiments that investigate the predicted consequences of variation in the P300 response for overt behavior provide tests of the theories of the functional significance of the P300. As a result, the theories are refined or revised, until, eventually, we are confident enough to proceed to use the ERP measure to monitor information processing activities (see, for example, Coles et al., 1985; Kramer, Sirevaag, & Braune, 1987).

3.5 The role of neurophysiology

As we noted earlier, there are serious problems in trying to infer the source of a given ERP component from its distribution on the scalp. Thus, the identification of a particular ERP component with a particular intra-cranial source must depend on a variety of converging methodologies including magnetic and intra-cranial recordings in humans and animals, neuropsychological research and lesion studies in animals.

What are the consequences of knowing the intra-cranial source for theorizing about the functional significance of an ERP component? We should emphasize that by "functional significance" we mean a specification of the information processing transactions that are manifested by the component, not its neurophysiological significance.

In the case of the P300, there has been considerable speculation that at least one source lies in the hippocampus (Okada, Kaufman, & Williamson, 1983; Halgren et al., 1980). If, in fact, this speculation is supported by future research, we could incorporate what is known about the psychologically relevant functions of the hippocampus into our theory of the P300 (and vice versa). In this regard, it is interesting to note the similarity between theories of hippocampal function that emphasize its role in memory (O'Keefe & Nadel, 1978) and the theory of P300 outlined in the previous sections that emphasizes its relationship to "context updating."

Note that we can articulate a theory of the functional significance of P300 without any knowledge of its neural origin. However, when its neural origin is known, then the theory can be refined and developed on the basis of what is known about the underlying neural structures. In this sense, then, neurophysiological knowledge may be useful, but not necessarily critical, to the psychophysiological enterprise. Of course, such knowledge is critical for those who wish to use ERPs as a sign of neurophysiological function.

3.6 Using the measure: Psychophysiological inference

In the preceding sections, we have considered the ways in which ERP researchers establish the functional significance of the ERP components. In a sense, this research can be considered as establishing the validity of ERP components as measures of particular psychological activities. In this section, we consider the ways in which the measures are used to make inferences about psychological processes. We shall review a series of inferential steps, from the crudest to the most sophisticated, that depend to a greater and greater extent on assumptions about the functional significance of the ERP. For the purposes of elucidating the inferential process, we shall consider an experiment in which subjects are run in two different conditions.

3.6.1 Inference 1: Conditions are different. At the most fundamental level, we can ask whether or not the two conditions are associated with different responses. The analytic procedure necessary to answer this question would involve a univariate or multivariate analysis of variance (with condition and timepoint as factors). Given that such an analysis yields a significant condition by timepoint interaction, we can infer that the conditions are different. If we assume that the ERP is a sign of brain activity and/or that it reflects some psychological process, then we can infer that the brain activity, and associated psychological processing, are different in the two conditions.

3.6.2 Inference 2: Conditions differ at a particular time. The second level of inference concerns the time at which the two conditions differ. This inference could be made on the basis of post-hoc tests of the significant condition by time interaction. It would take the form of "by at least ms X, processing of stimuli in condition A is different than processing of stimulus in condition B". This kind of inference is frequently made in studies of selective attention where an important theoretical issue concerns the relative time at which an attended event receives preferential processing. As with the most primitive form of inference, we need only assume that the ERP is a reflection of some aspect of psychological processing.

3.6.3 Inference 3: Conditions differ with respect to the latency of some process. For this level of inference, additional assumptions and measurement operations must be made. First, we must assume that the latency of a particular ERP component is related to the latency of a particular psychological process. For example, we argued previously that P300 does not occur until the evaluation process is complete. Second, we must adopt a procedure to identify the component in question and to measure its latency (see section 3). Then, we use an analytic procedure (analysis of variance, t-test, etc.) to evaluate the difference between the conditions with respect to the component latency. As a result of this procedure, we make the inference that the conditions differ with respect to the timing of process X.

3.6.4 Inference 4: Conditions differ with respect to the degree to which some process occurs. At the most complex level, we can use ERPs to infer that a particular process occurs to a greater degree under one condition than under another. In this case, we must assume that a particular ERP component is a manifestation of process X. We must further assume that changes in the magnitude of the component correspond directly to changes in the degree to which the process is invoked. Then, we must devise a suitable measurement procedure to identify and assess the magnitude of the component. Finally, we can proceed with the usual inferential test and determine whether or not the conditions differ with respect to the degree of process X.

4. A Selective Review of ERPs

In this section, we consider some of the research on the functional significance of a variety of different ERP components. We begin with a discussion of ERPs that occur prior to a marker event. This is followed by a brief overview of sensory and cognitive ERP components that occur after the marker event.

4.1 Event-preceding potentials

4.1.1 Movement-related potentials. One class of event-preceding potentials includes those that are apparently related to the preparation for movement. These potentials were first described by Kornhuber and Deecke (1965) who found that, prior to voluntary movements, a negative potential develops slowly, beginning some 800 ms before the initiation of the movement (see Figure 11). These "Readiness Potentials" (or "Bereitschaftspotentials") were distinguished from those that followed the movement, the "Reafferent Potentials." In a condition in which a similar, but passive movement was involved, only post-movement potentials were observed. Both readiness and reafferent potentials tend to be maximal at electrodes located over motor areas of the cortex. Furthermore, some components of the potentials are larger at electrode locations contralateral to the responding limb (at least for hand and finger movements). Indeed, this kind of lateralization has become an important criterion for movement-related potentials.

The investigation of movement-related potentials has developed along several different paths, including (a) the discovery and classification of different components of the movement-related potential (for reviews, see Brunia, Haagh, & Scheirs, 1985, and Deecke et al., 1984), (b) analysis of the neural origin using the scalp topography of ERPs (e.g., Vaughan, Costa, & Ritter, 1972) or magnetic field recordings (e.g., Okada, Williamson, & Kaufman, 1982; Deecke, Weinberg & Brickett, 1982), (c) analysis of the functional significance of different components (see Brunia et al., 1985, and Deecke et al., 1984, for reviews), and (d) recording of movement-related potentials in special populations (e.g., in mentally retarded children, Karrer & Ivins, 1976, and in Parkinson's patients, Deecke, Englitz, Kornhuber, & Schmidt, 1977). In general, these studies confirm that the potential described by Kornhuber and Deecke is generated, at least in part, by neuronal activity in motor areas of the cortex and is a reflection of processes related to the preparation and execution of movements.

Recently, movement-related potentials have been applied to the investigation of human information processing. In particular, we have used the measure to index the commitment to a specific motor response in choice reaction-time paradigms (see Bosco et al., 1986; De Jong, Wierda, Mulder, & Mulder, in press; Gehring, Gratton, Coles, & Donchin, 1986; Gratton, Coles, Sirevaag, Eriksen, & Donchin, in press; Kutas & Donchin, 198[illegible]

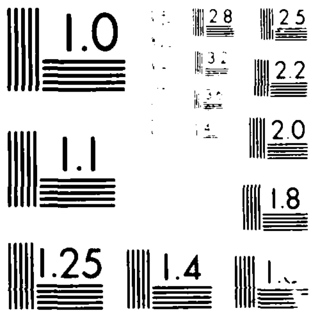
We have demonstrated that the speed and accuracy of a subject's reaction time response is, in part, related to the degree of prior preparation manifested by the movement-related potential.

4.1.2 The contingent-negative variation (CNV). The CNV was first described by Walter, Cooper, Aldridge, McCallum, & Winter (1964) as a slow negative wave that occurs during the foreperiod of a reaction time task (see Figure 12). The wave tends to be largest over central (vertex) and frontal areas. Researchers investigating the functional significance of CNV have manipulated several aspects of the S1-S2 paradigm, including the subject's task, the discriminability of the imperative stimulus, foreperiod duration, stimulus probability, presence of distractors, etc. The component has been variously described as related to expectancy, mental priming, association, and attention (for a review, see Donchin, Ritter, & McCallum, 1978; Rohrbaugh & Gaillard, 1983).

A central controversy in research in this area over the past decade concerns whether the CNV consists of one, or several, functionally distinct, components. A further, but related, question is whether the late portion of the CNV (just prior to the imperative stimulus) reflects more than the process of motor preparation as the subject anticipates making a response to the imperative stimulus. This controversy was raised by Loveless and his co-workers (e.g., Loveless & Sanford, 1974; see also, Connor

AD-A191 244 THE EUENT-RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING A..(U) ILLINOIS UNIU CHAMPAIGN COGNITIUE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 2/10

UNCLASSIFIED 29 FEB 88 CPL-88-1 AFOSR-TR-88-0316 F/G 5/8 NL


1.0
2.8
2.5
2.2
2.0
1.1
1.8
1.25
1.4

& Lang, 1969) who argued that the CNV consists of two components, an early orienting wave (the O-wave) and a later expectancy wave (the E-wave). Subsequent research by these investigators led them to argue that the E-wave is a readiness potential and reflects nothing more than motor preparation. Research by Rohrbaugh et al. (1976) and by Gaillard (1978; see also Rohrbaugh & Gaillard, 1983) also supports this interpretation. However, the question of the functional significance of the latter component (the E-wave) remains controversial. Some investigators have claimed that, because a late E-wave is evident even in situations in which no overt motor response is required, the E-wave has a significance over and above that of motor preparation. However, it is clear that even though the overt motor response requirement may be removed from these situations, attention to a stimulus necessarily involves some motor activity associated with adjustment of the sensory apparatus. Perhaps the most persuasive arguments for a non-motor role for the late CNV comes from a recent study of Damen and Brunia (1987). These authors found evidence for a motor-independent wave that precedes the delivery of feedback information in a time-estimation task.

4.2 Sensory components

The presentation of stimuli in the visual, auditory, or somatosensory modality elicits a series of voltage oscillations that can be recorded from scalp electrodes. In practice, sensory potentials can be elicited either by a train of relatively high frequency stimuli or by transient stimuli. In the former case, the ERP responses to different stimuli overlap in time. The waveforms have quite fixed periodic characteristics that are driven by the periodic stimulation, and are therefore referred to as "steady-state" (see Regan, 1972). In the case of transient stimuli, the responses from different stimuli are separated in time.

For both steady-state and transient potentials the potentials appear to be obligatory responses of the nervous system to external stimulation. In fact, the earlier components of all sensory potentials (within, say, 100 ms) are invariably elicited whenever the sensory system of interest is intact. In this sense, they are described as "exogenous" potentials. They are thought to represent the activity of the sensory pathways that transmit the signal generated at peripheral receptors to central processing systems. Therefore, these components are "modality specific," that is, they differ both in waveshape and scalp distribution as a function of the sensory modality in which the eliciting stimulus is presented. As would be expected of manifestations of primitive sensory processes, the sensory components are influenced primarily by stimulus parameters such as intensity, frequency, etc. For a review of these components, see Hillyard, Picton, and Regan (1978).

For clinical purposes, sensory evoked potentials are used in the diagnosis of neurological diseases (i.e., demyelinating diseases, cerebral tumors and infarctions, etc.). Of particular diagnostic importance are the auditory potentials (diseases involving the posterior fossa), and the steady-state visual potential (multiple sclerosis). Auditory potentials can also be used to diagnose hearing defects in uncooperative subjects.

Since most sensory potentials appears to be insensitive to psychological factors, they have not been used extensively in the study of psychological processes. We should note, however, that there have been reports that some of the middle-latency components (between 10 and 100 ms after stimulus) may reflect selective attention (McCallum, Curry, Cooper, Pocock, & Papakostopoulos, 1983; Michie, Bearpark, Crawford, & Glue, 1987). The relationship between the N100 component (in the visual, auditory, and somatosensory modality) and attention will be reviewed later in this section.

4.3 The early negativities

Several negative components have been described in the period between 100 and 300 ms after the presentation of an external stimulus. In this section, we will examine two families of negative components that have been associated with selective attention and elementary feature analysis. Although these components have been grouped because of functional similarities and latency into a few large sub-groups (N100s, N200s, etc.), their scalp distribution and morphology vary as a function of the modality of the eliciting stimulus. Therefore, these potentials may be considered to lie at the interface between "purely" exogenous (described in section 4.2) and "purely" endogenous components (described later in this section).

4.3.1 The N100s. First indications that ERPs could be used to investigate attentional processes came from studies in which the ERP response to attended stimuli was compared to that to unattended stimuli (e.g., Eason et al, 1964; Hillyard, Hink, Schwent, & Picton, 1973). These kinds of studies suggested that attended stimuli are associated with a more negative ERP between 100 and 200 ms. Subsequent research has been concerned with two issues: (a) the use of ERPs to test theories of selective attention, and (b) the nature of the attentional effect on ERPs.

"Selective attention" refers to the ability of the human information processing system to selectively analyze some stimuli and ignore others. Two metaphors have been associated with selective attention, that of filtering (see Broadbent, 1957) and that of resources (see Kahnemann, 1973, Norman & Bobrow, 1975). Filtering theories have focussed on the debate about the locus of the filter: does filtering occur at an early, perceptual level (early selection theories, Broadbent, 1957) or at later stages of processing (late selection theories, Deutsch & Deutsch, 1963)? According to the resource metaphor, selective attention is a mechanism by which the system allocates more resources to process information coming through a particular attended channel than through other unattended channels.

In a typical paradigm (Hillyard, Hink, Schwent, & Picton, 1973), four types of stimuli are presented. The stimuli (tones) differ along two dimensions (location and pitch), each having two levels (left vs. right ear and standard vs. deviant pitch). The subject is instructed to attend to stimuli at a particular location and to detect target tones of a deviant pitch (e.g., left ear tones of high pitch). To investigate attention effects, ERPs to standard tones occurring in the attended location (channel) are compared to those to standard tones in the unattended channel.

Using this paradigm, Hillyard and his colleagues have observed a larger negativity with a peak latency of about 100-150 ms for stimuli presented in the attended channel (see Figure 13 which shows data from a similar experiment by Knight, Hillyard, Woods, & Neville, 1981). The moment in time at which the waveforms for attended and unattended stimuli diverge is considered as the time at which filtering starts playing a role. Hillyard and his colleagues originally interpreted their data in terms of a modulation of the sensory N100 response (Hillyard et al., 1973). Thus, they considered these data as evidence for early selection.

Later research has shown that the difference between the ERP waveforms for the attended and unattended channels cannot be considered only in terms of the amplitude variation of a peak with a latency between 100 and 150 ms (N100). Rather, it can be characterized by the superimposition of a negative component lasting several hundred milliseconds (Naatanen, 1982). This component has been labelled "Processing Negativity". The onset latency of the Processing Negativity is related to the difficulty of the discrimination between the attended and the unattended channel (Hansen & Hillyard, 1983). The Processing Negativity interpretation of the selective attention effect is consistent with the resource metaphor. Rather than indicating a filtering process in the information processing flow, the Processing Negativity might reflect a selective allocation of processing resources to the attended channel.

Similar variations in the amplitude of ERP components with a latency of 100 to 200 ms have been observed for visual (Harter & Aine, 1984) and somatosensory (Desmedt & Robertson, 1977) stimuli. In these modalities, however, the difference between the attended and unattended channels is characterized more by an amplification of the ERP peaks than by the superimposition of a sustained negativity (Harter & Aine, 1984; Michie, Bearpark, Crawford, & Glue, 1987). Again, this may be viewed either in terms of early filtering that reduces the processing of irrelevant information, or in terms of extra resources devoted to process the critical information.

This review is necessarily an oversimplification of the complexity of ERP we have loosely referred to as "N100". For example, in a recent review Naatanen and Picton (1987) identified no less than six different components that are active around the time of N100!

4.3.2 The N200s. While the amplitude of the N100 (or of the Processing Negativity) appears to reflect the selection of information from a particular perceptual channel, the amplitude of the N200 component reflects the detection of deviant features. As with "N100," "N200" is used to refer to a family of components, one for each modality, that are similar in function and latency. Thus, different N200s can be observed for the visual modality (with maximum at the occipital electrode) and for the auditory modality (with maximum at the central or at the frontal electrode).

Squires, Squires, and Hillyard (1975) manipulated stimulus frequency and task relevance independently and found that the N200 was larger for rare stimuli, regardless of their task relevance. A similar observation has been made by Naatanen (1982) in the kind of paradigm used by Hillyard and his colleagues that we described in the previous section. Note that, to isolate N200 effects, the comparison of interest is between the rare targets and the frequent non-targets, rather than between the attended and the unattended channels (as is the case with N100 effects). N200s will be observed to rare stimuli presented on either channel: therefore the N200 does not require selective attention for its appearance (Squires, Squires & Hillyard, 1975)[3]. Not only does the amplitude of the N200 depend on stimulus probability, the latency of the N200 component is dependent on the difficulty of the discrimination between target and non-target stimuli (Naatanen, 1982). Furthermore, the amplitude of the N200 is also proportional to the difference between frequent and rare stimuli (see Figure 14). Therefore, Naatanen proposed that the N200 reflects the operation of an automatic "mismatch detector," and he labelled this component "mismatch negativity." Since the N200 appears to be related to the automatic detection of surprising (rare) events, this component has been related to the "orienting reflex" (see Naatanen & Gaillard, 1983). Furthermore, since the N200 is related to the automatic processing of rare features, it may be a reflection of the automatic stage of feature analysis proposed by some recent theories of perception (see Treisman & Gelade, 1980).

The N200 has been used in the investigation of mental chronometry. In particular, Ritter, Simson, Vaughan, and Macht (1982) and Renault (1983) have observed that the latency of this

---

[3] Note that, in this paradigm, the P300 is only elicited by rare targets presented on the attended channel. Thus, the P300 is only elicited by task-relevant stimuli, while the N200 appears to be "automatically" elicited by rare stimuli regardless of their task-relevance.

component covaries with reaction time. The high correlation between N200 latency and reaction time may reflect the importance of automatic feature discrimination processes (signalled by the N200) in determining the latency of the overt response. However, the subtraction technique used by Ritter et al. (1982) to derive their measures of N200 must be interpreted with caution, since the latencies of the components in the original waveforms differ. Furthermore, motor potentials, which are characterized by a large negativity, will also covary quite strictly with reaction times. Thus, it is important to disambiguate the N200 component from motor potentials when the former component is used in the study of mental chronometry.

4.4 The late cognitive components

In this section, we review a sample of the research dealing with two major endogenous components, the P300 and the N400. For reasons of space we do not discuss in detail other late components, particularly a group of "Slow Waves." At the present time, the functional significance of the slow waves is largely unknown. However, see Sutton and Ruchkin (1984) and the research on the O-wave (Section 4.2.2) for further information.

4.4.1 The P300. In Section 3.4, we reviewed research on the antecedents of P300, and we briefly alluded to attempts to evaluate a theory of P300 by studying its consequences. In this section, we focus on a series of studies that have sought to investigate the relationship between P300 and the memorability of events that elicit it.

As we noted earlier (Section 3.4), unexpected events that are relevant to the subject's task elicit large P300s. This led Donchin (1981; Donchin & Coles, in press) to formulate the "context updating" hypothesis of the functional significance of P300. This hypothesis allows one to generate predictions about the consequences of the elicitation of a large P300 component. The context updating hypothesis assumes that the elicitation of a P300 reflects a process involved in the updating of representations in working memory. Rare or unexpected events should lead to an updating of the current memory schemas, because only by so doing can an accurate representation of the environment be maintained. The updating process may involve the "marking" of some attribute of the event that made it "distinctive" with respect to other events. This updating of the memory representation of an event is assumed to facilitate the subsequent recall of the event, by providing valuable retrieval cues, so that the greater the updating that follows an individual event, the higher the probability of later recalling that event. P300 amplitude is assumed to be proportional to the degree of updating of the memory representation of the event. Therefore, as the updating process is supposed to be beneficial to recall, P300 amplitude should also predict the subsequent recall of the eliciting event.

The relationship between P300 and memory has been tested in various paradigms. For example, Karis, Fabiani, and Donchin (1984) recorded ERPs to words presented in series that contained a distinctive word (an "isolate") (cf. von Restorff, 1933). The isolation was achieved by changing the size of the characters in which the word was displayed. As is well documented (Cimbalo, 1978; von Restorff, 1933; Wallace, 1965), isolated items are better recalled than are comparable non-deviant items (the von Restorff effect). The isolated items, being rare and task-relevant, can be expected to produce large P300s. Thus, we could predict that the recall variance would be related to the very factors that are known to elicit and control P300 amplitude. Karis et al. (1984) found that the magnitude of the von Restorff effect depends on the mnemonic strategy employed by the subjects. Rote memorizers (i.e., subjects who rehearse the words by repeating them over and over) showed a large von Restorff effect, and poor recall performance, relative to elaborators (i.e., subjects who combine words into complex stories or images in order to improve their recall). For all subjects, isolates elicited larger P300s than non-isolates. For rote memorizers, isolates subsequently recalled elicited larger P300s on their initial presentation, than did isolates that were not recalled. This relationship between recall and P300 amplitude was not observed in elaborators (see Figure 15). It is noteworthy that the amplitude of a frontal-positive slow wave was correlated with subsequent recall in the elaborators, suggesting that this component may be related to the degree of elaborative processing.

Karis et al. (1984) interpreted these data as evidence that all subjects "noticed" the isolated words and reacted by updating their memory representations and producing large P300s. The differences among the subjects emerged when subjects tried to memorize the stimuli by using different types of rehearsal strategies. These strategies interacted with the retrieval processes: when subjects used rote strategies, changes in the stimulus representation, induced by the isolation and manifested by P300, made it easier to recall the word. For the elaborators, whose recall depended on the networks of associations formed as the series were presented, the effects of the initial memory activation and updating manifested by P300 were not noticeable, because they were overshadowed by the more powerful elaborative processing that occurred after the time frame of P300.

The interpretation of the data of the Karis et al. (1984) study capitalized on different strategies used by different subjects. The hypothesis that the relationship between P300 amplitude and subsequent recall indeed depends on the mnemonic strategy used by the subject was tested in a subsequent study in which strategy was manipulated by instruction on a within-subject basis (Fabiani, Karis, & Donchin, 1985; see also Fox & Michie, 1987). Strategy instructions were effective in manipulating the performance of the subjects. When instructed to use rote strategies, subjects recalled fewer words, and displayed a larger von Restorff effect, than when they used elaborative strategies. Analyses of the ERPs also supported our predictions. When subjects were instructed to use rote strategies, the P300s elicited by words subsequently recalled were significantly larger than those elicited by words subsequently not recalled. Such a relationship was not observed when subjects used elaborative strategies. In addition, when rote strategies were used, subjects recalled the size of the words better than in the elaborative condition. This suggests that size is a "distinctive" attribute of the memory representation of the word, such as to help recall in the case of rote instructions.

In another experiment, Fabiani, Karis, and Donchin (1986a) employed an incidental memory paradigm (an "oddball" task) to reduce the use of elaborative strategies. The results confirmed the prediction that stimuli that were subsequently recalled had larger P300s at the time of initial presentation than those for stimuli that were not recalled.

All the studies described above use either the von Restorff paradigm or the oddball paradigm. However, the memory effect is not limited to isolated or rare items. It is also present for the non-isolates and for the frequent items in the oddball experiment.

In several recent studies, memory paradigms have been used that do not capitalize on stimuli for which the P300 is expected to be enhanced, that is, paradigms in which neither the distinctiveness nor the probability of occurrence of the stimuli to be memorized are manipulated. A seminal study in this respect is that by Sanquist, Rohrbaugh, Syndulko and Lindsley (1980) which found that larger amplitude P300s (or late positive components) were elicited, in a same-different judgment task, by stimuli that were correctly recognized in a subsequent recognition test. Johnson, Pfefferbaum, and Kopell (1985) recorded ERPs in a study-test memory paradigm. They reported that the P300 associated with subsequently recognized words was slightly, but not significantly, larger than that elicited by non-recognized words. Paller, Kutas, and Mayes (1985) recorded ERPs in an incidental memory paradigm in which the subjects were asked to make either a semantic or a nonsemantic decision, and were subsequently, and unexpectedly, tested for their recognition or recall of the stimuli. They found that ERPs elicited during the decision task were predictive of subsequent memory performance, in that the P300 (late positive complex) was larger for words subsequently recalled or recognized than to words not recalled or recognized. Similarly, Paller, McCarthy, and Wood (1987) recorded ERPs in two semantic judgment tasks, which were followed by a free recall and by a recognition test. ERPs to words later remembered were more positive than those to words later not remembered, even though the memory effect was smaller for recognition than for recall. Neville, Kutas, Chesney, and

Schmidt (1986) recorded ERPs to words that were either congruous or incongruous with a preceding sentence in a task in which subjects were asked to judge whether or not the word was congruent with the sentence. They found that the amplitude of a late positive component (P650) predicted subsequent recognition.

Taken together, the experiments described in this section suggest that the P300, as well as other late components of the ERP, can have an important role in illuminating some of the cognitive processes that accompany stimulus encoding, and that are usually opaque to traditional techniques. While the first group of studies conducted in our laboratory has focussed on the interaction between distinctiveness and rehearsal strategies in determining the ERP/memory relationship, the other studies have shown that such a relationship is widespread over a variety of different memory paradigms. It is important to emphasize that memory is a complex phenomenon that can be influenced by a multitude of variables and that can be probed with a number of different tests. Thus, it is unlikely that a single component of the ERP can be identified as the "memory component." It is much more likely, and indeed more interesting, that a series of ERP components will prove to be significant in different memory tasks.

4.4.2 The N400. The N400 component of the ERP was first described by Kutas and Hillyard (1980a), who recorded ERPs in a sentence reading task. In this paradigm, words are presented serially and the subject is asked to read them silently in order to answer questions about the content of the sentence at the end of the experiment. In two studies reported by Kutas and Hillyard (1980a), 25% of the sentences ended with a semantically incongruous (but syntactically correct) word. These incongruous words elicited an N400 component that was larger than that elicited by words that were congruous with respect to the meaning of the sentence. Furthermore, the amplitude of the N400 appeared to be proportional to the degree of incongruity: moderately incongruous words ("he took a sip from the waterfall") had a smaller N400 than strongly incongruous words ("he took a sip from the transmitter"). Kutas and Hillyard (1982) reported that the N400 to incongruous endings was slightly larger and more prolonged over the right than the left hemisphere.

These results have been replicated and extended repeatedly, using variations of the sentence reading paradigm described above. The aim of these studies has been to determine whether the N400 is a manifestation of a distinctively semantic process, or whether it is elicited by other kinds of deviance. Kutas and Hillyard (1984) found that the amplitude of the N400 was inversely related to the subject's expectancy of the terminal word (cloze probability), while it was insensitive to sentence constraints (i.e., to the number of possible alternative endings). Kutas and Hillyard (1980b) recorded ERPs to the terminal word in a sentence. This word could be semantically or physically deviant, or not deviant. They interpreted their data as indicating that an N400 followed semantic deviation, while a late positive complex (P300) followed physical deviation[4]. In addition, Kutas and Hillyard (1983) inserted a number of semantic and grammatical anomalies in prose passages. They found that large N400s were associated with the semantic anomalies embedded in the text, while the negativities recorded to grammatical errors were inconsistent and had a scalp distribution different from that of the N400. Kutas, Lindamood, & Hillyard (1984) found that anomalous words that were semantically related to the sentence's "best completion" ("the pizza was too hot to drink") elicited smaller N400s than anomalous words unrelated to the best completion ("the pizza was too hot to cry"). This suggests that the degree of semantic relatedness is an important determinant of the N400 (see Figure 16).

A large N400 component is also evoked by semantic anomalies presented in the auditory modality (McCallum, Farmer, & Pocock, 1985), or in anomalies embedded in American Sign Language (ASL) gestures (Neville, 1985). However, Macar and Besson (1987) did not find N400 responses to anomalous endings of musical tunes.

N400-like components have also been recorded in paradigms other than sentence reading. For example, Fischler, Bloom,

---

[4] We should note that there is some controversy concerning whether or not a P300 follows the occurrence of an N400 in the case of semantic deviation. For present purposes, it is sufficient to note that the physical deviation was not followed by an N400.

Childers, Roucos, and Perry (1983) recorded ERPs in a sentence verification paradigm. In this paradigm, sentences are presented in segments ("a robin / is / a bird"), and two dimensions are orthogonally manipulated: whether the sentences are positive or negative ("is," "is not"), and whether they are true or false. The subject is required to indicate whether the sentence is true or false. A large negativity was elicited by false affirmative ("a robin / is / a tree") and true negative ("a robin / is not / a tree") sentences, that is by sentences in which the first and last element were semantically unrelated. In a similar study (Fischler, Childers, Achariyapaopan, & Perry, 1985) subjects were required to learn the occupations of fictitious people. They were then asked to indicate whether sentences of the type "John / is / a dentist" were true or false. They found that false statements were associated with larger negativities than true statements.

Rugg recorded N400-like components in a semantic priming paradigm (Rugg, 1985) and in a rhyme judgment task (Rugg & Barrett, submitted), although there are some differences between the scalp distribution of the negativity recorded in these paradigms and that recorded in the sentence reading task.

In general, research reviewed in this section suggests that there is a brain response (the N400) that is specifically sensitive to the violation of semantic expectancies. Measures of this component, then, should prove useful in testing theories and models relating to semantic priming (e.g. Van Petten & Kutas, 1987).

5. Potential Applications of ERP Research

In this section we consider the possible contribution of ERP research to problems that arise in various branches of psychology, with a particular emphasis on social psychology. Before we review some specific research examples, we need to consider two general issues related to the application of ERP research.

First, as we noted earlier, for some ERP components there is sufficient knowledge about their physiological significance for them to be used as markers of physiological functioning. This knowledge has led to the use of ERPs in neurological diagnosis (see, for example, Halliday, 1987 for a review). Second, we have devoted much of this chapter to a discussion of the psychological significance of ERP components - that is, we have attempted to show that the components are manifestations of particular cognitive activities. One implication of this is that, to the extent that questions that arise in any branch of psychology can be recast in terms of questions about cognitive activities, there is at least the possibility that the measurement of ERPs will be useful.

In the following sections, we review several examples of how such research has proceeded or might proceed. For reviews of other applications of ERP research to cognitive and engineering psychology, see Donchin, Karis, Bashore, Coles, and Gratton (1986) and Donchin, Kramer, and Wickens (1986).

5.1 ERPs and emotion

5.1.1 Emotion and cognition. In previous sections of this chapter, we have emphasized the view that ERPs can be regarded as manifestations of cognitive processes. For the most part, the ERP literature contains few studies of ERP-emotion relationships. This tendency to emphasize the cognitive rather than the emotional is probably due to the belief that the psychophysiological analysis of emotions is more the province of students of the autonomic nervous system. However, recent attempts to re-evaluate the distinction between cognition and emotion call into question the justification for their psychophysiological separation. For example, if the experience of emotion is attributed, at least in part, to the process of cognitive appraisal (e.g. Arnold, 1960), then the so-called "cognitive" ERPs have the potential to be useful in the analysis of appraisal processes. Similarly, if emotions are believed to influence cognitive processes (e.g., Bower, 1981), it may be that ERPs will be useful in describing such an influence.

These kinds of ideas are behind the recent analysis of ERPs in situations in which subjects anticipate and receive emotional stimuli (Johnston, Miller, & Burleson, 1986; Klorman & Ryan, 1980; Simons, MacMillan, & Ireland, 1982) and in studies of subjects suffering from emotional dysfunction (Miller, 1986; Yee & Miller, in press). In the following sections, we review two other kinds of studies in which there is evidence of concern with cognitive-emotion issues.

5.1.2 Lie detection. In preceding sections, we have emphasized that ERPs can be used to explore situations involving emotion, if these situations can be viewed in terms of the information processing activities. In the case of lie detection, studies using autonomic measures have similarly emphasized the importance of cognitive factors. Thus, rather than propose a specific emotional "lie-response", these studies have sought to use the measures to determine whether an individual has specific knowledge (cf. the guilty-knowledge test, Lykken, 1974; Podlesny & Raskin, 1977).

In the case of studies of Lie Detection using ERPs, researchers have used a paradigm in which stimulus words or phrases relating to a crime would be categorized in one way if the information they represented was unknown to the individual - but in another way, if the information was known (see Farwell & Donchin, 1986). The role of the ERPs, then, is to identify the categorization rule being used by the subject. This is accomplished by using a variant of the "oddball" task. Three classes of stimuli are presented: (a) words or phrases related to the crime (the probe stimuli); (b) words of phrases unrelated to the crime and unknown to the subject (irrelevant stimuli); and (c) words or phrases unrelated to the crime, but which are known to the subject (target stimuli). The subject is instructed to count the target stimuli.

Thus, the critical question is whether the subject will categorize the probe stimuli as irrelevant (indicating no knowledge) or as target stimuli (indicating knowledge). This question is answered by setting up the probabilities of the different stimuli such that the probe stimuli would elicit a large response (P300) if they are categorized as targets. Farwell and Donchin (1986) have shown that, if the probabilities of irrelevant, probe and target stimuli are .67, .17, and .16 respectively, the procedure is effective in determining the categorization rule used by the individual. In this way, the presence of guilty knowledge can be determined.

In this example, the role of the ERP is to determine whether a particular set of stimuli is classified by the subject as belonging to one of two other categories. In principle, then, it could be used in any situation in which the experimenter is interested in identifying classification rules.

5.1.3 Bargaining. Bargaining, like most forms of social interaction, involves both cognitive and emotional processes. Thus, bargaining presents an interesting topic of study for the analysis of cognitive-emotional relationships.

In the study to be described in this section (Karis, Druckman, Lissak, & Donchin, 1984), the focus was on the cognitive ERP responses to events believed to have emotional consequences. The subject was engaged in a "bargaining" session with the computer. The object of bargaining was the purchase of a used car. The subject and the computer alternated in making their concessions, until they agreed on a price. The subject's objective was to buy the car at the lowest possible price. ERPs were recorded while the subject viewed the computer's concession.

An interesting aspect of this study was that the computer adopted either one of two strategies in making its concessions: it could be either "generous" or "stingy." When the computer was in the generous mode it would concede 80% of the subject concession, while when it was in the stingy mode it would concede only 20% of the subject concession. Thus, it was important to the subject to guess the computer's strategy, in order to take advantage of it.

Switches in the computer's strategy (which occurred randomly) reliably elicited P300s that were larger than those recorded to trials in which the strategy remained the same. This can be viewed in terms of cognitive processes: To the extent that P300 represents a process of context updating, it seems reasonable that the subject would need to update his/her mental schema at the moment in which a task relevant event changes. Another interesting aspect of this study is that a photograph of the subject's face was taken every time the subject saw the computer's concession. The facial expression of the subject was then analyzed in terms of the emotions displayed. It is interesting to note that, on some occasions, the subject smiled when the concession was generous. Thus, the appraisal of a concession (as manifested by the ERP) was, in some cases, associated with the facial expression of emotion.

5.2 ERPs and "pay-off" manipulations

The social interaction between experimenter and subject can have an influence on the strategy adopted by the subject in performing his or her assigned task. Subjects approach experimental situations with hypotheses about the goals of the experiment and, most importantly, the pay-offs associated with different goals. Information about these pay-offs can be derived from explicit instructions as well as more implicit forms of communication between subject and experimenter.

In most ERP studies, we try to constrain the subject's task so as to reduce the options available to the subject. In this way, assumptions can be made about the processes used by the subject to perform the task.

In several studies, instructions or pay-off schedules have been deliberately manipulated in order to alter the subject's strategy in performing the task. In this case, the ERP focus is on understanding how strategic changes are implemented. An example of this approach has been described in section 4.4.1 where we saw that mnemonic strategies, manipulated by instruction, were important determinants of the P300/recall relationship and, by implication, of the role of the rehearsal of distinctiveness cues in memory.

In the following two sections, we review other examples of this kind of approach.

5.2.1 Decision-making. The way in which humans reach decisions is an important problem for social psychologists. Most theories of decision-making agree that subjective probability and pay-off structure are fundamental determinants of human choices (Fishbein & Ajzen, 1975).

Karis, Chesney, and Donchin (1983) investigated the role of these two dimensions in determining the amplitude of P300. In this study, subjects had to predict which of three stimuli (the numbers 1, 2, or 3) would next occur. The probability of the three stimuli were .45, .10, and .45, respectively. The subjects were rewarded for correct predictions following one of two schedules, run in different blocks. In one schedule ("all-or-none"), subjects were only rewarded for correct predictions. In another pay-off schedule ("linear"), subjects were not only rewarded for correct choices, but also received half bonuses when their predictions were off by one (i.e., a 1 occurred when a 2 was predicted). Thus, in the "linear" schedule, the optimal strategy was to predict 2 (objectively the rare event).

Karis et al. found that although subjects did indeed vary their overt predictions following the pay-off schedules, their

P300s were always larger for the rare stimuli (the 2s). That the subjects were aware of the probability structure in each of the two conditions was confirmed by subjective reports. These results indicate that P300 reflects subjective probability rather than a conscious decision about which stimulus was most rewarding.

Karis et al. also studied variations in subject's strategy. In particular they were able to quantify the "riskiness" of a particular choice, by estimating the expected loss with respect to the optimal behavior. They observed that P300 amplitude was proportional to the riskiness of the prediction. They interpreted this result in terms of differences in the relevance of the information conveyed by the stimulus in cases of high and low risk choices.

5.2.2 Speed-accuracy instructions. Kutas, McCarthy, and Donchin (1977) analyzed the relationship between P300 and reaction time, when subjects were given speed or accuracy instructions. While the timing of the reaction time response in relation to the P300 changed with instruction, the latency of the P300 itself was only slightly affected. Thus, in this case, the effect of instructions was to change the coupling between the processes associated with P300 and those associated with overt behavior. In terms of information processing, it appears that speed instructions lead the subject to generate an overt response before information has been fully processed.

A similar modulation of the relationship between the processing of stimulus information and the timing of the overt response has been described by Gratton et al. (in press; see also Gehring et al., 1986). In this case, measures of the readiness potential (see section 4.1.1) were diagnostic of the strategic mode of the subject (accurate versus fast).

5.3 ERPs and communication

In this chapter, we have considered ERPs as a tool for investigating cognitive processing. In particular, we have emphasized how ERPs can be used to obtain information about cognitive functions which would not be available in other ways. In somewhat more extreme terms, we may consider ERPs as a communication device.

An interesting example of this approach is given by a recent study by Farwell, Donchin, & Kramer (1986). These investigators have devised a technique that allows people to communicate by exploiting the relationship between ERPs and attention. Subjects in this study were presented with a matrix of letters that were illuminated in rapid succession. By examining the ERP responses to each of the flashes, they were able to determine the letters to which the subjects were attending. Thus, letter by letter, the subjects were able to communicate entire words, by means of their ERPs. This type of communication device may be particularly useful for cases in which people cannot communicate in other ways, such as in cases of patients that have lost any motor ability as a consequence of trauma and disease.

5.4 Summary

In the examples reviewed in this section, we have seen how ERPs can be used to study or, in one case, enhance social behavior, broadly defined. The underlying theme of these examples has been that, if the question of interest can be considered in terms of the information processing transactions that occur, ERPs may be a useful tool. This is because ERPs can be regarded as manifestations of information processing activities.

Upon analysis, there appear to be several other instances in which such a conception is quite feasible, such as when the interest of the investigator is on the cognitive processing of social information. For example, the recent development of "social cognition" as a sub-discipline within social psychology has led to the treatment of a subset of "social processes" as involving the operation of the cognitive apparatus on social information (see, for example, Wyer & Srull, 1984). The only restriction in the application of ERPs to this domain is that the information be presented discretely so that segments of the EEG can be identified that are time-locked to the events of interest.

In other cases, considerable efforts may be needed in order to apply the model of cognitive psychophysiology. For example, evaluations of social interactions would require not only that the interactions be structured so as to permit discrete events to be identified. One would also need to incorporate repetitions of the "same" event into the interaction (if averaging is necessary to extract stable ERPs) and, most importantly, to identify the putative cognitive processes associated with the events. The imposition of this kind of structure may result in a distortion of the essence of the interaction process.

On the basis of the foregoing examples and discussion, there appears to be good reason to expect considerable benefit from an application of a cognitive psychophysiology based on ERPs in various areas within psychology, particularly social psychology. Just as problems in cognitive psychology are proving more tractable when psychophysiological measures are added to those measures traditionally available, problems in social psychology may prove to be similarly more tractable. However, the benefits of such an application will only accrue if the psychophysiological measures are derived under appropriate circumstances and if such circumstances can be created without distortion of the phenomenon of interest.

6. References

Allison, T., Wood, C. C., & McCarthy, G. (1986). The central nervous system. In M. G. H. Coles, S. W. Porges, & E. Donchin (Eds.), Psychophysiology: Systems, processes, and applications (pp. 5-25). New York, NY: Guilford.

Arnold, M. B. (1960). Emotion and personality. New York: Columbia University Press.

Beatty, J., Barth, D. S., Richer, F., & Johnson, R. A. (1986). Neuromagnetometry. In M. G. H. Coles, S. W. Porges, & E. Donchin (Eds.), Psychophysiology: Systems processes, and applications (pp. 5-25). New York, NY: Guilford.

Berger, H. (1929). Uber das Elektrenkephalogramm das menchen. Archiv fur Psychiatrie, 87, 527-570.

Bosco, C. M., Gratton, G., Kramer, A. F., Wickens, C. D., Coles, M. G. H., & Donchin, E. (1986). Information extraction and components of the event-related potential. Psychophysiology, 23, 426. (Abstract)

Bower, G. H. (1981). Mood and memory. American Psychologist, 36, 129-148.

Broadbent, D. E. (1957). A mathematical model for human attention and immediate memory. Psychological Review, 64, 205-215.

Brunia, C. H. M., Haagh, S. A. V. M., & Scheirs, J. G. M. (1985). Waiting to respond: Electrophysiological measurements in man during preparation for a voluntary movement. In H. Heuer, U. Kleinbeck, & K. H. Schmidt (Eds.), Motor behavior. Programming, control and acquisition (pp. 35-78). Berlin: Springer-Verlag.

Buchwald, J., & Squires, N. (1983). Endogenous auditory potentials in the cat: A P300 model. In C. Woody (Ed.), Conditioning (pp. 503-515). New York: Plenum Press.

Cimbalo, R. S. (1978). Making something stand out: The isolation effect in memory performance. In M. M. Grunneberg, P. E. Morris, & R. N. Sykes (Eds.), Practical aspects of memory (pp. 101-110). New York: Academic Press.

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., Kramer, A. F., & Miller, G. A. (1986). Principles of signal acquisition and analysis. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), Psychophysiology: Systems, Processes, and Applications (pp. 183-221). New York, NY: Guilford.

Coles, M. G. H., Gratton, G., & Gehring, W. J. (1987). Theory in cognitive psychophysiology. Journal of Psychophysiology, 1, 13-16.

Connor, W. H., & Lang, P. J. (1969). Cortical slow-wave and cardiac rate responses in stimulus orientation and reaction time conditions. Journal of Experimental Psychology, 82, 310-320.

Csepe, V., Karmos, G., & Molnar, M. (1987). Effects of signal probability on sensory evoked potentials in cats. International Journal of Neuroscience, 33, 61-71.

Damen, E. J. P., & Brunia, C. H. M. (1987). Changes in heart rate and slow brain potentials related to motor preparation and stimulus anticipation in a time estimation task. Psychophysiology, De Jong, R., Wierda, M., Mulder, G., & Mulder, L. J. W. (0000). The timing of response preparation. Journal of Experimental Psychology: Human Perception and Performance.

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

Deecke, L., Englitz, H. G., Kornhuber, H. H., & Schmitt, G. (1977). Cerebral potentials preceding voluntary movement in patients with bilateral or unilateral parkinson akinesia. In J. E. Desmedt (Eds.), Attention, voluntary contraction, and event-related cerebral potentials. Progress in Clinical Neurophysiology, vol. 1 (pp. 151-163). Basel: Karger.

Deecke, L., Weinberg, H., & Brickett, P. (1982). Magnetic fields of the human brain accompanying voluntary movements. Bereitschaftsmagnetfeld. Experimental Brain Research, 48, 144-148.

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.

Desmedt, J. E., & Robertson, D. (1977). Differential enhancement of early and late components of the cerebral soamtosensory evoked potentials during forced-paced cognitive tasks in man. Journal of Physiology, 271, 761-782.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. Psychological Review, 70, 80-90.

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 Potentials Research (EPA-600/9-77-043) (pp. 501-510). Washington, D.C.: U.S. Govt. Printing Office.

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

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

Donchin, E., & Coles, M. G. H. (0000). Is the P300 component a manifestation of context updating? Behavioral and Brain Sciences.

Donchin, E., Gratton, G., Dupree, D., & Coles, M. G. H. (0000). 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., & Heffley, E. (1978). Multivariate analysis of event-related potential data: A tutorial review. In D. Otto (Eds.), 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., Heffley, E., Hillyard, S. A., Loveless, N., Maltzman, I., Ohman, A., Rosler, F., Ruchkin, D., & Siddle, D. (1984). Cognition and event-related potentials: II. In R. Karrer, J. Cohen, & P. Tueting (Eds.), Brain and information: Event-related potentials. Annals of the New York Academy of Sciences, Vol. 25 (39-57).

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. (1986a). Cognitive psychophysiology and human information processing. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), Psychophysiology: Systems, processes, and applications (pp. 244-267). New York: Guilford Press.

Donchin, E., Kramer, A. F., & Wickens, C. D. (1986b). 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 (pp. 702-718). New York: Guilford Press.

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

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. (1979). The time constant in P300 recording. Psychophysiology, 16, 53-55.

Eason, R. G., Aiken, L. R., Jr., White, C. T., & Lichtenstein, M. (1964). Activation and behavior: II. Visually evoked cortical potentials in man as indicants of activation level. Perceptual and Motor Skills, 19, 875-895.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (0000). 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.

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

Fabiani, M., Karis, D., & Donchin, E. (1986a). P300 and recall in an incidental memory paradigm. Psychophysiology, 23, 298-308.

Fabiani, M., Karis, D., & Donchin, E. (1986b). P300 and memory. In W. C. McCallum, R. Zappoli, & F. Denoth (Eds.) Cerebral psychophysiology: Studies in event-related potentials (pp. 63-69). Amsterdam: Elsevier. (Supplement 38 to Electroencephalography and Clinical Neurophysiology.)

Farwell, L. A., & Donchin, E. (1986). The "Brain Detector:" P300 in the detection of deception. Psychophysiology, 23, 435. (Abstract)

Farwell, L. A., Donchin, E., & Kramer, A. F. (1986). Talking heads: A mental prosthesis for communicating with event-related potentials of the EEG. Psychophysiology, 23, 435. (Abstract)

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W., Jr. (1983). Brain potentials related to stages of sentence verification. Psychophysiology, 20, 400-409.

Fischler, I., Childers, D. G., Achariyapaopan, T., & Perry, N. W., Jr. (1985). Brain potentials during sentence verification: Automatic aspects of comprehension. Biological Psychology, 21, 83-105.

Fishbein, M., & Ajzen, I. (1975). Belief, attitude, intention and behavior. Reading, MA: Addison-Wesley.

Fox, A. M., & Michie, P. T. (1987). P3 and memory relationship under instructed learning strategy conditions. Paper presented at the 4th International Conference on Cognitive Neurosciences, June, Paris-Dourdan, France.

Gaillard, A. (1978). Slow brain potentials preceding task performance. Doctoral Dissertation. Soesterberg, The Netherlands: Institute for Perception (TNO).

Gehring, W., Gratton, G., Coles, M. G. H., & Donchin, E. (1986). Response priming and components of the Event-Related Brain Potential. Psychophysiology, 23, 437-438. (Abstract)

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

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. (1987). A procedure for using multielectrode information in the analysis of components of the Event-Related Potentials: Vector Filtering. Submitted.

Gratton, G., Coles, M. G. H., Sirevagg, E., Eriksen, C. W., & Donchin, E. (0000). Pre- and post-stimulus activation of response channels: A psychophysiological analysis. Journal of Experimental Psychology: Human Perception and Performance.

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.

Halliday, A. M., Butler, S. R., & Paul, R. (Eds.) (1987). A textbook of clinical neurophysiology. Chichester, England: Wiley.

Harter, M. R., & Aine, C. J. (1984). Brain mechanisms of visual selective attention. In R. Parasuraman & D. R. Davies (Eds.), Varieties of attention (pp. 293-321). London: Academic Press.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. Science, 182, 177-180.

Hillyard, S. A., Picton, T. W., & Regan, D. (1978). Sensation, perception, and attention: Analysis using ERPs. In E. Callaway, P. Tueting, & S. H. Koslow (Eds.), Event-related brain potentials in man (pp. 223-321). New York, NY: Academic Press.

Horst, R. L., & Donchin, E. (1980). Beyond averaging II: Single trial classification of exogenous event-related potentials using Step-Wise Discriminant Analysis. Electroencephalography and Clinical Neurophysiolgy, 48, 113-126.

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

Johnson, R., Jr. (1986). A triarchic model of P300 amplitude. Psychophysiology, 23, 367-384.

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., Pfefferbaum, A., & Kopell, B. S. (1985). P300 and long-term memory: Latency predicts recognition time. Psychophysiology, 22, 498-507.

Johnston, V. S., Miller, D. R., & Burleson, M. H. (1986). Multiple P3s to emotional stimuli and their theoretical significance. Psychophysiology, 23, 684-694.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

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., 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.

Karrer, R., & Ivins, J. (1976). Steady potentials accompanying perception and response in mentally retarded and normal children. In R. Karrer (Ed.), Developmental psychophysiology of mental retardation (pp. 361-417). Springfield, IL: Thomas.

Klorman, R., & Ryan, R. M. (1980). Heart rate, contingent negative variation, and evoked potentials during anticipation of affective stimulation. Psychophysiology, 17, 513-523.

Knight, R. T., Hillyard, S. A., Woods, D. L., & Neville, H. J. (1981). The effects of frontal cortex lesions on event-related potentials during auditory selective attention. Electroencephalography and clinical Neurophysiology, 52, 571-582.

Kornhuber, H. H., & Deecke, L. (1965). Hirnpotentialanderungen bei Wilkurbewegungen und passiven Bewegungen des Menschen: Bereitschaftpotential und reafferente Potentiale. Pflugers Archives fur die gesammte Physiologie, 248, 1-17.

Kramer, A. F. Sirevaag, E. J., & Braune, R. (1987). A psychophysiological assessment of operator workload during simulated flight missions. Human Factors, 29, 145-160.

Kutas, M., & Donchin, E. (1980). Preparation to respond as manifested by movement-related brain potentials. Brain Research, 202, 95-115.

Kutas, M., & Hillyard, S. A. (1980a). Reading senseless sentences: Brain potentials reflect semantic incongruity. Science, 207, 203-205.

Kutas, M., & Hillyard, S. A. (1980b). 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., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. Memory and Cognition, 11, 539-550.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. Nature, 307, 161-163.

Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum & J. Requin (Eds.), Preparatory states and processes (pp. 217-237). Hillsdale, NJ: Erlbaum.

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

Lorente de No, R. (1947). Action potential of the motoneurons of the hypoglossus nucleus. Journal of Cellular and Comparative Physiology, 29, 207-287.

Loveless, N. E., & Sanford, A. J. (1974). Slow potential correlates of preparatory set. Biological Psychology, 1, 303-314.

Lykken, D. T. (1974). Psychology and the lie detector industry. American Psychologist, 29, 725-739.

Macar, F., & Besson, M. (1987). An event-related potential analysis of incongruity in music and other non-linguistic contexts. Psychophysiology, 24, 14-25.

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.

McCallum, W. C., Curry, S. H., Cooper, R., Pocock, P. V., & Papakostopoulos, D. (1983). Brain event-related potentials as indicators of early selective processes in auditory target localization. Psychophysiology, 20, 1-17.

McCallum, W., C., Farmer, S. F., & Pocock, P. V. (1984). The effects of physical and semantic incongruities on auditory event-related potentials. Psychophysiology, 24, 449-463.

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

Michie, P. T., Bearpark, H. M., Crawford, J. M., & Glue, L. C. T. (1987). The effects of spatial selective attention on the somatosensory event-related potential. Psychophysiolgy, 24, 449-463.

Miller, G. A. (1986). Information processing deficits in anhedonia and perceptual aberration: A psychophysiological analysis. Biological Psychiatry, 21, 100-115.

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 event-related potential (ERP). In A. W. K. Gaillard & W. Ritter (Eds.), Tutorials in ERP research: Endogenous components (pp. 119-141). Amsterdam: North Holland.

Naatanen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. Psychophysiology, 24, 375-425.

Neville, H. J. (1985). Biological constraints on semantic processing: A comparison of spoken and signed languages. Psychophysiology, 22, 576. (Abstract).

Neville, H. J., Kutas, M., Chesney, G., & Schmidt, A. L. (1986). Event-related brain potentials during initial encoding and recognition memory of congruous and incongruous words. Journal of Memory and Language, 25, 75-92.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. Cognitive Psychology, 7, 44-64.

Nunez, P. L. (1981). Electric fields of the brain: The neurophysics of EEG. London: Oxford University Press.

O'Keefe, J., & Nadel, L. (1978). The hippocampus as a cognitive map. Oxford: Oxford University Press.

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, 416-426.

Okada, Y. C., Williamson, S. J., & Kaufman, L. (1982). Magnetic fields of the human sensory-motor cortex. International Journal of Neurophysiology, 17, 33-38.

Paller, K. A., Kutas, M., & Mayes, A. R. (1985). An investigation of neural correlates of memory encoding in man. Psychophysiology, 22, 607. (Abstract)

Paller, K. A., McCarthy, G., & Wood, C. C. (1987). ERPs predictive of later performance on recall and recognition tests. Paper presented at the 4th International Conference on Cognitive Neurosciences, Paris-Dourdan, France.

Podlesny, J. A., & Raskin, D. C. (1977). Physiological measures and the detection of deception. Psychological Bulletin, 84, 782-799.

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

Regan, D. (1972). Evoked potentials in Psychology, Sensory Physiology, and Clinical Medicine. New York: Wiley.

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.

Ritter, W., Simson, R., Vaughan, H. G., Jr., & Macht, M. (1982). Manipulation of event-related potential manifestations of information processing stages. Science, 218, 909-911.

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: Endogeneous 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.

Rosler, F. (1983). Endogenous ERPs and cognition: Probes, prospects, and pitfalls in matching pieces of the mind-body problem. In A. W. K. Gaillard, & W. Ritter (Eds.), Tutorials in event-related potential research: Endogenous components (pp. 9-35). Amsterdam: Elsevier.

Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. Psychophysiology, 22, 642-647.

Rugg, M. D., & Barrett, S. E. (0000). Event-related potentials and the interaction between orthographic and phonological information in a rhyme-judgment task. Submitted for publication.

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.

Scherg, M., & von Cramon, D. (1985). Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. Electroencephalography and clinical Neurophysiology, 62, 32-44.

Simons, R. F., MacMillan, F. W., & Ireland, F. B. (1982). Anticipatory pleasure deficit in subjects reporting physical anhedonia: Slow cortical evidence. Biological Psychology, 14, 297-310.

Simson, R., Vaughan, H. G., Jr., & Ritter, W. (1976). The scalp topography of potentials associated with missing visual and auditory stimuli. Electroencephalography and Clinical Neurophysiology, 40, 33-42.

Sirevaag, E. J., Kramer, A. F., Coles, M. G. H., & Donchin, E. (1984). P300 amplitude and resource allocation. Psychophysiology, 21, 598-599. (Abstract)

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., & 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., 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.

Starr, A., & Farley, G. R. (1983). Middle and long latency auditory evoked potentials in cat. II Component distribution and dependence on stimulus factors. Hearing Research, 10, 139-152.

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

Sutton, S., Tueting, P., Zubin, J., & John, E. R. (1967). Information delivery and the sensory evoked potentials. Science, 155, 1436-1439.

Sutton, S., & Ruchkin, D. S. (1984). The late positive complex. Advances and new problems. In R. Karrer, J. Cohen, and P. Tueting (Eds.), Brain and Information: Event-Related Potentials. Annals of the New York Academy of Sciences, 425, 1-23.

Towle, V. L., Heuer, D., & Donchin, E. (1980). On indexing attention and learning with event-related potentials. Psychophysiology, 17, 291. (Abstract)

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

Van Petten, C., & Kutas, M. (1987). Ambiguous words in context: An event-related analysis of the time course of meaning activation. Journal of Memory and Language, 26, 188-208.

Vaughan, H. G., Costa, L. D., & Ritter, W. (1972). Topography of the human motor potential. Electroencephalography and Clinical Neurophysiology, 25, 1-10.

Verleger, R. (0000). Event-related potentials and memory: A critique of the context updating hypothesis and an alternative interpretation of P3. Behavioral and Brain Sciences.

Von Restorff, H. (1933). Uber die Wirkung von Bereichsbildungen im Spurenfeld. Psychologische Forschung, 18, 299-342.

Wallace, W. P. (1965). Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. Psychological Bulletin, 63, 410-424.

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.

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, 298-308.

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.

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

Wyer, R. S., & Srull, T. K. (1984). Handbook of social cognition. Hillsdale, NJ: Erlbaum.

Yee, C. M., & Miller, G. A. (0000). Emotional information processing: Modulation of fear in normal and dysthymic subjects. Journal of Abnormal Psychology.

7. Figure Legends

Figure 1. Schematic representation of the operations involved in the recording of Event-Related brain Potentials. From left to right: (a) Top view of the head, indicating the placements of five electrodes (Fz, C3, Cz, C4, and Pz) from which EEG is recorded - note that other locations are also frequently used; (b) the EEG signal is then transferred to an amplifying and filtering system; (c) the amplified and filtered signal may be stored temporarily on an analog magnetic tape; (d) the analog signal is then converted into a digital signal by sampling the potential at a high frequency (usually at least 100 Hz) by an Analog-to-Digital converter; (e) the digitally transformed signal may be stored on a digital store device (magnetic tape or disc); (f) finally, ERPs are extracted from the digitized EEG signal by averaging point-by-point across a large sample of trials (more than 20).

Figure 2. A schematic representation of ERP components elicited by auditory, infrequent target stimuli. The three panels represent three different voltage x time functions: the left bottom panel shows the very early sensory components (with a latency of less than 10 ms); the left top panel shows the middle latency sensory components (with a latency of between 10 and 50 ms); and the right panel shows late components (latency exceeding 50 ms). Note the different voltage and time scales used in the

three panels, as well as the different nomenclatures used to label the peaks (components). (Copyright 1979, Plenum Publishing Corporation. Adapted with permission of the author and publisher from Donchin, 1979.)

Figure 3. Schematic representation of configuration of neurons whose simultaneous polarization does or does not result in a potential detectable by a distant electrode. The electric fields generated by polarization of neurons organized in layers (such as those shown in the left panel) add together to form a powerful field that can be detected from a distant (e.g., scalp) electrode ("open field"). The fields generated by neurons organized concentrically (such as those shown in the right panel) cancel each other to produce a very small field that cannot be detected by a scalp electrode ("closed field"). (Copyright 1947, Alan R. Liss, Inc. Reprinted with permission of the author and publisher from Lorente de No, 1947.)

Figure 4. ERPs elicited by counted, rare tones (upper panel). The data recorded with four different high-pass filter settings ("time constant") are superimposed. Stimulus occurrence is indicated by an S on the time scale. Calibration pulses (lower panel) are plotted on the same voltage x time scale as the ERPs. Note the reduction in amplitude and deformation of the ERP waveshape produced by progressively shorter time constants, that

reduce low frequency activity. (Copyright 1979, The Society for Psychophysiological Research. Reprinted with permission of the author and publisher, from Duncan-Johnson & Donchin, 1979.)

Figure 5. Schematic representation of an ERP waveform, indicating different procedures for component quantification. Three types of peak measures are indicated. The peak latency is obtained by measuring the interval (in ms) between the external triggering event and a positive or negative peak in the waveform. The base-to-peak amplitude measure is obtained by computing the voltage difference (in microvolts) between the voltage at the peak point and a baseline level (usually the average pre-stimulus level). The peak-to-peak amplitude measure is obtained by computing the voltage difference (in microvolts) between the voltage at the peak point and the voltage at a previous peak of opposite polarity. The area measure is obtained by integrating the voltage between two timepoints.

Figure 6. Example of the application of Principal Component Analysis (PCA) to the study of ERPs. A series of ERP waveforms (whose grandaverage is plotted in the upper panel of the figure) are decomposed in several constituent components, whose time courses (component loadings) are shown in the lower panel of the figure. The component loadings are then used to determine the degree to which a particular component is present

in each of the ERP waveforms. (Copyright 1982, Elsevier Science Publishers. Reprinted with permission of the author and publisher from Duncan-Johnson & Donchin, 1982.)

Figure 7. Example of the application of Vector Filtering to the study of ERPs. Waveforms from three electrode locations, shown in the upper panel, are used to determine the contribution of a particular component (e.g. P300), which is characterized by a specific profile of amplitudes at the different electrode locations. The result of this operation is shown in the lower panel.

Figure 8. Grand-average ERP waveforms at Pz from 10 subjects for counted (high - left column) and uncounted (low - right column) stimuli (tones), with different a priori probabilities. The probability level is indicated by a percentage value beside each waveform. Waveforms from a condition in which the subjects were instructed to ignore the stimuli are are also presented for a comparison. The occurrence of the stimulus is indicated by a black bar on the time scale. Positive voltages are indicated by downward deflections of the waveforms. Note that P300 amplitude is inversely proportional to the probability of the eliciting stimulus ("probability effect"), and, at the same probability level, P300 is larger for counted than uncounted stimuli ("target effect"). (Copyright 1977, The

Society for Psychophysiological Research. Reprinted with permission of the author and publisher from Duncan-Johnson & Donchin, 1977.)

Figure 9. ERP waveforms at Pz averaged across subjects for three different semantic categorization tasks. The solid line indicates ERPs obtained during a task in which the subjects had to distinguish between the word DAVID and the word NANCY (the FN condition). The dotted line indicates ERPs obtained during a task in which the subjects had to decide whether a word presented was a male or a female name (the VN condition). The dashed line indicates ERPs obtained during a task in which the subjects had to decide whether a word was or was not a synonym of the word PROD (SYN condition). These three tasks were considered to involve progressively more difficult discriminations. Note the latency of P300 peak is progressively longer as the discrimination is made more difficult. (Copyright 1977, The AAAS. Adapted with permission of the author and publisher from Kutas, McCarthy, & Donchin, 1977.)

Figure 10. Accuracy of reaction time responses given at different latencies ("speed-accuracy functions") for trials with "fast" and "slow" P300. Response latency (defined in terms of the onset of the EMG response) is plotted on the abscissa. The probability that a response would be correct is plotted on the

ordinate. Note that the probability of giving a correct response increases as response latency increases. At very short response latencies, responses are at a chance level of accuracy (.5). At long response latencies, responses are usually accurate. The speed/accuracy function for those trials with P300 latency shorter than the median latency ("fast P300" trials) are indicated by solid lines. The speed/accuracy function for those trials with P300 latency longer the median latency ("slow P300" trials) are indicated by dashed lines. Note that, for each response latency, the probability of giving a correct response is higher when the P300 on that trial (reflecting the speed of stimulus processing on that trial) is fast than when it is slow. (Copyright 1985, The American Psychological Association. Reprinted with permission of the publisher from Coles, Gratton, Bashore, Eriksen, & Donchin, 1985.)

Figure 11. Typical movement related potential (recorded from a central electrode - Cz) preceding a voluntary hand movement. Note that the potential begins about 1 sec before the movement (indicated by the dashed vertical line). The potential can be subdivided into different components as follows: N1 (RP - Readiness Potential, BSP - Bereitschaftspotential); N2 (MP - Motor Potential); and the P2 (RAF - Reafferent Potential). (Copyright 1980, Elsevier Science Publishers. Adapted with

permission of the author and publisher from Kutas & Donchin, 1980.)

Figure 12. Schematic representation of a typical contingent negative variation (CNV) recorded from Cz. The CNV is the negative portion of the wave between the presentation of the warning and imperative stimuli. The early portion of the CNV is labelled "O-wave" (or Orienting wave), while the late portion is labelled "E-wave" (or Expectancy wave). (Copyright 1983, Elsevier Science Publishers. Adapted with permission of the author and publisher from Rohrbaugh & Gaillard, 1983.)

Figure 13. The effect of attention on early components of the auditory event-related potential recorded at the central electrode (Cz). The left panel shows ERPs for tones presented in the left ear. Note that, the difference between the ERPs to attended tones (solid line) versus those for unattended tones (dashed line) consists of a sustained negative potential. A similar difference can be seen for tones presented to the right ear (see right panel). (Copyright 1981, Elsevier Science Publishers. Adapted with permission of the author and publisher from Knight et al., 1981.)

Figure 14. The effects of deviance on "mismatch negativity". A standard (80 db) tone was presented on 90% of the trials and a

deviant tone (57, 70 or 77 db, in different blocks) was presented on 10% of the trials. The ERP to the standard is indicated by the thin line in each panel; the ERP to the deviant tone is indicated by the thick line. As the degree of mismatch between stimuli increases, the mismatch negativity also increases. (The magnitude of the difference between standard and deviant ERPs increases.) (Copyright 1987, The Society for Psychophysiological Research. Adapted with permission of the author and publisher from Naatanen & Picton, 1987.)

Figure 15. ERPs elicited by "isolated" words that were later recalled (solid line) or not-recalled (dashed line). The left column shows ERPs for subjects who used rote mnemonic strategies; the right column shows ERPs for subjects who used elaborative strategies. Note that the amplitude of P300 is related to subsequent recall for the rote memorizers, but not for elaborators. (Copyright 1986, Elsevier Science Publishers. Reprinted with permission of the publisher from Karis, Fabiani, & Donchin, 1986b.)

Figure 16. The effects of anomalous sentence endings on the N400. The ERPs (from Pz) depicted in the figure were recorded following visual presentation of words that varied in their relationship to the previous words in the sentence. For example, for sentences such as "The pizza was too hot to ...", three

endings were possible. Best completion: "eat"; Related anomaly: "drink"; Unrelated anomaly: "cry". Note that the N400 component is only present for anomalies, and is larger for unrelated than for related anomalies. (Copyright 1984, Lawrence Erlbaum Publishers. Adapted with permission of the author and publisher from Kutas, Lindamood, & Hillyard, 1984.)

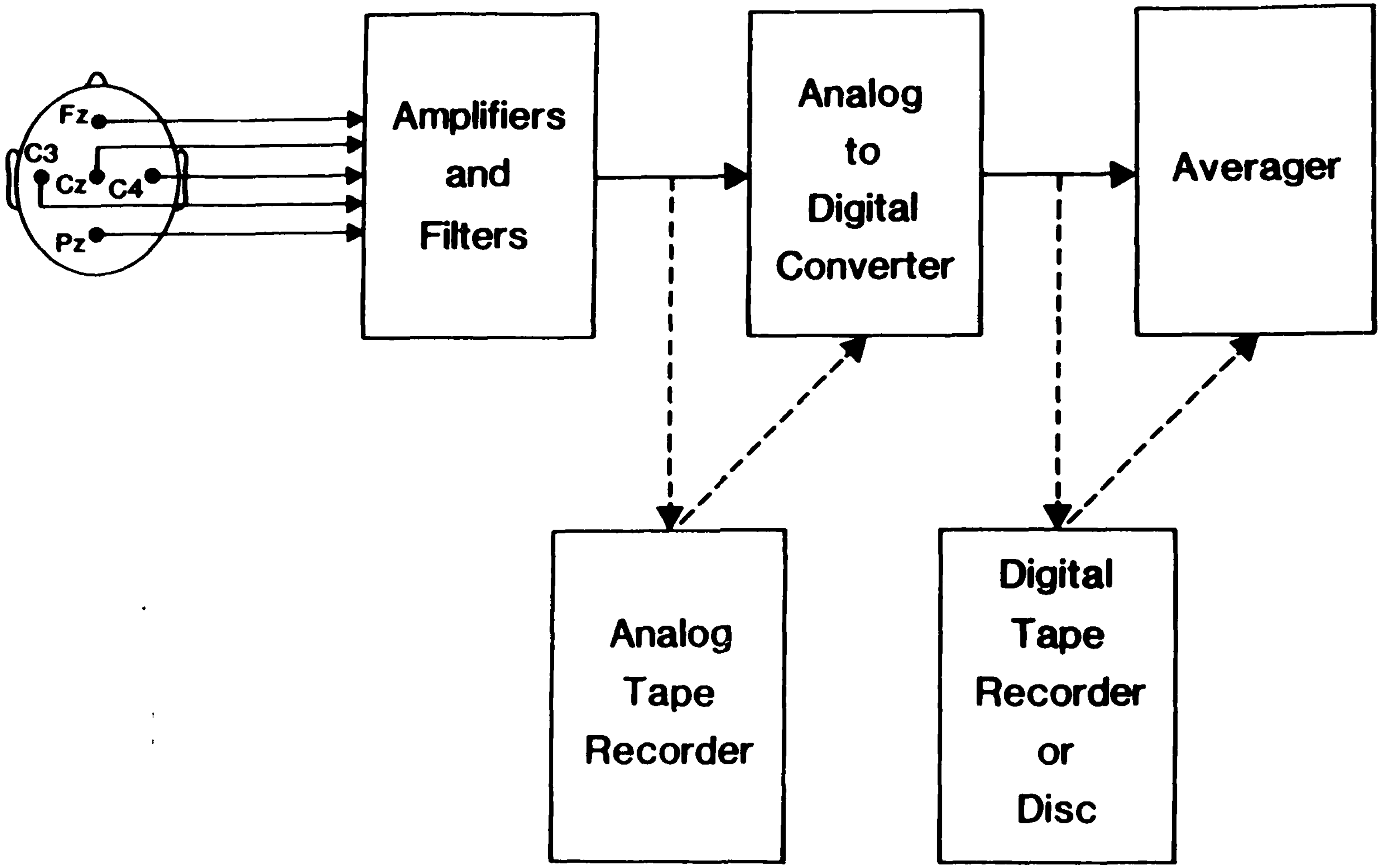
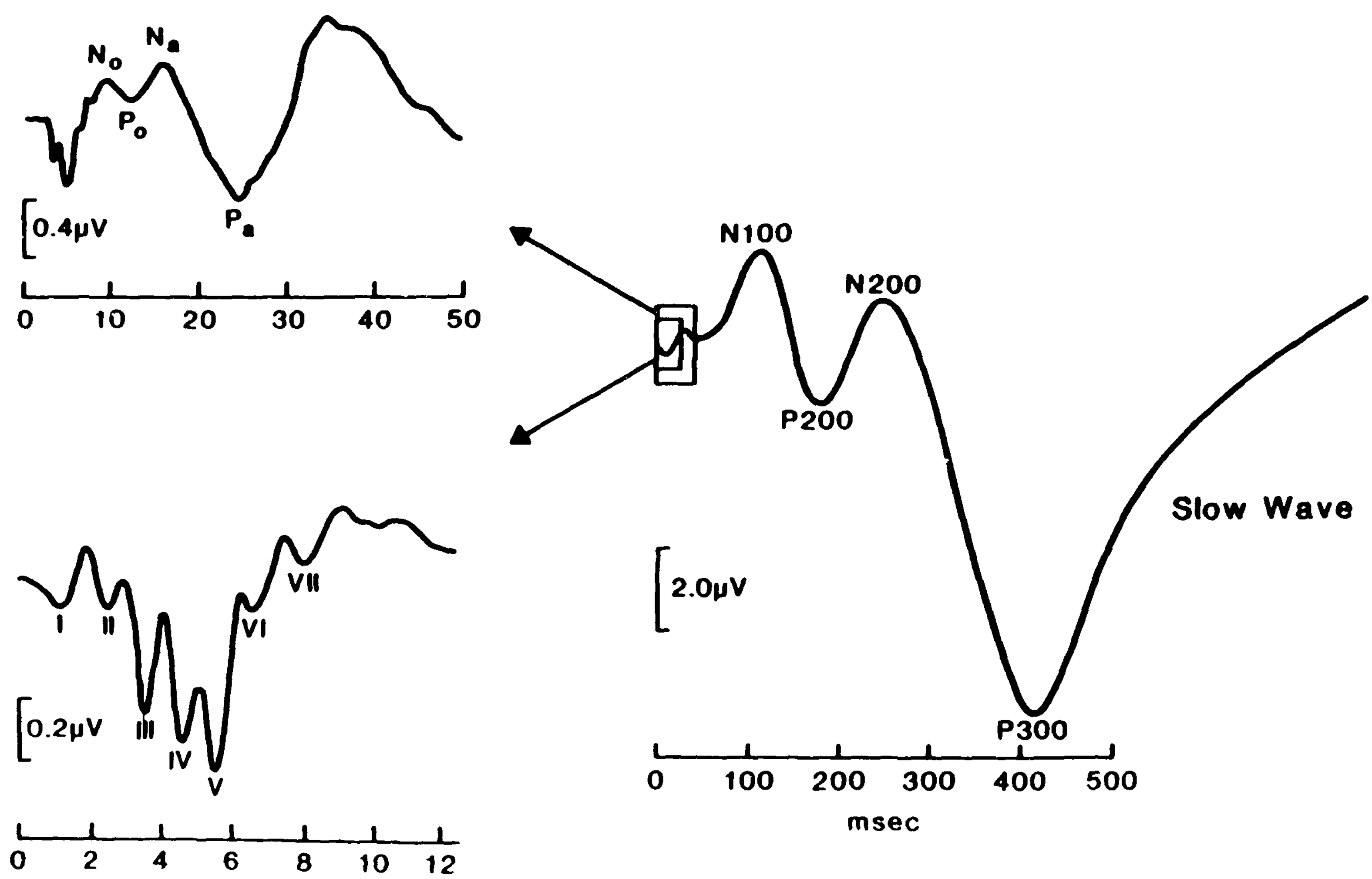

Fz
C3
Cz
C4
Pz
Amplifiers and Filters
Analog to Digital Converter
Averager
Analog Tape Recorder
Digital Tape Recorder or Disc

FIGURE 1



No
Na
Po
Pa
0.4μV
0
10
20
30
40
50
I
II
III
IV
V
VI
VII
0.2μV
0
2
4
6
8
10
12
N100
N200
P200
P300
Slow Wave
2.0μV
0
100
200
300
400
500
msec

FIGURE 2

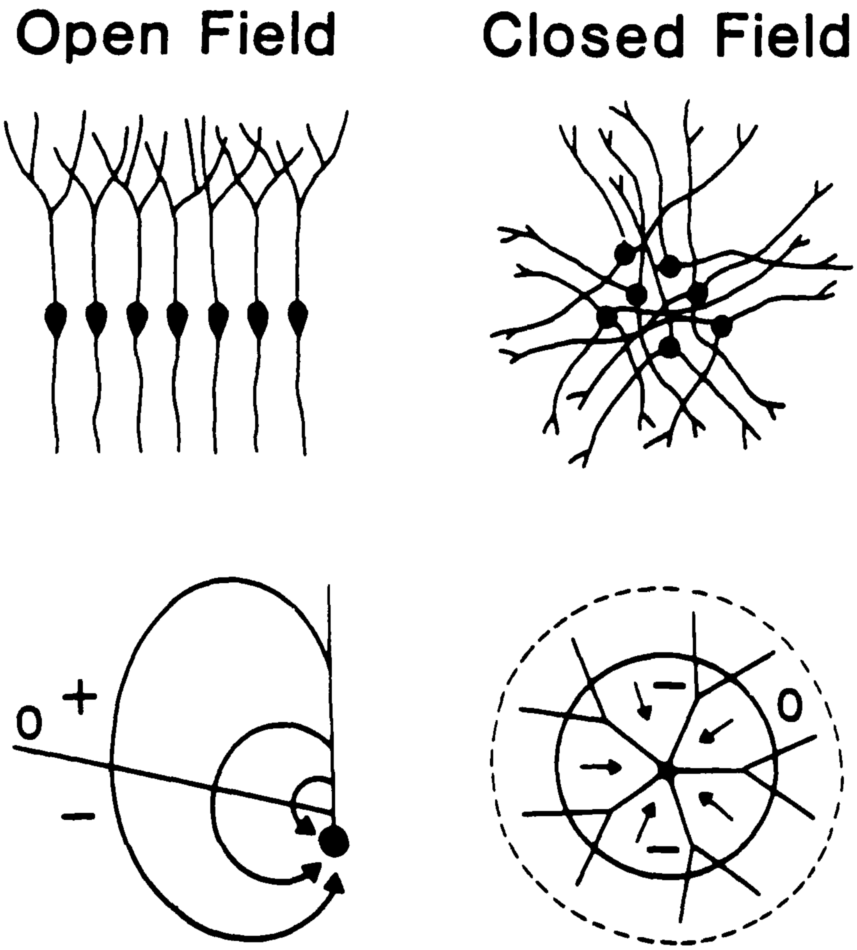


FIGURE 3

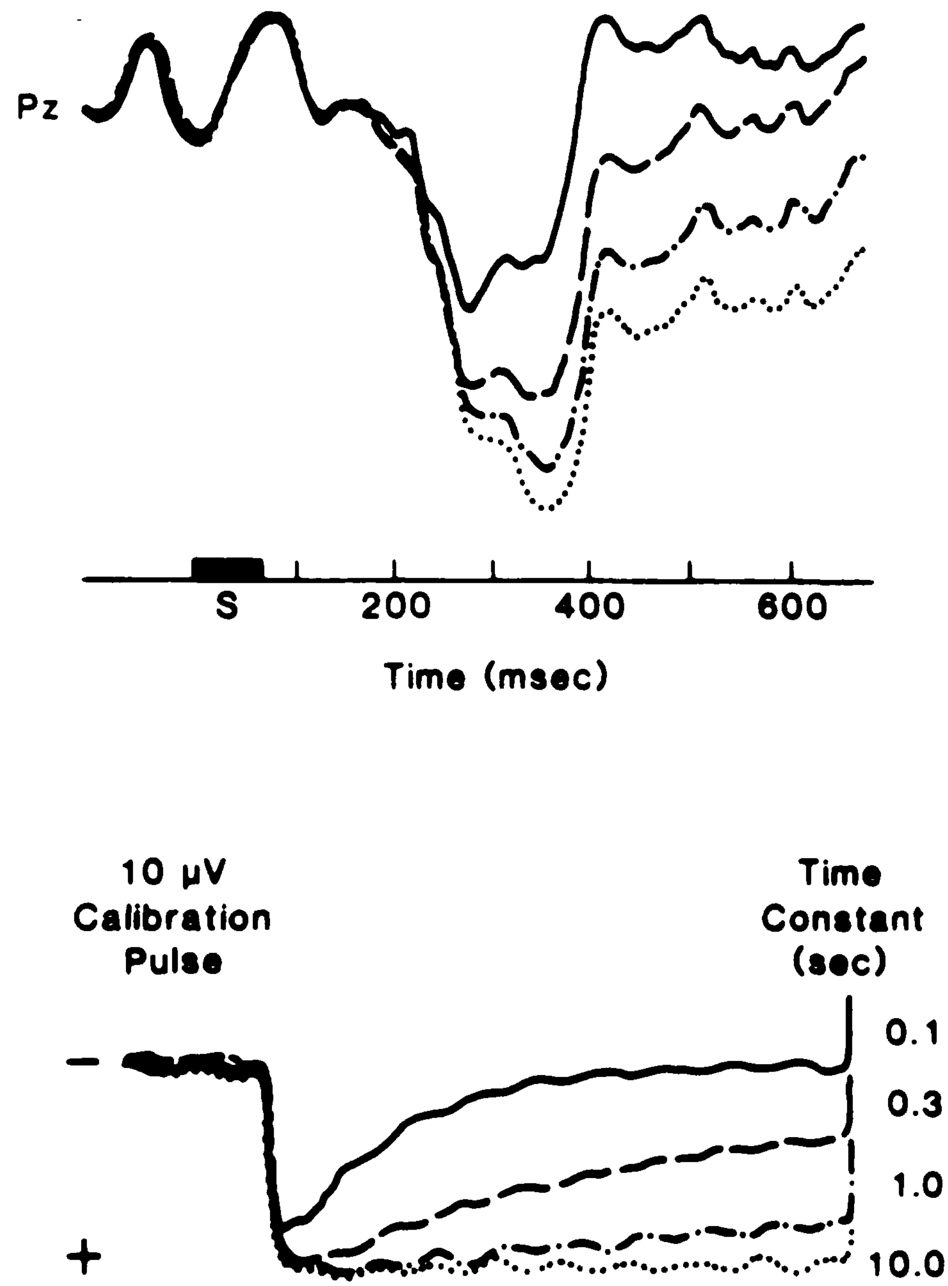

Pz
S
200
400
600
Time (msec)
10 µV
Calibration
Pulse
Time
Constant
(sec)
0.1
0.3
1.0
10.0
−
+

FIGURE 4

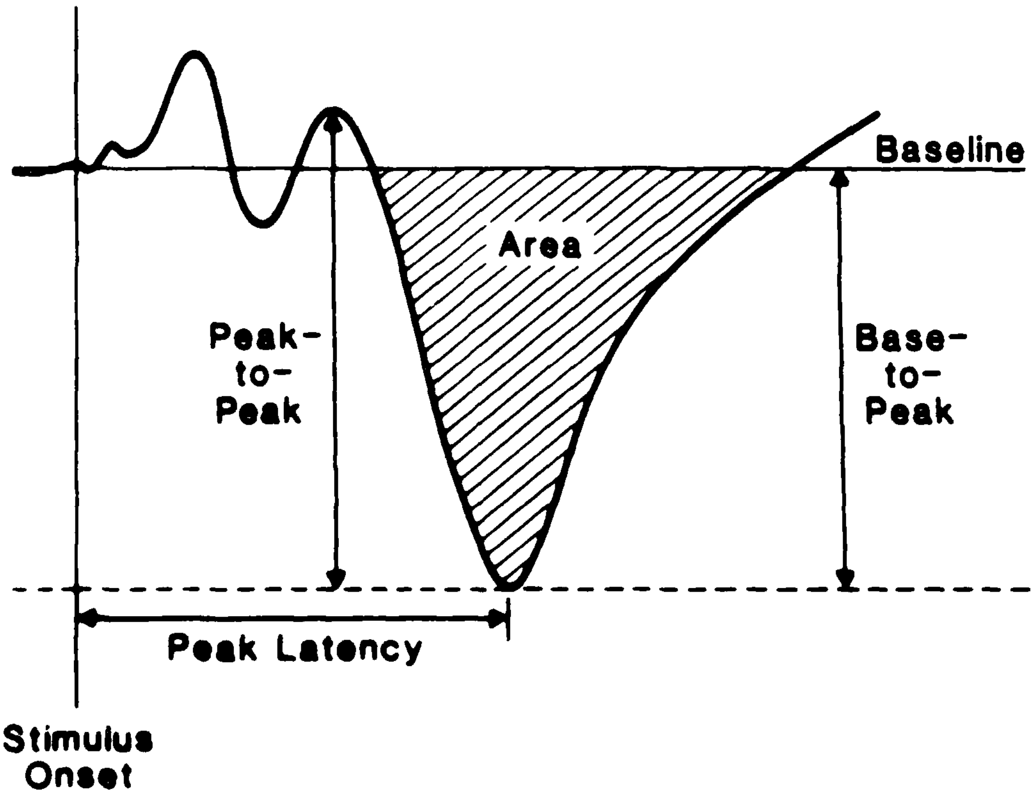

Baseline
Area
Peak-
to-
Peak
Base-
to-
Peak
Peak Latency
Stimulus
Onset

FIGURE 5

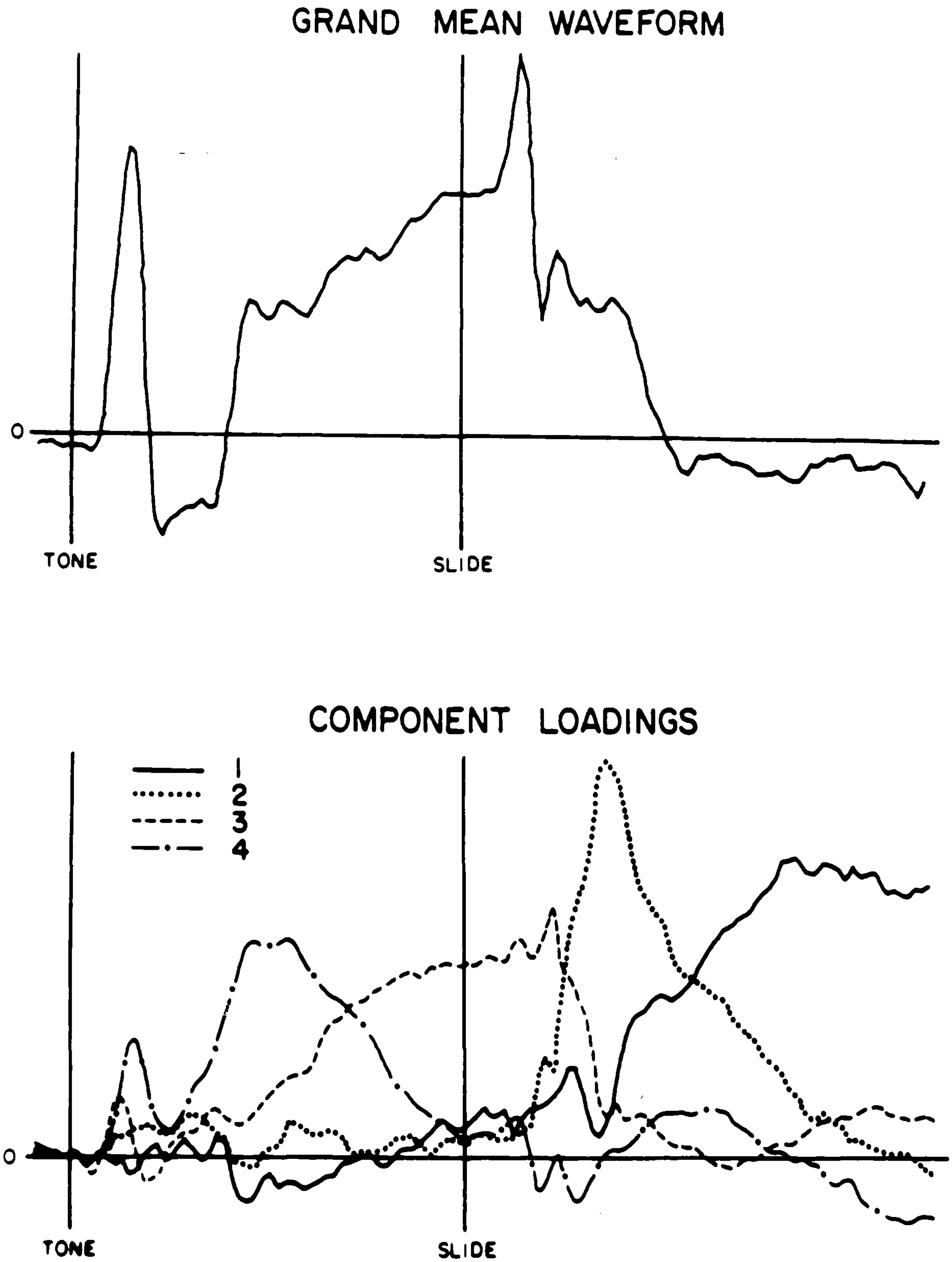

GRAND MEAN WAVEFORM
0
TONE
SLIDE
COMPONENT LOADINGS
1
2
3
4
0
TONE
SLIDE

FIGURE 6

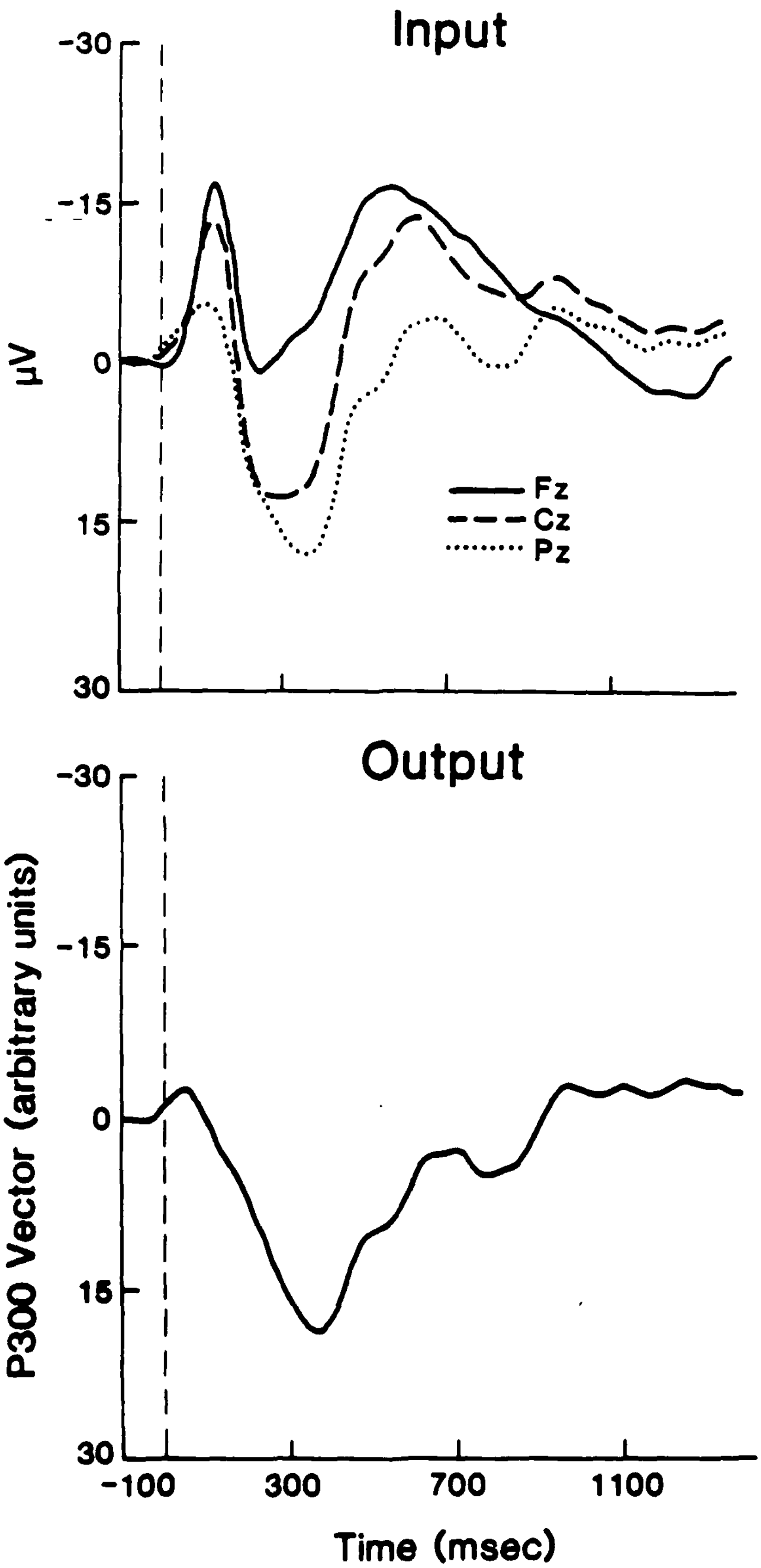

Input
-30
-15
0
15
30
μV
Fz
Cz
Pz
Output
-30
-15
0
15
30
P300 Vector (arbitrary units)
-100
300
700
1100
Time (msec)

FIGURE 7

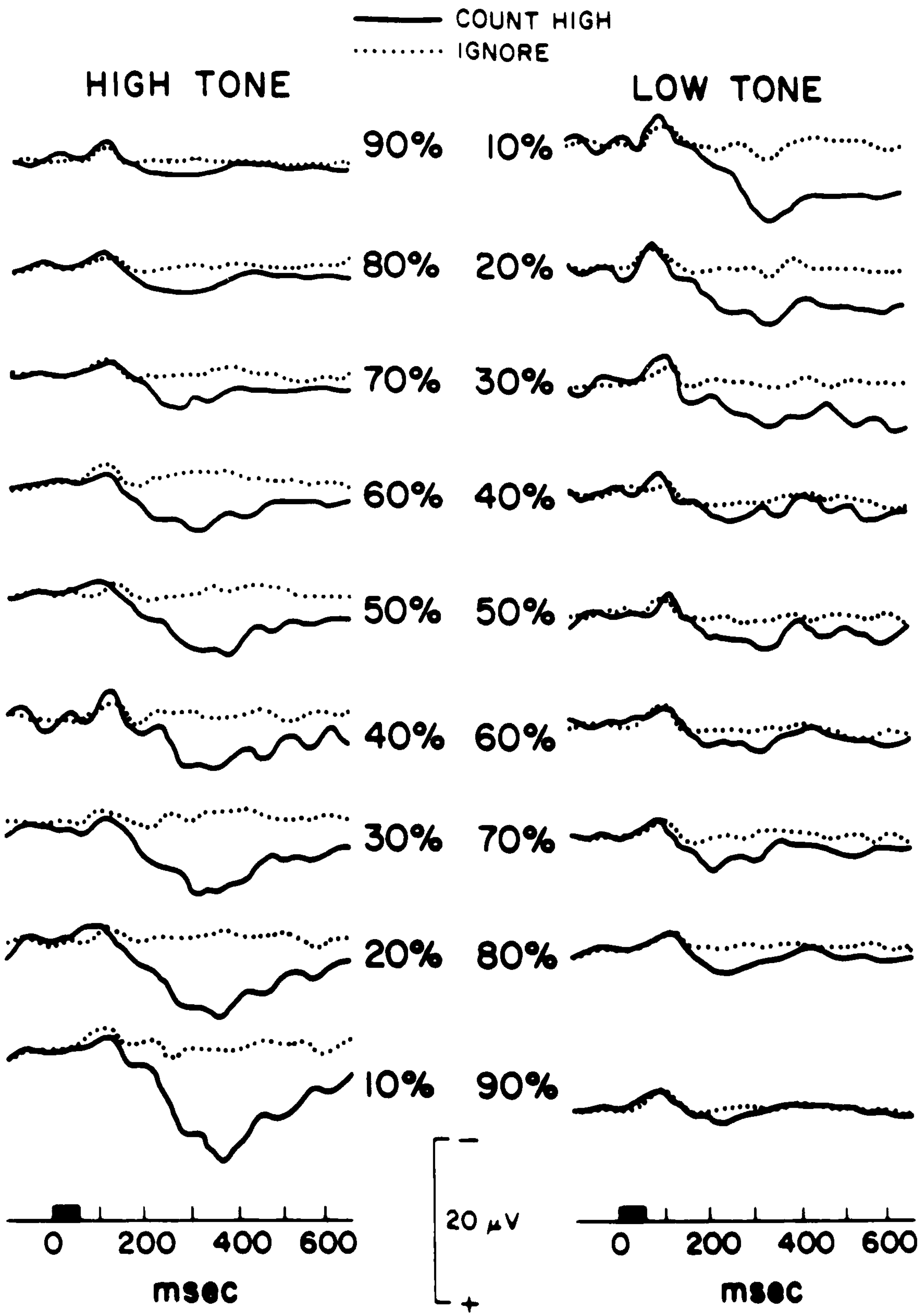

COUNT HIGH
IGNORE
HIGH TONE
LOW TONE
90%
80%
70%
60%
50%
40%
30%
20%
10%
10%
20%
30%
40%
50%
60%
70%
80%
90%
0
200
400
600
msec
20 μV
−
+
0
200
400
600
msec

FIGURE 8

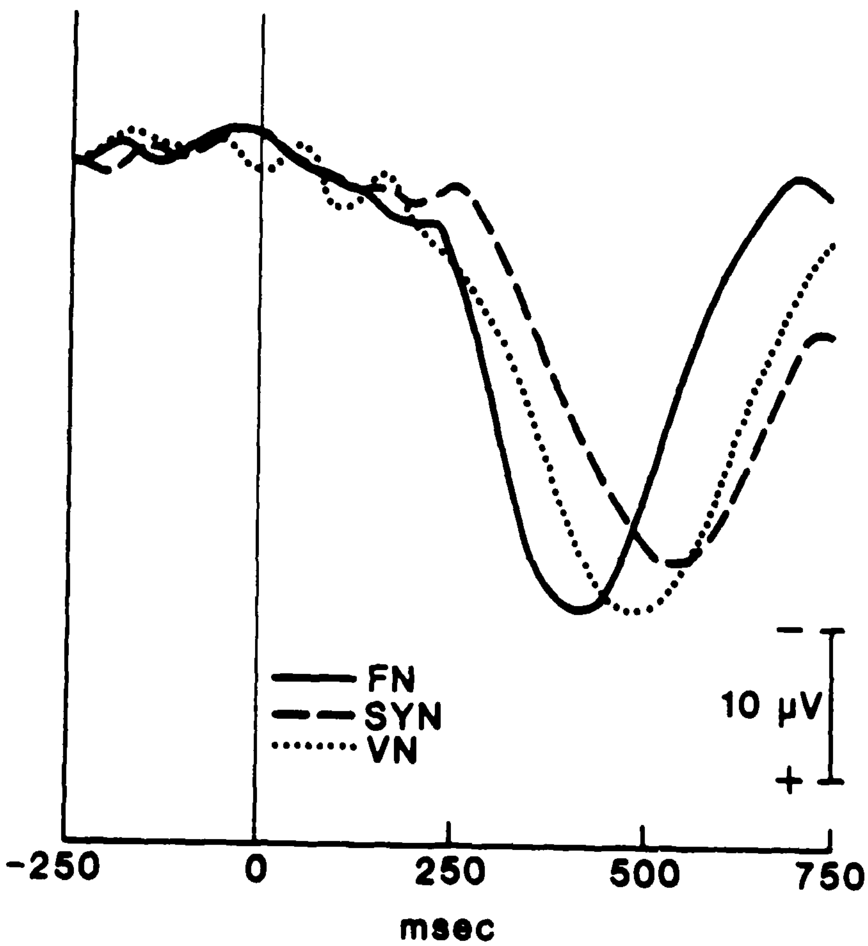


FIGURE 9

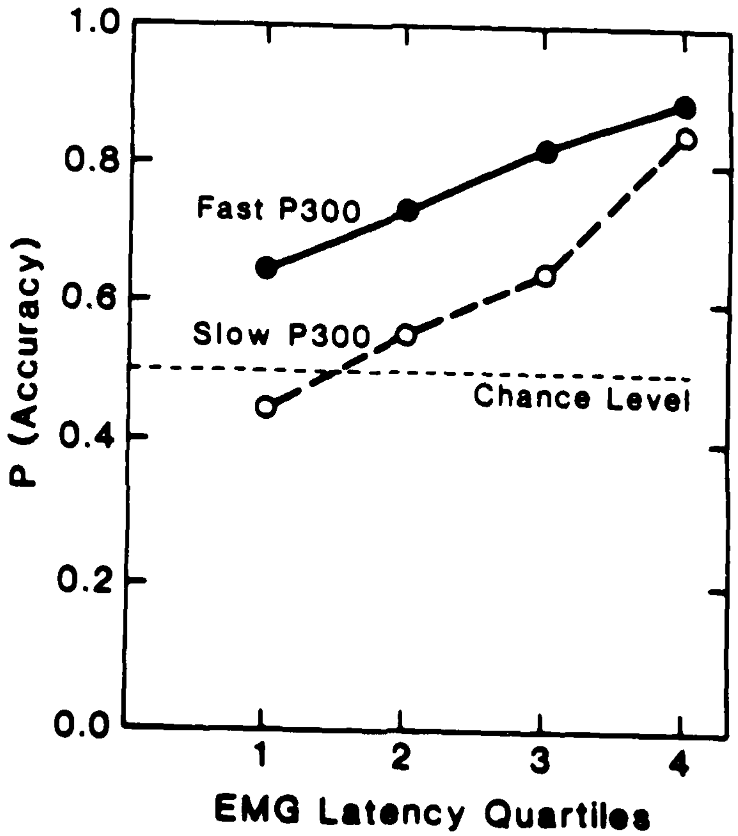

1.0
0.8
0.6
0.4
0.2
0.0
Fast P300
Slow P300
Chance Level
P (Accuracy)
1
2
3
4
EMG Latency Quartiles

FIGURE 10

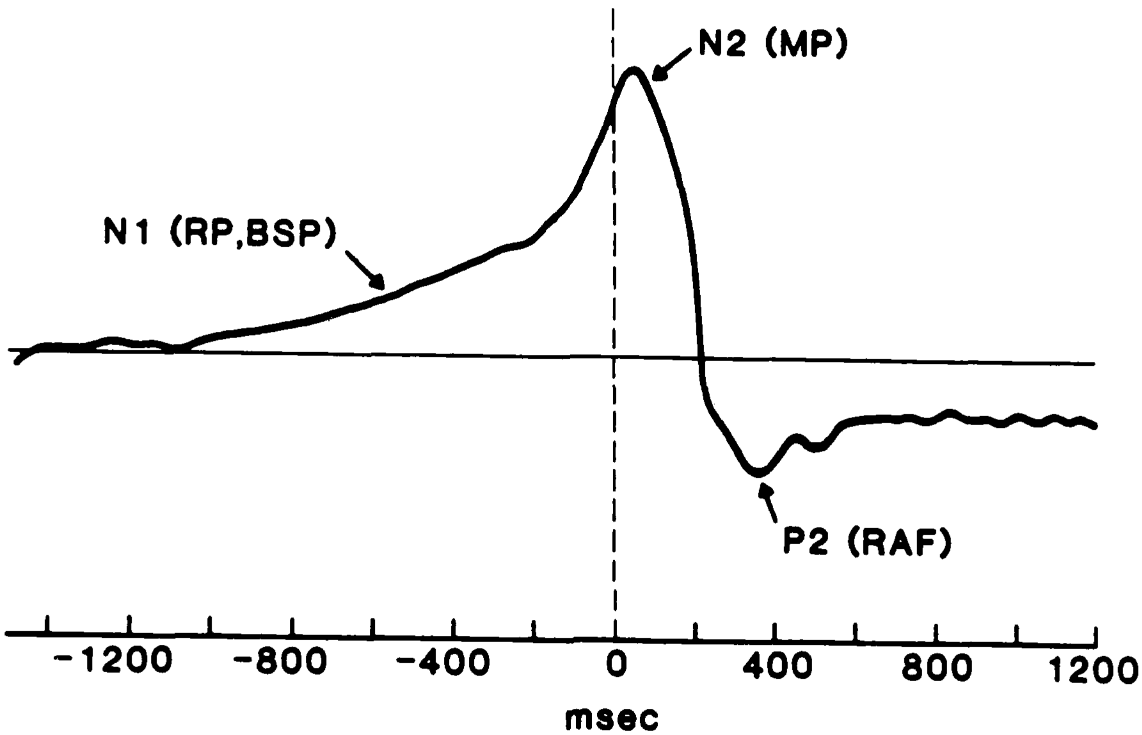


FIGURE 11

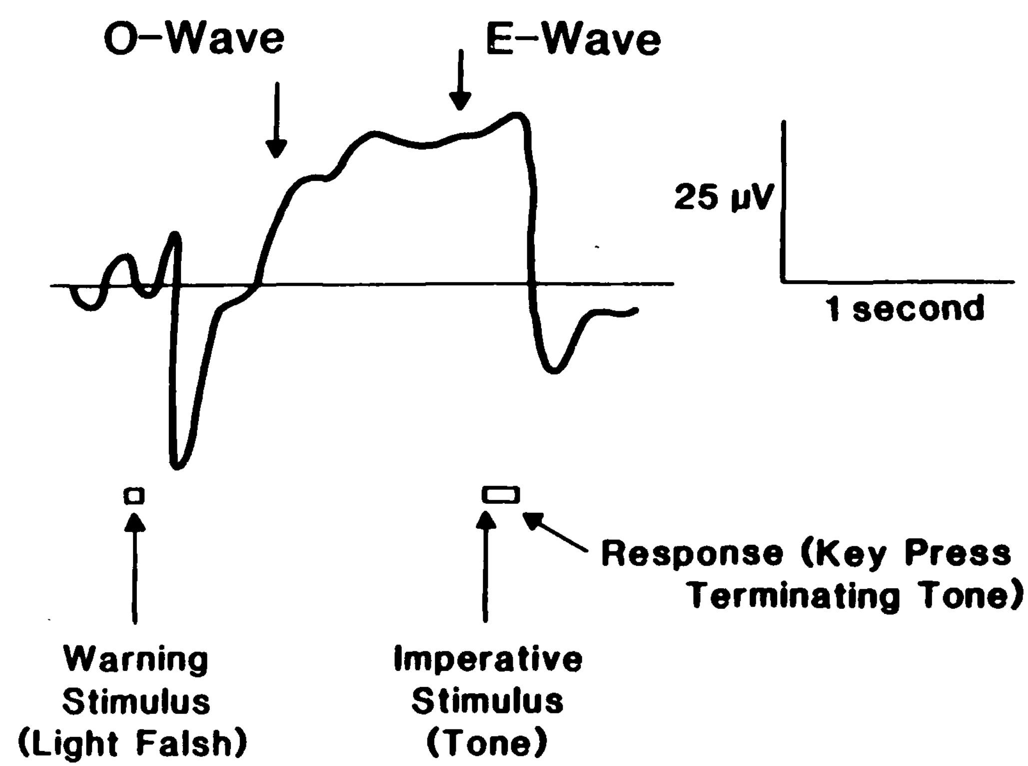

O-Wave
E-Wave
25 µV
1 second
Warning Stimulus (Light Falsh)
Imperative Stimulus (Tone)
Response (Key Press Terminating Tone)

FIGURE 12

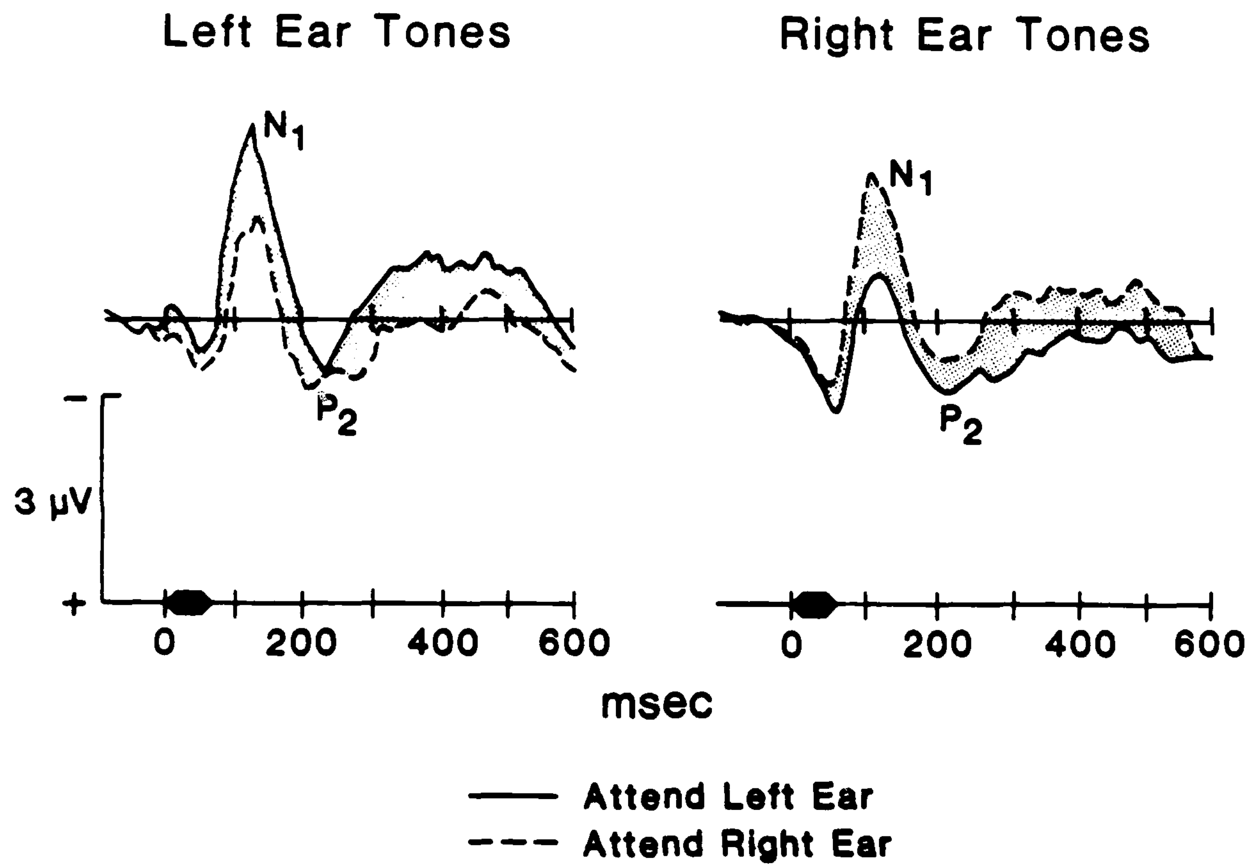

Left Ear Tones
Right Ear Tones
N1
P2
N1
P2
3 μV
−
+
0
200
400
600
0
200
400
600
msec
Attend Left Ear
Attend Right Ear

FIGURE 13

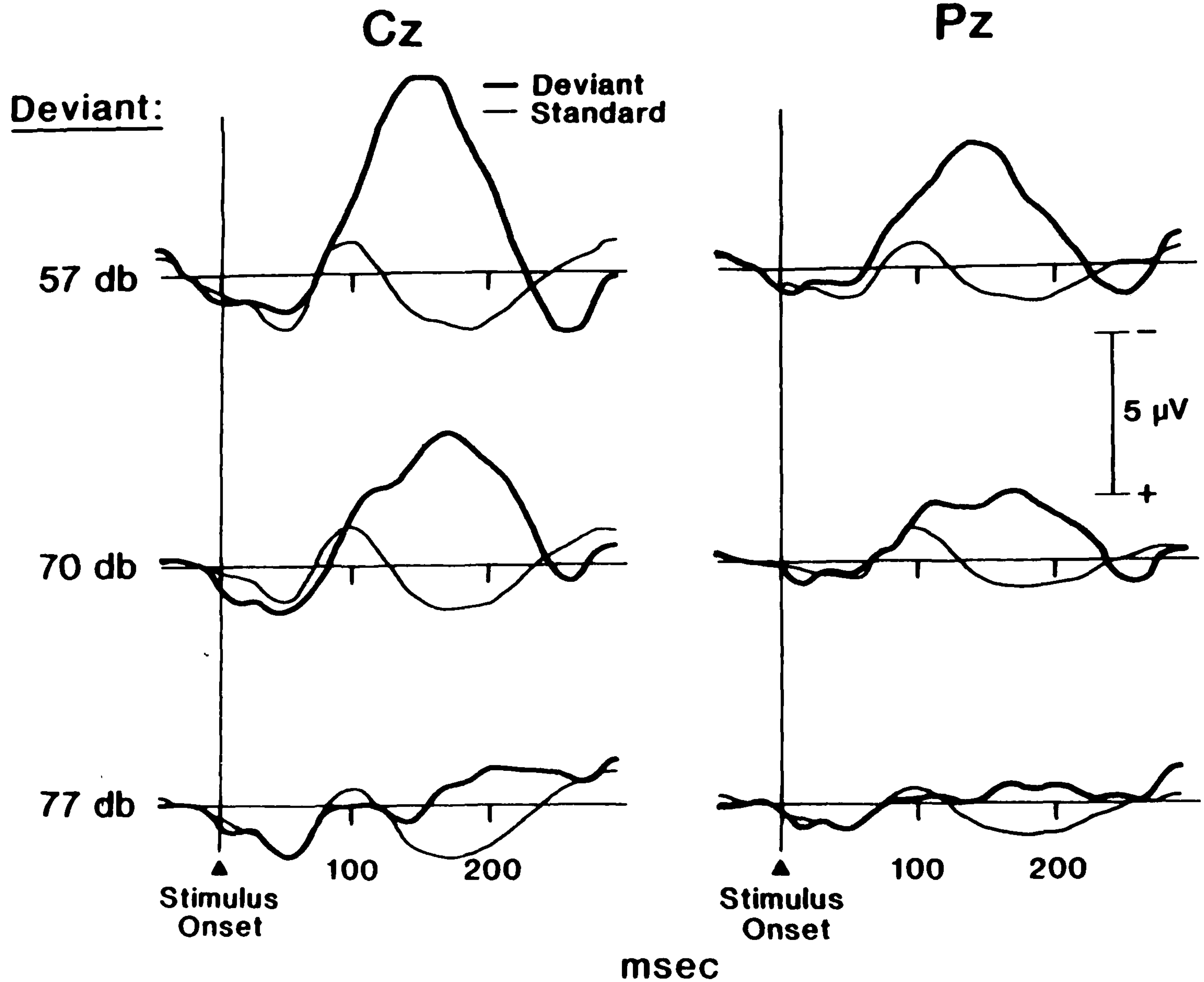

Cz
Pz
Deviant:
Deviant
Standard
57 db
70 db
77 db
5 μV
−
+
100
200
Stimulus Onset
msec

FIGURE 14

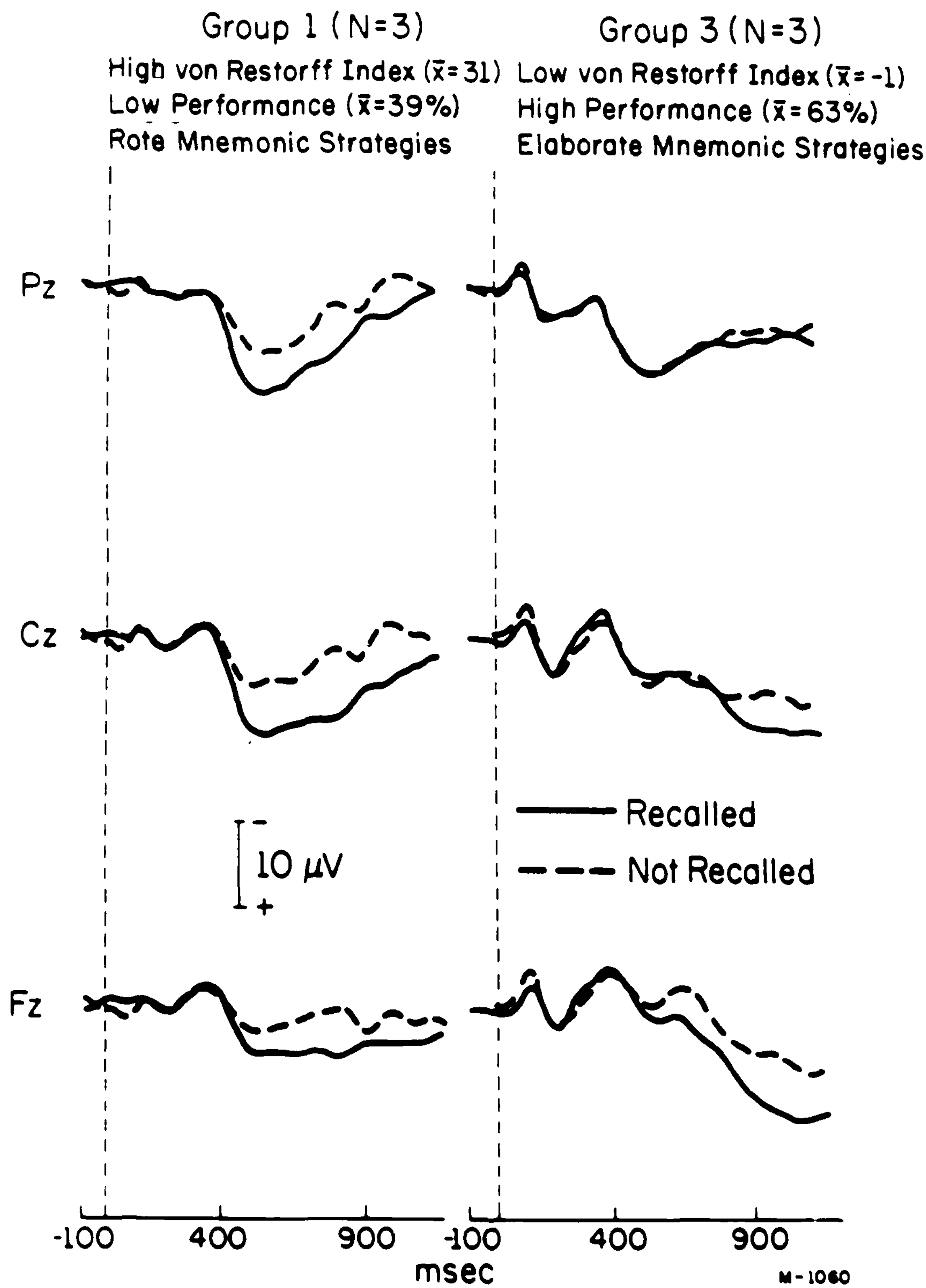

Group 1 (N=3)
High von Restorff Index (x̄=31)
Low Performance (x̄=39%)
Rote Mnemonic Strategies
Group 3 (N=3)
Low von Restorff Index (x̄=-1)
High Performance (x̄=63%)
Elaborate Mnemonic Strategies
Pz
Cz
Fz
-
10 μV
+
Recalled
Not Recalled
-100
400
900
-100
400
900
msec
M-1060

FIGURE 15

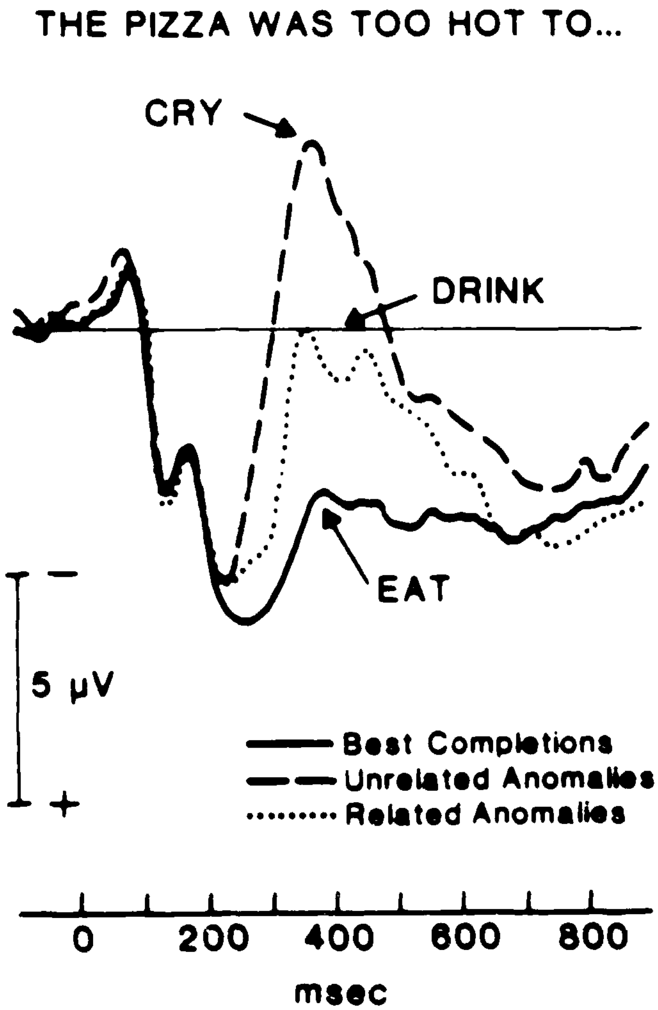

THE PIZZA WAS TOO HOT TO...
CRY
DRINK
EAT
5 µV
−
+
Best Completions
Unrelated Anomalies
Related Anomalies
0
200
400
600
800
msec

FIGURE 16

Theory in cognitive psychophysiology[1]

Michael G. H. Coles, Gabriele Gratton, & William J. Gehring
Cognitive Psychophysiology Laboratory, University of Illinois

Keywords: theory, psychophysiology, cognitive psychophysiology, event-related potentials.

I. Introduction

In this paper we shall confine our attention to the role of theory in cognitive psychophysiology. The basic assumption of the cognitive psychophysiological approach is that cognitive activity is implemented in the nervous system by means of physiological changes. Given this assumption, it is possible that a variety of psychophysiological measures may prove useful in the exploration of cognitive function. However, we shall focus our attention on measures of the event-related brain potential (ERP) (for an example of another physiological system, see van der Molen, Somsen, & Orlebeke, 1985).

We have argued previously (see, for example, Coles & Gratton, 1985, 1986; Donchin, Coles & Gratton, 1984; and see also Donchin, 198[illegible]) that cognitive psychophysiology represents an attempt to understand human cognitive function in which traditional 'cognitive' measures (reaction time, error rate, etc.) are supplemented by psychophysiological measures. In trying to understand human cognitive function we adopt a general information processing conception, in which a particular behavioral output, following some input, is attributed to the activities of various intervening processes. Given this framework, it has been useful to try to account for variability in behavioral output in terms of variability in a subset of the intervening processes. Output variability may be observed as a function of (stimulus condition), state, subject personality, age, or subject strategy vary. In seeking to account for output variability, we propose that the psychophysiological measures are sensitive to particular aspects of human information processing (the 'intervening processes' mentioned above). It is our working hypothesis that the traditional measures of behavioral output depend on a larger set of processes than do particular psychophysiological measures. Thus, psychophysiological measures should enable us to localize the source of output variability. We should emphasize that, while our strategy has been to try to account for behavioral output, our primary interest is more general—that is, we want to know how the information processing system works.

Note that in the preceding paragraph, we have explicitly made a number of theoretical assumptions. We adopted the 'computer metaphor' and assumed that human behavior, in response to a stimulus, is determined by the activities of several intervening processes. The number, nature, and mode of interaction of these intervening processes were not specified. However, different theories in cognitive psychology provide a variety of alternative, competing specifications which can be used both to guide, and be tested by, research in cognitive psychophysiology. We will now review these 'guidance' functions of cognitive theory.

II. The role of theory

In claiming that psychophysiological measures

[illegible] was supported by a grant from NIMH (MH41445) and by AFOSR contract F49620-83-0144, to Dr. Al Fregly. [illegible] is supported by a National Science Foundation Graduate Fellowship. The ideas presented in this article [illegible] discussions with our colleagues over a period of several years, and owe much to the Director of the Cognitive [illegible], Dr. Emanuel Donchin.

[illegible] can be interpreted as stages in discrete serial models of information processing (e.g. Sanders, 1980). As will [illegible] we do not necessarily wish to subscribe to a particular view of the way in which elements of the information [illegible] we prefer the more neutral term 'processes' rather than 'stages' to refer to the activities that transpire

[illegible] 603 East Daniel Street, Champaign, IL 61820, USA

[illegible]

cognitive activity. Statement [illegible] specification. Statement [illegible] specification about [illegible]

What is the basis for these theoretical statements? First, we may adopt a theory of cognitive psychology, such as the serial stage model (Sternberg, 1969; Sanders, 1980). This theory proposes that overt behavior is dependent on particular intervening information processing activities or stages. Then, we may design studies in which we choose particular manipulations that are known to affect particular information processing activities. For example, stimulus complexity affects stimulus evaluation; stimulus-response compatibility affects response execution processes. Then, we observe the effects of these manipulations on both the psychophysiological measure (P300 latency) and on overt behavior. Measures of overt behavior are taken as a manipulation check. If we observe that P300 latency is influenced by the complexity manipulation, but not by the compatibility manipulation (while reaction time is affected by both manipulations), we infer that P300 latency is related to stimulus evaluation time (Magliero, Bashore, Coles, & Donchin, 1984; McCarthy & Donchin, 1981).

[illegible] theory can be clearly discerned. First, we have a hypothetical construct that [illegible] relationship between an experimental manipulation and overt behavior. Second, we have a theoretical statement about the relationship between the hypothetical construct and the physiological measure. Third, this leads to a prediction about the relationship between the measure and overt behavior.

III. Physiological data

So far, we have seen how we arrive at theoretical statements about the functional significance of our psychophysiological measure. In a sense, we have described the validation process used to justify the ascription of psychological "meaning" to our measure. However, the measure is, after all, a measure of physiological function. How does the physiological status of the measure interact with its psychological status?[1]

First, we should point out that ascription of

[1] Space does not permit a discussion of the [illegible] between a construct and a physiological variable [illegible] the difference between manifestations and correlates, see Donchin (1979, 1981) and Heffley, Coles, and Gaillard (1986).

IV. Testing cognitive theories

[illegible] significance of an ERP [illegible] used to use measures of [illegible] alternative theories. In [illegible] application of this approach [illegible] found in the work of a number [illegible] laboratories (see, for instance, Duncan-Johnson & Kopell, 1981; Kutas & Van [illegible], in press; Muller, 1988; Naatanen, [illegible]).

In our laboratory, we have recently focussed on the continuous flow model of information processing (Eriksen & Schultz, 1979) and on the role of response priming in warned reaction time tasks. We have been able to elucidate the nature of communication between stimulus evaluation and response execution processes

V. Concluding remarks

[illegible]

[illegible] psychophysiological [illegible] such theory involves state[illegible] the relationship between [illegible] constructs and psychophysiological [illegible] In [illegible] the constructs are [illegible] from cognitive theory. In other [illegible] they may well be borrowed from other domains. However, wherever the constructs [illegible] it appears that psychophysiological theory does not exist independently of some other parent theory. To be sure, we may have concerns that our parent disciplines do not share, but these other concerns are, in turn, shared with other disciplines. Thus, our interest in the relationship between our measures and underlying neural activity is shared with biophysics and physiological psychology. Furthermore, while we may discover empirical relationships that are specific to the psychophysiological domain (e.g., "law of the initial values," Wilder, 1950), the theoretical status of these relationships derives, not from any psychophysiological theory, but from either psychological or physiological theories. For these reasons, it appears that it is the configuration of interests and methodologies that defines cognitive psychophysiology, and this configuration gives us the ability to contribute to solutions to a variety of problems in different domains.

References

[illegible] Psychophysiology [illegible] processes [illegible] S. Butler & I. Martin [illegible] New [illegible] Beck [illegible] England: Croom Helm.

[illegible] M. G. H. [illegible] (1986). Cognitive [illegible] states and processes. [illegible] R. J. Hockey, A. W. K. Gaillard, & M. G. H. Coles (Eds.), *Energetics and Human Information Processing* [illegible] Dordrecht: Martinus Nijhoff.

[illegible] M. G. H. [illegible] Bashore, T. R., Eriksen, [illegible] A psychophysiological [illegible] flow model of human [illegible] *Journal of Experimental* [illegible] *and Performance* [illegible]

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

[illegible] E. [illegible] Surprise! ... Surprise? *Psychophysiology* [illegible]

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.

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

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

Fabiani, M., Karis, D., & Donchin, E. (1986). P300 and recall in an incidental memory paradigm. *Psychophysiology, 23*, 298–308.

Gehring, W. J., Gratton, G., Coles, M. G. H., & Donchin, E. (1986). Response priming and components of the event-related brain potential. *Psychophysiology, 23*, 437–438.

Gratton, G., Coles, M. G. H., Sirevaag, E., Eriksen, C. W., & Donchin, E. (1987). Pre- and post-stimulus priming of response channels. In preparation.

Hockey, G. R. J., Coles, M. G. H., & Gaillard, A. W. K. (1986). Energetical issues in research on human information processing. In G. R. J. Hockey, A. W. K. Gaillard, & M. G. H. Coles (Eds.), *Energetics and Human Information Processing* (pp. 3–21). Dordrecht: Martinus Nijhoff.

Kutas, M., & Van Petten, C. (in press). Event-related brain potential studies of language. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in Psychophysiology, Volume 3*. Greenwich: JAI press.

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.

Miller, G. A. (1986). Information processing deficits in anhedonia and perceptual aberration: A psychophysiological analysis. *Biological Psychiatry, 21*, 100–115.

Naatanen, R. (1982). Processing negativity: An evoked potential reflection of selective attention. *Psychological Bulletin, 92*, 605–640.

Sanders, A. F. (1980). Stage analysis of reaction processes. In G. E. Stelmach, & J. Requin (Eds.), *Tutorials in Motor Behavior*. Amsterdam: North-Holland.

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.

van der Molen, M. W., Somsen, R. J. M., & Orlebeke, J. F. (1985). The rhythm of the heart beat in information processing. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in Psychophysiology, Volume 1*. Greenwich: JAI press.

Wilder, J. (1950). The law of initial values. *Psychosomatic Medicine, 12*, 392–399.

Psychophysiology and the Electronic Workplace
Edited by A. Gale and B. Christie

Article 4.

12

Heart Rate and Sinus Arrhythmia

MICHAEL G.H. COLES AND ERIK SIREVAAG

INTRODUCTION

Heart rate (HR) refers, literally, to the rate at which the heart beats; while sinus arrhythmia refers to variability in heart rate. Measures of HR and sinus arrhythmia have been, or could be, used in several ways to explore the *interaction between* the operator and the machine in the electronic workplace. First, these measures can provide insights into psychological processes such as attention, arousal, stress and mental workload — insights that are not always available from measures of overt behavior (performance) or subjective report (compare the ERP approach: see Kramer, Chapter 9 of this volume; Coles and Gratton, 1985). Second, since the cardiovascular system is intimately involved in the provision of energy for physical work, measures of its function can provide insights into physical workload (see, for example, Kalsbeek and Ettema, 1963; Vogt *et al.* 1973). Third, because of their relationship to pathophysiological conditions such as hypertension and coronary heart disease, the measures may aid in the identification of those characteristics of the electronic workplace that might lead to pathophysiology (Obrist, 1984).

We begin by considering the relationship between the two measures and the physiological systems that underlie them. Next, we discuss measurement procedures and various measurement related problems. Then, we review the theoretical approaches that have guided the interpretation of the measures. This is followed by a review of specific research studies involving the measurement of HR and/or sinus arrhythmia in work environments. Finally, we consider why much of the research in this area has not been particularly fruitful and point to strategies that may prove to be more useful in understanding operator–machine interactions.

PHYSIOLOGICAL BASES OF THE MEASURES

The human heart contracts approximately once every second. The goal of this tireless activity is to maintain a steady supply of oxygen and nutrients to the cells of the body, as well as to remove the waste products resulting from metabolic activity.

Structurally, the heart is subdivided into four chambers: two ventricles and two atria. Each individual heartbeat actually consists of a series of four contractions. Venous blood returns to the heart through the right atrium. The first contraction pumps the blood into the right ventricle. Here, the second contraction forces the blood out of the heart to the lungs where carbon dioxide is removed and replaced with oxygen. The arterial blood re-enters the heart through the left atrium and is pumped into the left ventricle. The final contraction in the sequence expels the blood via the aorta to the body. These events produce the characteristic pattern of peaks and troughs present in measures of the electrocardiogram (EKG or ECG, see below).

Most structures of the autonomic nervous system (ANS) are under joint control of both the sympathetic (SNS) and the parasympathetic (PNS) nervous systems. The heart is no exception. Activation of the SNS increases the firing rate of the pacemaker producing an increase in HR. Catecholamines released from sympathetic nerve endings are the transmitters that chemically mediate this process.

Parasympathetic (vagal) nerve endings also innervate the cardiac pacemaker. The release of acetylcholine from these cells inhibits the firing of the pacemaker, producing a decrease in HR. Thus, changes in HR can result from changes in either the PNS, the SNS, or both. This makes it difficult to calculate the individual contributions of either the SNS or the PNS to overall patterns of HR lability. One technique to overcome this problem is to inject drugs which selectively block activity in either the PNS or the SNS. Another, non-invasive method capitalizes on the fact that some aspects of cardiovascular activity (other than HR) are influenced by only one branch of the ANS. For example, the influence of respiratory activity on HR (respiratory sinus arrhythmia) is controlled by a gating mechanism which decreases the vagal influence on the heart during inspiration.

From the foregoing, it is evident that HR is multiply determined in the sense that it is influenced by both branches of the autonomic nervous system. It is also multiply determined in the sense that a variety of homeostatic and psychological factors can influence the rate at which the heart beats. These factors exert their influence through both sympathetic and parasympathetic mechanisms.

Homeostatic mechanisms affecting HR include baroreceptors and chemoreceptors located primarily in the carotid sinus and aortic arch. These receptors detect changes in blood pressure and blood gas composition respectively. They transmit ascending *information to regions of the brainstem and hypothalamus* which control cardiac function. As a result, HR is reflexively controlled as a function of blood pressure and blood gas changes. The changes in HR that are

associated with more psychologically relevant activity are presumed to result from higher-order control initiated by certain cortical and forebrain structures. This control, in some cases, emanates from the motor cortex and is related to the metabolic demands being placed upon the cardiovascular system. However, cortical effects on HR independent of somatic activity have also been obtained.

PROBLEMS OF MEASUREMENT AND SAMPLING

Most modern methods of recording HR involve the detection of the electrical activity of the heart muscles that contract during a heart beat. When electrodes are placed at appropriate locations on the body and attached to amplifiers and a polygraph, the resulting signal conforms to the traditional EKG or ECG in which P, QRS, and T wave components are observed. These components correspond to depolarization of the atrial muscles, depolarization of the ventricles, and repolarization of the ventricles respectively. When psychophysiologists consider HR, they generally focus on the rate at which R waves occur.

To derive measures of rate from the ECG signal, two methods are commonly used. First, the ECG can be fed to a rate measuring device that converts the intervals between R waves to an analog signal whose value depends on the reciprocal of the R-R interval. This analog signal can be displayed on a polygraph and its value assessed by hand-scoring techniques. [illegible]

[illegible]

In the frequency domain, one is interested in describing the variability of HR over a particular time period. Using Fourier transform procedures and other complex statistical procedures (see Mulder and Mulder Hajonides van der Meulen, 1973; Porges, 1984), the variability can be described in terms of a power frequency spectrum. This is the way in which sinus arrhythmias can be assessed. For example, Porges (1984) identifies that component of the frequency spectrum corresponding to respiratory frequency (0.12–0.40 Hz) as defining respiratory sinus arrhythmia, and he argues that the power at this frequency [illegible] a measure of the vagal influence on the heart.

It might be thought that the concept of respiratory sinus arrhythmia [illegible] coupling between cardiac and respiratory activity and that this should be assessed more directly than by analysis of the HR power spectrum alone. Indeed, at one time, Porges considered the measure of coherence between respiratory and cardiac activity to be the best measure of respiratory sinus arrhythmia. However, because of the variable lag between respiration and HR this measure proved to be less satisfactory than the simpler [illegible]. The other major frequency bands in the HR spectrum include the following: [illegible] the 'fundamental physiological frequency' (Loos, 1968) that may be related to spontaneous fluctuations in blood pressure (Mulder and Mulder Hajonides van der Meulen, 1973; Luczak and Laurig, 1973) and [illegible] frequency band that may be associated with thermoregulatory processes [illegible] Tam, and Calaresu, 1975).

Although this frequency decomposition is the preferable procedure, [illegible] in the assessment of variability have been employed. For example, the [illegible] deviation or range of HR over a particular time period have both been [illegible] quantify variability. While the latter measures have the advantage of [illegible] and relatively easy to obtain, they have the obvious disadvantage [illegible] discriminate among different sources of variability.

In measuring HR and HR variability, several problems [illegible]

What units should be used?

As we mentioned above, at the level of a single observation [illegible] HR and IBI are reciprocally related. However [illegible] mean, change score, or some other statistic to describe [illegible] tion, the units are not simply related (see Jennings [illegible] 1974). For example, a change in HR from [illegible] equivalent to a decrease in IBI of 143 ms. [illegible] from 120 to 130 bpm [illegible]

Should the data be expressed in real time or biological time?

Since the frequency with which [illegible] value of the measure [illegible]

numbers of observations for a given time period [illegible] This creates a particular problem in the analysis of phasic [illegible] most do that the course of a phasic response to [illegible] rather than the number of heart beats. [illegible] the investigator typically derives [illegible]

What should be done about individual differences?

[illegible]

How should artifacts be treated?

[illegible]

Arousal

the psychophysiology of human performance has been considered. The theory's origins can be traced to the work of Duffy (195[illegible]), Lindsley [illegible] [illegible] and Malmo (1959). The theory proposes, among other things, that there is a dimension of arousal on which various psychological states can be [illegible]. Thus, as arousal increases, the individual moves from sleep [illegible] alertness, to frantic excitement. At the high end of the dimension [illegible] can be considered to be "under stress." The relationship between [illegible] performance is described by an inverted U, such that [illegible] which performance is optimal lies somewhere between [illegible] dimension. Thus, the performance decrements that accompany [illegible] are understood in terms of a shift in arousal toward the [illegible] dimension. On the other hand, performance decrements [illegible] under stress conditions (noise, etc.) are understood in terms [illegible] arousal.

Easterbrook (1959) has attempted to account for these effects [illegible] performance in terms of a change in the range of stimuli utilized by [illegible] the performance of their task. At low arousal levels, subjects [illegible] enough; at high levels, they are overselective. While the [illegible] to account for some of the observed relationships between arousal and performance, it cannot handle all the data (Kahneman, 1973).

One of the central problems with arousal theory involves the [illegible] independent measure of arousal such that an individual's [illegible] inverted U function can be unequivocally determined. Otherwise [illegible] in performance that result from a manipulation of arousal [illegible] there is improvement, the manipulation can be said to move [illegible] below the optimal level towards it; if there is no change, the [illegible] said to move the individual from one side of the optimal [illegible] other hand, if there is a decrement, the manipulation [illegible] individual to a level of arousal that is higher than the optimum [illegible] 1960s, it appeared that psychophysiological [illegible] the critical independent assessment of arousal [illegible] another, almost all psychophysiological measures [illegible] arousal measures (for example, skin conductance, the [illegible] the electromyogram, HR, blood pressure, respiration [illegible] reasons why this approach seems to have failed. First, the [illegible] determined and are thus not pure measures of [illegible] has argued, these psychophysiological measures [illegible] them all to be regarded [illegible] Furthermore, under some conditions [illegible] conductance appear to be actually dissociated [illegible] Lacey, 1967). From this viewpoint, the [illegible] arousal level will depend on the [illegible] satisfactory situation and [illegible] turn later.

Motivation

[illegible]

Intake rejection

[illegible]

cycle changes can be subsumed under one hypothesis (for example, Coles, 1984). Second, the kinds of HR changes observed by the Laceys are clearly independent of gross changes in motor activity (see, for example, Brunia, 1984). Thus, the HR effects cannot be discounted as mere byproducts of the changes in motor activity, an association that would be expected because of the role of the cardiovascular system in muscular work. Third, research has begun to explore the interrelationship between cardiac and cortical events predicted by the afferent feedback hypothesis (see, for example, Sandman, 1984). Fourth, some of the research stemming from the Laceys' work has direct relevance for the electronic workplace. For example, Coles (1972) found that tonic HR differentiated between two visual search tasks requiring detection of items with different degrees of discriminability. Since visual search is a common activity in offices, it may be that HR could be used as an index of attention in this setting. Fifth, there has been an explosion of research in a related area, orienting response (OR) theory, to which we now turn.

Orienting response theory

The orienting response (OR) was considered by Pavlov (1927) to be a complex of behavioral and physiological responses to discrete events. The primary function of the OR was to heighten the organism's sensitivity to environmental events (see also Sokolov, 1963). In this sense, there appears to be a relationship between the concept of orienting and *that of intake. Indeed,* it was this relationship that led Graham and Clifton (1966) to attempt a rapprochement between orienting theory and the intake-rejection hypothesis. Although Sokolov (1963) had proposed that HR increased during orienting, Graham and Clifton argued, based on an extensive review of the literature, that the cardiac component of orienting was deceleration. Thus, the same kind of HR response, deceleration, was seen as an aspect of orienting and environmental intake.

It is important to note that, in general, OR research has focused on phasic responses to non-signal events (events with no relevance to the subject's task), and the principal criterion for a response to be classified as an OR was that it habituated with stimulus repetition (see, for example, Graham, 1984): that is, as the same physical stimulus is repeated, so the response to that stimulus decreases. OR researchers distinguish between adaptation and habituation. Following habituation, the OR will be reinstated if a different stimulus is presented, and subsequent re-presentation of the original stimulus will again evoke the OR. These observations indicate that the decrement in responsiveness is due to an active process of non-responding rather than to a passive fatigue-like process. Thus, it must be concluded that the organism builds some sort of memory representation of the repeated stimulus, against which all incoming stimuli are compared (Siddle, 1983).

These characteristics of the OR (a decelerative HR response that habituates [illegible]

with stimulus repetition) are usually contrasted with those of the defensive response (DR). The DR occurs to intense stimuli and can be characterized by an accelerative response that does not habituate. The functional significance of the DR is to reduce the sensitivity of the organism to environmental events (compare [illegible] Sokolov, 1963).

[illegible] OR research was extended to the [illegible] (Siddle, 1983) [illegible] OR [illegible] within the framework of contemporary [illegible]. Thus, we find attempts to identify the OR with the [illegible] (Ohman, 1979) [illegible] of input processing (Graham, 1984) [illegible] that make up the electronic workplace are likely to be [illegible] contemporary cognitive and information processing concepts [illegible] in OR research seems more likely to be of relevance to [illegible] than that based on the Pavlovian tradition (see also Spinks and [illegible], 1983). Similarly, to the extent that it is grounded in contemporary [illegible] psychology theory, a psychophysiological analysis of workload is [illegible]. We shall now review this area.

Workload: physical and cognitive

[illegible]

Bobrow, 1975). The allocation of this limited processing commodity (whatever it is) to the performance of a given task is determined by the motivation, skill, and strategy of the operator as well as by the demands of the task. The latter can, in turn, be manipulated either by changing characteristics of the task (such as reducing stimulus discriminability, or changing the pacing of the task for example), or by varying the minimal level of acceptable performance on the task.

An important distinction with respect to the concept of workload is that between automatic and controlled processing (Shiffrin and Schneider, 1977). It has been proposed that controlled tasks require the allocation of resources that are limited, while automatic tasks do not require such resources. On the basis of this distinction, it can be seen that the idea of mental workload is more applicable to tasks involving controlled processing. In fact, this concept underlies much of Mulder's (1980) work relating his measure of sinus arrhythmia to workload.

The above conceptualization of cognitive workload emphasizes the informational transactions presumed to underlie cognitive events. Sanders (1979) distinguishes this type of definition from those in which factors such as emotions, tensions, and frustrations are allowed to contribute to the mental load a given situation imposes on the operator. Of course in situations where such affective variables remain constant across the tasks of interest, the two definitions are equivalent. However, because HR is multiply determined, any measure of mental workload obtained through an examination of the heart will most likely be sensitive to both cognitive as well as emotional factors (see the ERP approach by Kramer, Chapter 9 of this volume, for a discussion of a measure sensitive solely to the information-processing demands of a task). The coupling of affect and cognition may not always be undesirable, however, for in the electronic workplace job-related stress is often an important factor affecting performance (Hockey, Gaillard, and Coles, 1986). In such situations, HR may provide a valuable tool to aid in the estimation of the total load placed upon an operator by the work environment.

Given that mental workload is difficult to define, it should not be surprising that it is even more difficult to quantify. Measures of task performance are not suitable workload metrics for at least two reasons: they cannot be generalized across tasks requiring different performance measures, and increased task difficulty is often not reflected in single task performance decrements, even though operators indicate subjectively that task difficulty changes. For this reason, several alternative techniques have been proposed for workload assessment. One approach stresses the advantages of dual task paradigms (Knowles, 1963; Rolfe, 1971; Brown, 1978). In these studies, increased workload on a primary task is inferred from changes in the performance of a concurrent secondary task. Thus, because resources are limited, secondary task performance decrements are presumed to result from the drain on resources entailed by increases in the workload of the primary task.

An alternative approach is to examine physiological measures in the context of

variations in mental workload (see, for example, Berlyne, 1960; Howitt, 1968; Roscoe, 1978; Wierwille, 1979). In general, these measures are presumed to reflect overall levels of arousal. However, because the cardiovascular system may be capable of modulating cortical activity (see above for a discussion of the afferent feedback hypothesis of the Laceys), several studies (reviewed below) have examined changes in the HR in the hope that they can be related more directly than simple measures of arousal to the information-processing activities implemented by the CNS.

Active passive coping

Over the past 20 years or so, Paul Obrist and his colleagues (see review by Obrist, 1984) have been pursuing a line of research that is based on a somewhat different theoretical perspective from those described previously. They have been particularly concerned with individual differences in cardiovascular reactivity and their relationship to pathophysiology.

Following the lead of the Laceys (Lacey and Lacey, 1962), Obrist has explored the notion of individual response stereotypy, that is, over a wide range of situations, individuals are consistent in their patterns of autonomic responding. Thus, the cardiovascular systems of some individuals are always overreactive. This pattern is exacerbated if the eliciting conditions are those evoking sympathetic activation in all individuals. One example of such a situation is that of active coping, which, for Obrist and his colleagues (1978), requires the subject to make a behavioral response to avoid painful electric shocks.

In pursuing these findings, Obrist attempted to determine whether individuals who show high HR reactivity, especially under conditions of active coping, are at risk for developing the pathophysiological condition of hypertension. With his colleagues, he found an association between a known risk factor (family history of hypertension) and hyperreactivity (Hastrup, Light, and Obrist, 1982).

What are the implications of Obrist's work for the electronic workplace? Clearly, the work environment provides a setting in which individuals are exposed to stressful conditions with which they must cope. Some of these stressors may lead to excessive sympathetic reactivity in all individuals, but particularly those who tend to be overactive sympathetically regardless of the situation, that is, they have a stereotypic profile of high sympathetic reactivity. For these individuals, the workplace may be literally unhealthy. Thus, one could foresee a situation in which individuals would be screened for overreactivity before being placed in high stress environments.

While more research is needed to prove the link between hyperreactivity, active coping, and hypertension, Obrist's work clearly has relevance to those interested in evaluating the consequences of particular types of stress, and in selecting individuals who can handle particular stressors without developing pathophysiological conditions.

A SELECTED REVIEW OF KEY RESEARCH PAPERS

In this section, we review studies involving the use of measures of HR and sinus arrhythmia in analyzing the effects of different facets of the work environment on the operator.

Physical workload

In a pioneering set of studies, Kalsbeek and Ettema (reviewed in Kalsbeek, 1969) demonstrated the sensitivity of measures of HR and sinus arrhythmia to increases in physical workload. In these studies, mean HR increased and sinus arrhythmia decreased with increases in physical load. This pattern of results occurred when baseline values collected during a rest condition were compared with a condition involving dynamic work, as well as a condition requiring only static work (for example, passively holding a weight with an outstretched arm). Similarly, Vogt et al. (1973), found that cardiac activity changed reliably with increased physical workload defined in terms of both increased muscular demands as well as increased temperature stress. Without question, then, HR and sinus arrhythmia have emerged as valid indicators of physical workload.

Cognitive workload

The attempt to gauge increased cognitive load by the examination of cardiac activity has, unfortunately, proven more problematic than the estimation of physical workload. To investigate the sensitivity of cardiac measures to increased cognitive load, Kalsbeek and Ettema (1963, 1965) developed an auditory discrimination task in which signals were presented at either 60 or 90 per cent of the maximum speed at which individual subjects could perform the task without error. This manipulation of the baud rate of stimulus presentation did not alter mean HR, but a significant effect on sinus arrhythmia was obtained. Thus, sinus arrhythmia scores were lower in the 90 per cent condition than the 60 per cent and both conditions were associated with lower HR variability than during a rest period.

A number of subsequent studies have attempted to confirm the sensitivity of cardiac measures to workload manipulations both in the laboratory and in more applied settings. The vast majority of these studies compare tonic measures of cardiac activity collected during conditions presumed to differ along some difficulty dimension with a baseline established during a rest condition.

An experiment by Mulder and Mulder-Hajonides van der Meulen (1973) is a case in point. The purpose of this experiment was to evaluate several techniques for estimating sinus arrhythmia in an auditory discrimination task similar to the Kalsbeek and Ettema experiment described above. In the Mulders' task, five

different rates of stimulus presentation were employed (20, 30, 40, 50, and 60 binary choices/min). While all five conditions displayed significantly lower sinus arrhythmia scores when compared with rest conditions, the most reliable measure of sinus arrhythmia discriminated only between the most extreme conditions (20 choices/min versus 60 choices/min).

Of course, these presentation-rate studies confound physical and cognitive workload, for increased baud rates require an increase in responses/min as well as decisions/min. Mulder (1980) recognized this problem and has subsequently conducted an extensive series of studies in which cognitive demands were varied while physical demands remained constant. In these studies Mulder focussed on the 0.06–0.14 Hz frequency band of the cardiac interval signal. This frequency band is relatively insensitive to changes in the rate or pattern of a subject's respiratory activity, and is therefore, more likely to be related to changes in blood pressure control mechanisms.

Mulder claims that the amplitude of a spontaneously oscillating component within this frequency band (0.10 Hz) is sensitive to increased task complexity. This claim is supported by research showing that the 0.10 Hz component changes with cognitive workload in conditions in which contrasts have been made between divided and focussed attention, the number of comparisons in a semantic memory task, and differential levels of practice. Mulder concludes that most of the variance in the oscillations of the 0.10 Hz component can be explained by two factors: habituation to the environment, and the amount of controlled (as opposed to automatic) processing demanded by the task.

Self-paced versus machine-paced work

A research topic related to the question of stimulus presentation rate discussed above also serves to illustrate some of the problems of using measures of cardiac activity as metrics of cognitive load. An important question in many applied settings concerns the costs and benefits of self-paced versus machine-paced work. In other words, will operators find tasks that are externally paced (such as most assembly line tasks) more or less fatiguing than tasks that allow them to set their own pace? The question seems simple enough at first glance, however, an examination of attempts to resolve the issue by the analysis of cardiac activity points out some of the major problems currently plaguing the attempt to relate the output of the heart with the cognitive load.

A study of Manenica (1977) concluded that self-paced work was more demanding than machine-paced work. In this experiment, subjects were trained to perform a simple light repetitive assembly task. Stimuli were presented along a moving conveyor belt. In the machine-paced conditions, subjects were required to pick up the stimulus (a roller bearing), process it and replace it on the conveyor belt before the stimulus left a tolerance zone. The self-paced condition [illegible] identical except that stimuli were allowed to stack up in a position in the middle of the tolerance zone. The rate of stimulus presentation during the machine-paced condition was adjusted so that it equalled the subject's mean output rate under self-paced conditions. Thus, the output rate for the two conditions was held constant. The particular method of measuring cardiac activity employed by this study centered around the analysis of the time between successive R-R intervals. Three statistics were used to analyze the cardiac data: the mean, the standard deviation, and D (a measure of consistency between adjacent R-R intervals). In the self-paced condition, HR was higher, less variable, and more consistent. Manenica claimed that an internal pacing mechanism operated in the self-paced condition, and that the increased load due to self-pacing results from the requirement to maintain this internal pacemaker.

In contrast, a study by Sharit, Salvendy, and Deisenroth (1982; see also Sharit and Salvendy, 1982) found no evidence for an internal pacing mechanism during self-paced work. Instead, they claim that uncertainty factors during machine-paced work produced greater mental load than that during self-paced work. In this study, two different conditions were developed in an attempt to build on the work of the Laceys described above. Thus, both machine-paced and self-paced presentation modes were employed during a visual detection task, and during a visual task requiring mental arithmetic. On the basis of the Laceys' work the former task was classified as 'external' and the latter as 'internal'. These labels supposedly defined the direction of attention. In accordance with the intake-rejection hypothesis, the experimenters expected to find cardiac decelerations associated with the external task, and accelerations associated with the internal task.

The data were examined to determine whether an effect of stimulus presentation mode upon the pattern of cardiac activity was present in each type of task. A measure termed the 'degree of change in cardiac response' was derived. This measure represented the average over all trials of the difference in milliseconds (ms) between two designated IBIs from each trial. The data were interpreted as indicating a greater degree of cardiac deceleration for self-paced as opposed to machine-paced work in both the external and the internal condition. The smaller decelerations for the machine-paced task were interpreted by these investigators as evidence that machine-paced work involves increased cognitive (internal) processing.

These two studies provide a good example of just how difficult it is to resolve conflicting results based on cardiac data. The paradigms used to answer basically the same question are drastically different. In Manenica's experiment, a task designed to mimic conditions in many applied settings (that is, the assembly line) was employed. In the latter study, two different tasks were chosen on the basis of a large body of psychophysiological studies that have followed the work of the Laceys. Thus, one study appears to have a great deal of face validity enabling the results to be extrapolated beyond the laboratory, while the other study can cite a large body of converging evidence obtained in other laboratories to aid in the interpretation of the data. In addition to employing vastly different paradigms, different techniques were used to evaluate cardiac activity. This problem is a pervasive one in the field. Thus, an evaluation of the research in the area of pacing is complicated by the fact that different measures, collected during different tasks, produce different patterns of results.

A study by Salvendy and Humphreys (1979) further complicates the issue. In this experiment, cardiac measures failed to discriminate between machine-paced and self-paced work during a marking–stapling task. These authors did find, however, that subject preference and performance with respect to the two schedules were related to a variety of individual difference variables. Thus, subjects identified as introverted, reserved and trusting preferred machine-paced work, while those identified as extraverted, outgoing, and suspicious preferred the self-paced condition. *It appears that age should also be considered as a factor when one considers subjects' preferences.* For example, Salvendy and Pilitsis (1971) found that older subjects preferred machine-paced conditions, while the opposite was true for younger workers. Thus, even if conflicting results concerning the cognitive load entailed by different modes of pacing had not been obtained by different investigators, there are data indicating that subjective workload (at least) may be determined more by individual differences in age and personality than by external pacing constraints *per se*.

Workload in applied settings

As we have seen, experimenters appear to have encountered great difficulty [illegible] using cardiac activity to assess cognitive load. However, they have been more successful when examining settings involving both physical and cognitive workload. For example, in a study of [illegible] (1986), HR reliably distinguished between takeoff, level flight, and landing phases of a flight task [illegible]. Another set of studies has related increased HR [illegible] to risk [illegible] as well as physical danger and [illegible] task [illegible] (Roscoe [illegible]; Roman [illegible] 1965; [illegible]). [illegible]

RECOMMENDATIONS

The preceding selective review reveals that despite some promising results measures of cardiac activity have not, as yet, proved to be particularly useful in elucidating the interaction between the operator and the work environment. This is particularly true of research dealing with cognitive aspects of work related tasks. It may be less true of physical workload. In this case, there are strong *a priori* reasons for expecting cardiovascular measures to be useful, and indeed the empirical research supports this contention. Many of the failures in the cognitive realm can be attributed to several factors, and these may be remedied in future research and practice.

Recommendations for research

(a) It has often been difficult to generalize knowledge of the functional significance of HR, derived from laboratory settings, to the electronic workplace. This is, at least in part, the responsibility of any basic research effort. In the future, we recommend that investigators consider the potential application of their research when designing their studies. For example, if more 'veridical' situations are used in the laboratory, and if more attention is paid to analyzing complex task situations into their constituent components, then there is a greater probability of transfer of research knowledge.

(b) Research of otherwise high quality has shown that measures of HR and sinus arrhythmia differentiate between rest and task conditions, but are not particularly sensitive to variations within conditions. More attention needs to be paid to measures of sinus arrhythmia, particularly the 0.10 Hz component, as this seems to be one of the more promising measures. For example, we need to know precisely the type of workload to which the measure is sensitive. Investigators should also take great care when designing their studies to ensure that physical and cognitive workload are not confounded.

(c) Research should take account of the existence of individual differences, both in terms of demographic and personality variables, and in terms of physiological factors. We need to know more about the interaction between individual characteristics and responses to particular types of stressors.

Recommendations for practice

(a) Practitioners must perform a careful analysis of the task so that [illegible] critical components can be examined in isolation.

(b) Once the task analysis has been performed, practitioners [illegible]

standard measurement procedures derived from laboratory studies, appreciate the multiple determinants of the cardiovascular system and translate this appreciation into action by considering the relevance of the various theoretical viewpoints to their particular measures in their specific situation.

REFERENCES

[illegible]

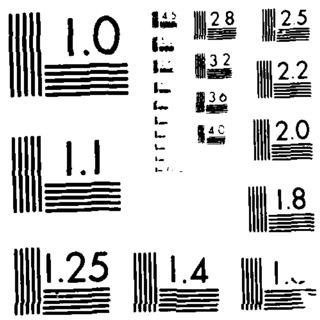
Graham, F.K. and Clifton, R.K. (1966) Heart rate change as a component [illegible] response. *Psychological Bulletin*, **65**, 305–320.
Gray, J.A. (1975) *Elements of a Two-process Theory of Learning*. New York [illegible] Press.
Hashimoto, K. (1964) The application of heart rate (electrocardiac [illegible]) ergonomic research — estimation of the driver's workload [illegible] the new Tokaido line in Japan. *Internal Report, Laboratory* [illegible] *Railway Labor Science, Japanese National Railways*.
Hastrup, J.L., Light, K.C. and Obrist, P.A. (1982) [illegible] diovascular response to stress in healthy young adults. [illegible] **19**, [illegible] 622.
Hebb, D.O. (1955) Drives and the C.N.S. (conceptual nervous system). [illegible] *Review*, **62**, 243–254.
Hockey, G.R.J., Gaillard, A.W.K. and Coles, M.G.H. (1986) [illegible] *Information Processing*. Dordrecht: Martinus Nijhoff.
Howitt, J.S. (1968) Flight deck workload studies in civil transport aircraft. [illegible] *Conference Proceedings*, **56**, 1–2.
Howitt, J.S., Balkwill, J.S., Whiteside, T.C.D. and Whittingham, P.D.G. [illegible] preliminary study of flight deck workloads in civil air transport aircraft. [illegible] *Research Committee Report No. 1240*.
Jennings, J.R., Stringfellow, J.C. and Graham, M. (1974) A comparison of [illegible] distributions of beat-by-beat heart rate and heart period. *Psychophysiology*, **11**, 207–210.
Kahneman, D. (1973) *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.
Kalsbeek, J.W.H. (1969) Sinus arrhythmia and the dual task method in [illegible] load. In W.T. Singleton, J.G. Fox, and D. Whitfield (eds) *Measurement of* [illegible] *Work: An Appraisal of Physiological and Psychological Criteria in Man-Machine Systems*, pp. 101–113. London: Taylor and Francis.
Kalsbeek, J.W.H. and Ettema, J.H. (1963) Scored regularity of the heart rate pattern [illegible] the measurement of perceptual and mental load. *Ergonomics*, **6**, 306.
Kalsbeek, J.W.H. and Ettema, J.H. (1965) Sinus arrhythmia and the measurement of mental load. *Communication at the London Conference of the British Psychological Society*, December.
Kimmel, D., van Olst, E.H., and Orlebeke, J.F. (eds) (1979) *The Orienting Reflex in Humans*. Hillsdale, NJ: Erlbaum.
Knowles, W.B. (1963) Operator loading tasks. *Human Factors*, **5**, 155–161.
Lacey, J.I. (1967) Somatic response patterning and stress: some revisions of activation theory. In M.H. Appley and R. Trumbull (eds) *Psychological Stress: Issues in Research*, pp. 14–37. New York: Appleton-Century-Crofts.
Lacey, J.I., Kagan, J., Lacey, B.C., and Moss, H.A. (1963) The visceral level: situational determinants and behavioural correlates of autonomic response patterns. In P.H. Knapp (ed.) *Expression of Emotions in Man*. New York: International Universities Press.
Lacey, J.I. and Lacey, B.C. (1962) The law of initial value in the longitudinal study of autonomic constitution: reproducibility of autonomic responses and response patterns over a four-year interval. *Annals of the New York Academy of Sciences*, **98**, 1257–1290, 1322–1326.
Lacey, J.I. and Lacey, B.C. (1978) Two-way communication between the heart and the brain: Significance of time within the cardiac cycle. In E. Meyer, and J. Brady (eds) *Research in the Psychobiology of Human Behavior*, pp. 59–73. Baltimore, MD: Johns Hopkins University Press.

[illegible]

Rottell Smith, H.P. (1967) The effect of stress on civilian transport pilots. [illegible] (ed.) *Report of the 7th Conference for Aviation Psychology* [illegible] European Association for Aviation Psychology.
Salvendy, G. and Humphreys, A.P. (1979) Effects of personality, perceptual difference and pacing of a task on productivity, job satisfaction and physiological stress. *Perceptual and Motor Skills*, **49**, 219–222.
Salvendy, G. and Pilitsis, J. (1971) Psychophysiological aspects of paced and unpaced performance as influenced by age. *Ergonomics*, **14**, 703–711.

AD-A191 244 THE EVENT-RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING A..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 3/10

UNCLASSIFIED 29 FEB 88 CPL-88-1 AFOSR-TR-88-0316 F/G 5/8 NL



MICROCOPY RESOLUTION TEST CHART

Sanders, A.F. (1979). Some remarks on mental load. In N. Moray (ed). *Mental Workload: Its Theory and Measurement*. New York: Plenum Press.

Sandman, C. (1984). Afferent influences on the cortical evoked response. In M.G.H. Coles, J.R. Jennings, and J.A. Stern (eds), *Psychophysiological Perspectives: Festschrift for Beatrice and John Lacey*, pp. 242–259. New York: Van Nostrand Reinhold.

Sharit, J. and Salvendy, G. (1982). External and internal attentional environments II. Reconsideration of the relationship between sinus arrhythmia and information load. *Ergonomics*, **25**, 121–132.

Sharit, J., Salvendy, G., and Deisenroth, M.P. (1982). External and internal attentional environments I. The utilization of cardiac deceleratory and acceleratory response data for evaluating differences in mental workload between machine-paced and self-paced work. *Ergonomics*, **25**, 107–120.

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

Siddle, D. (ed.). (1983). *Orienting and Habituation Perspectives in Human Research*. New York: Wiley.

Sokolov, E.N. (1963). *Perception and the Conditoned Reflex*. Oxford: Pergamon Press.

Spinks, J.A. and Siddle, D. (1983). The functional significance of the orienting response. In D. Siddle (ed.), *Orienting and Habituation Perspectives in Human Research*, pp. 237–314. New York: Wiley.

Vogt, J.J., Meyer-Schwertz, M. Th., Metz, B., and Foehr, R. (1973). Motor, thermal, and sensory factors in heart rate variation: a methodology for indirect estimation of intermittent muscular work and environmental heat loads. *Ergonomics*, **16**, 45–60.

Wierwille, W.W. (1979). Physiological measures of aircrew mental workload. *Human Factors*, **21**, 549–574.

BEHAVIORAL AND BRAIN SCIENCES

FINAL DRAFT FOR COMMENTATORS
No further alterations to be made
except for typographical errors.
(We would be grateful if commentators
pointed these out separately.)

PRECOMMENTARY ON VERLAGER'S CRITIQUE OF THE CONTEXT UPDATING MODEL

Is the P300 componenent a manifestation of context updating?

Emanuel Donchin and Michael G.H. Coles
Cognitive Psychophysiology Laboratory
University of Illinois at Urbana-Champaign
603 East Daniel Street
Champaign IL 61820

ABSTRACT: To understand the endogenous components of the ERP we must use data about the information processing function of the underlying brain activity. These hypotheses in turn generate testable predictions about the consequences of the component. We review the application of this approach to the analysis of the P300 component, whose amplitude is controlled multiplicatively by the subjective probablity and the task relevance of the eliciting events and whose latency depends on the duration of stimulus evaluation. These and other factors suggest that the P300 is a manifestation of activity occuring whenever one's model of the environment must be revised. Tests of three predictions based on this "context updating" model are reviewed. Verleger's critique is based on a misconstrual of the model as well as on a partial and misleading reading of the relevant literature.

I. Introduction

In the Forward to his "New Essays on the Psychology of Art," Arnheim (1986) makes the following observation,

> "Articles in professional journals profit from a permanence of their own, albeit one tied to the time of publication. A discipline grows like a tree, one on which the function of every new twig is determined by its place in the whole. Each contribution justifies itself by addressing a question that the profession has put on the agenda at that time."

It would have been useful if this view had guided Verleger (target article, this issue) as he reviewed the literature on the P300 component of the Event Related Brain Potential (ERP). Verleger treats the body of publications concerned with the P300 as if it had all been created simultaneously. He singles out selective pieces of writing originating from different historical contexts to patch together the theory he advocates while launching a rather sharp attack on two versions of "context updating," which he attributes to Donchin and his colleagues. However, neither of these versions is a valid or even a reasonable approximation to our ideas. The independence of his discussion from the historical context may be what makes it so difficult for Verleger to see how inaccurately he represents our views.

One example will illustrate the inadequacy of Verleger's survey of the literature. Investigators of the P300 are keenly aware that at least two attributes of the component - its latency and its amplitude - must be measured in any study. Furthermore, it is clear that the amplitude and the latency are each affected by somewhat different experimental manipulations and that the variance in amplitude and the variance in latency present rather different theoretical challenges. A substantial literature has accumulated since Ritter, Simson and Vaughan (1972) published the first systematic study of the relationship between the latency of the P300 and reaction time (RT). Indeed, our account of the functional significance of the P300, which Verleger labels the "context updating" model, evolved in part from an attempt to account for the seemingly counterintuitive dissociation between P300 latency and RT.

It makes little sense, therefore, to examine the context updating hypothesis devoting as little attention as Verleger does to the latency of P300. His concentration on the variance in P300 amplitude and his decision to interpret the words "working memory" and "expectancy" in the narrowest possible sense creates two straw men. These are then subjected to vigorous criticism in an ostensible refutation of our model. It should be pointed out that not only is Verleger attacking the wrong models, his attempted critique is weakened by his treatment of empirical data. Important segments of the literature are ignored (the literature on P300 latency is but one example). When data are discussed, Verleger offers speculative interpretations that merely serve to fit his preconceptions.

This precommentary begins with a review of the explanatory tasks an ERP investigator must undertake at different phases of the investigation. We then present our own view of the "context updating" model, reviewing the data on which we base this view as we proceed. Where appropriate, we also comment on Verleger's interpretations of the data. We have devoted a portion of this precommentary to a metatheoretical discussion of our approach to the study of ERP components.

2. The Endogenous Components of the ERP

2.1. Some general remarks on ERPs

The application of signal averaging methodology to the study of brain activity which is time locked to specific events made possible the study of event related brain potentials (ERPs) in awake behaving human subjects. ERPs are records of the activity of neuronal ensembles activated in a fixed temporal relation to an event whose time of occurrence can be made known to a computer and whose geometry allows their individual fields to summate and generate a field large enough to be recordable from the scalp. (For a discussion of the biophysical and neurophysiological basis of research on the ERP see Allison 1984; Allison, Wood and McCarthy 1986; Nunez 1981.)

The ERP leaves much to be desired as a window on the brain. The neurophysiological significance of the fields is a matter of controversy. The specific intracranial sources that generate the potentials are, in the main, unknown. Furthermore, it is quite possible that the neuronal ensembles whose activity is recorded in this fashion do not constitute distinct and unique neuroanatomical entities (Allison et al 1986, are particularly eloquent on the deficiencies of ERPs as measures of brain activity). Nonetheless, ERPs have been studied with vigor for the past two decades and have engaged the interest of many investigators. This interest is driven by the one advantage the ERP has over other manifestations of brain activity, namely, that it can be measured in the awake, behaving human. The recordings can be done without recourse to invasive procedures and at a relatively low cost. The work has also received considerable momentum from the fact that such recordings are highly sensitive to a host of experimental manipulations. Whether one varies the physical parameters of the stimulation, the class of subjects used or the cognitive operations the subjects apply to the eliciting events, the ERPs show patterns that are stable in and between individuals and occasions. This stability of pattern in the ERP has proved useful in the development of various diagnostic procedures in neurology, audiology and psychiatry, (e. g. Cracco and Bodis-Wollner 1986). However, in this precommentary we will focus on the possibilities and problems that arise when one attempts to use ERPs as tools in the study of cognitive function.

In the middle 60s it became apparent that the ERPs contain "endogenous" components whose variance appears to be relatively insensitive to the physical attributes of stimuli. In fact, the eliciting event need not even be a physical stimulus (Sutton, Tueting, Zubin and John 1967). Endogenous components are sensitive to variation in the information processing activities imposed on the subject by the assigned tasks (see Donchin, Ritter and McCallum 1978, for a discussion of the distinction between exogenous and endogenous components).

2.2. Classes of Psychophysiological assertions

Both Verleger's target article and this precommentary are concerned with an endogenous component which was first described by Sutton, Braren, Zubin and John (1965) who reported that stimuli which "resolve the subject's uncertainty" elicit an ERP characterized by a large positive component, with a latency to the peak of about 300 msec, which will be referred to here as the P300.[1] The enterprise launched by Sutton has yielded considerable information about the P300[2]. An examination of the literature reveals that investigators concerned with the P300 make assertions which can be classified into one of the following categories.

2.2.1. Statements about antecedent conditions

These are statements about the functional relationships between various attributes of the P300, such as its amplitude and latency, and assorted independent variables. For example, the claim that the amplitude of P300 is inversely proportional to the subjective probability of the eliciting events, given constraints on the task relevance of the events, is a statement about the conditions antecedent to the elicitation of the P300. The antecedent conditions of the P300 have been studied in detail and there is enough consensual data to support a substantial number of statements about its antecedent conditions. (For detailed reviews of the relevant literature see Pritchard, 1981; Johnson, in press). Such statements are necessary precursors of a theory of the P300. However, a collection of such statements does not constitute an adequate theory.

2.2.2. Statements about Applications

Even though statements about antecedent conditions are not a theory of the P300, they can serve as the basis for an application of the P300 in some domain of human action where a measurable index of some construct or variable is needed. Thus, for example, the observation that P300 latency depends on processing time has led to the suggestion that P300 latency may discriminate the demented from the depressed (Goodin, Squires and Starr 1978). The validity of this diagnostic use of the P300 depends solely on empirically demonstrable diagnostic power and does not require a theoretical understanding of the component, its origin or its functional significance. Similarly, the fact that P300 amplitude reflects the subject's focus on one of several concurrently assigned tasks makes it possible to use it as a measure of "mental workload" (Donchin, Kramer and Wickens 1986a). From an applied perspective such assertions are quite useful. However, they treat the ERP component as an index and do not necessarily shed light on its functional significance.

2.2.3. Generalizations about antecedent conditions

Several investigators have attempted to integrate the many empirical statements about the diverse independent variables which affect the P300 into a coherent statement that will identify the commonality among these statements. It is generally assumed that the diversity of effects is deceptive and that there is a common denominator to the various conditions under which the P300 is elicited. It would clearly be useful to have such a generalization so that one could predict the specific effec of any given manipulation on P300. This approach is illustrated by Johnson's (1986) proposal tha the amplitude of P300 is controlled by an additive and multiplicative relationship between several variables. The interpretation advocated by Verleger, namely, that "P3s are evoked by stimuli which are awaited when subjects are dealing with repetitive, highly structured tasks" constitutes another such generalization about the antecedent conditions of P300. Even though such generalizations are

---

[1] Components of the ERP are labeled by a letter indicating polarity and a number indicating their modal latency However, other nomenclatures exist and this very same component may be labeled P3 by some authors, including Verleger, and P3b by others.

[2] In keeping with the usage in Verleger's target article we will ignore the controversy regarding "multiple positivities." It is quite likely that a variety of processes manifest themselves on the scalp concurrently with those manifeste by P300, although the ardor with which this proposition is embraced varies across laboratories. However, a P300-lik component appears in everyone's pantheon of components and it is to this component that we direct our attention in thi precommentary.

valuable (and necessary precursors to a theory), they leave the functional significance of the P300 unexplained.

2.2.4. Hypotheses about functional significance

When we refer to the "functional significance" of the P300 we imply that the potentials recorded from the scalp, to which we refer as the P300, are manifestations of the intracranial activity of a specific functional entity. As we have noted elsewhere (Donchin, Karis, Bashore, Coles and Gratton 1986b), it is not necessary that the brain activity recorded as the P300 originate in a single intracranial "source." Nor is it necessary that the neuronal ensembles whose activity contributes to the P300 be associated with a single neuroanatomical entity. Many brain structures may be involved and several "sources" (in the sense of specific dipoles) may contribute to the component. All that is necessary is that this variety of neuronal activities that participate in a specifiable information processing task have the geometry and temporal characteristics so that they contribute to the P300. The ultimate goal of a theory of P300 is, we suggest, the elucidation of the nature of this task.

The brain is evidently an organ whose large variety of activities can be described in neurophysiological terms. There is no question that the ultimate understanding of the brain must include the detailed explication of its functioning as a mechanism constructed of neurons. However, this neuronal machine is an information processing device and the myriad of biochemical interactions that it performs must accomplish tasks that can be described in information processing terms. The level of resolution one adopts to describe the brain is dictated by one's purpose. No level is inherently more interesting or more valid than other levels; each descriptive tier is optimal for some purposes and quite the opposite for others. We choose, in the present context to focus on descriptions described strictly in information processing terms.

2.2.4.1. The Whirring Noise analogy

Within this framework we view the P300 as a surface manifestation of an internal information processing operation. As theorists our goal is to infer the nature of that underlying processing. Hence statements about the antecedent conditions of the component do not serve our purpose; they merely tell us what needs to happen outside the system to call forth the activation of the component. That, however, fails to explain what the brain is actually doing, functionally speaking, when the component is active.

An example may help. Assume you have a computer that always makes a whirring noise when activated. One may find out that to observe the whirring noise a computer must be attached to a power source, that it must be equipped with a diskette drive and that a diskette must be loaded into the drive. We also note that the whirring noise is generated whenever the computer has been reset following a sufficiently serious error. These and other such statements are collectively a description of the antecedent conditions for that whirring noise. However, none of these individually, or all together, account for the whirring noise in a satisfactory manner because none specifies the purpose of the process from which the noise emanates.[3] Through an accumulation of such antecedent conditions we may conclude that the whirring noise is heard whenever there appears to be a need for additional data not immediately available in the computer's memory. This will remain a statement about the antecedent conditions because it still does not suggest what is the function of the process manifested by the noise. One may however, after appropriate research, infer that the whirring noise is generated whenever data must be transmitted between the computer's memory and an external

[3] Note the distinction between the function of the noise itself of which there is none and the function of the process generating the noise as a byproduct. This is an apt analogy to the ERP as we are not claiming any functional significance to the potentials we record on the scalp. Rather, we assume that what we record are byproducts of the activation of processes which do have functional significance.

storage device. The noise is generated by a process invoked in the course of this data transmission, namely, when the mechanical device used to whirl the diskette around, or to move the reading head about, is active. These are statements about the functional significance of the process manifested by the noise. The process is very specific. The activation of the diskette motors is performed in the service of data transmission between the CPU and the diskette. It has an actual effect on the system only if other component processes are activated but its own functional significance can be described, to an extent, without reference to these other processes. Furthermore, it will turn out that whereas we can bring about this whirring noise whenever we boot the computer, there are many other occasions that have little functional similarity to the specifics of starting the computer's day which also cause the whirring noise to appear because all these diverse processes share a need to access the diskette, they all activate the motor and all activations generate the same whirring noise.

This analogy illustrates a distinction that appears not to have been made clearly enough in our previous writings. Note that the noise is a manifestation of an information processing operation involving the transmission of the data from a device called a diskette into memory. However, the generator of the noise - namely, the actions of the motor - is not itself necessary for the transmission of the information. Thus, for example, it is possible to simulate a diskette in memory. No whirring noise will be generated when the computer exchanges data with these memory-based simulated diskettes. In other words, under a specific set of circumstances one can obtain an external manifestation of an information processing operation because the side effects of some ancillary process are observable. This side effect, or the ancillary process which it manifests, is not identical with the totality of the information processing operation, it merely makes it observable. This distinction is important because Verleger's interpretation of our model is that the P300 is a manifestation of the updating process per se whereas we have tried carefully to assert that it is a manifestation of a process invoked in the service of the updating process, not necessarily the updating per se.

We view the P300, and all ERP components, much the same way we view the whirring noise in the foregoing example. We assume that the component is generated as a byproduct of the activation of a neural process whose function is the transformation of information. A satisfactory explanation of the component will, accordingly, describe the specific processes executed by the activity of which the component is a manifestation rather than by describing what it takes to generate the component, or to control its attributes. Note that neither of Verleger's construals of our view, his "Update" and his "Expect," satisfies this condition. Both purport to specify the rather limited conditions under which a P300 can be elicited but neither specifies its functional significance. "Context updating" as we interpret it attempts to identify a specific type of cognitive process which underlies the P300.

2.3. The Concept of 'Psychological Processes?'

2.3.1. Introductory remarks.

Verleger begins his target article by promising to "clarify what psychological process is associated with the P300 component." His clarification consists of the suggestion that:

> "P3s are evoked by stimuli which are awaited when subjects are dealing with repetitive, highly structured, tasks. Effects on P3 can be modeled by four parameters: (1) Whether or not stimuli match the expectancy for closing events. (2) The number of open epochs. (3) The degree of capacity invested in the epoch (positive correlation) (4) The degree of processing required by the stimulus. (negative correlation) (p. 2)."

This hardly seems to be the description of a psychological process as one ordinarily understands the term. Verleger is not alone in using "psychological process" so loosely. There have been many claims in the literature associating P300, and other ERP components, with various psychological concepts. The efforts range from identifying a description of antecedent conditions with psychological concepts to rather grandiose attempts to associate P300 with such global concepts as "attention" or "memory." It is important to be precise in describing such associations, in particular with regard to the level of the concepts used.

2.3.2. The hierarchy of concepts

There is a hierarchy of descriptors of psychological processes and we claim that ERPs are manifestations of activities related to an elementary level in this hierarchy. Descriptors of the highest level are terms such as motivation, emotion, memory and learning. These terms, indubitably psychological, describe broad categories of interactions between the organism and the environment. The instantiation of any such function may call forth a diverse set of information processing activities. These functions may accordingly require the activation of such "psychological" processes as "choice," "recall," "decision," or "discrimination." Such descriptors refer to specific information-processing functions each of which can be marwilled in the service of one or another of the functions described at the higher levels. Functions at the second level of the hierarchy require, in turn, the use of more elementary functions such as "encoding" or "mismatch detection." Descriptive terms at all levels of the hierarchy refer to "psychological" processes.

A computer analogy may clarify the distinction[4] Assume that one can examine all the programs in a computer system over a period. It is clear that these programs can be classified into different categories such as: accounting packages, games, word processors, statistical programs, and so on. This is a level of description that focuses on the function of the programs, the goals of the system as it interacts with the user or the world. At this level of description all word processing programs can be considered equivalent even though they may be quite different in their details.

Computer programs can also be discussed in a way that cuts across these goal oriented descriptions. A lower level describes the functions that must be executed in the system - with no necessary commitment to implementation. As illustrations consider: acquire data, store data, detect errors, back up data, invert matrix or produce graphic displays. The important point is that this class and the one just above it are not congruent. That is, each of the higher levels uses subsystems of the lower kind. Thus data may be acquired to subserve "accounting" or "data management." The descriptions that focus on the system's goals are orthogonal to the descriptions that focus on the actions the system must perform to accomplish those goals. This is even truer for a still lower descriptive level referring to operations at the implementational level (e.g. move index register I to the accumulator, or release all I/O channels). [See also Anderson: "Methodologies for Studying Human Knowledge" *BBS* 10(3) 1987.]

The analogy extends to descriptors of psychological processes in that there are descriptors that refer to the general class of goals used by the system. "attention" "memory" "love" and so forth are in this sense analogous to "accounting programs." At a lower level we have "psychological processes" such as "encoding" and at an even lower level we may have elementary information processing operations. It is not clear to which category of processes Verleger refers when he speaks of "psychological processes." It would seem, from his use of the term "expectancy," that he is referring to "psychological processes" at the highest level of the hierarchy. Yet the data on the endogenous components of the ERP suggest that they are "related" to processes that are describable at an elementary level of the hierarchy just discussed. Thus, we are not trying to "correlate" a brain wave with a high level psychological process, as implied for example by Churchland (1986), but rather to ask whether the brain wave, viewed as the scalp manifestation of information processing, does indeed behave in the way the byproduct of such a processor would behave, given certain manipulations.

---

[4] For those who are skeptical about analogies to computers we note that we are using this comparison merely to illustrate the use of different descriptive levels when discussing an information processing system.

It may be argued that the study of byproducts is of no inherent interest. This is a shortsighted view. We suggest that as long as there is a valid relation between the process of interest and its measurable byproducts, what matters is the degree to which the measurements can be placed at the service of interesting theoretical questions. Consider, for example, the electroretinogram (ERG). Although clearly the byproduct of an ancillary process, the ERG is evidently a manifestation of the activation of the retina. The function it manifests is the processing of the light patterns that fall on the retina. It is thus a critical component of the visual system. One may find, upon study, that ERGs are elicited when one directs one's gaze with love at one target and with hate at another. This does not mean that the ERG is "psychologically meaningless." It means rather that the psychological process manifested by the ERG is more elementary than "love" and "hate" in the sense that the function in which it participates, namely, "vision," may be involved in both of these higher level psychological processes.

We have examined in detail the nature of the enterprise in which we are engaged as we theorize about the P300 because Verleger is attributing to us an agenda we eschew. This error distorts his reading of the literature and focuses his attention on trivialities. He seems to be persuaded that the task we have undertaken is to identify a high level concept such as "surprise" or 'expectancy" or "memory" with the components. He then seeks out evidence that would be relevant to the demonstrations he feels we ought to be conducting. He then attacks the validity of the evidence, thereby missing the most important corollary of the emphasis on the search for the functional significance of the component, namely, the emphasis on the consequences of the component rather than on its antecedents.

2.4. The search for consequences

The concern with the consequences of ERP components is a crucial distinction between our approach and Verleger's. The issue concerns the process by which one validates theoretical assertions about an ERP component. It would appear that for Verleger the prime criterion for evaluating a theory is the post-hoc plausibility of the generalizations one makes about the antecedent conditions. This is evident from his treatment of (his construal of) the context updating model and from the way he supports his own model. Consider, for example, his treatment of the "Update" model, which he summarizes as follows:

> "Each stimulus has its representation ('trace') in working memory, which decays over time. After stimulus evaluation, the trace of the presented stimulus is updated to its optimal upper-limit value and it is this updating that is reflected by P3. If the most recent presentation of the present stimulus was not too long ago, then the amount of updating required is small, and so is P3. But, if that presentation was not too recent, then the amount of updating, and likewise the P3, has to be larger." (p. 4)

Ignore for the moment the fact that we have never implied that the P300 is a manifestation of the updating of specific memory traces. It is interesting to examine how Verleger evaluates the evidence he sees as relevant to this version of context updating.

Verleger accepts two studies as "direct tests" of the hypothesis. In one, Klein, Coles and Donchin (1984) reported that subjects with perfect pitch do not produce a P300 in an auditory oddball[5] task.

[5] The term "oddball" applies to all tasks which satisfy the following conditions: (a) The subject is presented with a series of events, where "event" can be defined in a complex and abstract manner as long as the subject can distinguish one event from another. (b) A classification rule is defined which categorizes each of the events into one of two categories. (c) The subject is assigned a task whose performance depends on applying the categorization rule to the events. (d) Events in the two categories are chosen at random, that is, they constitute a Bernoulli sequence. In general, a fifth condition has also obtained, namely, (e) the probability of one of the categories is substantially lower than that of the other category. It is this focus on the rarity of one class of events that gave the paradigm its label ("oddball.") And, in

The other is the series of studies of the von Restorff effect published by Karis, Fabiani and Donchin (1984), and Fabiani, Karis and Donchin (1986). It is instructive to note the discrepancy between the actual reports of these studies and the way they are described by Verleger. Of the Klein et al study Verleger says that the investigators "presented two go/no-go tasks to control subjects and to music students who could identify the absolute pitches of tones (p. 5)" It is not merely a quibble to point out that Verleger's implication that the control subjects were not music students is not consistent with the report by Klein et al that "All subjects were students of music at the University of Illinois (p. 1306)" Neither is it a quibble to note that the subject's self report of their ability to make absolute judgments of pitch was not the sole criterion for the results reported. In fact, all subjects were tested for their ability to name tones. The amplitude of the P300 was negatively correlated ($r = -.63$) with their actual performance on the Lockhead and Byrd test (1981) regardless of their self perception of their skill as perceivers of absolute pitch. These correlations are presented in Table I of Klein et al (page 1306). Of course, had Verleger not ignored these crucial details, he might have been less inclined to propose that Klein et al's results can be explained by a difference in motivation between control subjects and AP subjects ("who probably knew that they were selected because of their special skills. (p. 5))" It should be noted that Ebner, Schutz and Lucking (1986) report a successful replication of the Klein et al. study.

Verleger's presentation of Karis et al's and Fabiani et al's findings is similarly inaccurate. The focus of the study by Karis et al was the relation between the amplitude of P300 and recall in the context of the Von Restorff paradigm. The interesting aspect of the paradigm is that a subject is not presented merely with a "list of words," as Verleger states it, but rather with a list of words some of whose members are deviant from the rest of the list. The recall of such "isolates" is enhanced, as reported by von Restorff (1933). Karis et al's abstract describes the actual results as follows:

> "there were strong relationships between the Von Restorff effect, over all recall performance, mnemonic strategies, and the association between components of the ERP and recall performance. The overall recall performance of subjects who reported simple (rote) strategies was low, but they showed a high Von Restorff effect. For these subjects the amplitude of P300 elicited by words during initial presentation predicted later recall. In contrast, subjects who reported complex mnemonic strategies remembered a high percentage of words and did not show a von Restorff effect. For these subjects P300 did not predict later recall, although a 'slow wave' component of the ERP did (p. 177)."

This pattern of results provides striking support for the "context updating" model as we construe it. It is conceivable that others, Verleger included, will not be as persuaded as we are by these data, but it would seem incumbent on anyone commenting on the study to consider the data as reported rather than a rather selective and inaccurate simplification.

We will return to the specifics of these studies later as we examine the evidence for the context updating model. However, we note here that in his analysis of what he calls "direct evidence" for "Update," Verleger is primarily concerned with the degree to which the studies demonstrate that manipulations of "working memory traces" affect the P300. This tendency is even more apparent in his discussion of the "indirect evidence." Verleger enumerates a series of topics relating to changes in P300 amplitudes in different experimental conditions. Thus he notes, inter alia, that "Sequence effects are flexible," and that "Probability might affect P3 independently of sequence" or that "There is a target effect." All these assertions are correct enough. As a substantial portion of the data base he adduces derives from work conducted in our laboratory we are quite content to accept the validity of the empirical claims. But, unlike Verleger, we find all these data consistent with the context

---

general, the rarer the event, the larger the P300 it elicits (Duncan Johnson and Donchin 1977). However, it has been known for some time that rarity is neither necessary nor sufficient to elicit a P300. (See Fabiani, Gratton, Karis and Donchin (in press), for a review of the paradigms used in P300 experiments.)

updating model as we construe it. We would not adduce them as conclusive evidence for the model, however, because the data cited by Verleger invariably involve statements about the antecedent conditions, and the relations of these to the theory is necessarily post-hoc.

The structure of Verleger's argument appears to be the following. A model of the information processing system is constructed in the form of a block diagram (see Verleger's Fig 2, p. 4). The data on P300 are than considered and an association between the P300 and some element in the (essentially structural) model is derived. The data considered in forming the association between the model and the P300 are assertions about antecedent conditions. In this phase of theorizing one seeks a plausible account of the known data which would integrate them as fully as possible in a reasonable model of the human information processing system. This is fine, except that it is entirely circular for data that were used in forming the model to serve as evidence for the model. After all, if the model was created because it provided a plausible account of the antecedent conditions then any assessment of the model against these data tests nothing more than the theorist's flair for plausible post hoc explanations.

Theory in psychophysiology, to the extent that it ventures ideas about the functional significance of a component, must seek its validation in its predictive power rather than in its post-hoc plausibility. This predictive power can be tested because assertions about the functional significance of components ought to generate testable predictions about the consequences of the components. Posner appears to have been the first to note the need to focus on "consequences" in psychophysiological theory. In the proceedings of a conference held in 1973, he is quoted as remarking (in speaking of the P300) that "We should be able to describe the consequences of its use" (McCallum and Knott 1976, p. 224). The concept of the 'consequences' of the use of an ERP component derives from viewing the component as a manifestation of an elementary processing function. As Donchin et al (1986b) put it:

> "We assume that the consistency displayed by an ERP component in its relation to experimental manipulations is due to the fact that the activity it manifests is generated by the invocation of a distinct component of the information-processing system. A description of the component - if it is to bear meaning for the cognitive scientist - must identify the transformations that the algorithmic component, or the "subroutine" performs. (p. 247)"

The metaphor of the "subroutine" is important here. Subroutines (these days the metaphor of choice might have been the "procedure" or "production") are information processing entities that can perform a function in the service of a diverse array of higher level programs. The programs calling the subroutine may bear no resemblance to each other except that at one time or another they require the service performed by the subroutine. The interpretation of P300 as a "general purpose processor" by Donchin, Kubovy, Kutas, Johnson, and Herning (1973) was an attempt to cope with the bewildering diversity of circumstances in which the P300 can be elicited.[6] By 1973 a partial listing of the circumstances under which P300 can be elicited would have included stimuli that resolve uncertainty (Sutton et al 1965); stimuli that are task relevant (Donchin and Cohen 1967); subsets of stimuli used in a signal detection paradigm (Hillyard, Squires, Bauer and Lindsey 1971), or any first stimulus in a

[6] Donchin et al (1973) used the following words: " In view of the diversity of the P300 function that our data reveal, it seems that a reasonable approximation is to assume that P300 reflects the activity of a general-purpose processor which is invoked on demand by a host of data processing requirements. (p. 322)." This formulation appears to have confused a number of authors who interpreted us to be saying that P300 is a generalized processor with no specific function. The intent, however, was quite the opposite. The phrase "general-purpose," for which we used a Floating Point Processor as an example, was meant to imply a device with quite a specific function that could, however, subserve a larger variety of higher level functions. As noted above, the retina is a device that is specialized for the acquisition and initial processing of information carried by light. Yet it is a quite general device, in the sense that it can be used to look at, look for or look after whatever it is that the person's task requires at the time.

series, even if it carries no significance (Ritter, Vaughan, and Costa 1968).

Much of the theoretical effort at the time was directed at indentifying an overarching concept that would subsume this broad range of conditions. The task has proven unwieldy and virtually every attempt to solve the problem has left part of the observations unexplained. It was in this context that Donchin et al (1973) suggested that rather than calling for an explanatory concept at the highest level of the hierarchy of concepts it may be useful to assume that the P300 is a manifestation of an entity whose place in the hierarchy is much lower; this would be consistent with its apparent multipurpose nature.

Information processors are defined by the transformations they perform on information. They have inputs and outputs and their functional description must specify the transformation that has been applied to the input to generate the output. This is the description we seek of the processes underlying ERP components. It is critical to understand that the ERP itself is not the output of that process. Like the whirring noise in our previous example, the ERP is a byproduct of the activity of the processor which we can use as an index. The implications of the view that the ERP is not the principal output of the processor underlying it were stated by Donchin et al (1986) as follows:

> "If the ERP is not the critical output of the subroutine, then it is clear that a description of the relationship between the variables that control attributes of the subroutine can not in itself constitute an adequate description of the subroutine. The transformations we seek to describe are between the proper inputs and the proper outputs of the subroutine. Therefore the study of the antecedent conditions can only provide clues from which one must derive hypotheses regarding the functional significance of the routine. These hypotheses must be stated in terms of the effect the activation of the subroutine has on the recipients of its output. Clearly, to be of value the output of a processor must act as input to some other stage of the system. (p. 247)"

The emphasis on the study of the consequences of the component is a direct corollary of this approach. If the ERP manifests an information processor and indexes its utilization, and if we have a notion of what the transformations are which the processor performs, then we ought to be able to predict what would happen after the activation of the processor as a function of its degree of activation. The clearer our hypothesis about the component's function the clearer the predictions and the sharper the test of the model.

3. The Antecedent conditions for the P300

3.1. Interim summary

To summarize the ideas presented thus far we note that:

(1) We assume that the P300 is a manifestation on the scalp of brain activity that performs a specific information processing function. The activity may occur in a number of different structures concurrently, in a complex pattern.

(2) We assume that the brain activity in question implements some transformation on information that can be described in terms of a model of human information processing (a "cognitive" model). The P300 may well be the byproduct of a process ancillary to the underlying function. Thus, there may be two distinct explanatory burdens, only one concerned with the principal function manifested by the processes of which the P300 is a component.

(3) Theorizing about an ERP component consists of specifying candidate functional roles (specified as transformations of information). These theories must be able to predict the "consequences" of an ERP component. The theories are tested by the degree to which these predictions are confirmed.

(4) There is, of course, the need to ensure that measures of the ERP components are reliable and valid. This is a difficult but not an impossible task and in this precommentary we will assume

that the methodological problems are surmountable. For detailed discussions of the methodology see Fabiani et al (in press), Coles, Gratton, Kramer and Miller (1986), Glaser and Ruchkin (1976).

(5) Hypotheses about the functional role of the activity manifested by a component must be derived from studies of its antecedent conditions. We will therefore begin our presentation of the context updating model with a review of what is known about the antecedent conditions of the P300.

3.2. Event Probability and the P300

3.2.1. On the difference between observation and interpretation

Two contradictory assertions can be made. First, we note that it is neither necessary nor sufficient for an event to be the rare member of a task-relevant Bernoulli sequence for it to elicit a P300. We immediately note that this caveat notwithstanding, the evidence is overwhelming that in the main the rarer the event the more likely it is to elicit a P300. Furthermore, the amplitude of the P300 elicited by such rare events will be inversely related to the probability of the event. Both statements are supported by a considerable body of data and both seem to be accepted by Verleger. A number of authors have thoroughly documented the relationship (for a recent review see Johnson, in press). It would be useful, however, to comment on the way these data have accumulated with particular attention to those aspects which relate to Verleger's discussion of the role of "expectancy."

One of the difficulties in assessing the effect of event probability on the P300 is that investigators have not always distinguished between statements about the manipulations which are directly under the investigators' control, and statements that interpret the manipulations in terms of psychological constructs. The story is made even more complicated because, as Sutton (1969) has noted, it is how subjects exercise their options that determines how fully the experimenter controls the independent variables. The problem appeared with the very first paper in which the P300 was reported by Sutton et al (1965). The results of that classic experiment are summarized by the assertion that P300 is elicited by stimuli which "resolve the subject's uncertainty." However, it is only inferred that the subject's "uncertainty" was manipulated. What was directly under Sutton's control was the extent to which subjects were informed in advance which stimulus would be presented. Subjects were also instructed to predict the next stimulus. It is entirely accurate, therefore, to report that "when subjects are not informed in advance which of the two possible stimuli will occur, and they have predicted one or the other, the ultimate appearance of the event will elicit a P300."

It is plausible, but not self evident, that subjects were uncertain as to which stimulus would occur and hence that the occurrence of the stimulus did resolve the uncertainty. However, a number of assumptions that go beyond a report of the investigator's actions are made when the concept of "uncertainty" is invoked. It is assumed, for example, that the subjects do indeed monitor the series, maintain a memory of their predictions, and attend to the stimuli so that their uncertainty can be resolved. Whether or not such assumptions are valid depends on the effectiveness with which the investigator has controlled the experimental environment. Furthermore, it is incumbent on the investigator to provide evidence that such control has been exercised. (We do not question Sutton et al's interpretation as they did exercise the necessary controls.)

To the extent one reviews the relationship between P300 amplitude and specific experimental manipulations, the pattern of results is consistent and replicable across many laboratories. On the other hand, when results are recast in terms of explanatory psychological constructs such as "uncertainty," "surprise" or "expectancy," controversy arises, and it does so largely because the constructs used are not consensually defined. Verleger, for example, while accepting the validity of most relevant data, proceeds to argue that it is not necessary that stimuli be "unexpected" for them to elicit a P300. We will examine the specifics of this argument. But, it would be useful to begin with an examination of how the term "expectancy" was used by the Donchin and his colleagues.

3.2.2. Prior Probability and the P300

In a wide array of circumstances (though by no means in all), manipulating the probability of an attended event (a "task relevant" event) will affect the amplitude of the P300 elicited by the event. True, in their original experiments Sutton's group made the two events in the series equiprobable and varied the information the subject had about the next event. However, Sutton used the phrase "resolution of uncertainty" in the formal information-theoretic sense. There, resolution of uncertainty is synonymous with "delivery of information," and information is considered a function of the probabilities of the different "messages." Thus, it was natural to examine the effect that variation of the relative probabilities of the events would have on P300 amplitude. In subsequent work (e.g Tueting, Sutton and Zubin 1970; Friedman, Hakerem, Sutton and Fleiss 1973), the probability with which events occurred was varied and the amplitude of the P300 was again shown to be inversely related to the probability of the eliciting event. A parametric study of the odd-ball paradigm by Duncan-Johnson and Donchin (1977), using a larger number of probabilities, demonstrated a monotonic, inverse, relationship between P300 amplitude and stimulus probability. (See Donchin 1979, and Sutton 1979, for reports of the status of the issue as it stood in 1977; Campbell, Courchesne, Picton and Squires, 1979, provide a review from a somewhat different vantage point which arrives at the same conclusions.)

3.2.3. The role of subjective probability

It is important to note, however, that in many of the studies considered relevant to the "probability" or "expectancy" issue what is being controlled is not the prior probability, defined by the sequence generating rules used by the experimenter, but rather the degree to which the event is likely from the subject's point of view. Horst, Johnson and Donchin (1980), for example, recorded the P300 while subjects were engaged in a "paired associates" learning task. In each series of trials the subject was expected to learn which nonsense syllable (traditionally labeled a CVC for Consonant-Vowel- Consonant) must be produced in response to a CVC presented by the computer. On each trial the subject responded with a CVC and the computer then indicated which CVC would have been the correct response on that trial. These feedback presentations are referred to as the "response" CVCs by Horst et al. The P300 elicited by these CVCs was their object of study.

As the pairings of the CVCs was entirely arbitrary, the subject had to learn the correct response by observing the response CVCs. Initially, the subject's own responses were almost entirely erroneous, but as the trials progressed, subjects did learn the pairings, and their responses became almost entirely accurate. Thus, during the early phases of learning the response CVC almost always indicated that the subject had erred, whereas in the latter stages the response almost always indicated that the subject was right. Note that no particular CVC was presented frequently, yet the absolute probability of a given CVC did not predict its amplitude. One might be tempted to predict that the P300 would vary with the degree to which the CVC indicated whether or not the subject had erred, but this was not the case. It turns out that the amplitude of the P300 was determined by the interaction between the subject's confidence in the correctness of the response and the actual outcome of the trial.

Horst et al required the subjects to report their confidence before each response CVC appeared. The confidence ratings were analyzed in great detail and were normalized so that each subject's use of the confidence scale was taken into account in the analysis. The results were summarized as follows:

> "Our data indicate that the amplitude of the P300 elicited by the outcome and the subject's expectancy concerning this outcome. Neither confidence by itself nor whether the "response" CVC confirmed or disconfirmed the subject's three-letter response accounted for the variance in P300. Rather, P300 amplitude depended on the degree to which the outcome of each trial

> was unexpected. The lower the subjective probability assigned to the outcome, the larger was the elicited P300. These data thus strengthen the claim that P300 amplitude is dependent on the subjective probability associated with the ERP-eliciting event. (p. 483)"

Note that the term "unexpected" is defined operationally in terms of the discrepancy between the subjects' reports of their confidence and the actual outcome of the trial. The observed relationship between P300 amplitude and the interactive relationship between outcome and confidence is an empirical finding whose validity depends on how competently the experiment was conducted. Once we speak of "subjective probability," which in the papers published by Donchin and his colleagues between 1970 and 1980 was taken to be synonymous with "expectancy," we enter the realm of interpretation. The term "expectancy" was defined by Squires, K.C., Wickens, C. D., Squires, N.K. and Donchin, E. (1976) as referring to a hypothetical construct that accounted for their observation that even though the probability of events predicted the amplitude of P300 when measured from the ERP averaged over trials, there was substantial variability in the amplitude across trials. Much of this variance (78%) could be accounted for by assuming that P300 is an inverse function of "expectancy" and that this expectancy varies according to a three component model which combines an exponentially decaying memory with the prior probability of events and with the effect of alternating sequences. A critical implication of this model has been that it is the "subjective" rather than the prior probability that determines the amplitude of P300. It is to test this hypothesis that Horst et al measured the interaction between subjects' confidence and actual events. They present their interpretation of the results as follows:

> "Our notion of subjective probability implies that subjects apply their knowledge about a given situation to form differential expectancies (subjective probabilities) for the various events that might occur. These expectancies, being derived from external information that is filtered by subjects' perceptual biases, stored in a fallible memory, and tainted by the individual's predilections, are "subjective" in that they need not accurately reflect the objective probabilities with which events occur...Information processing triggered by the occurrence of an event is affected by the expectancy associated with that event. An aspect of the processing invoked by unexpected events is reflected in P300 amplitude p. 484)."

We dwell on the Horst et al study because it receives detailed attention from Verleger, who begins by stating that:

> "Another line of evidence frequently quoted in support of the claim that P3 reflects surprise is that disconfirming feedback evoked larger P3s than confirming feedback (Squires et al 1973; Horst et al 1980; De Swart, Kok and Das Smaal, 1981). p. 12"

This is a very partial and therefore quite misleading summary of all three studies. In the case of Horst et al (1980), it is patently incorrect, because these investigators clearly denied an exclusive relationship between the status of the stimulus as a confirming or disconfirming event. Rather, they attempted to measure directly the subject's expectations for confirmations and disconfirmations and reported that both confirming and disconfirming feedback will at times elicit a large P300 and at other times a small one depending on the subject's level of confidence. Horst et al's analysis is quite detailed. It may not be entirely persuasive and it is always possible that there are flaws in it and in the assumptions that underlie it. We would welcome a valid critique of these analyses, but that is not what Verleger's erroneous description seems to represent.

In a subsequent reference to Horst et al (1980), Verleger uses a strategy he employs elsewhere in his target article. Instead of dealing with the reported data, he surmises what subjects may have imagined while participating in the experiment. His language is instructive, as it illustrates the level of his critical analysis of empirical data:

> "In the Horst et al study, four of the six subjects were experienced, and all subjects were instructed 'We want to correlate your confidence rating with your brain waves.' In conclusion,

subjects could easily have imagined or even known exactly which outcomes were the object of the study, so that they expected those outcomes with greater interest. p. 13"

This interpretation is clearly at variance with the data. It is entirely unclear how the subjects' imagination could have controlled the outcomes with such precision when the subjects could not possibly know until they learned the correct pairings which outcomes to expect. Perhaps Verleger is implying here that the subjects were more interested in outcomes that violated their predictions, but this is unlikely because the outcome per se did not determine the P300 amplitude. Rather, it was the interaction between outcome and confidence that was the controlling variable. Furthermore, Verleger fails to deal seriously with the very detailed trial to trial analysis conducted by Horst et al. He likewise ignores the complex pattern of results reported by Squires et al (1973), focusing instead on the fact that the experimenters served as subjects. He also fails to take into account the very many replications of the signal detection work which followed that seminal experiment and which repeatedly found the pattern of results (see, for example, Campbell et al 1979; Hillyard et al 1971).

3.2.4. The role of Expectancy

It would appear then that the effect of "expectancy" on the amplitude of the P300 cannot be denied if "expectancy" operationally defines describes the conditions of stimulation and their interactions with the instructions to the subject. This does not mean that the relationship will hold for every construal of the term "expectancy" as a psychological construct. Such construals must be made in the constraints of a psychological theory and their validity has to be tested by their predictive power rather than by their apparent post-hoc plausibility or implausibility.

It is also important to emphasize that at no time did we assert that the subjective probability associated with events is the sole determinant of P300 amplitude. Indeed, it became evident quite early (Donchin and Cohen 1967) that whereas to a first approximation P300 amplitude was strongly affected by the rarity of the eliciting event, it was not possible to account for all its variance merely by specifying the prior probability of the eliciting stimuli. The need to go beyond "objective" or prior probabilities became evident with the discovery of the effects of the preceding stimulus sequence on the P300, first reported by Squires et al (1976).

As Verleger pays much attention to these data, it is worth reminding the reader what Squires et al reported: The amplitude of P300, when examined in a Bernoulli series governed by prior probability P, varies as a function of the sequence of stimuli presented on the preceding trials. When plotted as a "tree," these sequential effects mimicked the sequential effects often observed in reaction times (Falmagne 1965; Audley 1973). Squires et al suggested that the variable which controls the amplitude of P300 is the subjective probability, rather than the prior probability of events. They proposed that this subjective probability, or "expectancy," varies as a function of a decaying memory for each stimulus in the series, the global probability P and an "alternation" factor that was required to account for unique responses to stimuli that immediately followed an alternating pattern. The model has proved quite powerful and, as Verleger notes, the results have been widely replicated.

Squires, Petuchowsky, S., Wickens, C. D. and Donchin, E. (1977), demonstrated that the sequential effects appear in both visual and auditory oddball series and that the model presented by Squires et al (1976) accounts for the results in both modalities provided one takes into account their differential decay rates. Duncan-Johnson and Donchin (1977) reported that sequence effects are additive with the effects of prior probability. Johnson and Donchin (1980) demonstrated that when the subject counts one of three equally probable stimuli the two uncounted stimuli enter the sequence with the probability of their category (i.e. P(uncounted) = .66). We could multiply such details from our laboratory and those of others. The story that emerges is rather complex, but the complexity in no way justifies Verleger's assertion that "Limits were reported on the validity of these findings (p. 4)" Perhaps Verleger is not quite precise in his use of the word "validity." If he means "generality," we are surely in agreement. Clearly, the sequential effects will be observed only if the circumstances are appropriate and it is very important to specify these limits. However, the existence of these limits does not reflect on the validity of the findings.

To consider one example of such limits: Verleger cites the finding by Duncan-Johnson and Donchin (1982) that "sequence effects were eliminated when a warning signal was interspersed that allowed for predicting the next stimulus above chance level p. 6)." This is not quite an accurate report of Duncan-Johnson and Donchin's results. The study presented subjects with an oddball task in which each of the stimuli in the series was preceded by a warning stimulus in all experimental conditions. However, in some series the warning stimulus provided no information about the probability with which the imperative stimulus on that trial would occur. These series are contrasted with series in which each warning stimulus provided information about the probability distribution of the imperative stimuli. In other words, in some series the sequence of stimuli was the subjects' sole source of information about the probability distributions; in others the sequence carried no such information because the distribution was determined afresh by each warning stimulus on each trial. Note that the structure of the sequence and the prior probabilities (over the entire series) were the same in both series.

The experiment in question was designed to test the hypothesis that the sequential effects are observed because the subjects are extracting information about the probability of events in the environment from the preceding sequence. The data demonstrated that stimuli do indeed affect the P300 elicited by a succeeding event if, and only if, the subject has no other source of information about the probabilities. This appeared to support rather nicely the model presented by Squires et al (1976), and enhanced, rather than cast doubt on, the validity of the basic observations on the sequential effects. Furthermore, the results are quite consistent with the suggestion that the P300 is a manifestation of a process involved in the maintenance of a model of the environment. The "flexibility" of the effect can be construed as damaging to the context updating model only in the framework of the model which Verleger labels "Update." If P300 requires (as Verleger implies we suggested) the updating of specific traces in a buffer - an updating with no relationship to a larger purpose - then indeed the relationship should show no flexibility. But, as will be seen below, and as should be evident from the literature, "Update" is really just a caricature of our model.

3.3. Task Relevance

Even though the effect of subjective probability discussed is powerful and has received a prominent share of investigative attention it was quite evident from the earliest studies that at least one other factor must be taken into account, namely, "task relevance." The simplest demonstration of this fact is the observation that no P300 is elicited when subjects "ignore" the very same series of events whose rare members elicit a large P300 when they attend to the events. For example, in the case of the the ERPs elicited by the stimuli used by Duncan-Johnson and Donchin (1977) when their subjects were solving a word puzzle, virtually no P300s are elicited by the rare events. Yet these very same stimuli, when counted, elicit large P300s. As Verleger states by way of a critique, "There is a target effect (p. 6)." On this we agree. But what is difficult to see is how this observation, which we consider a critical element in the evidence for the context updating model, is used by Verleger in support of his negative view.

The sensitivity of P300 amplitude to the allocation of the subject's resources to the task in which the eliciting stimuli are embedded has been examined by Donchin and his colleagues in an extensive series of studies attempting to develop a measure of mental workload based on the amplitude of P300 (Donchin et al 1986a; Gopher and Donchin 1986). The strategy in most of these experiments has been to require subjects to perform the oddball task concurrently with another "primary" task whose demands on their attentional resources could be varied. The results, most of which are ignored by Verleger, have been rather consistent. The more demanding the primary task, the lower the amplitude of P300 elicited by the concurrently presented oddball stimuli (see also Rösler 1983).

Here again, upon careful examination the pattern of results proves rather specific. Not every increase in demand by the primary task has an effect on the P300 that is elicited by secondary task stimuli. Isreal, Wickens, Chesney and Donchin (1980), have shown that to be effective the demands must be concentrated in the perceptual domain. At least, it was clear from their data that increasing the motor demands of the primary task does not affect P300. Kramer, Wickens and Donchin (1985) examined the nature of the resources which must be in demand for an effect to be observed; they have shown that the interactions can be controlled by manipulating the relationship in visual space between the stimuli associated with the two tasks.

The general implication of these data is that the processing of which P300 is a manifestation is sensitive to the use of a limited capacity system which can be deployed in the service of one task or another. Strong evidence in support of this suggestion was provided by Wickens, Kramer, Vanasse and Donchin (1983) see also Sirevaag, Kramer, Coles and Donchin 1986), who measured concurrently the P300s associated with both the primary and the secondary task. It turns out that as the difficulty of the primary task is increased the amplitude of these P300s changes in a reciprocal fashion. That is, whereas P300s associated with the secondary task decrease in amplitude, the P300s associated with the primary task increase. The sum of the two concurrently recorded P300s was constant over the range of difficulties used for the primary task.

As the data have been reviewed in detail elsewhere, and as they received but little of Verleger's attention, we will not examine the effect of task relevance in further detail. Suffice it to note that one *cannot provide a satisfactory account of the antecedent conditions for P300* without considering task relevance. An important analysis of the interaction between probability and task relevance has been provided by Johnson (in press), who reviews the evidence for a multiplicative relationship between these variables. Johnson's review highlights the fact that task relevance operates only to the extent that the subject extracts the relevant information from the stimulus. He suggests that the meaningfulness of the stimulus -- that is, its task relevance defined externally to the subject -- can only operate to the extent that the subject extracts the necessary data. This is an extension of Ruchkin and Sutton's (1978) observation that "equivocation" will reduce the amplitude of P300.

This well taken point notes that it is "subjective" task relevance, just as it is "subjective" probability, that is the critical variable. In other words, the locus of the effects we are studying is the subject's mind (or brain) rather than the apparatus that generates the experimental conditions. Probability is filtered by the subject's concept of the probability of events and information is effective only to the extent that it is retrieved, processed and used. This view is highly consistent with our view of P300 as the manifestation of processing which is very sensitive to the cognitive operations elicited by events when subjects perfrom certain tasks. Any view as narrow as "Update," or "Expect" in the form enunciated by Verleger is necessarily implausible.

3.4. The Latency of P300

Whereas the information about the effects of probability and task relevance that we discussed above played a critical role in the development of our interpretation of the P300, a different body of data, one to which Verleger pays little attention, served as the key element in the development of the context updating model. This is an extensive literature concerned with the latency of the P300. The interest in this problem was triggered by the obvious dissociation between reaction time and P300 latency. Here again, the empirical data are clear, but their interpretation has been controversial.

In 1987 it may be difficult to recall that the dissociation between the latency of the P300 and reaction time was at one time considered a major impediment to the interpretation of the component. Donchin et al (1978), Donchin (1984) and Tueting (1978) review the issue. There were two central concerns. First, conflicting data have been presented about the correlation between P300 latency and RT. Ritter et al (1972), showed quite convincingly that the two variables, when taken on a trial by trial basis, are positively correlated. Other investigators, however, failed to observe the correlation.

A second problem was the general expectation that P300 would precede the reaction. This expectation was based on the notion that processing begins with a stimulus and ends with a response. If P300 (so went the argument) is a "correlate" of information processing, it must take place before the response, or else one would have to say that it cannot reflect processing, because all processing was presumed to precede the response.

Both of these difficulties led Donchin (1979, see also Tueting 1978), to suggest that it may be best to:

> "postulate the existence of multiple, parallel processors, continually active. Stimuli impinge on this stream of processing and modulate and activate in different ways this complex matrix. Multiple responses are evoked by a given stimulus. Some with immediate consequences, others will not be manifested for some time, exercising their effect through changes in the subject's strategies rather than through their effects on the immediate responses. If P300 is associated with processes that are more related to strategic rather than to tactical information processing, then the degree to which it('s latency) correlates with RT will depend on the strategies the subject tend to adopt. A dissociation between RT and P300 latency becomes, in this context, far more interesting than associations between these variables. (Donchin 1984, p. 114)"

These assertions were presented concurrently with evidence reported by Kutas, McCarthy and Donchin (1977) that the correlation between P300 latency and RT varies with the level of accuracy demanded of a subject. This result is hardly surprising. If P300 is elicited by an event as a function of the category to which that event is assigned then P300 will be elicited only after categorization and this latency would be proportional to categorization time. The overt reactions whose time is measured may or may not be elicited on the basis of a full analysis of the stimulus. The prevalence of fastguesses when speed of reaction is stressed is well known.

Thus, the variance in RT is affected by many factors. Only a subset of these factors affects the latency of P300. A striking confirmation was provided by Renault and his coworkers (Ragot and Renault 1981), who found, in a choice RT task, that when subjects respond with their hands crossed their RTs are greatly increased. This source of a substantial portion of the variance in RT had no effect whatsoever on the latency of the P300. These data confirmed the proposal made by Kutas et al (1977) and by McCarthy and Donchin (1981) that P300 latency is insensitive to variables related to the execution of the response.

The pattern of the relationship between RT and P300 latency has been reviewed in detail elsewhere (see Hillyard and Kutas 1983; Duncan- Johnson and Donchin (1977), Donchin et al (1986b); and for more recent work extending this approach and combining it with studies of response preparation, see Coles and Gratton 1986; Gratton, Coles, Sirevaag, Eriksen and Donchin, in press). For our present purpose it is sufficient to note one critical implication of these data, namely, that the processing manifested by the P300 is apparently not necessary for the execution of the overt responses the subject must make to the stimulus. On occasion, especially when subjects are accurate, the responses are delayed with respect to P300 and the correlation between P300 latency and RT increases (see Kutas et al 1977 as well as Coles, Gratton, Bashore, Eriksen and Donchin, 1985). In other words, when the variance in RT is shared with the variance in P300 latency, the correlation between these two measures will be positive and fairly large. But, in the main, a dissociation between these two measures is observed frequently, suggesting that overt responses can be emitted without awaiting the processing manifested by the P300. It is these data that suggest most powerfully that the P300 is a manifestation of processing used in maintaining our model of the environment.

4. The context updating Model

4.1. Summary of preceding argument

The task we undertake, as we outline the context updating model, has been well summarized by Ro"sler (1983),

> "To make progress towards finding an exact definition of the functional state indicated by P300, one has to do something more than just manipulate input, context or output variables. Therefore, the additional step is to abstract from the directly observed variables and the particular experimental settings one common 'denominator', an intervening variable processing construct, which can be taken as the possible if one relies on another source of information besides the directly observable facts. One must add some assumptions about the information processing activities performed by the brain; in short, one must formulate, more or less explicitly, a cognitive theory. (p. 17)"

In the immediately preceding sections of this precommentary we outlined the "directly observable facts." Before we outline the "cognitive theory" that may account for these facts we briefly summarize those pieces of the puzzle we consider crucial to the evolution of our model.

(1) The amplitude of P300 is controlled, in a multiplicative fashion, directly by the task-relevance of the eliciting event and inversely by the subjective probability of the event. We note here that even though Verleger argues that "P3s are not evoked by unexpected stimuli," his argument consists of a conjecture about the psychological state of the subject. That is, Verleger is arguing that rather than being surprised by the eliciting stimuli the subject is actually awaiting those stimuli. There is very little evidence, however, that would support speculations about the subjects' state of mind at the time the eliciting stimuli are presented. The closest any one has come to directly measuring the subjects' psychological states in these experiments have been the assessment of their confidence in their predictions, guesses or judgments (e.g. Hillyard et al 1971; Horst et al, 1980) and, as we noted above, these suggest that surprising events do elicit a larger P300. However, whatever one concludes with respect to this issue, there is hardly room for doubt concerning the strictly empirical assertion that, in a range of experimental parameters, reducing the probability of an event will increase the amplitude of the P300 elicited by the event. This effect of probability is tempered by its multiplicative interaction with the effect of task-relevance (Johnson, in press).

(2) Overt responses required of the subject on a given trial can be emitted prior to the appearance of the P300. If the response reflects a discrimination among stimuli, then responses emitted prior to the appearance of the P300 are more likely to be inaccurate. The correlation between the latency of the P300 and the RT, computed over trials, is not necessarily significant. Thus, different factors appear to control the variance in these two dependent measures. In other words, the P300 is relatively decoupled from the motor output the subject must generate in response to the eliciting event.

(3) The P300 seems to manifest activity in a limited capacity system whose use in the service of different tasks is under relative control by instructions. There is evidence that the total amplitude of P300 in the presence of concurrent tasks is constant.

(4) The amplitude of P300 seems to be sensitive to the intervals between successive stimuli. One can account for the sequential effects in P300 amplitude by assuming that it is inversely proportional to the decay of the representation established by preceding occurrences of the eliciting events. Furthermore, the ubiquitous effect of probability on the amplitude of the P300 in an oddball paradigm is drastically reduced, if not eliminated, when the interval between events is increased (Heffley, Wickens and Donchin, 1978; Fitzgerald and Picton 1981).

4.2. The Model

As we proceed to develop a "cognitive" theory to identify the processing function manifested by the P300 taking the above into account, we find the dissociation between P300 latency and RT to be the most compelling piece of the puzzle. It is a particularly important datum because it eliminates from consideration the rather large class of information processing activities involved in selecting and executing the overt response to the event which elicits the P300. The idea at the core of the context updating model is that P300 is invoked in the service of "strategic" rather than "tactical" information processing (Donchin et al 1978). According to the strategic/tactical distinction information is processed in parallel interacting streams. The "tactical" stream is concerned with the processes that select a specific set of responses at any instance, that is, it controls overt behavior in an immediate sense: actions. The "strategic" stream is involved in the planning and control of behavior, such things as the long term setting of priorities, the setting of biases, the mapping of probabilities on the environment or the deployment of attention. These are merely illustrations of the vast realm of metacontrol processes that must be implemented as the infrastructural support of overt behavior. The specific physical events occurring on a trial, be they stimuli or responses, serve as data for this stream of processing. We suggest that the P300 is a manifestation of processes in the strategic stream.

Note that the evidence pointing to such a meta-control role for the P300 is strong even though it yields no clue to the specific function of the process of which the P300 is a byproduct. The key to the puzzle contributed by the latency data is that it directs our attention to processes which are not critical to the immediately executed response. Three aspects of the data supply this clue. First, the latency of P300 often exceeds the RT. The obvious conclusion is that the process it manifests is not a necessary condition for the emission of the response. Second, the correlation between P300 latency and RT depends on the level of accuracy the subject is striving to achieve. This suggests that the elicitation of P300 depends on a more extensive extraction of information about the event and on the processing of that information. Finally, we note that the latency of the P300 appears to be unaffected by processes associated with the execution of the response (McCarthy and Donchin 1981). This again suggests that the function manifested by the P300 is associated with the evaluation of stimuli.

When we use such words as "associated" or "reflects" in referring to the relationship between P300 and, say, "stimulus evaluation," we do not intend to imply that the P300 is in fact a byproduct of stimulus evaluation. Such a suggestion would be inherently invalid because stimulus evaluation is not a unitary processing entity but a multiplicity of processes. The data on P300 latency allow a choice between two classes of processing functions - those involved in the decoding and assessment of information and those associated with the execution of overt responses. The P300 seems to be independent of the latter category. It seems reasonable to assume that its function once identified will belong to the former category, which we loosely label "stimulus evaluation." Additional support for this view comes from the studies of the behavior of P300 in dual-task studies. Isreal et al (1980), as well as Kramer, Wickens and Donchin (1983), have reported that to be effective in reducing the amplitude of the P300 elicited by secondary task stimuli the load of the primary task must be increased in the perceptual rather than the motor domain. Thus, the general class of processes of which the P300 is a manifestation probably involves a perceptual evaluation. The specific selection in this category, however, must be made on the basis of additional data. Once we accept P300 to be associated with "strategic," future oriented, processing we must seek a more precise definition of the specific process.

It is at this point that the fact that the P300 is often associated with events whose probability is low begins to affect our thinking. What function could one reasonably ascribe to a processor activated by rare events, one whose amplitude is directly proportional to rarity? Two assumptions enter our interpretation of these observations. Both are tenuous and are justified largely by their plausibility and their heuristic value. The first assumption is the weakest link in our entire structure. It is that the amplitude of the P300 is a measure of the extent to which the processor manifested by the P300 is activated, or "utilized." This assumption is basic to the ERP literature and Verleger appears to accept it. It must be admitted, however, that we have no direct physiological evidence for this or any other interpretation of changes in the amplitude of the P300. It is plausible to assume that the larger the potential the larger the voltages generating the field which is being recorded and that this size is a measure of the activation of the tissue from which we record. This, unfortunately, is not necessarily the case as Allison et al (1986) point out. One hopes that the various attempts to identify the intracranial sources of the P300 will provide data that will allow an interpretation of changes in the amplitude of the P300 (e.g. Halgren, Stapleton, Smith and Altafullah 1986; Wood, McCarthy, Squires, Vaughan, Woods, and McCallum 1984; Okada, Kaufman and Williamson 1983). For the time being, however, we must admit that the assumption is entertained primarily for its heuristic value and its merit must be evaluated according to its utility as a guide for future research.[7]

The second assumption we make here is a cognitive one: A class of metacontrol processes is concerned with maintaining a proper representation of the environment. The concept of a "model of the environment" is of course not new. It has assumed many guises, ranging from the Sokolovian "neuronal model" through contemporary notions of "schema" and "mental models" (Johnson-Laird 1983). [See also Kosslyn et al.: "On the Demystification of Mental Imagery" *BBS* 2(4) 1979 and Arbib: "Levels of Modeling of Mechanisms of visually guided Behavior" *BBS* 10(3) 1987.] There are very meaningful differences among all these theoretical concepts. However, they have one general idea in common, namely, that part of the totality of representations and skills potentially available to the organism at any time is in a state of heightened availability and that this high-availability segment is sensitive to the conditions of the environment (the "context") and to the needs imposed by the tasks at hand.

The idea that a model of the environment is used in evaluating incoming information and in selecting responses is quite common among cognitive theorists. The level of detail with which such concepts are presented varies and the specificity with which models are articulated is also quite variable. However, it is really not critical at this stage to specify how the "contextual model" is implemented. Although this is an interesting subject, we need not choose among the many candidates. The only assumption we need to make is that, however the model is implemented, there must be mechanisms that will maintain it as an accurate model. No model (or schema) would be of any use unless it was possible for it to adjust dynamically to the environment. When the "context" changes, the model must be revised. A model lacking an updating component would be useless because it would fail to meet its most basic requirement, namely, that it should reflect the ongoing context. As contexts are continually in flux, the model must undergo constant revisions. Novelty, surprise and the occurrence of improbable events must somehow be integrated into it, either by revising its mapping of probabilities on the environment or by rejecting the importance of the event and remaining unchanged.

The sensitivity of P300 to the probability of events adds plausibility to the suggestion that it is associated with maintaining the schema. Note that whether the decisive fact is that the stimulus is "unexpected" or is that it is associated with "closure" makes no difference at this stage of the argument. In either case we must hypothesize a functional role for the manifested process. We suggest that the process will be one involved in the maintenance of the schema, or the model, of the current context. This involves no commitments about how this goal is achieved; it leaves the details of the operation unspecified. We note, however, the importance of "task relevance" in the elicitation of the P300. This fact must constrain speculations about the underlying processing.

[7] It should be noted, though, that the assumption does gain credence from the convergence of the very large number of studies conducted under its guidance.

Clearly, not all improbable changes in the environment lead to the updating of the context. Only those segments of the context that are central to the tasks performed by the subject are likely to bring about the changes in the model of the environment. The observed changes are associated with an optional limited-capacity system. The constancy of the amplitude of P300 across tasks as the demands of the tasks are changed reciprocally (Wickens et al 1983) is strongly supportive of the association of the process with a limited capacity system. Two studies by Johnson and Donchin (1978 1982) illustrate the considerations that enter into the evolution of this model.

Johnson and Donchin (1978) asked subjects to press a button exactly one second after a light flashed. The subject was informed on each trial by one tone if his time estimate was correct and by another it was not. The intensity difference between the tones was the independent variable in the experiment. It turned out that the larger the difference between the tones, the larger the P300 they elicited. The very same tones, when used in an oddball task (that is, when the subject was instructed to count the number of tones in a series), elicited a P300 whose amplitude was independent of the intensity difference between them. Johnson and Donchin interpreted these data as indicating that the amplitude of P300 is proportional to the extent that the stimuli provide the subject with information that can help in the performance of the task. However, the degree to which the information is utilized is to an extent under the subject's control. During the time estimation task the extraction of information from the tones was not mandatory and the information was less likely to be used if the discrimination between the two tones placed a higher load on the subject. When directly instructed to count tones the subjects had no such option and the tones were used regardless of their discriminability. Within the framework of context updating we would suggest that the subjects maintained, as part of their model of the context, a basis (a template?) against which they estimated time. When the information provided by the tones was useful in maintaining and revising this model the tones were so used, as indicated by the large P300. This interpretation was bolstered by the fact that the lower the P300, the more errors the subjects committed and the less likely they were to improve.

In a subsequent study, Johnson and Donchin (1982) presented subjects with an oddball series in which the probabilities were changed in midcourse. The series might start with the probability of the high tone at .33 and then, in 40 to 80 trials, the probabilities of the two tones would be reversed so that the probability of the high tone became .66. This sequence of reversals continued over a long series so that when viewed over the entire series the probabilities of the two stimuli were essentially equal. There were two experimental conditions. The subjects were first run without knowing that the probabilities were changing. They were told to count high (or low) tones as they do in all other oddball paradigms. In the second condition the subjects were told that there was a continuing sequence of reversals in the probability of the two tones and they were assigned the task of detecting when the probabilities actually changed. The aspect of the result relevant to the present discussion concerns the behavior of the P300 during the "transition" periods which were defined as the segment of the series between the time the computer actually adopted the reversed probability and the time the subject detected the probability shift.

As Johnson and Donchin used identical stimulus sequences in the two conditions it was possible to compare the P300 elicited when the subjects knew about the reversals and when they did not. When the P300 is elicited on the five trials immediately preceding the subject's report of the change, it turns out that during the unwitting sequences the amplitude of the P300 is determined entirely by the immediately preceding sequence of stimuli. On the other hand, when the subject was trying to identify the time at which the operating parameters of the environment changed, the amplitude of P300 in those five trials increases gradually from trial to trial, reaching a peak just prior to the subject's report that the probability has changed. Immediately after this report the enhancement of P300 disappears and the amplitude is determined entirely by the global probability and the sequential dependencies as reported by Squires et al (1976). In other words, as the stimulus delivers information that is usable in the subject's structuring of the task relevant portion of the environment, the P300 is enhanced. As the evidence accumulates and the pressure for accepting the change increases, so does the amplitude of the P300, reaching a peak just at the time the change is accepted and the subject is made sufficiently aware of the change to report it.

In summary, the context updating model asserts nothing more than that the P300 is elicited by processes associated with the maintenance of our model of the context of the environment. The association of the P300 with novelty, its sensitivity to the decaying strength of representations, its dependence on a limited capacity system and its strong dependence on the relevance of the eliciting event to the task lead to the suggestion that the processing it manifests is invoked whenever there is a need to revise the organism's model of the context. We are assuming that the larger the amplitude of the P300, the larger the change in the model. This statement of our model is admittedly not precise enough. We have no commitments about the nature of the system that implements the contextual model or the process by which the context updating is implemented. However, we think it is premature to offer an answer to these questions. We need to obtain much more evidence on the P300 and its relation to context updating before we can begin to discuss such details. However, even at this stage the context updating model is powerful enough to generate interesting experimental predictions about the consequences of the P300.

4.2.1.

In the next section we will examine some of these predictions. First, however, we must examine Verleger's description of the context updating model. Two versions of the model are subjected to Verleger's critique. The first, the "Update" version, is arrived at by a tortuous route which leads Verleger from a statement of the model as presented above to an insistence that we specify "what exactly was meant by 'expectancies,' 'strategies,' 'schema,' 'model,' and 'context.' (p. 4)" The definitions are given whenever we conduct an empirical investigation. For example, "expectancy" was operationally defined by Squires et al (1976). In any event, Verleger proceeds to use out-of-context remarks about working memory and representations to produce the following version of our model

> "Each stimulus has its representation ('trace') in working memory, which decays over time. After stimulus evaluation, the trace of the presented stimulus is updated to its optimal upper-limit value and it is this updating that is reflected by P3. p. 5"

Verleger is clearly attributing much more specificity to us than we feel competent to provide. We have definitely not described a direct association between P300 and any specific operation involved in the maintenance of the contextual model. We certainly would not claim to know that P300 is a manifestation of the updating of traces. In the present state of our knowledge we cannot go beyond the more general suggestion that P300 is invoked because of the need to update the context. The narrow definition according to which Verleger focuses on the very act of trace updating leads him to see inconsistencies where none exist. Thus, for example, the flexibility of the sequential effects is only a problem if one assumes that P300 is uniquely related to trace updating. This poses no problem, however, if the need to update the context arises in some circumstances and not in others. Surely Verleger's discovery that "The largest P3s were evoked by the last relevant item of every trial" is precisely what has been reported by Donchin and Cohen (1967), or by Rohrbaugh, Donchin and Eriksen (1974). We suggest that these events are the ones impelling the system to update the model of the context.

A somewhat closer approximation to our views is given by Verleger when he discusses his "Expect" version of context updating. He acknowledges that the model as presented by Donchin (1981) "goes beyond Update's interpretation of expectancies being representations of stimuli in working memory." He sees that we are suggesting "strategies and expectancies may involve working memory, but they do not hold any privileged relation to working memory." However, he immediately proceeds to narrow the definition so that it is again a caricature, rather than a valid portrayal, of our model. Verleger's strategy as he discusses "Expect" is to suggest that subjects "expect the frequent stimulus" and are consequently "surprised" by the rare stimulus. First portraying the context updating model as an claim about the relationship between violations of expectancies and P300, Verleger proceeds to use the substantial body of data suggesting that stimuli need not be entirely unexpected to counter his own version of our model. This is as misleading as his treatment of "Update." We have repeatedly stated that low probability is neither a sufficient nor a necessary condition for eliciting P300 or for controlling its amplitude. It is inherent in the multiplicative relationship between rarity and task relevance that stimuli with high relevance will elicit a large P300 even when highly probable.

4.3. Predictions from the Model

4.3.1. The amplitude of P300 and response bias

The most important corollary of the context updating model is that it directed our attention to the study of the possible effect of eliciting P300 on subsequent performance through the restructuring of the subject's model of the environment. This approach was guided by the realization articulated earlier in this precommentary that theories about an ERP component must be tested by assessing the consequences of eliciting the component. If the process manifested by the component is involved in the restructuring of the subject's model of the environment we ought to be able to measure effects that can be attributed to such changes and their magnitude should be proportional to the amplitude of the P300.

As an illustration, consider the experiment reported by Donchin, Gratton, Dupree and Coles (in press). In this study names were displayed on a screen monitored by the subjects, who were instructed to press one of two buttons in response to names of males and the other button in response to names of females. Female names appeared 80% of the time and subjects, who were urged to perform quickly, were strongly biased to press the button indicating a female name. In general, subjects hardly ever erred in response to the female names whereas they often pressed the wrong button in response to a male name. These erroneous responses were clearly "fast guesses." The ERPs elicited by the rare male names showed an interesting pattern. Their P300 latency was about 100 msec greater than that of rare names correctly responded to. In other words, a fast guess delayed the P300 by about 100 msec.

What was the subject doing in these 100 msec? Obviously, no direct answer is available. It is plausible, however, that this extra period was used to assess the consequences of the error trial so that proper adjustments could be made to the operating context to minimize such errors in the future. We reasoned that if the P300 was, as our model assumes, a manifestation of processing associated with the adjustment of the operating context, then it is reasonable to predict that the larger the P300 the larger the adjustment. The most likely adjustment after the recognition of an error would be a revision of the biases so that errors would be less likely in the future. "If this hypothesis is correct then the characteristics of the P300 elicited on error trials should predict variations in response criteria in the following trials" (Donchin et al., in press).

We accoprdinglly examined all the trials on which the rare stimulus was presented and on which the subjects erred and sorted these according to their response on the trial on which the next rare name was presented. There could be pairs in which an error was followed by an error and pairs in which the error was followed by a correct response. The critical finding was that the amplitude of the P300 elicited by the first member of the pair was larger if the response to the second member was correct. On the assumption that the subject responds in error to male names because he is biased to respond to the more frequent female names, we propose that a correct response to the male names implies that the subject has adjusted the bias as a consequence of processing associated with the error. It is not implausible to suggest that the amplitude of the P300 is a manifestation of a process associated with this adjustment. This interpretation received further confirmation from the observation that the

RT associated with the first female name after a male name is inversely proportional to the amplitude of the P300 elicited by that male name.

Note that we are not suggesting that the P300 is a manifestation of the bias adjustment per se. The precise relationship between bias adjustment and the process manifested by the P300 is undetermined at this point. All we are claiming is that in the complex of processes triggered by the recognition that an error has been committed (which appears to delay the elicitation of the P300), an operation whose effectiveness is correlated with the amplitude of the P300 takes place and this operation predicts rather reliably whether or not the subject will commit yet another error.

4.3.2. Perfect Pitch and P300

Another prediction derived from our model was tested by Klein et al (1984). We predicted that subjects with perfect (or "absolute") pitch would not display a P300 when tested with an auditory oddball series but would show a P300 when tested with a visual one. The results confirmed our prediction. We did not invent the idea that such subjects maintain permanent templates in their memory for acoustic stimuli. This happens to be a common account of the Perfect Pitch phenomenon. As Klein et al noted :

> "The weight of the evidence suggests that individuals with Absolute Pitch (AP) have access to a set of internal 'standards' that allows them to fetch the name of a tone without comparing the representation of the tone to a recently fetched representation of a standard (Siegel 1974)."

Lockhead and Byrd (1981; see also Harnad 1987) review the relevant literature and describe a method for assessing the degree to which an individual does indeed have this skill. We arrived at the prediction that AP subjects would not use the mechanism manifested by the P300 by combining this common account of AP skill with our current view of P300. We proceeded to test music students, with and without AP, and the data clearly confirmed the prediction.

As we noted earlier, Verleger overlooks the fact that all subjects were music students and that the amplitude of P300 was proportional to each subject's actual skill as an AP perceiver rather than to the subjects' reports about his skill. Verleger also asserts:

> "If tones have permanently resident representations for AP subjects, why not other well-known stimuli such as colors, the words 'push' and 'wait,' synonyms and names? Since these words evoke P3s Klein et al assumption is improbable, p. 8"

Verleger seems unwilling to accept the overwhelming evidence that people with AP have a rather unique ability that cannot be explained simply by assuming that they have overlearned a specific set of well-known stimuli. As we tested our subjects in the way Lockhead and Byrd did the differences struck us with remarkable force. The subjects without AP were all highly qualified, advanced, music students. They had excellent relative pitch judgment. That is, if allowed to hear a standard tone and given its name, they could proceed to name any other tone with accuracy. However, without such prior exposure to the standard, subjects with relative pitch could only name correctly a small percentage of the tones. Subjects with AP named between 82% and 94% of the tones without access to a reference tone. It is inappropriate to compare the way AP subjects process tones to the way we process other "well-known" stimuli. The evidence suggests that the AP skill relies on a unique comparison with a permanently available standard. Our data show that these subjects, as predicted by the context updating model, draw on the processes underlying the P300 to a significantly smaller extent that they do when processing visual stimuli.

We note in passing that in the course of discussing the empirical data Verleger makes a number of statements that suggest he may have a rather limited model of the cognitive system. For example, in his discussion of Klein et al's paper, he wonders why the subjects appear to make the visual comparisons by relying on a retrieved standard. He than writes:

"If Klein et al's subjects were unable to form the abstract categories 'H' and 'S' but had to store the specific physical features, how can other subjects combine stimuli into targets and nontargets irrespective of physical features? Since this is obviously possible, the assumption is also improbable. p. 8"

Here again Verleger does not quote us accurately. We did not say that subjects were unable to form the categories 'H' and 'S.' Rather, we suggested that when a displayed character is assigned to one or another category an encoding process must precede the categorization and must use a feature detection scheme that takes into account the large variety of fonts and shapes in which the characters may be presented. That such a feature-detection stage, possibly preattentive, precedes the higher-level categorizations has been demonstrated rather persuasively. For a comprehensive review of much of the relevant data the reader is referred to Treisman (1986). It is also becoming increasingly clear that much of the data on categorization can be interpreted without assuming that subjects form "abstract categories" represented by prototypes; they may be using an exemplar based scheme that is very dependent on feature encoding schemes (Medin and Wattenmaker 1987; see Harnad 1987 for further kinds of representational model). Thus, it seems the cognitive assumptions of Klein et al are consistent with at least some current theories of categorization.

Equally strange is Verleger's notion of Working Memory. We quote in full, so the reader can assess Verleger's line of thinking:

"Working Memory (WM) contains anything that has happened during the previous, say, ten seconds. Then WM will be relatively empty when a trial starts. Storing the first stimulus in WM will change WM content from zero to one, storing the second stimulus will change its content from one to two, storing the last word of a seven word sentence will change its content from six to seven. p. 10"

Verleger appears to view WM as a simple push down stack which can be "relatively empty" at any time and into which simple representations of stimuli are pushed, and from which they are popped, as they are presented and used. This is, of course, a rather extreme simplification of the WM model of those who have been using the concept. There is almost no relation between Verleger's WM and the dynamic and complex system proposed, for example, by Baddeley (1987), in which executive controllers oversee and integrate the activity of articulatory loops and visual scratch pads, all the while monitoring the environment for relevant inputs. Our own view of the system was summarized by Donchin (1981) as follows:

"The schema may be conceptualized as a large and complex map representing all available data about the environment. It is the reservoir of information that is necessary for performing whatever tasks require active processing at any time. Some of the information has just been delivered and may still reside in various stages of dynamic memory. Other information resides in longer term memory and may need to be made available, on a temporary basis, for integration into the overall schema. The system is quite fluid. There must be some priority weighting system that is associated with the schema and determines which of its aspects is related to which tasks. Representations decay because of misuse or because of shifting strategies and tasks. New information is brought in. Choices are made in the process of using this schema. When there is need, the model is revised by building novel representations through the incorporation of incoming data into the schema based on long term memory data. It is likely that it is this updating process that we see manifested by the P300. p. 508."

4.3.3. P300 and the von Restorff effect

It is this more dynamic view of 'working memory' that we espouse as we consider the functional explanation of the patterns of P300 data and it is this approach that guided the last set of predictions about the P300 we will discuss. We refer to studies of the relation between the P300 amplitude elicited by stimuli and the probability that those stimuli will be recalled. Donchin (1981) noted that there may be a relation between theories of memory predicting that distinct items will be best remembered and the fact that the P300 is elicited by distinctive events. It accordingly seemed promising to examine the extent to which the amplitude of P300 -- interpreted either in this framework, as a measure of distinctiveness, or, alternatively, as a measure of the restructuring of the current model of the environment -- might predict the probability that items would be remembered.

The paradigm that proved particularly effective for testing this prediction was devised by von Restorff (1933). She presented subjects with a series of items, some distinct in some respect. These "isolates," she reported, were remembered far better than the other items in the series. Karis et al (1984) reasoned that these isolates would, by virtue of their very distinctiveness, elicit a P300. They further reasoned that the amplitude of these P300s would vary and that this variance would be correlated with the subjects' ability to recall the isolates.

To test this prediction Karis et al presented each subject with 40 series of 15 words each. In 30 of these lists one of the words, in one of the middle positions, was displayed in characters that were either larger or smaller than the characters used for the other items in the list. The subjects were asked to recall the items immediately after each list was presented, as well as at the end of the entire series of presentations. At a later time, the subjects' ability to recognize the previously presented words was also tested. Average ERPs were obtained separately for the recalled and unrecalled words. The pattern of results was rather complex. For present purposes it is sufficient to note that the amplitude of the P300 was indeed larger for recalled items. However, this was in the main true only for those subjects who used rote memorization to aid their recall. The relation between P300 and recall did not hold for subjects who used elaboration techniques as a tool to aid their recall.

The interaction between recall strategies on the one hand and the P300/recall relationship on the other was confirmed in a subsequent study by Fabiani et al (1986), who replicated the Karis et al study, except that this time the subjects were explicitly instructed to use one strategy or the other. All subjects were instructed (in a properly counterbalanced way) to use all strategies. It was found that the aforementioned interaction between P300, recall strategies and recall holds in subjects. In another experiment, Fabiani, Karis and Donchin (1986) used an incidental learning paradigm in which subjects had no reason to use elaborative strategies. For all subjects the amplitude of the P300 elicited by subsequently recalled items was larger than the amplitude elicited by items that were not recalled. Paller, Kutas and Mayes (1987) report similar results using a "level of processing" paradigm. Confirming findings were also reported by Fox and Michie (1987).

To what extent do these data provide support for the context updating model? If the model is construed as an claim about the class of functions the P300 is most likely to manifest, then the strong relationship between P300 amplitude and subsequent recall is certainly consistent with the suggestion that the process underlying the P300 is involved in the creation and maintenance of representations. In its simplest form this assertion anchors the P300 in the family of processes that make "distinctiveness" so powerful a factor in determining recall and recognition of items. Even more important is the interaction between P300 and the subject's recall strategies. The importance of this result is that it places constraints on the range of functions that might be associated with the P300. When stripped of the necessary complexities, the fact that the effect we observe is restricted to subjects who use rote memory suggests that the interaction in which the P300 process plays a role must take place at one of the earlier levels of stimulus encoding. It is the representation of the actual physical stimulus, rather than some subsequently encoded representation, that is affected by the intensity with which the process manifested by the P300 is activated. Fabiani et al (in preparation) tested a corollary of this interpretation. They assessed the degree to which subjects actually recall the size of the original stimuli presented during the study period. It turns out that when subjects use rote memory they are more likely to recall the original size of the word than are the very same subjects when they "elaborate." The implication is clear - the P300's effect is exercised at the early representational stage of stimulus encoding.

The details of the process need further elaboration. Here, as in previous discussions, we do not find it necessary to push the theoretical analysis beyond the depth needed to move ahead with our experimental program. Indeed, we view theory largely as a guide, a very rough guide, to our experimental program. Rather than propose all-encompassing systems as accounts of the P300, systems in which every item is fully defined and every operation entirely predictable, we prefer theories that at any stage of the research are sufficiently precise to allow the next stage of the program to establish itself.

5. Summary

In conclusion, we remain persuaded that our approach to the study of the endogenous components of the ERP is sound. Our goal is to understand the functional significance of these components. In this precommentary we focused on the P300 because Verleger's critique was concerned with our views of this component. However, our remarks are more general in their intent. We believe that accounts of these phenomena should take the form of hypotheses about the information processing functions of the activities manifested by the component. These should be derived from a knowledge of the antecedent conditions of the components; however, an enumeration of these conditions, no matter how general, does not in itself constitute a theory of the component. Once a theory is proposed, it should be assessed, not by arguments about the plausibility of post hoc explanations, but by predictions about the consequences of the component.

Verleger's two versions of our model do not represent our views adequately because they are based on a very narrow and simplistic construal of terms such as "expectancy" and "working memory." The theory Verleger proposes is not a theory but rather a generalization about the antecedent conditions of the P300. The target article also ignores rather important segments of the literature that are crucial to an understanding of the P300, and those studies that are discussed are mostly misrepresented and misinterpreted.

In this precommentary we present the context updating model as we actually construe it and we review some of the studies that have been generated by our version of the model. It will be seen that whereas the model is indeed vague about the specific mechanisms manifested by the P300, it is specific with respect to the classes of functions likely to be manifested by the P300. More important, the model has served as a tool, guiding our empirical investigations into a number of exciting research avenues and leading to a number of interesting discoveries. This, to us, is the ultimate test of a successful model.

Acknowledgmets

We gratefully acknowledge the comments of Bill Gehring, Gabriele Gratton, Monica Fabiani, Larry Farwell, Greg Miller and Erik Sirevaag as well as of Stevan Harnad. This paper was prepared while the first author was a Visiting Professor of Psychology at New York University and the support of the Department of Psychology and the Neuromagnetism Laboratory at NYU is gratefully acknowledged. Support was also provided by AFOSR (Contract #F-49620-83-0144, Al Fregly, Project Director) and by NIMH (Grant #1R01-MH41445).

References

Allison, T. (1984) Recording and interpreting evoked potentials. In: Cognitive Psychophysiology. ed. E. Donchin. pp 1-36. Erlbaum Associates.

Allison, T., Wood, C. C., & McCarthy, G. (1986) The central nervous system. In: Psychophysiology: Systems, Processes, and Applications, ed. M.G.H. Coles, E. Donchin, & S.W. Porges. pp 5-25.

Guilford Press.

Arnheim, R. (1986) New essays on the psychology of art. University of California Press.

Audley, R.J. (1973) Some observations on theories of choice reaction time: Tutorial review. In: Attention and Performance IV. ed. S. Kornblum. Academic Press.

Baddeley, A. (1987) Working memory. Oxford University Press.

Campbell, K. B., Courchesne, E., Picton, T. W., & Squires, K. C. (1979) Evoked potential correlates of human information processing. Biological Psychology 8:45-68.

Churchland, P. S. (1987) Neurophilosophy: Toward a unified science of the mind-brain. MIT Press.

Coles, M. G. H. & Gratton, G. (1986) Cognitive psychophysiology and the study of states and processes. In: Energetics and human information processing. ed. G. R. J. Hockey, A. W. K. Gaillard, & M. G. H. Coles. pp. 409-424. Martinus Nijhof.

Coles, M. G. H., Gratton, G., Kramer, A., & Miller, G. A. Principles of signal acquisition and analysis. In: Psychophysiology: Systems, Processes, and Applications. ed. M.G.H. Coles, E. Donchin, & S.W. Porges. pp 183-221. Guilford Press.

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.

Cracco, R. Q. & Bodis-Wollner, I. (1986) Evoked potentials. Liss.

De Swart, J.H., Kok, A. & Das-Smaal, E. A. (1981) P300 and uncertainty reduction in a concept identification task. Psychophysiology 18:619-629.

Donchin, E. (1979) Event-related brain potentials: A tool in the study of human information processing. In: Evoked brain potentials and behavior. ed. H. Begleiter. pp. 13-88. Plenum Press.

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

Donchin, E. (1984) The dissociation of electrophysiology and behavior: A disaster or a challenge? In: Cognitive Psychophysiology. ed. E. Donchin. pp. 107-118. Erlbaum Associates.

Donchin, E. & Cohen, L. (1967) Average evoked potentials and intramodality selective attention. Electroencephalography & Clinical Neurophysiology 22:537-546.

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. To appear in: Neurophysiology and psychophysiology: Experimental and clinical applications. ed. M. Kietzman, G. Galbraith, & E. Donchin. Erlbaum Associates.

Donchin, E., Karis, D., Bashore, T., Coles, M. G. H., & Gratton, G. (1986b) Cognitive Psychophysiology and human information processing. In: Psychophysiology: Systems, processes, and applications. ed. M.G.H. Coles, E. Donchin, & S. W. Porges. pp. 244-267. Guilford Press.

Donchin, E., Kramer, A. F., & Wickens, C. (1986a) Applications of brain event- related potentials to problems in engineering psychology. In: Psychophysiology: Systems, processes, and applications. ed. M. G. H. Coles, E. Donchin, & S. W. Porges. pp. 702-718. Guilford Press.

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, W. C. (1978) Cognitive psychophysiology: The endogenous components of the ERP. In: Event related potentials in man. ed. E. Callaway, P. Tueting, & S. H. Koslow. pp. 349-412. Academic Press.

Duncan-Johnson, C. C. & Donchin, E. (1977) On quantifying surprise: The P300 component of the event-related brain potential as an index of information processing. Psychophysiology 14:456-457.

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.

Ebner, A., Schutz, K., & Lucking, C. H. (1986) The late positive complex (P300) in normal subjects with and without absolute pitch. The Third International Evoked Potentials Symposium, Berlin-West.

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 potential. In: Advances in psychophysiology, Vol II. ed. P. K. Ackles, J. R. Jennings, & M. G. H. Coles. JAI Press.

Fabiani, M., Karis, D., & Donchin, E. (1986) P300 and recall in an incidental memory paradigm. Psychophysiology 23:298-308.

Falmange, J. C. (1965) Stochastic models for choice reaction time with applications to experimental results. Journal of Mathematical Psychology 2:77-174.

Fitzgerald, P. G. & Picton, T. W. (1981) Temporal and sequential probability in evoked potential studies. Canadian Journal of Psychology 355:188-200.

Fox, A. M. & Michie, P. T. (1987) P3 and memory relationship under instructed learning strategy conditions. Fourth International Conference on Cognitive Neuroscience. Dourdan, France.

Friedman, D., Hakerem, G., Sutton, S., & Fleiss, J. L. (1973) Effect of stimulus uncertainty on the pupillary dilation response and the vertex evoked potential. Electroencephalography & Clinical Neurophysiology 34:475-484.

Glaser, E. M. & Ruchkin, D. S. (1976) Principles of neurological signal analysis. Academic Press.

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.

Gopher, D. & Donchin, E. (1986) Workload - An examination of the concept. In: Handbook of perception and human performance, Vol II. ed. K. Boff, L. Kaufman, & J. P. Thomas. pp. 41.1-41.49. John Wiley and Sons.

Gratton, G., Coles, M. G. H., Sirevaag, E., Eriksen, C. W., & Donchin, E. (In press) Pre- and post-stimulus activation of response channels: A psychophysiological analysis. Journal of Experimental Psychology: Human Perception and Performance.

Halgren, E., Stapleton, J. M., Smith, M. & Altafullah, I. (1986) Generators of the human scalp P3(s). In: Evoked potentials. ed. R. Q. Cracco & I. Bodis-Wolner. pp. 269-284. Liss.

Harnad, S. ed. (1987) Categorical perception: The groundwork of cognition. Cambridge University Press.

Heffley, E., Wickens, C., & Donchin, E. (1978) Intramodality selective attention and P300--Reexamination in a visual monitoring task (abstract). Psychophysiology 15:269-270.

Hillyard, S. A. & Kutas, M. (1983) Electrophysiology of cognitive processing. Annual Review of Psychology 34:33-61.

Hillyard, S. A., Squires, K. C., Bauer, J. W., & Lindsey, P. N. (1971) Evoked potential correlates of auditory signal detection. Science 172:1357-1360.

Horst, R. L., Johnson, R. Jr. & Donchin, E. (1980) Event related brain potentials and subjective probability in a learning task. Memory and Cognition 8:476-488.

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 22:212-224.

Johnson, R. Jr. (1986) A triarchic model of P300 amplitude. Psychophysiology 23:367-384.

Johnson, R. Jr. (In press) The amplitude of the P300 component of the event-related potential: Review and synthesis. In: Advances in psychophysiology, Vol III. ed. P. K. Ackles, J. R. Jennings,

& M. G. H. Coles. JAI Press.

Johnson, R. Jr. & Donchin, E. (1978) On how P300 amplitude varies with the utility of the eliciting stimuli. Electroencephalography & Clinical Neurophysiology 44:424-437.

Johnson, R. Jr. & Donchin, E. (1980) P300 and stimulus categorization: Two plus one is not so different than 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.

Johnson-Laird, P N. (1983) Mental models. Harvard University Press.

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 perfect pitch process tones without producing a P300. Science 223:1306-1309.

Kramer, A. F., Wickens, C. D., & Donchin, E. (1983) Analysis of the processing requirements of a complex perceptual-motor task. Human Factors 25:597-621.

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

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

Lockhead, G. R. & Byrd, R. (1981) Practically perfect pitch. Journal of the Acoustical Society of America 70:387-389.

McCallum, W. C. & Knott, J. R. (1976) The responsive brain. John Wright.

McCarthy, G. & Donchin, E. (1981) A metric for thought: A comparison of P300 latency and reaction time. Science 211:22-80.

Medin, D. & Wattenmaker, W. D. (1987) Category cohesiveness, theories and cognitive archeology. In: The ecological and intellectual basis of categorization. ed. U. Neisser. Cambridge University Press.

Nunez, P. (1981) Electric fields of the brain: The neurophysics of EEG. Oxford University Press.

Okada, Y. C., Kaufman, L., & Williamson, S. J. (1983) The hippocampal formation as a source of slow endogenous potentials. Electroencephalography & Clinical Neurophysiology 55:417-426.

Paller, K. A., Kutas, M., & Mayes, A. R. (1987) Neural correlates of encoding in an incidental learning paradigm. Electroencephalography & Clinical Neurophysiology 67:360-371.

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.

Ritter, W., Simson, R., & Vaughan, H. G. Jr. (1972) Association cortex potentials and reaction time in auditory discrimination. Electroencephalography & Clinical Neurophysiology 33:547-555.

Ritter, W., Vaughan, H. G. Jr. & Costa, L. D. (1968) Orienting and habituation to auditory stimuli: a study of short term changes in average evoked responses. Electroencephalography & Clinical Neurophysiology 25:550-556.

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

Rosler, F. (1983) Endogenous ERPs and cognition: Probes, prospects, and pitfalls in matching pieces of the mind-body puzzle. In: Tutorials in event related potential research: Endogenous components. ed. A. W. K. Gaillard & W. Ritter. pp. 9-35. North-Holland.

Ruchkin, D. S. & Sutton, S. (1978) Equivocation and P300 amplitude. In: Multidisciplinary perspectives in event-related potential research. ed. D. Otto. pp. 175-177. US Government Printing Office, Washington, D.C.

Siegel, J.A. (1974) Sensory and verbal coding strategies in subjects with absolute pitch. Journal of Experimental Psychology 103:37-44.

Sirevaag, E., Kramer, A., Coles, M. G. H., & Donchin, E. (1984) P300 amplitude and resource allocation. Psychophysiology 21:598-599.

Squires, K. C., Petuchowsky, S., Wickens, C. D., & Donchin, E. (1977) The effects of stimulus sequence on event related potentials: A comparison of visual and auditory sequences. Perception and Psychophysics 22:31-40.

Squires, K. C., Wickens, C. D., Squires, N. K. & Donchin, E. (1976) The effects of stimulus sequence on the wave-form of the cortical event-related potential. Science 193:1142-1146.

Sutton, S. (1969) The specification of psychological variables in an average evoked potential experiment. In: Average evoked potentials: Methods, results and evaluations. ed. E. Donchin and D. B. Lindsley. pp. 237-297. (NASA SP-191). US. Government Printing Office.

Sutton, S. (1979) P300 - Thirteen years later. In: Evoked brain potentials and behavior. ed. H. Begleiter. pp. 107-126. Plenum Press.

Sutton, S., Braren, M., Zubin, J. & John, E. R. (1965) Evoked-potential correlates of stimulus uncertainty. Science 150:1187-1188.

Sutton, S. Tueting, P., Zubin, J., & John, E.R. (1967) Information delivery and sensory evoked potentials. Science 155:1436-1439.

Treisman, A. (1986) Properties, parts, and objects. In: Handbook of perception and human performance Vol II. ed. K. Boff, L. Kaufman, & J. P. Thomas. pp. 35.1-35.70. Wiley.

Tueting, P. (1978) Event-related potentials, cognitive events, and information processing: A summary of issues and discussions. In: Multidisciplinary Perspectives in Event-Related Potential Research. ed. D. Otto. pp. 159-169. US Government Printing Office.

Tueting, P., Sutton, S., & Zubin, J. (1970) Quantitative evoked potential correlates of the probability of events. Psychophysiology 7:385-394.

von Restorff, H. (1933) Uber die wirkung von bereischbildungen im spurenfeld. Psychologische Forschnung 18:299-342.

Wickens, C. D., Kramer, A. F., Vanasse, L. & Donchin, E. (1983) Performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information-processing resources. Science 221:1080-1082.

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: Two case studies. Annals of the New York Academy of Sciences 425:681-721.

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: P300 and fast guesses

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.), *Neurophysiology and Psychophysiology: Experimental and Clinical Applications*. 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 <u>count</u> condition). On other occasions the subject indicates the occurrence of one of the categories by pressing one of two buttons (a <u>Reaction Time</u>, 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 words, 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 study. 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%.

P300, RT, and Accuracy

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 Neuropnysiology, 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 v`n 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

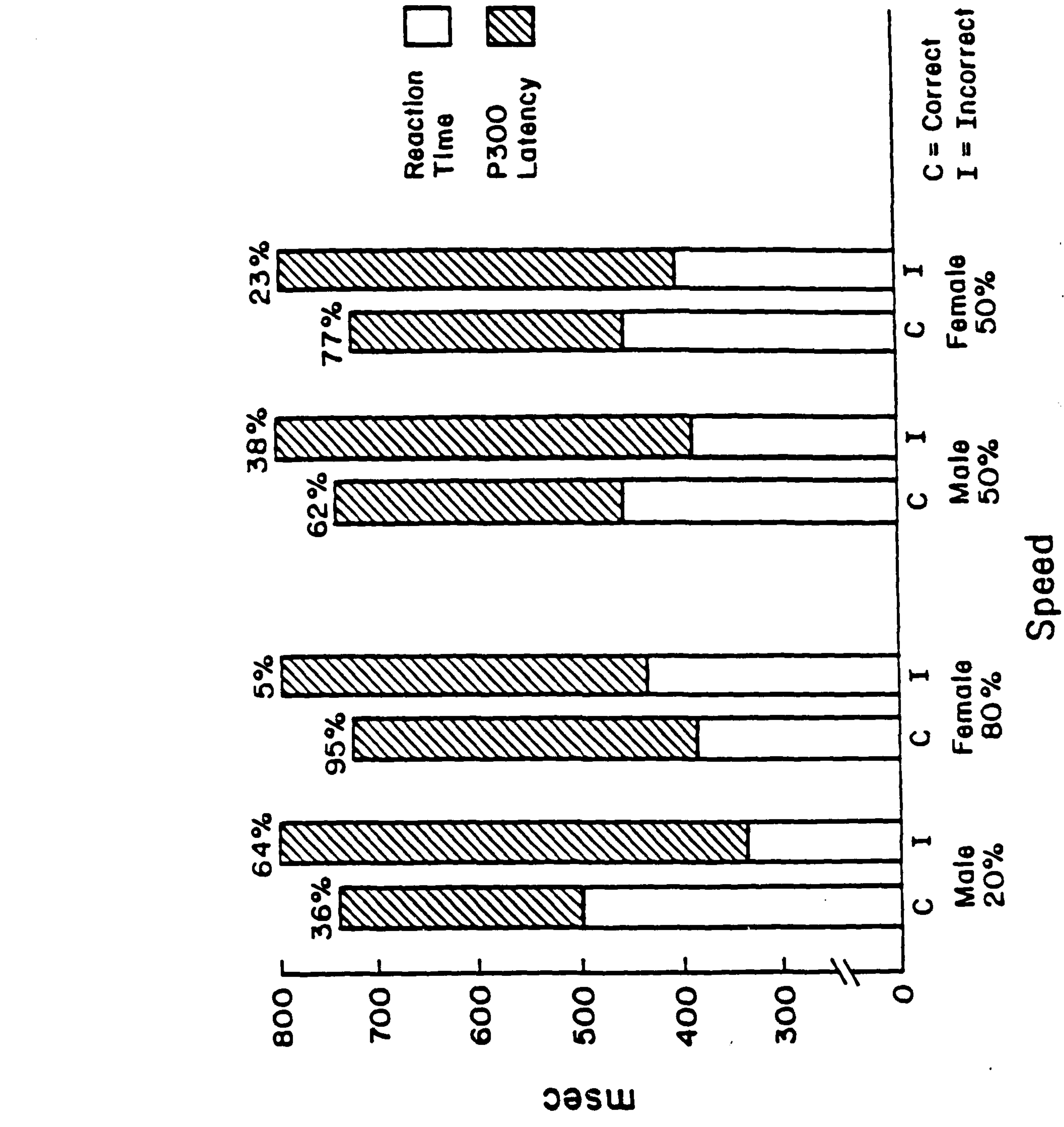
Figure 1. Reaction time and P300 latency for each probability (20/80 and 50/50), stimulus (male and female) and response (correct - C - and incorrect - I) condition. The frequency of correct and incorrect responses for each condition is also indicated.

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.

Fig. 1



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

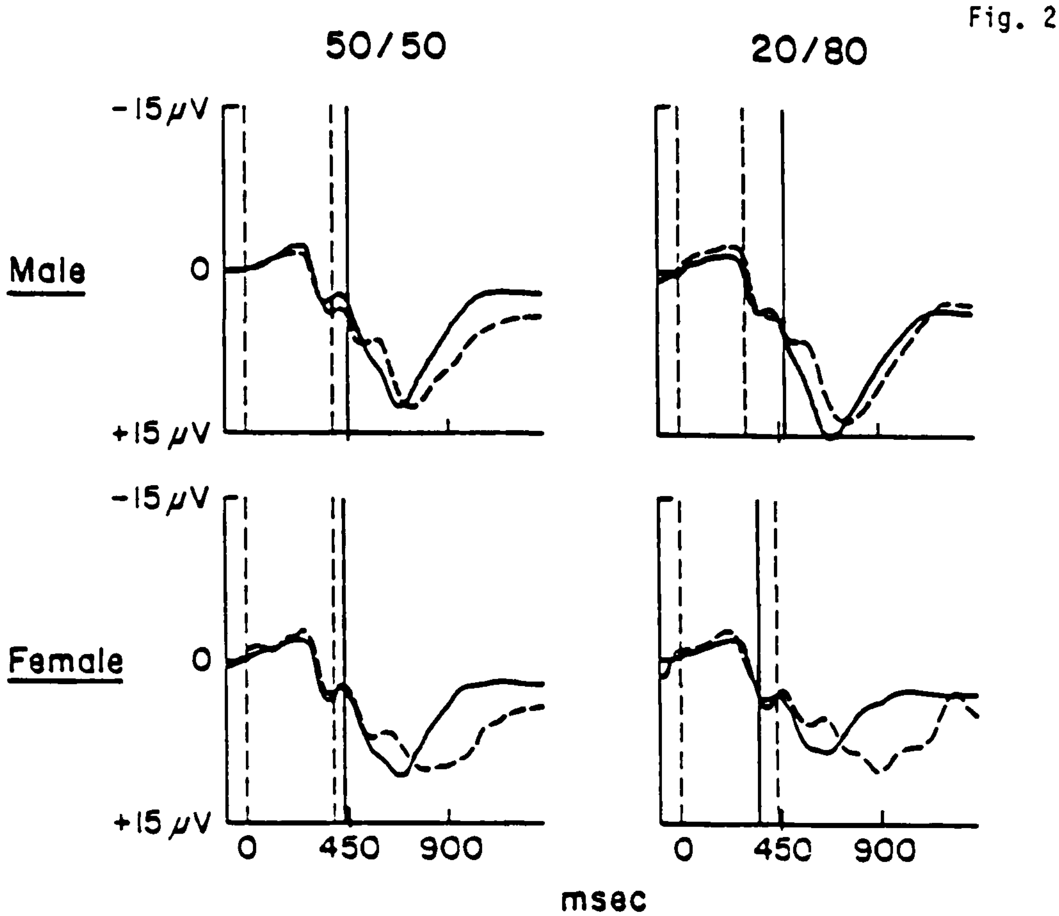
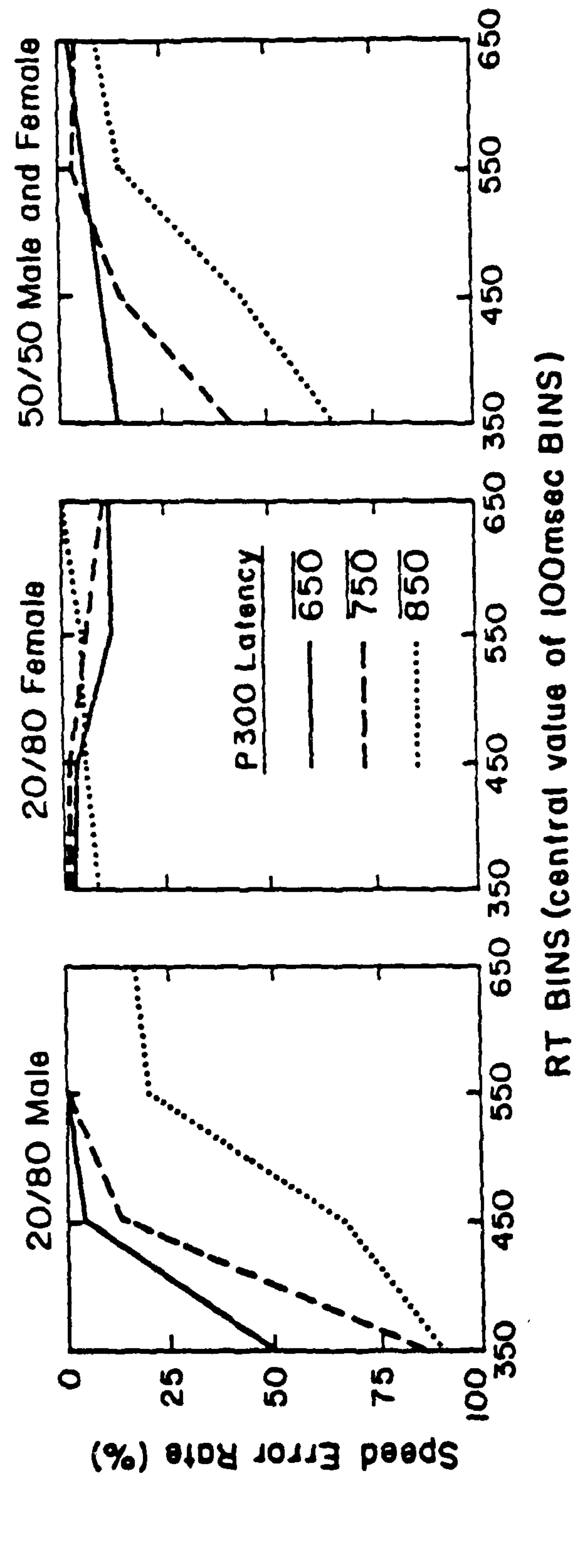


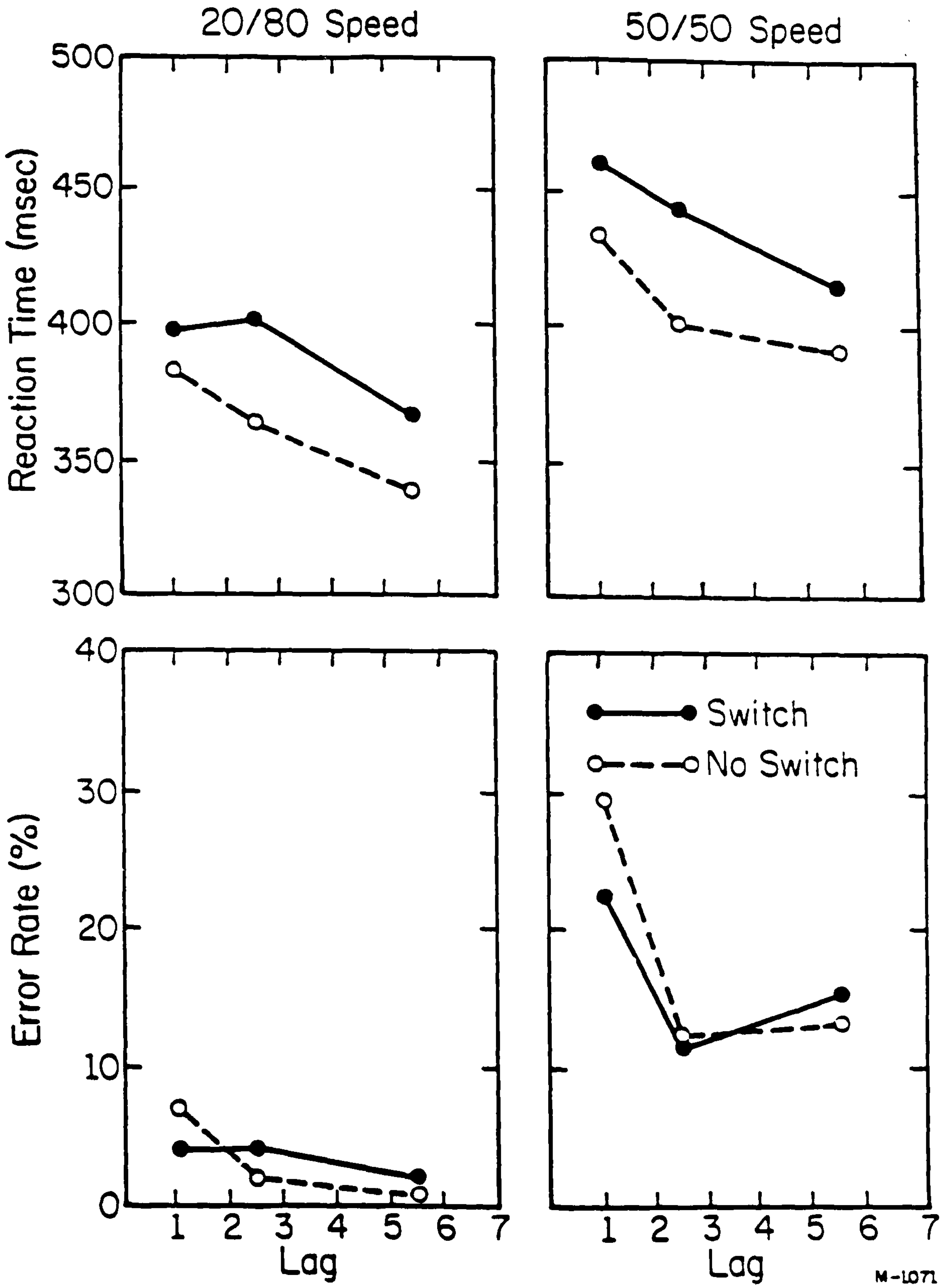
Fig. 2
50/50
20/80
-15 μV
Male
0
+15 μV
-15 μV
Female
0
+15 μV
0
450
900
0
450
900
msec

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
1 2 3 4 5 6 7
Lag
M-1071

Article 7.

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]

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 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

Advances in Psychophysiology, Volume 2, pages 1-78.

ISBN: 0-89232-644-1

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.

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" 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 experi-

mental 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 occurrences 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 subcomponents (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* references 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 brain stem 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 last 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 re-

sponse 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, and 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 statements 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 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, 1986). 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 is 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),[1] 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 & 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, e.g., Ragot, 1984; 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 (1985; see also Donchin, 1979).

Vector analysis (Gratton, Coles, & Donchin, 1983b, 1987) 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 ., 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, & 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, and 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, and Hillyard (1984) on monkeys. Thus, although P300-like activity has been observed fr 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 postsynaptic 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 o. these multiple late positivities, several investigators claim that P300 is not a unitary process (Courchesne, Hillyard, & Galambos, 1975; Friedman, Vaughan, & Erlenmeyer-Kimling, 1981; Picton & Stuss, 1980; Rösler & Sutton, 1986). 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 those 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, and 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 informa tion 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 subprocess 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 nontargets). 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 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.) 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.

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 ----------		
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 ----------				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)
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 μv 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 effect disappears at long ISIs (> 3 sec)	—	—
B. stimuli			
1. modality	visual larger (0–5 μv)	visual longer (100 ms)*	Cz usually more positive for visual?
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	—	—	—
B. nature of processing and transformation required	selective attention required; positive (old) items 5 μv or more larger in recognition or Sternberg paradigms; reduction (5–10 μv) with increase in workload (in dual tasks); 0–5 μv 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 μv increase with bonus for correct guess/prediction	—	—
6. Subject characteristics	decreases 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).

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 & Heffley, 1979): (a) peak measures do not allow a discrimination between overlapping components (this criticism also apply to area measures); and (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 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 num-

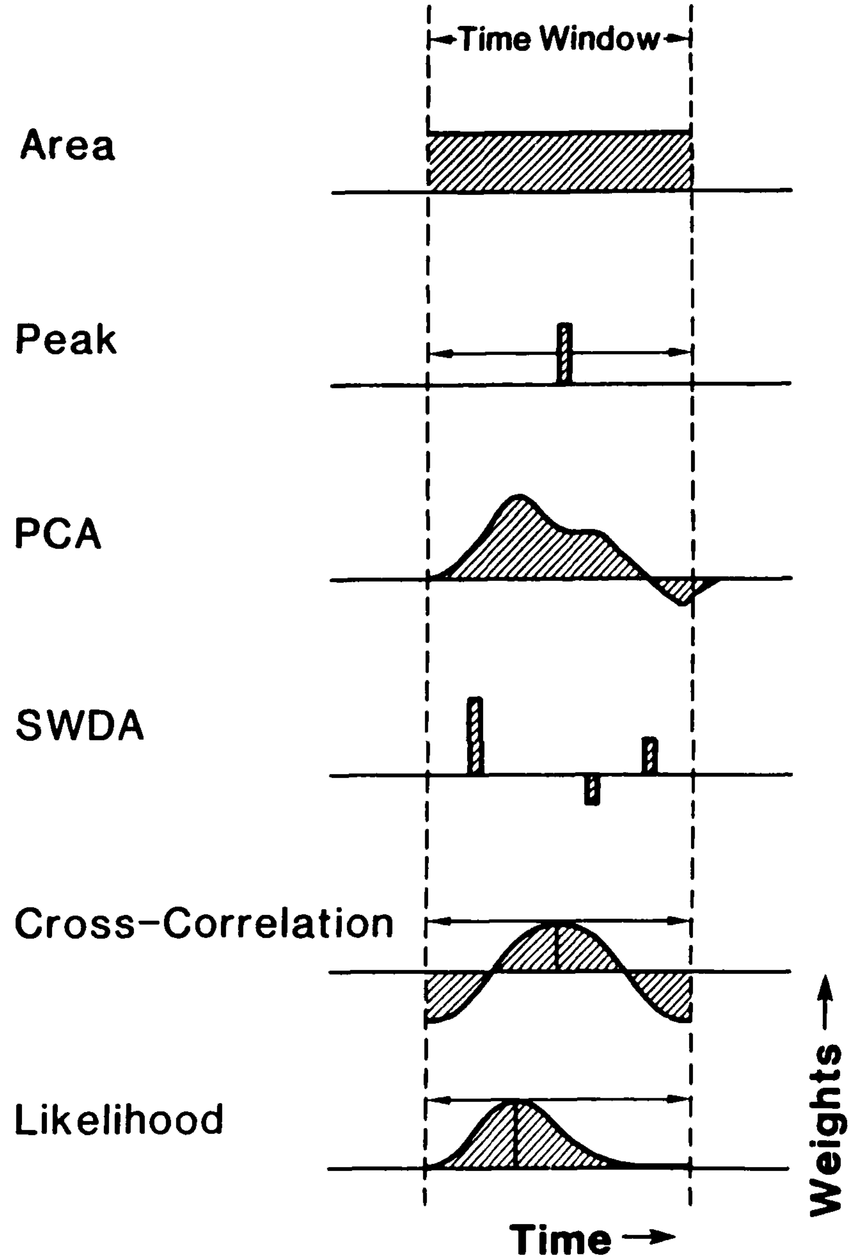


Figure 1. Graphic representation of the vectors of weights corresponding to several operational definitions of P300. The arrows indicate the possibility of shifting the vector of weights. In these cases P300 latency can be measured

ber 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, and Villegas (1964; see also John, Ruchkin, & Vidal, 1978), and Skrandies and 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, then 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 & 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 waveform classified as P300. Such procedure 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, 1987) 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 (1987; 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,' 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 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

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)

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. 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, 1987).

Table 4a. Number of Studies Using Various Techniques for P300 Measurement (N = 34)

Base-to-Peak only (or Peak 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

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, and Donchin (1983b, 1987). 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,* 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 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 nontarget (noncounted) stimuli (probability and target effects).

In Figure 4 we present the average Pz waveforms for each individual subject in the H-S oddball. Waveforms for the "count rare" conditions 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 nontarget) 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.

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.

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

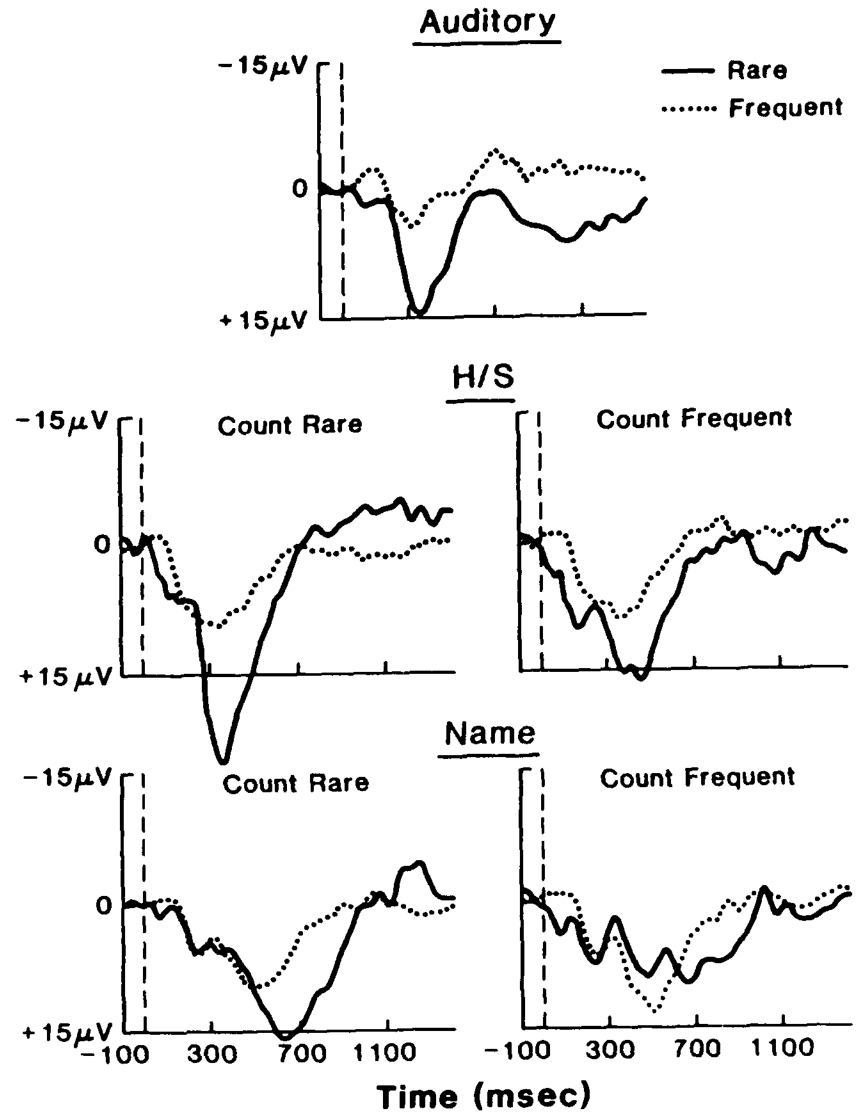


Figure 2. Average Pz waveforms for rare (20%) and frequent (80%) stimuli from the first session of one subject (#1) in the five different oddball tasks. (a) auditory RT task involving pitch discrimination, (b) visual count rare task involving letter discrimination, (c) visual count frequent task involving letter discrimination, (d) visual count rare task involving name gender discrimination, and (e) visual count frequent task involving name gender discrimination

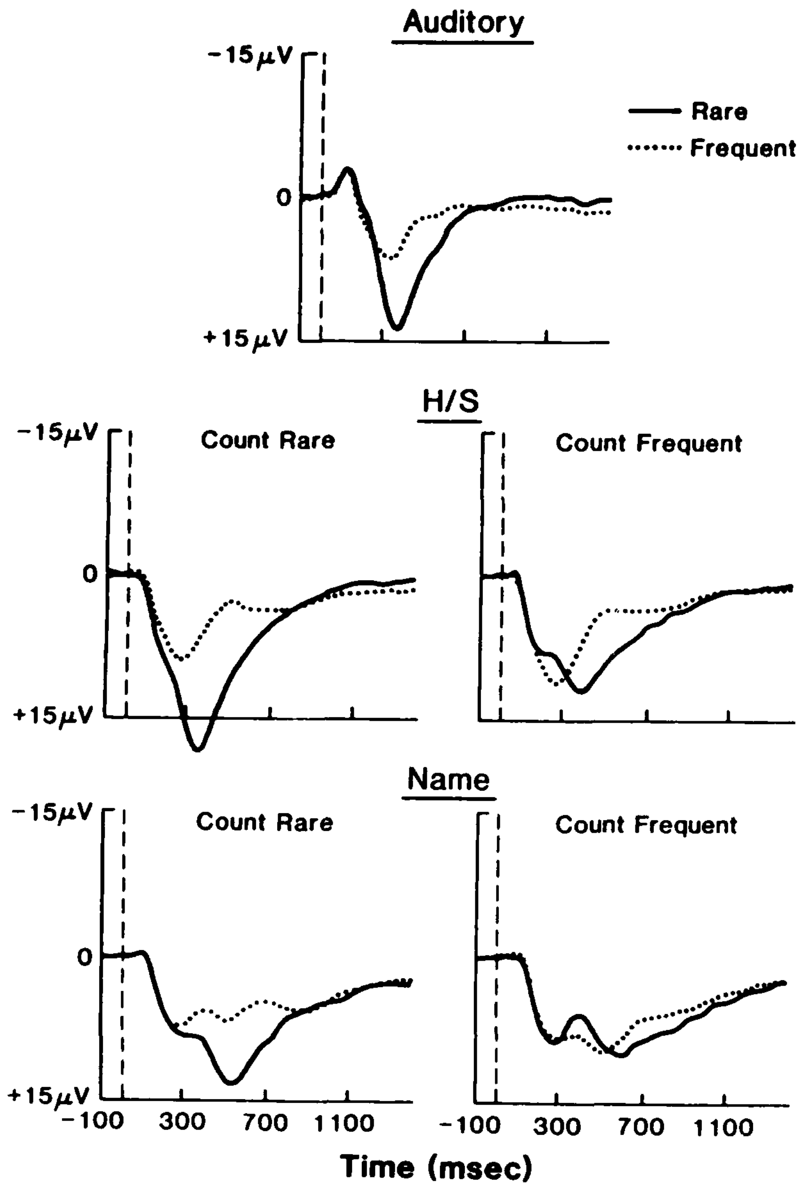


Figure 3. Grandaverage Pz waveforms (n = 49) for rare (20%) and frequent (80%) stimuli from the first session in the five different oddball tasks, (a) auditory RT task involving pitch discrimination, (b) visual count rare task involving letter discrimination, (c) visual count frequent task involving letter discrimination, (d) visual count rare task involving name gender discrimination, and (e) visual count frequent task involving name gender discrimination.

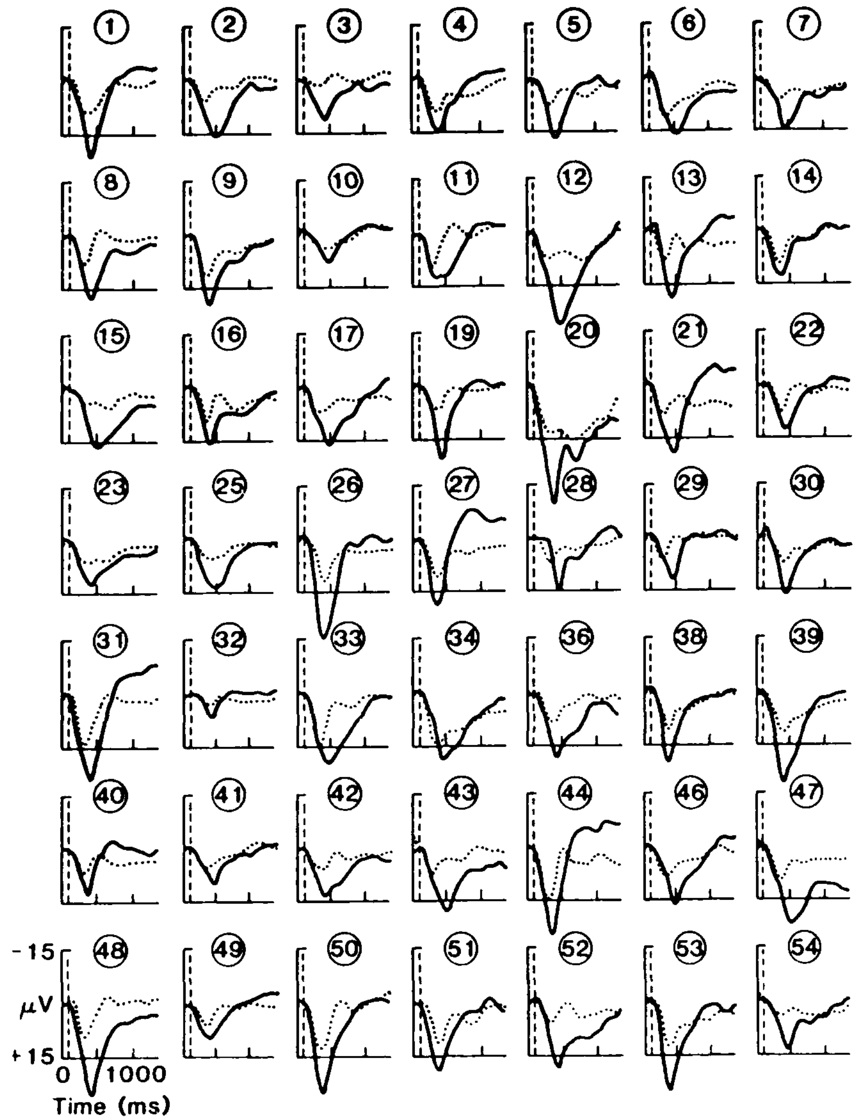


Figure 4a. Average Pz waveforms of each individual subject in the letter discrimination (H-S) task for rare (20%, solid line) and frequent (80%, dashed line) stimuli from the first session. Waveforms from the count rare condition are presented in this figure.

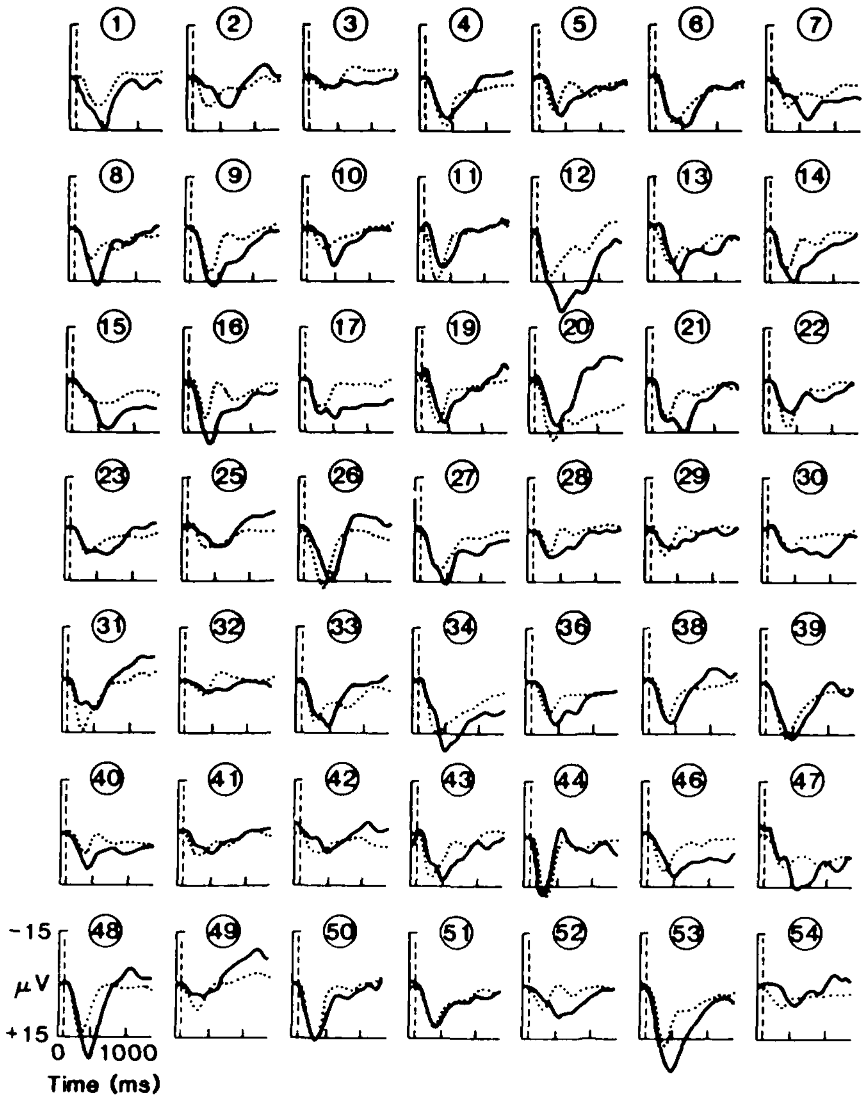


Figure 4b. Average Pz waveforms of each individual subjects in the letter discrimination (H-S) task for rare (20%, solid line) and frequent (80%, dashed line) stimuli from the first session. Waveforms from the count frequent condition are presented in this figure.

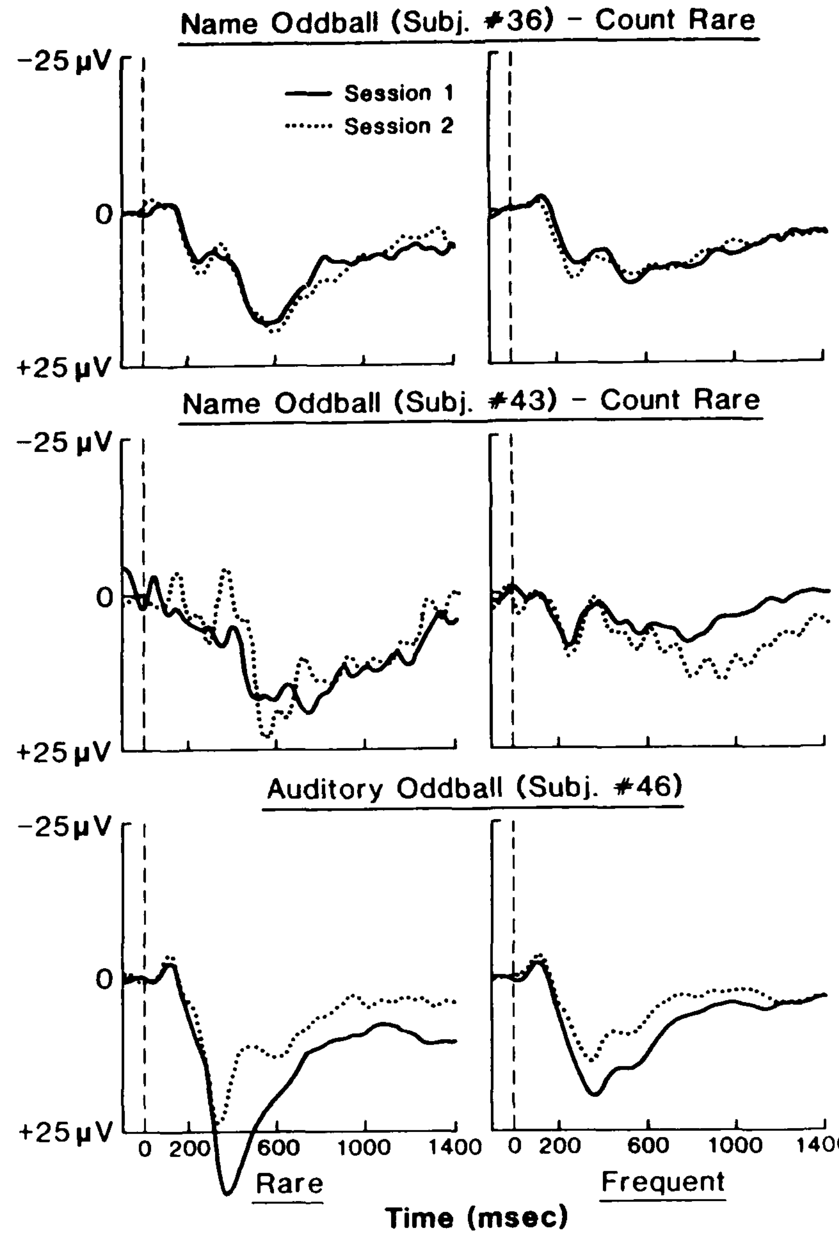


Figure 5. Examples of average Pz waveforms for rare and frequent stimuli for three different subjects and two sessions. The first session is represented by solid lines, the second by dashed lines.

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 Hz). 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 subgroups, 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 techniques for estimating P300 latency (Gratton, Kramer, Coles, & Donchin, 1987), 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 (1987) 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.' 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 windows. 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 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 differences 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 × 2 stimuli × 3 electrodes) and 204 waveforms for the H/S and Name oddball tasks (17 subjects × 2 tasks × 2 stimuli × 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 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 × 2 sessions × 2 stimuli × 3 electrodes), and 408 waveforms for the H/S and Name oddball tasks (17 subjects × 2 sessions × 2 tasks × 2 stimuli × 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 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 nontarget stimuli (task × 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 × 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 × 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 & Heffley, 1979).

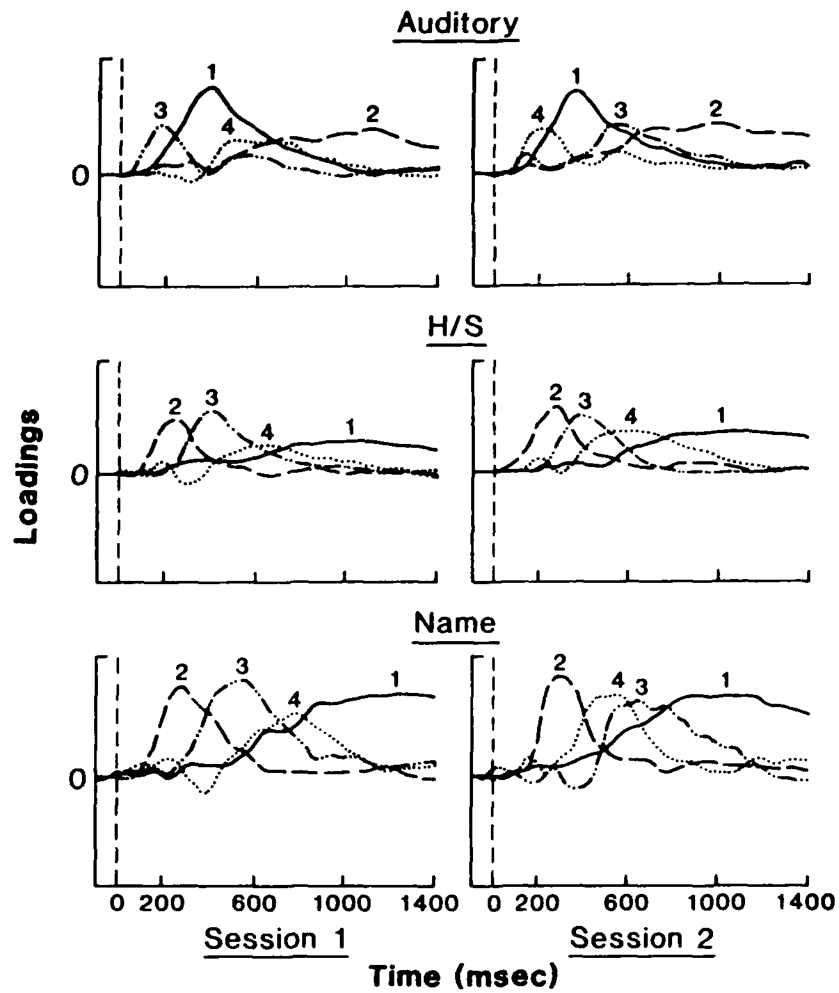


Figure 6. Reliability study: component loadings for the PCA's run separately for each session (left, session 1 - right, session 2). Component loadings for the Auditory oddball are at the top, the H/S oddball is in the middle, and the Name oddball is at the bottom.

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

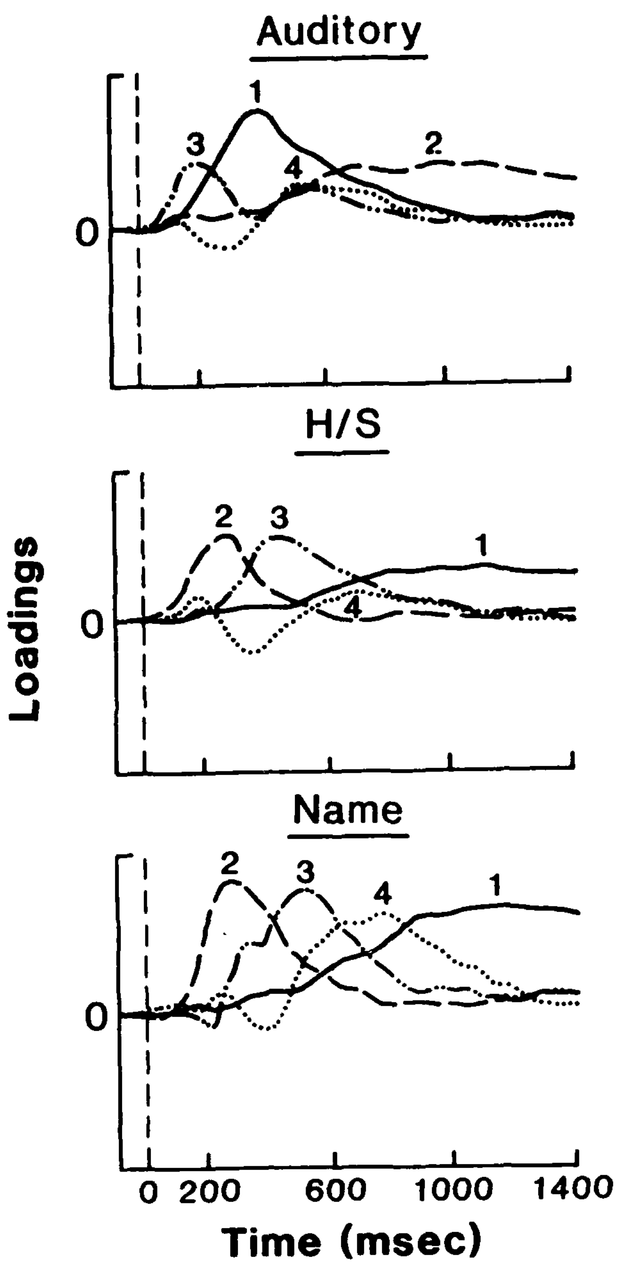


Figure 7. Reliability study: component loadings for the PCA's run on both sessions. Component loadings for the Auditory oddball are at the top, the H/S oddball is in the middle, and the Name oddball is at the bottom.

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 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.

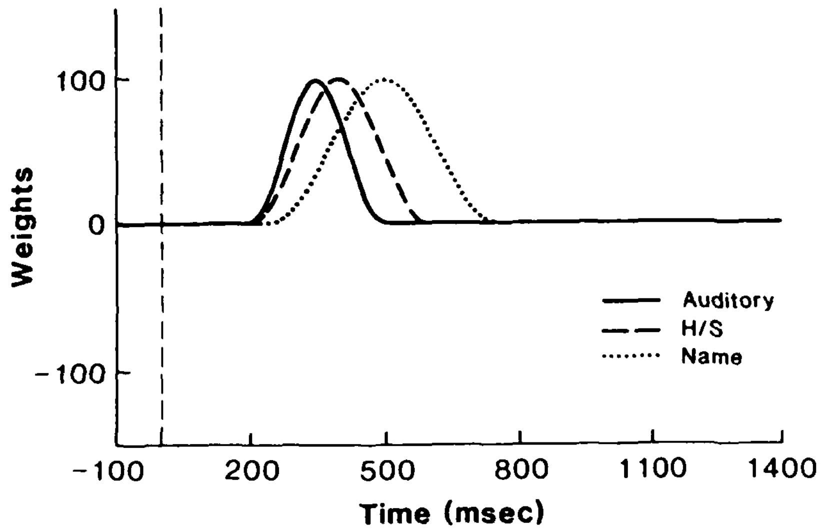


Figure 8. Reliability study: "time weights" used for the "simulated visual inspection" of the waveforms.

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. scored 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 5 presents a summary of the P300 measures we obtained with each 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 × stimulus conditions (i.e., Auditory rare target, Auditory frequent target, H/S rare target, H/S rare nontarget, H/S frequent target, H/S frequent nontarget, Name rare target, Name rare nontarget, Name frequent target, and Name frequent nontarget). 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 nontarget).

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 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 nontarget 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 nontarget 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. nontarget) are shown in Table 7.

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

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

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.

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 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).

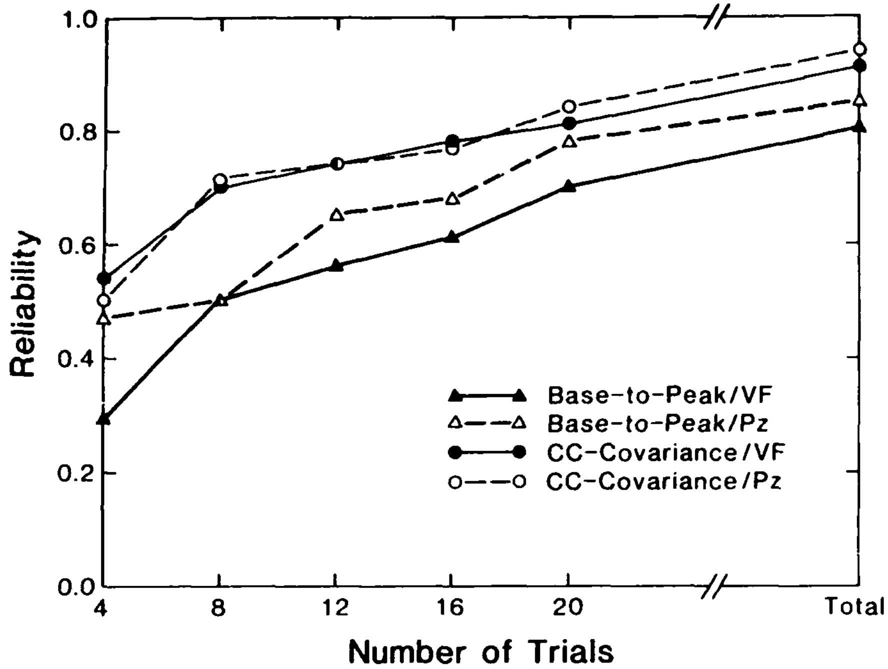


Figure 9. Reliability study: reliability of four P300 amplitude measures (Base-to-peak on Vector filtered data, Base-to-peak at Pz, CC-covariance on Vector filtered data, CC-covariance at Pz) as a function of the number of trials used for the analysis. The measures were taken on single trials and then averaged. Only frequent trials were used for the analysis.

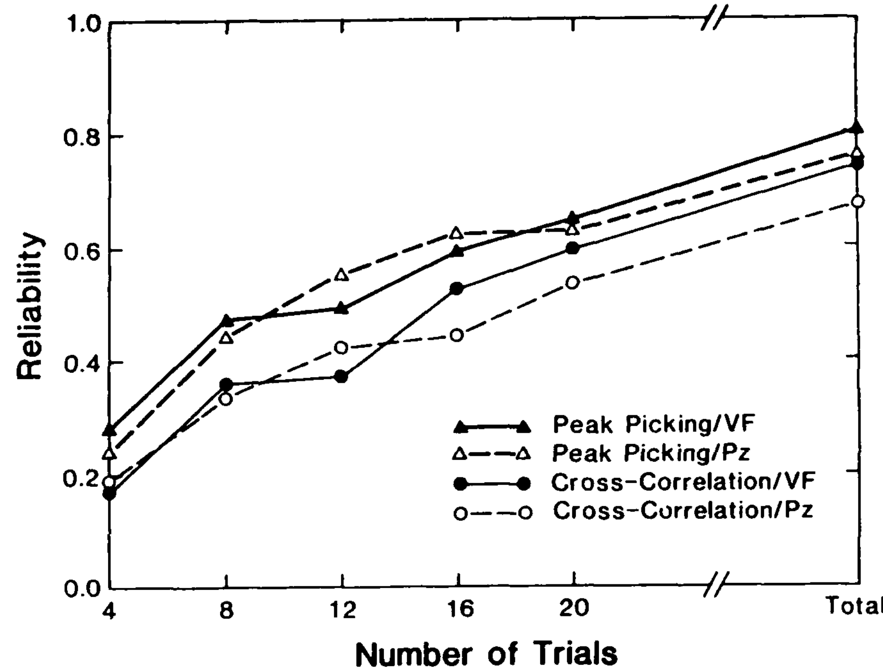


Figure 10. Reliability study: reliability of four P300 latency measures (Peak-picking on Vector filtered data, Peak-picking at Pz, Cross-correlation on Vector filtered data, Cross-correlation at Pz) as a function of the number of trials used for the analysis. The measures were taken on single trials and then averaged. Only frequent trials were used for the analysis.

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. 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 sessions, 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 a sign of inconsistency of the procedure used to measure P300. It may very well be the case that the time

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

lapse affects subjects to varying degrees. This "subject × 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 that 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 nontarget) are also shown.

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 reliability 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

Table 9. Between Session Reliabilities

Parameter	Range	All	Aud.	H/S	Name	Rare	Freq.	Targ.	NTarg.
		Medians							
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

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.

As exemplified in some of these plots, very high or very low reliability indices may be due to the effect of a small subsample of subjects with very deviant values (outliers). To detect outliers, we applied an algorithm to the data entered in each scatter plot. We

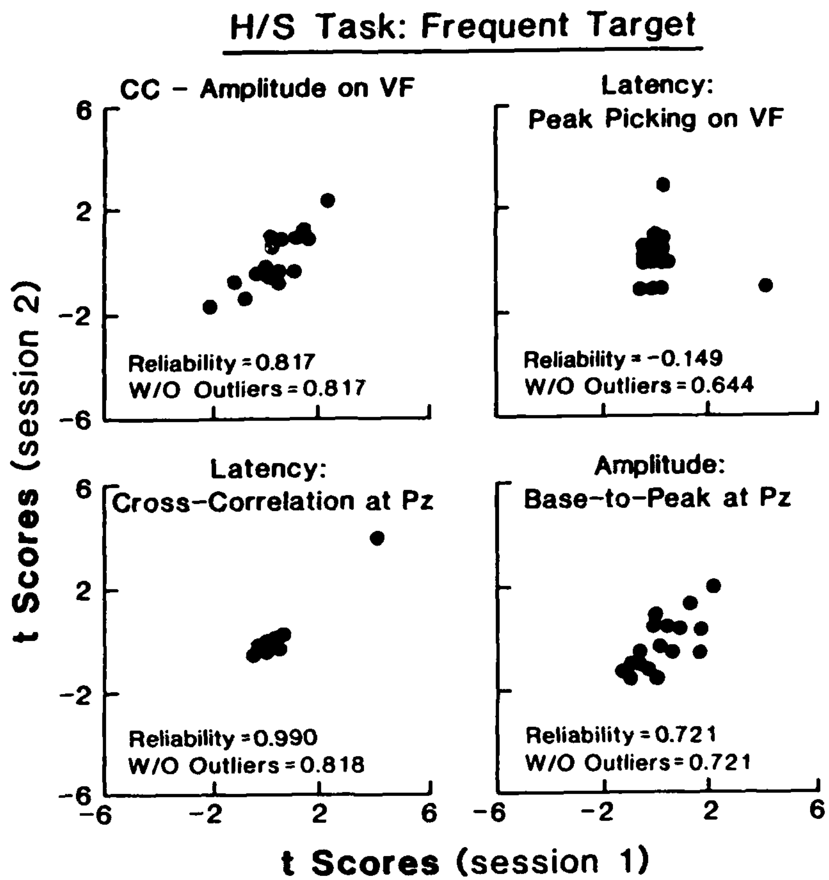


Figure 11. Examples of scatter plots of P300 parameters (t-scores) measured at the first and second session on 17 subjects. The plots refer to data obtained in the H/S oddball, count frequent condition. The upper left plot corresponds to estimates of P300 amplitude obtained with CC-amplitude on Vector 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.

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.

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 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

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 *Average*	Within Session *Single Trials*	Between Session *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
Latency:	Likelihood			.43

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 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 (1987), 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* subjects participated in the reliability study).

A similar caveat should be considered in *selecting* a *template* to be used 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 neces-

sary 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, 1987). 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, 1986; 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 generalizations 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 (1987) 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 components 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 (1985), 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, 1987) 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 procedures may strengthen 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, counter

balanced or random experimental designs should be used. Data from different sessions should be analyzed separately and pooled together only when yielding similar results.

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 "classical" 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 *rare*, while the other is *frequent*. 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 *target*) 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 nontarget 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 the 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, 1986).

In a classic oddball paradigm, P300 and 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 nontargets (target effect), and the ERPs to rare nontargets with those to frequent nontargets (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 nontarget 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, 1986; 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 than 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 nontargets, and focuses on examining the difference between nontargets on the attended channel versus nontargets on the nonattended 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 differences 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/nogo 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 nogo 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 nogo trials, compared with go trials, Cz became much more positive, while Pz remained constant (and Cz thus became equal to Pz). In a go/nogo 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 nogo trials. An explanation centering around differences in motor potentials between go and nogo 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

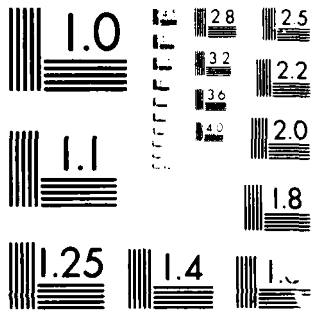
ERPs have been recorded to weak auditory, visual, and somatosensory stimuli in detection paradigms, although auditory paradigms are the most common. The typical paradigm involves a warning stimulus, followed less than a second later by a threshold-level stimulus on half of the trials. A few seconds later (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%. Intertrial 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 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

AD-A191 244 THE EVENT-RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING A..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 4/10

UNCLASSIFIED 29 FEB 88 CPL-88-1 AFOSR-TR-88-0316 F/G 5/8 NL


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4

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, after which 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, 1986) 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; 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 preceded 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 component. 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 referred to the "N2-P3a" component, as if these were 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 poststimulus, 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 later for visual than auditory stimuli, but there was no statistically significant difference in amplitude. Somatic stimuli also elicited P300s that were later 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 audit..., 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

NOTES

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 characteristics 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, and Gratton, 1986, 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.

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). New York: Van Nostrand Reinhold.

Coles, M. G. H., Gratton, G., Kramer, A. F., & Miller, G. A. (1986). Principles of signal acquisition and analysis. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), *Psychophysiology: Systems, processes, and applications* (pp. 183–221). New York: 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, DC: 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* (pp. 155–178). 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, DC: 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. (1986). Cognitive pschophysiology and human information processing. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), *Psychophysiology: Systems, processes, and applications* (pp. 244–267). 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. (1986). 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* (pp. 702–718). New York: Guilford Press.

Donchin, E., Ritter, W., & McCallum, W. 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. (1986). P300 and memory. In W. C. McCallum, R. Zappoli, & F. Denoth (Eds.), *Cerebral psychophysiology: Studies in event-related potentials. Supplement 38 to Electroencephalography and Clinical Neurophysiology*, 63–69.

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, pp. 301–339). 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. (1987). *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. (1987). *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, England: John Wright and Sons, Limited.

Hillyard, S. A., & Hansen, J. C. (1986). Attention: Electrophysiological Approaches. In M. G. H. Coles, E. Donchin, & S. W. Porges, *Psychophysiology: Systems, processes, and applications* (pp. 227–243). 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, October). *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). New York: Academic Press.

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. (1986). Multiple P300s in feedback tasks. In W. C. McCallum, R. Zappoli, & F. Denoth (Eds.), *Cerebral psychophysiology: Studies in event-related potentials. Supplement 38 to Electroencephalography and Clinical Neurophysiology*, 52–57.

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 aircrafts* (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-ty: : 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 judgments. *Psychophysiology, 22*, 334–344.

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, DC: American Psychological Association.

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. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology, 62*, 203–208.

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, September). *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 paired 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). Amsterdam: Elsevier-North Holland.

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. W., Woods, D., Stuss, D. T., & 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, DC: 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. (1986). On how to interbreed endogenous ERP components with cognitive constructs. In W. C. McCallum, R. Zappoli, & F. Denoth (Eds.), *Cerebral psychophysiology: Studies in event-related potentials. Supplement 38 to Electroencephalography and Clinical Neurophysiology*, 51–92.

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*, 259–315.

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). Evoked-potential correlates of stimulus uncertainty. *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*. New York: 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). New York: 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.

Journal of Experimental Psychology: Human Perception and Performance. In press.

Pre- and post-stimulus activation of response channels: A psychophysiological analysis

by

Gabriele Gratton, Michael G. H. Coles, Erik J. Sirevaag,
Charles W. Eriksen, and Emanuel Donchin

Department of Psychology
University of Illinois at Urbana-Champaign

Running Title: Response channels and psychophysiology

Abstract

The present study examined mechanisms of response activation in a choice reaction time task. Six subjects were instructed to respond differentially to the central letter of one of four arrays, HHHHH, SSHSS, SSSSS, and HHSHH, which were presented in a random sequence. Instructions emphasized speed. Measures of the Event-Related brain Potential (ERP), and of electromyographic activity (EMG) were obtained on each trial.

Conditional accuracy functions (i.e., accuracy conditional on latency) were characterized by three phases. For very fast responses, accuracy was at chance level for all arrays, suggesting that subjects were guessing. For intermediate latency responses, accuracy was above chance if the noise was compatible with target, and below chance if it was incompatible. This pattern of data suggests that these responses were based on partial stimulus analysis. For slow responses, accuracy was above chance for all arrays, suggesting that these responses were based on complete stimulus analysis.

Motor-related brain potentials provided further information about mechanisms of response activation. The occurrence and the accuracy of fast responses could be predicted by examining motor potentials preceding the presentation of the array. Furthermore, measures of the motor potentials in the period following the presentation of the array suggested that both responses were energized by array presentation, that partial analysis of stimulus information could activate responses, and that the level of response activation at the time of the triggering of the EMG response was constant for trials with different response latencies. The data are discussed within the framework of a response channel conception.

The present paper reports results of an investigation of the mechanisms that lead to the activation of responses in a choice reaction time (RT) paradigm. It is usual to define "responses" as the mechanical events (e.g., switch closures) that are recorded by the experimenter. However, these events are, in fact, the consequence of a complex chain of causally related phenomena that can be traced back through the spinal cord and the motor areas to various pre-setting mechanisms. While closing the switch is the goal imposed on the subject by the experimenter, the real task of the subject is to accomplish the movement that will result in the switch closure. Thus, to understand how the switch comes to be closed, it is arbitrary to choose the point of switch closure as the level at which the analysis should be conducted. Indeed, as we hope to illustrate, we are led to a much richer understanding of the processes involved in a choice RT task when we broaden our research focus to include different levels of the response process rather than restricting ourselves to the switch closure.

Support for this approach can be found in studies of response preparation and response competition, which have both been interpreted in terms of "subthreshold" response activation, that is, activation of the response system which is not sufficient to trigger an overt response. Thus, response preparation can be conceptualized as an energizing phenomenon (Posner, 1978; Requin, 1985), by which response structures are activated at a subthreshold level. Indeed, when Requin (1985) recorded from the motor cortex of monkeys, he observed an increase in the firing rate of neurons during the foreperiod of RT tasks. Some of these neurons showed a further increase in firing rate during the execution of the overt response. Response competition, on the other hand, can be viewed as the reciprocal inhibition of competing response structures (see Eriksen & Schultz, 1979;

Sherrington, 1906), in which the subthreshold activation of one response leads to a delay in the execution of another. This interpretation of response competition has been supported by the results of studies in which electromyographic (EMG) responses were recorded in association with overt responses (see Coles, Gratton, Bashore, Eriksen, & Donchin, 1985; Eriksen, Coles, Morris, & O'Hara, 1985).

The concept of response channels

The idea that there can be various levels of response-related activity in the nervous system can be conceptualized in terms of the notion of "response channel." We use this term as a heuristic device to refer to that complex of structures whose activities are more or less directly related to the mechanical event which is defined as the overt response. A similar conception can be found in Gaillard (1978) and Naatanen and Merisalo (1977).

We consider the response channels to be distinct from the stimulus evaluation system, whose function is to analyze stimulus information, resulting in activation of the appropriate response channel. However, the level of activity of the response channel may be affected by other factors such as response priming or bias and response competition.

Within this framework, different degrees of involvement of response structures are viewed as different degrees of response channel activation. Thus, response preparation can be represented as an increase in the level of activation of a response channel, while response competition is a decrease in the level of activation in one response channel as a result of the increase in the level of activation of another channel. The involvement of particular response structures can be represented by different thresholds of response channel activation. Thus, there is a threshold for the initiation of muscle activity, as revealed in the EMG, and a higher threshold for the initiation of an overt movement (see Coles et al., 1985).

Measuring the Activation of Response Channels

Information about the level of activation of the response channels can be gained by studying the activity of neural and muscular structures involved in the response, as well as by examining the overt response. In our previous work (Coles et al.. 1985; Eriksen et al., 1985), we recorded both the EMG activity and the overt response (a squeeze of a dynamometer) to determine the time at which the level of response channel activation reached two particular thresholds (that for the EMG response and that for the squeeze response). While this procedure allowed us to identify two points in the response channel activation function, it did not provide a continuous, analog description of the function.

In the present study, we used a component of the event-related brain potential (ERP), the readiness potential, as a measure of response channel activation. This component was first described by Kornhuber and Deecke (1965) in a study of self-paced movements. These authors recorded electrical activity from scalp electrodes and observed a negative, ramp-like potential, that began some 800 ms prior to the execution of a movement and peaked just after the onset of the movement. Later research has shown that this potential is largest at scalp electrodes placed over the hemisphere contralateral to the hand responsible for the movement (e.g., Kutas & Donchin, 1974, 1977, 1980; Vaughan, Costa, & Ritter, 1972). Furthermore, several investigators (e.g., Kutas & Donchin, 1980; Rohrbaugh, Syndulko, & Lindsley, 1976) have described a similar lateralized potential that develops during the foreperiod of simple warned RT tasks. The peak of this lateralized potential, which occurs following the imperative stimulus, appears to be time-locked to the overt response. A similar lateralized potential develops after the imperative stimulus in choice RT tasks, when response choice depends on information provided by the imperative stimulus, but before the imperative stimulus when the choice depends on information provided by the warning stimulus (Kutas and Donchin, 1980). Thus, the potential appears to become lateralized when a choice is made about the responding hand.

These observations led Kutas and Donchin to argue that the lateralization of the readiness potential can be viewed as an index of specific motor preparation.1 Subsequent reports (e.g., Okada, Williamson, & Kaufman, 1982), that have used neuromagnetic techniques to relate this potential to motor cortex activity, also lend credibility to the suggestion that the readiness potential is a good candidate for a psychophysiological measure of response channel activation. Therefore, we propose to use the lateralization of the readiness potential as a measure of the relative activation, or priming, of response channels prior to stimulus presentation. Further variations in the lateralization of the readiness potential occurring after the presentation of the stimulus should reflect the relative activation of the response channels as a consequence of the processes of stimulus evaluation.

Pre-stimulus activation

Pre-stimulus activation refers to activation of a response channel prior to the presentation of the imperative stimulus. When pre-stimulus activation occurs, the corresponding response should be facilitated, because the activation level at the time of stimulus presentation is closer to the overt response threshold. In extreme cases of pre-stimulus activation, small increases in activation that may occur as a consequence of mere stimulus presentation are sufficient to raise the activation level above the threshold and trigger the response. In this way, the probability of emitting a response in the absence of stimulus information will be enhanced. Responses of this type are commonly labelled "fast guesses." According to this view, then, fast guesses occur when there is an increased level of response channel activation prior to the imperative stimulus. Such a hypothesis is compatible with explanations of the effects of instructional sets, priming or warning stimuli, and manipulations of response probability, on the accuracy and latency of responses (see, for instance, Posner, 1978). More generally, this hypothesis, the variable baseline hypothesis, interprets variability in response bias in terms of variability in the baseline level of response channel activation.

This hypothesis differs from that proposed by Grice, Nullmeyer, and Spiker (1982) who argued that fast guesses are a consequence of variations in the setting of the response criterion. According to this "variable criterion hypothesis," fast guesses occur when the response criterion is set to such a low level that small fluctuations in the activation of the response channel (occurring after the stimulus) are sufficient to cross the criterion, or threshold, and trigger the response.

We propose that measures of the lateralized readiness potential can be used to evaluate these two hypotheses. In particular, if trials on which a fast guess occurs are associated with a greater pre-stimulus activation than slow response trials, then we should observe a larger lateralized readiness potential during the foreperiod of fast guess trials. On the other hand, if fast guesses occur when the response criterion is set to a lower level than for slow response trials, there should be less lateralization of the readiness potential at the time the overt response is initiated on fast guess trials.

Post-stimulus activation

Post-stimulus activation refers to the activation of the response channels that occurs as a consequence of the processing of the imperative stimulus. In our previous work (Coles et al., 1985, Eriksen et al., 1985), we have found evidence to suggest that responses can be activated in a parallel and continuous fashion as stimulus evaluation proceeds. We used a choice RT paradigm in which the stimulus was made up of a central target letter flanked by noise letters (cf. B. Eriksen & C. Eriksen, 1974). The noise letters could either call for the same response as the target (compatible noise) or for the opposite response (incompatible noise). Analysis of EMG and overt response measures revealed that, on some trials, both responses were initiated (Coles et al., 1985; Eriksen et al., 1985). These double responses occurred more often when incompatible noise was presented, suggesting parallel processing of target and noise information. Furthermore, conditional accuracy functions (i.e., accuracy conditional on latency) revealed that when subjects responded quickly to incompatible noise stimuli, they tended to make errors. Indeed, they erred more than would be predicted on the basis of chance responding (Coles et al., 1985). On the other hand, slow responses to incompatible noise stimuli were associated with high accuracy. The accuracy of responses on compatible noise trials was always greater than chance. These data suggest that responses are sometimes given on the basis of a preliminary analysis of stimulus information, whose output is dominated by the noise letters. In the case of incompatible noise, the output of this preliminary analysis leads to the incorrect response, since it is dominated by the noise letters.

These data support the thesis that response channels can be activated by partial analysis of stimulus information. Therefore, the lateralization of the readiness potential in the period following the presentation of the imperative stimulus, and preceding the overt response, should provide evidence of the influence of this partial analysis. In particular, if the imperative stimulus contains incompatible noise, we should observe activation of the incorrect response.

The present experiment

We have proposed that several factors can affect the activation of response channels. These include pre- and post-stimulus response activation. Furthermore, we have argued that the lateralized readiness potential can be used as a measure of the relative activation of response channels.

These ideas were explored in the present experiment. In particular, we used electrophysiological measures to investigate the mechanisms responsible for fast guesses, and to provide a description of the influence of evaluation processes on the response system. To this end, subjects performed a version of the noise/compatibility task (cf. Coles et al., 1985). They were instructed to respond quickly, and a large number of trials were collected, so as to provide a reasonable sample of fast guesses. Thus, in the present experiment we replicated the basic design of the Coles et al. (1985) study. However, rather than focussing on the temporal relationship between two measures of peripheral response activation (EMG and overt movement) we concentrated on a more central measure of response activation. In particular, we recorded the readiness potential from lateral scalp electrodes to derive an analog measure of the relative activation of response channels.

Method

Subjects

Six graduate students (2 females) at the University of Illinois, were paid $3.50 an hour, plus bonuses, for participation. The subjects (between 23 and 25 years old) had normal, or corrected to normal, vision and hearing. The subjects were not informed of the precise purposes of the experiment.

Stimuli

Each trial was initiated by a warning tone (1000 Hz, 50 ms duration, 65 dB amplitude), generated by a Schlumberger sine-square audio generator (model SG-18A) and administered binaurally through headphones. One thousand ms after the warning tone, a five letter array (HHHHH, SSHSS, SSSSS, or HHSHH) was presented on a DEC VT-11 CRT display for 100 ms. The interval between two consecutive trials varied randomly between 3500 and 5500 ms. The subject sat facing the screen at a distance of one meter, such that the angle subtended by each letter was approximately 0.5 degrees. Thus, the angle subtended by the whole array was approximately 2.5 degrees. A fixation dot, placed .1 degrees above the location of the target letter, remained visible throughout the experiment.

Procedure

Each subject took part in one practice and three experimental sessions. The subjects received 17 blocks of trials during each of the four sessions. In the first 15 blocks (made up of 80 trials each), one of the four arrays was presented on each trial. In the last two blocks (made up of 40 trials each), one of only two arrays (either HHHHH and SSHSS, or SSSSS and HHSHH) was presented on each single trial. Thus, in the last two blocks the central stimulus was always the same, making these blocks equivalent to a "simple RT" task rather than the "choice RT" task used in the first 15 blocks.2 In each case, the stimulus arrays were equiprobable (.25 for each stimulus array in the first 15 blocks, .50 in the last two blocks), and their presentation order was randomized. Each subject was instructed to respond to one of the two target letters (H or S) with one hand (left or right), and to the other target letter with the other hand, by squeezing a zero-displacement dynamometer. The association between target letter and responding hand was consistent for each subject and counterbalanced across subjects. Speed was emphasized over accuracy and subjects were trained, by means of verbal feedback, to respond at a speed which produced error rates that ranged between .10 and .20.

Overt Response Measurement

When subjects squeezed the zero-displacement dynamometers (Daytronic Linear Velocity Force Transducers, Model 152A, with Conditioner Amplifiers, Model 830A, see Kutas & Donchin, 1977), a voltage was generated which was proportional to the force applied to the transducer. This signal was digitized at 100 Hz for 1000 ms following array presentation, giving a continuous recording of the force output of both hands following each stimulus. A Schmitt-trigger could be set to any preselected force level. When the force reached a prescribed criterion, the system recorded the occurrence of an overt "criterion" response. Before the practice session, the value of each subject's maximum squeeze force was determined for each hand separately. Criterion values for each subject were set at 25% of the maximum force applied by that subject for that hand. During the practice sessions only, a click was presented over a loud-speaker whenever the force exerted on the transducer crossed the criterion.

Psychophysiological Recording

The electroencephalogram (EEG) was recorded from Fz, Cz, Pz, (according to the 10/20 system, Jasper, 1958), C3' (4 cm to the left of Cz), and C4' (4 cm to the right of Cz) referenced to linked mastoids, using Burden Ag/AgCl electrodes. Vertical and horizontal electrooculographic activity (EOG) was recorded using Beckman biopotential Ag/AgCl electrodes placed above and below the right eye, and at 2 cm external to the outer canthus of each eye. Ground electrodes were placed on the forehead. The EMG was recorded by attaching pairs of Beckman electrodes on both right and left forearms, using standard forearm flexor placements (Lippold, 1967). For EEG and EOG electrodes the impedance was less than 5 KOhm, for EMG, below 20 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 of 8 s for the high pass filter. The EMG signals were conditioned using a Grass Model 7P3B preamplifier and integrator combination. The preamplifier had a half 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 sec. For each psychophysiological measure (EEG, EOG, and EMG) and each trial, the derived voltages were digitized at 100 Hz for 2100 ms, starting 100 ms prior to the presentation of the warning tone, and ending 1000 ms after the presentation of the array.

Data Reduction

Motor 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 "criterion response." This response criterion was used for on-line feedback and RT was defined as the latency at which this criterion was crossed. However, a different response classification system was used for most of the analyses. In particular, a measure based on the onset latency of the EMG response was used.

The onset latency of the EMG response was evaluated on each trial according to the following procedure (cf. Coles et al., 1985). First, a minimum criterion value was established for each subject to discriminate the EMG response from random variations in background EMG. Since the background EMG was typically characterized by very small amplitude activity, the EMG response was readily identifiable. When the integrated EMG exceeded the minimum criterion, an EMG response was defined as having started and the latency of this activity was noted. EMG responses in both arms could be observed on the same trial. -In the present paper, we will focus on analyses based on the onset latency of the first EMG response.

ERP measures. For each single trial, the EEG data were corrected for both vertical and horizontal ocular movement artifacts using a modification of the procedure described in Gratton, Coles, and Donchin (1983). The corrected single trial data were then stored for further analysis, including computation of averages and analysis of ERP component parameters on each single trial. For the single trial analysis, the data from the five scalp electrodes (Fz, Cz, Pz, C3' and C4') were smoothed using a low pass digital filter (high frequency cut-off point at 3.14 Hz, two iterations), and the baseline level was subtracted by averaging the first ten points of the epoch (corresponding to 100 msec).

The readiness potential was assessed using the ERP waveforms recorded at C3' and C4'. These electrodes are placed on scalp regions close to brain motor areas. Previous research has shown that the amplitude of the readiness potential is maximum at these locations when squeeze responses are required (Kutas & Donchin, 1977; Bashore, McCarthy, Heffley, Clapman, &

Donchin, 1982).

Results

The results section will be divided into several parts. First, we describe the basic conditional accuracy functions for the two noise/compatibility conditions. Second, we present data concerning the influence of pre-stimulus response activation mechanisms on the latency and accuracy of overt responses. Third, we analyze the mechanisms of post-stimulus activation by studying the influence of stimulus evaluation on the relative activation of the two responses. Fourth, we consider the possibility that responses are released when relative response activation reaches a fixed criterion. Finally, we describe an analysis of the absolute, rather than relative, activation of each response.3

Conditional Accuracy Functions

Conditional accuracy functions were derived for compatible and incompatible conditions, and for each of the six subjects, by computing response accuracy for each of seven 50 ms response latency bins ranging from 100 to 449 ms (cf. Lappin & Ditsch, 1972a, 1972b; Pachella, 1974). The accuracy and latency of each "response" were defined either in terms of the first criterion squeeze (Figure 1, left panels) or onset of the first EMG response (Figure 1, right panels). Average functions are shown in the upper panels of Figure 1, while the proportions of responses for each latency bin are shown in the lower panels.

Insert Figure 1 About Here

The conditional accuracy functions based on the two response definitions are qualitatively similar, suggesting that both measures are manifestations of the activity of the same underlying system. However, there are many fewer criterion squeeze responses than EMG responses with a latency shorter than 200 ms. (In fact, only three subjects had criterion squeeze responses with a latency less than 150 ms.) This supports the view that EMG responses represent a more sensitive measure of response activation than the squeeze response (see also Coles et al., 1985; Eriksen et al., 1985). Furthermore, EMG responses are less affected by response competition than are squeeze responses (see Coles et al., 1985 and footnote 3). These two characteristics of the EMG onset measure make it more suitable than the criterion squeeze response as an index of peripheral response activation. Therefore we used the EMG measure in the remainder of this paper as the basis for response classification.

Analysis of the conditional accuracy functions based on EMG onset revealed that the proportion of correct responses increased with increasing response latency, $F(6,30) = 59.78$, $p<.001$, and that subjects were more often correct on compatible than incompatible trials, $F(1,5) = 144.40$, $p<.001$. The most interesting aspect of these functions, however, was that they differed in shape, as reflected in the significant interaction between compatibility and response latency, $F(6,30) = 9.96$, $p<.001$. As can be seen in Figure 1, the most noticeable difference between the functions occurred at response latencies of between 150 and 249 ms. For responses in the 150-199 and 200-249 latency bins, accuracy was significantly greater than chance (.50) for compatible arrays, $t(5)=4.25$ and 15.72 (one-tailed), but significantly below chance for incompatible arrays, $t(5)=-3.47$ and -2.34 (one-tailed). This differential pattern of response accuracy, which was evident in all subjects, suggests that, for these response latencies, the noise letters are controlling response accuracy (cf. Coles et al., 1985).

Since the noise letter in incompatible arrays calls for the incorrect response, responses for these arrays tended to be less accurate than chance. On the other hand, very fast responses (with a latency of less than 150 ms), tended to have a chance level of accuracy regardless of the compatibility of the array. For compatible and incompatible arrays, response accuracy was not significantly different from chance (50%), $t(5) = 0.80$ and -1.22, respectively. Since the accuracy of these responses was unaffected by stimulus information, they could be classified as "fast guesses." Conversely, slow responses, with a latency greater than 300 ms, tended to be correct regardless of the compatibility of the array. As we have argued previously, the accuracy of these slower responses was controlled by analysis of target letter information (cf. Coles et al., 1985).

Pre-stimulus activation

In this section, measures of the lateralized readiness potential are used to describe the pattern of response activation processes that occurs during the foreperiod (between warning tone and array presentation). In particular, we focus on the question of whether the accuracy of fast guess responses is under the control of an activation process that occurs before array presentation. Such an analysis should reveal whether these fast responses are associated with a higher baseline level of response activation, as predicted by the variable baseline hypothesis.

To examine the activation process, we computed the difference in voltage between the two laterally placed electrodes (C3' and C4') for every time point on each trial beginning 100 ms before tone onset and extending through the time of array presentation. As we noted earlier, the potential at these electrode sites is more negative on the side contralateral to the responding hand, when subjects are informed which hand to use to respond (see Kutas & Donchin, 1980). Thus, we predicted that the direction and degree of laterality observed during the foreperiod would be related to the nature of the response to the array. The values of lateralization were computed on the basis of which hand happened to be correct on a particular trial, so that negative values (upward deflections in Figures 2, 4, and 5) always indicate more negativity at the electrode site contralateral to the correct response.

To compare lateralization waveforms preceding fast and slow responses, we sorted trials according to EMG onset latency (using 50 ms bins ranging from 100 to 399 ms), response accuracy (correct or incorrect response) and noise/compatibility (compatible or incompatible noise). To obtain stable estimates of the lateralized brain potential, a data base of at least 50 trials must be used to derive average waveforms. In fact, the shortest EMG onset latency bin (100 to 149 ms) for all types of trial, and later EMG onset latency bins for incorrect compatible trials, did not match the 50 trial criterion. For this reason we had to restrict our analysis to the remaining cells (EMG onset latency bins between 150 and 399 ms for correct compatible trials, and for correct and incorrect incompatible trials). Note that the shortest EMG onset latency bin for which average lateralization waveforms were computed was 150 to 199 ms. Although the conditional accuracy data indicated that fast guess trials were most evident in the 100 to 149 ms bin, a substantial number of the trials in the 150 to 199 ms bin are also fast guesses, since accuracy is still quite close to chance for this bin.

Figure 2 shows average laterality values from incompatible trials during the foreperiod for correct and incorrect response trials with different response latencies (150-199 ms in the lower panel and 300-349 ms in the upper panel). As with the conditional accuracy functions, we used the EMG response to define both response accuracy and response latency.

Insert Figure 2 About Here

Inspection of Figure 2 suggests that the readiness potential preceding fast responses becomes lateralized during the foreperiod and that the direction of the laterality is related to the correctness of the subsequent response. This inference is supported by several analyses. First, when laterality is defined in terms of the mean absolute value of the C3'-C4' difference in the last 100 ms of the foreperiod, fast responses are associated with a greater degree of laterality than slow responses (the analysis was restricted to correct responses and to latency bins ranging between 150 and 399 ms because of unstable ERP data on other bins), $F(4, 20)=3.31$, $p<.05$. Further analysis revealed a significant linear trend in lateralization as a function of EMG onset latency bin, $F(1, 20)=10.45$, $p<.01$. Second, the direction of the laterality is related to response accuracy. Fast correct responses are associated with laterality in the direction of the correct response (greater negativity at the electrode location contralateral to the correct response), while fast incorrect responses are associated with laterality in the direction of the incorrect response (greater negativity at the electrode location ipsilateral to the correct response), $F(1, 5)=19.54$, $p<.01$. (This analysis was restricted to incompatible trials, because there were not enough incorrect responses to compatible arrays to yield stable averages.)

Finally, we classified each individual trial into one of three categories: trials for which the readiness potential was larger over the motor cortex contralateral to the correct response ("contralateral" trials); trials for which the readiness potential was larger over the motor cortex ipsilateral to the correct response ("ipsilateral" trials); and trials for which there was no significant difference between the two sides ("non-lateralized" trials).4 We computed the conditional accuracy functions separately for each of these categories of trials. The average conditional accuracy functions for compatible and incompatible trials computed in this way are shown in Figure 3.

Insert Figure 3 About Here

We predicted that the accuracy of fast guesses would be greater than chance when the correct response was prepared in advance of stimulus presentation - that is, for contralateral trials - and below chance when the incorrect response was prepared in advance of stimulus presentation - that is, for ipsilateral trials. This prediction was confirmed. An analysis of variance conducted on the conditional accuracy functions, for which accuracy was used as a dependent variable and noise compatibility, response latency bin and trial category were used as factors, indicated a significant main effect of trial category, $F(2,10) = 5.66$, $p < .05$, as well as a significant 2-way interaction between trial category and response latency bin, $F(10,50) = 2.88$, $p < .01$. Subsequent analyses revealed, for the main effect of trial category, a significant linear trend with decreases in accuracy from Contralateral to Non-Lateralized to Ipsilateral trials, and, for the interaction, that this linear trend was significant only for the first response latency bin (between 100 and 149 ms).

Taken together, these data suggest that subjects sometimes chose and activated a particular response during the foreperiod before the array was presented, although, objectively, each response was equiprobable. When a fast response was emitted on these trials, its accuracy depended on whether the subject happened to have chosen the correct response. This is consistent with our interpretation that these trials are fast guesses. However, the data presented so far are not sufficient to conclude that fast guesses occur when there is a high level of pre-stimulus response activation. Data bearing upon this issue will be presented later, when we consider a measure of individual response channel activation (the single-sided readiness potential). The presence of variable levels of response activation during the foreperiod and their relationship to response accuracy is consistent with the variable baseline hypothesis of response activation. In fact, it appears that the relative activation of response channels in the foreperiod is related to the accuracy of a fast guess.

Post-stimulus activation

In this section, we show how measures of the readiness potential are sensitive to the modulation of response channel activation by stimulus evaluation processes. In particular, these measures support the view that stimulus-related response activation can occur prior to complete stimulus evaluation.

As described above, we computed the difference between the electrical potential recorded from the two lateral electrodes (C3' and C4') on each trial. However, we now focus on the post-stimulus period from array presentation until the end of the recording epoch 1000 ms later. Trials were sorted as a function of array compatibility and average lateralization values were computed. The resulting average waveforms are shown in the middle panel of Figure 4. Note that these waveforms include all trials regardless of response accuracy or latency.

Insert Figure 4 About Here

If the response channels are continually influenced by information coming from stimulus evaluation processes, then we would expect that post-array values of laterality would reveal this influence. We argued previously on the basis of the conditional accuracy functions (reproduced in the upper panel of Figure 4) that responses with a latency between 150 and 249 ms appeared to be affected more by the noise letters than by the target letter, suggesting an influence of preliminary phases of the stimulus evaluation process on response channels. In fact, the average lateralization waveform for incompatible noise exhibits a "dip" towards incorrect response activation between 150 and 250 ms post-stimulus (see middle panel of Figure 4). This dip corresponds to greater negativity at the scalp site contralateral to the incorrect response. Note that the waveforms shown in the middle panel of Figure 4 correspond very closely to the conditional accuracy functions shown in the upper panel. The correlations computed between the average laterality (sampled at the beginning of each response latency bin) and conditional accuracy values for each of the seven bins were .93 for compatible and .94 for incompatible arrays ($p<.01$). When these correlations were computed on an individual subject basis, the mean value for compatible arrays was .87 (.05 confidence interval .79-.92) while that for incompatible arrays was .89 (.05 confidence interval .82-.94).

These data point to a remarkable convergence between two quite different procedures used to gain insights into the influence of the evaluation process on the response channels. One procedure is based on measures of scalp activity and the other on the latency and accuracy of responses.

While both procedures converge in their description of the time-course of evaluation processes, the accuracy of this description is compromised by the fact that both procedures are based on data aggregated over trials. Because of this, it is not clear whether the functions depicted in the upper and middle panels of Figure 4 represent the nature and time course of the evaluation process on individual trials. The figures suggest that information about the array is accumulated gradually and that this information is continuously available to response systems. However, the gradually increasing continuous function that characterizes conditional accuracy curves can be generated by aggregating trials for which the output of stimulus evaluation is, in fact, discrete but variable in latency over trials. Similarly, the description provided by the aggregated laterality data could be the result of averaging trials for which a discrete change in laterality occurred (in the direction of the response that was executed) but for which the change varied in latency. To demonstrate that indeed response activation occurs as stimulus evaluation proceeds, we need to demonstrate the influence of different phases of the stimulus evaluation process on the response system within the same trial.

To this end, we partitioned the laterality data as a function of response latency and accuracy. Since we were particularly concerned with finding evidence for an influence of preliminary evaluation processes on the response systems, we focussed on the "dip" in laterality seen in the waveform for incompatible arrays in the middle panel of Figure 4. We argued above that this "dip" represents early activation of the incorrect response due to the influence of the incompatible noise letters. We reasoned that if we could find evidence for this "dip" on trials in which the first peripheral response given was correct, then we could infer that, indeed, on these trials at least, preliminary evaluation processes could result in preliminary response activation. Furthermore, the evidence for an influence of evaluation processes on response systems should be most evident for relatively slow trials, since fast trials are influenced by pre-stimulus activation effects, as we saw earlier. Thus, we focussed our analysis on relatively slow, correct response trials only (with a response latency of 300-349 ms). On these trials, by definition, the EMG activity was first observed on the correct side. The lower panel of Figure 4 gives the laterality values for these trials.

The "dip" can be clearly seen in the waveform for the incompatible arrays while a simultaneous increase in laterality toward the correct response can be seen in the waveform for the compatible arrays. The difference between compatible and incompatible arrays in the area of the waveform between 170 and 210 ms after the array presentation was significant, $t(5)=2.67$, $p<.05$ (one-tailed), as was the difference between the same area measure for incompatible arrays and zero, $t(5)=-2.78$, $p<.05$ (one-tailed).

Thus, the measure of the lateralized readiness potential suggests that processing of the noise letters can result in preliminary incorrect response activation, even though the correct overt response is ultimately given presumably as a result of a later occurring analysis of the target letter. This influence is evident for long latency responses for which pre-stimulus activation effects are minimal.

Response execution

Grice's variable criterion hypothesis (Grice et al., 1982) leads to the prediction that the level of activation of the response channels at the moment of the response should be lower for fast than slow responses, while a fixed criterion hypothesis predicts a constant level, regardless of response latency. Figure 5 shows the average laterality waveforms from 100 ms before the warning stimulus until the time of the response. Waveforms for the compatible incorrect trials are not included because of insufficient data (see above).

Insert Figure 5 About Here

To evaluate the merit of the variable criterion hypothesis, we focus on the laterality value at the time an EMG response was observed (indicated by vertical lines in Figure 5).5 An analysis of variance of these data revealed no significant effects of compatibility ($F(1, 5) = 3.02$, $p > .05$) or response latency (both the main effect of response latency bin and the interaction of response latency bin by response accuracy were not significant, $F(3,15) = 0.13$, and $F(3,15) = 0.46$, respectively), but a significant effect of accuracy, $F(1,5) = 133.39$, $p < .001$. In fact, the mean laterality values were -0.71 and +0.65 microvolts for correct and incorrect responses respectively. These data indicate, then, that the absolute laterality value at the moment the EMG response is triggered is constant regardless of response latency and that, as expected, the direction of laterality is related to response accuracy.

Individual response channel activation

The measure of the lateralization of the readiness potential we have used so far provides only an index of the <u>relative</u> activation of the response channels. This relative measure cannot reveal generalized facilitation of both responses, nor can it distinguish cases in which one response is activated from cases in which the other response is inhibited. A related problem is that, conceptually, it is reasonable to expect the occurrence of a fast guess to be dependent on the absolute level of activation of a response channel. Similarly, the emission of an EMG response may depend on the absolute level of response channel activation.

For these reasons, it would be useful to obtain a <u>separate</u> measure of the activation of each response channel. To derive such a measure we have to solve the methodological problem of isolating ERP components. In fact, the potential observed at each of the lateral scalp electrodes (C3' and C4') is given by the sum of potentials generated locally (i.e., the single-side readiness potential) and of other potentials generated in other regions of the brain and propagated by volume conduction. Therefore, to obtain "pure" measures of the potential associated with the activation of just one response channel, we need to remove the influence of the other potentials. Note that when we measure the lateralization of the readiness potential (i.e., the difference between C3' and C4'), the influence of other brain potentials, not associated with a specific motor response, is eliminated by the subtraction procedure.

We approached the problem of isolating the single-sided readiness potential by using the Vector Filter procedure (Gratton, Coles, & Donchin, 1987). A description of the procedure and its derivation is given in the Appendix.

Examples of the waveforms obtained with the procedure are shown in Figure 6. The waveforms represent the estimated activation of correct and incorrect response channels throughout the experimental epoch for four response latency bins. Average waveforms were computed separately for each compatibility condition and for correct and incorrect trials. The 8 waveforms for the compatible correct condition shown in Figure 6 are representative of the 32 waveforms available.

Several aspects of this figure are noteworthy. First, during the foreperiod, there is greater activation of the correct side for fast than for slow trials. To determine whether the absolute amplitude of the readiness potential predicts if a fast guess will occur, we computed the probability that a fast response (i.e., a response with a latency shorter than 200 ms) would be emitted as a function of the level of negativity prior to the array at the electrode contralateral to the responding hand. The level of negativity prior to the array was assessed by computing the average single-sided readiness potential in the last 100 ms of the foreperiod at the electrode site contralateral to the correct response (when the probability of emitting a correct response was concerned) and at the electrode contralateral to the incorrect response (when the probability of emitting an incorrect response was concerned). We sorted the trials according to their prior negativity into 5 categories, ordered from a large to a low level of negativity prior to the array. The prediction that a fast response was more likely to occur on trials with a larger negativity prior to the array was confirmed, $F(4, 20) = 2.96$, $p < .05$. The probability ranged from .144 to .108 from the high to the low level of negativity. A post-hoc analysis revealed a significant linear trend in probability as a function of prior negativity, $F(1, 20) = 7.36$, $p < .02$. Thus, the probability of a fast response was a function of the level of negativity prior to the array at the electrode contralateral to the responding hand.

A second interesting aspect of Figure 6 is that there appear to be a fixed level of activation of the response channels at the moment the EMG response is triggered. Thus, the picture of response activation processes provided by Figure 6 (and by comparable waveforms for the other conditions) confirms the results reported in the previous sections (see Figure 5).

Insert Figure 6 About Here

The derivation of single-sided readiness potential measures enabled us to focus on one additional question, that is, the effect of mere stimulus presentation on response channels. One striking common aspect of all the individual response channel functions (those shown in Figure 6 and 24 other comparable waveforms from incompatible and incorrect trials) is that both correct and incorrect response channels show an increase in activation between 50 and 150 ms after the array. In the case of very fast response trials, when the level of activation of a response channel is close to the criterion due to priming in the foreperiod, this increase in activation appears to raise the function above the threshold for the emission of an EMG response.

Discussion

In this paper, we have presented an analysis of response channel activation in a warned, choice RT paradigm, in which the stimulus information was sometimes conflicting. We proposed that activation can be modulated by several mechanisms including those related to pre-stimulus activation or bias and those related to stimulus evaluation processes.

Measures of the lateralized readiness potential indicate that, on some trials, subjects appeared to select and activate particular responses during the foreperiod before the imperative stimulus (array) was presented. In fact, fast responses (with a latency of less than 200 ms) were associated with large and significant lateralization of the readiness potential during the foreperiod. If the lateralization was in the direction associated with the correct response, the response tended to be accurate. If it was in the direction associated with the incorrect response, the response tended to be inaccurate. For these fast responses, the lateralization of the readiness potential was similar to that observed in simple RT conditions when the subject knows in advance the hand to be used to make a response (cf. Kutas & Donchin, 1980). Thus, we infer that a fast guess was preceded by selection and activation of a response in advance of stimulus presentation. Furthermore, when pre-stimulus selection and activation occurred, the mere presentation of the stimulus appeared to trigger a response. This "mere presentation" effect may be due to a "response energizing" process, such as that proposed by Grice et al. (1982) and Sanders (1981). Preliminary evidence for such a process is provided by the analysis of single-sided readiness potential measures, which suggests that, regardless of response latency, both correct and incorrect response channels exhibited increased activation following stimulus presentation. However, this increased activation might also be attributed to some stimulus-processing activity, not completely eliminated by our procedure for isolating the single-sided readiness potential (see Appendix).

After stimulus presentation, responses appear to be activated as information about the stimulus becomes available - unless, of course, a fast guess is made. In particular, conditional accuracy functions and lateralized readiness potential values suggest a tendency for incorrect response activation between 150 and 250 ms after the presentation of an incompatible array. Because we were able to demonstrate the presence of this incorrect response activation for correct trials of relatively constant latency (300-349 ms), we inferred that, at least on some trials, stimulus evaluation processes may affect the response systems before the evaluation processes are complete.

Finally, we have shown that responses (EMG activity) occur when response activation achieves a particular fixed level. Analyses of both the lateralized and single-sided readiness potential at the time of response onset indicated that the amplitude of the readiness potential was fixed for all response latencies.

These results have implications for various issues in the study of human information processing. First, they provide support for a "variable-baseline/fixed-criterion" hypothesis of response activation, in contrast to a "fixed-baseline/variable-criterion" hypothesis. Specifically, we have demonstrated that a measure that is intimately related to response processes (the readiness potential) varies in the foreperiod and that this variability is related to the speed and accuracy of responses. We have also shown that the same measure appears to have a fixed value at the time of the response, regardless of response latency. While the latter observation suggests that a fixed level of activation of some central structure determines whether a peripheral response will be activated, it is of course possible that variable criteria operate at earlier phases of response activation, although our measures of activation provide no evidence about the operation of variable criteria. It should be noted that the portion of ERP waveform that is associated with the moment at which the EMG response is triggered may reflect the execution of the response (a fixed phenomenon) rather than the preparation of the response (a variable phenomenon).

However, recent neurophysiological data-question the distinction between response execution and preparation. In fact, Requin (1985) showed that there are three classes of neurons in the motor cortex that fire during a warned RT trial. Some fire only during the foreperiod, others fire only during the response, and still others fire during both the foreperiod and the response. While our scalp recordings cannot distinguish among the activity of these three classes of neurons, we should note that Requin (1985) proposes that the three classes operate as an integrated system. We consider this neurophysiological system as constituting a central portion of the response channel. Thus, response preparation and response execution can be represented by different aspects of the response channel activity.

A second issue addressed by the present experiment concerns the nature of the stimulus evaluation process. The interference determined by irrelevant noise presented visually in close proximity to target information has been studied extensively (for a review see Johnston & Dark, 1986). In the present study, conditional accuracy and lateralization data converge in suggesting that there are at least two phases in stimulus evaluation. During the first phase (between 150 and 250 ms) the correct response tends to be activated when the noise letters are compatible, and the incorrect response tends to be activated when the noise letters are incompatible. Later in time, the correct response is activated regardless of the nature of the noise letters. As we have noted previously (Coles et al., 1985), the first phase appears to correspond to an analysis of all the letters or features in the array without regard to their location, while the second phase may be related to the association of the letters with their location. This two-phases conception of stimulus evaluation processes is consistent with the feature integration theory of Treisman and Gelade (1980, see also

Treisman, Sykes, & Gelade, 1977), and with the "zoom-lens" analogy proposed by Eriksen and Yeh (1985, see also Naatanen, 1982). Our data appear to be consistent with the interpretation that location plays an important role in visual attention (Johnston & Dark, 1986). However, since we did not manipulate information about location, our data cannot serve to determine whether this role is qualitatively different from that of other stimulus dimensions (as proposed by the zoom-lens analogy) or similar to that of other stimulus dimensions (as proposed by the feature-integration theory).

The conditional accuracy functions also reveal that accuracy is between .90 and .95 for incompatible arrays even at long response latencies, when accuracy might be expected to approximate unity. This suggests that the stimulus evaluation system sometimes fails to correctly identify the target letter, perhaps because its operation terminates when some degree of certainty about the stimulus is reached, or because its performance is sometimes data-limited.

A third issue addressed by the present experiment concerns the possibility that stimulus information can affect response-related processes before the evaluation process is completed. Our measure of preliminary response activation (at a central level) showed that incorrect responses can be partially activated on the same trials for which a correct overt response is given later. These data are inconsistent with a single-stage decision model that places the decision stage at an earlier level than that manifested by the readiness potential measures. Furthermore, our previous findings (Coles et al., 1985) are inconsistent with a model that places a single-decision stage at an earlier level than the peripheral response (EMG or squeeze). In the earlier study, we found that both correct and incorrect peripheral responses could be activated on the same trial. Taken together, these data argue against single-decision stage models (e.g. Sternberg, 1969) because information is apparently transmitted from stimulus evaluation to response systems at at least two moments in time. This is consistent with continuous flow (Eriksen & Schultz, 1979; Grice et al., 1982) or multiple-decision-stage models (Miller, 1982).

In conclusion, we have illustrated how the concept of response channels and the measurement of their activation can provide insights into the mechanisms involved in a warned, choice RT task. Our data suggest that response channels are continuously active, and that when this activity crosses a threshold, a peripheral response is emitted. Reaction times measure the latency at which the threshold is crossed and the overt response is emitted. However, they provide little information about the behavior of the response channel activation function in the period preceding the moment at which the overt response is emitted. In our previous work, we used measures of the EMG to provide a second point in the description of the response channel activation function (Coles et al., 1985; Eriksen et al., 1985). In the present experiment, we extended our measurement repertoire to include a continuous measure (the lateralization of the readiness potential) and we have illustrated its power in providing insights into various aspects of the human information processing system.

Authors' note

Some of the data reported in this study were presented at 25th annual meeting of the Society for Psychophysiological Research, Houston (TX), 1985.

The study was supported in part by a contract from the Air Force Office of Scientific Research, Contract #F49620-83-0144, Al Fregly, Project Director; by a grant from the National Institute of Mental Health, #1R01-MH41445; and by a Public Health Service Research Career Program Award K6-MH-22014 to C. W. Eriksen. We wish to thank Ritske De Jong, Monica Fabiani, and Gordon Logan for their comments on earlier versions of this manuscript. The manuscript also benefited from the comments of Jeff Miller, Risto Naatanen, and two anonymous reviewers.

Requests for reprints should be sent to Gabriele Gratton, University of Illinois, Psychology Department, 603 E. Daniel St., Champaign, Illinois, 61820.

Footnotes

1. Another negative component, labelled Contingent Negative Variation (CNV, Walter, Cooper, Aldridge, McCallum, & Winter, 1964) can be observed in the foreperiod of warned RT tasks (see, for instance, Coles et al., 1985). The distinction between the readiness potential and the CNV has recently been debated (see Rohrbaugh & Gaillard, 1983). In particular, the possibility of a CNV in the absence of an overt response has been discussed. The arguments related to this debate are beyond the scope of this paper. We note that, in general, the literature about CNV has focussed on its relationship to "attentional" or "alertness" states, while that on the readiness potential has focussed on its relationship with motor responses. Since we will consider the scalp negativity preceding the stimulus in terms of its relationship with the motor response, we will use the label "readiness potential" to refer to it.

2. The simple RT condition was used to determine whether in fact a scalp lateralization of the readiness potential could be observed under conditions of extreme response bias. Since such a result was in fact obtained, we will not discuss this condition further.

3. In our previous study (Coles et al., 1985), we analyzed both EMG and squeeze onset latencies. Squeeze and EMG latency data from the present study replicated those obtained previously for the "random-warned" condition. In particular, for incompatible arrays relative to compatible arrays, we observed (a) longer RTs (by 41 ms), $\underline{F}(1,5)=59.16$, $\underline{p}<.001$, (b) longer intervals between EMG and squeeze onsets (by 6 ms), $\underline{F}(1,5)=7.66$, $\underline{p}<.05$, and (c) more trials for which both squeeze responses were initiated, $\underline{F}(3,15)=24.30$, $\underline{p}<.001$. Furthermore, on trials where both squeeze responses were initiated, the EMG to squeeze interval was prolonged by 18 ms, $\underline{F}(2,$

10)=32.52, $p<.001$. Thus, data from the present experiment confirm that response competition is a contributing factor to the differences in mean RTs between compatible and incompatible noise stimuli.

4. Estimates of the magnitude of the lateralized readiness potential on each single trial were obtained at each electrode by computing the average EEG activity of the 100 ms preceding the appearance of the stimulus array. Single trials were then classified into one of three categories (IPSILATERAL, CONTRALATERAL, and NON-LATERALIZED), by comparing the readiness potential at C3' and C4'. The voltage difference between these two electrodes was transformed into standard scores (for each subject). If the standardized difference value between the two electrodes was between +0.5 and -0.5, the trial was said to be "non-lateralized." If the difference was larger than this criterion, the trial was said to be "ipsilateral" if there was more negativity at the electrode ipsilateral to the hand corresponding to the correct response, and "contralateral" if there was more negativity at the electrode contralateral to the hand corresponding to the correct response.

5. To account for the transmission delay between the cortex and the muscle, we actually sampled the value of the lateralized readiness potential at the beginning of each EMG onset latency bin -- that is, on the average, 25 ms before the onset of EMG activity.

References

Bashore, T. R., McCarthy, G., Heffley, E. F., Clapman, R. M., & Donchin, E. (1982). Is handwriting posture associated with differences in motor control? An analysis of asymmetries in the Readiness Potential. Neuropsychologia, 20, 327-346.

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.

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

Eriksen, C. W., Coles, M. G. H., Morris, L. R., & O'Hara, W. P. (1985). An electromyographic examination of response competition. Bulletin of the Psychonomic Society, 23, 165-168

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

Eriksen, C. W., & Yeh, Y-Y. (1985). Allocation of attention in the visual field. Journal of Experimental Psychology: Human Perception and Performance, 11, 583-597.

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.

Gaillard, A. (1978). Slow brain potentials preceding task performance. Doctoral Dissertation.

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. (1987). A multivariate approach to the analysis of scalp distribution: Vector Analysis. Submitted for publication

Gratton, G., Kramer, A. F., Coles, M. G. H., & Donchin, E. (1987). Simulation studies of the latency measures of components of the event-related brain potential. Submitted for publication.

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

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

Johnston, W. A., & Dark, V. S. (1986). Selective attention. Annual Review of Psychology, 37, 43-75.

Kornhuber, H. H., & Deecke, L. (1965). Hirnpotentialanderungen bei Wilkurbewegungen und passiven Bewegungen des Menschen: Bereitschaftpotential und reafferente Potentiale. Pflugers Archives fur die gesammte Physiologie, 248, 1-17.

Kutas, M., & Donchin, E. (1974). Studies of squeezing: The effect of handedness, the responding hand and response force on the contralateral dominance of the readiness potential. Science, 186, 545-548.

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., & Donchin, E. (1980). Preparation to respond as manifested by movement-related brain potential. Brain Research, 202, 95-115.

Lappin, J. S., & Disch, K. (1972a). The latency operating characteristic: 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 characteristic: II. Effects of 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. 249-297). Amsterdam: North-Holland.

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.

Naatanen, R. (1982). Processing negativity: An evoked-potential reflection of selective attention. Psychological Bulletin, 92, 605-640.

Naatanen, R., & Merisalo, A. (1977). Expectancy and preparation in simple reaction time. In S. Dornic (Ed.), Attention and Performance II (pp. 115-138). Hillsdale, NJ: Erlbaum.

Okada, Y. C., Williamson, S. J., & Kaufman, L. (1982). Magnetic fields of the human sensory-motor cortex. International Journal of Neurophysiology, 17, 33-38.

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, N.J.: Erlbaum.

Requin, J. (1985). Looking forward to moving soon: Ante factum selective processes in motor control. In M. I. Posner & O. S. M. Marin (Eds.), Attention and Performance X (pp. 147-167). Hillsdale, NJ: Erlbaum.

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.

Sanders, A. F. (1981). Stage analysis of reaction processes. In G. E. Stelmack & J. Requin (Eds.), Tutorials in motor behavior. Advances in Psychology (Vol. 1, pp. 331-354). Amsterdam: North Holland.

Sherrington, C. S. (1906). The integrative action of the nervous system (2nd edition, 1947). New Haven, CT: Yale University Press.

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. 336-361). Hillsdale, NJ: Erlbaum.

Vaughan, H. G., Costa, L. D., & Ritter, W. (1972). Topography of the human motor potential. Electroencephalography and Clinical Neurophysiology, 25, 1-10.

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.

Appendix

The potential recorded at each of the two lateral electrodes (C3' and C4') can be considered as being given by the sum of the readiness potential generated locally (that is, in the region of the brain close to the electrode) and of scaled fractions of potentials generated in other parts of the brain and that are propagated by volume conduction. The other potentials may include the readiness potential generated underneath the contralateral electrode, as well as other ERP components. This can be expressed by the following formulae:

(1a) V(C3') = a1*V(C3't) + a2*V(C4't) + b1*V(ERP1) + ... + bn*V(ERPn)

(1b) V(C4') = a1*V(C4't) + a2*V(C3't) + b1*V(ERP1) + ... + bn*V(ERPn)

where:

V(C3') and V(C4') are the potentials recorded at C3' and C4';

V(C3't) and V(C4't) are the "true" readiness potentials generated underneath C3' and C4';

V(ERP) are potentials associated with other ERP components that may affect the recordings;

a1, a2, b1, and bn, are scaling factors.

By subtracting (1b) from (1a) we get (2):

(2) V(C3')-V(C4') = (a1-a2) * [V(C3't)-V(C4't)]

Thus, apart from a scaling factor, the difference between the potentials observed at the lateral electrodes is equal to the difference between the "true" readiness potentials.

While the derivation of the differential measure is quite straightforward, the separate measurement of the readiness potential on each side is more problematic. In fact, (1a) and (1b) can be transformed in (3a) and (3b):

(3a) a1*V(C3't) = V(C3') - a2*V(C4't) - b1*V(ERP1) - ... - bn*V(ERPn)

(3b) a1*V(C3't) = V(C3') - a2*V(C4't) - b1*V(ERP1) - ... - bn*V(ERPn)

From (3a) and (3b) it is evident that, in order to obtain the value of the readiness potential on each side, the contribution of components with a significant influence on C3' and C4' must be estimated and subtracted from the voltages recorded at C3' and C4'. Gratton, Coles, and Donchin (1987) have developed a procedure (Vector Filter) that allows the investigator to separate the contribution of ERP components characterized by different distribution over the scalp electrodes. This technique is based on the development of a series of "spatial filters," each specific for a different component. These filters are given by a series of weights, one for each electrode. The filtered data are represented by linear combinations of the values observed at each electrode. If sets of weights are chosen to be orthogonal, the corresponding filtered components will be orthogonal (although not necessarily uncorrelated). The Vector Filter procedure can result in a reduction of the contribution of overlapping components, although it cannot guarantee their complete elimination. Examples of the gains in component identification and measurement obtained by applying the Vector Filter procedure are given in Gratton, Kramer, Coles, and Donchin (1987) and in Fabiani, Gratton, Karis, and Donchin (in press).

In the analysis of the data of the present experiment we chose sets of weights according to the following criteria. First, they should suppress the contamination due to the contralateral readiness potential. For this reason, the two sets of weights used for estimating independently the readiness potential on each side were selected so as to be orthogonal (i.e., their cross-product was equal to 0). Second, they should be such as to suppress the influence of other ERP components. In particular, we filtered out the contribution of components with an exclusively midline distribution. This was accomplished by giving equal weights to the midline electrodes (Fz, Cz, and Pz). Third, to eliminate general components (with an equivalent influence at all sites), the sum of each set of weights was equal to 0.

The Vector Filters used were the following:

(4a) C3't = 0.888*C3' - 0.114*C4' - 0.258*Fz - 0.258*Cz - 0.258*Pz

(4b) C4't = 0.888*C4' - 0.114*C3' - 0.258*Fz - 0.258*Cz - 0.258*Pz

Note that the sum of squares of each set of weights is equal to 1, so that the Vector Filter amounts to a rotation of axes in the space defined by the electrode locations.

To provide some indication of the validity of this procedure, we computed the average correlation across trials, over the six subjects, between C3' and C4' (in the last 100 ms before array presentation) before and after applying Vector Filter. As can be seen by comparing (3a) and (3b), the "undesired" components (i.e., the contralateral readiness potential, as well as the other ERP components) will produce similar effects on C3' and C4', and will therefore tend to increase the correlation between the potential recorded at these two electrodes. The average correlation was .83 $\pm$.03 before correction, and .00 $\pm$.08 after correction. Note that the latter correlation could have been different from 0 even though orthogonal sets of weights were used. This would occur if the procedure under- or over-estimated the effects of volume conduction, or if there was some "structural" or "functional" association between the brain electrical activity generated beneath the two lateral electrodes. The latter might occur under conditions of "response competition" or "response energizing."

Figure Legends

Figure 1. Conditional accuracy functions (upper panels) and proportion of trials for each response latency bin (lower panels) for compatible (solid) and incompatible (dashed) trials. Values on the ordinate correspond to the midpoint of each response latency bin. Data for trials classified on the basis of criterion squeeze response are shown in the left panels, and for trials classified on the basis of the EMG onset in the right panels. For the conditional accuracy functions based on EMG onset latency, standard errors are indicated by the vertical bars. For criterion squeeze response functions, only three subjects had trials in the 100-149 ms latency bin.

Figure 2. Grand-average ERP waveforms depicting the voltage difference between the scalp electrodes contralateral and ipsilateral to the correct response for correct (solid) and incorrect (dashed) response trials, from the incompatible condition. The waveforms in the upper panel are averages for trials with response latencies between 300 and 349 ms, the waveforms in the lower panel refer to trials with response latencies between 150 and 199 ms. Upward deflections indicate lateralization (larger negativity) toward the scalp electrode contralateral to the correct response, downward deflections indicate lateralization toward the scalp electrode ipsilateral to the correct response (and contralateral to the incorrect response).

Figure 3. Conditional accuracy functions, averaged over compatible and incompatible conditions, as a function of the lateralization of the readiness potential during the last 100 ms of the foreperiod.

Figure 4. Upper panel. Conditional accuracy functions for compatible and incompatible trials. Middle panel. ERP waveforms of the lateralized readiness potential following array presentation, averaged separately for all compatible and for all incompatible trials. Lower panel. ERP waveforms

of the lateralized readiness potential following array presentation, for compatible and incompatible correct trials with a response latency between 300 and 349 ms.

Figure 5. ERP waveforms of the lateralized readiness potential for the period beginning 100 ms before the warning tone and ending at the time of the peripheral response to the array. Separate waveforms are shown for compatible correct (upper panel), incompatible correct (middle panel), and incompatible incorrect trials (lower panel), for four response latency bins. The solid vertical lines in each panel indicate the response latency for each bin. The upper and lower horizontal lines in each panel indicate the inferred thresholds for correct and incorrect muscle response emission.

Figure 6. ERP waveforms representing the correct (solid) and incorrect (dashed) single-side readiness potentials for compatible correct trials. Separate average waveforms are shown for four response latency bins. The dashed vertical lines after array presentation indicate the response latency for each bin. The upper horizontal line in each panel indicate the inferred threshold for muscle response emission.

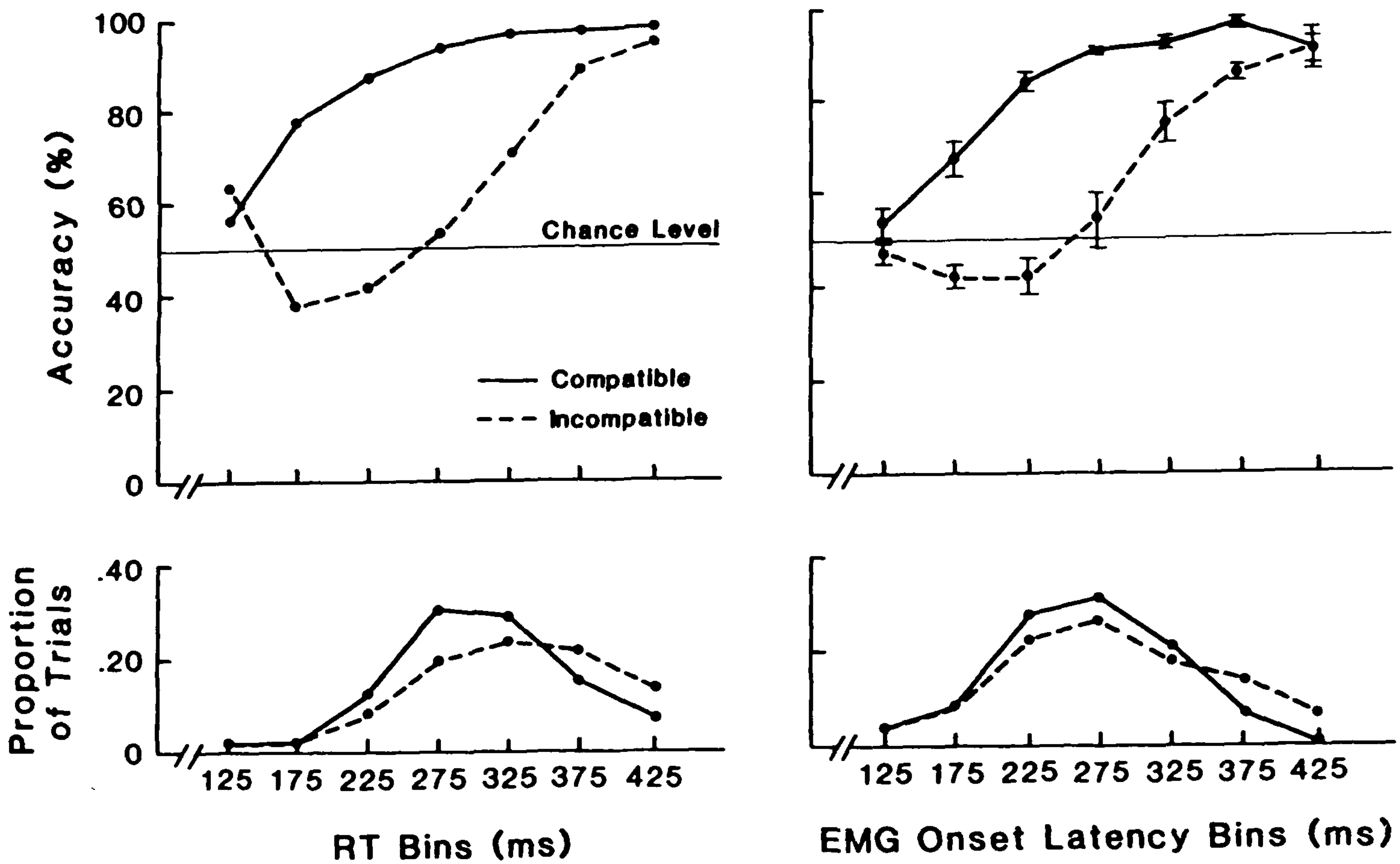
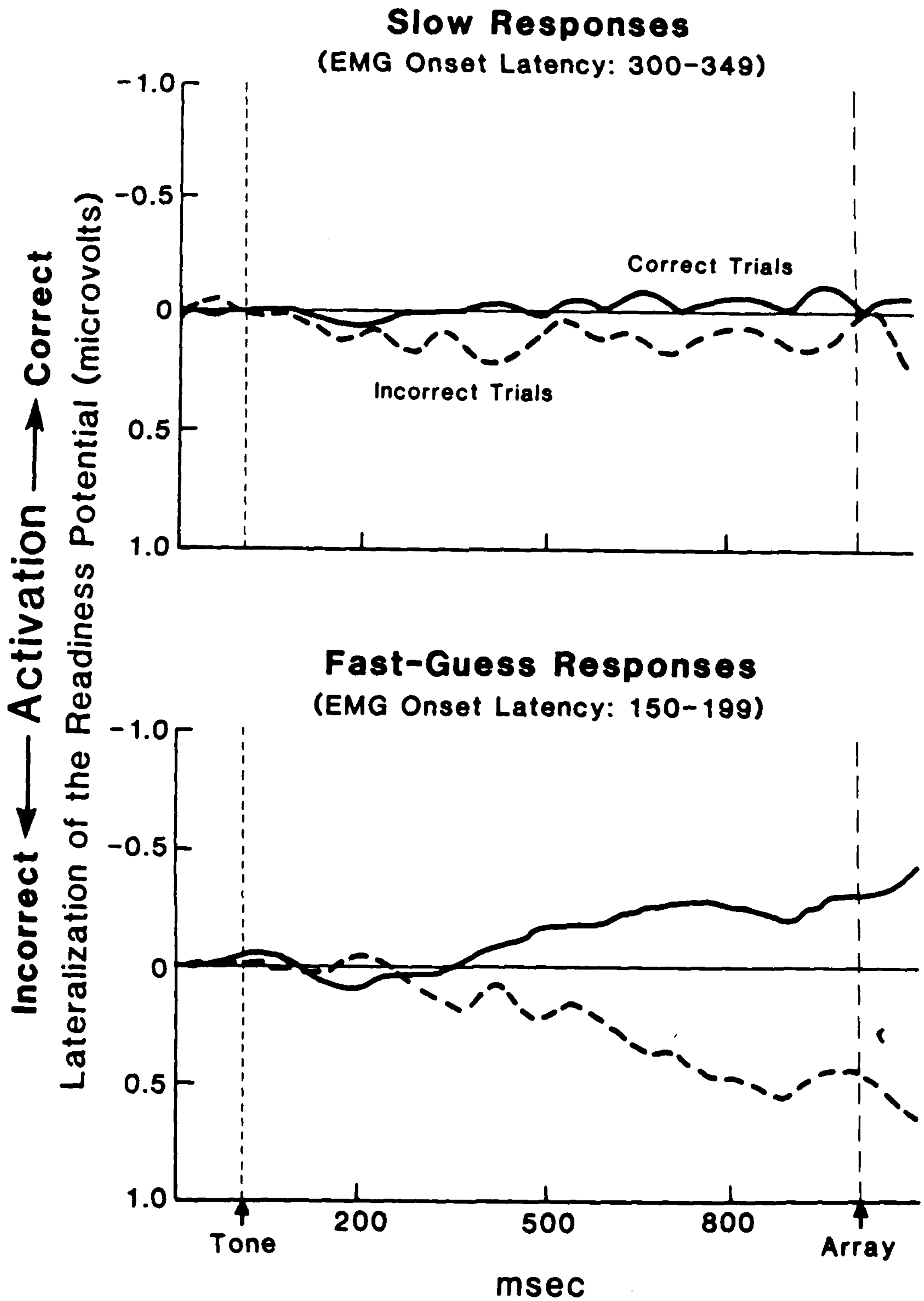
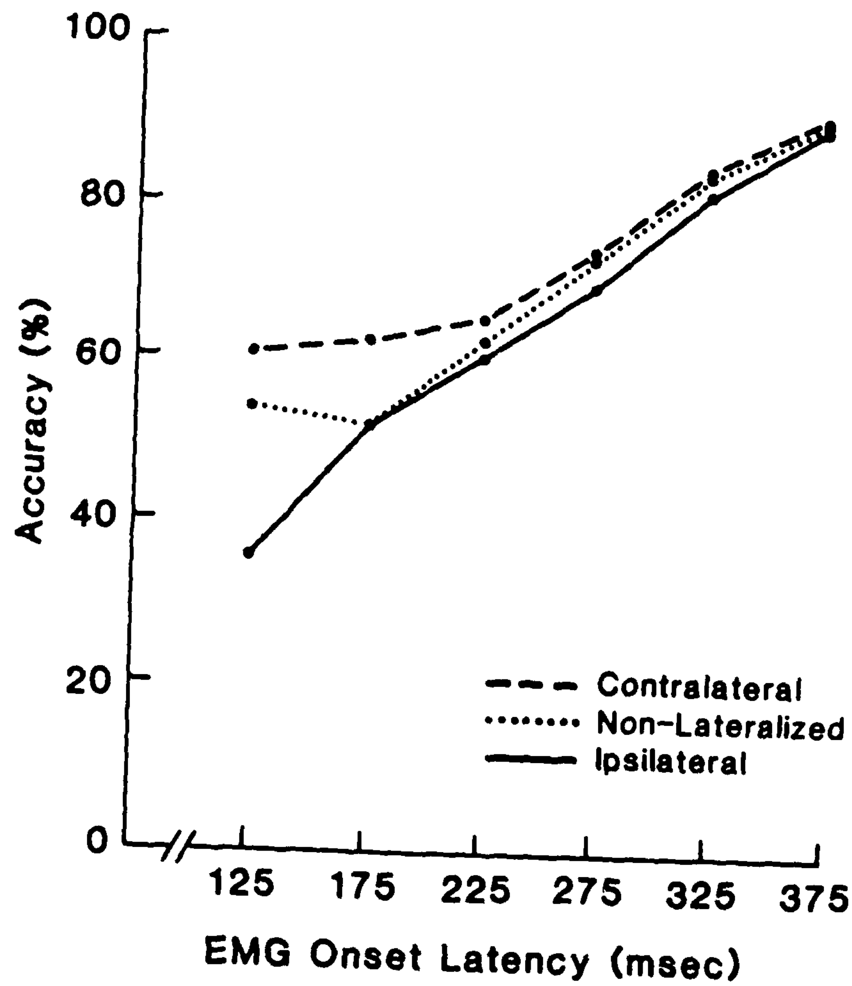

100
80
60
40
20
0
Accuracy (%)
Chance Level
Compatible
Incompatible
.40
.20
0
Proportion of Trials
125 175 225 275 325 375 425
RT Bins (ms)
125 175 225 275 325 375 425
EMG Onset Latency Bins (ms)

Fig. 1



Slow Responses
(EMG Onset Latency: 300-349)
Correct Trials
Incorrect Trials
Fast-Guess Responses
(EMG Onset Latency: 150-199)
-1.0
-0.5
0
0.5
1.0
Incorrect ← Activation → Correct
Lateralization of the Readiness Potential (microvolts)
Tone
200
500
800
Array
msec

Fig. 2



100
80
60
40
20
0
Accuracy (%)
Contralateral
Non-Lateralized
Ipsilateral
125
175
225
275
325
375
EMG Onset Latency (msec)

Fig. 3

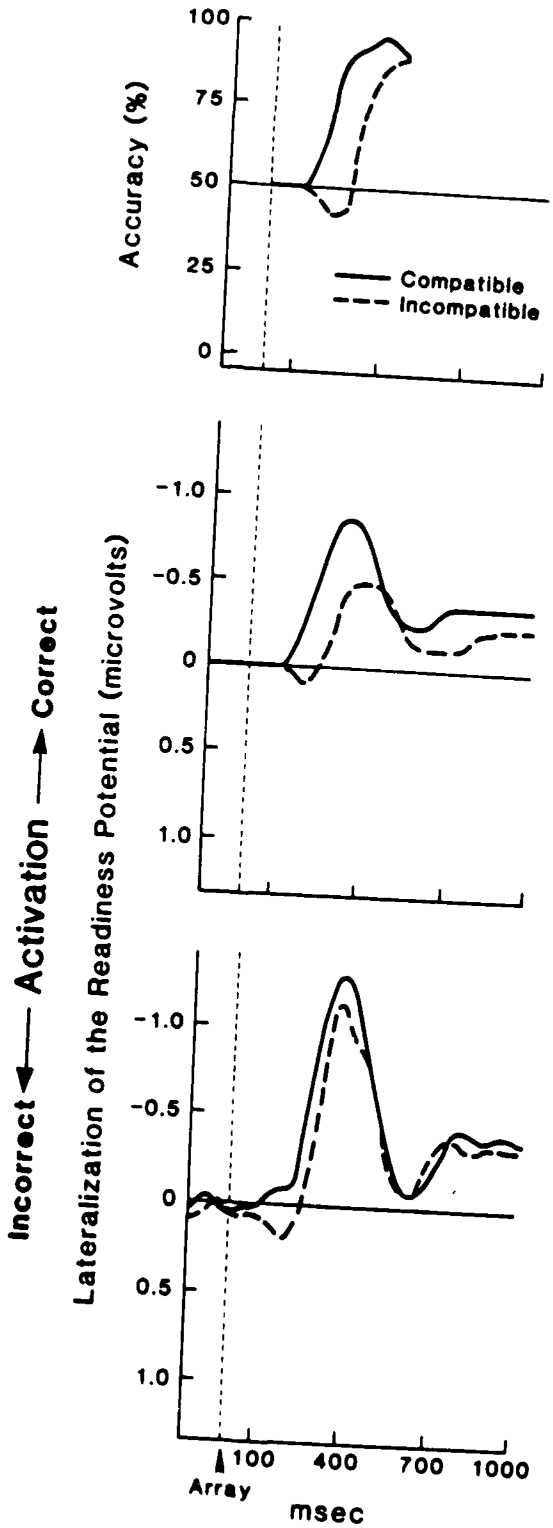
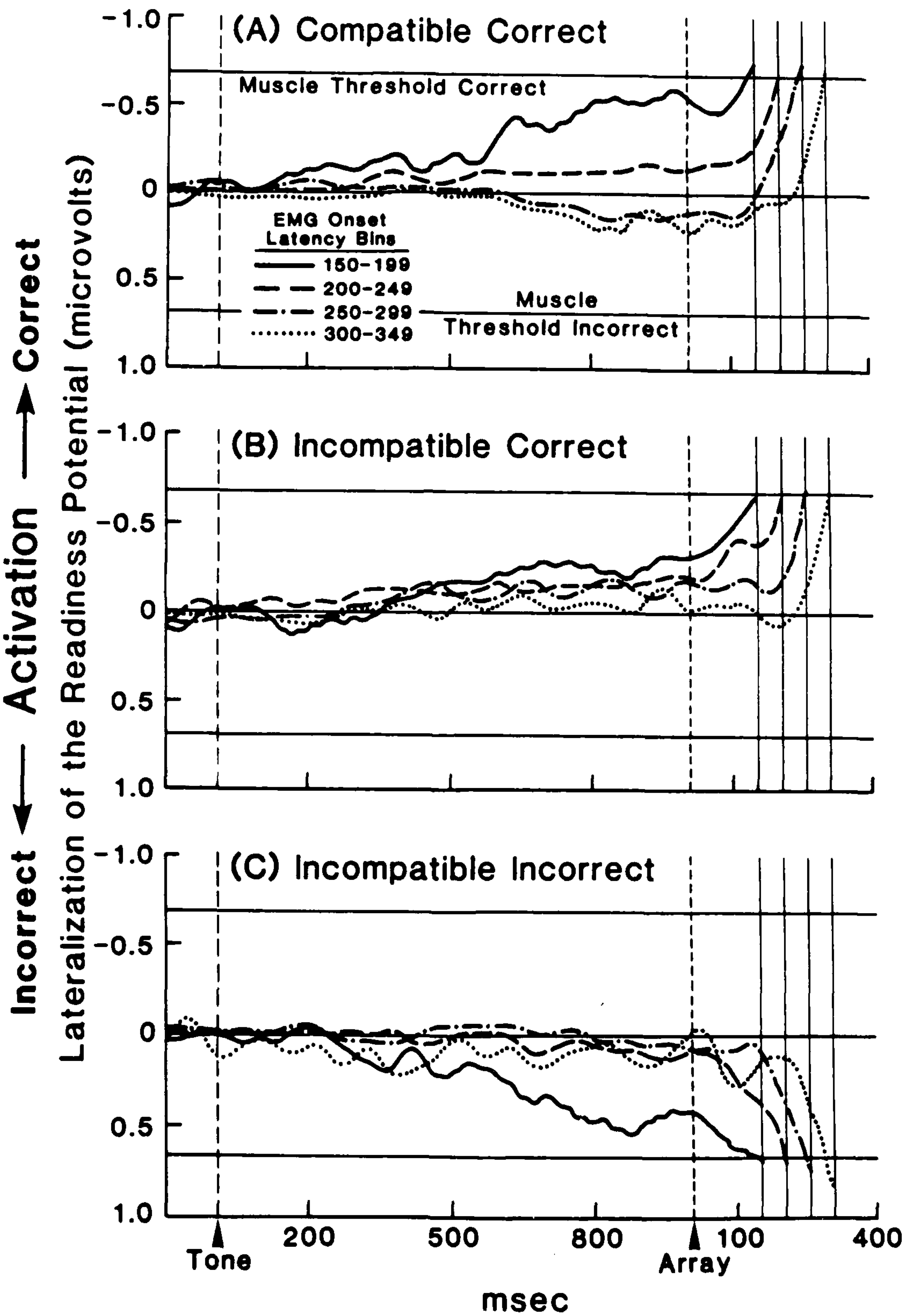

Accuracy (%)
100
75
50
25
0
Compatible
Incompatible
Incorrect ← Activation → Correct
Lateralization of the Readiness Potential (microvolts)
-1.0
-0.5
0
0.5
1.0
100
400
700
1000
Array
msec

Fig. 4



(A) Compatible Correct
Muscle Threshold Correct
EMG Onset
Latency Bins
150-199
200-249
250-299
300-349
Muscle
Threshold Incorrect
(B) Incompatible Correct
(C) Incompatible Incorrect
-1.0
-0.5
0
0.5
1.0
200
500
800
100
400
Tone
Array
msec
Incorrect ← Activation → Correct
Lateralization of the Readiness Potential (microvolts)

Fig. 5

Compatible Correct

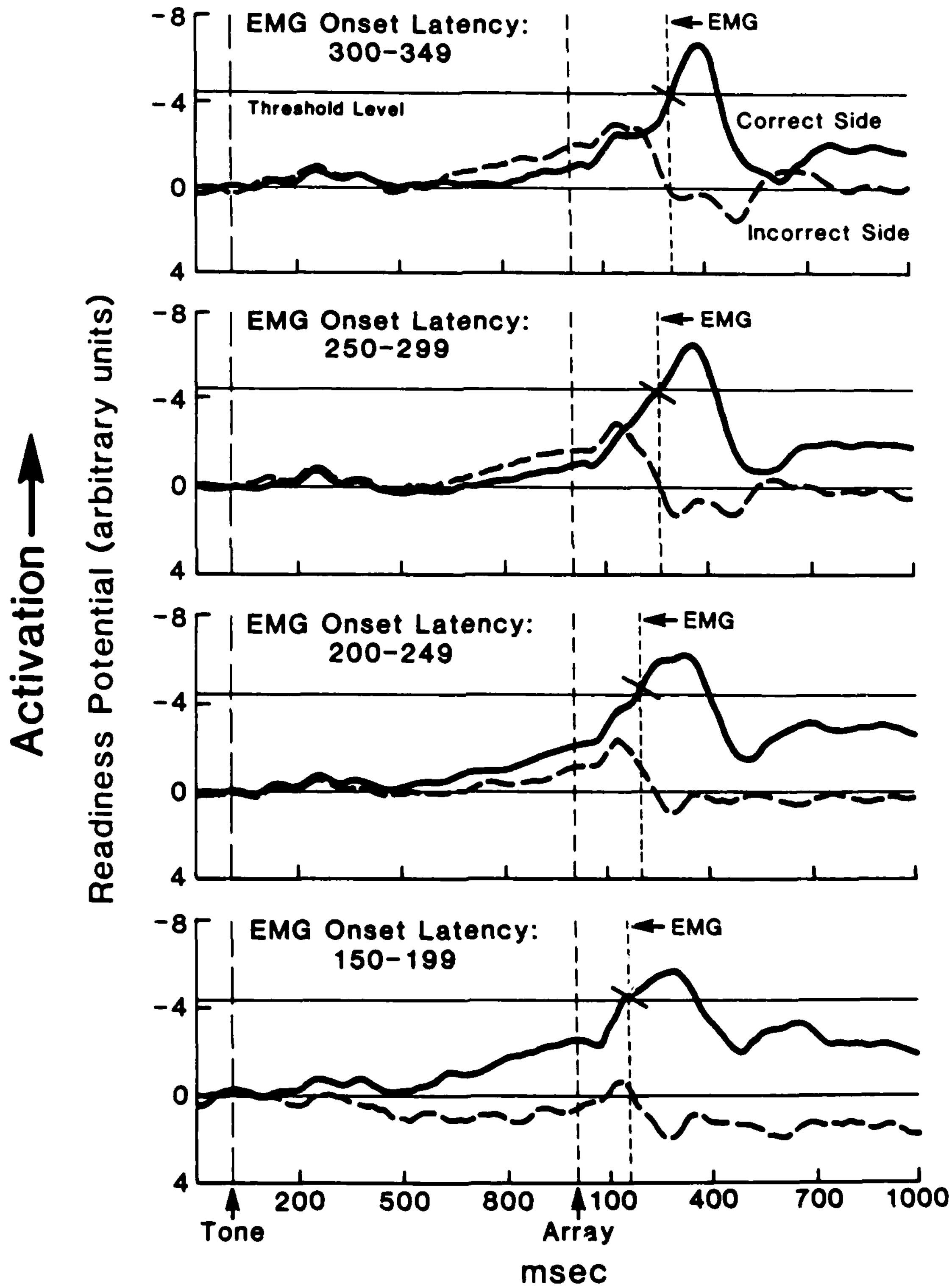


Fig. 6

Article 9.

Psychophysiology and the Electronic Workplace
Edited by A. Gale and B. Christie

9

Event-Related Brain Potentials

ARTHUR F. KRAMER

INTRODUCTION

In many cases, automation has served to increase our productivity and unburden us from the requirement of performing mundane, repetitive tasks. However, an important issue concerns the degree to which automation has unburdened the human operator from duties as a controller of low level system parameters by increasing the demands on supervisory duties such as decision-making and information management (Wickens and Kramer, 1985). As an example of the changing demands, consider the task of a worker in a manufacturing plant. In the past, it was relatively easy to determine the workload and efficiency of an individual employed in the operation of a drill press machine. The precision and speed with which the operator made his measurements and performed the drilling operation could easily be determined by a time study analysis of the process and an examination of the finished product. However, in modern day manufacturing plants this individual and his co-workers would be replaced with a series of computer controlled drill presses that perform measurements and adjustments of the process on the basis of information obtained from their sensors. The human operator would be relegated the task of monitoring the process and intervene only infrequently when an abnormality in the system had been detected. At this point the operator would be called upon to diagnose the malfunction and begin the sequence of procedures required to return the system to a normal state of operation. Thus, the human operator in an automated system spends a great deal of time engaging in information processing activities which are not readily accessible with our traditional measurement techniques. How then do we evaluate the demands these tasks impose upon the limited processing capabilities of the human operator? What constitutes a measure of efficient performance given that the operator is only infrequently called upon to make a manual response? How do we assess the strategies that operators employ to cope with the different task requirements of normal system monitoring and the occasional but usually critical detection and diagnosis of system abnormalities?

One method which has been employed in an attempt to provide access to the information processing required of operators of highly automated systems is the verbal protocol (Bainbridge, 1974). Although, in many situations verbal protocols have been found useful in providing information about the strategies and planning activities of human operators, there have been other instances in which operators have been unable to provide a coherent description of their cognitive activities (Broadbent, 1977). Another approach to the assessment of the processing demands imposed upon operators of modern day systems is the use of psychophysiological techniques. In this chapter one specific psychophysiological technique, the event-related brain potential (ERP), will be introduced and its sensitivity to perceptual, cognitive and motor demands will be evaluated.

OVERVIEW OF ERPs

This section describes the component structure of ERPs, briefly illustrates the advantages and pitfalls of different measurement procedures, and discusses the current state of knowledge of the neuroanatomical substrates of different ERP components. The following section describes the current and potential applications of ERPs to problems in human factors engineering.

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. This temporal relationship between the ERP and the eliciting stimulus or response is what differentiates ERPs from the ongoing electroencephalographic (EEG) activity (see Chapter 10 of this volume for a description of EEG). Thus, the EEG provides a measure of the tonic state of the organism while ERPs reflect phasic changes related to the processing of specific events.

Description and definition of ERP components

It is important to recognize the componential nature of the ERP. Early studies which investigated the effects of different stimulus parameters on the ERP treated the waveform as a unitary entity, measuring the amplitude over the entire recording epoch. Many of these results were difficult to interpret. The effects of experimental manipulations tend to be quite specific to a few components and a combination of measures may obscure the relevant variance. Although there is a degree of controversy as to the proper identification and definition of components, ERPs have generally been viewed as a sequence of separate but sometimes temporally overlapping components which are influenced by some com-

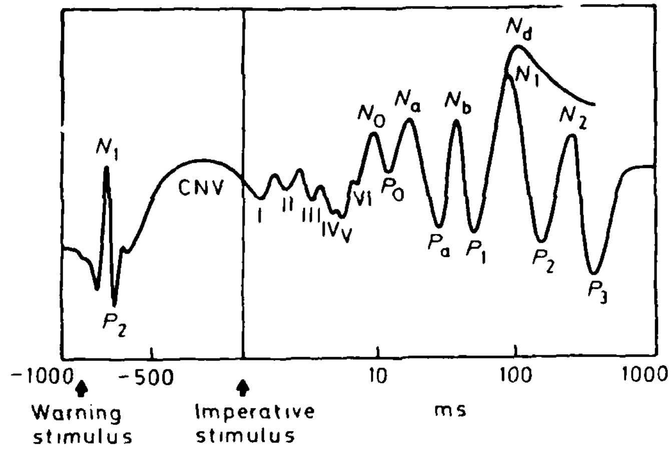


Fig. 1 Schematic representation of an auditory event-related potentials (ERPs) illustrating prestimulus components as well as exogenous and endogenous components. Negative components are represented by an upward deflection in the waveform. (After Kutas and Hillyard, 1984, reprinted from *Handbook of Cognitive Neuroscience* by permission of the University of Illinois.)

bination of the physical parameters of the stimuli and psychological constructs such as expectancy, task relevance, attention and memory. Figure 1 presents the series of components which are normally recorded with the presentation of an auditory stimulus. Similar diagrams can be drawn for visual and somatosensory modalities.

Components are typically labeled with an 'N' or a 'P' denoting negative or positive polarity, and a number indicating their minimal latency measured from the onset of an eliciting event (for example, N100 is a negative going component which occurs at least 100 ms after a stimulus). Components may be categorized along a continuum from exogenous to endogenous. The exogenous components represent an obligatory response of the brain to the presentation of a stimulus. These components are usually associated with specific sensory systems, occur within 200 ms of a stimulus, and are primarily sensitive to the physical attributes of stimuli. For example, exogenous visual potentials are influenced by the intensity, frequency, hue, patterning and location of the stimulus in the visual field. The exogenous components have been successfully used in clinical settings to monitor the functional integrity of the nervous system during surgical procedures, to assess changes in the nervous system as a result of maturation and aging, and to help diagnose various types of neuropathology including tumors, lesions and demyelinating diseases such as multiple sclerosis (Starr, 1978; Stockard, Stockard, and Sharborough, 1979). The endogenous components, on the other hand, occur somewhat later than the exogenous components and are not very sensitive to changes in the physical parameters of stimuli, especially when these changes are not relevant to the task. Instead these components are primarily influenced by the processing demands of the task imposed upon the subject. In fact, endogenous components can even be elicited by the absence of a stimulus if this 'event' is relevant to the subject's task. The strategies, expectations, intentions and decisions of the subject as well as task parameters and instructions account for the majority of the variance in the endogenous components.

The importance of the componential nature of the ERP in the assessment of organismic state and information processing has made it imperative that components be clearly defined. The labeling of different peaks and troughs in Figure 1 suggests that some basis exists for the categorization of ERP components. The attributes of the ERP that have served as definitional criteria include: the distribution of voltage changes across the scalp, latency range, polarity, sequence, and the sensitivitty of components to manipulations of instructions, task parameters and physical changes in the stimulus (Donchin, Ritter, and MCallum, 1978; Kramer, 1985). The scalp distribution refers to the relative amplitude and polarity of the 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 range depends on the experimental manipulations as well as the specific component. For example, the components occurring within 10 ms of the presentation of a stimulus, the brainstem evoked potentials, are influenced by both organismic and stimulus variables but their latency range is only a few ms. On the other hand, the latency range of the P300 component depends on the processing requirements of the task and can span several 100 ms. The sensitivity of components to specific experimental manipulations is perhaps the most important of the definitional criteria. In fact, it has been suggested that components with different scalp distributions but a similar relationship to task parameters or instructions be defined as the same component (Ritter, Simson, and Vaughan, 1983). The emphasis on the functional relationship between components and experimental manipulations underscores the importance of recording ERPs in situations in which other measures of performance may also be obtained.

The preceeding exposition was not intended to suggest that ERP components can be easily defined on the basis of a tidy set of criteria. On the contrary, each of the criteria is fraught with several measurement and interpretative problems and the interaction between them is quite complex. However, given that we are aware of the limitations of the methodology, the judicious use of ERPs in conjunction with other approaches can provide useful insights into the informational transactions of the brain. The next section describes some of the issues which must be addressed during the measurement of the ERP components.

Measurement of ERP components

The identification of measurement of ERP components is complicated by a relatively low signal-to-noise ratio in the single trial data. The ERP is typically defined as that portion of the waveform which is time-locked to a stimulus or response. All other variability, not time-locked to the discrete event, is considered noise (for instance, ongoing EEG). Unfortunately the noise distribution possesses an amplitude of between 50 and 100 microvolts (μV) while ERP amplitudes range from 1 to 15 μV. Several methods have been proposed to enhance the signal-to-noise ratio. The most commonly employed procedure is signal averaging. The basic method consists of the repetition of a large number of essentially identical trials. These single trials are then averaged. The degree of attenuation of the noise is inversely proportional to the square root of the number of single trials in the sample. The averaging procedure is usually not much of a burden in a laboratory setting since repetition of trials is required to calculate mean reaction times.

Although the signal averaging procedure serves as an efficient method of reducing the noise and thereby making the extraction of the components easier, there are several assumptions which must be met prior to its use (Coles, *et al.*, 1986). These assumptions include; (1) the ERP components must be temporally invariant over repeated presentations of the stimulus; (2) the morphological characteristics of the ERP must be invariant over trials; and (3) the noise must not be systematically related to the components. However, if the temporal invariance assumption is not met, iterative cross-correlation procedures can be employed to temporally align the single trial waveforms prior to averaging (Woody, 1967). Other procedures such as filtering also serve to reduce the amplitude of the noise relative to the amplitude of the ERP provided that the two distributions are not highly correlated.

Although in many cases the averaging procedure is an adequate way to extract the ERP from the background noise, there are situations in which an analysis of the single trials would be desirable. For example, if the objective was to use the ERP as a real-time index of the processing demands imposed upon a human operator, averaging would be clearly inappropriate. One technique which has been successfully employed to distinguish between ERPs on a single trial basis is stepwise discriminant analysis (SWDA). The goal of the SWDA procedure is the selection of a subset of variables which maximise the intergroup separation. In terms of its application to the analysis of ERPs, the variables are the time points along the waveform. The objective is to discriminate between ERPs elicited by different stimuli or different experimental manipulations (for example, levels of processing demand).

Once the ERP has been extracted from the background noise, the characteristics of the individual components must be quantified. The simplest, and most frequently employed method to accomplish this function is the selection of the largest or smallest voltage value within a prespecified temporal window. Thus, in this case a component is defined as a single peak or trough in the waveform. The peak selection procedure provides a measure of the amplitude and latency of the component. The principal advantages of this procedure are its intuitive appeal and computational simplicity. Peak measurement algorithms represent a direct analog of visual inspection of the waveform, with the added advantage of an easily standardized procedure. However, there are also several disadvantages of the peak measurement procedure. The definition of a component in terms of a single point in addition to being fairly arbitrary for slow components, ignores the information which is provided by the morphology of the waveform. Peak measurement techniques also fail to provide information concerning component overlap. The measurement of a single point does not permit the assessment of the actual number of temporally overlapping components which may jointly be responsible for the voltage recorded at a specific time point. A multivariate technique which has been used to effectively deal with the problem of component overlap is the principal component analysis (PCA) procedure. The PCA also defines a component in terms of a series of time points which are weighted to reflect their contribution to the variance of that component, thus avoiding the definition of a component as a single point (Donchin and Heffley, 1979).

Neuronal substrates of ERPs

The reasons for recording ERPs are probably as numerous as the number of components that have been catalogued. However, it is possible to illustrate the range of variability by describing the two extremes. On the one hand are those investigators who are specifically interested in using ERPs as tools to aid in our understanding of human information processing. In this case, ERPs can be viewed as another dependent variable no different from measures of reaction time and percentage correct. In fact, it does not really matter where in the brain the ERPs are generated. On the other end of the continuum, the interest is in using ERPs to enhance our understanding of the anatomy and physiology of the brain. In this case, the manipulation of task parameters and experimental instructions is of interest only insofar as they further our understanding of neural function. Instead the primary goal is to localize the source of different ERP components. Of course, most investigators would subscribe to a combination of these views; that the optimal use of ERPs requires the knowledge of both the neuroanatomical sources and the functional significance of ERPs in terms of human information processing. Thus, in actuality these two disparate views are complementary rather than competitive. In the present section our knowledge of the neuronal substrates of ERPs will be described. The subsequent sections will be concerned with the use of ERPs as a tool in the study of human information processing in complex, person–machine systems.

The electrical potentials recorded at the scalp represent a small subset of the neuronal transactions of the brain. These scalp-recorded ERPs primarily reflect the summation of synchronous neuronal activity. The ability to observe this activity is constrained by the geometry of the cell groups generating the electrical field, the conductivity of the brain tissue, and the distance of the generator source from the recording electrodes (Wood and Allison, 1981). For example, cell groups which are arranged in an open field pattern, with their cell bodies and processes aligned in parallel, generate electrical fields that can be recorded at a considerable distance. However, the arrangement of cells in a closed field, with their cell bodies and processes aligned in a radial pattern, generate an electrical field that can only be recorded at the center of the field. Thus, this type of activity would not be reflected in the scalp-recorded ERP but could only be detected if an intracranial electrode was located at the center of the field. The degree to which the electrical fields generated within the brain are accurately represented at the scalp is also dependent upon the depth of the generating source(s). For sources located in the cortex, the absolute amplitude recorded at the scalp will be large, but small changes in the location of the electrode will produce large changes in the amplitude of the ERP. However, for more distant sources the absolute amplitude recorded at the scalp will be small, but changes in the location of the recording electrode will have relatively small effects on the amplitude of the ERP. Thus, the conclusions drawn about the efficiency of the brain as a volume conductor are dependent upon the geometry of the cell populations as well as the distance of the generator sources from the recording sites.

Although there is some knowledge of the neuroanatomical sources of the early sensory components, there is little information on the locus of the later, endogenous components. For example, on the basis of clinical and experimental evidence the brainstem evoked potentials can be localized to specific anatomical structures along the ascending sensory pathways (Goff, Allison, and Vaughan, 1978). In most cases, these components appear to be generated by a single source. However, in the case of the later endogenous components it is generally assumed that the electrical fields of several generators summate to produce the component recorded at the scalp. Thus, the problem of source localization is complicated by the requirement to determine the number of sources as well as their interactions.

The distribution of voltages recorded at the scalp is used as one criterion for the definition of an ERP component. It is generally assumed that a scalp distribution which is invariant across repeated stimulus presentations implies a specific and fixed set of neuronal generators. Thus, one method to derive the source of a scalp-recorded ERP would be to extrapolate back from the surface potential field to the source location. Unfortunately, there is no unique solution for this 'inverse' mapping from a source field to a neuronal generator. However, analytical solutions have been derived which enable the calculation of surface potential fields on the basis of a proposed set of generators (Wood et al., 1984). These 'direct' procedures perform an iterative best-fit solution between the calculated and empirical fields, given specific assumptions about the location and orientation of the source(s). Other techniques which have also proven useful in providing information on the neuroanatomical sources of scalp-recorded ERPs include: intracranial recordings in humans, studies of the effects of brain lesions on ERP components, the recording of evoked magnetic fields, and the use of animal models of ERP phenomena. Although none of these techniques in and of itself is sufficient to unequivocally localize the neuronal sources of scalp recorded ERPs, each of the techniques provides a useful source of converging information.

HUMAN FACTORS ENGINEERING: A PSYCHOPHYSIOLOGICAL APPROACH

The discussion thus far has focused on the technical issues involved in the definition, measurement and interpretation of ERPs. In the present section the emphasis is on the utilization of ERPs as tools in the study of problems in the design of tasks and the assessment of human performance and cognition in the electronic workplace. By advocating the use of ERPs to investigate human information processing in complex tasks, we are not suggesting that ERPs supplant the more traditional behavioral and subjective measurement techniques. If ERPs provided information which was completely redundant with that obtained from other measures there would be no justification for employing this complex and time-consuming technique. However, there are situations in which ERPs can be used to augment the information provided by other measures. Several examples will be discussed in which the dissociation as well as the correspondence between ERPs and other assessment procedures has enhanced our understanding of the capabilities, limitations and strategies of human operators.

Display image quality

A major concern in the design of present-day interactive systems is the interface between the hardware, in most cases a computer system, and the human operator. This problem is not unique to specific work environments but covers the range from the word processor in an office to the pilot of a high performance aircraft. In many cases the system requirements dictate that a vast amount of alphanumeric and graphic information be presented to the operator via some type of video display terminal. An important question in the realm of interface design concerns the level of resolution which is necessary for the operator to efficiently perform the assigned tasks. The answer to this question is at least partially dependent on the types of tasks to be performed. For example, the detection of a flashing warning light can be performed with much lower resolution of the display systems than the detection of velocity changes of objects moving across the screen. This in turn can be performed with lower resolution than the precise reading of several rapidly changing quantitative variables. However, in all of these situations it is important to have the capability to predict how changes in specific display parameters affect the extraction and processing of the visual information. In the case of one of the most commonly used display devices, the raster scan display, numerous psychophysical investigations have been conducted to assess the effects of display parameters such as the number of raster lines per unit distance, noise level, and the bandwidth of the video signal on the performance of various types of tasks (Grether and Baker, 1972; Pearson and Pearson, 1985). Complementary information concerning the effects of display image quality on visual processing has been obtained through subjective assessment procedures.

The sensitivity of the visual evoked potential (VEP) to changes in the physical parameters of stimuli such as intensity, spatial frequency, hue and orientation suggests that it might provide a useful index of the image quality of video displays. VEPs have been used successfully as an objective measure of visual acuity. Decreasing the clarity of focus of a checkerboard pattern on the retina through the use of a distorting lens leads to a decrease in the amplitude of the VEP. On the other hand, insertion of a lens which corrects refractive errors enhances the amplitude of the VEP (Harter and White, 1968). Thus, the amplitude of the VEP may be used as a measure of the clarity of focus of a visual stimulus.

In many cases, the VEP is elicited by the infrequent occurrence of a brief stimulus. Thus, a patterned field may be presented once a second for a duration of 100 ms. These 'transient' VEPs can be decomposed into a series of separate components which occur within 200 ms of the presentation of the eliciting stimulus. The effects of changes in the physical parameters of the stimuli can be assessed in terms of the latency and amplitude of these components. Another method of eliciting VEPs is to present a continuously flickering stimulus at a rate of between 10 and 50 Hz. After a few seconds of stimulation the brain begins to produce a VEP with a roughly sinusoidal waveshape at the same frequency as the eliciting stimulus. These 'steady state' VEPs (SSVEP) can be quantified by decomposing the waveform into its fundamental and harmonic frequency components. The effects of changes in the physical stimulus parameters can then be specified in terms of the phase and amplitude of the SSVEP components.

Both the transient and steady state VEPs have advantages and each has been used successfully to assess the functional integrity of the visual system as well as to evaluate changes in the physical parameters of stimuli. The complexity of the transient VEP permits the mapping of latency and amplitude changes of different components onto the manipulation of stimulus duration, contrast, orientation and spatial frequency. However, the small amplitude of these components relative to the ongoing EEG makes it necessary to relate changes in the physical aspects of the stimulus to an average VEP composed of several minutes of stimulus presentations. The major advantages of the SSVEPs is that these potentials can be reliably defined after only a few seconds of stimulation.

Gomer and Bish (1978) evaluated the effects of two display system parameters, horizontal resolution and gray shade levels, on the amplitude of the SSVEP. Subjects viewed binocularly a checkerboard pattern which reversed at a rate of approximately ten times per second. Gray shade level was varied by manipulating the modulation contrast of the checkerboard pattern. Displays with seven, five and three gray shades were employed in the study. Horizontal resolution was varied by low-pass filtering the video signal, producing a display with either 185, 305 or 955 raster lines. Figure 2 illustrates the effects of the manipulation of the display parameters on the SSVEP. Both horizontal resolution and the number of gray shades had a significant effect on the amplitude of the SSVEP, with resolution exerting a greater influence than gray shade. The amplitude of the SSVEP increased with increasing clarity of the visual display. These results suggest that SSVEPs may provide an accurate and rapid measure of the effects of stimulus parameters on the resolution of a visual display. Additional research employing more sensitive analytical techniques will be required to investigate a larger set of display factors in a wider variety of tasks.

Visual selective attention

The increasing levels of automation in the workplace have shifted the role of the human operator from that of a manual controller of low level system parameters to that of a decision-maker and information manager. One implication of this change in roles has been the increasing demands placed upon operators to extract and process a huge amount of visually presented information (Wickens, 1984). Depending upon the specific system in which the human is involved, the time-lag between the extraction of this data from a visual display and an overt control action indicating that the information has been processed may be on the order of several seconds, minutes or even hours. Thus, a major concern in these semiautomated systems is the timely assessment of the degree to which the human operator has perceived and processed the task relevant information. This becomes especially important in adaptive systems that attempt to assess the operator's mental model of the state of the system and present information in a form compatible with this model.

A method of evaluating the operator's proficiency in dealing with the available information is to infer processing capabilities and strategies on the basis of overt response sequences. This is usually adequate in situations in which the time-lag between the presentation of the information and the subsequent control action is short, and the system is functioning normally. However, with longer time-lags or abnormal system operations, a more timely assessment method is needed. One solution has been to examine operators' eye scanning strategies in an attempt to decompose the mental processes which intervene between the

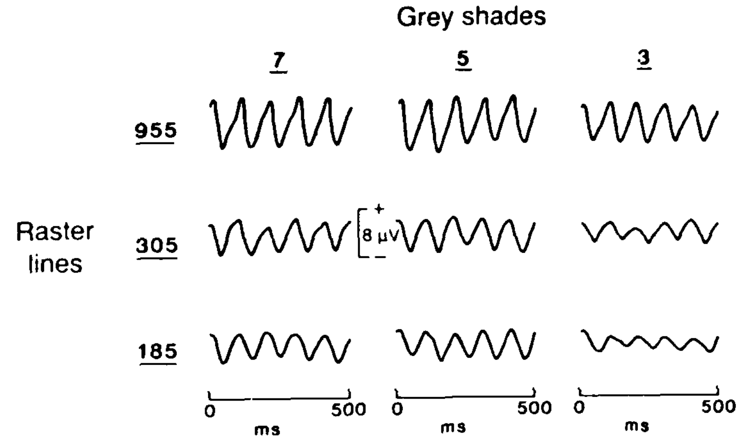


Fig. 2 Average SSVEPs recorded from one observer in each of the experimental conditions. Note that the amplitude of the SSVEP increased with increases in the number of gray shade levels. (After Gomer and Bish, 1978, from *Human Factors*, **20**, 589–596, by the Human Factors Society, Inc. and reproduced by permission.)

control actions (Moray and Rotenberg, 1985). In this method fixation has been equated with the allocation of attention. Thus, when we focus our eyes on an object it is assumed that we are attending to and processing the information available in that object (for example, size, color, shape, orientation). In many cases this assumption appears to provide an adequate representation of operators' information extraction and processing strategies. However, it has also been shown that we are capable of directing our attention to objects in the visual environment in the absence of eye movements (Posner, 1980). For example, an operator may be fixating one display, while concurrently extracting information from another, peripherally located display. Even in situations in which attention is allocated solely within a fixation, it is difficult to determine which of the properties of the display the operator is processing (Neisser and Becklen, 1975). Thus, although eye movements provide an adequate account of attentional allocation in some situations, they do not provide a fine-grained analysis or peripheral processing or resolve the degree to which different attributes of a fixated object are processed.

One line of evidence for the dissociation between fixation and attention has been provided by studies of visual ERPs. In a typical experiment, subjects are instructed to fixate their eyes on the center of a CRT while focussing their attention on either the right or the left visual field. Stimuli are presented in a random order to both fields. The subject's task is to respond to an occasional target among the more frequent standards in the attended field. The infrequently occurring targets are distinguished from the standards by a simple physical feature such as intensity, size or color. ERPs are elicited by the presentation of these stimuli and averages are computed for the targets and standards within the attended and unattended visual fields. The basic effect obtained in this paradigm is an enhancement of a series of positive and negative peaks in the ERPs elicited by the stimuli in the attended field (Eason, Harter, and White, 1969). Furthermore, the size of the enhancement appears to depend on the extent to which subjects are able to focus their attention on a specific location in the visual field. The effect is diminished in cases in which subjects must divide their attention between the right and left visual fields (Van Voorhis and Hillyard, 1977). These effects are most consistently obtained for two exogenous components, the P100 and N100, which occur between 100 and 200 ms post-stimulus.

Other ERP components have been found to discriminate between targets and standards within an attended location. Thus, while the P100 and N100 are enhanced for any stimulus in an attended field, the P300 component is largest for the target stimulus in an attended location. This pattern of data suggests that the early exogenous components, the P100 and N100, and the endogenous P300 component reflect two hierarchically ordered levels of selection. It appears that the P100 and N100 components reflect selection between locations in the visual field while the P300 is sensitive to differences among stimuli within an attended location.

Additional evidence for a hierarchy of cue selections has been found in studies which have varied dimensions of visual stimuli other than location, such as the orientation, size, shape, brightness or color of an object. In one study, subjects were instructed to respond selectively to a grating of a specific spatial frequency and orientation (Previc and Harter, 1982). The stimuli presented to the subjects were composed of one of two spatial frequencies and were oriented either horizontally or vertically. Thus, any one of four stimuli could occur on any trial but only one of the gratings required a response. The ERP can be characterized in terms of feature-specific and grating-specific effects. Stimuli with the same spatial frequency as the target grating were distinguished from stimuli with a different spatial frequency by an enhanced negativity starting at 175 ms and peaking at 225 ms post-stimulus. A similar but slightly later effect was found for orientation. The differences between ERPs elicited by relevant and irrelevant gratings was represented by an enhanced negativity at 375 ms as well as a larger P300 component.

These results provide converging evidence for the proposal that stimuli are initially selected on the basis of their separable features and only later are these features combined and processed as a single object (Treisman and Gelade, 1980). Furthermore, the ERPs have demonstrated a sensitivity to the direction of attention *independent* of eye movements. On the basis of the ERP results, it appears that the processing of stimulus features such as color, size and brightness are hierarchically dependent upon prior selection for location.

The studies described thus far have investigated the sensitivity of ERPs to

selective attention in fairly restrictive visual environments. In these paradigms, subjects have been instructed to maintain fixation on the center of a CRT while directing their attention to a peripheral location. Thus, the generalizability of these results to real-world visual processing, in which the position of objects as well as that of the observer is dynamic, is difficult to assess. However, the findings of several recent studies suggest that ERPs can be employed as an index of operators' focus of attention within a complex, dynamically changing display, and more specifically, to a particular attribute of a single object (Kramer, Wickens, and Donchin, 1985).

Cognitive workload

The study of cognitive workload has traditionally involved an evaluation of changes in task performance with increases in the difficulty of a task or in situations in which two or more tasks are concurrently performed. Thus, increased workload has been inferred from a deterioration in performance. However, this is not to say that the level of workload is synonymous with an operator's proficiency in performing a task. For example, imagine a situation in which two pilots perform a series of difficult flight manoeuvres. Under normal conditions, both pilots execute the manoeuvres with a high level of proficiency and their flight performance is indistinguishable. Is this to say that their workload is also equivalent? Now imagine the same two pilots flying the same manoeuvres while concurrently attempting to diagnose the cause of an intermittent engine problem. Although their performance is equivalent under normal conditions, one pilot may cope adequately with the additional demands, while the second pilot's flight performance may deteriorate as he troubleshoots the failure. This hypothetical scenario illustrates the distinction between performance and workload. Cognitive workload can be described as the cost of performing one task in terms of a reduction in the capacity available to perform *additional* tasks. This cost may be inferred from performance tradeoffs when two difficult tasks are time-shared. It has been found useful to conceptualize human capacity as represented by a finite pool of resources available for time-sharing among concurrently performed tasks. In early versions of this class of models, the hypothetical resources were viewed as *undifferentiated*, implying that all tasks draw resources from the same pool (Kahneman, 1973).

The resource model of human capacity underlies the secondary task technique of workload assessment. In this procedure, a subject is assigned two tasks, a primary task that is to be performed as well as possible and a secondary task which is to be performed to the extent that it does not interfere with the performance of the primary task. It is assumed that the demands imposed upon the subject by the primary task will be reflected in the performance on the secondary task. Increasing primary task difficulty presumably 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. Although the secondary task technique has been extensively used in the evaluation of cognitive workload, it presents a number of practical and theoretical problems (Brown, 1978; Ogden, Levine, and Eisner, 1979). For example, in many cases the performance of the secondary task intrudes upon the performance of the primary task rendering the interpretation of the resource tradeoffs extremely difficult. Thus, it would be useful to have a secondary task which is sensitive to changes in primary task difficulty but does not require an overt response.

The P300 is an endogenous ERP component which is recorded within 300–800 ms following the presentation of a task-relevant stimulus. One task in which P300 is readily elicited is the oddball paradigm. In a study using this paradigm, subjects were instructed to count covertly the total number of high-pitched tones in a series of high and low pitched tones (Duncan-Johnson and Donchin, 1977). In different conditions, the relative probability of the two tones was manipulated. The amplitude of the P300s elicited by the tones increased monotonically with decreases in stimulus probability. In another condition in which subjects performed a word puzzle and ignored the tones, P300s were not elicited. Thus, the amplitude of the P300 is determined by a combination of the task relevance and subjective probability of the eliciting event.

These results as well as others have led to the proposal that P300 is a manifestation of the context updating that occurs whenever an event calls for a revision of our mental model of the environment (Donchin, Ritter, and MCallum, 1978). This updating is invoked only if 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? Will the amplitude of the P300 reflect the graded changes in the difficulty of a primary task? If so, then the P300 *may serve as an index of the resource demands*, and hence, the cognitive workload imposed upon a human operator by a task. The P300 might also offer a solution to the problems of secondary task intrusion into primary task performance, *since the ERP eliciting tones occur intermittently*, are easily discriminable, and do not require an overt response.

Wickens, Isreal, and Donchin (1977) conducted an experiment in which subjects *were instructed to perform a compensatory tracking task concurrently* with an ERP eliciting oddball task. Difficulty was manipulated by varying the number of dimensions to be tracked. P300 was significantly attenuated with the introduction of the tracking task. However, *increasing tracking difficulty failed to* produce any further reduction in P300 amplitude. Figure 3 illustrates the results from a similar study in which the bandwidth of the forcing function was varied (Isreal *et al.*, 1980a). Again, the amplitude of the P300 was *reduced with the*

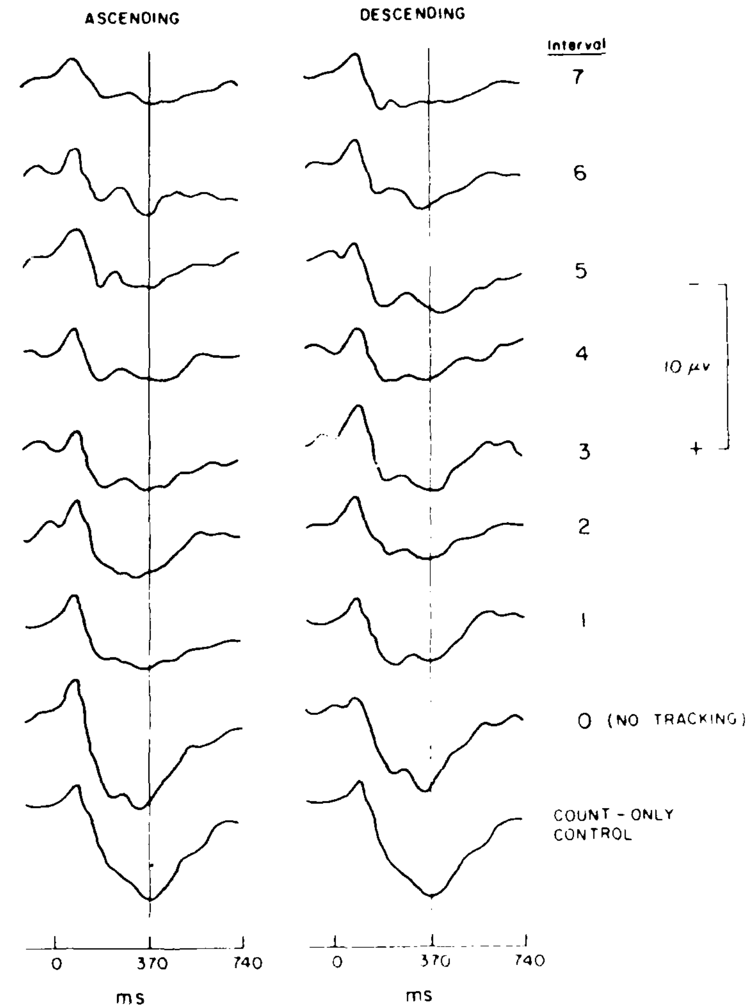


Fig. 3. Average parietal ERPs, elicited by equiprobably counted tones, for each bandwidth level and count-only control conditions for both ascending and descending blocks of trials. (From Isreal *et al.*, 1980, reproduced by permission of *Psychophysiology*.)

introduction of the tracking task but no further attenuation was demonstrated when the *forcing function* bandwidth was *increased*. Thus, *in both studies* tracking difficulty failed to significantly influence the amplitude of the P300s elicited by the oddball task. The authors interpreted these results in terms of a *multiple resource model of processing resources*. They argued that the resources demanded when the bandwidth or dimensionality of the tracking task was increased were separate from those underlying the P300.

Multiple resource models suggest that processing resources are not undifferentiated, but instead that they can be defined on the basis of separate dimensions. In one such model (Wickens, 1980) processing resources have been represented by three dichotomous dimensions: stages of information processing (perceptual/central and response), modalities of processing (visual and auditory), and codes of processing (verbal and spatial). Attempts to perform concurrently tasks which require processing resources from the same modalities, codes or stages of processing generally result in larger decrements in performance than does the concurrent performance of tasks which require resources from different structures. The proposal that P300 is sensitive to a specific aspect of information processing is consistent with other research which has shown that P300 latency is determined by stimulus evaluation time and is largely independent of the time required for response selection and execution (McCarthy and Donchin, 1981). If the manipulation of the bandwidth and dimensionality of a tracking task demand resources associated largely with response selection and execution processes then P300 amplitude should not reflect fluctuations in performance. However, if the perceptual aspects of a task were varied, the amplitude of the P300 elicited by a secondary task would be expected to covary with primary task difficulty.

Isreal *et al.* (1980b) tested this hypothesis by combining the oddball task as a secondary task with a visual monitoring task which served as the primary task. Subjects monitored a simulated air traffic control display and counted the total number of course changes of a relevant class of aircraft. The difficulty of the monitoring task was manipulated by varying the number of aircraft traversing the display. As can be seen from Figure 4, the P300s elicited by the oddball task decreased in amplitude with increases in the difficulty of the monitoring task. Since the primary task did not require an overt response, the data of Isreal *et al.* (1980b) provide strong support for the proposal that P300 amplitude is sensitive to the perceptual demands of a task.

The studies described above have shown that P300 is a sensitive measure of a subset of the processing demands of a task. P300s elicited by secondary task probes decrease in amplitude with increases in the perceptual/central processing difficulty of a primary task. Thus, the P300s appear to mimic the allocation of resources presumed to underlie dual-task performance. However, the secondary task P300s provide only a partial picture of the resource tradeoffs. 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 relation between the resources allocated to the primary and secondary tasks. If the P300 truly reflects the resource tradeoffs that occur during dual task performance, then it should be possible to demonstrate that P300s elicited

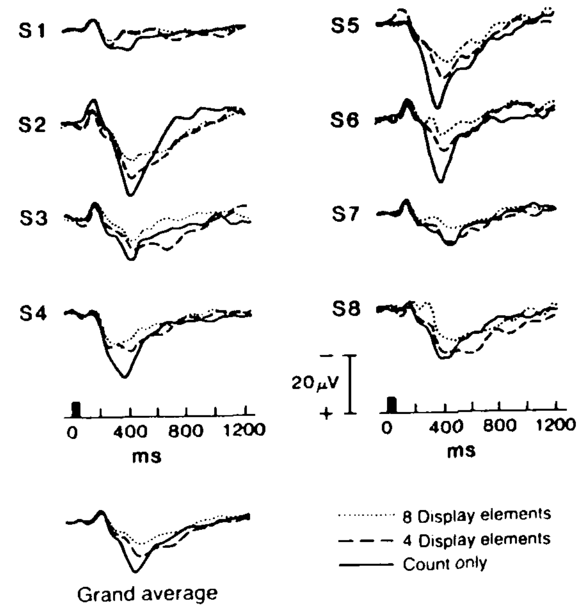


Fig. 4 Single subject and grand average ERPs elicited by counted secondary task probes presented alone and concurrently with a visual monitoring task which required the detection of course changes of simulated aircraft. Difficulty was varied by manipulating the number of symbols traversing the CRT. (From Isreal *et al.*, 1980, reproduced by permission of *Psychophysiology*.)

by primary task events increase in amplitude with increase in difficulty of the primary task.

A series of studies have obtained results which confirm this hypothesis. In one such study, subjects were required to perform a pursuit step tracking task concurrently with an auditory oddball task (Wickens *et al.*, 1983). ERPs were elicited by changes in the spatial position of the target in the tracking task and the presentation of the tones in the oddball task. Difficulty was varied by manipulating two variables in the tracking task: the predictability of the positional changes of the target and the control dynamics. The ordering of difficulty was validated by measures of tracking performance and subjective ratings of tracking difficulty. The ERP results are illustrated in Figure 5. Consistent with previous studies, P300s elicited by the discrete secondary task events decreased in amplitude with increases in the difficulty of the primary task. On the other hand, increasing the difficulty of the tracking task by decreasing the stability of the control dynamics and the predictability of the target resulted in a systematic increase in primary task P300 amplitude. The reciprocal relation between P300s elicited by primary and secondary task stimuli as a function of primary task difficulty is consistent with

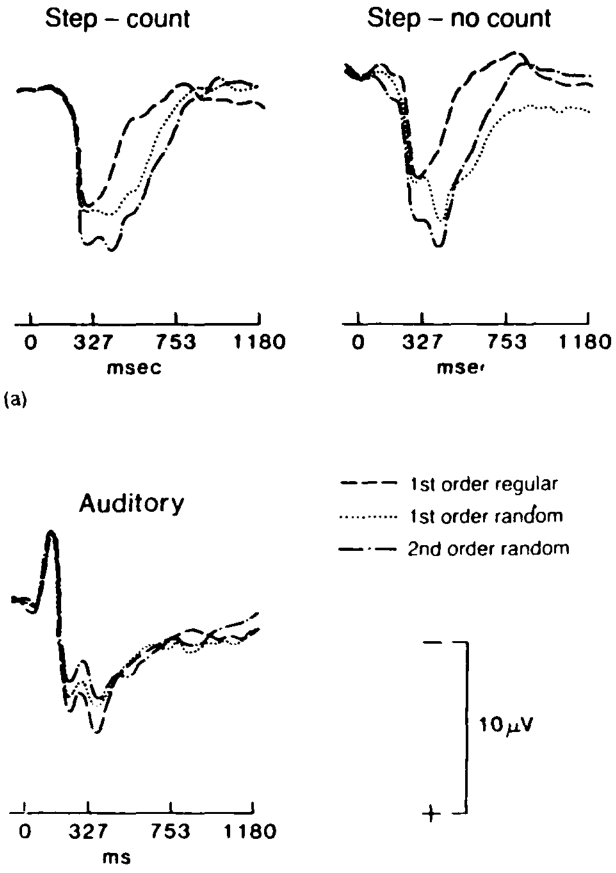


Fig. 5 Grand average ERPs elicited by discrete events in both the primary and secondary tasks. Note that in the primary task (a) the large positive deflection in the waveforms increases in amplitude with increases in the difficulty of the primary task, while the ERPs elicited by the secondary task (b) probes decrease in amplitude with increases in tracking difficulty. (From Wickens, *et al.*, 1983, *Science*, **221**, 1080–1082, with permission of the American Association of the Advancement of Science, copyright 1983.)

the resource tradeoffs presumed to underlie dual-task performance decrements.

An additional benefit of using P300 to assess resource demands is illustrated by the step/no count panel in Figure 5. In this condition, subjects performed the tracking task without counting any extraneous events. ERPs were elicited by changes in the spatial position of the target in the tracking task. The ordering of P300s as a function of tracking difficulty was identical to the condition in which

subjects counted the primary task events. These data suggest that inferences from the P300 about resource allocation and therefore workload can be made in the total absence of a secondary task requirement. This is a considerable advantage if workload is to be assessed unobtrusively in real-time environments.

Models of skill development have proposed that the modulation of some hypothetical resources underlie the overt, measurable improvement in performance (Hirst *et al.*, 1980; Schneider and Shiffrin, 1977). Increased proficiency in performing a task is described in terms of a reduced demand for these processing resources. Given that P300 reflects the resource tradeoffs that occur when two difficult tasks are time shared, one might hypothesize that P300 will also reflect the changing resource demands with practice. Within a dual-task scenario, the increasing automaticity of a primary task should release resources which could then be allocated to the secondary task. Thus, if P300s were elicited by secondary task events, we would expect them to increase in amplitude with increased practice on the primary task. Kramer, Wickens, and Donchin (1983) conducted an experiment in which subjects performed a target acquisition task concurrently with an auditory oddball task. When subjects were relatively inexperienced with the task, the P300s elicited by the secondary task probes were small. However, P300s increased in amplitude with increases in the amount of practice on the target acquisition task. Further evidence for the systematic relation between P300 and changes in the skill level of subjects has been found in both single and dual-task paradigms (Kramer *et al.*, 1986; Natani and Gomer, 1981; Rösler, 1981).

Mental chronometry

The timing of mental processes has been of concern to psychologists, physiologists and astronomers for over 100 years. Initial interest in the topic grew out of the desire to accurately measure the transit time of stars. Two observers, watching the same star move from one location to another would often find that their estimates differed by several hundred ms. Upon further investigation, the variability in temporal estimates was found to be due to differences in reaction times among observers.

Interest in accounting for these differences among individuals led to the development of several methodologies designed to decompose reaction time into its constituent processes. Donders (1909) had subjects perform a series of tasks of varying complexity. In one task subjects made a simple response to a single stimulus (task A). In a second task subjects were instructed to make a different response to each of several stimuli (task B) while in a third task subjects were required to discriminate among several stimuli and respond to only one of them (task C). Donders proposed that each of these tasks required specific mental processes and that by subtracting the reaction time for one task from that of another, the time necessary to complete a specific process could be derived. For example, it was assumed that task A required a *simple reaction time process* while task C also required a stimulus categorization process. Thus, by subtracting the reaction time of task A from the reaction time of task C, the time necessary for stimulus categorization could be determined.

Unfortunately, some of the assumptions upon which the methodology was based proved untenable. For instance, the procedure presupposed that the deletion of one process from a task would not influence the way in which other processes were executed. This assumption of '*pure insertion*' was found to be incorrect. However, this problem was resolved by the additive factors methodology (Sternberg, 1966). In this procedure, Sternberg suggested that different mental processes could be distinguished on the basis of patterns of additivity and interactions among experimental variables. When two experimental variables were found to interact (that is, stimulus intensity and the presence/absence of a mask) they were said to influence a common stage of processing while two variables that were additive in their effects on reaction time (that is, stimulus intensity and stimulus–response compatibility) were said to influence separate stages.

Both Donders' and Sternberg's methodologies imply that human information processing may be conceptualized in terms of a number of stages, from the encoding of the stimulus array to the execution of a response. Furthermore, Sternberg's serial stage model assumes that reaction time is the sum of a number of non-overlapping processing stages and that the function performed by each stage is independent of the duration of preceding stages. Although recent empirical results and theories have suggested that information processes may be executed in a continuous fashion rather than in discrete stages, the importance of Donders' and Sternberg's methodological contribution to the study of mental chronometry should not be underestimated (Eriksen and Schultz, 1979; McClelland, 1979).

Regardless of whether one adopts a discrete stage or continuous model of information processing, one problem inherent in distinguishing the contribution of different processes to a particular task or subject's strategy is the inferences that must be drawn from traditional measures. That is, measures of system output such as reaction time and accuracy are used to infer changes in perceptual, cognitive and motor processes. Clearly, it would be desirable to possess indices of a subset of the processes which are reflected in these output measures.

A number of ERP components have been described in terms of their sensitivity to specific mental processes. For instance, two negative components have been associated with the priming of response channels. Both the readiness potential (RP) and contigent negative variation (CNV) occur prior to a voluntary response or primed stimulus and tend to reflect the amount of response preparation. The RP has been shown to reflect primarily motor processes and is largest in amplitude contralateral to the responding limb (Kutas and Donchin, 1980). The CNV appears to be composed of two separate processes, an orienting compo-

nent to a warning stimulus and a response preparation component (Rohrbaugh and Gaillard, 1983). Evidence suggests that the P300 component reflects stimulus evaluation processes but is relatively insensitive to response selection and execution (McCarthy and Donchin, 1981). The N200 component appears to be sensitive to physical mismatch between subsequent stimuli while the N400 reflects semantic incongruity (Kutas and Hillyard, 1980; Näätänen, Simpson, and Loveless, 1982).

These ERP components used in conjunction with traditional measures such as reaction time and accuracy have proven to be quite useful in explicating the locus of interactions among task parameters. One paradigm in which this multivariate approach has resolved a long standing theoretical debate is the Stroop task (Stroop, 1935). In the standard Stroop paradigm subjects are instructed to name the ink color in which a word is printed. 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. The theoretical *debate has concerned* the locus of this interaction.

Duncan-Johnson and Koppel (1981) reasoned that since the latency of the P300 component is sensitive to factors that 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 latter 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.

Eriksen and Schultz (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. 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 *et al.*, 1985).

In a recent study, Coles *et al.*, (1985) augmented measures of reaction time and accuracy with measures of the electromyogram (EMG) as well as with the latency of two components of the ERP, the P300 and the 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 and Schultz, 1979).

Although the experiments described within the present section have been exclusively conducted in laboratory settings, the results have important practical implications. The section on cognitive workload (pp. 209–215) described how ERPs recorded in conjunction with traditional measures might be used to assess the resource demands of operational tasks. The research and theories described in the current section suggests that mental processes also play a role in human information processing. Thus, the demands imposed on the human operator can be conceptualized in terms of both the resources required to perform a task and the mental processes or transformations that must be performed on the data. The joint ERP-performance based approach, allows for a fine-grained analysis of both these components of information processing.

RECOMMENDATIONS

This chapter has provided a brief glimpse of the pitfalls and potential advantages of employing ERPs in the assessment of human performance and cognition. Truly, there are many situations in which traditional measures can provide adequate answers to our questions, thereby rendering the costly and time-consuming ERP methodology unnecessary. However, there are other cases in which the issues have proven difficult to resolve with our current battery of measurement techniques. It is in these situations that ERPs can be most profitably employed.

Recommendations for future research

Several examples have been discussed in which ERPs, employed in conjunction with other assessment techniques, have enhanced our understanding of human information processing. At best, the experiments performed thus far have provided an initial glimpse of the processes under study. Further advances in the use of ERPs as tools in the investigation of cognitive processes depend upon (a) the development and validation of multivariate statistical techniques that will allow us to more reliably extract and measure ERP components; and (b) the continued explication of the functional significance of ERP components (Donchin, 1981). In addition to the research areas described in the present paper,

other substantive areas in which ERPs offer the potential for significant contributions include: the assessment of cognitive deficits in aging, the analysis of language processes, and the evaluation of response parameters of keyboards and other data input and retrieval devices.

Recommendations for practice

Clearly, the use of ERPs to address problems in human factors engineering is in its infancy and at present the procedure has been almost exclusively employed in a laboratory environment. If ERP measures are to truly be of use to designers of complex, semiautomated systems, then the validity of this approach must be assessed in extralaboratory situations such as simulators and ultimately operational settings.

We can distinguish two different scenarios in which ERPs could potentially be of use in the future. First, ERPs might be used in an *offline* context to explicate the processing demands of a task, the adequacy of a visual or auditory display, or the individual differences in processing strategies exhibited by different groups of system controllers or supervisors. To some degree, this line of research has already begun, at least in the fairly sterile laboratory environment. Second, ERPs might be used in an *online* context to detect momentary fluctuations in attentional state or processing demands and to adjust the level of system automation on the basis of these measures. In our opinion, the offline scenario appears to be the most realistic contribution we could expect from ERP technology in the short term. Potential online uses of ERPs will have to await a more detailed understanding of the relationship between single trial ERP components and cognitive processes as well as the development of more accurate and reliable signal extraction techniques.

ACKNOWLEDGEMENTS

Preparation of this chapter was supported by NASA Ames Research Center under contract No. 1-5-25-477 with Dr. Michael Vidulich as technical monitor and by the Air Force Office of Scientific Research under contract No. F49620-83-0144 with Dr. Alfred Fregly as technical monitor. I am indebted to Michael Coles, Demetrios Karis, and Erik Sirevaag for helpful comments on an earlier draft of this manuscript.

REFERENCES

Bainbridge, L. (1974). Analysis of verbal protocols from a process control task. In E. Edwards and F. P. Lees (eds), *The Human Operator in Process Control*. London: Taylor and Francis.

Broadbent, D. E. (1977). Levels, hierarchies, and the locus of control. *Quarterly Journal of Experimental Psychology*, **29**, 181–201.

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

Coles, M.G.H., Gratton, G., Bashore, T.R., Eriksen, C.W., and 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., Kramer, A., and Miller, G. (1986). Principles of signal acquisition and analysis. In M.G.H. Coles, E. Donchin and S. Porges (eds), *Psychophysiology: Systems, Processes and Applications*. New York: Guilford Press.

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

Donchin, E. and 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, DC: US Government Printing Office.

Donchin, E., Ritter, W., and 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.

Donders, F.C. (1909). On the speed of mental processes. (English translation of the original 1868 article). *Acta Psychologica*, **30**, 412–431.

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

Duncan-Johnson, C.C. and Koppel, B.S. (1981). The stroop effect: brain potentials localize the source of interference. *Science*, **214**, 938–940.

Eason, R., Harter, M., and White, C. (1969). Effects of attention and arousal on visually evoked cortical potentials and reaction time in man. *Physiology and Behavior*, **4**, 283–289.

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.

Eriksen, C.W., Coles, M.G.H., Morris, L.R., and O'Hara, W.R. (1985). An electromyographic examination of response competition. *Bulletin of the Psychonomic Society*, **23**, 165–168.

Goff, W.R., Allison, T., and Vaughan, H.G. (1978). The functional neuroanatomy of event-related potentials. In E. Calloway, P. Tueting, and S.H. Koslow (eds), *Event-Related Brain Potentials in Man*. New York: Academic Press.

Gomer, F.E. and Bish, K.G. (1978). Evoked potential correlates of display image quality. *Human Factors*, **20**, 589–596.

Grether, W.F. and Baker, C.A. (1972). Visual presentation of information. In H.P. Van Cott and R.G. Kinkade (eds), *Human Engineering Guide to Equipment Design*. Washington, DC: US Government Printing Office.

Harter, M.R. and White, C.T. (1968). Effects of contour sharpness and check size on visually evoked cortical potentials. *Vision Research*, **8**, 701–711.

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

Isreal, J.B., Chesney, G., Wickens, C.D., and Donchin, E. (1980a). P300 and tracking difficulty: evidence for multiple resources in dual task performance. *Psychophysiology*, **17**, 259–273.

Isreal, J.B., Wickens, C.D., Chesney, G., and Donchin, E. (1980b). The event related brain potential as an index of display monitoring workload. *Human Factors*, **22**, 211–224.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice Hall.

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

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

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

Kramer, A.F., Schneider, W., Fisk, A., and Donchin, E. (1986). The effects of practice and task structure on components of the event related brain potential. *Psychophysiology*, **23**, 33–47.

Kutas, M. and Donchin, E. (1980). Preparation to respond as manifested by movement related brain potentials. *Brain Research*, **202**, 95–115.

Kutas, M. and Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, **207**, 203–205.

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.

Moray, N. and Rotenberg, I. (1985). Experimental studies of process control. In R.W. Swezey (ed.), *Proceedings of the Human Factors Society 29th Annual Meeting*. Baltimore, MD.

Näätänen, R., Simpson, M., and Loveless, N.E. (1982). Stimulus deviance and evoked potentials. *Biological Psychology*, **14**, 53–98.

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

Neisser, U. and Becklen, R. (1975). Selective looking: attending to visually specified events. *Cognitive Psychology*, **7**, 480–494.

Ogden, G.D., Levine, J.W., and Eisner, E.J. (1979). Measurement of workload by secondary tasks. *Human Factors*, **21**, 529–548.

Pearson, J.E. and Pearson, L.E. (1985). The effects of bandwidth compression and image quality on image interpreter performance. *Human Factors*, **27**, 345–353.

Posner, M. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, **32**, 3–25.

Previc, F.H. and Harter, M.R. (1982). Electrophysiological and behavioural indicants of selective attention to multifeature gratings. *Perception and Psychophysics*, **32**, 465–472.

Proctor, R.W. (1981). A unified theory for matching task performance. *Psychological Review*, **88**, 291–326.

Ritter, W., Simpson, R., and Vaughan, H.G. (1983). Event related potentials and two sequential stages of information processing in physical and semantic discrimination. *Psychophysiology*, **20**, 168–179.

Rohrbaugh, J.W. and Gaillard, A.W.K. (1983). Sensory and motor aspects of the contingent negative variation. In A.W.K. Gaillard and W. Ritter (eds), *Tutorials in Event Related Potential Research: Endogenous Components*. Amsterdam: North-Holland.

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

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

Starr, A. (1978). Sensory evoked potentials in clinical disorders of the nervous system. *Annual Review of Neuroscience*. Palo Alto, California: Annual Reviews Inc.

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

Stockard, J.J., Stockard, J.E., and Sharbrough, F.W. (1979). Brainstem auditory evoked potentials in neurology: methodology, interpretation, clinical application. In M.J. Aminoff (ed.), *Electrophysiological Approaches to Neurological Diagnosis*. New York Churchill Livingstone.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* **18**, 643–662.

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

Van Voorhis, S.T. and Hillyard, S.A. (1977). Visual evoked potentials and selective attention to points in space. *Perception and Psychophysics*, **1**, 54–62.

Wickens, C.D. (1980). The structure of attentional resources. In R. Nickerson and R. Pew (eds), *Attention and Performance VIII*, Hillsdale, NJ: Erlbaum.

Wickens, C.D. (1984). *Engineering Psychology and Human Performance*. Columbus, Ohio: Charles Merrill.

Wickens, C.D. and Kramer, A. (1985). Engineering Psychology. In M.R. Rosenzweig and L.W. Porter (eds), *Annual Review of Psychology*. Palo Alto, CA: Annual Reviews Inc.

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

Wickens, C., Kramer, A., Vanasse, L., and Donchin, E. (1983). The reciprocity of primary and secondary task resources: Evidence from the P300 component of the ERP. *Science*, 1080–1082.

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

Wood, C.C., McCarthy, G., Squires, N.K., Vaughan, H.G., Woods, D.L., and McCallum, C.W. (1984). Anatomical and physiological substrates of event related potentials. In R. Karrer, J. Cohen, and P. Tueting (eds), *Brain and Information: Event Related Potentials*., New York: New York Academy of Sciences.

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

HUMAN FACTORS, 1987, 29(2), 145–160

A Psychophysiological Assessment of Operator Workload During Simulated Flight Missions

ARTHUR F. KRAMER,[1] ERIK J. SIREVAAG, *and* ROLF BRAUNE, *Institute of Aviation, Aviation Research Laboratory, University of Illinois, Champaign, Illinois*

Previous research has indicated that components of the event-related potential (ERP) may be used to quantify the resource requirements of complex cognitive tasks. The present study was designed to explore the degree to which these results could be generalized to complex, real-world tasks. The study also examined the relations among performance-based, subjective, and psychophysiological measures of operator workload. Seven male volunteers, enrolled in an instrument flight rule (IFR) aviation course at the University of Illinois, participated in the study. The student pilots flew a series of IFR flight missions in a single-engine, fixed-based simulator. In dual-task conditions subjects were also required to discriminate between two tones differing in frequency and to make an occasional overt response. ERPs time-locked to the tones, subjective effort ratings, and overt performance measures were collected during two separate 45-min flights differing in difficulty. The difficult flight was associated with high subjective effort ratings, as well as increased deviations from the command altitude, heading, and glideslope. The P300 component of the ERP discriminated among levels of task difficulty, decreasing in amplitude with increased task demands. Within-flight demands were examined by dividing each flight into four segments: takeoff, straight and level flight, holding patterns, and landings. The amplitude of the P300 was negatively correlated with deviations from command headings across the flight segments. In sum, the findings provide preliminary evidence for the assertion that ERP components can be employed as metrics of resource allocation in complex, real-world environments.

INTRODUCTION

The importance of the explication of mental workload in operational environments has been underscored in recent years by a number of conferences convened to examine the topic as well as by several methodological and theoretical reviews of the literature (Frazier and Crombie, 1982; Gopher and Donchin, 1986; Moray, 1979; O'Donnell and Eggemeier, 1986, Williges and Wierwille, 1979). Although mental workload is acknowledged to be a problem in many manual control and supervisory tasks, there has been a failure to reach a consensus on its definition or measurement. This, in part, is due to the multidimensional nature of the concept (Eggemeier, 1980; Johannsen, Moray, Pew, Rasmussen, Sanders, and Wickens, 1979).

One set of models that appear to offer a useful framework in which to conceptualize mental workload are the multiple-resource

[1] Requests for reprints should be sent to Arthur F. Kramer, Department of Psychology, University of Illinois at Urbana-Champaign, 603 E. Daniel St., Champaign, IL 61820.

models of attentional allocation (Freidman and Polson, 1981; Navon and Gopher, 1979; Wickens, 1980, 1984). In these models, human capacity is represented by a number of finite pools of resources available for time-sharing among concurrently performed tasks. The models predict that tasks that require the same types of processing resources will be more poorly time-shared than will tasks that require different resources. Wickens (1980) has proposed that resources may be defined by three dichotomous dimensions: stages of processing (perceptual/cognitive and response), codes of processing (verbal and spatial), and modalities of processing (auditory and visual). Within such a framework, *mental workload* can be described as the cost of performing one task in terms of a reduction in the capacity to perform additional tasks, given that the two tasks overlap in their resource demands.

The research presented here derives from an extensive series of investigations that have demonstrated the utility of event-related brain potentials (ERPs) in the assessment of residual capacity during the acquisition and performance of a variety of perceptual-motor tasks (Donchin, Kramer, and Wickens, 1986; Kramer, 1987). The focus of the present study was to employ ERPs in conjunction with measures of overt performance and subjective indices of mental workload in order to monitor changes in resource demands that occur during a complex real-world task. The task involved "flying" an instrument flight rule (IFR) plan in a single-engine aircraft simulator.

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). This temporal relationship between the ERP and the eliciting stimulus or response is what differentiates ERPs from the ongoing electroencephalographic activity. Thus, the electroencephalograph (EEG) provides a measure of the tonic state of the organism, whereas ERPs reflect phasic changes related to the processing of specific events.

It is important to recognize the componential nature of the ERP. ERPs have generally been viewed as a sequence of separate but sometimes temporally overlapping components that are influenced by a combination of the physical parameters of the stimuli and such psychological constructs as expectancy, task relevance, attention, and memory (Kramer, 1985). Components are typically labeled with an *N* or a *P*, denoting negative or positive polarity, and with a number indicating their minimal latency measured from the onset of an eliciting event (e.g., N100 is a negative-going component that occurs at least 100 ms after a stimulus).

ERP components may be categorized along a continuum from exogenous to endogenous. The exogenous components represent an obligatory response of the brain to the presentation of a stimulus. These components are usually associated with specific sensory systems, occur within 200 ms of a stimulus, and are primarily sensitive to the physical attributes of stimuli. For example, exogenous visual potentials are influenced by the intensity, frequency, hue, patterning, and location of the stimulus in the visual field. The exogenous components have been successfully used in clinical settings to monitor the functional integrity of the nervous system during surgical procedures, to assess changes in the nervous system as a result of maturation and aging, and to help diagnose various types of neuropathology including tumors, lesions, and demyelinating diseases, such as multiple sclerosis (Starr, 1978; Stockard, Stockard, and Sharbrough, 1979).

The endogenous components, on the other hand, occur somewhat later than the exogenous components and are not very sensitive to changes in the physical parameters of stimuli, especially when these changes are not relevant to the task. Instead, these components are primarily influenced by the processing demands of the task imposed on the subject. In fact, endogenous components can even be elicited by the absence of a stimulus if this "event" is relevant to the subject's task. The strategies, expectancies, intentions, and decisions of the subject, along with task parameters and instructions, account for the majority of the variance in the endogenous components. One of the dependent variables in the present study, the P300, is a typical example of an endogenous component. (For a comprehensive review of the P300, see Pritchard, 1981.)

One might ask why ERPs should be used to monitor changes in resource demands, given that several technically simpler approaches to the assessment of skill acquisition and mental workload have already been implemented. Although numerous performance-based measures of mental workload exist, they suffer from several drawbacks. First, some of the measurement techniques require subjects to perform a secondary task, which frequently interferes with the performance of the task of interest (Knowles, 1963; Rolfe, 1971; Wickens, 1979). This is clearly unacceptable in an operational environment in which the safety of the operator must be assured. Even in the laboratory setting it is difficult to determine which of the two tasks generated an observed performance decrement, since performance on the two tasks is easily confounded. Second, performance-based measures of mental workload provide an output measure of the operator's information-processing activities (e.g., RT, accuracy). Thus, at best, performance measures provide only an indirect index of cognitive function. Third, performance measures do not always correlate highly with the actual workload of the tasks (Brown, 1978; Dornic, 1980; Ogden, Levine, and Eisner, 1979).

For example, imagine a situation in which two pilots perform a series of difficult flight maneuvers. Under normal conditions, both pilots execute the maneuvers with a high level of proficiency, and their flight performance is indistinguishable. Is this to say that their workload is also equivalent? Now imagine the same two pilots flying identical maneuvers while concurrently attempting to diagnose the cause of an intermittent engine problem. Although their performance is equivalent under normal conditions, one pilot may cope adequately with the additional demands, whereas the second pilot's flight performance may deteriorate as he or she troubleshoots the abnormality. This pattern of results would suggest that the second pilot was operating under higher levels of workload even during the normal flight condition.

Several recent studies have illustrated the usefulness of the ERP, and more specifically the P300 component, as an index of processing resources (Horst, Munson, and Ruchkin, 1984; Isreal, Chesney, Wickens, and Donchin, 1980; Kramer, Wickens, and Donchin, 1983, 1985; Natani and Gomer, 1981; Strayer and Kramer, 1986; Wickens, Kramer, Vanasse, and Donchin, 1983). The general paradigm employed in these studies requires subjects to perform two tasks concurrently. One task is designated as primary and the other task as secondary. Subjects are instructed to maximize their performance on the primary task and devote any additional resources to the performance of the secondary task.

Primary tasks have included system monitoring, decision making, and manual control. Secondary tasks have required subjects to discriminate between tones of different frequencies or lights of different intensities. In general, the response demands of the secondary probe tasks have been minimal, requiring subjects either to covertly count the total number of one type of event or to respond to an occasional target probe.

ERPs are elicited by events either in one or in both of the tasks. Increases in the perceptual/cognitive difficulty of the primary task result in a decrease in the amplitude of the P300s elicited by the secondary task. Conversely, P300s elicited by discrete events embedded within the primary task increase in amplitude with increases in primary task difficulty. Furthermore, changes in response-related demands of a task have no influence on the P300 (Isreal et al., 1980).

The reciprocal relationship between P300s elicited by primary and secondary task stimuli is consistent with the resource tradeoffs presumed to underlie dual-task performance decrements (Kahneman, 1973; Navon and Gopher, 1979; Sanders, 1979; Wickens, 1980). That is, resource models predict that as the difficulty of one task is increased, additional resources are reallocated to that task in order to maintain performance, thereby depleting the supply of resources that could have been used in the processing of other tasks. Thus, the P300 appears to provide a measure of resource trade-offs that can only be inferred from more traditional performance measures. Furthermore, P300s elicited by secondary task events are selectively sensitive to the perceptual/cognitive demands imposed on the operator. This selective sensitivity may be especially useful in decomposing the changing processing requirements of complex tasks (Kramer et al., 1983).

The goal of the present experiment was to augment the conclusions drawn from the studies cited above by demonstrating that the dual-task ERP paradigm could be employed in a complex real-world situation to provide information concerning mental workload and residual capacity. Student pilots performed a series of dual-task flight missions. In each case, the primary task consisted of performing a specified flight scenario under IFR flight conditions. The difficulty of the primary task was varied in two ways. *Between-mission difficulty* was manipulated by varying the direction and speed of wind conditions, the severity of turbulence, and the probability of a subsystem failure during a critical portion of the mission. A second way that difficulty was manipulated might be labeled *within-mission difficulty*. In this case, we capitalized on the different levels of processing demand inherent in the flight task (i.e., straight and level versus approach to landing).

The secondary task consisted of a concurrently performed go/no-go auditory discrimination task, in which subjects pressed a button in response to the presentation of one of two tones. ERPs associated with the secondary task tones, overt performance measures from the flight task and discrimination task, and subjective indices of task difficulty were examined to assess the extent to which the manipulations of primary task difficulty modulated the mental workload associated with the flight task.

METHOD

Subjects

Seven right-handed male volunteers were paid for their participation in the study. All of the subjects were student pilots enrolled in an aviation course at the University of Illinois' Willard Airport. Prerequisites for the course include two semesters of basic visual flight rule (VFR) training, the possession of a private pilot's license, and two introductory courses in IFR flight skills. Thus, the student pilots were proficient in VFR flight skills and had a basic familiarity with both IFR flight skills and precision landing approach techniques. All subjects were between the ages of 20 and 26, and had normal hearing and normal or corrected-to-normal vision.

Simulator and Stimulus Generation Equipment

The ILLIMAC flight system used in the experiment consisted of a fixed-based flight simulator that was designed around the INTEL 8086 digital microprocessor. The simulator was developed at the Institute of Aviation at the University of Illinois. The flight equations were produced by the 16-bit 8086 microcomputer. The digital information was converted to analog voltages to drive the simulator instrumentation. In the present study, the simulator was configured to mimic the flight characteristics of the Beechcraft Sport 180, a single-engine aircraft with fixed gear and a fixed propeller. The ILLIMAC flight panel contained the instrumentation and navigational radios required for instrument flight conditions. Flight performance measures were digitized at the rate of 30 Hz and were transferred via an RS232 link to a DEC PDP 11/73 human-performance/electrophysiological laboratory computer (Heffley, Foote, Mui, and Donchin, 1985).

The auditory stimuli employed for the secondary task were produced by an audio-generator and binaurally presented to the pilots through headphones. ERPs, flight performance data, and secondary task RTs were recorded on magnetic tape for off-line analysis.

Tasks

A graphic illustration of the basic flight task is presented in Figure 1. The flight began on Runway 32 at Willard Airport. The pilot was instructed to climb to 3000 feet at a constant rate of 500 feet per minute and to intercept the 062 radial from the Champaign very-high-frequency omnidirectional radar beacon (VOR) after turning to a heading of 090 deg. At the 12 mile distance measuring equipment point (DME), the pilot was instructed to roll 30 deg to the left and fly one minute outbound on a 032 heading in order to set up a holding pattern with an inbound heading of 242 deg. Three standard holding patterns were flown prior to continuation of the flight. Next, the pilots were instructed to track the 216 radial direct to the Veals (the outer marker for the ILS approach to Runway 32 at Champaign) non-directional beacon (NDB), maintaining an airspeed of 135 knots. At the 6-mile DME point (Veals), the pilots were to turn outbound in order to intercept the localizer inbound to Runway 32. At the 8.2-mile DME point, a turn was to be made to a 091 heading. After flying for two minutes outbound, the pilot executed a procedure turn to intercept the localizer inbound to Runway 32. When the localizer was intercepted, the pilot tracked 316 deg at 2600 feet towards the middle marker, where the flight was terminated. Performance measures included deviations from assigned heading, airspeed, altitude, and glideslope.

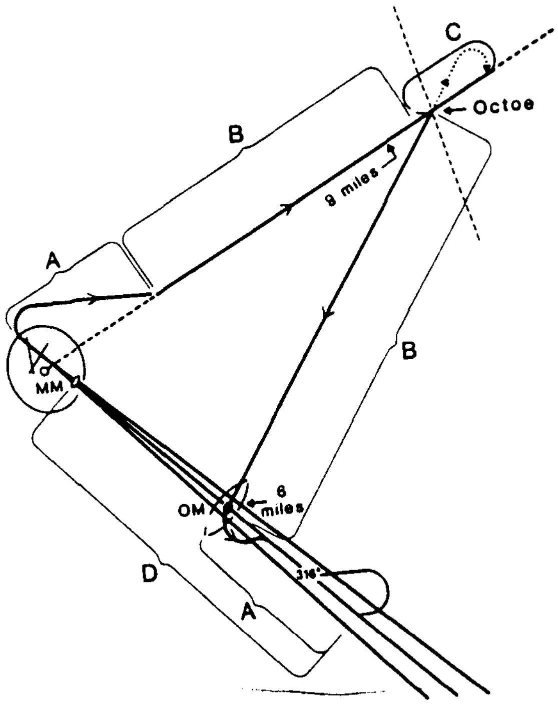


Figure 1. *A schematic illustration of the 45-min flight mission performed by the student pilots.*

This IFR flight plan—roundtrip from

Champaign to Octoe intersection—formed the framework within which the processing demands imposed upon the student pilots were investigated. (See Champaign approach plate AL-709, ILS Runway 32 for additional details.) Thus, the flight segments constituted the primary task in this experiment, and subjects were instructed to maximize their performance of the flight missions.

In addition to the primary task, subjects performed a concurrent secondary task. This task required the student pilots to monitor a Bernoulli sequence of auditory stimuli presented binaurally through headphones. Two different tone frequencies (1000 Hz and 1500 Hz) were used. The 1000-Hz tone was designated as the target tone for four subjects; for the remaining three subjects, the 1500-Hz tone was designated as the target. Targets were presented on 30% of the trials. Subjects were instructed to respond to targets by depressing a switch located on the left side of the control yoke. The response required only a minimal movement of the thumb. Both speed and accuracy were emphasized in the instructions. Non-targets, which were presented 70% of the time, did not require a response. Tones were 50 ms (including a 10-ms rise/fall time, 65 dB) in duration, and were presented every 1.4 to 1.7 s.

ERP Recording System

Electroencephalographic (EEG) activity was recorded from three midline sites (Fz, Cz, and Pz according to the International 10/20 system; see Jasper, 1958) and referred to linked mastoids. Beckman Biopotential Ag-AgCl electrodes filled with Grass electrode paste were attached to all scalp sites and also to a forehead ground. In addition, identical electrodes were placed above and below the subject's right eye to evaluate electrooculographic (EOG) activity in the vertical plane. All electrode impedances were maintained below 10 kohms.

The EEG and EOG channels were amplified by Grass Model 12A5 amplifiers with a 10-s time constant and an upper half amplitude of 35 Hz, 3-dB/octave rolloff. The recording epoch for all channels was 1300 ms, beginning 100 ms prior to the presentation of secondary task tones. The data channels were digitized every 5 ms and were digitally filtered off-line (−3 dB at 6.27 Hz; 0 dB at 14.29 Hz) prior to further analysis. Artifactual contributions to the EEG from EOG activity were evaluated and eliminated off-line by submitting the data to an eye-movement correction procedure (Gratton, Coles, and Donchin, 1983).

Procedure

Each subject flew a total of four 45-minute missions in the fixed-base ILLIMAC flight simulator. In the first session, the students flew the flight course twice. These flights served to familiarize the subjects with the IFR flight plan and the dynamics of the simulator. Both of the missions were flown under the easy flight conditions (no wind, turbulence, or subsystem failures). Since the flights were considered as practice, the performance data will not be presented here.

In the second session, subjects again flew the flight path illustrated in Figure 1. However, in this session one of the two flights included 30-mile/h winds from 270 deg, moderate turbulence, and a partial suction failure in the heading indicator during approach to landing. The presentation order of the easy and difficult flights in the second session was counterbalanced across subjects.

Prior to the flights, subjects were presented with a short block (30 trials) of tones in order to familiarize them with the tone frequencies among which they were subsequently asked to discriminate. Subjects then performed a short, single-task flight segment in which they flew a straight and level course for five minutes under calm atmospheric conditions (no wind or turbulence). This task was included to provide a baseline for subsequent subjective ratings of task difficulty. Subjects were instructed to assign a value of 100 to this flight segment and to compare each of the flight segments in the easy and difficult flights with this baseline when assigning their subjective difficulty ratings (Gopher and Braune, 1984). Subjects then performed the auditory discrimination task alone for 200 trials. This condition was included to provide single-task estimates of performance and ERPs on the secondary task.

Following the discrimination task, subjects performed the flight tasks concurrently with the tone-discrimination task. A certified flight instructor (CFI) was present during each flight to instruct the subjects on the flight scenarios and to evaluate their performance. Upon completion of each of the two flights, subjects were asked to rate the difficulty of the flight as a whole, as well as each of the individual flight segments. Each of the flight missions lasted approximately 45 min. Subjects received a 15-min rest break between flights.

Finally, subjects again performed 200 auditory discriminations under single task conditions. Because the ERPs and performance data from the two single task discrimination conditions were not significantly different, they were averaged together prior to further analysis.

RESULTS AND DISCUSSION

Flight Performance and Probe Discrimination Data

The flight performance data were collected to assess the validity of the difficulty manipulations, both within and across missions. The flight scenario has been described in terms of the mission requirements (see Figure 1). For purposes of the statistical analysis, the mission was partitioned into four flight segments, on the basis of common operational demands (as judged by a panel of CFIs). Segment A comprised both the takeoff and the preparation for final approach to landing. The two straight and level flight segments were combined into Flight Segment B, and the three holding patterns into Segment C. The final approach to landing and flight along the glideslope were combined into Segment D.

Two measures of flight performance, heading and altitude deviation, were recorded in all flight segments in both easy and difficult missions. These indices were submitted to a two-way repeated measures analysis of variance. One additional measure, deviation from the glideslope, was recorded in both the easy and difficult scenarios in the final flight segment. Table 1 presents the mean values of the flight performance measures for both easy and difficult flights.

All three of the flight performance measures indicated that our between-mission experimental manipulations successfully influenced the difficulty of the flight task. When the student pilots were required to fly the 45-min mission with high winds, moderate turbulence, and a subsystem failure during approach to landing, their deviations from command altitude increased, $F(1,6) = 6.6$, $p < 0.05$; their ability to track the glideslope accurately decreased, $F(1,6) = 8.0$, $p < 0.05$; and their deviations from assigned headings increased, $F(1,6) = 9.1$, $p < 0.05$; relative to the mission flown under easy flight conditions. Performance was also influenced by mission segment, irrespective of between-mission difficulty (see Table 2). Subjects were more accurate at maintaining their assigned headings in Segments B and C than they were in Segments A and D, indicating that takeoff and landing were more difficult for the students than was straight and level flight or holding patterns, $F(3,18) = 4.3$, $p < 0.05$. However, students' performance on the altitude measure did not differ as a function of flight segment. Furthermore, no significant interactions were obtained for any of the flight measures. Thus, flight performance measures most strongly discriminated between easy and difficult flights, whereas intersegment differences were found for only a subset of the performance measures.

Performance measures for the ERP-eliciting probe task are presented in Table 1 for the between-mission comparison and in Table 2 for the comparison across flight segments. Subjects were uniformly accurate across missions and flight segments, with response accuracy ranging from 89% to 93%. Neither RTs nor accuracies differed as a function of flight segments or missions ($p > 0.05$). Thus, any differences among the ERPs elicited in different flight conditions cannot be attributed to the subjects' failure to perform the probe task in the more difficult flight missions or segments.

Subjective Workload Ratings

Table 1 presents the subjective workload ratings for the between-mission comparisons, and Table 2 displays the ratings for each of the flight segments. The subjective ratings were collected in a manner described by Gopher and Braune (1984). Subjects initially flew a 5-min straight and level flight path and were instructed to assign this segment a workload rating of 100. Each of the flight segments in the easy and difficult flights was then rated relative to the straight and level segment. Subjects made their ratings after each mission and were permitted to assign any numerical value to their estimates of subjective workload. The ratings were normalized prior to statistical analysis.

Subjects rated the flight mission with high winds, moderate turbulence, and a partial suction failure in the heading indicator during instrument landing system (ILS) approach as having a significantly higher workload than the flight without wind, turbulence, or subsystem failures, $F(1,6) = 18.7$, $p < 0.01$. Subjective ratings also discriminated among flight segments, $F(1,6) = 6.0$, $p < 0.01$. Moreover, the subjects estimated Segments A, C, and D to be equally difficult, whereas Segment B—the straight and level portion of the flight—was estimated to be easier than the other three segments, $F(1,6) = 9.8$, $p < 0.01$. No significant interactions were obtained for the workload measures.

A comparison of the flight performance measures and the subjective workload ratings suggests that, for the most part, the pilots' subjective estimates corresponded well with their performance on the flight task. Both the performance measures and the subjective ratings discriminated between

TABLE 1

Mean (and Standard Deviation) Simulator Performance, Probe Discrimination Reaction Time and Accuracy, and Subjective Workload Ratings for Easy and Difficult Flights

Measures	Flight Missions	
	Easy Flight	Difficult Flight
Heading deviation (deg)	1.86 (0.67)	3.08 (0.90)
Altitude deviation (feet)	40.3 (17.8)	70.9 (37.8)
Subjective workload ratings	115.2 (12.7)	137.8 (21.6)
Probe reaction time (ms)	580.0 (139.5)	604.0 (119.3)
Probe accuracy (percentage correct)	92.4 (6.0)	89.0 (7.2)
Glideslope deviation (deg)	0.35 (0.21)	0.74 (0.38)

TABLE 2

Mean (and Standard Deviation) Simulator Performance, Probe Discrimination Reaction Time and Accuracy, and Subjective Workload Ratings for the Four Flight Segments

Measures	Flight Segments			
	Segment A	Segment B	Segment C	Segment D
Heading deviation (deg)	2.89 (1.03)	1.87 (1.06)	1.92 (0.81)	3.19 (1.44)
Altitude deviation (feet)	53.6 (32.2)	60.1 (12.5)	41.9 (11.2)	67.4 (58.6)
Subjective workload ratings	128.6 (18.5)	117.2 (10.1)	128.4 (13.1)	133.0 (21.7)
Probe reaction time (ms)	604.0 (101)	564.0 (135)	584.0 (127)	617.0 (146)
Probe accuracy (percentage correct)	89.4 (7.3)	93.1 (6.0)	92.2 (6.4)	89.6 (6.4)

easy and difficult missions. Flight segments were also differentiated by subjective and objective measures. However, in this case the heading deviation measure indicated that the students' performance in both straight and level flight and holding patterns was superior to their performance during takeoff and landing, whereas subjective workload ratings suggested that holding patterns were perceived to be as difficult as takeoffs and preparation for landing.

Event-Related Potentials

Figure 2 presents the grand average ERPs elicited by the target tones in the single- and dual-task conditions. In Figure 3, the ERPs recorded at Pz are overplotted for the three conditions. Three different components can be discerned by visual inspection of the waveforms. The earliest is a frontally maximal, negative-going deflection that occurs between 100 and 200 ms post-stimulus. ERP components are traditionally defined in terms of their latency relative to a stimulus or response, scalp distribution, and sensitivity to experimental manipulations (Donchin, Ritter, and McCallum, 1978; Sutton and Ruchkin, 1984). By virtue of its latency and scalp distribution, this component can be defined as the N100. A second frontally maximal negative component is also visible in the waveform. This component reaches its peak negativity in the 200- to 300-ms latency range, and will henceforth be referred to as the N200. A large positive-going deflection peaking between 300 and 500 ms post-stimulus can also be identified by visual inspection of the waveforms. This component increases in amplitude from the frontal to the parietal recording site and appears to discriminate among levels of task difficulty. The component will be labeled the P300.

The single trial ERPs were corrected for eye-movement artifacts and then averaged within experimental conditions. The three ERP components described above were quantified by measures of peak latency and amplitude. For the P300 component, the latency measure was obtained from the single trials. The amplitude measure of the P300 and the latency and amplitude measures for the other ERP components were derived from the average waveforms. Area measures for the N100, N200, and P300 components were obtained from the 100 to 200, 200 to 300, and 300 to 500 ms latency ranges, respectively. Peak latencies were defined as the largest positive or negative deflection within

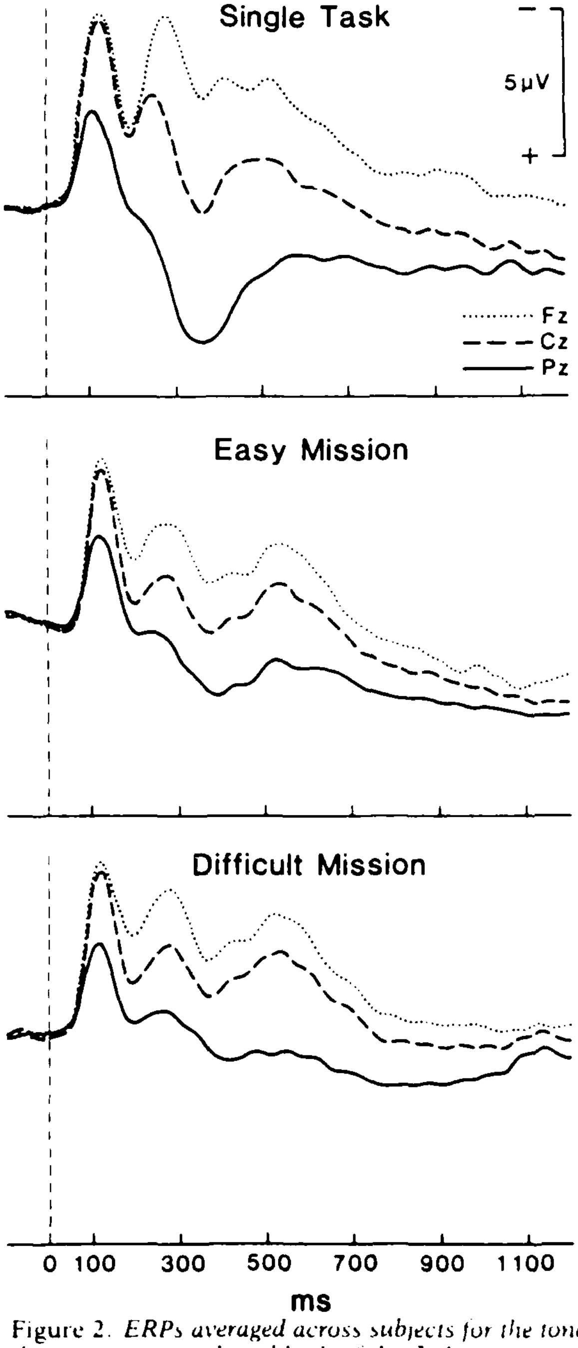


Figure 2. *ERPs averaged across subjects for the tone discrimination task and both of the flight missions.*

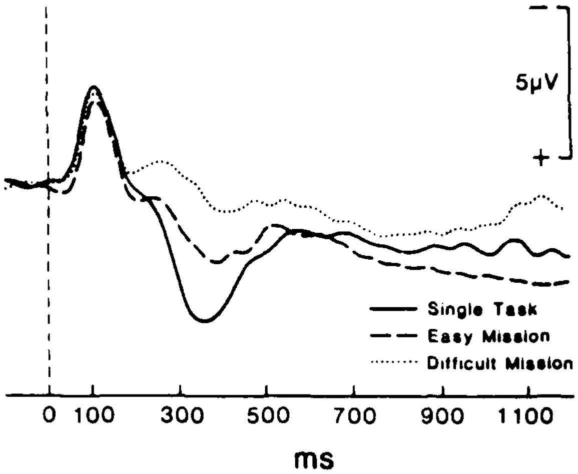


Figure 3. *Parietal grand average ERPs overplotted for the tone discrimination task alone and for both of the flight missions.*

the predefined latency ranges. Two separate analyses were performed on the ERP data. In the between-mission analysis, ERP component measures were submitted to three-way, repeated-measures analyses of variance (3 Single/Dual Tasks × 2 Probe Types × 3 Electrodes). Three factors were also entered into the within-flight analyses of variance (4 Flight Segments × 2 Probe Types × 3 Electrodes).

P300 Component

Further support for the identification of the large positive deflection with the P300 component was provided by the significant main effects for electrode, $F(1,6) = 7.9$, $p < 0.05$, and probe type, $F(1,6) = 9.8$, $p < 0.05$. P300s were larger for the low-probability tone than they were for the high-probability tone. The amplitude of the P300 increased from the Fz to the Cz to the Pz electrode site. Both of these effects are consistent with previous findings and have been used as definitional criteria for the P300 (Donchin et al., 1978; Kramer, 1985). One of the major questions in the present study was the extent to which P300 amplitude would discriminate among levels of workload imposed on the student pilots by the flight tasks. The main effect of flight mission indicated that P300 amplitude was sensitive to the task demands of the different missions, $F(2,12) = 4.7$, $p < 0.05$. Post hoc comparisons further indicated that P300s elicited by the tones in the discrimination task were largest in the single-task conditions; of intermediate amplitude when the students were flying with no wind, turbulence, or subsystem failures; and smallest with high winds, turbulence, and a heading indicator failure (for all comparisons, $p < 0.05$; see Table 3). This systematic decrease in the amplitude of the P300s elicited by the tone discrimination task alone and combined with the flight missions mimics the resource trade-offs presumed to underlie multitask performance.

Additional support for the sensitivity of P300 to changes in the resource demands of a task was found in an analysis of the effects of stimulus sequence on the P300s elicited in single- and dual-task conditions. Previous studies have shown that there is a systematic relationship between the amplitude of the P300 elicited by a stimulus on a given trial and the sequence of stimuli that precede it (Johnson and Donchin, 1978; Squires, Petuchowski, Wickens, and Donchin, 1977; Squires, Wickens, Squires, and Donchin, 1976). When a particular trial is preceded by the alternate stimulus (in the present experiment, a low tone followed by a high tone, or vice versa), a large P300 is elicited. However, when a given trial represents a repetition of a previous stimulus, a relatively small P300 is obtained. A model proposed to account for this "sequential" effect has emphasized the importance of three controlling factors: (1) the global probability of each event, (2) the specific structure of the prior sequence, and (3) the memory for event frequency in the prior sequence (Squires et al., 1976). Within the domain of complex, real-world tasks, we might expect that as the resource demands of a task increase, the memory capacity available to maintain the sequential effect might decrease. This, in turn, should decrease the difference between the amplitude of the P300s elicited by alternations and repetitions.

Figure 4 presents the grand average ERPs elicited by stimulus alternations and repetitions for single- and dual-task conditions. As can be seen from the figure, there is a large difference between the amplitude of the P300s in the tone discrimination task. Both the overall amplitude as well as the difference between the alternation and repetition P300s appears to decrease when the tone discrimination task and flight task are performed concurrently. Further decreases in amplitude are apparent when comparing the easy and difficult conditions.

The differences between these conditions

TABLE 3

Mean (and Standard Deviation) N100 and P300 Amplitude (Arbitrary Units) and Latencies (ms) for the Tone Discrimination Alone and Combined with the Two Flight Missions

		Flight Missions	
Measures	*Single Task*	*Easy Flight*	*Difficult Flight*
*N100 amplitude	1139 (480)	650 (268)	650 (235)
*N100 latency	126 (17)	121 (12)	122 (13)
**P300 amplitude	1954 (903)	902 (516)	482 (536)
**P300 latency	354 (26)	441 (63)	468 (62)

* Obtained from Fz electrode site
** Obtained from Pz electrode site

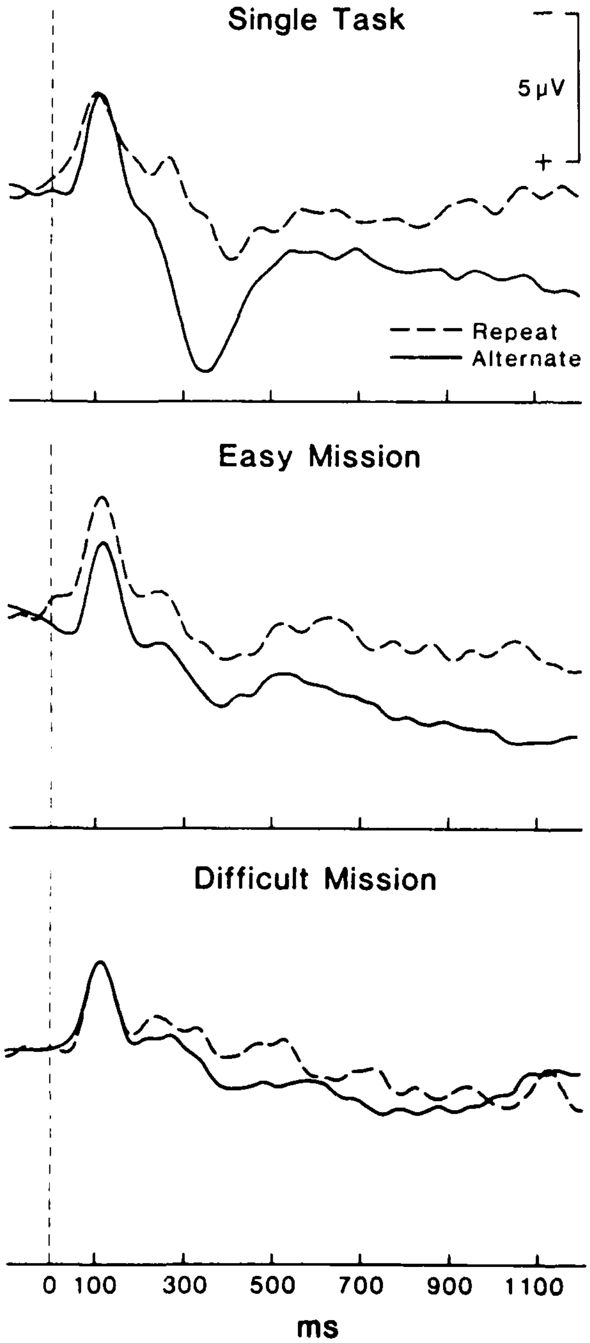


Figure 4. *Parietal grand average ERPs elicited by stimulus repetition and alternation for single- and dual-task conditions.*

were quantified by obtaining measures of the amplitude of the P300 and submitting these values to a three-way, repeated-measures analysis of variance (2 Stimulus Sequences × 3 Single/Dual Tasks × 2 Probe Types). Main effects were obtained for stimulus sequence, $F(1,6) = 25.9$, $p < 0.01$, and task, $F(2,12) = 4.2$, $p < 0.05$. More interesting, however, was the significant interaction between task and sequence, $F(2,12) = 4.0$, $p < 0.05$, which suggested that the overall amplitude as well as the difference between alternations and repetitions decreased with increasing task difficulty. Moreover, post hoc comparisons indicated that although the difference between alternations and repetitions was significant for the tone-discrimination task and the easy flight, alternation and repetition P300s did not differ significantly in the difficult flight condition. Thus, since the global probability and the sequential structure of the ERP eliciting tones was uniform across conditions, these results may imply that residual memory capacity decreased with increased task difficulty.

A main effect of flight mission was also found for the P300 latency variable, $F(3,18) = 11.3$, $p < 0.01$. Post hoc comparisons indicated that this effect could be attributed to the significant difference between single and dual tasks, ($p < 0.05$). However, latencies did not differ between the two flight missions. This finding suggests that the difference in P300 amplitude between the two flight missions cannot be explained by increased latency variability in the difficult flight. Given that single-task P300 latencies are often shorter than dual-task latencies, it appears safe to conclude that the P300 component successfully discriminated between single and dual tasks as well as between the two versions of the flight task.

Thus far we have described the effects of between-mission difficulty on the amplitud and latency of the P300 component. Our second analysis compares ERP components elicited during the four flight segments. It was predicted that the P300s elicited by the tones in the more difficult flight segments would be smaller than those recorded during the easier flight segments, reflecting increased processing demands in the more difficult conditions. Although the within-mission comparisons yielded weaker effects for the performance and subjective measures than did the between-mission analysis, an ordering of the flight segments could be ascertained. The straight and level flight segment (B) and the holding pattern (C) were flown with smaller heading deviations than the takeoff (A) and landing (D) components of the missions. The student pilots also rated the straight and level segment to be subjectively easier than the other three segments. Although the ordering of the P300 amplitudes was consistent with these measures (the mean amplitudes of Segments A through D were 209, 525, 508, and 283, respectively), the main effect for flight segment did not attain statistical significance ($p > 0.05$). However, a small but significant correlation was obtained between the amplitude of the P300 and the deviation-from-command heading ($R = -0.27$), indicating that the amplitude of P300 decreased with increases in heading deviation. The latency of P300 did not differ across flight segments.

Negative Components

Two different negative components are apparent in the waveforms presented in Figure 1. The earliest is a frontally maximal deflection, which occurs between 100 and 200 ms post-stimulus. This component, labeled the N100, has been found to be sensitive to the allocation of attention to physical attributes of stimuli such as frequency, loudness, and location. N100s are larger for any stimulus possessing the attended attribute than they are for stimuli possessing other attributes (Hillyard and Hansen, 1986). The amplitude of the N100 is also influenced by the number of information sources that must be attended simultaneously (Parasuraman, 1976; Schwent and Hillyard, 1975). N100s decrease in amplitude as subjects are required to monitor additional sources of information.

In the present experiment, an important question concerns the degree to which the N100 reflects the transition from single to dual tasks, as well as the increase in workload imposed on the student pilots by the more difficult flight task. To this end, measures of N100 amplitude and latency were submitted to a three-way, repeated-measures analysis of variance (3 Single/Dual Tasks × 2 Probe Types × 3 Electrodes). A significant main effect was obtained for the task variable, $F(2,12) = 5.0$, $p < 0.05$. Post hoc comparisons indicated that this effect could be attributed to larger N100s in the single-task condition ($p < 0.05$). The amplitude of the N100s did not differ between the easy and difficult flight missions. No other main effects or interactions were significant for N100 amplitude or latency measures. The latency and amplitude measures for the later frontally negative component, the N200, were also submitted to a three-way, repeated-measures analysis of variance. None of the main effects or interactions were significant for this component.

CONCLUSIONS

The results of the present experiment provide preliminary support for the assertion that components of the ERP can provide sensitive and reliable measures of the task demands imposed upon operators of complex, real-world systems. One component of the ERP in particular, the P300, varied in a systematic manner in response to the demands of different versions of the flight task. Relative to conditions in which the tone discrimination task was performed alone, the amplitude of the P300 decreased when the student pilots performed the easy version of the flight task. Further decreases in P300 amplitude were observed when the difficulty of the flight task was increased through the manipulation of wind speed, turbulence, and the probability of subsystem failures. Although the within-mission effects were not as dramatic as the between-mission comparisons, a small but significant correlation between P300 amplitude and heading deviation was obtained across flight segments. Another component of the ERP, the N100, varied in amplitude as a function of the number of tasks that the subjects performed.

The sensitivity of the P300 to the processing demands of the different flight missions is noteworthy for several reasons. First, the changes in the amplitude of the P300 as a function of task demands mimics the modulation of resources presumed to underlie variations in operator performance. Resource models predict that as task demands increase, additional resources will be allocated to the high-priority task, thereby withdrawing resources from tasks of lesser importance (Navon and Gopher, 1979; Wickens, 1980). The amplitude of the P300s elicited by the secondary probe task decreased with increases in the difficulty of the flight task. Other studies have found that P300s elicited by primary task events increase in amplitude with increases in task demands (Kramer et al., 1985). Thus, it appears that the P300 provides a measure of the hypothetical resources that can only be inferred from more traditional measurement techniques.

A second point concerns the use of the secondary task procedure in the assessment of mental workload. The resource demands of a primary task are usually inferred from decrements in secondary task performance (Ogden et al., 1979). However, a particular difficulty of the secondary task methodology is the intrusion of the secondary task into primary task performance, thereby complicating the interpretation of the performance decrements. In the present study, our secondary probe task required a relatively simple discrimination and an occasional overt response. In fact, secondary task performance did not discriminate among levels of primary task difficulty. Secondary task reaction time and accuracy were uniformly high in all conditions. On the other hand, the P300 elicited by the probe stimuli did discriminate among the demands of the flight task. Thus, the ERP-eliciting probe task provides a sensitive metric of resource demands without intruding upon the performance of the task of interest, a clear advantage in operational settings. However, this is not to imply that even a relatively nonintrusive secondary task is an ideal workload assessment procedure in complex, real-world systems. Clearly, a more acceptable solution would be the elicitation of ERP components by primary task events, thereby negating the requirement for any type of secondary task. Such a procedure has been successfully employed in the laboratory, and we are currently exploring its efficacy in operational settings (Kramer, Wickens, Vanasse, Heffley, and Donchin, 1981; Sirevaag, Kramer, Coles, and Donchin, 1984).

A third point concerns the nature of the metric. Although ERP components certainly qualify as physiological measures, they are somewhat unique in that they are selectively sensitive to a subset of processing demands. Autonomically mediated measures such as heart-rate variability, respiration, blood pressure, and skin conductance are influenced by ambient environmental conditions, anxiety, and physical exertion, in addition to mental workload (Wierwille, 1979). Therefore, these measures are sensitive to workload in general but are not diagnostic in the sense of identifying the source of the processing demands. The P300 is sensitive to information-processing demands, and, more specifically, is influenced by perceptual/cognitive demands but not by motor processes.

Thus, unlike other physiological measures, ERP components are quite diagnostic.

The unique characteristics of endogenous ERP components make them well suited to serve as indices of mental workload. However, since workload appears to be best represented as a multidimensional rather than a scalar quantity, ERPs can best fulfill their role by augmenting other workload metrics. Only through the joint use of a number of measurement techniques can we expect to elucidate the mental workload of human operators.

ACKNOWLEDGMENTS

This research was supported by NASA-Ames Research Center under contract No 1-5-25477 with Michael Vidulich as technical monitor, and by the Air Force Office of Scientific Research under contract No F49620-83-0144, with Alfred Fregly as technical monitor We are indebted to Gabriele Gratton and Amir Mane for their helpful comments on a earlier draft of this manuscript We gratefully acknowledge the programming assistance of Mark Klein and Brian Osmond

REFERENCES

Brown I D (1978) Dual-task methods of assessing workload *Ergonomics, 21*, 221-224

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 Kramer, A , and Wickens, C (1986) Applications of brain event-related potentials to problems in engineering psychology In M G H Coles, E Donchin, and S W Porges (Eds), *Psychophysiology Systems, processes and applications* (pp 702-718) New York Guilford

Donchin, E , Ritter, W and McCallum, W C (1978) Cognitive psychophysiology The endogenous components of the ERP In F Callaway, P Tueting, and S Koslow (Eds) *Event related brain potentials in man* (pp 349-412) New York Academic Press

Dornic, S (1980) Language dominance, spare capacity and perceived effort in bilinguals *Ergonomics, 23*, 369-377

Eggemeier, F T (1980). Some current issues in workload assessment In *Proceedings of the Human Factors Society 24th Annual Meeting* (pp 669-673) Santa Monica, CA Human Factors Society

Frazier, M L , and Crombie, R B (1982) *Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics* (AFFTC-TR-82-5) Edwards Air Force Base, CA Air Force Systems Command

Freidman, A , and Polson, M C (1981) The hemispheres as independent resource systems Limited capacity processing and cerebral specialization *Journal of Experimental Psychology: Human Perception and Performance, 7*, 1030-1058.

Gopher, D., and Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors, 26*, 519-532.

Gopher, D., and Donchin, E. (1986). What is workload: An introduction. In K. Boff, L. Kaufman, and J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp 41.1-41.49). New York: Wiley.

Gratton, G., Coles, M G. H., and Donchin, E (1983) A new method for off-line removal of ocular artifact *Electroencephalography and Clinical Neurophysiology, 55*, 464-484.

Heffley, E., Foote, B., Mui, T , and Donchin, E (1985). PEARL II: Portable laboratory computer system for psychophysiological assessment using event-related brain potentials *Neurobehavioral Toxicology and Teratology, 7*, 339-407.

Hillyard, S. A , and Hansen, J. C (1986). Attention Electrophysiological approaches In M. G. H. Coles, E. Donchin, and S. W. Porges (Eds.), *Psychophysiology Systems, processes and applications* (pp 227-243) New York Guilford.

Horst, R. L., Munson, R. C , and Ruchkin, D. S (1984). Event-related potential indices of workload in a single task paradigm In *Proceedings of the Human Factors Society 28th Annual Meeting* (pp. 727-731) Santa Monica, CA: Human Factors Society

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

Jasper, H. H (1958). The ten-twenty electrode system of the international federation *Electroencephalography and Clinical Neurophysiology, 10*, 371-375

Johannsen, G., Moray, N., Pew, R., Rasmussen, G., Sanders, A., and Wickens, C (1979) Final report of the experimental psychology group In N Moray (Ed), *Mental workload Its theory and measurement* (pp 101-116). New York: Plenum

Johnson, R. E., and Donchin, E (1978). Does P300 amplitude depend on the expectancy for physical stimuli or for stimulus categories? *Psychophysiology, 15*, 262

Kahneman, D (1973) *Attention and effort* Englewood Cliffs, NJ: Prentice Hall

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

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

Kramer, A. F (1987). Cortical evoked potentials In A Gale and B. Christie (Eds.), *Psychophysiology and the electronic workplace*. Sussex, England: Wiley

Kramer, A. F., Wickens, C D., and Donchin, E (1983) An analysis of the processing demands of a complex perceptual-motor task. *Human Factors, 25*, 597-622.

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

Kramer, A. F., Wickens, C. D., Vanasse, L., Heffley, E , and Donchin, E (1981). Primary and secondary task analysis of step tracking: An event-related brain potentials approach. In *Proceedings of the Human Factors Society 25th Annual Meeting* (pp. 43-47). Santa Monica, CA Human Factors Society.

Moray, N. (1979). *Mental workload: Its theory and measurement*. New York: Plenum.

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

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

O'Donnell, R. D., and Eggemeier, F. T. (1986). Workload assessment methodology. In K. Boff, L. Kaufman, and J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 42.1-42.49). New York: Wiley.

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

Parasuraman, R. (1976). Auditory evoked potentials and divided attention. *Psychophysiology, 15*, 460-465.

Pritchard, W. S. (1981). Psychophysiology of P300. *Psychological Bulletin, 89*, 506-540.

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

Rolfe, J. M. (1971). The secondary task as a measure of mental load. In W. T. Singleton, J. G. Fox, and D. Whitfield (Eds.), *Measurements of man at work* (pp. 135-148). London: Taylor and Francis.

Sanders, A. F. (1979). Some remarks on mental load. In N. Moray (Ed.), *Mental workload: Its theory and measurement* (pp. 41-78). New York: Plenum.

Schwent, V. L., and Hillyard, S. W. (1975). Evoked potential correlates of selective attention with multichannel auditory inputs. *Electroencephalography and Clinical Neurophysiology, 38*, 131-138.

Sirevaag, E., Kramer, A. F., Coles, M. G. H., and Donchin, E. (1984). P300 amplitude and resource allocation. *Psychophysiology, 21*, 598-599.

Squires, K. C., Petuchowski, S., Wickens, C. D., and Donchin, E. (1977). The effects of stimulus sequence on event related potentials: A comparison of visual and auditory sequences. *Perception and Psychophysics, 22*, 31-40.

Squires, K. C., Wickens, C. D., Squires, N. K., and Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science, 193*, 1142-1146.

Starr, A. (1978). Sensory evoked potentials in clinical disorders of the nervous system. *Annual Review of Neurosciences* (pp. 103-127). Palo Alto, CA: Annual Reviews, Inc.

Stockard, J. J., Stockard, J. E., and Sharbrough, F. W. (1979). Brainstem auditory evoked potentials in neurology: Methodology, interpretation and clinical application. In M. J. Aminoff (Ed.), *Electrophysiological approaches to neurological diagnosis* (pp. 370-413). New York: Churchill-Livingstone.

Strayer, D. L., and Kramer, A. F. (1986). Psychophysiological indices of automaticity and attentional resources. *Psychophysiology, 23*, 466.

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

Wickens, C. D. (1979). Measures of workload, stress and secondary tasks. In N. Moray (Ed.), *Mental workload: Its theory and measurement* (pp. 79-100). New York: Plenum.

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson and R. Pew (Eds.), *Attention and performance VIII* (pp. 239-258). Hillside, NJ: Erlbaum.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman and D. R. Davies (Eds.), *Varieties of attention* (pp. 63-102). New York: Academic Press.

Wickens, C. D., Kramer, A. F., Vanasse, L., and Donchin, E. (1983). The reciprocity of primary and secondary task resources: Evidence from the P300 component of the ERP. *Science, 221*, 1080-1082.

Wierwille, W. W. (1979). Physiological measures of aircrew mental workload. *Human Factors, 21*, 575-594.

Williges, R. C., and Wierwille, W. W. (1979). Behavioral measures of aircrew mental workload. *Human Factors, 21*, 549-574.

Effects of Foveal Task Load on Visual-Spatial Attention: Event-Related Brain Potentials and Performance

Arthur F. Kramer, Erik J. Sirevaag and Paul R. Hughes

Department of Psychology
University of Illinois
Champaign, Illinois

Acknowledgment

The research was supported by NASA Ames Research Center under contract number 1-5-25-477 with Dr. Michael Vidulich as technical monitor and by the Air Force Office of Scientific Research under contract number F49620-79-C-0233 monitored by Dr. Al Fregly. The authors wish to thank Charles Eriksen, Marta Kutas, Greg Miller, Raja Parasuraman and Mick Rugg for their helpful comments on an earlier draft of this paper. Requests for reprints should be sent to Arthur F. Kramer, Department of Psychology, 603 East Daniel Street, University of Illinois, Champaign, Illinois, 61820.

PSYCHOPHYSIOLOGY (in press)

Abstract

The effects of foveal task difficulty on the processing of events in the visual periphery were investigated through an analysis of event-related brain potentials and performance measures. Subjects performed a foveally presented continuous monitoring task both separately and together with an arrow discrimination task which was presented at three different retinal eccentricities. The subjects detected occasional failures in the monitoring task while also responding to designated targets in the left and right visual fields. The analysis of the event-related brain potentials elicited by discrete events in the arrow discrimination task indicated that the amplitude of the N190 and P300 components decreased with both the introduction of the foveal task and an increase in its difficulty. The N160 component was sensitive to the distribution of attention within a task but was uninfluenced by dual-task demands. These findings suggest that the N160 reflects the distribution of attention to different spatial locations within a task while the N190 may index the distribution of general purpose perceptual resources. P300 appears to index the allocation of perceptual/central processing resources. The implications of the results for models of resource allocation and attentional gradients are discussed.

DESCRIPTORS: Selective and divided attention, dual-tasks, resource allocation, Event-related brain potentials (ERP), visual-spatial attention, N160, N190, P300.

Introduction

Research conducted to examine the sensitivity of event-related brain potentials (ERPs) to attentional processes has been mainly performed within two different paradigms (Donchin, Kramer and Wickens, 1986; Hillyard and Kutas, 1983; Kutas and Hillyard, 1984). Moreover, these paradigms are generally associated with theoretical frameworks that focus on different aspects of attention. In the present study, we have designed a hybrid paradigm that allowed us to investigate the interaction between two varieties of attentional processes; dual-task processing demands and intratask selective attention. Prior to describing our task and hypotheses in more detail we will outline the paradigms and theoretical perspectives from which our study is derived.

One of the most frequently used paradigms is a modification of the classic dichotic listening/selective looking task (Becklen and Cervone, 1983; Cherry, 1953; Broadbent, 1958; Moray, 1959; Neisser and Becklen, 1975). In the ERP based version of this task subjects are instructed, in different conditions, to (a) attend to information in one of two or more channels and ignore information presented on the other channels, or (b) to divide their attention among the channels (Eason, Harter and White, 1969; Eason and Ritchie, 1976; Hillyard, Hink, Schwent and Picton, 1973). Channels are generally defined in terms of levels along a physical dimension of a stimulus (eg. ear, pitch, loudness, location, color, orientation). Within each of the channels events can be further categorized on the basis of physical or semantic differences. The subjects are instructed to respond each time one of two events occurs on the attended channel. The event that requires a response is labelled the target and usually occurs with a lower probability than the standards.

Although the specific ERP components that are recorded in the dichotic listening/selective looking paradigm differ somewhat as a function of modality (Hillyard, Simpson, Woods, Van Voorhis and Munte, 1984), a number of components have been extensively examined across modalities. These components include an early negativity that peaks at approximately 80-120 msec in the auditory modality and from 140 to 170 msec in the visual modality. Attention effects include both amplitude modulations of these peaks as well as broader, sometimes overlapping negativities which are known as Nd in the auditory domain and selection negativity in visual attention. A later positivity occurring at approximately 300 to 700 msec post-stimulus and referred to as the P300 has been found to be sensitive to attentional manipulations in the auditory and visual domains (Harter and Aine, 1984; Naatanen, 1982; Picton, Campbell, Baribeau-Braun and Proulx, 1978). Several of the experimental results obtained in the examination of the sensitivity of the negativities and the P300 component to the manipulation of attentional processes are of particular relevance in the present study.

First, it is clear from experimental results that a hierarchy of selective mechanisms exist. The amplitude of the N100 component discriminates between the processing of information in attended and unattended channels, while later negativities and the P300 discriminate among target and non-target items within the attended channel (Hillyard, Munte and Neville, 1985; Naatanen, 1982). Since channels have been defined in terms of levels along physical dimensions, this pattern of results provides strong support for early selection models of attention which posit that information in the ignored channels is attenuated by a filter that is tuned on the basis of physical stimulus features (Broadbent, 1982; Johnston and Dark, 1982; Treisman, 1969; Treisman and Riley, 1969). This filtering strategy has been termed stimulus set selection (Broadbent, 1958).

The subsequent negativities and P300 component also suggest that further filtering of irrelevant information takes place later in the information processing system. This filtering may reflect response set mechanisms posited by late selection models of attention (Keele, 1973; Deutsch and Deutsch, 1963; Duncan, 1980; Shiffrin, McKay and Shaffer, 1976). Thus, the ERP results are quite consistent with both empirical findings as well as recent models that suggest that attentional selection is relatively flexible and can take place on the basis of either physical features or semantic content depending on task constraints and subject strategies (Bashinski and Bacharach, 1980; Beck and Ambler, 1973; Dark, Johnston, Myles-Worsley and Farah, 1985; Francolini and Egeth, 1980; Kahneman and Henik, 1981; Kahneman and Treisman, 1984; Pashler, 1984).

Another important finding obtained in the ERP based dichotic listening/selective looking paradigm was the demonstration of attentional gradients for a variety of physical dimensions. In the visual modality, N100s elicited by unattended events have been found to systematically decrease in amplitude relative to attended events with increasing physical distance (Mangun and Hillyard, 1987) as well as increased spatial frequency separation (Harter and Previc, 1978). Similar differences have been obtained with respect to changes in frequency in auditory paradigms (Alho, Sams, Paavilainen and Naatanen, 1986). These ERP results are consistent with several recent psychophysical studies that have obtained evidence for attentional gradients in both two (Eriksen and Yeh, 1985; LaBerge and Brown, 1986; Shulman, Wilson and Sheehy, 1985) and three dimensional space (Downing and Pinker, 1985).

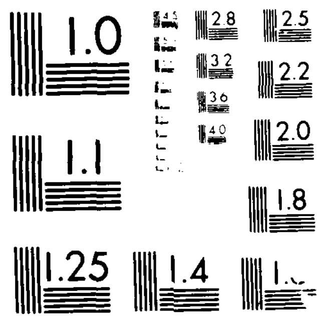
A Zoom Lens model has been proposed to account for attentional gradients in visual space (Eriksen and Saint James, 1986). This model suggests that attention can operate along a spatial continuum ranging from tightly focused to widely distributed. When attention is focused, items in the visual field are processed sequentially but with a high degree of precision. When attention is distributed, items are processed in parallel but with a lower degree of precision. Attentional gradients are presumed to result from a gradual dropoff in processing resources towards the boundary of the attentional focus. The ERP data suggest that the mechanism underlying the phenomenon of attentional gradients operates at a relatively early stage of processing.

The Zoom Lens model employs a resource metaphor to account for the differences in speed and precision of processing during focused and divided attention. A fixed quantity of resources is proposed to be available for processing of events in the visual field. Resources are evenly distributed with a low density during divided attention, while being concentrated with a high density during focused attention. Thus, the density of resources limits the rate and extent of processing of events within the focus of attention. It is noteworthy that ERP results are quite consistent with the resource metaphor. The amplitude of the N100 has been found to be largest when elicited by events in attended channels, intermediate in amplitude in divided attention conditions, and smallest when elicited by unattended events during focused attention (Hink, Van Voorhis, Hillyard and Smith, 1977; Okita, 1979; Parasuraman, 1978, 1985). Furthermore, the summed amplitude of the N100s remains constant between conditions of focused and divided attention (hink et al., 1977; Van Voorhis and Hillyard, 1977). Thus, the amplitude of the N100 component appears to reflect the density of resources allocated to events as well as the fixed quantity of resources available for processing.

Although the examination of ERPs within the dichotic listening/selective looking paradigm has produced a wealth of valuable information concerning the mechanisms underlying focused and divided

AD-A191 244 THE EVENT-RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING A..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 5/10

UNCLASSIFIED 29 FEB 88 CPL-88-1 AFOSR-TR-88-0316 F/G 5/8 NL



MICROCOPY RESOLUTION TEST CHART

attention, research in this paradigm has been limited to the investigation of intratask attentional processes. Thus, one important question not addressed within this research program is the degree of generalizability of attentional phenomena to multitask environments. The issue of attentional tradeoffs between, rather than within, tasks has been addressed in the dual task paradigm. Within the dual task paradigm subjects perform two tasks both separately and together. Dual task demands are manipulated by varying the difficulty and/or the processing priorities of one or both of the tasks. Although dual task ERP studies have concentrated mainly on the sensitivity of P300 to changing attentional demands, several noteworthy findings have been obtained.

When subjects are instructed to maximize their performance on one task and the difficulty of this task is manipulated, P300s elicited in the secondary task decrease systematically with increases in the difficulty of the primary task (Isreal, Chesney, Wickens and Donchin, 1980: Kramer, Wickens and Donchin, 1985; Kramer, Sirevaag and Braune, 1987; Lindholm, Cheatham, Koriath and Longbridge, 1984; Natani and Gomer, 1981). In conditions in which ERPs are also recorded in the primary task, the amplitude of the P300s increase with increases in the difficulty of this task (Horst, Munson and Ruchkin, 1984; Kramer, Wickens and Donchin, 1985; Sirevaag, Kramer, Coles and Donchin, 1984; Wickens, Kramer, Vanasse and Donchin, 1983). The reciprocal relationship between P300s elicited by primary and secondary task stimuli is consistent with the attentional tradeoffs presumed to underlie dual task performance decrements (Freidman and Polson, 1981; Kahneman, 1973; Navon and Gopher, 1979; Sanders, 1979; Wickens, 1980, 1984). That is, attentional resource models predict that as the difficulty of one task is increased, additional resources are reallocated to that task in order to maintain performance, thereby depleting the supply of resources that could have been used in the processing of other tasks. Resource models also predict that attentional tradeoffs should occur when subjects shift their emphasis from one task to another, even in the absence of difficulty manipulations. P300s have also been found to mimic the shifts in attentional resources expected with changes in the processing priorities of two tasks (Hoffman, Houck, MacMillian, Simons and Oatman, 1985; Kramer and Strayer, 1987). Thus, the P300 appears to provide a measure of attentional tradeoffs that can only be inferred from more traditional performance measures.

The ERP results obtained in the dichotic listening/selective looking and dual task paradigms have provided converging evidence for theories of attention as well as new insights into the mechanisms underlying attentional phenomena. However, there are a number of questions that remain to be addressed, particularly in regards to the interaction of the varieties of attention investigated in these paradigms. In the present study we will examine a number of questions relevant to the effects of intertask demands on the processing of stimuli in the visual field.

One such question concerns the level of processing at which attentional resources are transferred from one task to another during dual task performance. Although N100s have been found to decrease in amplitude from focused to divided attention conditions within a single task, it is unknown whether this effect generalizes from single task to dual task conditions. At a finer level of analysis, it becomes relevant to ask if and how the visual-spatial N100 attention effect is influenced by intertask demands. Is it the case that the amplitude of the N100s elicited by both attended and ignored events are reduced or does the reduction reflect a selective focusing of attention (eg. diminished ignored versus attended N100s)? This same question is relevant for the P300 component. Previous investigations of the sensitivity of the P300 to changes in intertask demands have focused solely on the P300s elicited by task relevant events. In the present study we examine the degree to which the sensitivity of the P300 to intralocation and interlocation selective processes is influenced by demands imposed from another task.

A second important question concerns the effects of intertask demands on visual-spatial attentional gradients. A similar issue has been addressed in performance based paradigms. In these studies subjects are required to selectively respond to brief presentations of items in the periphery of the visual field while also performing a foveal task. In a number of such studies, imposing a foveal task and increasing its difficulty has led to a decrement in the processing of information in the periphery, particularly at the most distant locations (Ikeda and Takcuchi, 1975; Williams, 1985). This phenomenon has been termed cognitive tunnel vision (Mackworth, 1965). Although these findings suggest the intriguing possibility that the functional visual field constricts with increased foveal task demands, several aspects of the research may limit the generalizability of this phenomenon.

First, cognitive tunnel vision has only been obtained in a subset of the studies examining the effects of foveal task load on peripheral processing. In many cases all stimuli in the periphery were more poorly processed with increased foveal task demands (Antes and Edwards, 1973; Holmes, Cohen, Haith and Morrison, 1977; Voss, 1981; Williams, 1982). Second, the effects have been found only in situations in which the foveal and peripheral stimuli were presented very briefly. Third, since only measures of accuracy or RT were collected, the level of processing at which this effect occurred could not be assessed. In the present study we will examine the effects of foveal task load on the selective processing of items in the periphery by increasing the difficulty of a foveally presented continuous monitoring task and examining the ERPs elicited by stimuli presented at different retinal eccentricities.

Method

Subjects Ten dextral graduate students (five male) were paid for their participation in the study. All of the subjects were between the ages of 20 and 26 and had normal or corrected to normal vision. None of the subjects had any prior experience with the tasks employed in the experiment.

Triangle Monitoring and Arrow Discrimination Tasks Subjects performed two tasks both separately and together. The tasks are graphically depicted in Figure 1. The foveal task required subjects to monitor a dynamic process for occassional failures. The multi-attribute process was represented by a triangle which changed shape as a function of three system variables. The triangle was displayed at the center of a CRT positioned 88 cm directly in front of the subject. The difficulty of the monitoring task was varied by manipulating the complexity of the rule for detecting a system failure. Two independent sums of six sine waves were used to update the length of segments I1 and I2. Note that the sum of these segments determined the length of the base of the triangle.

In the correlated monitoring condition, changes in the length of segment O (which determined the height of the triangle) were perfectly correlated with changes in segments I1 and I2. During a system failure, the length of segment O was driven by a third sum of sines function, uncorrelated with the changes in segments I1 and I2. Thus, a system failure was said to occur when the relationship between the height and the base of the triangle changed from a correlated to an uncorrelated state. The change from a normal to an abnormal system took place in the form of a ramp function over a 1200 msec epoch.

During the uncorrelated monitoring condition, the contingencies were reversed. In this condition, the subjects were to detect when the system changed from the uncorrelated state to one in which the height of the triangle depended upon changes in the I1 and I2 segments. The horizontal base (segment I1 and I2) of the triangle varied from 0.3 to 2.6 degrees of visual angle. The vertical height of the triangle (segment O) varied from 0.3 to 2.0 degrees of visual angle. The triangle dimensions were updated every 50 msec.

The number of failures that could occur on any block of trials varied from 10 to 15. The failures occurred with a stimulus onset asynchrony (SOA - the time between the end of one failure and the beginning of the next) which varied randomly between 5 and 50 sec. Following the onset of a system failure, subjects were given an interval of 6000 msec to signal the detection of the failure by depressing a button with the thumb of one hand. The assignment of response hand to the detection of triangle failures was counterbalanced across subjects. Responses occuring within this 6000 msec interval were scored as hits and the reaction time (RT) was recorded. If the subject failed to respond within this interval, a miss was scored and the parameters of the algorithm controlling the triangle were reset to the normal system state. Button press responses that occured when the triangle was not in a "failure" state were scored as false alarms (FAs). Based upon previous research it was expected that the detection of failures in the uncorrelated system would be substantially slower and more error prone than in the correlated system (Casey and Wickens, 1986).

In the parafoveal task subjects were required to selectively attend to either the right or left visual field and depress a response button whenever a prespecified target, an upward or downward pointing arrow, was presented. During a given block of trials, arrows pointing in the target direction, regardless of visual field, occurred with a probability of .20, arrows pointing in the non-target direction occurred with a probability of .80. The probability of an arrow occurring in either the right or left visual field was .50. Thus, the number of trials actually requiring a response (those that pointed in the target direction in the appropriate visual field) constituted 10% of the trials during a given block. Subjects' signaled the detection of a target by depressing a button with whichever thumb was not used to respond to the triangle monitoring task.

The arrows were presented at three different retinal eccentricities (defined in terms of deviations in degrees of visual angle from the center of the screen); "close" (1.4 degrees), "middle" (4.5 degrees), and "far" (9.1 degrees). During all conditions in which the arrow discrimination task was performed, a total of 200 arrow stimuli were presented with an inter-stimulus interval (ISI) that varied between 1.2 and 1.5 sec. Each arrow was presented for a 50 msec duration and subtended a visual angle of 1.3 degrees. RT, hits, misses, and FAs were collected for each experimental trial. Failures in the triangle monitoring task and targets in the arrow discrimination task were offset by at least 1500 msec in the dual task conditions.

Before each block of trials subjects were told which visual field to attend. Special emphasis was placed on the requirement that the eyes remain fixated at the center of the screen. Trials containing eye movements and blinks were not included in the analysis. Each subject performed an equal number of attend left and attend right blocks. The retinal eccentricity variable was similarly blocked. Thus, stimuli occurred at the same retinal eccentricity within the attend and ignore visual fields during a single block of trials. The direction of the target arrow was counterbalanced across subjects such that an individual subject always responded to target arrows pointing in the same direction. This combination of variables produced a total of 6 different single task arrow discrimination conditions (2 visual fields x 3 retinal eccentricities). When combined with the two levels of triangle monitoring (correlated and uncorrelated) a total of 12 dual task conditions resulted.

Insert Figure 1 About Here

ERP Recording Electroencephalographic (EEG) data were recorded from four midline sites (Fz, Cz, Pz, and Oz). In addition, activity at two non-standard lateral occipital sites was also recorded. These lateral leads were located 2.5 cm to the left (designated O1) and 2.5 cm to the right (designated O2) of the Oz position. All active EEG electrodes were referenced to linked mastoids. Two ground electrodes were positioned on the left side of the forehead.

Electrooculographic (EOG) activity in the vertical plane was monitored by electrodes placed above and below the subject's left eye. Horizontal EOG activity was monitored by placing an electrode at the outer canthus of each eye. Beckman Biopotential electrodes filled with Grass paste were utilized at all EEG, EOG, reference, and ground sites. Electrode impedances were maintained below 10 kohms.

The EEG and EOG were amplified with Van Gogh Model 50000 amplifiers (time constant 10 s and upper-half amplitude of 70 Hz). EEG and EOG channels were sampled every 5 msec, during a 1000 msec epoch beginning 100 msec prior to the onset of the stimuli in the arrow discrimination task. Single trials containing blinks or horizontal eye movements were rejected prior to statistical analysis.

Stimulus Generation and Data Collection Stimulus presentation and data acquisition were governed by a DEC PDP 11/73 human performance - electrophysiological laboratory computer (Heffley, Foote, Mui, and Donchin, 1985) interfaced with an Imlac graphics processor. Single trial and average EEG and EOG data were monitored on-line by use of a GT-44 display. The digitized data from the single trials were stored on magnetic tape for off-line analysis.

Procedure The experimental conditions were presented in a blocked fashion. Each block lasted approximately 5 min and required either single task triangle monitoring, single task arrow discriminations, or dual task performance.

Subjects participated in three experimental sessions each of which took place on separate days. The first session was considered practice. The data from this session were not analyzed. Each subject performed eleven blocks of trials in the practice session. Two of these blocks did not require responses, subjects merely monitored the movements of the triangle. When a failure occurred the word "FAILURE" was presented on the CRT immediately above the triangle. These conditions served to familiarize the subject with the correlated and uncorrelated systems. The remaining blocks required the subjects to respond to the detection of system failures. In some of the blocks subjects performed the triangle task alone, while in others they were required to monitor the triangle and respond to the stimuli in the arrow discrimination task.

ERPs, RTs and accuracy measures were recorded during the subsequent two sessions. During a given session subjects performed only one version of the monitoring task (correlated or uncorrelated). The presentation order was counterbalanced across subjects. Subjects performed 21 blocks of trials in each of the sessions. Three of these were single task monitoring (either correlated or uncorrelated). Six single task arrow discrimination blocks were also completed in each session (2 visual fields x 3 retinal eccentricities). Two replications of each of the six dual task blocks, the monitoring task paired with each of the six arrow discrimination conditions, completed the 21 blocks in each of the experimental sessions. The presentation order of single and dual tasks was counterbalanced across subjects and sessions.

Data Analysis Three ERP components were analyzed in this experiment (N160, N190, and P300). For the two negative components, amplitude and latency estimates were obtained by submitting the average data from each electrode to an algorithm which selected the most negative peak relative to the mean value of the pre-stimulus baseline in a latency window derived from visual inspection of the grand mean waveform (100 to 300 msec post-stimulus).

Because latency variability in single trial P300s can contribute to amplitude differences in average waveforms, estimates of P300 amplitude and latency were derived from the single trials. These estimates were derived using two different techniques. First, the waveforms from each of the electrodes were submitted to a covariance algorithm. This procedure computed the covariance of each waveform with a cosinusoidal waveform within a moving 500 msec window beginning 300 msec post-stimulus and ending at 800 msec post-stimulus. P300 latency was defined as the midpoint in the epoch which provided the maximum covariance, and P300 amplitude was estimated from the magnitude of the maximum covariance. Trials for which the maximum waveform/template correlation coefficient was less than .35 were discarded from the analysis. The second signal extraction technique employed in the estimation of P300 amplitude and latency was a base to peak measure on the single trials. The largest positive peak was selected in a 300 to 800 msec post-stimulus window. Since a comparison of the measures obtained via cross correlation and base to peak did not reveal any significant differences between the pattern of results, the F ratios will be reported for the base to peak measures. In order to protect against positive biases in the F tests the significance levels were computed with conservative degrees of freedom (Greenhouse and Geisser, 1959).

Results

Performance measures

Triangle Monitoring Task

RT was defined as the interval between the beginning of the "failure" and the subjects' keypress. Consistent with our predictions, subjects were more successful in detecting failures when monitoring the triangle in which the two base legs were correlated with the height than when monitoring the triangle in which the base legs and height varied independently. The data presented in Table 1 represents the average performance measures obtained in the dual task triangle monitoring conditions. Two separate ANOVA's were performed on the triangle monitoring data; a comparison of single and dual task performance and a separate dual task ANOVA.

Insert Table 1 About Here

In our instructions to the subjects we indicated that they should protect their triangle monitoring performance in the dual task conditions at the expense of their performance in the arrow discrimination task. Thus, it was emphasized that the triangle monitoring task was of primary importance and that their performance in this task should be maintained at single task levels. In order to interpret secondary task decrements as due to the allocation of scarce resources, rather than simply due to a tradeoff in performance in one task against the other, primary task performance must remain unchanged (Wickens, 1984). A comparison between single and dual task triangle monitoring measures indicated that subjects did protect their performance in the primary task. RT, hits, misses and FA's did not differ in the single and dual task monitoring conditions (all p's>.15). 1

The performance data obtained in the dual task conditions were analyzed in a series of three-way repeated measures ANOVA's (10 subjects x 2 tasks x 3 retinal eccentricities). As predicted, subjects were both faster ($F(1,9)=179.2$, $p<.01$) and more accurate ($F(1,9)=13.7$, 27.8, 19.7, $p<.01$ for hits, misses and FA's respectively) when monitoring for failures in the correlated than in the uncorrelated conditions. This same pattern of results was obtained in the single task triangle monitoring conditions. The retinal eccentricity of the arrow stimuli did not significantly influence any of the performance measures. Thus, the pattern of results obtained in the monitoring task indicated that (a) subjects were capable of maintaining their single task performance in the dual task conditions and, (b) that our manipulation of the degree of correlation between triangle dimensions was successful in influencing subjects' performance on the task.

Arrow Discrimination Task

Table 2 presents the average RT data for both the single and dual task conditions. Several trends in the data are noteworthy. First, RT increased with the introduction of the triangle monitoring task. Furthermore, this increase in RT was larger with the more difficult version of the monitoring task. Second, RT increased as the arrows were presented further in the periphery, especially in the single task condition. The results of a three-way repeated measures ANOVA (10 subjects x 3 task conditions x 3 retinal eccentricities) confirmed these observations. Subjects were significantly faster when performing in the single task than they were in the dual task conditions (F(2,18)=37.3, <.01). RT also increased with the retinal eccentricity of the arrow stimuli (F(2,18)=25.3, p<.01). The interaction between task condition and retinal eccentricity was also significant (F(4,36)=5.9, p<.05) indicating that the increase in RT was larger from the single task condition to the correlated dual task conditions than it was from the correlated to the uncorrelated conditions. Post-hoc comparisons indicated that the dual-task conditions were reliably different at the close and middle arrow positions.2

Insert Table 2 About Here

A similar pattern of results was obtained for misses. Subjects failed to respond to the target arrows more fequently in the correlated dual task conditions than in the single task conditions. Misses also increased from the correlated to the uncorrelated dual task conditions (F(2,18)=11.8, p<.01). Number of misses increased as a function of retinal eccentricity, with more misses occurring farther in the periphery (F(2,18)=6.5, <.05). FA's also increased as a function of the retinal eccentricity of the arrow stimuli (F(2,18)=6.6, P<.05).

The pattern of results obtained for the arrow discrimination task suggests that the demands imposed upon the subjects by the foveally presented triangle monitoring task influenced subjects' performance in the peripheral task. RT increased while accuracy decreased with the introduction of the monitoring task. Furthermore, the increase in the difficulty of the monitoring task produced additional decrements in the subjects' performance on the arrow discrimination task. These results when viewed in conjunction with the performance measures obtained from the triangle monitoring task, argue that subjects were compensating for increases in primary task demands by reducing their accuracy and increasing their response time in the secondary task.

Although the performance measures provide valuable information concerning the resource allocation policies adopted by the subjects to cope with the increases in task demands, they do not clarify the level within the information processing system at which these resources were exchanged. The performance measures are also silent on the degree to which processing resources are withdrawn from different elements in the visual field. Thus, it is conceviable that the processing of all peripherial stimuli may be reduced, the processing of stimuli in the unattended spatial locations may be reduced, or the processing of stimuli not sharing all of the features of the target stimuli may be curtailed. It is for answers to these questions that we turn to an analysis of the ERPs.

Event-Related Brain Potentials

The ERP section is organized around the issues of intra- and intertask attention as well as the components of the ERP that have been proposed to index different levels of attentional analysis. We begin by describing the effect of the experimental manipulations on the amplitude and latency of the visual N100 component. Since the N100 recorded at occipital sites has previously been found to differ in its sensitivity to attentional manipulations from N100s recorded at frontal, central and parietal sites, the N100s recorded from the occipital leads will be analyzed separately (Harter and Aine, 1984; Mangun and Hillyard, 1987; Rugg, Milner, Lines and Phalp, 1987; Van Voorhis and Hillyard, 1977). These components will henceforth be labelled according to their average latency, N160 for the component recorded at the frontal, central, and parietal sites and N190 for the occipitally recorded component. The amplitude and latency of the P300 component will be characterized in terms of their sensitivity to changes in processing demands and selective attention.

Preliminary analysis of the effects of visual field (attend left or right) on the ERP components did not produce any significant effects. Therefore, the ERPs elicited by the attended arrows in the left and right visual fields were averaged together prior to further analysis. The same procedure was followed for the averaging of the unattended stimuli. For the lateralized electrode placements (O1 and O2) preliminary analysis indicated that electrode site did not interact with other experimental variables. Therefore, since attentional effects have previously been shown to differ between ipsilateral and contralateral configurations, averages were produced for the attended and unattended stimuli in the contralateral and ipsilateral fields.

Insert Figures 2,3,4 & 5 About Here

Early Selective Attention and Processing Demands - N100 Components

The ERPs recorded in single and dual task conditions for standards at the Fz, Cz and Pz electrode sites and targets at the Pz site are presented in figures 2 through 5, respectively. As can be seen from the figures, the ERPs were composed of a number of positive and negative peaks. Of particular interest in the present study are the early negativity peaking at 160 msec post-stimulus, and the late positivity which is largest for the target stimuli and peaks at approximately 450 msec post-stimulus.

N160 Component The N160 component was quantified by computing base-peak measures of amplitude and latency for each of the single and dual task conditions. These values were then submitted to five-way repeated measures ANOVAs (10 subjects x 3 task levels x 3 retinal eccentricities x 2 attention conditions x 2 stimulus types) for Fz, Cz and Pz electrode sites.

As can be seen from figures 2 through 5, the N160s elicited by stimuli in the attended field were larger than the N160s elicited by stimuli in the unattended field. The effect of direction of attention was significant at both the Cz ($F(1,9)=5.9$, $p<.05$) and Pz sites ($F(1,9)=10.7$, $p<.01$) while being marginally significant at Fz ($F(1,9)=4.7$, $p<.06$). A significant two-way interaction between attention and retinal eccentricity was also obtained at the Cz ($F(2,18)=5.5$, $p<.05$) and Pz sites ($F(2,18)=6.9$, $p<.05$). The direction of the interaction is apparent from a comparison of the amplitude measures presented in Table 3. Post-hoc comparisons indicated that the attention effects were significant for the middle and far retinal eccentricities but not for the close position. Furthermore, the increase in the amplitude of the N160 as a function of eccentricity was significant only for the attended field stimuli. This pattern of results is consistent with Hillyard and Munte (1984) who found that when stimuli were presented in close proximity, N160s did not differentiate between attended and unattended locations. The results also appear to suggest that the increase in the attention effect with retinal eccentricity is due to an increase in the N160s elicited by the attended field stimuli rather than a decrease in the N160s elicited by stimuli in the unattended field.

Insert Table 3 About Here

Although it appears upon inspection of the grand average waveforms that the absolute amplitude of the N160s may have declined from single to dual task conditions this effect was not significant at any of the electrodes (all p's >.25). Furthermore, task level did not interact with the attention effect. Thus, the increased demands imposed upon subjects by the introduction of the monitoring task and the increase in its difficulty was not reflected in the level of attentional selection indexed by the N160 component. This pattern of results is particularly interesting when considered in light of other findings which have indicated that both auditory (Hink, Van Voorhis, Hillyard and Smith, 1977; Parasuraman, 1978) and visual N100's (Parasuraman, 1985; Van Voorhis and Hillyard, 1977) decrease in amplitude when subjects are required to divide their attention between two channels of information instead of focusing on a single channel. In the dual task conditions in the present experiment, however, subjects divided their attention between two tasks rather than two channels of information in a single task. Thus, the N160 may reflect the division of attention among events within a task but may be insensitive to inter-task demands.

A significant main effect for stimulus type was obtained at the Fz ($F(1,9)=12.3$, $p<.01$) and Cz sites ($F(1,9)=7.3$, $p<.05$). The larger N160s elicited by the targets may be due to the probability difference between the two classes of stimuli. This, in turn, would lead to a longer ISI between targets than standards thereby allowing less recovery for the standards (Picton, Woods and Proulx, 1978; Woods and Courchesne, 1986; Woods, Courchesne, Hillyard and Galambos, 1980). Another possibility is that the smaller number of trials in the target average may have resulted in greater amplitude variability for the targets than the standards. It is interesting to note that although there was a significant main effect for stimulus type, the lack of an interaction of stimulus type with other experimental variables indicates that the attentional manipulations had the same effect on the targets and the standards.

A small increase in N160 latency was observed from the frontal to the parietal site (160, 165, 169 msec for Fz, Cz and Pz, respectively). The only significant effect for N160 latency was an interaction between task level and attention condition ($F(2,18)=7.0$, $p<.05$) at Fz. Post-hoc tests revealed that the N160s elicited by ignored stimuli in the difficult dual task conditions were significantly longer (171 versus 158 msec) than the N160s elicited by the attended and ignored stimuli in the other single and dual task conditions.

Insert Figures 6, 7 & 8 About Here

N190 Component The N190 component was quantified by computing the base to peak measures of amplitude and latency for each of the single and dual task conditions for Oz as well as the ipsilateral and contralateral configurations. These values were then submitted to a six-way repeated measures ANOVA (10 subjects x 3 task levels x 3 retinal eccentricities x 2 attention conditions x 2 stimulus types x 3 electrode sites).

Consistent with the N160 findings, a significant main effect was obtained for attention ($F(1,9)=11.0$, $p<.01$). Stimuli occurring in the attended field elicited larger N190s than stimuli in the unattended field. However, the main effect of attention is qualified by several two- and three-way interactions. As can be seen from inspection of figures 6, 7 and 8, the attention effect is largest at the contralateral sites, somewhat smaller at Oz and absent at the ipsilateral sites ($F(2,18)=9.3$, $p<.05$). The two-way interaction between task level and attention was also significant ($F(2,18)=6.1$, $p<.05$). Post-hoc tests indicated that the introduction of the monitoring task decreased both the overall amplitudes of the N190's as well as the difference between attended and ignored stimuli. These effects as well as the interaction among electrode site, attention and task level

($F(4,36)=10.2$, $p<.05$) are illustrated by the grand mean N190 amplitude values presented in Table 4. Post-hoc comparisons indicated that although the amplitude of the N190s elicited by the attended stimuli were reduced for the dual task conditions at the contralateral site, the difference between attended and ignored conditions was significant for both the single and dual tasks. For the Oz site, the N190s were reduced in amplitude in the dual task conditions and the attention effect was no longer significant. For the ipsilateral site, the introduction of the monitoring task significantly decreased the amplitude of the N190s. The attention effect was nonsignificant in single and dual task conditions.

This pattern of results would appear to indicate that the increased demands imposed upon the subjects in the dual task conditions had their major impact on the amplitude of the N190s elicited by the attended stimuli. A secondary effect of the task demands was to diminish the differences between the attended and ignored stimuli for the contralateral and Oz sites. Thus, it appears that the selective attention effect reflected by the N190 is sensitive to inter as well as intra-task demands. These results contrast with the insensitivity of the N160 to inter-task demands thereby suggesting that N160 and N190 reflect different aspects of early selective attention.

Insert Table 4 About Here

The interaction among retinal eccentricity, attention and electrode ($F(2,18)=5.9$, $<.05$) is illustrated in Table 3. Post-hoc comparisons indicated that the differences between attended and ignored conditions were significant for the middle and far retinal eccentricities but not for the close position at the Oz site. For the contralateral site, the attention effect was significant at each of the retinal eccentricities. There was also a significant increase in the amplitude of the attended N190s and a decrease in the unattended N190s as a function of eccentricity for both the Oz and contralateral sites. Neither retinal eccentricity nor attention had significant effects on the N190s elicited at the ipsilateral sites. These results are interesting in that, for the most part, they are quite consistent with the changes in the N160 component. However, there is one notable exception. The increased attention effect with retinal eccentricity seems to be due to an increase in the N190s to attended field stimuli as well as a decrease in N190s elicited by stimuli in the unattended field. This contrasts with the N160 results in which the unattended N160s did not change as a function of retinal eccentricity. Thus, the N190 may reflect both enhanced processing of the attended stimuli as well as the diminished processing of the unattended stimuli while the N160 may index only the former process.

Other significant effects included electrode ($F(2,18)=6.9$, $p<.05$) and stimulus type ($F(1,9)=9.6$, $p<.05$). The amplitude of the N190 was larger at the contralateral than the ipsilateral or Oz sites. Targets elicited larger N190s than standards. The stimulus type variable did not interact with any of the other experimental factors.

N190 latencies were significantly longer when recorded contralateral to stimulus presentation than they were when recorded at the ipsilateral or Oz sites ($F(2,18)=5.3$, $p<.05$). The latency of the N190 component increased from the close and middle to the far retinal eccentricity ($F(2,18)=5.9$, $p<.05$; 183, 182 and 202 msec, respectively). The N190s elicited by the targets were significantly later than those elicited by the standards ($F(1,9)=16.0$, $p<.01$; 200 versus 177 msec). The interaction between stimulus type and task was also significant ($F(2,18)=7.6$, $p<.05$), indicating that for the target stimuli N190s elicited in the single task conditions were earlier than those elicited in the dual task conditions (181, 207 and 203 msec). The pattern of latencies obtained for the target N190s is consistent with the pattern of RTs recorded in the arrow discrimination task (see Table 2). Thus, it appears that by 180 msec post-stimulus, evidence of the effects of foveal task load on the processing of stimuli in the periphery has been provided by both the amplitude and the latency of the N190. It is interesting to note that the difference between single and dual task RTs is 66 msec while the difference in N190 latencies is 22 msec. These results suggest that although some of the increase in processing time can be accounted for by the attentional processes reflected in the N190, an additional 66% of the temporal cost attributed to the imposition of the monitoring task occurs in processes following the discrimination between attended and unattended events.

As suggested by Rugg et al. (1987), an important control for subjects' eye fixation during the performance of the arrow discrimination task is the magnitude of asymmetries in the latencies of the ERP components elicited by attended and unattended events. If subjects' fixations deviated towards the attended field it would be expected that the ipsilateral-contralateral differences would be larger for the unattended events due to the greater distance between fixation and stimulus presentation. Thus, if subjects had modified their direction of gaze as a function of the relevance of the stimuli in a visual field, a significant interaction between attention and electrode would be expected. This interaction was not significant for the N190 component.

Insert Figure 9 About Here

Late Selection and Processing Demands - P300 Component

The late positive component occurring between 300 and 800 msec post-stimulus was larger for the low probability targets than the high probability standards (compare figure 5 with figure 4) and increased in amplitude from the frontal to the parietal site with decreases in amplitude from Pz to Oz (see figure 9). It can also be seen in figures 2 through 5 that the late positivity elicited by the attended events is larger than the positivity elicited by the stimuli in the unattended visual field. The latency range, scalp distribution and sensitivity to manipulations of probability and attention are consistent with the criteria employed in the definition of the P300 component (Pritchard, 1981; Sutton and Ruchkin, 1984).

The P300 component was quantified by computing base to peak measures for each of the single and dual task conditions. These values were then submitted to six-way repeated measures ANOVAs (10 subjects x 3 task levels x 3 retinal eccentricities x 2 attention conditions x 2 stimulus types x electrodes). Separate ANOVAs were performed for the midline and lateral electrodes for the measures of P300 amplitude and latency.

Insert Tables 5 and 6 About Here

Midline Sites Consistent with the waveforms presented in figures 2 through 8, P300 amplitude was significantly larger for the attended than the unattended field stimuli ($F(1,9)=106.8$, $p<.01$). Of more importance, however, was the interaction between stimulus type and attention ($F(1,9)=62.4$, $p<.01$). Post-hoc tests indicated that P300s not only discriminated between attended and unattended visual fields but also between targets and standards within the attended field. Thus, while the early negativities reflected processes of interlocation selectivity, the P300 provided an index of both inter and intralocation selectivity. The interaction of this effect with electrode site is illustrated in Table 5 ($F(3,27)=12.6$, $<.01$). The differences between P300s elicited by target and standard stimuli in the attended and unattended fields is largest for Pz, somewhat smaller for the Cz and Oz sites, and smallest for Fz.

Table 6 presents the grand average P300 amplitudes for the attended and ignored stimuli at the three retinal eccentricities collapsed over targets and standards ($F(2,18)=5.8$, $p<.05$). This pattern of results is similar to that obtained for the N160 and N190 components. Post-hoc comparisons indicated that the amplitude of the attended field P300s increased with increased retinal eccentricity while the amplitudes of the unattended field P300s diminished. It is interesting to note, however, that for the P300 the attention effects are significant at each of the eccentricities. For the N160 and the N190 at Oz, attention effects were significant only for the middle and far positions. These results suggest that although attentional selectivity between the close positions may not have occurred early in processing, evidence for such a discrimination was provided by 400 msec post-stimulus.

Consistent with previous research (Isreal et al., 1980a; Kramer et al., 1983, 1985, Kramer, Sirevaag and Braune, 1987) the increased processing demands imposed upon the subjects by the introduction of the monitoring task and the increase in its difficulty from the correlated to the independent conditions had a significant effect on the amplitude of the P300s elicited in the arrow discrimination task ($F(2,18)=22.8$, $p<.01$). This decrement in P300 amplitude with increased processing demands is illustrated for the attended target stimuli in Figure 10. The amplitude of the P300s elicited in the arrow discrimination task decreased with the introduction of the monitoring task. Increases in the difficulty of the monitoring task served to further attenuate the amplitude of the P300. The decrease in P300 amplitude with processing demands was largest at the Pz site ($F(6,54)=6.5$, $p<.05$).

Insert Figure 10 and Table 7 About Here

Of particular interest in the present study was the influence of intertask demands on the intratask selective attention effect. Table 7 presents the mean P300 amplitude values for the significant attention x task interaction ($F(2,18)=5.9$, $p<.05$). As can be seen from the table, the imposition of the monitoring task resulted in a decrease in the amplitude of the P300s elicited by both attended and unattended field stimuli. However, increased monitoring demands within the dual task conditions resulted in decreases in the amplitude of the P300s elicited by the attended but not the unattended field stimuli. Thus, the reduction in the attention effect with increased intertask demands was more than twice as large for the attended field P300s than it was for the unattended field P300s (3.6 versus 1.4 microvolts). It is interesting to note that although the P300 is largest in amplitude for the target stimuli in the attended field, the reduced attention effect with increased task demands is not specific to the intralocation distinction. Instead the amplitude of P300s elicited by all events within the attended field are reduced with increased intertask demands.

P300 latencies were longer for the target stimuli than they were for the standards ($F(1,9)=27.5$, $p<.01$). The latency of the P300 also increased as a function of task level ($F(2,18)=7.9$, $p<.05$; 435, 466 amd 487 msec). Although the interaction between task level and retinal eccentricity was only marginally significant when using conservative degrees of freedom ($F(4,36)=4.7$, $p<.06$), the pattern of P300 latencies was quite similar to that obtained for RT. The differences in P300 latencies as a function of retinal eccentricity were attenuated in the dual task conditions.

Lateral Sites Consistent with the results obtained at the midline sites, P300 was larger for the attended than the unattended field events ($F(1,9)=55.4$, $p<.01$). Furthermore, the attention effect interacted with stimulus type ($F(1,9)=56.4$, $p<.01$), indicating that P300s were largest for the target stimuli in the attended visual field (see Table 5). Thus, consistent with the present results for the midline sites as well as previous findings, P300 amplitude discriminated between attended and unattended spatial locations as well as target and nontarget items within the attended location.

P300 amplitude decreased from the single task to the correlated dual task conditions ($F(2,18)=13.4$, $p<.01$). Increases in dual task difficulty also produced a decrement in P300 amplitude. The interaction between retinal eccentricity and attention was significant ($F(2,18)=8.5$, $p<.05$). Post-hoc comparisons indicated that the amplitude of attended field P300s increased with eccentricity while the amplitude of the unattended field stimuli diminished. Finally, consistent with results obtained by other investigators a main effect of recording site was obtained (Hillyard and Munte, 1984; Neville and Lawson, 1987). P300s were slightly larger (.34 microvolts) when recorded ipsilateral to stimulus presentation ($F(1,9)=7.4$, $p<.05$).

P300 latencies were longer for the targets than they were for the standards ($F(1,9)=65.6$, $p<.01$). The main effect of retinal eccentricity ($F(2,18)=6.9$, $p<.05$) as well as the interaction between eccentricity and task level ($F(4,36)=5.2$, $p<.05$) attained statistical significance. As can be seen in Table 2, the latency of the P300 increased with retinal eccentricity to a greater extent in the single than in the dual task conditions. Finally, P300s elicited by stimuli ipsilateral to the recording site were significantly eariler than those obtained at contralateral sites ($F(1,9)=5.8$, $p<.05$; 443 versus 457 msec).

Discussion

A major goal of the present study was the examination of the effects of intra- and intertask demands on ERP manifestations of attentional processes. However, in order to investigate the sensitivity of ERP components to attentional manipulations we were first required to provide evidence for resource tradeoffs. The pattern of RT and accuracy results obtained for the arrow discrimination task suggested that the demands imposed upon the subjects by the foveally presented triangle monitoring task influenced subjects' performance in the peripheral task. RT increased while accuracy decreased with the introduction of the monitoring task. Furthermore, the increase in the difficulty of the monitoring task produced additional decrements in the subjects' performance on the arrow discrimination task. With respect to resource models of attentional allocation, these results can be interpreted in terms of the withdrawal of resources from the arrow discrimination task to cope with the processing demands of the monitoring task (Freidman and Polson, 1981; Navon and Gopher, 1979; Sanders, 1979; Wickens, 1980). Thus, some proportion of the resources that had been used in the performance of the arrow discrimination task when performed alone were reallocated to the monitoring task in the dual task conditions. Additional resources needed to cope with increased demands were withdrawn from the arrow discrimination task when subjects were required to perform the uncorrelated version of the monitoring task.

The pattern of ERP results indicated that the amplitude of the N160 recorded at the central and parietal sites was insensitive to intertask demands, while the amplitude of the occipital N190 declined with both the introduction of the foveal task and an increase in its difficulty. These findings are interesting in several respects. First, they suggest that, consistent with previous arguements, the N100s recorded at central-parietal and occipital sites differ in their sensitivity to attentional manipulations both within and across tasks (Harter and Aine, 1984; Van Voorhis and Hillyard, 1977). The N160 appears to reflect the distribution of attention to different spatial locations within a task while the N190 is sensitive to both intra- and intertask demands, and therefore may reflect the distribution of general purpose perceptual resources. Second, the finding that N190 decreased as a function of intertask demands indicates that resource models which heretofor have been silent on the temporal locus of tradeoffs between tasks will need to address this issue (Freidman and Polson, 1981; Herdman and Freidman, 1985; Kinsbourne and Hicks, 1978).

One solution for these models would be to incorporate a limited resource attentional filter (Treisman, 1969). In the case of dual tasks, however, the resources would need to be shared not only among different elements within a task but also between tasks. Additional support for such a filter is provided by the task level x attention interaction. The decrese in the interlocation attention effect in the dual task conditions was primarily due to a reduced amplitude for the attended field N190s, suggesting that attention was withdrawn from the processing of items in the cued location to cope with the demands of the monitoring tasks. This pattern of results suggests that the attentional mechanism reflected by the N190 was capable of strategically allocating resources on the basis of processing priorities both within and across tasks.

It is interesting to note that one resource model in particular does address the issue of the temporal locus of resource tradeoffs between tasks. Wickens (1980) proposed a multiple resource model according to which attentional resources may be represented by three dimensions: stages of processing, codes of processing and modalities of processing. The stage dimension is further subdivided into perceptual/central and response resources. Thus, two tasks may compete for attentional resources at either of these stages. P300s have previously been used to indicate a demand for perceptual/central resources (Isreal et al., 1980b). Manipulation of response related demands have only a minimal effect on the amplitude and latency of the P300 (Isreal et al., 1980a, Magliero, Bashore, Coles and Donchin, 1984; McCarthy and Donchin, 1981; Ragot, 1984). The sensitivity of the N190 to intertask demands suggests that tradeoffs between tasks may occur substantially earlier than P300 and therefore may indicate the need for a further subdivision of the stage dimension (eg. perceptual, central and response demands). Thus, it appears that the N190 may index the allocation of perceptual resources while the P300 may provide a metric of a combination of perceptual and central processing resources.

Previous studies that have investigated the sensitivity of P300 to attentional tradeoffs between tasks have focused primarily on the P300s elicited by target stimuli (Kramer et al., 1983; Natani and Gomer, 1981; Lindholm et al., 1984; Sirevaag et al., 1984). In the present study we also examined the effects of dual task demands on the intra- and interlocation selective processes reflected by P300. Consistent with previous findings, a main effect for task level was obtained. The amplitude of the P300s elicited in the peripheral task decreased with the imposition of the monitoring task as well as with an increase in its difficulty from the correlated to the uncorrelated system. However, the manner in which P300s elicited by events in the attended and ignored locations were affected differed as a function of the level of dual task demands. The imposition of the monitoring task resulted in a decrease in the amplitude of P300s elicited by both attended and unattended field stimuli while only the P300s elicited by events in the attended field were influenced by increased dual task demands. These results may indicate that the pace of the peripheral task may not have been sufficient (ISI = 1.2 to 1.5 sec) to require tightly focused attention in the single task condition. However, with the imposition of the foveal task it was necessary for subjects to greatly reduce the attention they allocated to the unattended field stimuli.

It is interesting to note that although this pattern of results would imply that subjects were capable of strategically altering their distribution of attention to accomodate increased processing demands, the lack of a significant stimulus type x task level x attention interaction suggests that their flexibility may have been limited to interlocation selective processes. Since P300s are larger not only for attended field items but also for target items within an attended location, as they were in the single task condition in the present experiment, it might be expected that subjects would reallocate resources that had been used to process nontarget items in the attended field prior to reducing their attention to target items. Within Broadbentian (1982) nomenclature, it appears that although subjects were able to filter successfully on the basis of stimulus set (eg. location) to cope with external task demands, they did not limit their attention to specific items within an attended location. Future research will be necessary to determine if such hierarchical filtering is a viable strategy in other multi-task situations.

The Zoom Lens model has been proposed to account for differences in the speed and precision of processing during focused and divided attention in the visual modality (Eriksen and Saint James, 1986). It has been argued that a fixed quantity of attentional resources is available for the processing of items in the visual field and that the performance differences observed in focused and divided attention can be attributed to the density of resources allocated to task relevant events. Although the model was specifically designed to account for the distribution of attentional resources within a task, the results of the present study suggest that it might apply equally well to divided attention between tasks. The N160, N190 and P300 discriminated between cued and uncued visual fields during focused attention. The requirement to divide attention between two different tasks reduced the amplitude of the N190s and P300s elicited by items in the attended field. Performance also decreased from the focused to the divided attention conditions. This pattern of results is consistent with the prediction that a reduction in the magnitude of resources allocated to the processing of visual events, as indicated by the amplitude of the N190 and P300, results in a decrement in performance. The present results suggest that this prediction is fulfilled when attention is distributed between as well as within tasks.

Previous studies have found that N100s elicited by stimuli lateral to the attended location progressively decrease in amplitude with increasing distance (Mangun and Hillyard, 1987). These ERP results are consistent with other studies that have found increased RTs and decreased accuracies as unexpected stimuli that require a response are presented at increasing distances from an attended location (Downing and Pinker, 1985; Eriksen and Saint James, 1986; LaBerge and Brown, 1986). Taken together, the behavioral and ERP studies suggest that the spatial filter functions with a gradual dropoff of processing resources with increasing distance from the focus of attention. The results of the present study are consistent with this conception of the visual-spatial filter. The differences in the N160, N190 and P300's elicited by stimuli in the attended and unattended fields were magnified with increases in the spatial separation of the stimuli from 2.8 to 18.2 degrees.

There has been some controversy as to the effects of foveal task difficulty on the processing of information in the visual periphery. Proponents of the Cognitive Tunnel Vision model assert that the functional visual field constricts with increasing foveal task demands (Ikeda and Takcuchi, 1975; Mackworth, 1965; Williams, 1985). In terms of a joint ERP/performance analysis this would suggest that (a) with increases in the difficulty of the foveal task, performance in the peripheral should decrease in proportion to the distance of the stimuli from the fovea and, (b) the N160, N190 and P300 attentional gradient effects should interact with the difficulty of the foveal task such that the difference between the components elicited by the attended and unattended field stimuli should dropoff more quickly at the distant locations. Neither of these effects were obtained in the present experiment. Thus, our results provide support for the counter arguement that asserts that stimuli in the visual periphery suffer from the division of attention between a foveal and peripheral task, irrespective of their physical distance from the fovea. There were, however, several methodological differences between the present experiment and those that have obtained support for the Cognitive Tunnel Vision model. These differences included the use of a continuous rather than a discrete foveal task and a substantially longer presentation of peripheral stimuli (50 versus 10 msec). Although, given the differences in methodologies it is impossible for us assert that Cognitive Tunnel vision does not occur, the necessity for either of these conditions would appear to greatly limit the generalizability of the phenomenon.

References

Alho, K., Sams, M., Paavilainen, P. and Naatanen, R. (1986). Small pitch separation and the selective attention effect on the ERP. Psychophysiology, 23, 189-197.

Antes, J. and Edwards, D. (1973). Information processing in the visual periphery. Bulletin of the Psychonomic Society, 1, 351-353.

Bashinski, H. and Bacharach, V. (1980). Enhancement of perceptual sensitivity as the result of selectively attending to spatial locations. Perception and Psychophysics, 28, 241-248.

Beck, J. and Ambler, B. (1973). The effects of concentrated and distributed attention on peripheral acuity. Perception and Psychophysics, 14, 225-230.

Becklen, R. and Cervone, D. (1983). Selective looking and the noticing of unexpected events. Memory and Cognition, 11, 601-608.

Broadbent, D. (1982). Task combination and selective intake of information. Acta Psychologia, 50, 253-290.

Broadbent, D. (1958). Perception and Communication. London: Pergamon.

Casey, E. and Wickens, C. (1986). Visual display representation of multidimensional systems: The effect of information correlation and display integrality. Technical Report CPL-86-2. Cognitive Psychophysiology Laboratory. University of Illinois.

Cherry, C. (1953). Some experiments on the perception of speech with one and with two ears. Journal of the Acoustical Society of America, 25, 975-979.

Dark, V., Johnston, W., Myles-Worsley, M. and Farah, M. (1985). Levels of selection and capacity limits. Journal of Experimental Psychology: General, 114, 472-497.

Deutsch, J. and Deutsch, D. (1963). Attention: Some theoretical considerations. Psychological Review, 70, 80-90.

Donchin, E., Kramer, A. and Wickens, C. (1986). Applications of brain event-related potentials to problems in engineering psychology. In. M.G.H. Coles, S. Porges and E.Donchin (Eds.), Psychophysiology: Systems, Processes and Applications. New York: Guilford Press.

Downing, C. and Pinker, S. (1985). The spatial structure of visual attention. In M. Posner and O. Marin (Eds.), Attention and Performance XI. Hillsdale, N.J.: Erlbaum.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. Psychological Review, 87, 272-300.

Eason, R., Harter, R., and White, C. (1969). Effects of attention and arousal on visually evoked cortical potentials and reaction time in man. Physiology and Behavior, 4, 283-289.

Eason, R. and Ritchie, G. (1976). Effects of stimulus set on early and late components of visually evoked potentials. Paper presented at the 17th Annual Meeting of the Psychonomic Society. St. Louis, MO.

Eriksen, C. and Saint James, J. (1986). Visual attention within and around the field of focal attention: A zoom lens model. Perception and Psychophysics, 40, 225-240.

Eriksen, C. and Yeh, Y. (1985). Allocation of attention in the visual field. Journal of Experimental Psychology: Human Perception and Performance, 11, 583-597.

Francolini, C. and Egeth, H. (1980). On the nonautomaticity of "automatic" activation: Evidence of selective seeing. Perception and Psychophysics, 27, 331-342.

Freidman, A., and Polson, M. (1981). The hemispheres as independent resource systems: Limited capacity and cerebral specialization. Journal of

Experimental Psychology: Human Perception and Performance, 7, 1030-1058.

Greenhouse, S. and Geisser, S. (1959). On methods in the analysis of profile data. Psychometrika, 24, 95-112.

Harter, M., and Aine, C. (1984). Brain mechanisms of visual selective attention. In R. Parasuraman and R. Davies (Eds.), Varieties of Attention. New York: Academic Press.

Harter, R. and Previc, F. (1978). Size specific information channels and selective attention: Visual evoked potential and behavioral measures. Electroencephalography and Clinical Neurophysiology, 45, 628-640.

Heffley, E., Foote, B., Mui, T. and Donchin, E. (1985). PEARL II: Portable laboratory computer system for psychophysiological assessment using event-related brain potentials. Neurobehavioral Toxicology and Tetratology, 7, 399-407.

Herdman, C. and Freidman, A. (1985). Multiple resources in divided attention: A cross modal test of the independence of hemispheric resources. Journal of Experimental Psychology: Human Perception and Performance, 11, 40-49.

Hillyard, S., Hink, R., Schwent, V. and Picton, T. (1973). Electrical signs of selective attention in the human brain. Science, 182, 177-180.

Hillyard, S.A. and Kutas, M. (1983). Electrophysiology of cognitive processing. Annual Review of Psychology, 34, 33-61.

Hillyard, S. and Munte, T. (1984). Selective attention to color and location: An analysis with event-related brain potentials. Perception and Psychophysics, 36, 185-198.

Hillyard, S., Munte, T. and Neville, H. (1985). Visual-spatial attention, orienting, and brain physiology. In M. Posner and O. Marin (Eds.), Attention and Performance XI. Hillsdale, New Jersey: Erlbaum. (p. 63-84).

Hillyard, S., Simson, G., Woods, D., Van Voorhis, S., and Munte, T. (1984). Event-related brain potentials and selective attention to different modalities. In F. Reinoso-Suarez and C. Ajmone-Marsan (Eds.), Cortical Integration. New York: Raven Press.

Hink, R., Van Voorhis, S., Hillyard, S. and Smith, T. (1977). The division of attention and the human auditory evoked potential. Neuropsychologica, 15, 597-605.

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

Holmes, D., Cohen, K., Haith, M. and Morrison, F. (1977). Peripheral visual processing. Perception and Psychophysics, 22, 571-577.

Horst, R., Munson, R., and Ruchkin, D. (1984). Event-related potential indices of workload in a single task paradigm. In Proceedings of the Human Factors Society, 28th Annual Meeting. Santa Monica, California.

Ikeda, M. and Takeuchi, T. (1975). Influence of foveal load on the functional field. Perception and Psychophysics, 18, 255-260.

Isreal, J., Chesney, G., Wickens, C., and Donchin, E. (1980a). P300 and tracking difficulty: Evidence for multiple resources in dual task performance. Psychophysiology, 17, 259-273.

Isreal, J., Wickens, C., Chesney, G. and Donchin, E. (1980b). The event related brain potential as an index of display monitoring workload. Human Factors, 22, 211-224.

Johnston, W. and Dark, V. (1982). In defense of intraperceptual theories of attention. Journal of Experimental Psychology: Human Perception and Performance, 8, 407-421.

Kahneman, D. (1973). Attention and Effort. Englewood Cliffs, New Jersey: Prentice Hall.

Kahneman, D. and Henik, A. (1981). Perceptual organization and attention. In M. Kubovy and J. Pomerantz (Eds.), Perceptual Organization. Hillsdale, N.J.: Lawrence Erlbaum.

Kahneman, D. and Treisman, A. (1984). Changing views of attention and. automaticity. In R. Parasuraman and D. Davies (Eds.), Varieties of Attention. Orlando, Florida: Academic Press.

Keele, S. (1973). Attention and Human Performance. Pacific Palasides, California: Goodyear.

Kinsbourne, M. and Hicks, R. (1978). Functional cerebral space: A model for overflow, transfer and interference effects in human performance. In J. Requin (Ed.), Attention and Performance. Hillsdale, N.J.: Erlbaum.

Kramer, A.F., Sirevaag, E. and Braune, R. (1987). A psychophysiological assessment of operator workload during simulated flight missions. Human Factors, 29, 145-160.

Kramer, A. and Strayer, D. (1987). P300 operating characteristics: Performance/ERP analysis of dual-task demands and attention. In M. Kutas and B. Renault (Eds.), Proceedings of the 4th International Conference on Cognitive Neurosciences. Paris-Dourdan, France.

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

Kramer, A.F., Wickens, C.D., and Donchin, E. (1983). An analysis of the processing demands of a complex perceptual-motor task, Human Factors, 25, 597-622.

Kutas, M. and Hillyard, S. (1984). Event-related potentials in cognitive science. In M. Gazzangia (Ed.), Handbook of Cognitive

Neuroscience. New York: Plenum Press.

LaBerge, D. and Brown, V. (1986). Variations in size of the visual field in which targets are presented: An attentional range effect. Perception and Psychophysics, 40, 188-200.

Lindholm, E. Cheatham, C., Koriath, J. and Longbridge, T. (1984). Physiological assessment of aircraft pilot workload. Technical report AFHRL-tr-83-49. Brooks AFB.

Mackworth, N. (1965). Visual noise causes tunnel vision. Psychonomic Science, 3, 67-68.

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

Mangun, G. and Hillyard, S. (1987). The spatial allocation of visual attention as indexed by event related brain potentials. Human Factors, 29, 195-212.

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

Moray N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. Quarterly Journal of Experimental Psychology, 11, 56-60.

Naatanen, R. (1982). Processing negativity: An evoked potential reflection of selective attention. Psychological Bulletin, 92, 605-640.

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

Navon, D. and Gopher, D. (1979). On the economy of the human processing system. Psychological Review, 86, 214-255.

Neisser, U. and Becklen, R. (1975). Selective looking: Attending to visually specified events. Cognitive Psychology, 7, 480-494.

Neville, H. and Lawson, D. (1987). Attention to central and peripheral visual space in a movement detection task: an event-related potential and behavioral study. I Normal hearing adults. Brain Research, 405, 253-267.

Okita, T. (1979). Event-related potentials and selective attention to auditory stimuli varying in pitch and localization. Biological Psychology, 9, 271-284.

Parasuraman, R. (1978). Auditory evoked potentials and divided attention. Psychophysiology, 15, 460-465.

Parasuraman, R. (1985). Event related brain potentials and intermodal divided attention. In R. Swezey (Ed.), Proceedings of the Human Factors Society, 29th Annual Meeting. Baltimore, Maryland.

Pashler, H. (1984). Evidence against late selection: Stimulus quality effects in previewed displays. Journal of Experimental Psychology: Human Perception and Performance, 10, 429-448.

Pritchard, W. (1981). Psychophysiology of P300. Psychological Bulletin, 89, 506-540.

Picton, T.W., Campbell, K.B., Baribeau-Braun, J. and Proulx, G.B. (1978). The neurophysiology of human attention: A tutorial review. In J. Requin (Ed.), Attention and Performance VII. Hillsdale, New Jersey: Erlbaum Associates. (p. 429-467).

Picton, T., Woods, D., Proulx, G. (1978). Human auditory sustained potentials. II. Stimulus Relationships. Electroencephalography and Clinical Neurophysiology, 45, 198-210.

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

Rugg, M., Milner, C., Lines, C. and Phalp, R. (1987). Modulation of visual event related potentials by spatial and non-spatial visual selective attention. Neuropsychologica, 25, 85-96.

Sanders, A. (1979). Some remarks on mental load. In N. Moray (Ed.), Mental Workload: Its Theory and Measurement. New York: Plenum Press.

Shiffrin, R., McKay, D., and Shafer, W. (1976). Attending to 49 spatial positions at once. Journal of Experimental Psychology: Human Perception and Performance, 2, 14-22.

Shulman, G., Wilson, J. and Sheehy, J. (1985). Spatial determinants of distribution of attention. Perception and Psychophysics, 37, 59-65.

Sirevaag, E., Kramer, A.F., Coles, M.G.H. and Donchin, E. (1984). P300 amplitude and resource allocation. Psychophysiology, 21, 598-599.

Sutton, S. and Ruchkin, D.S. (1984). The late positive component: Advances and new problems. Brain and Information: Event-Related Potentials. New York: New York Academy of Sciences.

Treisman, A. (1969). Strategies and models of selective attention. Psychological Review, 76, 282-299.

Treisman, A. and Riley, R. (1969). Is selective attention selective perception or selective response? A further test. Journal of Experimental Psychology, 79, 27-34.

Van Voorhis, S. and Hillyard, S. (1977). Visual evoked potentials and selective attention to points in space. Perception and Psychophysics, 22, 54-62.

Voss, M. (1981). Narrowing of the visual field as an indicator of mental workload. In J. Morael and K. Kraiss (Eds.), Manned Systems Design. New York: Plenum Press.

Wickens, C.D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), Attention and Performance VIIIHillside, New Jersey:

Erlbaum.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman and R. Davies (Eds.), Varieties of Attention. New York: Academic Press.

Wickens, C.D., Kramer, A., Vanasse, L., and Donchin, E. (1983). Performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information-processing resources. Science, 221, 1080-1082.

Williams, L (1982). Cognitive load and the functional field of view. Human Factors, 24, 683-692.

Williams, L. (1985). Tunnel vision induced by a foveal load manipulation. Human Factors, 27, 221-227.

Woods, D. and Courchesne, E. (1986). The recovery functions of auditory event-related potentials during split second discriminations. Electroencephalography and Clinical Neurophysiology, 65, 304-315.

Woods, D., Courchesne, E., Hillyard, S. and Galambos, R. (1980). Recovery cyclyes of event-related potentials in multiple detection tasks. Electroencephalography and Clinical Neurophysiology, 50, 335-347.

Figure 1 A graphic representation of the triangle monitoring and arrow discrimination tasks. The different versions of the triangle monitoring task are also indicated for both normal and failure states.

Figure 2 Grand average waveforms elicited by the standards in each of the experimental conditions. The waveforms were recorded at the Fz site.

Figure 3 Grand average waveforms elicited by the standards in each of the experimental conditions. The waveforms were recorded at the Cz site.

Figure 4 Grand average waveforms elicited by the standards in each of the experimental conditions. The waveforms were recorded at the Pz site.

Figure 5 Grand average waveforms elicited by the targets in each of the experimental conditions. The waveforms were recorded at the Pz site.

Figure 6 Grand average waveforms elicited by the standards in each of the experimental conditions. The waveforms were recorded at the ipsilateral sites.

Figure 7 Grand average waveforms elicited by the standards in each of the experimental conditions. The waveforms were recorded at the Oz site.

Figure 8 Grand average waveforms elicited by the standards in each of the experimental conditions. The waveforms were recorded at the contralateral sites.

Figure 9 Grand average target waveforms for each of the midline sites in the single and dual task conditions.

Figure 10 Grand average target waveforms for each of the attended field stimuli for each of the task levels and retinal eccentricities.

Footnotes

1. Since the single and dual task monitoring conditions were not directly comparable due to the levels of retinal eccentricity of arrow presentation in the dual task conditions, three separate ANOVA's were performed. Single task monitoring performance was compared to monitoring performance when paired with close, middle and far arrow presentations. Single and dual task measures did not differ in any of these comparisons.
2. All post-hoc comparisons reported in the article are computed with the Bonferroni t-test and are significant at $p<.05$.

Table 1

Average RT (msec), number of hits, misses and FA's for the triangle monitoring task in the dual task conditions for each of the three retinal eccentricities of the arrow stimuli.

Measures	Correlated Condition			Uncorrelated Condition		
	Close	Middle	Far	Close	Middle	Far
Average RT	1720	1634	1837	4356	4194	4339
Average Hits	49	49	50	41	38	40
Average Misses	.6	.4	.5	3.8	3.6	3.0
Average FA's	1.1	1.6	1.6	12.7	13.3	12.5

Table 2

Average RT and P300 latency (in parentheses) in msec obtained at the lateral sites for the arrow discrimination task in both the single and dual task conditions for each of the retinal eccentricities.

	Retinal Eccentricities		
	Close	Middle	Far
Single Task	399 (386)	419 (435)	445 (461)
Dual Correlated	476 (455)	478 (463)	506 (476)
Dual Uncorrelated	494 (460)	492 (471)	506 (511)

Table 3

Grand average N160 amplitude measures in microvolts for Cz and Pz electrode sites and N190 measures at occipital recording sites for the attended and ignored stimuli at the three retinal eccentricities.

	Retinal Eccentricity					
	Close		Middle		Far	
Electrode Site	Attend	Ignore	Attend	Ignore	Attend	Ignore
Cz	1.9	1.7	2.3	1.8	2.6	1.5
Pz	2.0	1.7	2.4	1.9	2.8	1.6
Ipsilateral	2.1	2.0	2.4	2.5	2.3	2.2
Oz	2.3	2.1	2.9	1.9	3.1	1.7
Contralateral	2.8	2.2	3.2	1.9	3.6	1.7

Table 4

Grand average N190 amplitude measures in microvolts for Oz, ipislateral and contralateral sites for the three task levels and the attended and ignored stimuli

Task Level	Ipsilateral		Oz		Contralateral	
	Attend	Ignore	Attend	Ignore	Attend	Ignore
Single	2.7	2.6	3.0	1.9	3.8	1.9
Dual-Correlated	1.7	1.8	1.9	1.7	2.4	1.7
Dual-Independent	1.8	1.8	2.0	1.7	2.5	1.6

Table 5

Grand average P300 amplitude measures in microvolts at the midline and combined ipsi/contralateral sites for stimuli in the attended and nonattended fields.

	Targets		Standards	
Electrode	Attend	Ignore	Attend	Ignore
Fz	7.0	3.1	4.0	2.7
Cz	11.9	3.9	5.1	3.3
Pz	16.7	4.1	5.0	2.9
Oz	9.5	3.8	3.9	2.3
Ipsi/contra	9.4	2.9	2.9	1.9

Table 6

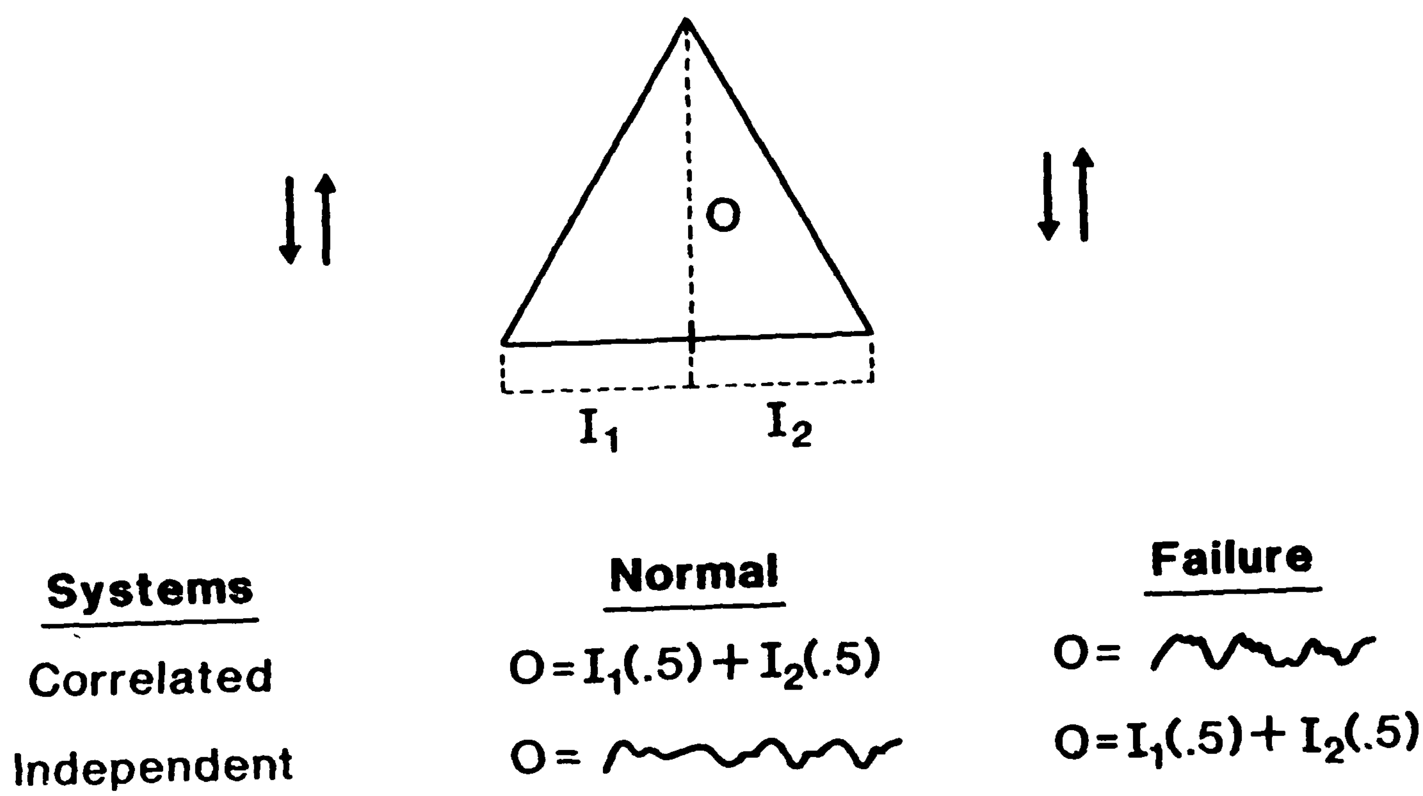
Grand average P300 amplitude measures in microvolts for the attended and ignored field stimuli at each of the retinal eccentricities. The amplitude values are collapsed over targets and standards.

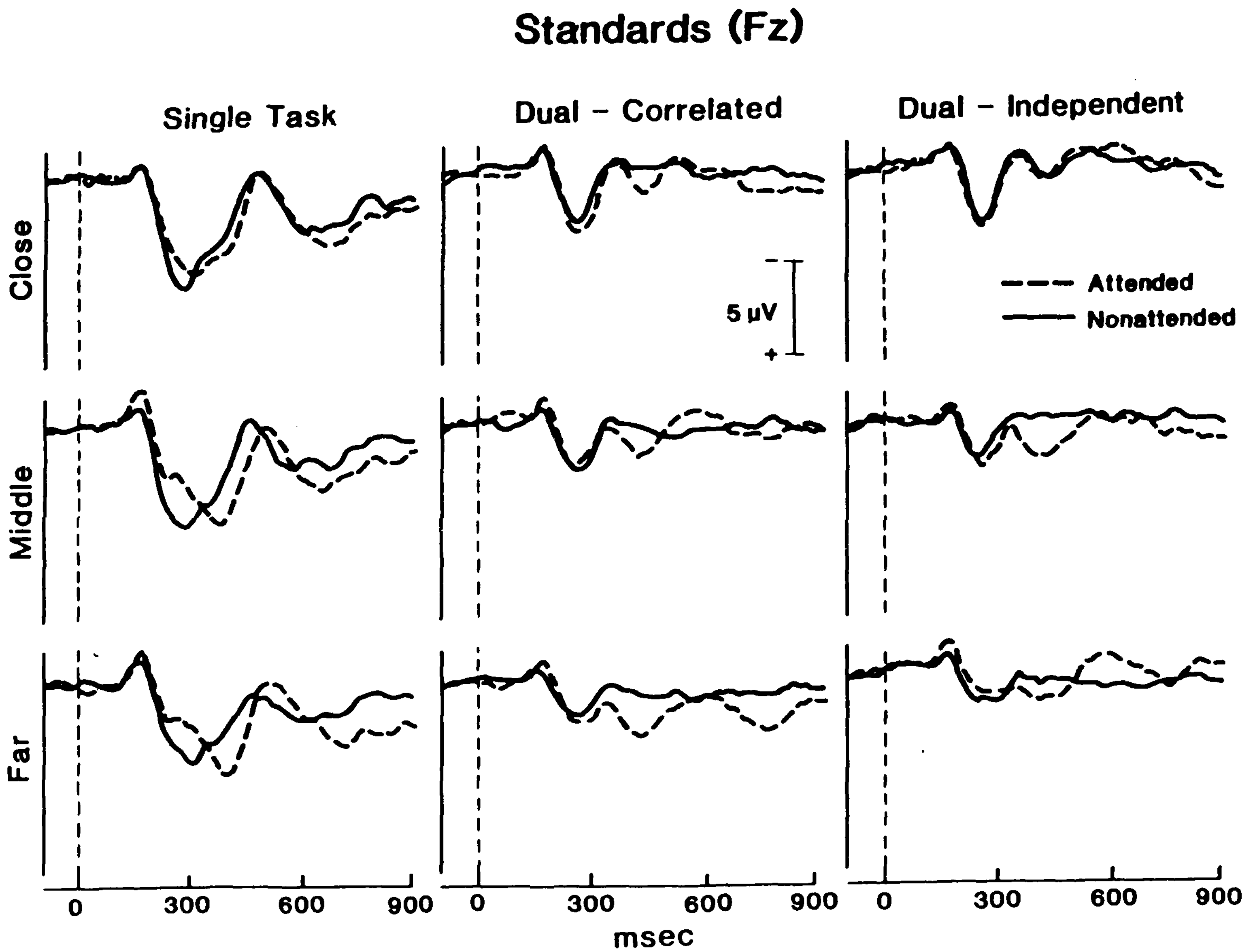
Retinal Eccentricity	Attend	Ignore
Close	7.4	3.5
Middle	7.8	2.7
Far	8.0	2.6

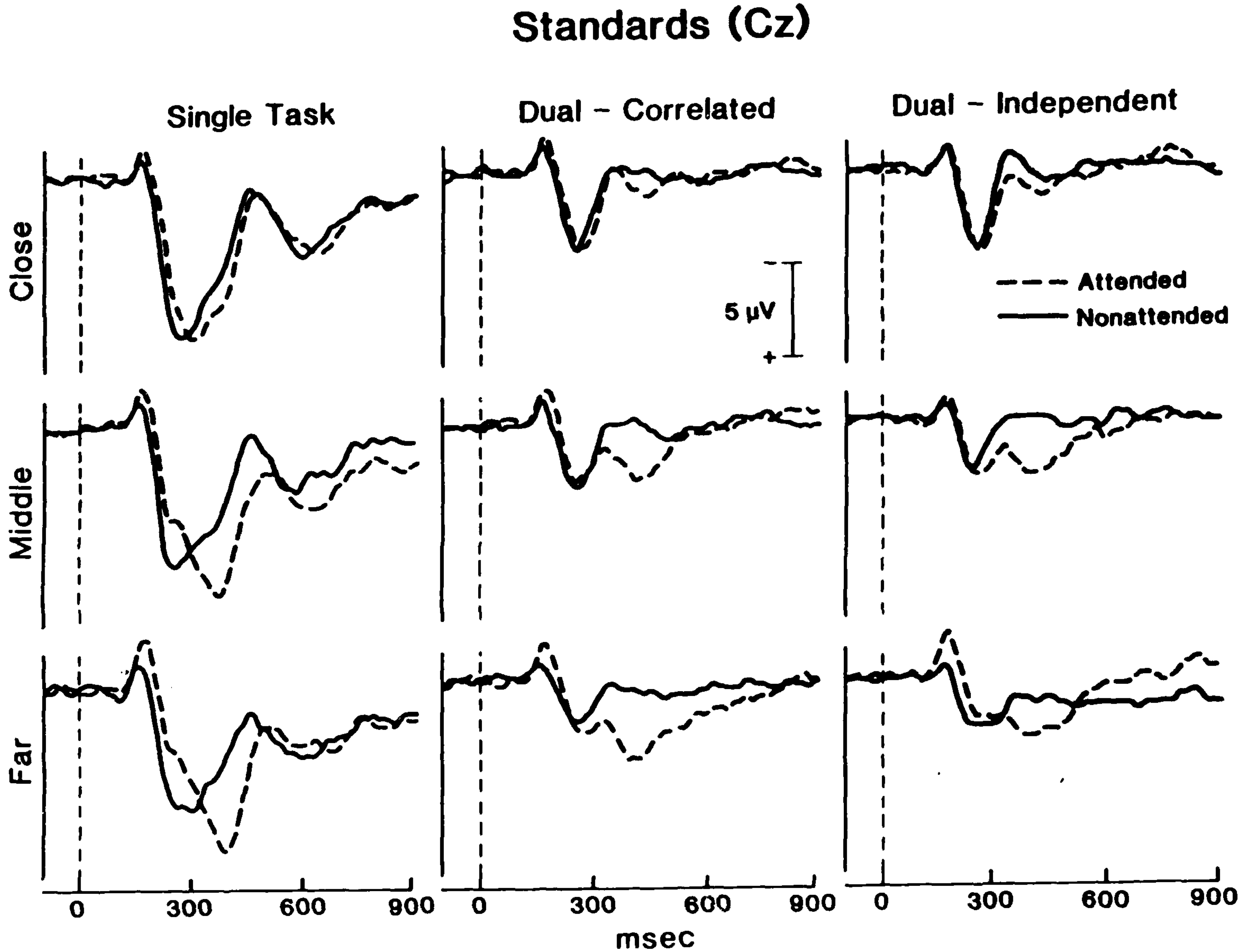
Table 7

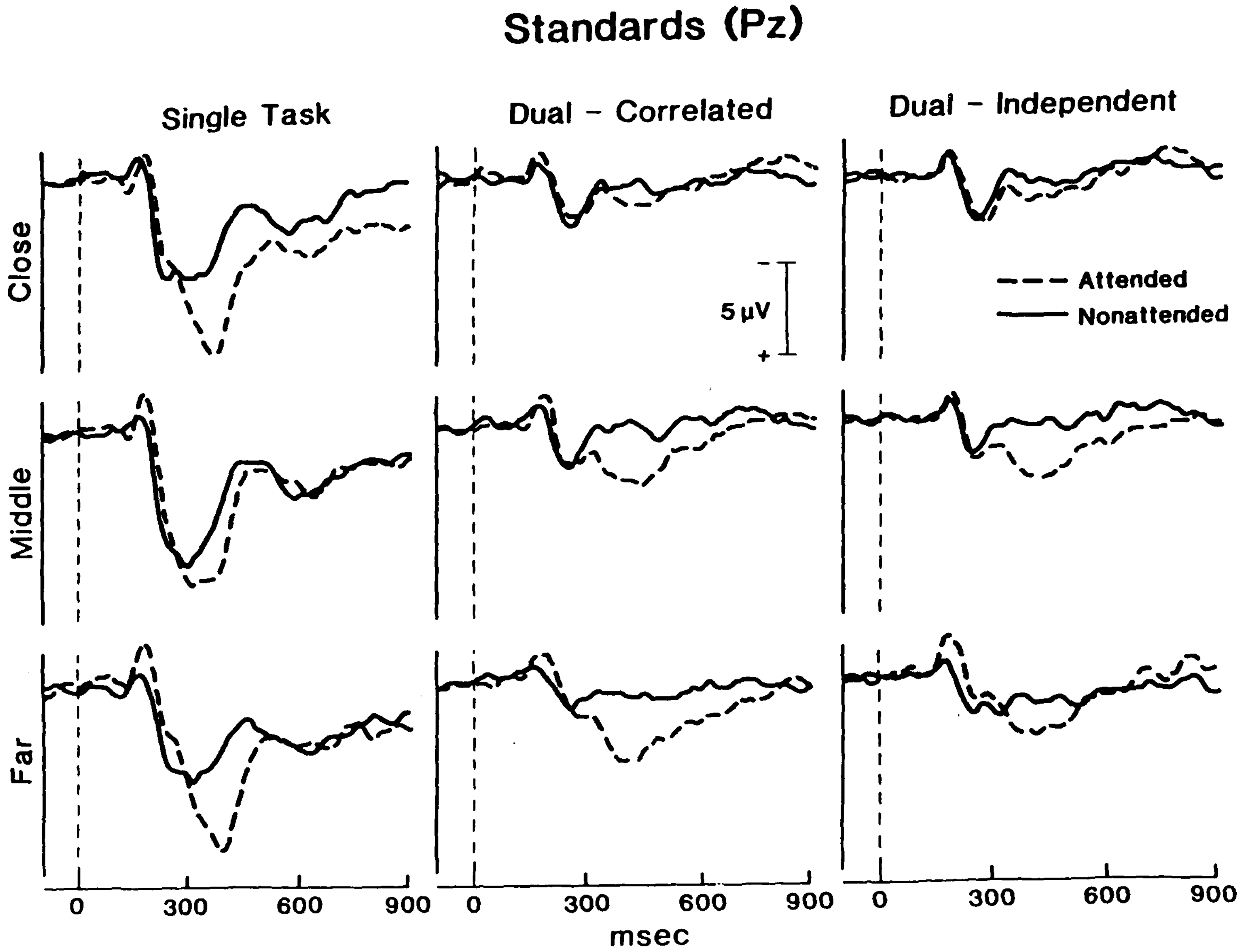
Grand average P300 amplitude measures in microvolts for the attended and ignored field stimuli for each of the task levels. The amplitude values are collapsed over targets and standards.

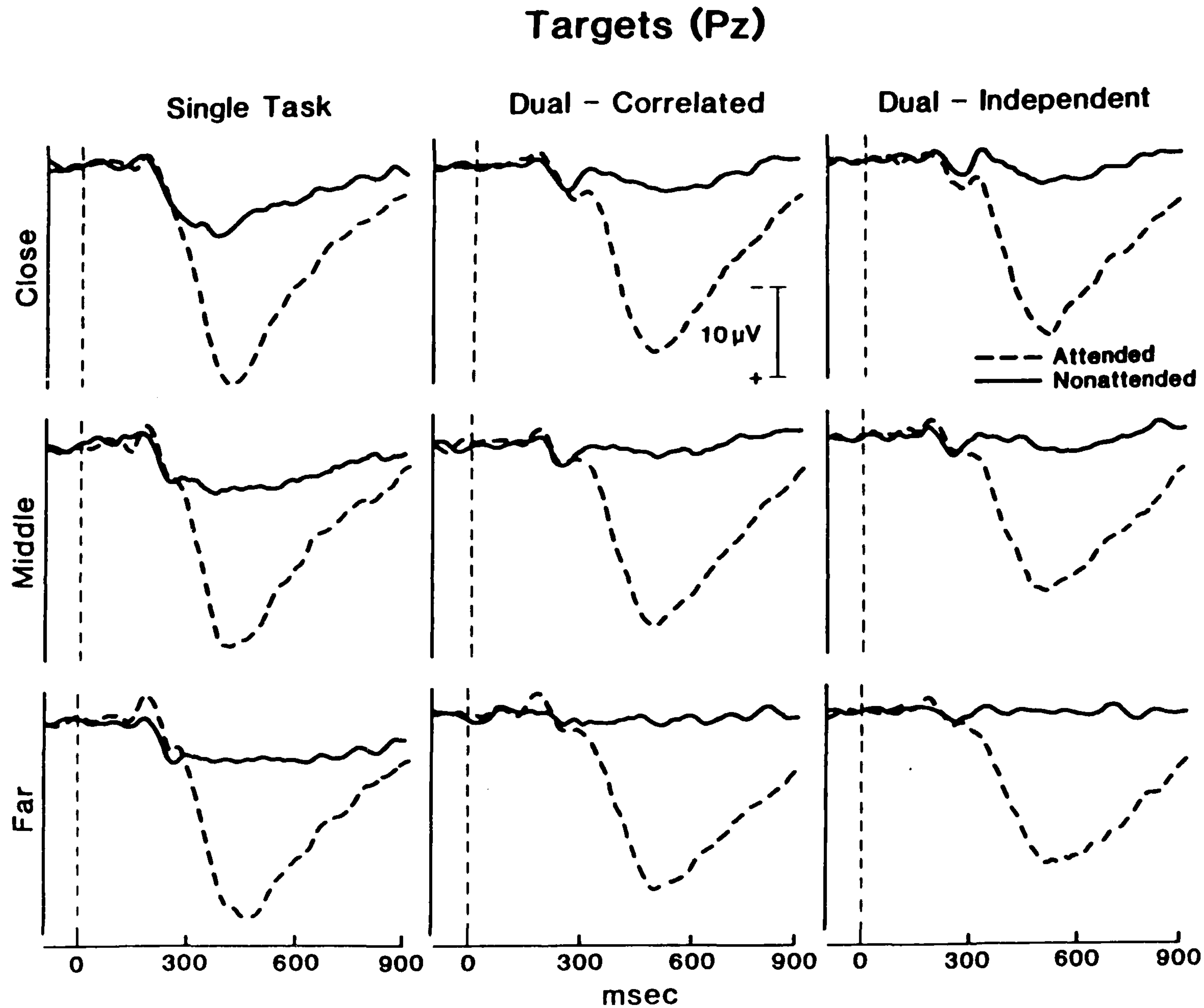
Task Level	Attend	Ignore
Single	11.6	3.5
Dual-Correlated	9.2	2.1
Dual-Independent	8.0	2.1

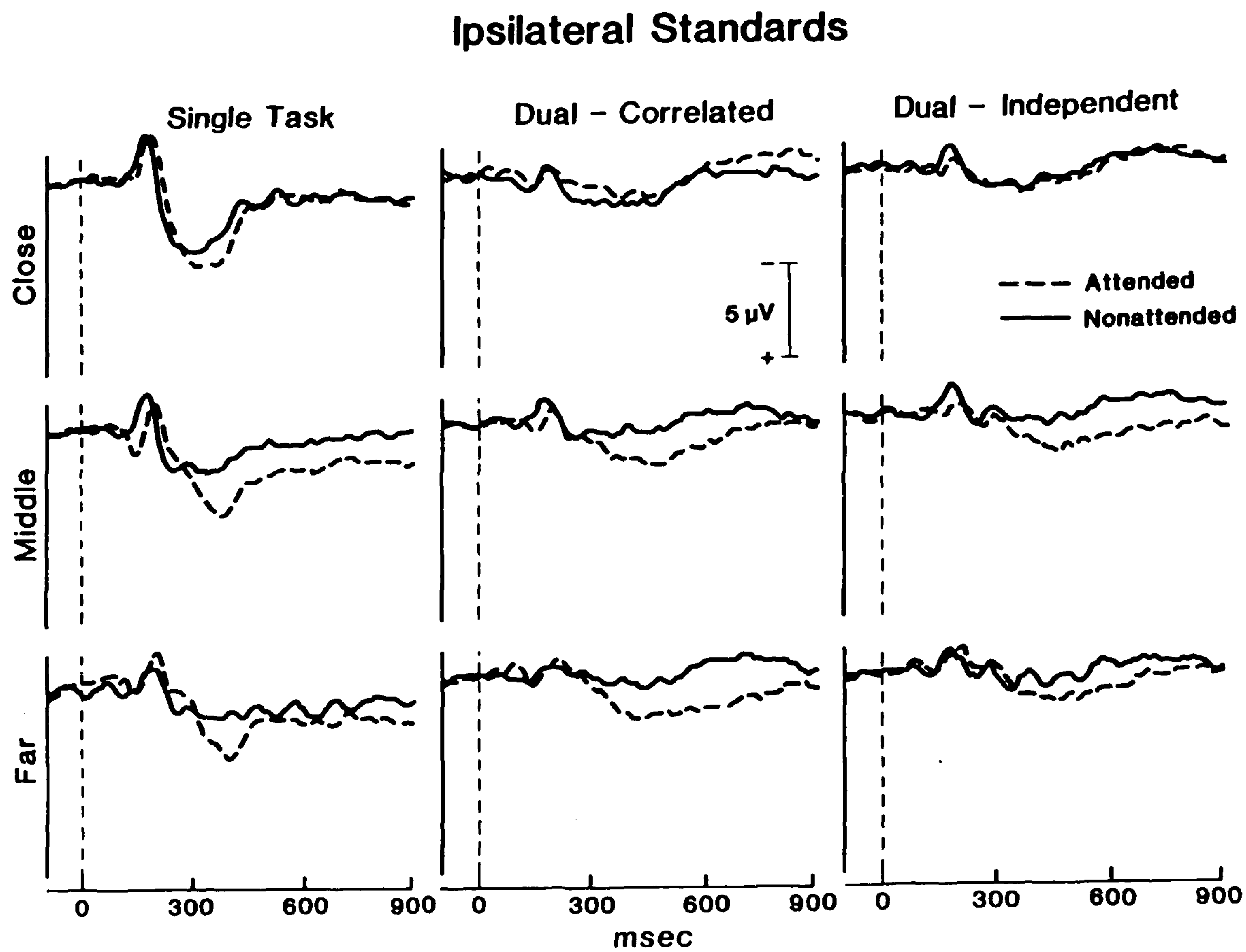

O
I_1
I_2
Systems
Normal
Failure
Correlated
$O = I_1(.5) + I_2(.5)$
O =
Independent
O =
$O = I_1(.5) + I_2(.5)$

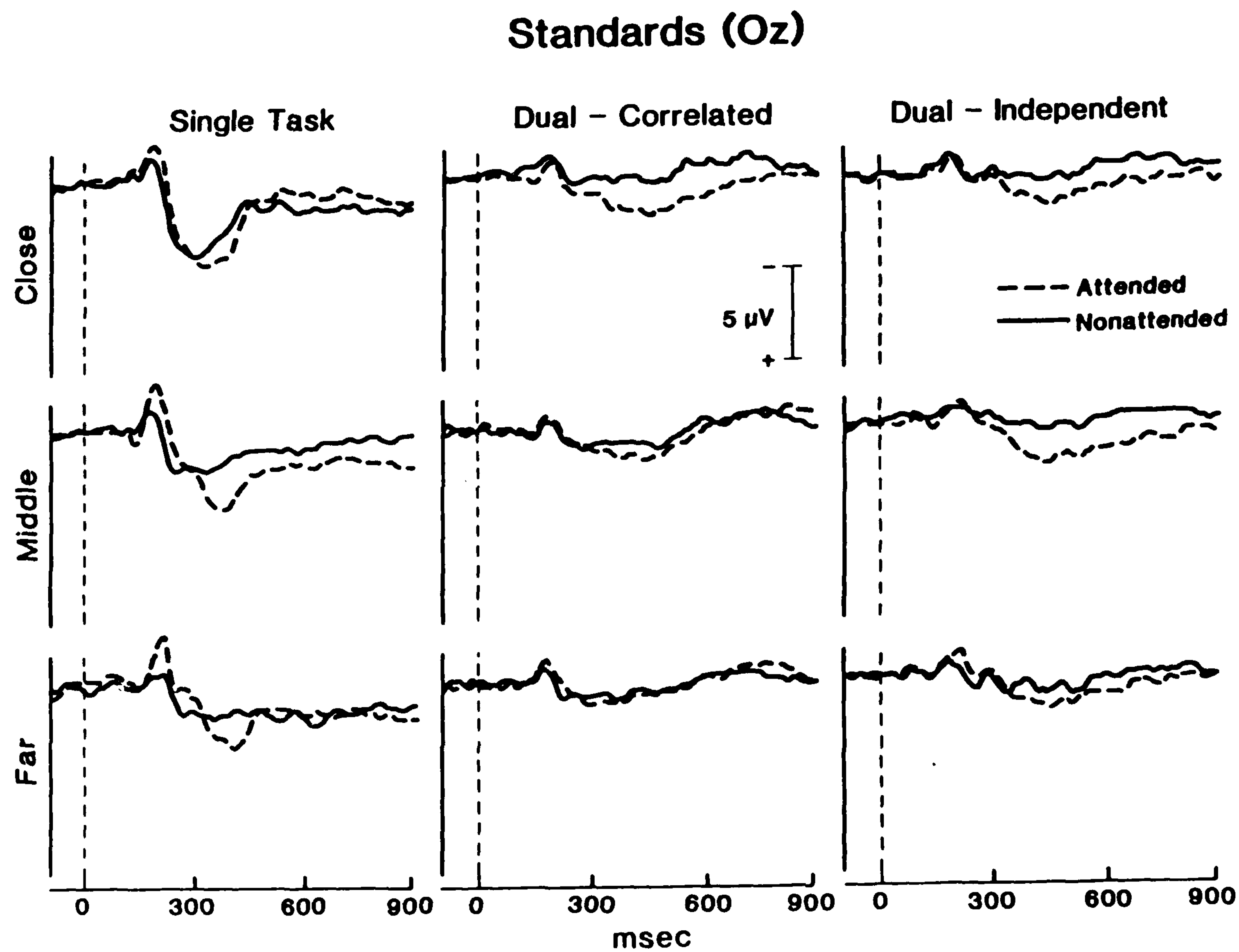

Standards (Fz)
Single Task
Dual - Correlated
Dual - Independent
Close
Middle
Far
5 µV
-
+
Attended
Nonattended
0
300
600
900
msec

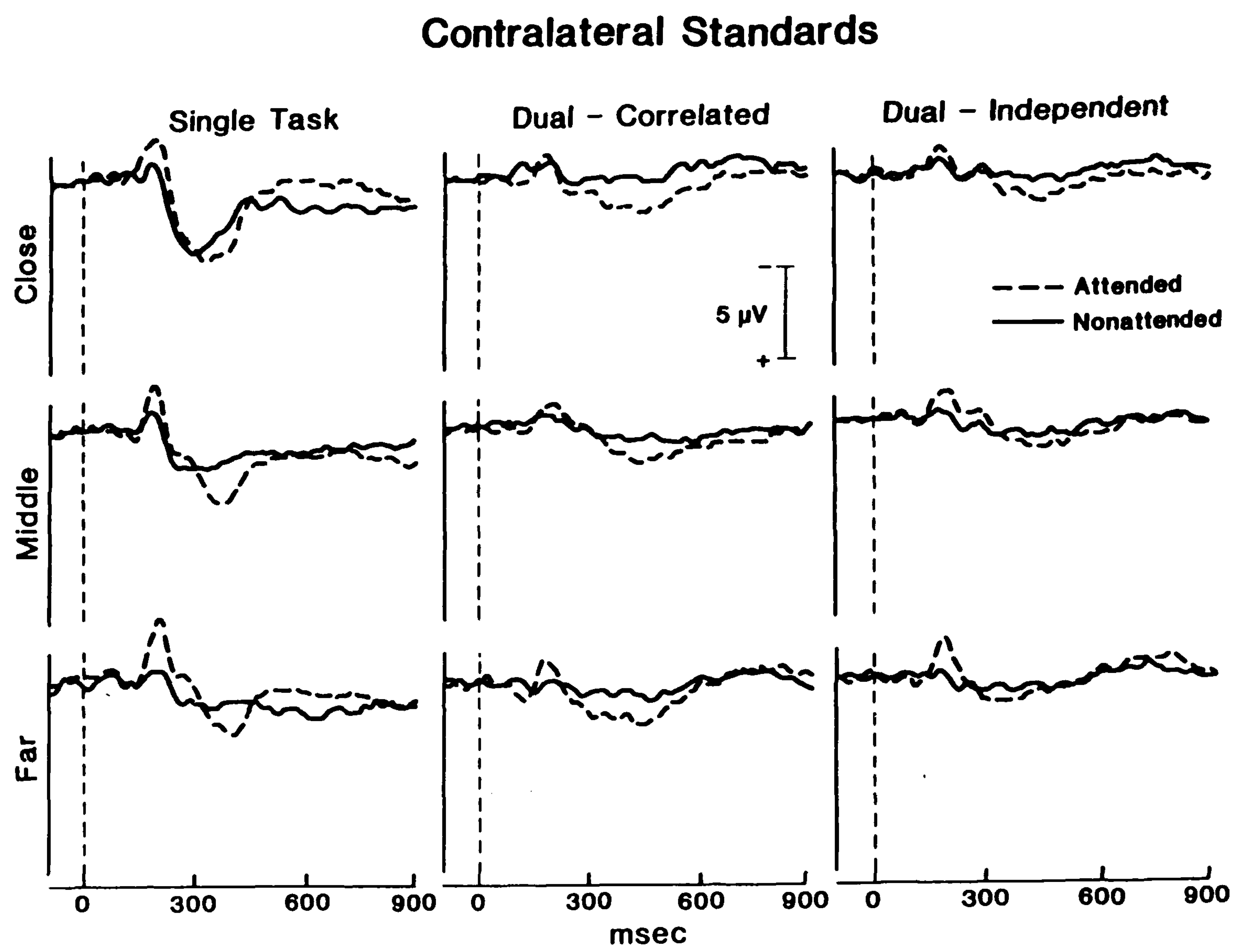

Standards (Cz)
Single Task
Dual - Correlated
Dual - Independent
Close
Middle
Far
5 µV
-
+
Attended
Nonattended
0
300
600
900
msec


Standards (Pz)
Single Task
Dual - Correlated
Dual - Independent
Close
Middle
Far
5 µV
- - - Attended
—— Nonattended
0
300
600
900
msec

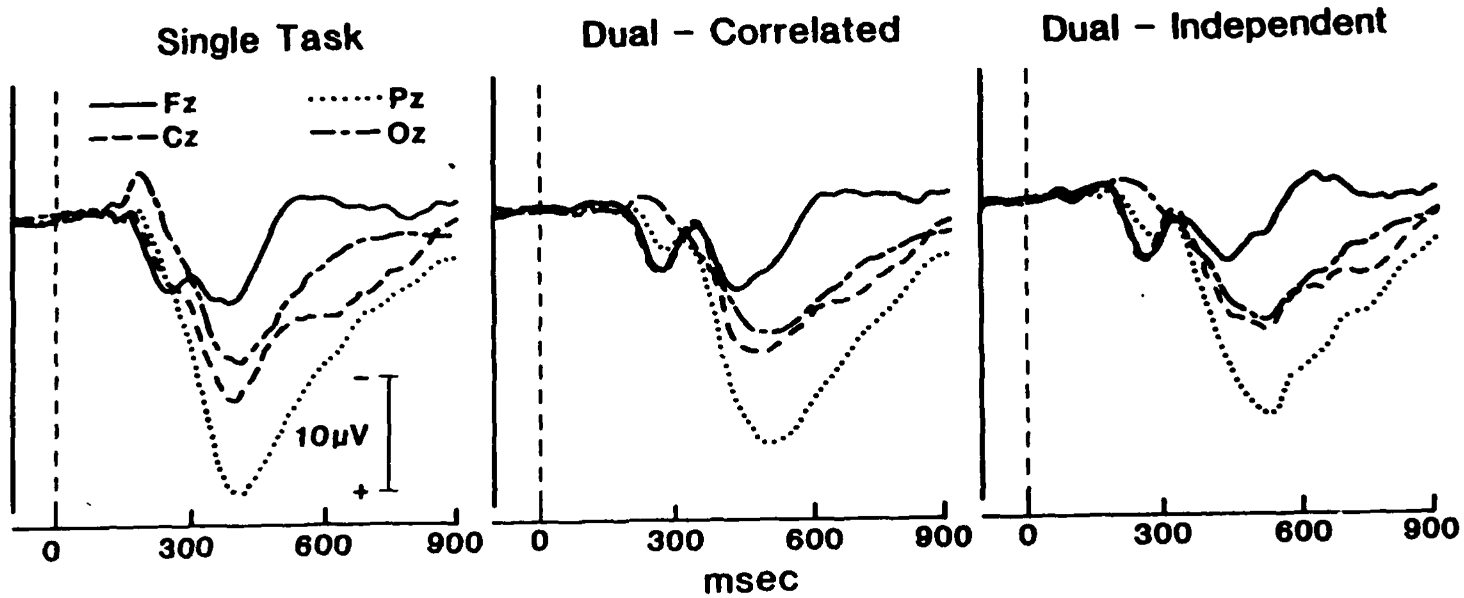

Targets (Pz)
Single Task
Dual - Correlated
Dual - Independent
Close
Middle
Far
10µV
-
+
Attended
Nonattended
0
300
600
900
msec


Ipsilateral Standards
Single Task
Dual - Correlated
Dual - Independent
Close
Middle
Far
5 μV
Attended
Nonattended
0
300
600
900
msec

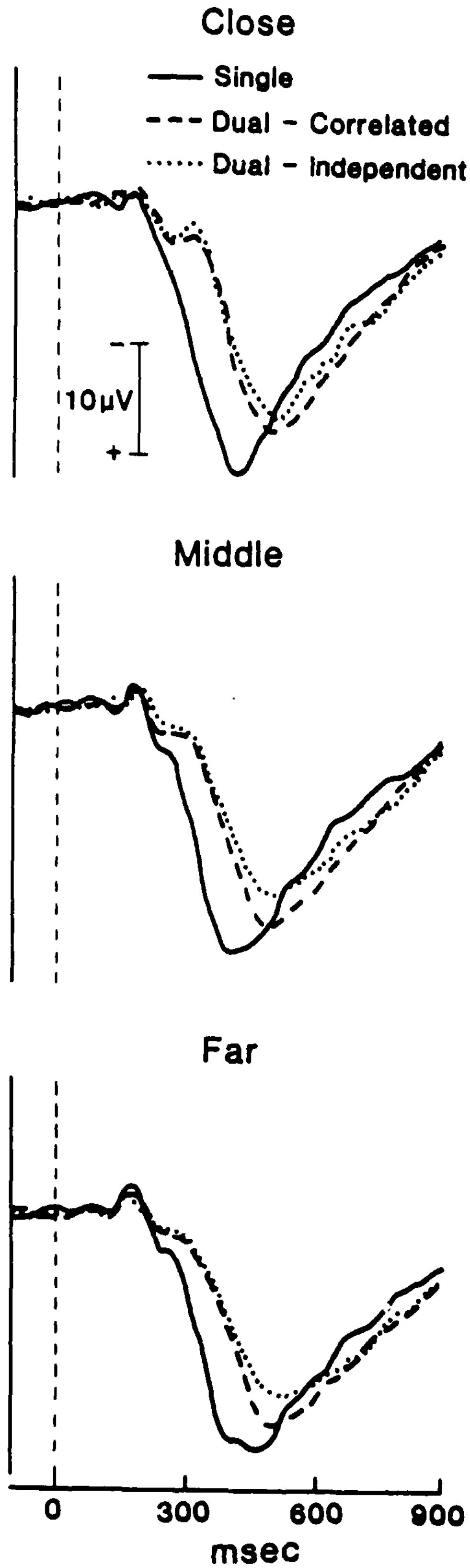

Standards (Oz)
Single Task
Dual - Correlated
Dual - Independent
Close
Middle
Far
5 μV
- +
Attended
Nonattended
0
300
600
900
msec


Contralateral Standards
Single Task
Dual - Correlated
Dual - Independent
Close
Middle
Far
5 µV
Attended
Nonattended
0
300
600
900
msec

Attended Targets – Scalp Distributions



Attended Targets (Pz)



Article 12.

Assessing the Development of Automatic Processing: An Application of Dual-Task and Event-Related Brain Potential Methodologies

Arthur F. Kramer and David L. Strayer

Department of Psychology

University of Illinois

Champaign, Illinois

Running head: DEVELOPMENT OF AUTOMATIC PROCESSING

Biological Psychology (In Press)

Abstract

Previous research has found that properties of automatic processing do not always co-occur, suggesting that the acquisition rates may differ. The present study investigated the acquisition rate of several of these properties by employing additive factors logic, dual task methodology, and event-related brain potentials. Seven subjects participated in a ten session experiment in which they performed two tasks, a visual/memory search task and a pursuit step tracking task, both together and separately. RT and P300 latency measures indicated that parallel processing of the display was achieved early in training in the consistent mapping condition. This processing was unaffected by dual task demands. An analysis of RT/P300 ratios suggests that another form of perceptual efficiency was achieved later in practice in both the varied and consistently mapped search tasks. This effect was larger in the consistent mapping condition. Reductions in the slope of the memory set function occurred significantly earlier for P300 latency than for RT, suggesting that the stimulus evaluation processes became automated more rapidly than the response selection components of memory search. Consistent with an analysis of the processing demands of the two tasks, the introduction of the tracking task and an increase in tracking difficulty produced equivalent interference during consistent and varied mapping conditions. Results are discussed in terms of models skill acquisition and component task automaticity.

Assessing the Development of Automatic Processing: An Application of Dual-Task and Event-Related Brain Potential Methodologies

In recent years a number of theories have been proposed to account for the quantitative and qualitative changes in performance and subject strategies that occur during the development of highly skilled behaviors (LaBerge, 1981; Logan, in press; Neumann, 1984; Posner and Snyder, 1975; Schneider and Shiffrin, 1977). It is interesting to note that many of the characteristics of these "automatic" processes were described over 80 years ago. Solomons and Stein (1896), in their introspective studies of reading and writing, described the development of automatic processing in terms of a rapid increase in the speed of performance, a reduction of the effort and attention required to perform the task, a lack of memory for automatically processed events, and poor development of automatic processing in situations in which difficult perceptual discriminations were necessary. Bryan and Harter's (1897, 1899) studies of professional telegraphers allowed them to add several other characteristics to this list. They described the highly skilled telegraphers' performance as being resistant to intrusions and highly stereotypical. Although a few additional characteristics have been added to the list in recent years (e.g., difficult to modify, activation without intention; Shiffrin and Dumais, 1981), the observations made at the turn of the century still provide the core set of attributes that typify automatic processing.

One of the goals of the present study is to provide a fine grained analysis of the changes in information processing that occur during the development of highly skilled behavior. To this end we have employed several methodological approaches, including dual-task techniques, additive factors logic and event-related brain potentials (ERP), in the assessment of the changes in subject's performance and processing strategies over an extended period of training. A second issue that is addressed in the present study is the degree to which automatic processing can be localized to specific information processing components within highly practiced tasks.

Development of Automatic Processing

A paradigm that has been extensively employed in the examination of automaticity is a modification of the Sternberg memory search task (1966, 1969). In this paradigm subjects are instructed to memorize a set of items. After memorization has been completed, subjects compare a series of visually presented probes to the members of the memory set. Subjects make one response if one of the visually presented items is from the memory set and another response if the probe(s) do not match any of the memory set items. Probes that match an item in the memory set are referred to as targets while probes that do not match a memory set item are labeled distractors. In an extensive series of studies, Schneider and Shiffrin (1977; see also Shiffrin and Schneider, 1977) have demonstrated that automatic processing develops in a consistent mapping condition (CM) in which targets are always selected from one set of items (e.g., letters A to M) and distractors are selected from another set of items (e.g., letters N to Z). Thus, the mapping of the stimuli to the responses does not vary over trials in the CM conditions. Non-automatic or controlled processing is employed in a varied mapping (VM) condition in which subjects are unable to consistently map stimuli to responses. In the VM conditions both targets and distractors are chosen from the same set of items (e.g., letters A to Z). Thus, targets and distractors exchange roles over trials in the VM conditions.

As a result of their studies, Schneider and Shiffrin (1977) proposed a two-process theory to account for changes in performance with practice. Automatic processing which developed as a result of extensive practice with consistent stimulus-response relations was characterized as fast, inflexible, difficult to suppress once learned, and not limited by short-term memory capacity or attention. It was suggested that controlled processing occurs in novel situations or in situations in which stimulus-response relations are inconsistent over time. Controlled processing was characterized as slow, serial and capacity limited. Asymptotic controlled processing requires little practice and is easily modified. The qualitative and quantitative differences between automatic and controlled processing modes have been demonstrated in search tasks with a variety of stimulus materials including; characters (Shiffrin and Schneider, 1977), words (Fisk and Schneider, 1983), categories (Fisk and Schneider, 1984a), spatial and temporal patterns (Eberts and Schneider, 1986; Myers and Fisk, 1987) and higher-order rules (Fisk and Oransky, in press; Kramer and Strayer, 1987a).

A number of criteria have been employed in the memory/visual search task to distinguish between automatic and controlled processing modes. One of these is referred to as the "zero slope" criterion. This criterion is satisfied when the memory or visual search slope is reduced to less than 10 msec per item in a character classification task (Schneider, 1985). This can be compared with a 40 msec slope in an equally practiced VM condition. Two other criteria were developed within the context of dual task experiments in which a CM task is paired with a VM task. One of the criteria deals with performance in the CM task. The "perfect time sharing" criterion is fulfilled when the performance in the dual task condition is equivalent to performance in the single task CM condition. Thus, in this case CM performance is insensitive to the imposition of the VM task or an increase in its difficulty. The second of the dual task criteria deals with performance in the VM task. The "intrusion" criterion is satisfied when performance in the VM task is reduced with the occurrence of a CM target, despite subjects' attempts to ignore the CM event. The co-occurrence of these criteria has been used to evaluate the internal consistency of automatic processing. However, several investigators have noted that these criteria do not always co-occur (Kahneman and Chajzyck, 1983; Paap and Ogden, 1981; Regan, 1981). This led Logan (1985) to suggest that automatic processing might be better characterized by a continuum than a dichotomy since the properties of automatic processing appear to develop at different rates during training.

Schneider's (1985) recent extension of the two-process theory (Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977) provides a description of the changes that occur in performance with the development of automaticity. According to Schneider, the transition takes place in four phases. In the first phase controlled processing predominates. This phase is characterized by the effect of memory load on performance in the search task. In phase two, controlled and automatic processing co-occur and performance is a mixture of the two processing modes. This phase occurs shortly after the introduction of consistent practice and is characterized by a flattening of the memory load function for the larger set sizes. It is proposed that the reduction in slope for the larger set sizes is due to the weak automatic processing finishing first when controlled processing is relatively slow. The third phase is characterized by a lack of effect of memory load on performance (i.e. zero slope). The controlled sequential operations are no longer necessary due to the increased strength of the automatic processing; however, subjects still allocate attention to perform the task. The fourth phase represents pure automatic processing and is characterized by a zero slope and perfect time sharing. In the present study we will examine the transitions in processing predicted by Schneider's (1985) model. This will be accomplished by using ERPs in conjunction with performance measures to localize the changes in information processing that take place during the development of automaticity. For example, the joint use of RT and P300 latency will allow us to distinguish between stimulus evaluation and response processing interpretations of changes in the Sternberg memory set slope and intercept with increased practice and dual task demands.

Component Task Automaticity

In a recent review of the automaticity literature, Jonides, Naveh-Benjamin and Palmer (1985) have suggested that, "the investigation of the task as a whole may lead one to conclude that the task is not performed automatically, when one or more component processes used in completing the task may, in fact, be automatic by acceptable criteria" (p. 163). Other investigators have also emphasized the importance of a careful analysis of task structure prior to the assessment of automaticity. For example, LaBerge (1975, 1981) distinguishes between two types of automatic processing that develop under different conditions, within system or perceptual automaticity and between system or association automaticity. Fisk and Schneider (1984b) argue that automatic processing can develop for consistent components of a task even when the entire task is not consistent. In an ingenious series of studies Logan (1979) used additive factors logic within a dual task paradigm to examine the relative automaticity of different component processes (see also Logan, 1978). In these studies VM and CM choice reaction time tasks were paired with a concurrent memory retention task. Initially, performance in both the CM and VM tasks interacted with increases in the difficulty of the concurrent task. However, following practice performance in the VM task continued to interact with increases in the difficulty of the concurrent task while performance in the CM task was additive with dual task demands. This reduced dual task interference suggests that the automatic processing did not place demands on the same limited capacity processes required by the concurrent memory retention task.

The interpretation of the dual task results in terms of overlapping demands is consistent with resource models of attention which assume that processing resources are limited in quantity and shareable between concurrently performed tasks (Kahneman, 1973; Norman and Bobrow, 1975; Kantowitz and Knight, 1976). Current formulations of resource theory hold that a number of processing units have their own supply of resources that can be shared by several ongoing cognitive operations (Freidman and Polson, 1981; Kinsbourne and Hicks, 1978; Navon and Gopher, 1979; Sanders, 1983). One such model proposed by Wickens (1980) argues that processing resources may be represented by three dimensions: stages of processing (perceptual/central and response), codes of processing (verbal and spatial) and modalities (visual and auditory). Tasks that place demands on the same limited capacity processes are predicted to be more poorly timeshared than tasks that do not overlap in their processing requirements.

The multiple resource interpretation of the pattern of dual task interactions is consistent with the observation that highly practiced CM tasks can be timeshared with less interference from some types of tasks (Hirst, Spelke, Channick and Neisser, 1980; Schneider and Fisk, 1982; Solomons and Stein, 1896) than others (Hoffman, Nelson, and Houck, 1983;

Hoffman, Houck, MacMillian, Simons, and Oatman, 1985). In fact, Hoffman et al. (1985) have suggested that CM memory/visual search tasks may still place demands on two separate limited capacity processes: one concerned with the episodic representation of the stimulus array and the other with the production of speeded responses. In the present experiment we will further examine the issue of component task automaticity by pairing a pursuit step tracking task which has previously been shown to place heavy demands on perceptual and motor (Vidulich and Wickens, 1981; Wickens, Derrick, Micallizi and Beringer, 1980) processes with the Sternberg memory search task. It is predicted that the imposition of the tracking task and an increase in its difficulty will interact with the intercept of the memory set function in the Sternberg task but will be additive with the slope. This prediction is based upon additive factors logic (Sternberg, 1969), which suggests that the intercept in the memory comparison task reflects both encoding and responses processes while the slope provides an index of the time required to complete memory comparison processes. Furthermore, we predict that the magnitude of the interaction between the tracking task and the Sternberg task will be the same for VM and CM conditions since it is the memory comparison process that becomes automated in the memory search task.

Event-Related Brain Potentials

The ERP is a transient series of voltage oscillations in the brain that can be recorded in response to the occurrence of a discrete event (Donchin, 1981; Regan, 1972). The ERP can be partitioned into a number of separate components identified by the polarity and approximate latency of the peak (Donchin, Ritter and McCallum, 1978). Two separate components, the N200 and P300, will be examined in the present study in an effort to aid in the explication of the processing changes that take place during the development of automatic processing in single and dual task conditions.

The P300 component of the ERP is represented by a positive voltage deflection maximal over the parietal scalp with a minimal peak latency of 300 msec post-stimulus (Sutton, Braren, Zubin and John, 1965). This ERP component is particularly useful in the study of automatic processing since its latency appears to be influenced by stimulus evaluation processes while being relatively unaffected by the processes of response selection and execution (Magliero, Bashore, Coles and Donchin, 1984; McCarthy and Donchin, 1981; Ragot, 1984). Thus, the use of P300 latency in conjunction with RT will allow us to isolate processing changes to pre and post response processes.

A number of studies have jointly examined reaction time and P300 latency in a VM Sternberg task (e.g.,Adam and Collins, 1978; Brookhuis, Mulder, Mulder, and Gloerich, 1983; Ford, Pfefferbaum, Tinklenberg, and Kopell, 1982; Ford, Roth, Mohs, Hopkins, and Kopell, 1979; Gomer, Spicuzza, and O'Donnel, 1976; Pfefferbaum, Ford, Roth, & Kopell, 1980; Strayer, Wickens, and Braune, 1987). The consensus of these studies is that both reaction time and P300 latency increase with memory load; however, the effects were greater for reaction time. This suggests that the reaction time slope contains both a stimulus evaluation component and a response-related component (cf. Marcel, 1976).

P300 amplitude has been found useful in the study of the allocation of processing resources among concurrently performed tasks. Within a dual task paradigm, P300 has been found to increase in amplitude with increased processing demands when elicited by task relevant events in a primary task. On the other hand, P300s elicited by secondary task events decrease in amplitude with increases in the perceptual/cognitive difficulty of a primary task (Isreal, Chesney, Wickens and Donchin, 1980; Isreal, Wickens, Chesney, and Donchin, 1980; Kramer, Sirevaag and Braune, 1987; Kramer, Wickens and Donchin, 1983, 1985; Sirevaag, Kramer, Coles and Donchin, 1984; Wickens, Kramer, Vanasse and Donchin, 1983). This pattern of changes in P300 amplitude is consistent with predictions of resource models of attentional allocation (Navon and Gopher, 1979). Thus, it appears that while P300 latency provides information concerning the mental chronometry of information processing, P300 amplitude is sensitive to changes in the resource demands of processes.

The N200 component of the ERP is sensitive to the degree of mismatch between stimulus events (Naatanen, Simpson and Loveless, 1982; Naatanen and Gaillard, 1983; Ritter et al., 1984). In addition to its sensitivity to physical mismatch, N200s have also been found to be elicited by orthographic, phonological and semantic mismatches (Kramer and Donchin, 1987; Sandquist, Rohrbaugh, Syndulko and Lindsley, 1980). Naatanen and Picton (1986) have suggested that the N200 component is composed of a number of temporally overlapping subcomponents, some of which are elicited only during controlled processing (e.g. N2b), and others which are evoked during automatic processing (e.g. mismatch negativity). In the present study the N200 will be used to provide an index of the changes in mismatch processing as a function of the level of automatic processing and dual task demands.

A number of recent studies of automatic processing have used ERPs as dependent measures. Hoffman, Simmons, and Houck (1983) and van Dellen, Brookhuis, Mulder, Okita, and Mulder (1985) employed ERPs to assess changes in information processing, both in terms of the timing of information processes as well as their resource demands, during automatic detection. Subjects searched visual arrays for predefined targets. Several interesting results were obtained. Both research groups found that in the VM visual search condition RT and P300 latency increased with larger visual arrays. However, the RT and P300 latency display size effects were significantly smaller in the CM than the VM conditions. These results suggest that both stimulus evaluation and response processes are reduced during automatic detection. Both groups of investigators also found that P300 amplitudes were not significantly different in the CM and VM conditions. Within a resource model framework these results argue that automatic processing does appear to consume at least one type of perceptual/cognitive processing resource. Finally, van Dellen et al. (1985) found that the difference between target and nontarget ERPs occurred much earlier in the CM than VM condition. These results were interpreted in terms of more efficient perceptual coding in the CM task. Kramer, Schneider, Fisk and Donchin (1986) obtained essentially the same P300 results in a memory search paradigm. Neither P300 latency nor RT was influenced by memory set size in the practiced CM conditions. P300 amplitude did not differ in the CM and VM tasks.

Hoffman et al. (1985) employed ERPs, and more specifically the P300, to provide converging evidence for the role of attentional resources in the automatic detection task. A Sternberg memory search task was paired with a dot detection task. Subjects were required to make a detection response for each task on every trial. Processing priority was manipulated by instructing subjects to emphasize one or the other task or treat them both equally. Performance measures and P300 amplitude showed a tradeoff as a function of processing priority, suggesting that with this particular task combination processing resources were employed during automatic processing.

The present study was designed to extend these results in several ways. First, we have recorded ERPs both during and after extensive practice in CM and VM memory search tasks in order to track the changes in information processing that accompany the development of automatic processing. Second, in addition to the manipulation of memory set size, we have employed a display size of two items arranged horizontally. This will allow us to examine the development of automatic processing in both visual and memory search components of the task. Third, we have employed a concurrent task, which on the basis of previous research, is predicted to interact selectively with perceptual and response but not memory comparison components of the search task. This will enable us to assess the effects of component specific dual task demands on the development of automatic processing in the search task.

Method

Subjects

Seven right handed persons (4 male and 3 female), aged 22 to 27 years, 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 either of the experimental tasks. All of the subjects had normal or corrected to normal vision. Each of the subjects participated in the ten experimental sessions.

Insert Figure 1 about here

Step Tracking and Sternberg Tasks

The single axis pursuit step tracking task is illustrated, along with the Sternberg probes, in Figure 1. The tracking display which consisted of the computer driven target and the subject controlled cursor was presented on a Hewlett Packard CRT which was positioned approximately 70 cm from the subjects. The rectangular target was 1.5 cm x 1.1 cm in size and subtended a visual angle of 1.2 degrees horizontally and .9 degrees vertically. The cursor consisted of one vertical and two horizontal .8 cm lines, and subtended a visual angle of 2.4 degrees horizontally and .9 degrees vertically. The target changed its position along the horizontal axis once every 3.0 sec and the subjects' task was to nullify the position error between the target and cursor. The target could jump anywhere along the horizontal axis. The magnitude and direction of the jump were randomly determined on each trial. The cursor was controlled by manipulating a joystick with the right hand. Single task tracking blocks were comprised of 100 step changes and lasted approximately five min. Although changes in the spatial position of the target were discrete events, the tracking task was performed continuously since the subjects were required to constantly manipulate the joystick to nullify the position error between the target and cursor. The high gain of the tracking system necessitated constant movement of the joystick to control the position of the cursor.

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.

$$X(t) = [(1-a)\int u(t)\,dt] + [(a)\iint u(t)\,dt]$$

where: u = stick position; t = time and a = difficulty level.

The task was conducted at two different levels of the system order manipulation: (1) in the first order (velocity) condition a was set to zero while (2) in the second order (acceleration) condition a was set to 1.0. Numerous investigators have validated the increasing difficulty associated with higher order control (Kramer et al., 1983; North, 1977; Trumbo, Noble and Swink, 1967; Vidulich and Wickens, 1981). Converging evidence employing Sternberg's additive factors paradigm indicates that the demands of higher order tracking are both perceptual and motor in nature, given the requirement to process higher derivatives of the error signal to maintain stable control (Wickens et al., 1980)

In the Sternberg task, subjects' were instructed to decide if one of two letters presented on the CRT belonged to a previously memorized set of letters. A match will henceforth be referred to as a positive trial while a mismatch will be labeled as a negative trial. Each set of thirty trials began with a six sec presentation of a memory set of either two, three or four letters. In the 30 trials that followed the presentation of each memory set, the subjects' task was to deflect a joystick in one direction if one of the two probe items matched an item from the memory set and in the opposite direction if neither of the letters were from the memory set. The joystick was manipulated with the left hand. The direction of the deflection of the joystick for the two responses was counterbalanced across subjects. The two probe items were presented simultaneously for a duration of 200 msec. The ISI was 3000 msec. Subjects were given 1500 msec to indicate their response. Responses prior to 150 msec and after 1500 msec following stimulus onset were scored as incorrect. Instructions emphasized both speed and accuracy.

Two variables served as blocking factors within the Sternberg task. First, subjects performed the task in both CM and VM conditions. In the CM condition, targets were always selected from one set of letters (G,J,N,X) while distractors were selected from another set of letters (P,H,Z,B,F,D,V,T). In the VM condition both targets and distractors were chosen from the same set of letters (P,H,Z,B,F,D,V,T). Targets and distractors exchanged roles over trials in the VM condition. The second blocking factor was the number of items in the memory set. Subjects performed the task with either two, three or four memory set items. A third factor, the probability of a positive or negative trial was fixed at .50 in each block. On a positive trial, one of the items was a target and the other was a distractor. The targets occurred equally often on the left and right. On a negative trial, both of the items were distractors. The Sternberg probes were presented within a rectangular frame that also served as the target in the tracking task.

In the dual-task blocks subjects concurrently performed the tracking task and the Sternberg task. Pursuit step tracking was defined as the primary task. Thus, in these conditions subjects were required to encode a set of memory items, respond to the presentation of the probes and minimize tracking error. Subjects employed the left joystick for their discrete responses in the Sternberg task and the right joystick for their continuous responses in the tracking task. Following each block of trials the subjects were informed of their RT and accuracy in the Sternberg task and the average root mean square (RMS) tracking error.

Insert Figure 2 about here

The temporal sequence of the trials in the dual-task conditions is graphically illustrated in Figure 2. The sequence proceeded as follows: The subjects began tracking changes in the spatial position of the target. Two spatial changes in the target occurred with an ISI of 1.5 sec. At this time, the Sternberg memory set was presented for 6 sec in the center of the

CRT, above the tracking task. Following presentation of the memory set, the changes in the spatial position of the tracking target alternated with the occurrence of the Sternberg probes. Thus, the ISI between events within each task was 3.0 sec while the ISI between Sternberg probes and the displacement of the tracking target was 1.5 sec. After the occurrence of 60 events, another memory set was presented, followed by another 30 Sternberg probes and 30 changes in the position of the target in the tracking task. Thus, each dual-task block was composed of two different Sternberg memory sets, 60 presentations of Sternberg probes and 62 changes in the position of the target in the tracking task. Dual-task blocks lasted approximately 200 sec. Subjects continuously performed the tracking task during each dual task block.

ERP 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. Beckman Biopotential electrodes affixed with Grass paste were used for scalp, mastoid and ground recording. Beckman electrodes, affixed with adhesive collars, were also placed below and supra-orbitally to the right eye to record electro-oculogram (EOG). Electrode impedances did not exceed 10 kohms.

The EEG and EOG were amplified with Van Gogh model 50000 amplifiers (time constant 10 sec and upper half amplitude of 35 Hz, 3dB octave roll-off). Both EEG and EOG were sampled for 1300 msec, beginning 100 msec prior to stimulus onset. The data was digitized every 10 msec. ERP's were filtered off-line (-3dB at 6.29 Hz, 0dB at 14.29Hz) prior to statistical analysis. Evaluation of each EOG record for eye movements and blinks was conducted off-line. EOG contamination of EEG traces was compensated for through the use of an eye movement correction procedure (Gratton, Coles and Donchin, 1982).

Design

A repeated measures, five way factorial design, was employed. The factors were task level (single task Sternberg, Sternberg with first order tracking, and Sternberg with second order tracking), the structure of the Sternberg task (CM or VM), memory set size (2, 3, or 4), session (1 and 10), and the type of Sternberg trial (positive or negative). The first four variables served as blocking factors, the fifth factor was varied within blocks. Subjects also performed single task tracking blocks with first and second order control dynamics.

Procedure

Each of the seven subjects participated in all of the experimental conditions. Ten experimental sessions, within a three week period, were required to complete the experiment. In sessions one and six through ten, subjects performed two replications of the dual and four replications of the single task conditions. Thus, in these sessions subjects completed 24 dual-task blocks comprised of 120 trials each and 32 single task blocks each composed of 60 trials. Dual-task blocks took approximately 3.5 min each while single task conditions required 1.5 min each. Subjects were permitted to take brief rest breaks between each of the blocks and longer breaks whenever necessary. Each of these sessions lasted approximately 2.5 hours.

In sessions two through five, subjects performed five replications of the six single task Sternberg conditions (task structure x set size). Thus, during these sessions subjects completed 30 blocks of trials, each requiring approximately 1.5 min. These sessions lasted 1 hour each. The order of presentation of experimental blocks was counterbalanced across subjects in each of the sessions in a latin square design. Subjects performed 20,160 Sternberg probe trials over the ten sessions, 10,080 each in the CM and VM conditions. The tracking task, in both single and dual-task conditions, was performed for approximately nine hours.

RT and accuracy measures were recorded in the Sternberg task throughout the ten sessions. RMS tracking error was recorded in session one and six through ten. ERPs were recorded during the presentation of the Sternberg probes in the first and last sessions.

Data Analysis

RT in the Sternberg task was defined as the interval between the appearance of the probes and the subject's keypress. RMS error in the tracking task was calculated every 50 msec during single and dual-task conditions. These data were averaged off-line according to experimental condition.

The single-trial ERPs acquired during single and dual-task Sternberg performance were averaged separately for each of the experimental conditions in sessions 1 and 10. Each of the single-subject averages was composed of at least 50 single trial ERPs. The amplitude and latency of the N200 and P300 components were quantified in the following manner. On each single-trial P300 amplitude was measured as the difference between the maximum positive deflection between 300 and 900 msec after the presentation of the probes and the baseline, that was defined as the average voltage recorded over the 100 msec epoch just preceeding the stimulus array (Coles, Gratton, Kramer and Miller, 1986). The latency was defined as the time at which the P300 reached its maximum amplitude. Since the signal to noise ratio for the N200 component is smaller than that for the P300, the amplitude and latency of the N200 were measured on the average waveforms for each of the experimental conditions. The N200 was defined as the maximum negative deflection occurring between 200 and 450 msec post-probe. Due to the enormity of the data base in the present study, analyses of P300 and N200 measures will be restricted to the Pz and Cz sites, respectively.

Results and Discussion

This section is organized in the following manner. First, the results for both of the single tasks will be presented for the first and last experimental sessions. This is done to demonstrate that we have replicated the effects of task structure and practice on measures of performance and to illustrate how psychophysiological measures can be used to explicate changes in cognitive processing. Second, we provide a comparison of the results obtained in the single and dual task conditions, contrasting the transition from single to dual tasks as well as examining the effects of an increase in the difficulty of the tracking task on Sternberg performance. Several dependent measures are examined in both the single and dual-task analyses. These include: RMS error in the tracking task, reaction time (RT), accuracy, measures of the amplitude and latency of the N200 and P300 components of the ERP and the RT/P300 ratio in the Sternberg task.

Insert Figure 3 about here

Single Tasks

Performance measures

A five-way repeated measures ANOVA (subjects x session x task structure x memory set size x response type) was performed on the RT data. Figure 3 presents the average RTs for the CM and VM conditions and the three memory set sizes in sessions 1 and 10. As suggested by the figure, subjects performed the Sternberg task more quickly in session 10 than they did in session 1 ($F(1,6)=19.9$, $p<.01$). RT was faster in the CM than in the VM conditions ($F(1,6)=52.0$, $p<.01$). The analysis also indicated a significant three-way interaction among session, task structure, and memory set size ($F(2,12)= 10.5$, $p<.01$), suggesting that memory set size did not have an effect on RT in the CM conditions in session 10. This interpretation of the three-way interaction was supported by a series of post-hoc comparisons.[1] In session 1, RTs increased as a function of set size for both CM ($F(1,6)= 26.7$, $p<.01$) and VM ($F(1,6)=74.8$, $p<.01$) conditions. In session 10, after subjects had received over 20,160 trials of practice, memory set size still produced a significant effect in the VM conditions ($F(1,6)=76.2$, $p<.01$). However, in the CM conditions, the size of the memory set did not have a significant effect on RT ($p>.65$). The differences in memory set effects between CM and VM conditions as a function of practice are further supported by the memory set slopes obtained in a series of regression analyses performed on these conditions. The slopes for the CM positive and negative conditions in sessions 1 and 10 were 39.6, 43.1, 1.3 and 2.1 msec, respectively. The slopes for the same VM conditions were 53.6, 71.5, 52.6 and 72.2, respectively.

The pattern of RTs produced in the CM and VM conditions is consistent with previous findings and fulfills the "zero slope" criterion for automaticity. (Logan, 1985; Shiffrin and Schneider, 1977; Shiffrin, Dumais and Schneider, 1984). Even extensive practice does not improve memory search performance when subjects are unable to consistently map stimuli to responses. However, when subjects are able to consistently map stimuli to responses, performance improves such that the time required to compare two probes to four items in memory does not significantly differ from the time required to compare the probes to two memory set items.

It is important to note that performance does improve in the VM conditions with practice.2 However, this decrease in RT was in the intercept not the slope. In the CM conditions both the intercept and memory set size slope decreased with practice. This observation of decreased RT in the VM conditions is not new. In fact, it has been hypothesized that the improvement in performance may be attributed to familiarization with the task instructions, equipment useage, or selection of strategies (Ackerman, 1987) or the automatization of consistent components of the task (Schneider, Dumais, and Shiffrin, 1984). However, at the present time these hypotheses remain untested. One strategy for examining these differences in performance is to analyze the changes in RT within the framework of Sternberg's (1966, 1969) additive factors methodology. Within such a framework the pattern of results obtained in the present study would suggest that either encoding and/or response processes become more efficient in both the VM and CM conditions, while the need for memory comparison processes diminishes in the CM but not in the VM conditions. The distinction between encoding and response demands will be addressed by the analysis of P300 latency.

Insert Table 1 about here

A further decomposition of the changes in processing with practice can also be provided by an examination of the response type variable. Although we did not explicity manipulate the number of display comparisons, as has been done in other investigations of automatic and controlled processing, we did present the target items on both the right and left side of the frame. On half of the positive probe trials a memory set item appeared on the left and a distractor on the right while this arrangement was reversed on the other half of the positive trials. Since the probes were presented within a 1.9 degree visual angle for 200 msec, it was unlikely that subjects moved their eyes to scan the display for a target. However, as is apparent from Table 1, subjects did appear to shift their attention from the left to the right when performing in the VM conditions. Thus, it took subjects 34 msec longer to respond to the target when it appeared on the right than it did when the target occurred on the left of the frame ($F(1,6)=6.2$, $p<.05$). In the CM conditions, RT was not influenced by the position of the probe ($p>.48$). It is interesting to note that this relationship between response type and task structure did not interact with the amount of practice. This would imply that the rapid and apparently parallel display processing strategy exhibited in the CM condition developed within the first 1440 trials of practice (session 1), far more quickly than the decrease in the memory set slope. Negative trials took longer to respond to than positive trials in both CM and VM conditions ($F(2,12)=10.22$, $p<.01$).

Insert Figure 4 about here

Figure 4 presents the average error rate for the CM and VM conditions and the three memory set sizes in sessions 1 and 10. The mean error rate across all of the experimental conditions was 1.5%. The pattern of results was quite similar to that obtained for RT. Subjects made significantly more errors when performing in the VM conditions than they did in the CM conditions ($F(1,6)=31.2$, $p<.01$). Error rate also increased as a function of memory set size ($F(2,12)=7.2$, $p<.01$). A significant three-way interaction among session x task structure x set size was also obtained ($F(2,12)=5.8$, $p<.05$), suggesting that subjects' error rate increased as a function of memory load for both CM and VM conditions in session 1, but only for VM conditions in session 10. Since higher error rates were associated with longer RTs, these data suggest that subjects were not trading speed for accuracy while performing the Sternberg task.

Single task tracking performance was evaluated by calculating RMS error for each condition and submitting this data to a three-way repeated measures ANOVA (subjects x sessions x system order). Subjects performed significantly better with first order dynamics than they did with second order dynamics ($F(1,6)=27.5$, $p<.01$). Subjects tracking performance also improved with practice ($F(1,6)=11.85$, $p<.01$). The interaction between system order and session was not significant.

Insert Figure 5 about here

Event-Related Potentials

The ERPs were recorded to address several issues. First, since the latency of the P300 component is influenced by the duration of stimulus evaluation processes but is relatively unaffected by response selection and execution processes, the joint use of RT and P300 latency were used to decompose the information processing demands of the Sternberg task as a function of task structure, practice and dual-task demands. Second, the amplitude of the P300 has been found to vary with the perceptual/cognitive demands of a task. Therefore, this measure was employed to provide an estimate of the resource costs of the tasks, both in isolation and when combined in the dual-task conditions. Finally, since the N200 component provides an index of the processing of mismatches, we would expect that it will assist us in explicating the effects of the experimental manipulations on the target/non-target decision.

Figure 5 shows the grand average ERPs, recorded at the parietal site, for each of the single task Sternberg conditions. ERP components are traditionally defined in terms of their latency relative to a stimulus or response, scalp distribution, and sensitivity to experimental manipulations (Donchin, 1981; Kramer, 1985; Sutton and Ruchkin, 1984). The large positive going deflection in the waveforms became increasingly positive from the Fz to the Pz electrode site ($F(2,12)=34.2$, $p<.01$) and the base to peak measures were maximal between 350 and 800 msec post-stimulus. Based on these criteria this positive deflection can be identified as the P300. The negative going deflection preceeding the P300 increased in negativity from the parietal to frontal site ($F(2,12)=22.4$, $p<.01$) and was maximal in amplitude between 200 and 400 msec post-stimulus. This component will be referred to as the N200.

Insert Figure 6 about here

P300 component The mean P300 latency values obtained in each of the single task conditions are presented in Figure 6. The pattern of latencies in the positive VM conditions is quite similar to that found for RT. The latency of the P300 increases as a function of memory set size in both sessions 1 and 10. The pattern of P300 latencies obtained in the positive CM condition is also quite similar to the pattern of RTs. In session 1, P300 latency increased from the set size 2 to set size 3 condition. In session 10, the effect of memory set size on P300 latency was greatly reduced.

These differences were quantified in a five-way repeated measures ANOVA (subjects x session x task structure x memory set size x response type). As suggested by the figure, P300 latency was significantly shorter in the CM conditions than it was in the VM conditions ($F(1,6)=27.1$, $p<.01$). A significant interaction between session, tasks structure and memory set size suggests that the memory set slope in the CM condition decreased from the first to the last session while the slope obtained in the VM conditions was not influenced by practice ($F(2,12)=4.9$, $p<.05$). This interpretation of the three-way interaction was supported by a series of post-hoc comparisons. Significant memory set effects were obtained for the CM condition in session 1 and the VM conditicns in sessions 1 and 10 ($p<.05$). The memory set effec: was not significant for the CM condition in session 10 ($p>.70$). The slopes for the CM positive conditions in sessions 1 and 10 were 14.6 and .28. For the VM conditions the slopes were 22.6 and 27.2.

The pattern of P300 latencies in the positive CM and VM conditions is consistent both with the pattern of RT results obtained in the present study and the reduced P300 slopes found in practiced CM conditions in previous studies (Hoffman, Simmons, and Houck, 1983; Hoffman et al., 1985; Kramer et al., 1986). It is interesting to note, however, that even in session 1 the slope in the CM condition is flat at the higher memory set sizes. This finding is consistent with phase 2 processing in Schneider's (1985) model of automaticity. In phase 2 of the model, automatic and control processes co-occur and the relatively weak automatic processing finishes before controlled processing only when many controlled processing comparisons must be made. Thus, there should be a flattening of the slope for the higher memory set sizes. The finding that P300 latency and not RT shows this flattening early in practice suggests that the component of stimulus evaluation processing influenced by the memory load manipulation becomes automated prior to the components of response related processing that are affected by memory load (Marcel, 1976).

Evidence for additional differences in processing as a function of task structure is provided in Table 1. Consistent with the pattern of RT's, P300s elicited by the VM targets presented on the left side of the display were 29 msec faster than those elicited by targets presented on the right side of the display ($F(1,6)=10.4$, $p<.01$). However, the P300s elicited in the CM conditions were uninfluenced by the position of the target in the display ($p>.35$). The relationship between response type and task structure did not interact with the amount of practice. Thus, the P300 results when viewed in conjunction with RT suggest that the probes were processed in parallel in the CM conditions while in the VM conditions it appears that subjects serially processed the two items. Furthermore, since the magnitude of the differences in RT and P300 latency were essentially the same in the VM conditions, the serial processing strategy cannot be attributed to differences in response criterion. In both VM and CM conditions negative trials elicited later P300s than positive trials ($F(2,12)=7.2$, $p<.01$).

An analysis of the performance data indicated that the RT intercept for the memory set size function significantly decreased for both the CM and VM conditions from session 1 to 10. Within the framework of Sternberg's additive factors logic this suggests more efficient utilization of either encoding and/or response processes. Given that P300 latency is primarily sensitive to stimulus evaluation processes, the joint examination of RT and P300 latency intercepts can aid in the localization of this practice effect.

The P300 latency intercept was uninfluenced by the level of practice ($p > .50$). This pattern of results suggests that the reduced RT intercepts obtained in the CM and VM conditions after practice can be attributed to a reduction in the time required to complete response processes.

Although the P300 latencies elicited by the target stimuli were consistent with the pattern of RTs, the negative or non-target P300s showed a different pattern of results (see figure 6). A significant four-way interaction among session, task structure, memory set size and response type ($F(4,24)=11.2$, $p < .01$) suggested that the memory set effect was diminished in the negative conditions. Post-hoc comparisons supported this observation. The memory set effect was not significant in any of the negative conditions. The diminished memory set effect in negative conditions has also been found in other studies (Adams and Collins, 1978; Pfefferbaum et al., 1985; Strayer et al., 1987; van Dellen et al., 1985). This insensitivity of P300 latency to the set size manipulation may be the result of a deadline strategy in the negative conditions.

The differences among the mean P300 amplitudes obtained in each of the single task conditions were quantified in a five-way repeated measures ANOVA (subjects x sessions x response type x memory set size x task structure). Consistent with previous studies (Ford et al., 1980; Hoffman, Simmons, and Houck, 1983; Kramer et al., 1986), P300s elicited by the targets were significantly larger than P300s elicited by non-targets ($F(2,12)=18.2$, $p < .01$). No other main effects or interactions were significant.

Insert Figure 7 about here

RT/P300 ratio A single trial ratio of RT to P300 latency was computed to determine changes in the relative proportion of post-stimulus evaluation processing during learning. Figure 7 presents the mean RT/P300 ratios for each of the single task conditions. A referent value in the figure is the solid horizontal line drawn at the RT/P300 ratio of 1.0. This ratio reflects the co-occurrence in time of the P300 peak latency and the RT response. Values larger than 1.0 indicate that RT was preceeded by P300 while values less than 1.0 indicate that P300 followed RT.

The effects of experimental manipulations on the RT/P300 ratio were quantified in a five-way repeated measures ANOVA (subjects x session x task structure x memory set size x response type). As suggested by figure 7, the RT/P300 ratio decreased with practice in both the CM and VM conditions ($F(1,6)=37.9$, $p<.01$). In the CM and VM conditions in session 1 the RT/P300 ratios were equal to or greater than 1.0. However, in session 10 all of the CM conditions and the smaller set sizes in the VM conditions produced ratios of less than 1.0. A significant two-way interaction between session and task structure ($F(1,6)=35.0$, $p<.01$) indicated that the reduction in the RT/P300 ratio was larger in the CM than the VM conditions. These results have several important implications. First, they suggest that the extraction of perceptual information becomes more efficient following practice. Subjects are capable of emitting faster and more accurate responses on the basis of less evaluation of the stimulus array. Second, the smaller RT/P300 ratios in the practiced CM conditions suggest that the process of information extraction is most efficient following consistent training. LaBerge (1981) has described this process of efficient perceptual encoding as within system automaticity.

A significant effect of response type indicated that non-targets required more post-stimulus evaluation processing than targets ($\underline{F}(1,6)=38.4$, $\underline{p}<.01$). However, the difference between targets and non-targets decreased with practice, especially in the CM condition ($\underline{F}(1,6)=7.2$, $\underline{p}<.05$). A decrease in the RT/P300 ratio memory set size effect was also obtained when subjects received substantial practice on the task ($\underline{F}(2,12)=9.6$, $\underline{p}<.01$). The increased RT/P300 ratio with larger set sizes and non-targets has previously been described in terms of the resolution of uncertainty that results from a re-checking of memory. It has been asserted that this process of re-checking occurs subsequent to the elicitation of P300 and therefore results in a lengthening of RT in the absence of changes in P300 latency (Ford et al., 1979). Assuming that such a re-checking process occurs, our results indicate that practice, and especially practice that leads to the development of automaticity, reduces the uncertainty of the memory comparison process (see also Kramer et al., 1986).

N200 component The mean N200 latency and amplitude values recorded at CZ were quantified in five-way repeated measures ANOVAs (subjects x sessions x response type x task structure x memory set size). The N200's elicited by the non-targets were significantly larger than the N200's elicited by the targets ($\underline{F}(1,6)=6.4$, $\underline{p}<.05$). This result is consistent with a large body of literature which has found that N200s are larger for mismatches than they are for matches (Naatanen, Simpson and Loveless, 1982; Naatanen and Gaillard, 1983; Ritter et al., 1984). The only other significant effect for N200 amplitude was session ($\underline{F}(1,6)=13.8$, $\underline{p}<.01$). N200's were larger in session 10 than they were in session 1. This effect may be attributed to a reduction in the latency variability of subjects' performance and ERP components with extended practice on the Sternberg task. It is interesting to note that task structure did not influence the amplitude of the N200's elicited by the stimuli. Thus, consistent with previous findings (Kramer et al., 1986) this pattern of results suggests that the mismatch detection processes reflected by N200 do not differ in automatic and nonautomatic tasks. No significant effects were obtained for N200 latency.

Dual-Task Interactions

The results presented in this section will provide a comparison of performance measures and ERP components obtained in the single and dual task conditions, contrasting the transition from single to dual tasks as well as examining the effects of an increase in the difficulty of the tracking task on Sternberg performance. Since the previous section has already dealt with the significant effects obtained within the single tasks, the current section will be confined to a presentation of those effects that interact with task level (e.g., single task, dual task easy tracking, dual task difficult tracking). This section will be organized in the same format as the single task section. Analyses will include Sternberg RT and error data, RMS tracking error, measures of the amplitude and latency of the P300 and N200 components and the RT/P300 ratio in the Sternberg task.

Insert Figures 8 and 9 about here

Performance measures

Figures 8 and 9 present the mean RT's obtained in the Sternberg task when performed in conjunction with the first order and second order versions of the Step Tracking task. The pattern of RTs obtained in the dual task conditions are quite similar to that obtained in the single task Sternberg condition. RTs increased with set size for the VM and CM conditions in session 1 and the VM conditions in session 10. Decreases in the intercept of the set size function can be observed for both the CM and VM conditions with practice. The RTs for the nontarget responses were longer than those for the target responses.

A six-way repeated measures ANOVA (subjects x task level x session x task structure x memory set size x response type) was performed on the RT data. A significant main effect was obtained for task level ($F(2,12)=4.5$, $p<.05$). Post-hoc comparisons indicated that RT increased both with the introduction of the tracking task and with an increase in its difficulty (p's $<.05$). A significant interaction between task level and memory set size was also obtained ($F(4,24)=4.9$, $p<.01$). Post-hoc comparisons indicated that the interaction was due to the reduction in memory set slope from the first to the second order tracking condition.

The finding of an increase in RT with the introduction of the tracking task and an increase in its difficulty when viewed in conjunction with the decrease in the memory set slope from the first to the second order dual task conditions suggests that the main effect of task level may be due to an increase in the memory set intercept. This hypothesis was confirmed by an analysis of the memory set intercepts. The intercept increased significantly from the single task to the dual first order condition as well as from the dual first order to the dual second order condition (p's $<.05$). Thus, consistent with previous findings the locus of the interaction of the two tasks appears to be during either encoding and/or response selection (Trumbo et al., 1967; Vidulich and Wickens, 1981; Wickens et al., 1980). An analysis of P300 latency will allow us to distinguish between an encoding and response demand interpretation of this effect.

The lack of an interaction of task level with the structure of the task (CM and VM conditions) and the amount of practice underscores the importance of a careful task analysis. Thus, even after 20,000 trials of practice and the attainment of the "zero slope" criterion of automaticity, subjects' response speed in the CM and VM conditions was similarly affected by dual task manipulations. This pattern of results provides additional support for the assertion that it is not tasks that become automatic but instead task components or processes (Jonides et al., 1985; Logan, 1985; Shiffrin, Dumais and Schneider, 1984).

The memory comparison process reflected by the memory set slope was not adversely affected by the introduction of the tracking task or an increase in its difficulty. In fact, an underadditive interaction between memory set size and task level was obtained suggesting an increased overlap in the memory comparison and response processes in the difficult dual task conditions (Pashler, 1984; Stanovich and Pachella, 1977). Thus, in the present study the use of dual task methodology has allowed us localize the development of automaticity to specific task components (see also Logan, 1978, 1979).

Insert Table 2 about here

Table 2 presents the average error rate for the dual task Sternberg conditions. The pattern of errors is quite similar to that obtained in the single task conditions. Error rates were higher in the CM than in the VM conditions. Errors increased with memory set size while decreasing with practice. A main effect of task level was obtained ($F(2,12)=8.0$, $p<.01$). Post-hoc comparisons indicated that the error rate increased with the introduction of the tracking task and an increase in its difficulty ($p's < .05$). A significant two-way interaction between task level and session ($F(2,12)=11.6$, $p < .01$) indicated that the task level difference decreased as subjects received practice on the tasks.

The pattern of RMS error in the tracking task was similar to that obtained in the single task condition. Subjects performed better with first order dynamics that they did with second order dynamics. RMS error diminished with practice. Tracking performance was unaffected by the transition from single to dual task conditions and the levels of memory load in the Sternberg task. Thus, subjects protected their "primary" task performance at the expense of their performance on the "secondary" task. This pattern of results is necessary in order to interpret secondary task decrements (Sternberg task) in terms of primary task (tracking task) demands (Wickens, 1984).

Insert Figures 10 and 11 about here

Event-Related Potentials

The grand average ERPs recorded at Pz in the dual task conditions are presented in Figures 10 and 11. Visual inspection of the waveforms suggests that the dual task ERPs are similar in morphology to those elicited in the single task conditions. The waveforms are characterized by a series of negative and positive peaks with a large amplitude positivity occurring in the range of 450 to 700 msec post-stimulus. The scalp distribution of the N200 and P300 components was also equivalent to that obtained in the single task conditions. The N200 increased in negativity from the parietal to the frontal site while the P300 became increasingly positive from frontal to the parietal site.

Insert Table 3 about here

P300 component The mean P300 latencies obtained at the Pz recording site in each of the dual task conditions are presented in Table 3. As can be seen by a comparison of the table with the single task P300 latencies in Figure 6, the single task effects of task structure, memory set size, practice, and response type are also found in the dual task conditions. The only significant effect of task level was an increase in the intercept of the memory set function from the single task Sternberg condition to the dual task condition ($F(2,12)=3.9$, $p<.05$). The intercept did not increase further when the difficulty of the tracking task was increased from first to second order control dynamics. These results serve to clarify the locus of the increase in the RT intercept with the introduction of the tracking task and an increase in its difficulty. Given that P300 latency is sensitive to factors that influence stimulus evaluation processes and relatively insensitive to response factors, the pattern of P300 latency and RT results suggest that the imposition of the tracking task influenced both encoding and response processes while the increase in tracking difficulty had its primary effect on response processes.

An analysis of P300 amplitudes was undertaken to determine the degree to which the demands imposed upon the subjects by the imposition of the tracking task and an increase in its difficulty would be reflected in the resources allocated to the Sternberg task. P300s′ elicited in the Sternberg task prior to extensive practice decreased in amplitude with the introduction of the tracking task and an increase in its difficulty ($F(2,12)=9.4$, $p<.01$). However, P300 amplitude was not influenced by task level after 25 hours of practice on the single and dual tasks. These results, interpreted within a resource theory framework, suggest that the resource requirements of the tracking task decreased with extensive practice such that the resource demands reflected by P300 amplitude were equivalent in the first and second order tracking conditions.

At first glance this interpretation may seem to be at odds with the finding that although tracking performance improved with practice, RMS error was still significantly larger for the second order than it was for the first order condition after practice. However, previous analyses of the resource requirements of system order manipulations have indicated that they consume both perceptual and response related resources (Trumbo et al., 1967; Vidulich and Wickens, 1981). Thus, the pattern of RMS error and P300 amplitude suggests that although the response related demands of the tracking task may have remained relatively constant over practice, the perceptual demands have diminished (see Kramer et al., 1983 for further support).

RT/P300 ratio The mean RT/P300 ratios obtained in the dual task conditions are presented in Table 4. As can be seen from the table, the major effects obtained in the single task conditions were replicated in the dual task conditions. RT/P300 ratio decreased with practice for both CM and VM tasks, although this effect was largest in the CM conditions. The ratios were larger for the nontargets than they were for the targets especially in the VM conditions. A significant two-way interaction between task level and session was obtained ($F(2,12)=10.3$, $p<.01$). Post-hoc comparisons indicated that this effect could be attributed to an increase in the RT/P300 ratio in the second order conditions. Thus, the RT/P300 ratios recorded on the single trials are consistent with the pattern of results obtained for the mean RT's and P300 latencies. Both sets of analyses suggest that the time required for response processes increased in the difficult dual task condition.

N200 component The only significant single/dual task effect for the N200 component was a main effect of task level ($F(2,12)=12.2$, $p<.01$). Post-hoc comparisons indicated that this was due to a reduced N200 amplitude in the second order conditions. Since this is the condition in which P300 amplitude was also smallest it is conceivable that the reduction in N200 amplitude was due to increased latency variability in this condition. Although it was impossible to evaluate the latency variability of the N200 since measures of its amplitude and latency were obtained from averages, an analysis of both P300 and RT single trial measures indicated an increased latency variability in the second order conditions.

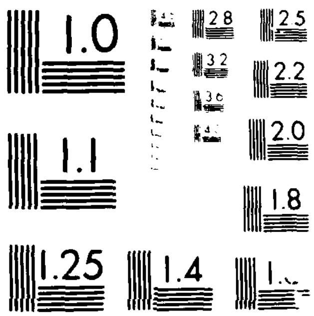
Conclusions

The present study had two main goals; to elucidate the acquisition rates of several of the properties of automatic processing and to examine the degree to which automatic processing can be localized to particular components of the search task. These questions were addressed by a combined methodological approach which included ERP components, dual-task manipulations and additive factors logic.

Schneider's (1985) recent extension of the two-process theory of automatic and controlled processing (Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977) can be used as a framework in which to evaluate the changes in processing obtained in the present study as a function of practice, task structure and dual task demands. The first phase of the

AD-A191 244 THE EVENT-RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING A..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 6/10

UNCLASSIFIED 29 FEB 88 CPL-88-1 AFOSR-TR-88-0316 F/G 5/8 NL


1.0
2.8
2.5
3.2
2.2
2.0
1.1
1.8
1.25
1.4

model is characterized by the effect of memory and visual display load on performance in the search task. Controlled processing is proposed to account for these effects. Since VM search tasks do not have a consistent relation between stimuli and responses their performance is predicted to remain at this level even with substantial training. Performance in the VM condition in the current study is consistent with the characteristics of phase 1. Even after 20,000 trials of practice, memory comparison and display search performance had not improved in the VM conditions. RT and P300 latency increased with memory set size and the display position of the target.

In phase two, controlled and automatic processing are proposed to co-occur with performance being a mixture of the two processing modes. It is argued that this phase occurs shortly after the introduction of consistent practice and is characterized by a flattening of the memory load function for the higher set sizes and the occurrence of the "pop out" effect in visual search. The reduction in the memory set slope at the larger set sizes is accounted for by postulating that the relatively weak automatic processing finishes before the controlled processing when many controlled processing comparisons must be made. It is suggested that the "pop out" effect or parallel processing of the visual array is due to the tuning of a perceptual filter for the consistently mapped targets (see also Hoffman, 1979; LaBerge, 1981). Evidence for both of these performance characteristics have been obtained early in training in the present study. It is interesting to note, however, that the decreased memory load slope for the higher set sizes was obtained for P300 latency but not RT during the first 1440 trials of practice. An analysis of data collected subsequent to the first session of training indicated that RT displayed the same change in slope after approximately 3,000 trials. This pattern of results suggest that the components of stimulus evaluation processing influenced by the memory load manipulation become automated at a faster rate than the components of response related processing that are affected by memory load. Thus, P300 latency may serve as an early marker of the automation of stimulus evaluation processing.

In addition to obtaining the reduced memory set slope in the CM conditions early in training, we also obtained evidence that is consistent with the previously reported "pop out" effect in visual search (Logan, 1985; Shiffrin and Dumais, 1981). In our display, targets could appear in either the right or left side of the frame. In the VM conditions subjects' RT and P300 latency was significantly longer when the target occurred on the right than when it was presented on the left side of the frame (34 and 29 msec differences for RT and P300, respectively). However, even in the first session RT and P300 latency did not differ as a function of whether the target occurred on the left or the right in the CM conditions. This pattern of results was uninfluenced by dual task demands or additional practice. Thus, consistent with Schneider's (1985) model it appears that rapid and apparently parallel display processing develops early in training in CM conditions.

With additional consistent practice performance is predicted to transition from phase 2 to phase 3. The third phase is characterized by a lack of effect of memory load on performance (i.e. zero slope). It is asserted that the sequential operations of controlled processing are no longer necessary. However, subjects still allocate attention to perform that task. The zero slopes obtained for RT and P300 latency in session ten are consistent with the transitions predicted by Schneider's model.

Although we did not record ERPs in sessions 2 through 9, it would appear, based upon the reduced P300 latency slopes for the larger set sizes in session 1, that the P300 latency slope may have attained the zero slope criterion prior to RT. Further research will be required to test this hypothesis.

In addition to attaining the zero slope criterion after substantial consistently mapped practice, we also found another interesting effect not predicted by Schneider's model. RT/P300 ratios were significantly smaller in session 10 than they were in session 1, suggesting that subjects' emitted fast and accurate responses with less evaluation of the stimulus array after substantial practice on the search task (see also van Dellen et al., 1985). It is interesting to note that although this reduction in RT/P300 ratio was larger in the CM condition, the reduction was also significant in the VM condition. Therefore, it appears that subjects become more efficient at extracting the relevant information from a visual display with practice. In some sense it is not surprising that this process improves for both CM and VM search tasks. The VM task is not inconsistent in all of its components, only stimulus-response mapping changes over trials. Thus, even in the VM task a limited set of items occurs in fixed positions on the display. It appears that this level of consistency is sufficient to enhance the process of information extraction, although additional consistency in the form of stimulus-response mapping results in added perceptual efficiency.

It is important to note that this improved efficiency in information extraction appears to differ in several respects from the parallel display processing manifested in session 1. First, the RT/P300 ratio effect was obtained in both CM and VM conditions while the parallel processing effect was obtained in the CM but not the VM conditions. Second, the parallel processing effect occurred early in practice while the changes in RT/P300 ratio were coincident with the zero memory set slope. Finally, the RT/P300 ratio was influenced by dual task demands while the parallel processing effect was insensitive to the introduction of the tracking effect as well as an increase in tracking difficulty. This pattern of results suggests that the improved processing efficiency manifested by the decreased RT/P300 ratio develops gradually and remains a limited-capacity process, at least within the level of training received in the present study. Further research will be necessary to determine whether this process becomes less sensitive to dual task demands with additional training.

Phase 4 processing occurs when automatic processing has developed sufficiently such that neither attention nor controlled processing is necessary. This phase is characterized by a resistance to dual task demands (i.e. perfect time sharing criterion). It has been asserted that, "Consistently mapped training appears to be a necessary condition for improvements in the subjects' ability to timeshare tasks" (Schneider and Fisk, 1982; p. 161). However, although consistently mapped training may be a necessary condition for improved timesharing it is not sufficient. For example, Schneider (1985) has qualified the conditions under which transitions occur from phase 3 to phase 4 processing by stating that, "Phase 4 processing may not operate effectively if stimuli are severely degraded" (p. 489).

In the present study we have extended the line of research which suggests that automatic processing develops for task components rather than tasks (Jonides et al., 1985; Logan, 1985) by showing that even after 20,000 trials of practice with two supra-threshold tasks, dual task performance decrements in the form of increased memory set intercepts were the same for

VM and CM conditions. The dual task interactions were predicted on the basis of previous data that have suggested that the processing demands of the tracking task and the memory comparison process do not overlap (Wickens et al., 1980). Thus, the spare processing capacity liberated by the automation of the comparison process would not be predicted to be of benefit to the processing performed in the tracking task (Wickens, 1980). It is interesting to note that we found perfect timesharing performance of the CM but not the VM search task with a recognition running memory task after the same amount of practice that subjects received in the present study (Kramer and Strayer, 1987b). Therefore, it cannot be argued that subjects were given insufficient training to achieve phase 4 processing in this study. Instead, this pattern of results would seem to argue for the inclusion of an additional qualification on the transition from phase 3 or capacity-limited to phase 4 or capacity-free processing. This transition appears to be limited (a) to task combinations in which the capacity liberated by the automatic processing may be used by the concurrent task -- the Multiple Resource view of processing resources and (b) the task stimuli must be presented at supra-threshold levels (Hoffman, Nelson, and Houck, 1983; Hoffman et al., 1985).

References

Ackerman, P. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. Psychological Bulletin, 102, 3-27.

Adam, N., & Collins, C.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.

Bryan, W. and Harter, N. (1897). Studies in the physiology and psychology of the telegraphic language. Psychological Review, 4, 27-53.

Bryan, W. and Harter, N. (1899). Studies in the telegraphic language. The acquisition of a hierarchy of habits. Psychological Review, 4, 345-375.

Coles, M.G.H., Gratton, G., Kramer, A. & Miller, G. (1986). 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.

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

Donchin, E., Ritter, W. and McCallum, W.C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting and S.H. Koslow (Eds.), Event Related Brain Potentials in Man. New York: Academic Press.

Eberts, R. & Schneider, W. (1986). Effects of perceptual training of sequenced line movements. Perception and Psychophysics, 39, 236-247.

Fisk, A.D., & Oransky, N.A. (In press). Examination of the role of "higher-order" consistency in skill development.

Fisk, A.D. & Schneider, W. (1983). Category and word search: Generalizing search principles to complex processing. Journal of Experimental Psychology: Learning, Memory and Cognition, 9, 177-195.

Fisk, A.D. & Schneider, W. (1984a). Memory as a function of attention, level of processing, and automatization. Journal of Experimental Psychology: Learning, Memory and Cognition, 10, 181-197.

Fisk, A.D. and Schneider, W. (1984b). Consistent attending versus consistent responding in visual search: Task versus component consistency in automatic processing development. Bulletin of the Psychonomic Society, 22, 330-332.

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., Mohs, R.C., Hopkins, L.F. & Kopell, B.S. (1979). Event-related potentials recorded from young and old adults during a memory retrieval task. Electroencephalography and Clinical Neurophysiology, 47, 450-454.

Freidman, A. and Polson, M.C. (1981). The hemispheres as independent resource systems: Limited capacity processing and cerebral specialization. Journal of Experimental Psychology: Human Perception and Performance, 7, 1030-1058.

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

Gratton, G., Coles, M.G.H. and Donchin, E. (1982). A new method for off-line removal of ocular artifact. Electroencephalography and Clinical Neurophysiology, 55, 468-484.

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

Hoffman, J. E. (1979) A two-stage model of visual search. Perception and Psychophysics, 25, 319-327.

Hoffman, J.E., Nelson, B. & Houck, M.R. (1983). The role of attentional resources in automatic detection. Cognitive Psychology, 15, 379-410.

Hoffman, J.E., Simmons, R.F. & Houck, M.R. (1983). Event-related potentials during controlled and automatic target detection. Psychophysiology, 20, 625-632.

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

Isreal, J.B., Chesney, G.L., Wickens, C.D. and 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. and 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.

Jonides, J., Naveh-Benjamin, M., & Palmer, J. (1985). Assessing automaticity. Acta Psychologica, 60, 157-171.

Kahneman, D. (1973). Attention and Effort. Englewood Cliffs, New Jersey: Prentice Hall.

Kahneman, D., & Chajzyck, 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.

Kantowitz, B.H., & Knight, J.L. (1976). Testing tapping time-sharing I: Auditory secondary task. Acta Psychologica, 40, 343-362.

Kinsbourne, M. and Hicks, R. (1978) Functional cerebral space. In J. Requin (Ed.), Attention and Performance VIII. New York: Academic Press.

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

Kramer, A.F., & Donchin, E. (1987). Brain potentials as indicies of orthographic and phonological interactions during word matching. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 76-86.

Kramer, A.F., Schneider, W., Fisk, A.D. and Donchin, E. (1986). The effects of practice and task structure on components of the event-related brain potential. Psychophysiology, 23, 33-47.

Kramer, A.F., Sirevaag, E.J., & Braune, R. (1987). A psychophysiological assessment of operator workload during simulated flight missions. Human Factors, 29, 145-160.

Kramer, A.F. and Strayer, D.L. (1987a). Consistency and the development of automatic processing. Technical Report. Department of Psychology, University of Illinois. Champaign, Illinois.

Kramer, A.F. and Strayer, D. (1987b). P300 operating characteristics: Performance/ERP analysis of dual-task demands and attention. In M. Kutas and B. Renault (Eds.), Proceedings of the 4th International Conference on Cognitive Neurosciences. Paris-Dourdan, France.

Kramer, A., Wickens, C.D. and Donchin, E. (1983). An analysis of the processing demands of a complex perceptual-motor task. Human Factors, 25, 597-622.

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

Laberge, D. (1975). Acquisition of automatic processing in perceptual and associative learning. In P. M. A. Rabbitt & S. Dornic (Eds.), Attention and

Performance V. New York: Academic Press.

LaBerge, D. (1981). Automatic information processing: A review. In J. Long and A. Baddeley (Eds.), Attention and Performance IX. Hillsdale, N.J.: Erlbaum.

Logan, G. (1978). Attention in character classification tasks: Evidence for the automaticity of component stages. Journal of Experimental Psychology: General, 107, 32-63.

Logan, G. (1979). On the use of concurrent memory load to measure attention and automaticity. Journal of Experimental Psychology: Human Perception and Performance, 5, 189-207.

Logan, G. (1985). Skill and automaticity: Relations, implications and future directions. Canadian Journal of Psychology, 39, 367-386.

Logan, G. D. (in press). Toward an instance theory of automatization. Psychological Review.

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.

Marcel, A.J. (1976). Negative set effects in character classification: A response retrieval view of reaction time. Quarterly Journal of Experimental Psychology, 29, 31-48.

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

Myers, G. and Fisk, A.D. (1987). Training consistent task components: Application of automatic and controlled processing theory to industrial task training. Human Factors, 29, 255-268.

Naatanen, R. & Gaillard, A.W.K. (1983). The orienting reflex and the N2 deflection of the ERP. In A.W.K. Gaillard and W. Ritter (Eds.), Tutorials in Event Related Potential Research: Endogenous Components. Amsterdam: North Holland Publishing Company.

Naatanen, R. and Picton, T.W. (1986). N2 and automatic versus controlled processes. In W.C. McCallum, R. Zappoli, F. Denoth (Eds.), Cerebral Psychophysiology: Studies in Event-Related Potentials. Holland: Elsevier Science.

Naatanen, R., Simpson, M. & Loveless, N.E. (1982). Stimulus deviance and evoked potentials. Biological Psychology, 14, 53-98.

Navon, D. and 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 components and demands as factors in dual-task performance. Aviation Research Laboratory. Report Number ARL-77-2/AFOSR-77-2. University of Illinois at Urbana-Champaign,

Neumann, O. (1984). Automatic processing: A review of recent findings and a plea for an old theory. In W. Prinz and A.F. Sanders (Eds.), Cognition and Motor Processes. Berlin: Springer-Verlag.

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. Journal of Experimental Psychology: Human Perception and Performance, 10, 358-377.

Paap, K.R. and Ogden, W.C. (1981). Letter encoding is an obligatory but capacity demanding operation. Journal of Experimental Psychology: Human Perception and Performance, 7, 518-527.

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

Posner, M.I. and Snyder, C.R. (1975). Attention and cognitive control. In R. Solso (Ed.), Information Processing and Cognition: The Loyola Symposium. Potomac, Maryland: Erlbaum.

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

Regan, D. (1972). Evoked Potentials in Psychology, Sensory Physiology and Clinical Medicine. London: Chapman and Hall.

Regan, J.E. (1981). Automaticity and learning: Effects of familiarity on naming letters. Journal of Experimental Psychology: Human Perception and Performance,7, 180-195.

Ritter, W., Ford, J., Gaillard, A., Harter, R., Kutas, M., Naatanen, R., Polich, J., Renault, B. & Rohrbaugh, J. (1984). Cognition and event-related potentials I. The relation of negative potentials and cognitive processes. In R. Karrer, J. Cohen & P. Tueting (Eds.), Brain and Information: Event-Related Potentials. New York: New York Academy of Sciences.

Sanders, A.F. (1983). Towards a model of stress and human performance. Acta Psychologica, 53, 63-97.

Sandquist, T.F., Rohrbaugh, J., Syndulko, K., & Lindsley, D.B. (1980). Electrocortical signs of levels of processing: Perceptual analysis and recognition memory. Psychophysiology, 17, 568-576.

Schneider, W. (1985). Toward a model of attention and the development of automatic processing. In M. Posner and O. Marin (Eds.), Attention and Performance XI. New York: Erlbaum.

Schneider, W., Dumais, S.T., & Shiffrin, R.M. (1984). Automatic and control processing and attention. In R. Parasuraman & R. Davies (Eds.), Varieties of Attention. (pp. 1-27). New York: Academic Press.

Schneider, W. & Fisk, A.D. (1982). Concurrent automatic and controlled visual search: Can processing occur without resource cost? Journal of Experimental Psychology: Learning, Memory and Cognition, 8, 261-277.

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

Shiffrin, R.M. & Dumais, S.T. (1981). The development of automatism. In J. R. Anderson (Ed.), Cognitive Skills and Their Acquisition. (pp. 111-140). Hillsdale, NJ: Erlbaum.

Shiffrin, R.M., Dumais, S.T. & Schneider, W. (1984). Characteristics of Automatism. In J. Long and A. Baddeley (Eds.), Attention and Performance IX. Erlbaum: Hillsdale, New Jersey.

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

Sirevaag, E., Kramer, A.F., Coles, M.G.H. and Donchin, E. (1984). P300 amplitude and resource allocation. Psychophysiology, 21, 598-599.

Solomons, L.M. and Stein, G. (1896). Normal motor automatism. Psychological review, 3, 492-512.

Stanovich, K.E. & Pachella, R.G. (1977). Encoding, stimulus-response compatibility, and stages of processing. Journal of Experimental Psychology: Human Perception and Performance, 3, 411-421.

Sternberg, S. (1966). High speed scanning in human memory. Science, 153, 652-654.

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

Strayer, D.L., Wickens, C.D. and Braune, R. (1987). Adult age differences in speed and capacity of information processing: 2. An electrophysiological approach. Psychology and Aging, 2, 99-110.

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

Sutton, S. and Ruchkin, D.S. (1984). The late positive component: Advances and new problems. Brain and Information: Event-Related Potentials. New York: New York Academy of Sciences.

Trumbo, D., Noble, M. and Swink, J. (1967). Secondary task interference in the performance of tracking tasks. Journal of Experimental Psychology, 73, 232-240.

van Dellen, H.J., Brookhuis, K.A., Mulder, G., Okita, T. & Mulder, L.J.M. (1985). Evoked potential correlates of practice in a visual search task. In D. Papakostopoulus (Ed.), Clinical and Experimental Neuropsychology. Croom Helms: Beckenham.

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

Wickens, C.D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), Attention and Performance VIII Hillside, New Jersey: Erlbaum.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman and R. Davies (Eds.), Varieties of Attention. New York: Academic Press.

Wickens, C.D., Derrick, W.D., Micallizi, J. and Beringer, D. (1980). The structure of processing resources. In the Proceedings of the 24th Annual Meeting of the Human Factors Society. Santa Monica, California.

Wickens, C., Kramer, A., Vanasse, L. and Donchin, E. (1983). The performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information processing resource. Science, 221, 1080-1082.

Acknowledgment

The research was supported by NASA Ames Research Center under contract number 1-5-25-477 with Dr. Michael Vidulich as technical monitor and by the Air Force Office of Scientific Research under contract number F49620-79-C-0233 monitored by Dr. Al Fregly. The authors wish to thank John Flach, Gordon Logan, Risto Naatanen and Bernard Renault for their helpful comments on an earlier draft of this paper. Requests for reprints should be sent to Arthur F. Kramer, Department of Psychology, 603 East Daniel Street, University of Illinois, Champaign, Illinois, 61820.

Footnotes

1 All post-hoc comparisons reported in this article are computed with the Bonferroni t-test and are significant at $p<.05$.

2 For the purposes of the present study, we defined the intercept as the average RT obtained in the memory set size two condition. This was necessary because of the flat slope in the practiced CM condition.

Table 1

Means (msec) and standard deviations (SD) for the three different response types in both CM and VM conditions. Note: NT is negative trial, PL is positive left trial, and PR is positive right trial. P300's have been recorded at Pz.

	CM			VM		
	NT	PL	PR	NT	PL	PR
RT	577	519	515	671	584	618
SD	67	51	40	71	50	61
P300 Latency	605	547	551	620	570	599
SD	67	47	53	64	54	47

Table 2

Mean error rates for all of the dual task Sternberg conditions. Note: S1 is session 1, S2 is session 10.

Memory Set Size	CM Dual First S1	CM Dual First S2	CM Dual Second S1	CM Dual Second S2	VM Dual First S1	VM Dual First S2	VM Dual Second S1	VM Dual Second S2
Targets:								
Set Size 2	1.0	.6	1.2	1.0	1.9	1.6	2.6	1.7
Set Size 3	1.4	.7	1.4	1.2	2.4	2.3	2.5	2.3
Set Size 4	1.4	.4	1.5	.8	3.6	2.2	3.7	2.3
Non-targets:								
Set Size 2	.9	.7	2.2	1.0	1.1	.8	4.7	1.8
Set Size 3	1.6	.7	2.2	1.1	3.4	1.5	5.0	2.1
Set Size 4	2.0	1.4	3.1	1.2	6.1	1.4	6.2	2.5

Table 3

Mean P300 latencies at Pz (msec) for all of the dual task Sternberg conditions.

Note: S1 is session 1, S2 is session 10.

Memory Set Size	CM				VM			
	Dual First		Dual Second		Dual First		Dual Second	
	S1	S2	S1	S2	S1	S2	S1	S2
Targets:								
Set Size 2	555	553	559	547	586	605	587	585
Set Size 3	597	555	600	558	606	621	605	610
Set Size 4	604	564	605	554	639	645	627	630
Non-targets:								
Set Size 2	618	615	642	610	624	640	606	636
Set Size 3	636	627	644	626	630	666	620	665
Set Size 4	641	627	637	629	638	681	648	673

Table 4

Mean RT/P300 ratios for all of the dual task Sternberg conditions. Note: S1 is session 1, S2 is session 10.

Memory Set Size	CM Dual First S1	CM Dual First S2	CM Dual Second S1	CM Dual Second S2	VM Dual First S1	VM Dual First S2	VM Dual Second S1	VM Dual Second S2
Targets:								
Set Size 2	1.04	.88	1.06	.93	1.04	.89	1.14	.98
Set Size 3	.99	.85	1.08	.94	1.12	.92	1.18	1.00
Set Size 4	1.12	.91	1.09	.96	1.17	.99	1.22	1.05
Non-targets:								
Set Size 2	1.04	.87	1.13	.90	1.12	.92	1.23	1.01
Set Size 3	1.08	.88	1.18	.93	1.25	.96	1.29	1.05
Set Size 4	1.25	.88	1.24	.94	1.40	1.11	1.36	1.08

Figure Captions

Figure 1. A graphic representation of the task configuration. The subjects controlled the cursor in the tracking task with the right hand and responded to the Sternberg probes with the left hand. Sternberg probes were presented in the rectangular target. Memory sets were displayed in the center of the display, directly above the tracking task.

Figure 2. The temporal structure of the tracking and Sternberg tasks both separately and together.

Figure 3. Average single task RTs for each of the experimental conditions in sessions 1 and 10.

Figure 4. Average single task error rates for each of the experimental conditions in sessions 1 and 10.

Figure 5. Grand average ERPs recorded at the parietal site for all of the single-task Sternberg conditions in sessions 1 and 10.

Figure 6. Average single task P300 latencies recorded at Pz for each of the experimental conditions in sessions 1 and 10.

Figure 7. Average single task RT/P300 ratios for each of the experimental conditions in sessions 1 and 10.

Figure 8. Average Sternberg RTs for each of the dual first order experimental conditions in sessions 1 and 10.

Figure 9. Average Sternberg RTs for each of the dual second order experimental conditions in sessions 1 and 10.

Figure 10. Grand average ERPs recorded at the parietal site for all of the dual-task first order Sternberg conditions in sessions 1 and 10.

Figure 11. Grand average ERPs recorded at the parietal site for all of the dual-task second order Sternberg conditions in sessions 1 and 10.

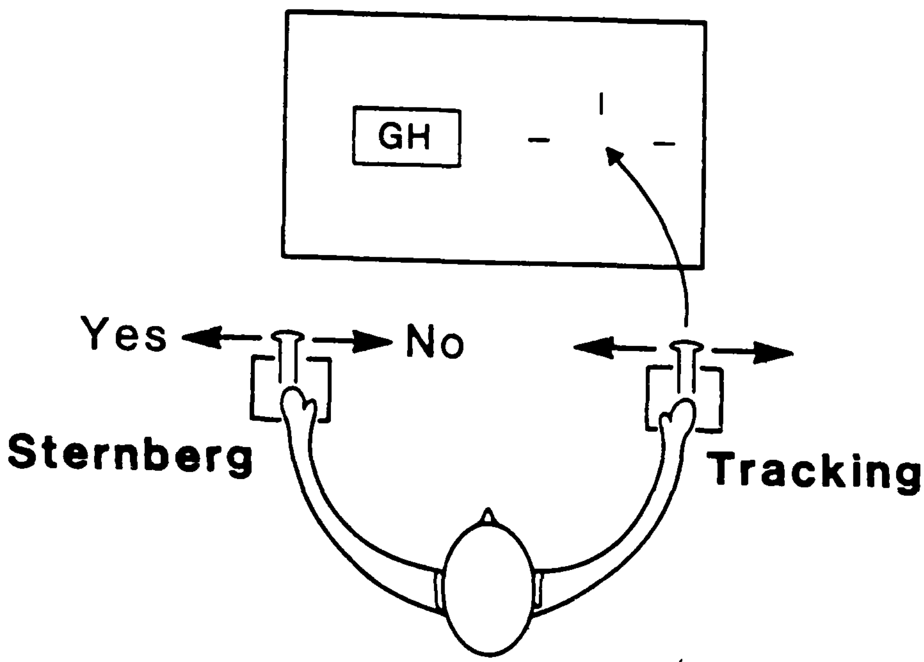


FIGURE 1

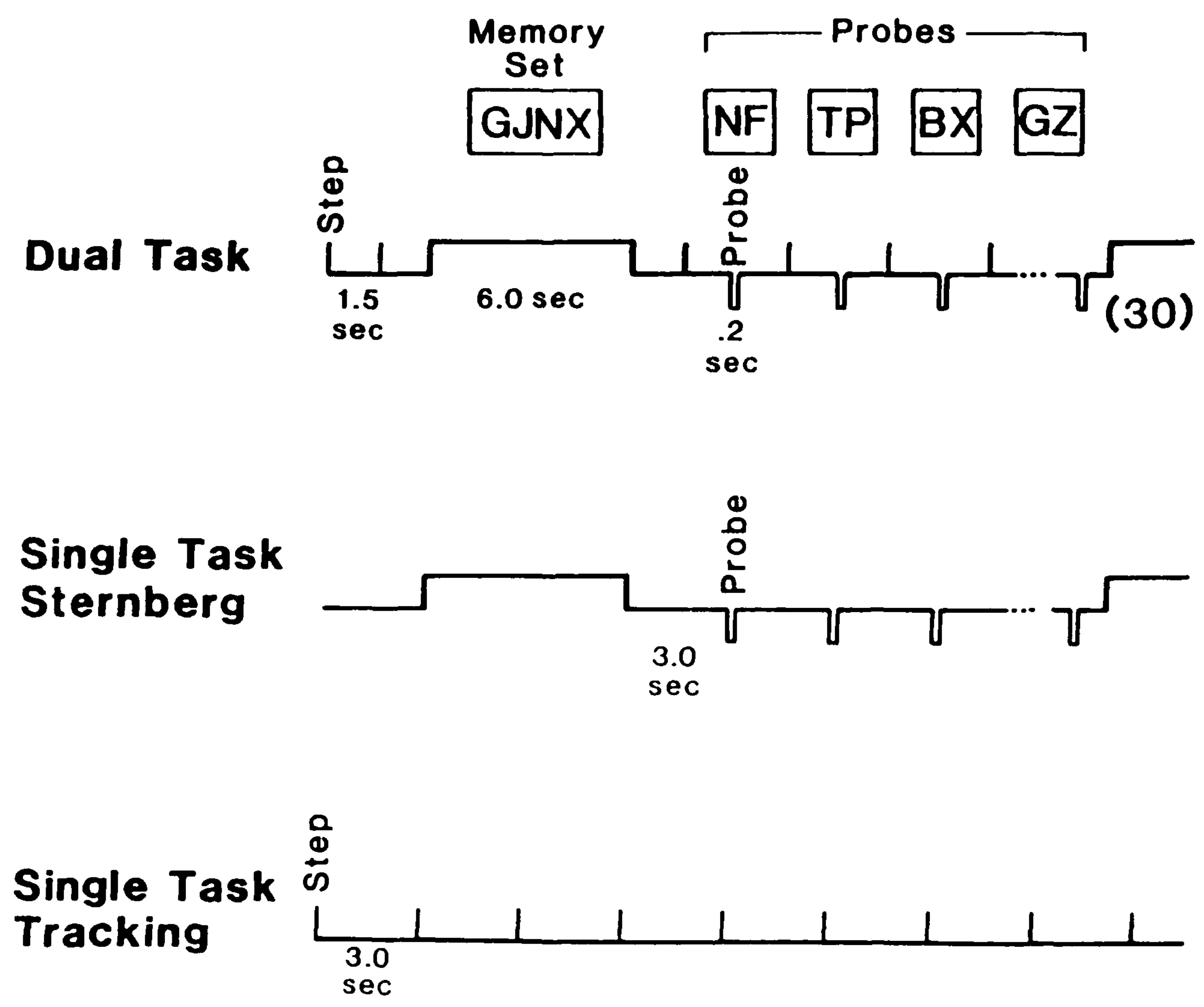

Memory Set
GJNX
Probes
NF
TP
BX
GZ
Step
Probe
Dual Task
1.5 sec
6.0 sec
.2 sec
(30)
Single Task Sternberg
Probe
3.0 sec
Single Task Tracking
Step
3.0 sec

FIGURE 2

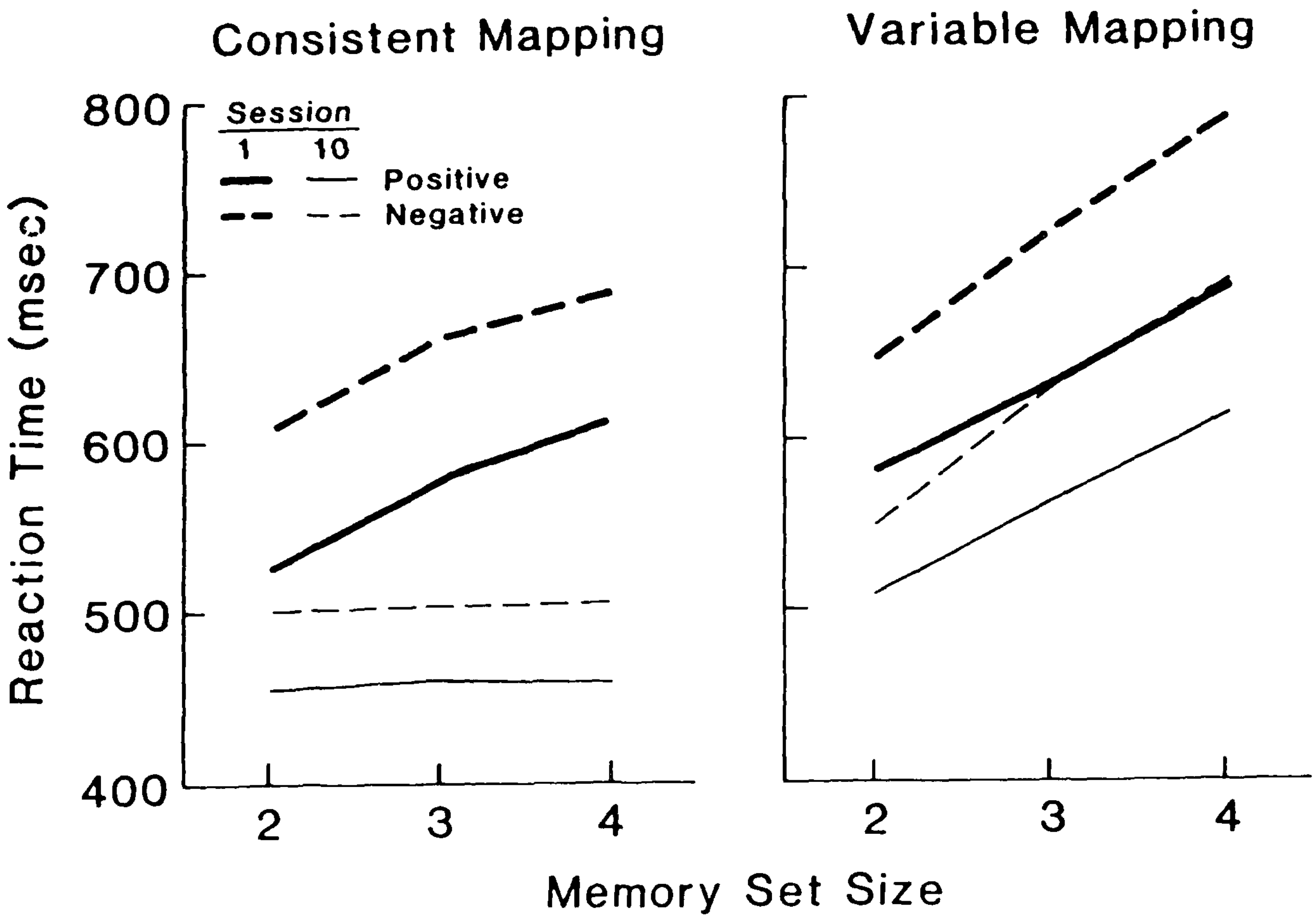

Consistent Mapping
Variable Mapping
Session
1
10
Positive
Negative
Reaction Time (msec)
800
700
600
500
400
2
3
4
Memory Set Size

FIGURE 3

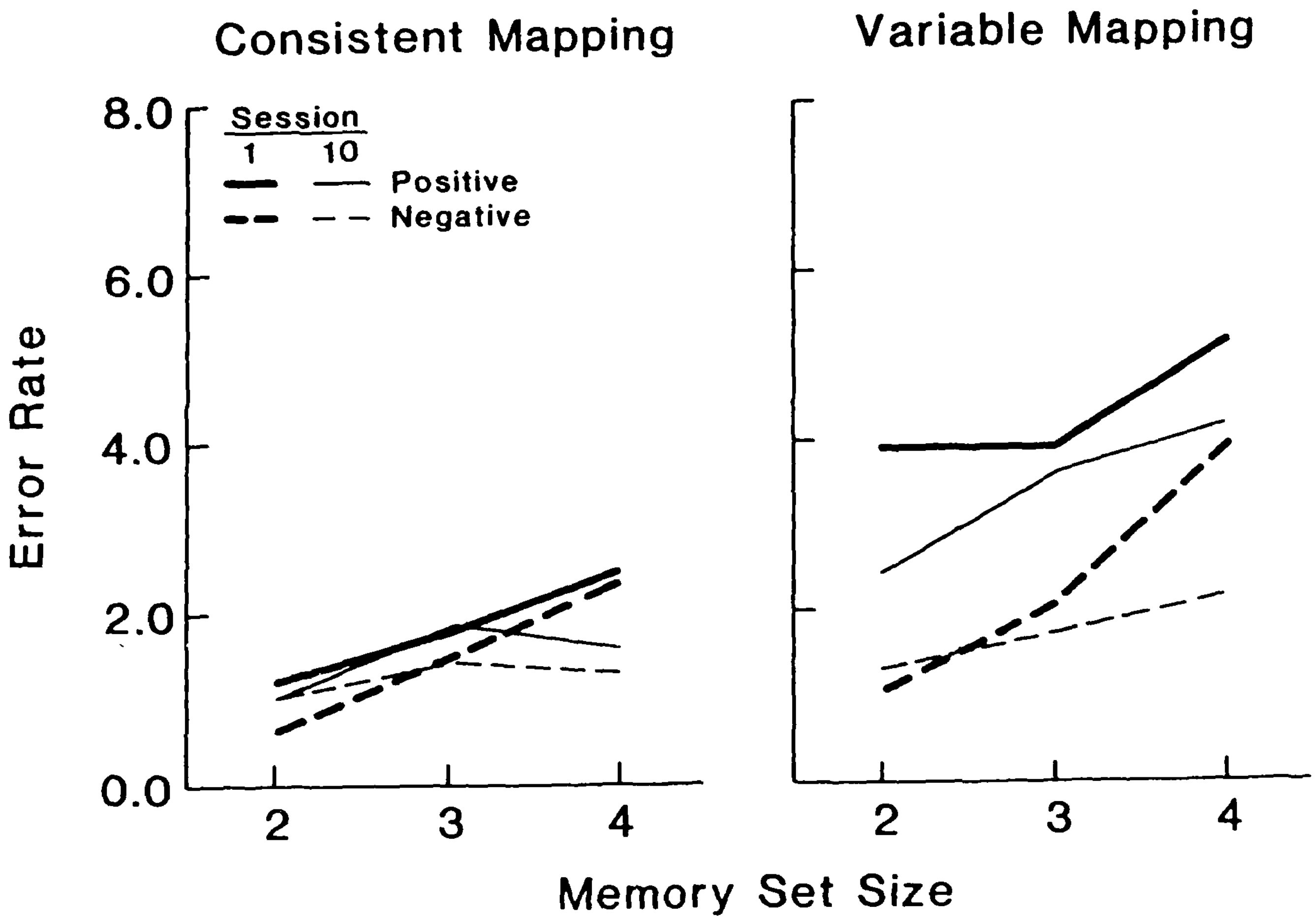
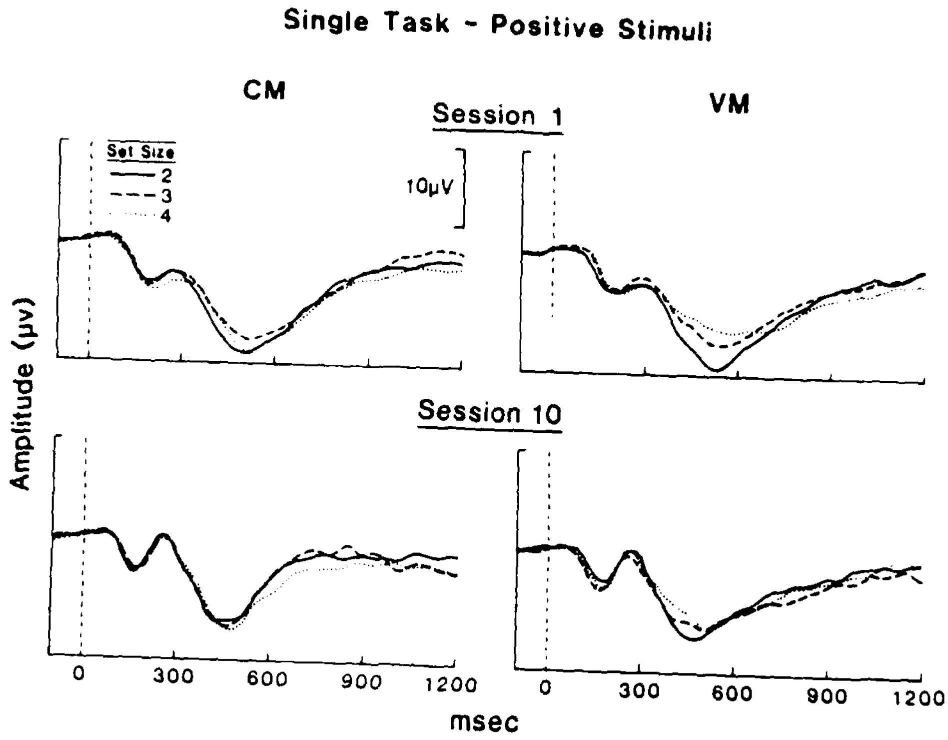
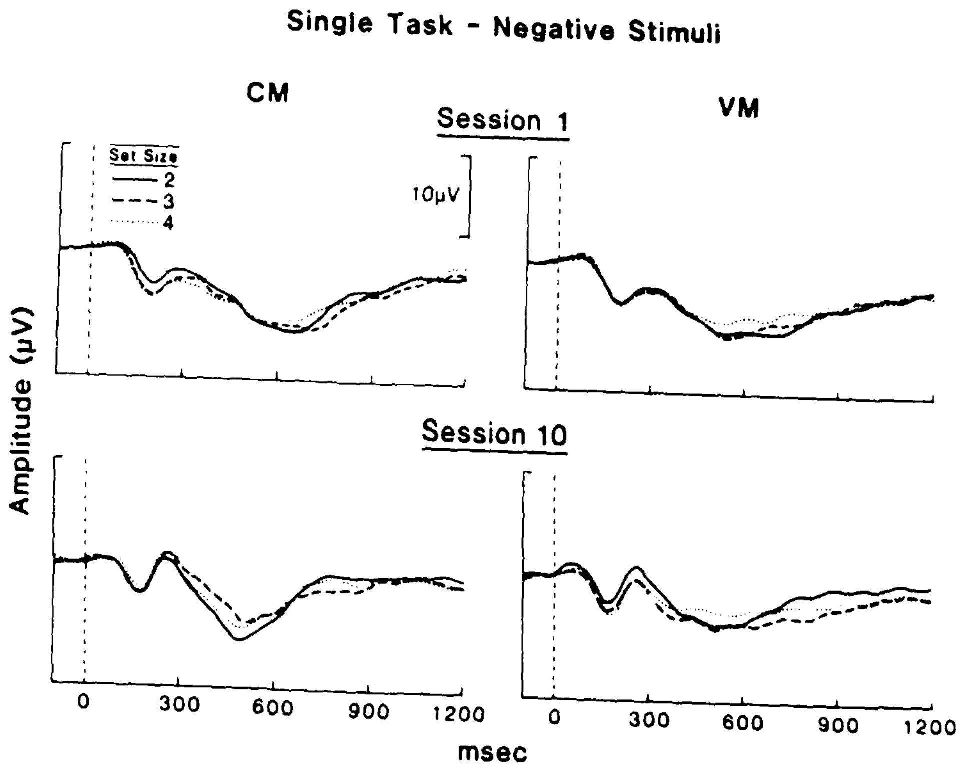

Consistent Mapping
Variable Mapping
Session
1
10
Positive
Negative
8.0
6.0
4.0
2.0
0.0
Error Rate
2
3
4
Memory Set Size

FIGURE 4



Single Task - Positive Stimuli
CM
Session 1
VM
Set Size
2
3
4
10μV
Amplitude (μv)
Session 10
0
300
600
900
1200
msec



Single Task - Negative Stimuli
CM
Session 1
VM
Set Size
2
3
4
10μV
Amplitude (μV)
Session 10
0
300
600
900
1200
msec

FIGURE 5

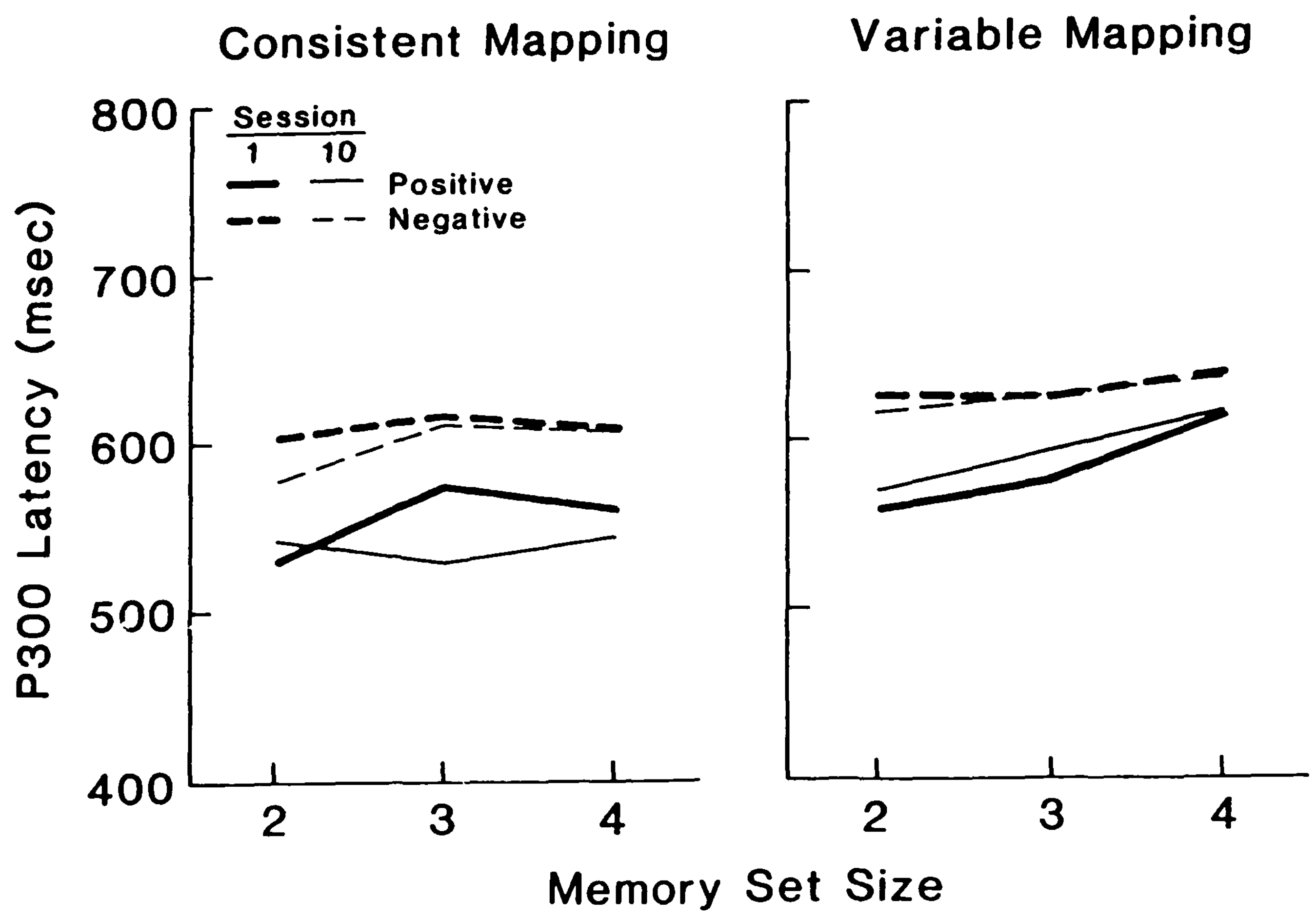

Consistent Mapping
Variable Mapping
Session
1
10
Positive
Negative
P300 Latency (msec)
800
700
600
500
400
2
3
4
Memory Set Size

FIGURE 6

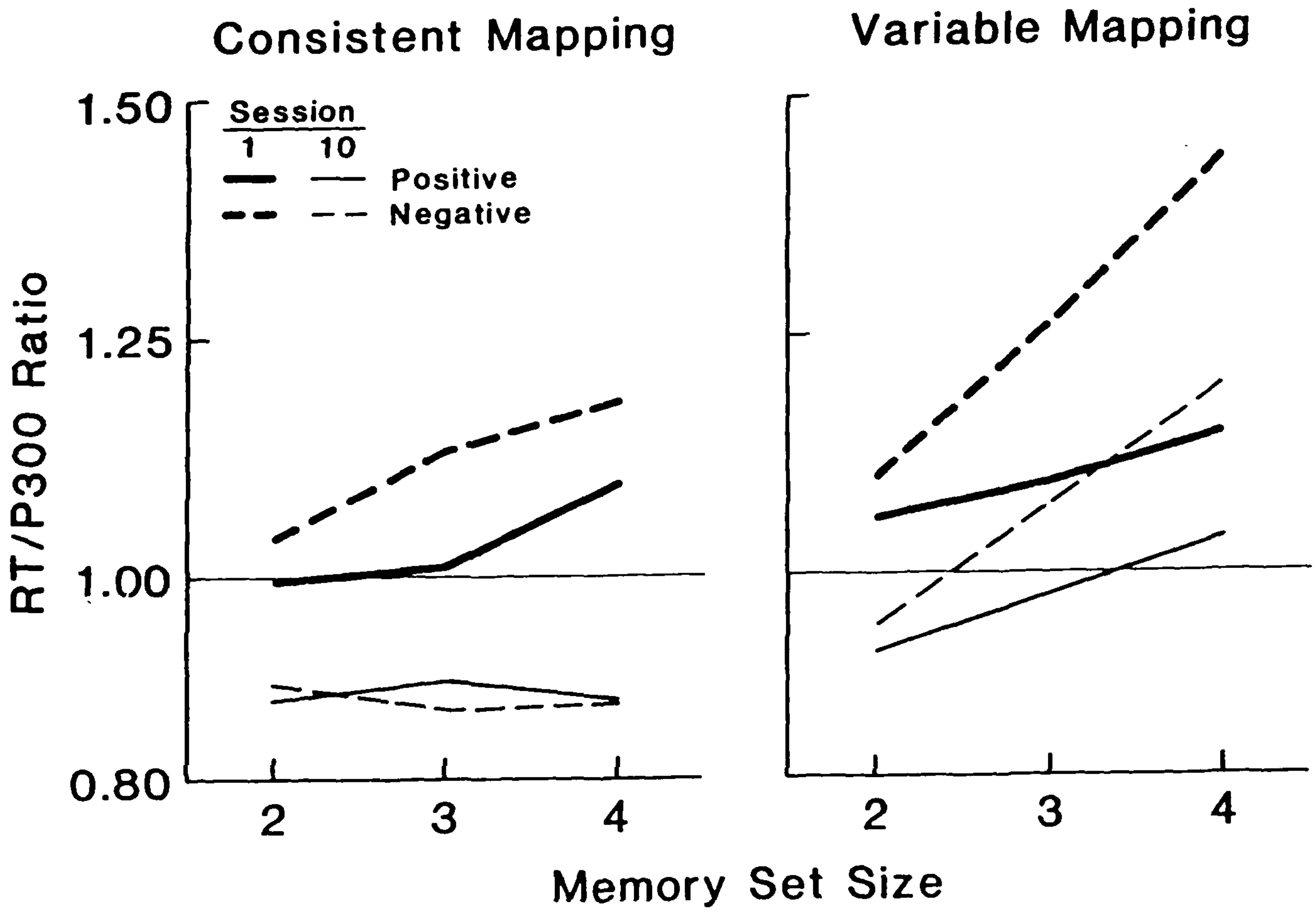

Consistent Mapping
Variable Mapping
Session
1
10
Positive
Negative
RT/P300 Ratio
1.50
1.25
1.00
0.80
2
3
4
Memory Set Size

FIGURE 7

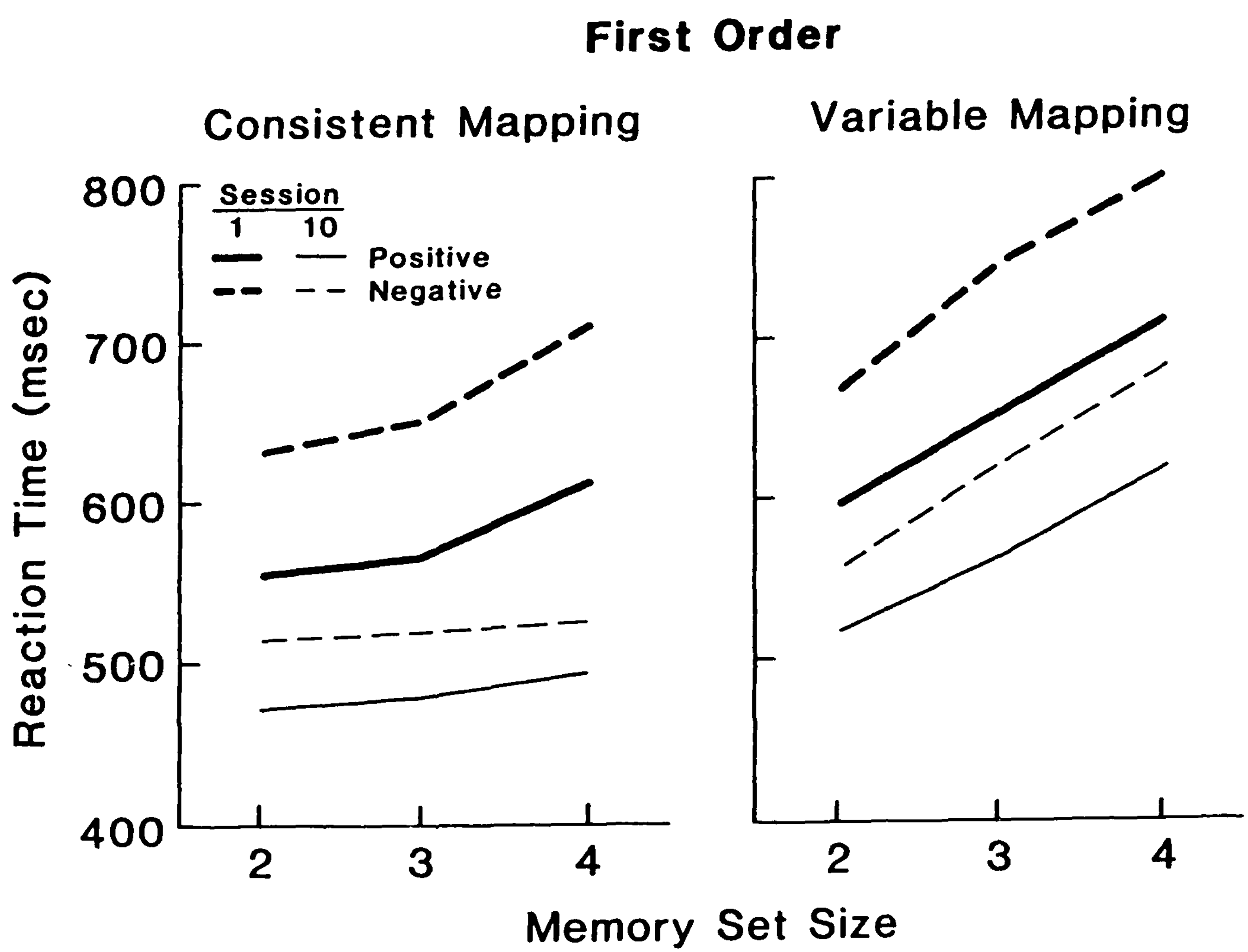

First Order
Consistent Mapping
Variable Mapping
Session
1
10
Positive
Negative
800
700
600
500
400
2
3
4
Reaction Time (msec)
Memory Set Size

FIGURE 8

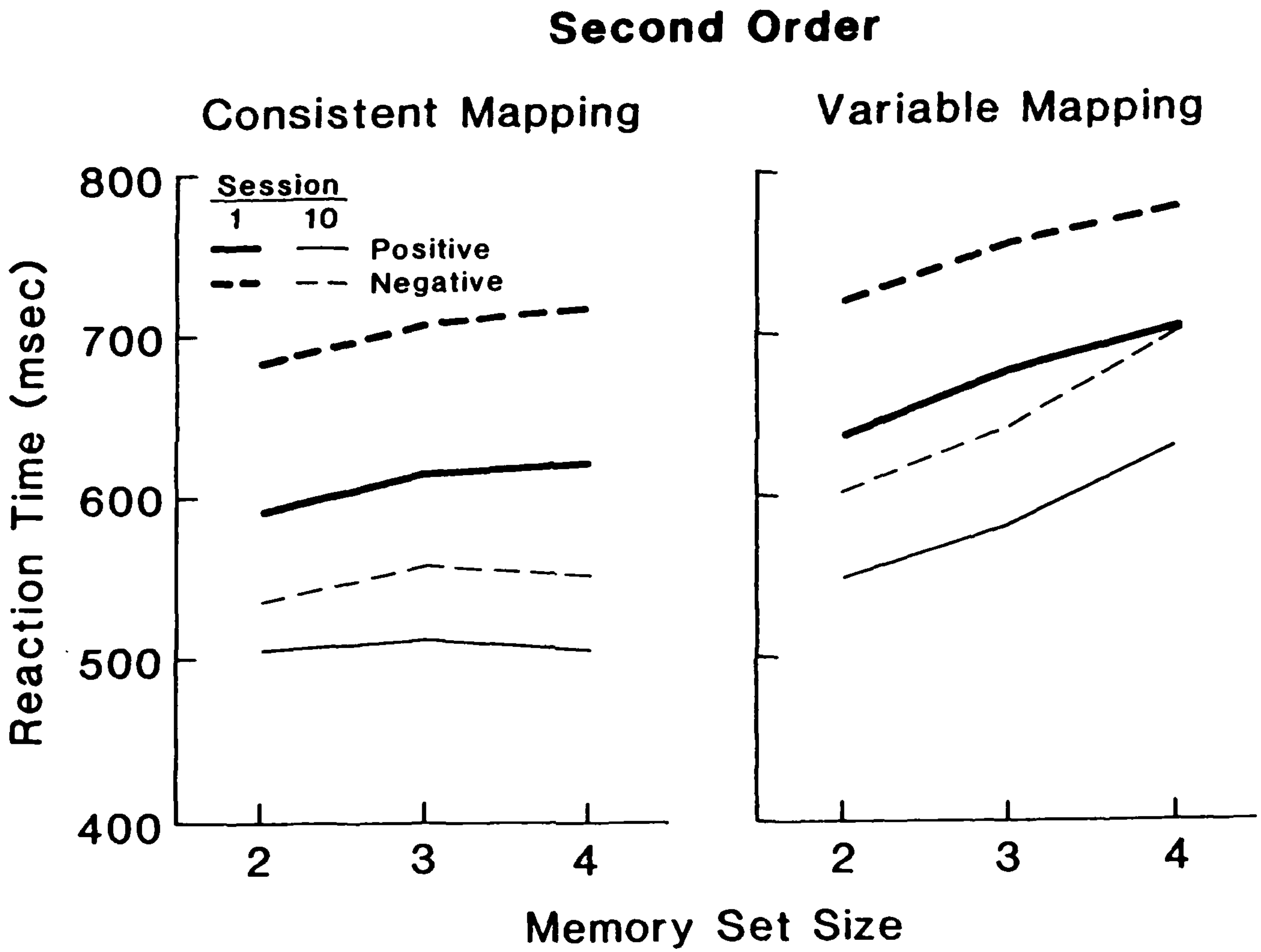

Second Order
Consistent Mapping
Variable Mapping
Session
1
10
Positive
Negative
Reaction Time (msec)
800
700
600
500
400
2
3
4
2
3
4
Memory Set Size

FIGURE 9

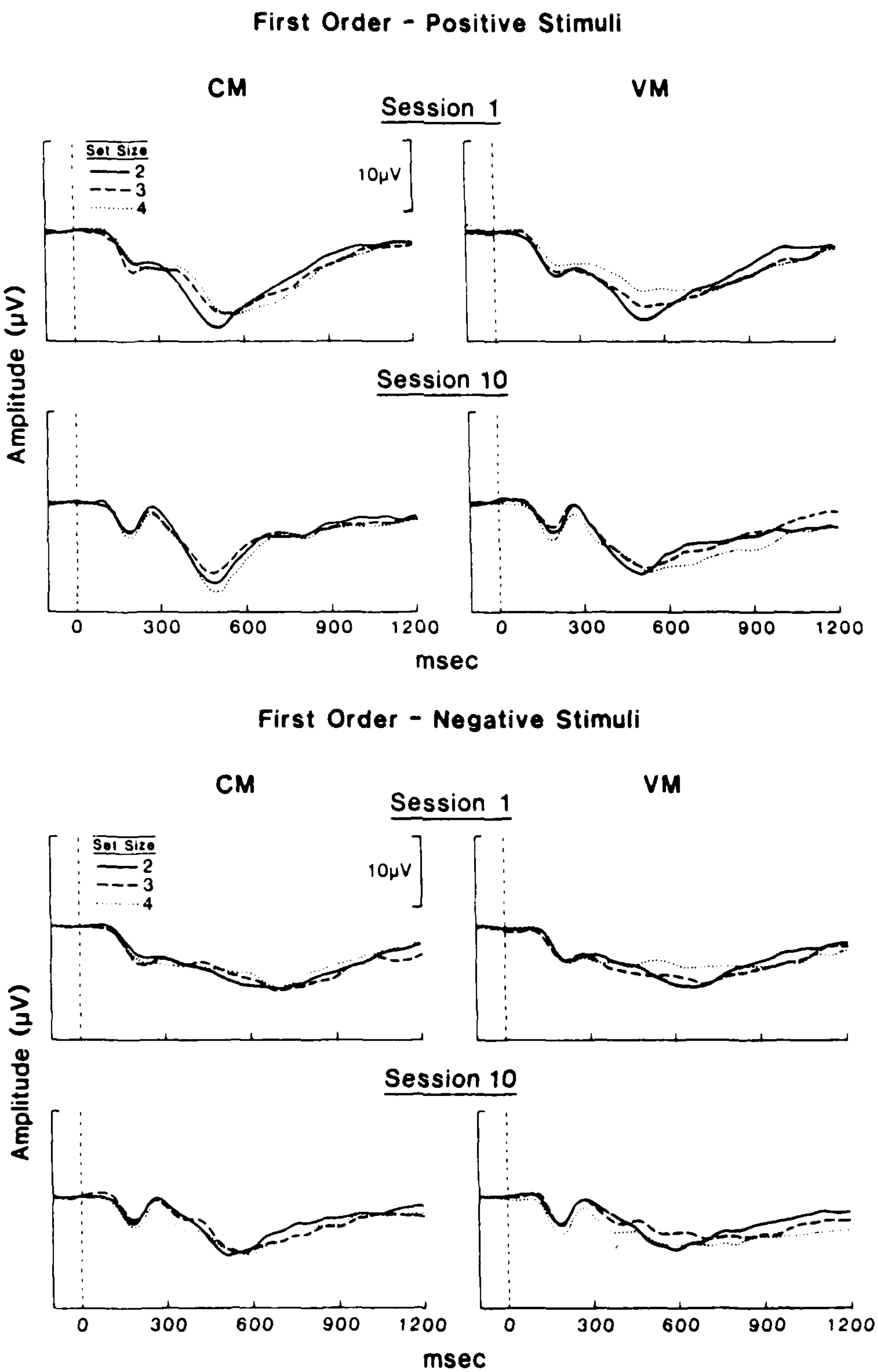

First Order - Positive Stimuli
CM
VM
Session 1
Set Size
2
3
4
10μV
Amplitude (μV)
Session 10
0
300
600
900
1200
msec
First Order - Negative Stimuli
CM
VM
Session 1
Set Size
2
3
4
10μV
Amplitude (μV)
Session 10
0
300
600
900
1200
msec

FIGURE 10

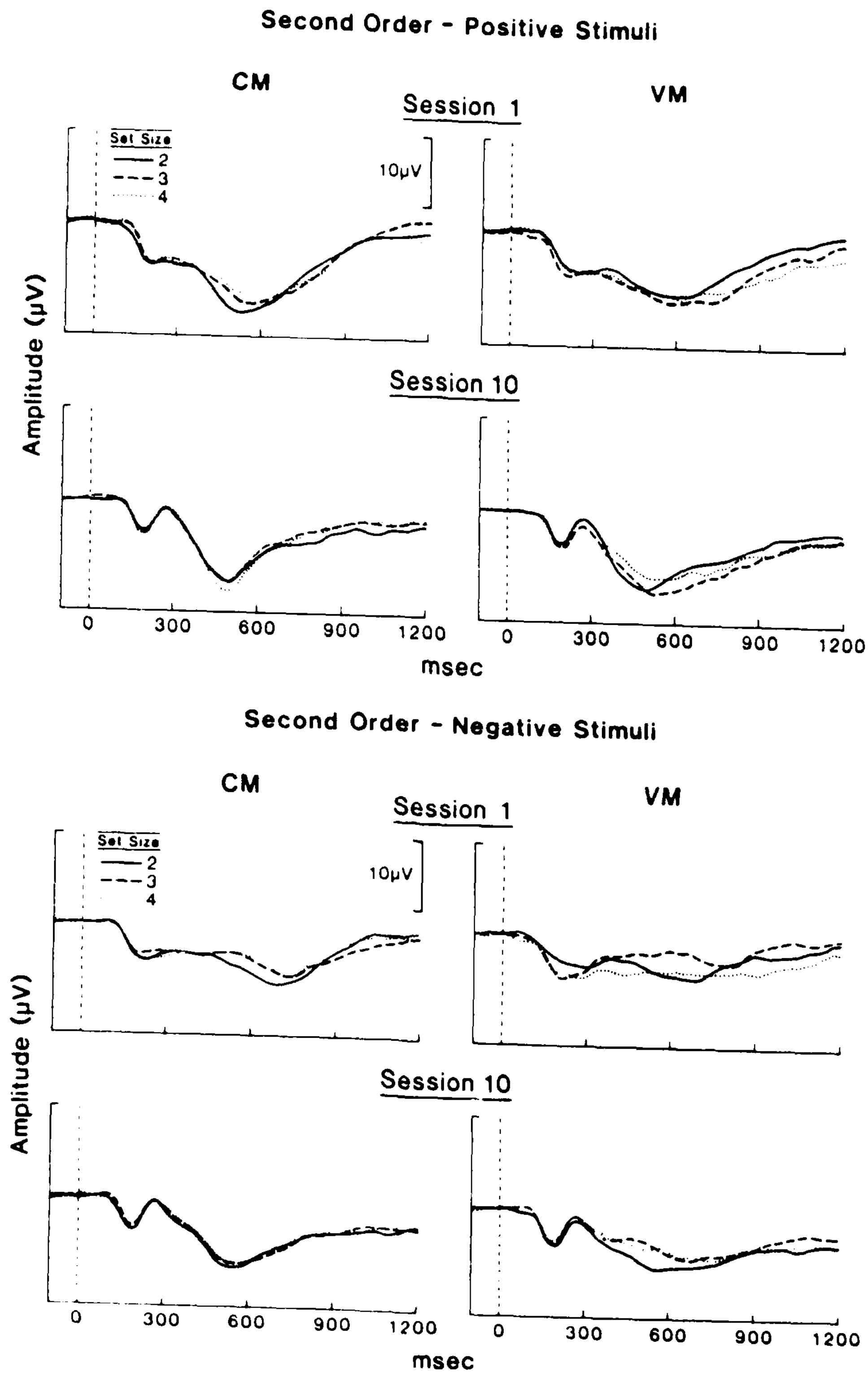

Second Order - Positive Stimuli
CM
VM
Session 1
Set Size
2
3
4
10µV
Amplitude (µV)
Session 10
0
300
600
900
1200
msec
Second Order - Negative Stimuli
CM
VM
Session 1
Set Size
2
3
4
10µV
Amplitude (µV)
Session 10
0
300
600
900
1200
msec

FIGURE 11

Article 1?.

ANNUAL REVIEW OF
Gerontology and Geriatrics

Volume 7, 1987

K. Warner Schaie, Ph.D.
SPECIAL VOLUME EDITOR

Carl Eisdorfer, Ph.D., M.D.
EDITOR-IN-CHIEF

SPRINGER PUBLISHING COMPANY
New York

1987

CHAPTER 1

Research in Geriatric Psychophys ،ogy

GREGORY A. MILLER
COGNITIVE PSYCHOPHYSIOLOGY LABORATORY
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, CHAMPAIGN, ILLINOIS

THEODORE R. BASHORE
DEPARTMENT OF PSYCHIATRY
MEDICAL COLLEGE OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA

LAWRENCE A. FARWELL
COGNITIVE PSYCHOPHYSIOLOGY LABORATORY
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, CHAMPAIGN, ILLINOIS

EMANUEL DONCHIN
COGNITIVE PSYCHOPHYSIOLOGY LABORATORY
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, CHAMPAIGN, ILLINOIS

Although psychophysiological data often bear on physiological questions, the present review will focus on what psychophysiology is able to say about the psychological aging of the individual, as distinguished (Birren & Cunningham, 1985) from biological and social aging. Much of the psychophysiological literature using elderly human subjects addresses one of two psychological questions, and this review is organized accordingly. First, what is the nature of cognitive changes with age? Second, what cognitive abnormalities develop in association with common psychiatric and neurological disorders in old age?

Work on this chapter was supported in part by AFOSR Contract #F49620-79-C-0233, Al Fregly, Project Manager. Noel K. Marshall, Ron D. Chambers, Frank Morrell, Leyla DeToledo-Morrell, and Thomas Hoeppner were collaborators during various phases of our own work with the aged, which was supported in part by NIH Grant AG03151. Kay Strayer's assistance in the execution of the project is greatly appreciated. The authors also wish to acknowledge the subjects who made the research possible.

Several emphases in this review should be noted at the outset. In terms of content, this review gives primary attention to an attempt to understand the locus of the pervasive cognitive slowing that comes with advanced age (Birren & Botwinick, 1955; Cerella, 1985; Salthouse, 1985a). This emphasis reflects the salience of the issue of slowing in the aging literature. Moreover, it is an issue that highlights the potential utility of psychophysiology for the study of aging, in that it exemplifies how psychophysiological measures can help dissect complex cognitive processes. In particular, the burgeoning literature on event-related brain potentials (ERPs) provides an impressive basis for specifying the stages of information processing that may mediate behavioral slowing, in partial disagreement with the reaction time (RT) literature.

Another emphasis will be most apparent in the discussion of psychophysiological research on psychiatric and neurological problems in the aged. As articulated elsewhere (Donchin & Bashore, in press; Donchin, Miller, & Farwell, 1986b), we suspect that the usual strategy of comparing samples of subjects recruited on the basis of clinical diagnosis places serious constraints on the richness, usefulness, and replicability of the data obtained. The present review will critique the available literature in light of this issue and will propose an alternative strategy of comparing subjects selected on the basis of psychophysiological functioning.

NORMAL AGING: COGNITIVE CHANGES

A number of components of the ERP have attracted interest in studies of healthy elderly adults. Although brain-wave measures are not necessarily more useful for studying cognition than are other physiological systems, most of the cognitive psychophysiological literature has employed ERPs. After a brief review of progress in other areas, this section will emphasize ERP evidence for the source of cognitive slowing in the elderly.

Electroencephalogram

The electroencephalogram (EEG) is the oldest measure of gross brain electrical activity. Whereas the ERP is recorded as a change in the ongoing EEG elicited by a specific event, the EEG has generally been recorded with subjects at rest or in some other static condition (e.g., Stage 2 sleep). EEG studies with the elderly have appeared steadily for some decades. It is well established that the dominant frequencies in the EEG decline with age (Goodin, 1985). Thus, in a state in which alpha waves (8–13 Hz) predominate, the peak frequency is likely to average around 10 Hz for young adults and 8 Hz for an older group (Obrist, 1979; but see Duffy, Albert, McAnulty, & Garvey, 1984). Attempts have been made over the years to relate this finding to a general decline in CNS arousal with advancing age. However, this picture is muddled, and the functional significance of this decline, if it exists, is not clear. These efforts have contributed little to our understanding of the aging process or of the psychological performance of the aged. Importantly, it has been suggested (Duffy et al., 1984; Katz & Horowitz, 1982) that much or all of the decline in peak alpha frequency is artifactual, reflecting the increasing prevalence of disease in elderly samples not carefully screened on health status.

In an earlier review, Marsh and Thompson (1977) suggested that there were already too many studies of EEG changes and aging, shedding too little light on fundamental issues. Their recommendations for future studies included the following: (a) assessment of the full EEG spectrum, not merely conventional EEG bands; (b) collection of longitudinal data; (c) measurement of EEG changes during specific tasks; and (d) evaluation of the contribution of CNS–ANS (autonomic nervous system) relationships to behavioral changes with age. In light of these needs, we may evaluate what has been accomplished in the decade since their review.

In general, interesting EEG research with the elderly in recent years has relied on new ways to combine and analyze data, rather than on innovations in conceptual approach. For example, Duffy et al. (1984) employed a 20-channel EEG recording array. As recommended by Marsh and Thompson (1977), they provided subjects with a series of 10 standard tasks in an attempt to characterize EEG changes with age. Their results were at variance with much of the previous literature in a number of ways, such as finding no effect of age on dominant alpha frequency and decreased rather than increased amounts of slower-frequency activity. On the other hand, many of their EEG measures correlated surprisingly well with neuropsychological measures of memory function. These investigators were also impressed with a decline in "alpha reactivity" found with increasing age. That is, alpha blocking, the difference between eyes-closed and eyes-open alpha levels, declined. In addition to the authors' suggestion that their unusual findings may result from the unusually good health of their older subjects (see also a subsequent section of the present chapter), it may be that the study's emphasis on active tasks makes comparisons with previous studies using subjects at rest inappropriate.

Dustman, LaMarche, Cohn, Shearer, and Talone (1985) recorded the EEG at rest from 40 young and 40 elderly adults. EEG power across the spectrum was less variable, as a function of recording site on the scalp, in the older sample. Furthermore, a measure of the EEG phase relationship between pairs of sites showed tighter coupling in this group. Dustman et al. (1985) interpreted their findings in terms of a decline in the functional autonomy of different brain regions with age. These data provide an interesting parallel to the increased uniformity across brain regions in the amplitude of the P300 measure, which will be discussed below.

Few of udies directly address the findings of Dustman et al. (1985). In a brief report, Giaquinto and Nolfe (1985) also noted greater consistency between brain regions in their elderly group. In contrast, Drechsler (1978) found decreased synchrony in EEGs of the elderly. Both of these papers provide a much less thorough data analysis and discussion than do Dustman et al. (1985), so it is difficult to integrate these studies. Of indirect relevance is a finding by Podlesny and Dustman (1982) that P300 amplitude, contingent negative variation (CNV) amplitude, heart rate deceleration, and reaction time were more strongly correlated in a young-adult sample than in an elderly group. Several papers have proposed that elderly subjects' slowed RT may result in part from reduced efficiency in the interaction between physiological systems (Marsh & Thompson, 1973, 1977; Thompson & Nowlin, 1973). This question of the existence and functional significance of multichannel synchrony warrants considerably more empirical attention before conclusions can be drawn because the field has not achieved consensus on how to quantify or draw conclusions about synchrony.

Increased or decreased consistency or linkage between brain areas is not necessarily problematic for the aging adult. It may be more or less appropriate, depending on the task at hand. It would be very useful, in line with Marsh and Thompson's (1977) recommendations, to repeat the Dustman et al. (1985) experiment with subjects engaged in a series of active tasks. This would be especially helpful if the tasks were chosen for an ability to foster differential regional activation of the brain, or if the tasks were known to benefit from greater or lesser regional independence. A finding of enhanced cross-regional coupling persisting in the elderly in the face of such challenges would greatly extend the data obtained at rest by Dustman et al. (1985). Such data would be important for addressing whether the increased coupling is related to a loss of functional capacity in the elderly. Duffy et al. (1984) collected appropriate data to address such questions but did not emphasize regional linkage in their analyses. Thus, such a study remains to be done.

Overall, limited progress has been made in the decade since Marsh and Thompson's (1977) recommendations. We can repeat their plea for more sophisticated experimental design. The few recent studies reviewed here are improvements over earlier papers, but more such work is needed. Furthermore, longitudinal studies are still lacking. Before additional EEG studies are pursued, however, a thoughtful reassessment of the questions to be addressed by EEG studies is sorely needed. Particularly in the absence of active task manipulations, the promise of continued EEG research appears limited. Conceptual development is also needed to guide future research; the generalized, nonspecific arousal notion that was popular in psychophysiological research decades ago has been thoroughly discredited (e.g., Davidson, 1978).

Contingent Negative Variation (CNV)

Beginning in the early 1970s, a series of papers reported on the contingent negative variation (CNV) in the aging population. The CNV is typically recorded as a negative-going voltage at the scalp that increases in the seconds during which a subject awaits an imminent event, expected to occur at a known time. It is associated with both general anticipatory processes to evaluate the event and with motor preparation to respond to the event (Rohrbaugh & Gaillard, 1983; Simons, in press). The CNV was one of the first ERP components to be assessed (see Marsh & Thompson, 1977, for a review).

Tecce, Cattanach, Yarchik, Meinbresse, and Dessonville (1982) tested a theory-based prediction about the CNV in the elderly. Older individuals have been found to be less able than young adults to switch attentional set (Birren, 1974). Tecce (1979) had found an enhanced CNV when subjects switch from a task with distraction to one without it. Tecce et al. (1982) hypothesized that the elderly might fail to show this CNV "rebound" if they were indeed less able to switch attentional focus. Results supported the authors' hypothesis for subjects aged 70 to 85 but not for those in the 55-to-69-year-old group. This CNV rebound failure in the oldest group resembled that reported for Alzheimer's patients (Tecce, Boehner, Cattanach, Branconnier, & Cole, 1981). In addition, the oldest group in the Tecce et al. (1982) study showed a deficient early CNV over the frontal lobes, also suggesting perseveration in this group. A final note of interest in this study is that, whereas P300 amplitude and latency differentiated the young group (18–32 years) from both older groups, CNV differentiated the oldest group from the other two groups, with the oldest group showing reduced CNV amplitude. Thus, CNV and P300 effects with advancing age not only are psychologically distinct but have their impact at different developmental stages.

Michalewski and associates (1980) found equivalent reduction in CNV amplitude for young and aged subjects when a distractor task was introduced, but, like Tecce et al. (1982), they noted that frontal CNV was reduced, across conditions, for the aged group. They suggested that this site-specific effect may indicate regional differences in the aging process. On the other hand, Podlesny and Dustman (1982) reported that CNV amplitude increased with age, which they interpreted as possibly reflecting weakening inhibitory function. In sum, the picture for simple CNV effects with age is unclear. It may be that, rather than being a main effect for age on the CNV, the aging process interacts with yet-to-be-specified variables in determining changes in the CNV.

CNV research with the elderly is not widely pursued at present. However, in light of new evidence that healthy older subjects in excellent physical shape may show few of the changes believed to be inevitably age-related (see below), the CNV may yet prove to be a valuable tool in studies of movement and cognitive preparation for movement.

P300

The P300 component of the ERP is the most frequently studied measure in recent psychophysiological studies of the elderly. This positive-going deflection of brain electrical activity occurs 300 msec or more after an event that is rare and/or task-relevant, particularly when the event requires the individual to update his or her working model of the current environment (Donchin, 1981; Donchin et al., 1986b). The increasing emphasis on the P300 is perhaps the best example of a trend noted a decade ago in a review of the aging/psychophysiology literature—"the shift from rather slowly responding parameters, such as galvanic skin response or skin temperature, to experiments that analyze rapid onset phenomena" (Storrie & Eisdorfer, 1978, p. 1489).

The appeal of the P300 lies partly in the consistency with which researchers have found its latency to increase with age and partly in its demonstrated relationship to relatively high order cognitive processes. The established cognitive significance of changes in P300 contrasts markedly with the "slowly responding parameters" originally favored, the cognitive significance of which is less richly established. Most specifically, explanations for the slowing of behavior with age that rely on peripheral processes or that implicate generalized CNS processes are challenged by data on the slowing of P300 latency with age. P300 research has now advanced to the point that it supports a significant revision in prevailing views on the basis of the slowing of behavior.

Findings from a series of studies in Arnold Starr's laboratory are generally credited with sparking the high interest in the P300 for studies of age-related changes in cognition. This early work sought to identify a scalp-recorded brain event that could contribute to the differential diagnosis of dementing disease, and this work has provided the impetus for studies of normal aged populations in which more challenging cognitive tasks have been performed. (A detailed review of this literature is available in Bashore, in press, summarized here; for earlier reviews, see Ford & Pfefferbaum, 1980, 1985; Goodin, 1985; Marsh & Thompson, 1977; Polich & Starr, 1984; Squires, Chippendale, Wrege, Goodin, & Starr, 1980; Squires, Goodin, & Starr, 1979.)

In the first study from Starr's group (Goodin, Squires, Henderson, & Starr, 1978a), neurologically normal subjects ranging in age from 6 to 76 years were tested in an auditory "oddball" task. In this task, rare tones of one pitch occur randomly in a train of tones of another pitch. A series of response components is readily distinguished in the ERPs obtained from this task, labeled according to voltage polarity and approximate latency (e.g., N100, P300). Enhanced P300s are reliably observed in response to the rare tones if the subject is asked to attend to (e.g., to count) them. Goodin et al. (1978) found systematic, linear increases in the latency of P200 (0.7 msec/year), N200 (0.8 msec/year), and P300 (1.8 msec/year) beginning at age 15, as well as concomitant reductions in the amplitude of P200 (0.2 μV/year, measured as the N100–P200 peak-to-peak value) and P300 (0.2 μV/year, measured as N200–P300). The latency of N100 was not found to change with age (N100 amplitude was not measured). Thus, age-dependent, linear changes in both amplitude and latency of components later than N100 were noted in this initial study. Moreover, the later the component, the larger the increase in latency seen with age, and the largest was for P300.

The results from this initial study have been replicated to a large degree in subsequent studies. That is, age-related increases in latency have been reported consistently in auditory oddball tasks for P200, N200, and P300 but not for N100 (e.g., Brown, Marsh, & LaRue, 1983; Donchin et al., 1986b; Ford & Pfefferbaum, 1985; Goodin, Squires, & Starr, 1978b; Picton, Stuss, Champagne, & Nelson, 1984; Syndulko et al., 1982). There have been some reports of failures to find age-related latency effects, at least in some paradigms (Michalewski, Patterson, Bowman, Litzleman, & Thompson, 1982; Picton et al., 1984; Podlesny & Dustman, 1982; Snyder & Hillyard, 1979). However, by far the majority of studies have demonstrated age-correlated slowing, and none has reported a decrease in P300 latency with age.

Amplitude reductions in P300 have also been reported in a number of studies (Brown et al., 1983; Donchin et al., 1986b; Goodin et al., 1978; Mullis, Holcomb, Diner, & Dykman, 1985; Picton et al., 1984; Podlesny & Dustman, 1982; Polich, Howard, & Starr, 1985), but this consistency is lacking for N100, P200, and N200 amplitude. Several studies also dispute the P300 amplitude decline with age (Beck, Swanson, & Dustman, 1980; Ford, Pfefferbaum, Tinklenberg, & Kopell, 1982; Pfefferbaum, Ford, Roth, & Kopell, 1980a). However, to some extent this may be an artifact of a more intriguing finding: a shift in the scalp distribution of the P300. As reported in the original paper by Goodin et al. (1978a), the amplitude appears to be reduced at the parietal lead and maintained or increased at the frontal lead with advancing age (Donchin et al., 1986b; Ford & Pfefferbaum, 1985; Mullis et al., 1985; Pfefferbaum, Ford, Roth, & Kopell, 1980b; Pfefferbaum, Ford, Wenegrat, Roth, & Kopell, 1984a; Picton et al., 1984; Smith, Michalewski, Brent, & Thompson, 1980; Tecce et al., 1982). Clear exceptions to this frontal shift have rarely been reported (e.g., Duffy et al., 1984, who used an unusual index of P300 amplitude).

Emerging from this research is some controversy about the form of the relationship between age and P300 latency. Whereas Goodin et al. (1978a) reported a linear increase in component latency, Brown et al. (1983) identified a curvilinear pattern, in which positively accelerating increases were apparent at age 45 for P200, N200, and P300, although little change was apparent in these components prior to that age. A reanalysis by these investigators of the P300 latency data from Goodin et al. (1978a), Polich et al. (1985, in preparation at the time), and Syndulko et al. (1982) identified similar age-dependent curvilinear trends in the latter two studies. Picton et al. (1984) failed, however, to find any evidence for curvilinearity in their P300 data.

The findings from these studies can be summarized as follows: Age-related increases in the latency of components beyond 100 msec poststimulus occur, covarying with component latency so that, in general, the later the component, the larger the age-related increase. However, although we can conclude with confidence that the latencies of certain components of the ERP do slow with age, whether this slowing occurs at a constant rate over time or accelerates at a certain point in the life span is unclear. In contrast to the consistency observed in component latency, changes in the amplitude of the various components are observed less consistently. Finally, there appears to be a topographic shift with age, an equalization of P300 amplitude across recording sites. Whether this shift in measured P300 amplitude is truly a change in the neural generator(s) of the classic P300 or a change in the generator(s) of some other, overlapping component remains to be determined.

A number of other P300 oddball studies have been conducted with elderly subjects in recent years, using visual stimuli (e.g., Beck et al., 1980; Donchin et al., 1986b; Picton et al., 1984; Snyder & Hillyard, 1979) and omitted stimuli (Donchin et al., 1986b; Michalewski et al., 1982; Picton et al., 1984). Some studies have added an RT requirement (e.g., Donchin et al., 1986b; Ford & Pfefferbaum, 1985; Mullis et al., 1985; Pfefferbaum et al., 1980b; Pfefferbaum et al., 1984a; Picton et al., 1984; Picton, Cerri, Champagne, Stuss, & Nelson, 1986; Podlesny & Dustman, 1982; Podlesny, Dustman, & Shearer, 1984). Conclusions to date are similar to those for the oddball studies that emphasized auditory stimuli and "count" tasks: a consistent (though not invariant) increase in P300 latency, as well as mixed evidence of amplitude decrement, with advancing age.

Two studies of P300 and RT in the elderly deserve extended review here, as they exemplify ERP research that goes beyond description of age-related changes and tests a cognitive hypothesis about aging. Ford, Duncan-Johnson, Pfefferbaum, and Kopell (1982) applied a sequential analysis to P300 amplitude, P300 latency, and RT data obtained in a two-choice RT task, to test the hypothesis that older individuals do not utilize prior information as effectively as do young individuals to anticipate subsequent stimulus events (Rabbitt & Vyas, 1980). This analysis technique had been used originally with RT data to infer the effect of expectancy on response speed (e.g., Kirby, 1976). In this analysis, the pattern of response latencies is assessed over a sequence of trials, typically five. It was commonly observed that repetitions of a stimulus produced faster RTs than did discontinuations and that, as the number of repetitions increased, RT became a bit faster. The interpretation offered has been that repetitions of stimulus events induce subjects to anticipate subsequent repetitions (i.e., establish a response set), confirmations of which produce faster RTs than do disconfirmations because the correct response has been prepared.

Squires, Wickens, Squires, and Donchin (1976) were the first to apply this analysis to ERP data. An important element of this study was its demonstration of the crucial role that subjective expectancy (rather than objective probability alone) plays in determining P300 amplitude. Squires et al. (1) showed that a larger P300 was produced when the preceding stimulus differed from the counted stimulus; and the longer the sequence of different stimuli that preceded the counted stimulus, the larger the P300 elicited by that stimulus. Conversely, the longer the sequence of similar stimuli (i.e., those that were also counted) preceding a counted stimulus, the smaller the P300 elicited by that stimulus. Thus, the pattern of amplitude change for P300 paralleled that reported for RT. These findings have been confirmed and extended in a number of subsequent studies.

It is important to note that Ford et al. (1982) found that the effects of stimulus sequence on P300 latency and amplitude were comparable in young and elderly subjects. Thus, P300 amplitude was larger in both groups in response to stimuli that discontinued rather than continued a stimulus sequence, and both P300 latency and RT were prolonged by discontinuations of a stimulus event. Discontinuations of a stimulus sequence produced larger P300s in the young and the old subjects, but the change was more pronounced in the older subjects. This disparity in P300 amplitude effects was not apparent when expectations were confirmed. Ford et al. suggested that disconfirmations of expectancies disrupt the processing efficiency of the elderly more than they do that of the young, but that confirmations of expectancies do not benefit the elderly more than the young. These findings imply that older subjects establish expectancies for future stimulus events from earlier events in much the same way as do young subjects and that they may be even more sensitive to prior stimulus events than are young subjects, contrary to the hypothesis of Rabbitt and Vyas (1980).

Several consistent findings have emerged from the P300/RT studies:

1. P300 latency is prolonged with age, and the increase is comparable to that reported in the oddball studies (1.0–2.0 msec/year).
2. Unlike the age-related decrease in P300 amplitude observed in the oddball studies, when an overt response to an auditory stimulus is required, there is no evidence of amplitude diminution with age.
3. The scalp distribution of P300, as reported in the oddball studies, is equipotential along the midline in the old, whereas it is maximal at Pz and reduces frontally in the young.
4. This age-related variation in the topography of P300 is paralleled by a reduction in the frontal negativity of the slow wave in the elderly (recall also the reduced slow frontal negativity reported by Tecce et al., 1982).
5. The elderly utilize prior stimulus information to establish expectancies for future events in much the same way as do the young.
6. RTs are comparable among the young and the old in these reasonably uncomplicated tasks. This last finding is notable, since studies in which RT is the only dependent measure of processing speed have consistently reported a slowing of response speed with age at all levels of processing

complexity (Birren, Woods, & Williams, 1979; Cerella, 1985; Kausler, 1982; Salthouse, 1985a, b), although there is strong evidence that the age effect on RT increases for more complex tasks (Cerella, Poon, & Williams, 1980).

Comparability of response latency in old and young subjects is typically reported only when the old subjects are in very good health and keep fit through regular exercise (see below). In most of the ERP/RT studies, subjects have been screened medically; whereas in the RT-only studies, health status questionnaires were typically used to screen candidates. It has now been documented that such questionnaires probably provide unreliable information (Siegel, Nowlin, & Blumenthal, 1980). It may very well be that healthier subjects were used in these ERP studies and that this is reflected in the comparability of response latencies across age groups. Indeed, this issue of underassessed health status is emerging as a significant confound in the psychophysiological literature on the aged (see below).

P300/RT Relationships in Memory Function

The most concrete contributions of ERP research to gerontology have emerged from five studies on memory scanning, P300 latency, and RT. Taken together, this work provides a basis for revising notions of the locus of cognitive slowing in the elderly. A sixth study suggests a specific memory deficit in those elderly individuals with a particularly small or even absent P300. In the five memory scanning studies, Sternberg's (1969) task was chosen to assess age-related declines in the speed of memory search (Ford, Roth, Mohs, Hopkins, & Kopell, 1979; Ford et al., 1982; Marsh, 1975; Pfefferbaum et al., 1980a; Strayer, Wickens, & Braune, 1987). The sixth study adapted a clinical test of memory function for use in the ERP laboratory (Donchin et al., 1986b; Farwell, Chambers, Miller, Coles, & Donchin, 1985).

In the Sternberg task, the subject is shown a list of stimuli such as the digits 1 to 9, one stimulus at a time. This list, the memory set, typically varies in length from one to six items. Following presentation of the last item in the memory set, the subject is presented a single stimulus, the test or probe item, and is required to make a choice RT response indicating whether or not the test item was in the memory set for that trial. Subjects are encouraged to place equal emphasis on speed and accuracy, and the task is designed for perfect recall (Sternberg, 1969). There are two basic versions of the task: fixed and varied set. In the former, the subject is shown the same memory set items on each trial in a block of trials. In the latter, the memory set items are different on each trial.

Sternberg (1969, 1975) and many others have shown that RT to the test item is a linear function of the number of items in the memory set and that the increase is

approximately 40 msec per item in young adults. The slope of this function was hypothesized by Sternberg (1969) to reflect the time needed to compare the test item with each item held in memory, what he has labeled "serial comparison time." Negative responses typically have longer latencies, but the slopes of the two functions are parallel. According to Sternberg (1969), this process of serial comparison is exhaustive. That is, the subject is presumed to scan through the entire memory set for both positive and negative items, irrespective of the location of the positive test item in the memory set. The intercept of the regression function is hypothesized to be an aggregate time comprising stimulus encoding, binary decision, and response translation and organization times, independent of comparison time.

Sternberg's (1969) componential analysis of the regression function derives from his assumption that the four stages hypothesized to mediate memory scanning are activated sequentially. Thus, a stage is not engaged until it receives the output of the processing completed at the stage immediately prior. In this model, the value of the intercept is the sum of the processing times of the three stages it includes, and the total RT is the sum of the times required for each stage to complete its processing. This assumption of serially engaged stages of processing gave rise to Sternberg's (1969) development of the Additive Factors Method, to isolate the stages influenced by different experimental manipulations. Experimental variables are inferred to activate different stages of processing if they produce main effects (i.e., are additive) and to activate the same stage if they produce interactive effects.

Although not unchallenged (e.g., Grice, Nullmeyer, & Spiker, 1982), these theoretical and research assumptions have had a major influence on cognitive psychology (see Lachman, Lachman, & Butterfield, 1979). Investigations in which the memory scanning time of young and old subjects has been compared reveal that older subjects have longer RTs, larger intercepts, and steeper slopes than do young adults (Anders & Fozard, 1973; Anders, Fozard, & Lillyquist, 1972; Eriksen, Hamlin, & Daye, 1973; Madden, 1982; Madden & Nebes, 1980; Maniscalco & DeRosa, 1983; Salthouse & Somberg, 1982a, 1982b). These findings have led to the conclusion that memory scanning is slower at all stages of processing in older adults.

Importantly, however, a different conclusion can be drawn from the ERP studies of age-related changes in memory scanning. The five psychophysiological studies completed to date have used a varied-set version of Sternberg's task to study age-related declines in memory scanning (Ford et al., 1979, 1982b; Marsh, 1975; Pfefferbaum et al., 1980a; Strayer et al., 1987). Although there is some disagreement in the findings from these studies, sufficient consistency exists to permit the additive-factors logic of Sternberg (1969) to be applied to the P300 latency data. First, the RT data are generally consistent with those from the RT literature for both the young and the old. That is, RT increases with increases in memory set size, it is longer in the old, and the slope is steeper and t... intercept of the RT-memory set size regression function larger in the old than in the young. Second, the slope of the P300 latency-memory set size regression function is comparable in the two groups, whereas the intercept of this function is larger in the old than it is in the young. Third, the interval between P300 latency and the response output is larger in the old, and the slope is steeper and the intercept is larger in the old for the RT–P300 latency difference regression function.

From these findings it is possible to partition the memory scanning process with greater precision than can be accomplished using RT measures alone. As noted above, Ford et al. (1979) were the first to do this analysis, and their paper provides an elegant example of this procedure. Their contribution was to use P300 latency as a measure of the timing of stimulus-related processing. The rationale for this was provided by the evidence that suggested the relationship between P300 latency and stimulus processing (e.g., Kutas, McCarthy, & Donchin, 1977).

In the Sternberg model, the slope of the RT-memory set size regression function is an estimate of the time taken at the serial comparison stage, and the intercept of this function is the sum of the times taken to complete stimulus encoding, to make the binary decision, and to activate the correct response. However, Ford et al. (1979) reasoned that the slope of the P300 latency-memory set size function, rather than that of the RT-memory set size function, provides an estimate of serial comparison time. They also argued that the intercept of the P300 latency function gives an estimate of the time required to encode the stimulus information. They reasoned further that the time between the peak of the P300 and the response output was a measure of the relative timing of response-related processes. The slope of this regression function was assumed to provide an estimate of binary decision time, whereas its intercept was assumed to reflect the time required to translate and organize the response.

Ford et al. (1979) found that the slope of the RT-memory set size function was steeper and the intercept larger in the old than in the young. As in the RT function, the intercept was larger in the old. Unlike the RT function, however, the slope of the P300 latency-memory set size function was comparable in the two groups. As would be expected from these data, the RT–P300 latency-memory set size function had a slope that was steeper and an intercept that was larger in the old. These findings led Ford et al. to the following conclusions: (1) Serial comparison time does not increase with age (P300 latency slope); (2) stimulus encoding processes are somewhat slower in the old than in the young (P300 latency intercept); (3) response-related processes are much slower in the old than in the young (RT–P300 latency intercept); (4) the old are less confident in their response selection decisions as task difficulty increases (RT–P300 latency slope).

Strayer et al. (1987) also fractionated memory scanning utilizing RT and P300 latency data, and they added another refinement by analyzing speed–

accuracy tradeoff functions for the RT data. Their data largely ... ated those of Ford, Roth, et al. (1979). Addition of the speed–accuracy tradeoff analyses suggested that the delays in response-related processes evident in the elderly could be attributed in part to their adoption of a more conservative response criterion. Strayer et al. argued that three components contributed equally to the overall slowing of the older subjects. The first, evident in the elevation of the P300 latency intercept, was inferred to be related to stimulus encoding. The second, indexed by the steeper slope of the RT function (they did not analyze the RT–P300 latency interval), was thought to be related to slowing of response-related processes with increases in memory set size (not to serial comparison because of the comparability of the P300 latency slopes). The third, derived from the speed–accuracy analysis, was thought to be produced by shifts in the response criterion (i.e., the older subjects were more conservative).

Importantly, the results of these two studies suggest that age-related slowing in memory scanning occurs at both stimulus encoding and response-related stages of processing. *The speed of serial comparison processes appears to be maintained with age*. Response-related slowing may be contributed to by strategy differences and by changes in the speed with which responses can be selected and executed, the relative contributions of which are unknown. The work of Ford et al. (1979) and of Strayer et al. (1987) provides elegant examples of how mental processes can be articulated utilizing ERP/RT measures. In the next section, a refinement of a model of age-related slowing in the rate of mental processing will be developed, using the rationale that guided these ERP researchers.

P300, RT, and the Slowing of Behavior

On the basis of independent analyses, both Salthouse (1985a, b) and Cerella (1985; Cerella et al., 1980) concluded that age-related slowing in response speed is multiplicative and that it is mediated by declines in the efficiency of CNS transmission. Both concluded that the decline is generalized across levels of central information processing. An assumption underlying these analyses is that RT is a reasonably pure measure of general nervous system integrity and as such can be decomposed into peripheral and central constituents using the appropriate experimental and quantitative procedures. However, our own analysis of P300 latency data (Bashore, in press) demonstrates an age-related increment in higher-order processing time that is addictive, not proportional. Since the timing of brain electrical activity provides a more direct measure of higher-order central engagement than does the timing of response output, we would argue, in agreement with Ford and Pfefferbaum (1985), that P300 latency provides a purer measure of age-related changes in higher-order CNS processes than does response latency. It follows, in contrast to the conclusions of Cerella (1985) and

Salthouse (1985a, b), that only certain elements of central information processing decline with age.

To pursue this view more fully, Bashore (in press) aggregated P300 latency and RT data from 33 studies that included 64 different experimental conditions in which age-related changes were analyzed. Importantly, the slopes of the RT and RT–P300 latency regression functions suggest that the age-related decline in mental processing speed is multiplicative, whereas that for P300 latency indicates an additive relationship. How can this difference be reconciled? Previous research has demonstrated that RT is extremely sensitive to variations in response strategy (e.g., Wickelgren, 1977), whereas P300 latency is reasonably insensitive (e.g., Kutas et al., 1977; Pfefferbaum et al., 1983; Strayer et al., 1987). There is compelling evidence that older subjects adopt more conservative response strategies than do young subjects (Salthouse, 1982; Strayer et al., 1987; but see Ford & Pfefferbaum, 1985, for a counterargument), and this conservatism may increase with increases in task complexity (Ford et al., 1979; Pfefferbaum et al., 1980a). Hence, the slope functions for these two measures may reflect changes in response strategy as processing complexity is increased; that is, nonanatomic factors contribute to the slope effect. Since P300 is generated from a neural source and its timing is not very sensitive to shifts in response strategy, we can infer that its latency reflects the engagement of central mechanisms uncontaminated by nonneural variables. In sum, rather than a generalized slowing of CNS function, the P300/RT data indicate that the "slowing of behavior" in the elderly is confined to response strategy and execution stages.

Some speculations can be offered to account for the additive linear function for P300 latency. It is conceivable that this additivity represents the sum of earlier processing times that may or may not be differentially changed with age. Since the literature indicates that N100 latency does not become prolonged with age (another ERP basis for arguing against the generalized CNS slowing hypothesis) and that its amplitude and latency vary comparably in the old and the young when a selective attention task is performed (Ford, Hink, Hopkins, Roth, Pfefferbaum, & Kopell, 1979), it is reasonable to hypothesize that components between N100 and P300 may be differentially affected by age when a rapid decision is made. Potential candidates are the Na and N200 components investigated in studies by Ritter, Vaughan, and colleagues (Ritter, Simson, Vaughan, & Friedman, 1979; Ritter, Simson, Vaughan, & Macht, 1982; Ritter, Vaughan, & Simson, 1983), which appear to reflect pattern recognition and stimulus categorization processes, respectively, engaged during the decision-making process itself. In contrast, the timing of P300 may be more closely related to the outcome of this prior processing, the function of which is to prepare for stimulus processing on future trials (e.g., Donchin, Karis, Bashore, Coles, & Gratton, 1986a). Thus, the timing of these earlier components may be more reflective of age-related declines in the speed of mental processes activated for immediate action, where the timing of the P300 may be more closely tied to the

engagement of processes essential to the performance of later, memory-dependent functions such as recognition of previous stimulus events (Johnson, Pfefferbaum, & Kopell, 1985).

P300 and Maintenance of Working Memory

Whereas the five P300 *latency* studies employing the Sternberg memory scanning task suggest a new view of the role of CNS changes in the *speed* of memory function and behavior with age, a recent study in our laboratory investigates the relationship of P300 *amplitude* to the *quality* of memory function in the elderly (Donchin et al., 1986b; Farwell et al., 1985).

Our study of P300 in the elderly included, as a first stage, the assessment of 53 healthy subjects, aged 60 to 82 years, in four variants of the oddball experiment. The results indicated that these subjects yielded an orderly data set in which rare stimuli elicited a clear and typical P300, albeit with a somewhat longer latency than we have observed in young adults. However, in about one-sixth of the sample, the P300 appeared to be absent or very small in all four tasks. This was quite surprising relative to our experience with hundreds of young adults run in similar studies.

To test the reliability of these observations, the four oddball tests were repeated with half of the original sample several months later. It turned out that the low-P300 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 reasonably good (.44, .51, .66, and .73).

A comparison of 19 young adults with the full elderly sample yielded findings consistent with those recorded in other laboratories, reviewed above. There was a highly significant P300 latency difference between our aged (539 msec) and young (455 msec) groups, averaged across tasks and target/nontarget. This is a slowing of 1.77 msec/year, within the range of values found by other investigators (see discussion above). As some other laboratories have reported, P300 amplitude was lower in the full elderly group, overall.

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 significant differences in reaction time. The groups also did not differ in P300 latency or response accuracy. Medical records were examined, data on educational background were collected, and various psychometric, sensory, and other tests were administered to the subjects. The high-P300 group was, on average, a year older than the low-P300 group and had a higher proportion of women. However, none of these differences proved statistically reliable. In consideration of our developing theory of P300 as a manifestation of an intracranial process associated with updating representations of information in working memory, it seemed possible that subjects who lack a P300 would 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, patterned after the Hebb-Corsi test used by Milner (1978) and her colleagues in evaluating the effects of severe temporal lobe damage on memory. On each trial in the task we used, the subject was presented with a digit followed by a second digit 1,000 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 accord with our theory, high- and low-P300 subjects differed in RT performance. Both groups were able to accomplish the task with high accuracy (88% for the "lows," and 93% for the "highs"; not significantly different). Mean RTs when reporting that a pair had been seen before were comparable for the high-P300 and low-P300 groups. The low-P300 subjects, however, were significantly slower than the high subjects *when reporting that a pair of digits had not been seen before*. Reaction time to nonrepeating digit pairs correlated with the P300 amplitude obtained in the original oddball study for the high and low groups combined (Spearman $r = -.49$, $p < .05$). We found a similar pattern of RTs in a small replication sample.

In summary, 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 external 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 small 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. It is becoming increasingly apparent that studies involving multiple, active tasks are needed to characterize age-related cognitive changes.

PSYCHIATRIC AND NEUROLOGICAL DISEASE IN THE ELDERLY

Although EEG assessment has played a role in clinical diagnosis for some decades (e.g., McAdam & Robinson, 1962), its clinical utility is variable. In the diagnosis of dementia, for example, EEG assessment is of little value (Goodin, 1985). This role has been eclipsed to a significant extent by the widespread use of computed tomography (CT) X-ray techniques. It further appears that the high anatomical fidelity available with nuclear magnetic resonance imaging (NMR or MRI) will further erode reliance on the EEG. Despite this progress in anatomical diagnosis, much remains to be accomplished with electrophysiological measures of brain function. Accordingly, the greater cognitive specificity that can apparently be achieved with ERP rather than EEG measures is moving the field away from the EEG and toward ERPs collected under increasingly precise cognitive demands.

In recent years, much of the focus of the literature on clinical applications of psychophysiology has shifted to two issues. First, there has been growing concern with potential utility in differential diagnosis, particularly with ERP measures. For example, it appears that simple, stimulus-determined visual ERPs show too much overlap between healthy and diseased groups to be of clinical use for individual patients (Coben, Danziger, & Hughes, 1983). On the other hand, components of the ERP that are sensitive to higher-order cognitive function appear to be more promising, with the latency of the P300 receiving most of the attention (discussed above). Second, data are appearing that implicate general physical health, as promoted by above-average exercise regimens, in substantially moderating the psychophysiological declines once thought to be inevitable with age.

The Specificity of P300 Changes

A group of recent studies captures well the dilemma encountered in attempts to apply ERP methodology to improve differential diagnosis in elderly samples. A number of studies have found a lengthened P300 latency associated with dementia in elderly subjects (Brown, Marsh, & LaRue, 1982; Goodin et al., 1978b; Pfefferbaum, Ford, Wenegrat, Roth, & Kopell, 1984b; Polich, Ehlers, Otis, Mandell, & Bloom, in press; Syndulko et al., 1982). Squires et al. (1980) reported considerably different distributions of P300 latency in a comparison of demented and (1) nondemented neurological patients, (2) psychiatric patients, and (3) healthy elderly subjects. Furthermore, clinician ratings of degree of cognitive impairment exam have been found to correlate with P300 latency (Lai, Brown, Marsh, & LaRue, 1983; Polich et al., in press). Thus, there is evidence that P300 holds considerable promise for the differential diagnosis of dementia. However, there is also evidence that P300 latency cannot distinguish among different types of dementia and cannot detect mild cases of dementia (Polich et al., in press).

Using a slightly more complex task, Pfefferbaum and colleagues (1984a, b) conducted an extensive study comparing 135 normal controls ranging in age from 18 to 90 with a group of demented patients, as well as with schizophrenics, depressives, and nondemented, cognitively impaired patients. Auditory and visual stimuli were employed, with both rare target (subjects responded with a button press) and rare nontarget (no response required) as well as frequent stimuli. P300s were elicited by both types of rare stimuli.

Like previous research, this study found that demented patients exhibited significantly prolonged P300 latencies when compared with normal subjects for both target and nontarget rare stimuli in both auditory and visual modalities. However, schizophrenics also showed significantly prolonged P300s. Nondemented, cognitively impaired patients and depressives seemed to exhibit somewhat longer P300 latencies than did normals but did not show as great an effect as the demented patients. Variability in P300 latency was greater in demented patients and in schizophrenics.

Amplitude effects were also observed in this study. Demented patients showed 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 had slower RTs than controls.

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 suggests that use of ERPs in differential diagnosis is problematic at best. Moreover, not all studies have found P300 latency differences in demented patients. With 42 demented elderly patients, 29 nondemented elderly patients, and 10 healthy young controls, Slaets and Fortgens (1984) found no significant difference in P300 latency between demented and nondemented patients performing an auditory oddball counting task. It is difficult to give these results weight, however, as a number of aspects of the results are highly unusual. The authors could identify a P300 in only 37% of their controls, 19% of their demented patients, and 20% of their nondemented patients. Only 35% of the patients could perform the instructed task. Of greatest concern was that nothing was done to deal with electrooculogram (EOG) artifact.

Of course, in considering this inconsistent literature, it is necessary to note that, too often, studies viewed as putative replications of previous work differ in important respects. 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 the failure of Pfefferbaum et al. (1984b) to replicate the work of Goodin et al. (1978b). Overall, Goodin's (1985) review of this and related studies concludes that P300 latency appears to be rather useful clinically because of what he sees as its high specificity for dementia.

General Health and the Inevitability of CNS Decline

The popular interest in health, exercise, and nutrition is beginning to filter into psychophysiological research on aging. A large literature has documented that individuals over 60 who have maintained aerobic fitness through systematic exercise for a number of years typically have cardiovascular and pulmonary systems that function comparably to those of young adults (Shepherd, 1978). There are no systematic studies, however, of the beneficial effects of such training on CNS functions, particularly those mediated in the brain. A handful of RT studies do suggest that mental processing speed (and presumably brain processing speed) is faster in older exercisers than it is in older nonexercisers; in some instances, it may be comparable to that of young adults (see Spirduso, 1980, 1982). A pilot study (Bashore, Heffley, & Donchin, unpublished data) assessed the relationship between age, aerobic fitness, and mental processing speed utilizing a task developed by McCarthy and Donchin (1981). We tested 10 men between 62 and 74 years of age on this task, 5 of whom jogged regularly (averaging about 20 miles a week) for at least 10 consecutive years. The ERPs of the older exercisers appeared to have a closer resemblance to those of young adults than to those of old nonexercisers. This correspondence was evident in the amplitude and latency of P300, as well as in the RT.

That this question is worthy of systematic investigation is suggested in a study completed recently by Dustman and colleagues (personal communication). They did comprehensive neuropsychological assessments, including ERP and RT measures, of old and young, exercisers and nonexercisers, and found, as we did, that older exercisers had RTs and P300 latencies that were significantly faster than their sedentary age peers and comparable to those of young adults. The suggestion from these two data sets is that the rate at which declines in brain processing efficiency occur with age may be alterable and that aerobic exercise may be one means by which this decline can be delayed well into later life.

CONCLUSIONS

In closing, we note a set of methodological and interpretive issues yet to be faced adequately in the aging/psychophysiology literature. First, the vast majority of psychophysiological studies with elderly subjects have been limited to a single physiological response system (e.g., heart rate or ERPs), perhaps with RT also recorded. Yet the value of multiple simultaneous measurement for characterizing psychological and physiological changes with age is clear. For example, the question of age-related increases in physiological consistency or decreases in variability, noted earlier in this review for EEG power and for P300 amplitude, has emerged also in quite different studies, involving a variety of cardiovascular measures (e.g., Gintner, Hollandsworth, & Intrieri, 1986). If age-related variability changes are stable, pervasive, individual-differences phenomena, might increased co ncy (decreased flexibility?) in P300 response predict sustained elevation in blood pressure? The short- and long-term implications of blood pressure changes in the elderly are particularly deserving of continuing research, given the high rate of blood pressure problems in this population. Questions about consistency across measures and across individuals within an age group deserve increased attention in future research.

As a second general issue, how does one decide that an age-related change is a "decline" in function, particularly without reference to a specific subject goal or task context? It is also crucial to determine the boundaries of an apparent functional decline. As a third and closely related issue, should increases in consistency, decreases in response amplitude, and decreases in reactivity be treated as equivalent types of "decline"? Consider the following example.

In a study of the habituation of the sweat response (skin conductance response, or SCR) and recall of slides of common and uncommon objects, elderly subjects had poorer recall performance and smaller SCRs (Plouffe & Stelmack, 1984). However, young and elderly subjects showed very similar relationships between novelty and SCR, SCR and recall performance, and habituation and other variables. Thus, while the attenuation of SCR paralleled attenuated memory function, the dynamic relationship between the experimental variables was fully intact in the elderly sample. It appears from this data set that *how* the relevant memory system and its physiological support are functioning is entirely normal in the older sample. That is, the older sample's decline in memory performance is not due to a complete loss of some capability. Rather, normal strategies are being used but with somewhat reduced success.

These results highlight a fourth issue: When two groups of subjects are found to differ on some observable measure, do we conclude that they employed different strategies, the same strategies but at different times, or the same strategies at the same time but with differing degrees of skill? It is of considerable theoretical significance whether a deviant group differs because it is incapable of pursuing a normal strategy or merely because it is capable of the normal strategy but chooses to employ a different one (Perlmutter & Mitchell, 1982). This issue is also of considerable practical significance for training and rehabilitation efforts.

Finally, more attention must be paid to the importance of showing appropriate *differential* deficits, not merely differences on one measure. This is a classic point from the testing and assessment literature in clinical psychology (e.g., Chapman & Chapman, 1973). If one group is thought to have certain generalized deficits, then demonstration of a difference with another group on a single measure may say nothing about that measure. For example, if one group is generally less able to maintain vigilance, then any task requiring vigilant attention will indicate inferior performance for that group when, in fact, there may be no specific deficit for that task. A related point is that care must be taken in matching tasks and subject groups on relevant variables. For example, a task

believed to measure a certain cognitive function may do so equally for two subject groups but may be more difficult for one group than the other for reasons having nothing to do with possible differences in that cognitive function. One group's poorer performance would then falsely suggest lower cognitive function. When multiple tasks are employed—a research strategy for which there is great need in the literature reviewed here—it is important to evaluate the tasks for their relative psychometric properties. For example, if two subject groups truly differ from each other an equal amount on two tasks, and the two tasks differ in reliability (in a test-theoretic sense), the task having the higher reliability will produce a larger group difference, purely as a statistical artifact. These issues of relative task difficulty and reliability have received essentially no attention in the aging/psychophysiology literature.

Psychophysiology is partaking actively of the rapidly growing interest in basic research in gerontology. Important contributions are already apparent in understanding the locus of the general slowing of behavior in old age, and much promise is apparent in analyzing cognitive function. Psychophysiology at present is struggling to establish itself as useful in clinical applications: In the language of clinical medicine, the sensitivity of a variety of psychophysiological measures is fairly well established, but there is no consensus yet on the specificity. It is certainly premature to draw conclusions about whether a useful degree of specificity can be established. As a suggestion applicable to any study of special populations (Donchin & Bashore, in press; Donchin et al., 1986b), a supplement to the usual strategy of selecting subjects on the basis of clinical diagnosis and then assessing possible differences in the laboratory would be to select subjects on the basis of laboratory deficits (e.g., specific differences in cognitive function) and then to look for other laboratory differences and for systematic clinical differences.

Only in cognitive ERP research have enough programmatic psychophysiological studies been conducted with aged samples to provide a significant contribution to what is known about the aging process. A number of methodological and conceptual issues have been raised here in the context of a brief review of a rapidly growing literature. It is hoped that greater attention to these issues will allow other facets of geriatric psychophysiology to develop a solid knowledge base.

REFERENCES

Anders, T. R., & Fozard, J. L. (1973). Effects of age upon retrieval from primary and secondary memory. *Developmental Psychology*, 9, 411–416.

Anders, T. R., Fozard, J. L., & Lillyquist, T. D. (1972). Effects of age upon retrieval from short-term memory. *Developmental Psychology*, 6, 214–217.

Bashore, T. R. (in press). Age-related changes in mental processing revealed by analyses of event-related brain potentials.

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*, 523–545.

Birren, J. E. (1974). Translations in gerontology—from lab to life. Psychophysiology and the speed of response. *American Psychologist, 29*, 808–815.

Birren, J. E., & Botwinick, J. (1955). Speed of response as a function of perceptual difficulty and age. *Journal of Gerontology, 10*, 433–436.

Birren, J. E., & Cunningham, W. (1985). Research on the psychology of aging: Principles, concepts, and theory. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging (2nd ed.)* (pp. 3–34). New York: Van Nostrand Reinhold.

Birren, J. E., Woods, A. M., & Williams, M. V. (1979). Speed of behavior as an indicator of age changes and the integrity of the nervous system. In F. Hofmeister & C. Mueller (Eds.), *Brain function in old age* (pp. 10–44). Berlin: Springer-Verlag.

Brown, W. S., Marsh, J. T., & LaRue, A. (1982). Event-related potentials in psychiatry: Differentiating depression and dementia in the elderly. *Bulletin of the Los Angeles Neurological Society, 47*, 91–107.

Brown, W. S., Marsh, J. T., & LaRue, A. (1983). Exponential electrophysiological aging: P3 latency. *Electroencephalography and Clinical Neurophysiology, 55*, 277–285.

Cerella, J. (1985). Information processing rates in the elderly. *Psychological Bulletin, 98*, 67–83.

Cerella, J., Poon, L. W., & Williams, D. M. (1980). Age and the complexity hypothesis. In L. Poon (Ed.), *Aging in the 1980s: Psychological issues* (pp. 332–340). Washington, DC: American Psychological Association.

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

Coben, L. A., Danziger, W. L., & Hughes, C. P. (1983). Visual evoked potentials in mild senile dementia of the Alzheimer's type. *Electroencephalography and Clinical Neurophysiology, 55*, 121–130.

Davidson, R. J. (1978). Specificity and patterns in biobehavioral systems: Implications for behavior change. *American Psychologist, 33*, 430–436.

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

Donchin, E., & Bashore, T. (in press). Clinical versus psychophysiological paradigms in the study of event-related brain potentials. In S. Harnad (Ed.), *The behavioral and brain sciences*.

Donchin, E., Karis, D., Bashore, T. R., Coles, M. G. H., & Gratton, G. (1986a). Cognitive psychophysiology and human information processing. In M. G. H. Coles, S. W. Porges, & E. Donchin(Eds.), *Psychophysiology: Systems, processes, and applications* (pp. 244–267). New York: Guilford Press.

Donchin, E., Miller, G. A., & Farwell, L. A. (1986b). The endogenous components of the event-related potential—a diagnostic tool? In D. F. Swabb, E. Fliers, M. Mirmiram, W. A. Van Gool, & F. van Haaren (Eds.), *Progress in brain research. Vol. 70: Aging of the brain and Alzheimer's disease* (pp. 87–202). Amsterdam: Elsevier.

Drechsler, F. (1978). Quantitative analysis of neurophysiological processes of the aging CNS. *Journal of Neurology, 218*, 197–213.

Duffy, F. S., Albert, M. S., McAnulty, G., & Garvey, A. J. (1984). Age-related differences in brain electrical activity of healthy subjects. *Annuals of Neurology, 16*, 430–438.

Dustman, R. E., LaMarche, J. A., Cohn, N., Shearer, D. E., & Talone, J. M. (1985). Power spectral analysis and cortical coupling of EEG for young and old normal adults. *Neurobiology of Aging, 6*, 193–198.

Eriksen, C. W., Hamlin, R. M., & Daye, C. (1973). Aging adults and rate of memory scan. *Bulletin of the Psychonomic Society, 1*, 259–260.

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).

Ford, J. M., Duncan-Johnson, C. C., Pfefferbaum, A., & Kopell, B. S. (1982). Expectancy for events in old age: Stimulus sequence effects on P300 and reaction time. *Journal of Gerontology, 37*, 696–704.

Ford, J. M., Hink, R. F., Hopkins, W. F., Roth, W. T., Pfefferbaum, A., & Kopell, B. S. (1979). Age effects on event-related potentials in a selective attention task. *Journal of Gerontology, 34*, 388–395.

Ford, J. M., & Pfefferbaum, A. (1980). The utility of brain potentials in determining age-related changes in central nervous system and cognitive functioning. In L. Poon (Ed.), *Aging in the 1980s: Psychological issues* (pp. 115–124). Washington, DC: American Psychological Association.

Ford, J. M., & Pfefferbaum, A. (1985). Age-related changes in event-related potentials. In P. Ackles, R. Jennings, & M. G. H. Coles (Eds.), *Advances in psychophysiology, Vol. 1* (pp. 301–339). Greenwich, CT: JAI Press.

Ford, J. M., Pfefferbaum, A., Tinklenberg, J. R., & Kopell, B. S. (1982). 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. 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.

Giaquinto, S., & Nolfe, G. (1985). The EEG and cognitive tasks in the aged. *Journal of Neurology, 232*, 447.

Gintner, G. G., Hollandsworth, J. G., & Intrieri, R. C. (1986). Age differences in cardiovascular reactivity under active coping conditions. *Psychophysiology, 23*, 113–120.

Goodin, D. S. (1985). Electrophysiological evaluation of dementia. *Neurologic Clinics, 3*(3), 633.

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.

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

Johnson, R., Pfefferbaum, A., & Kopell, B. S. (1985). P300 and long-term memory: Latency predicts recognition performance. *Psychophysiology, 22*, 497–507.

Katz, R., & Horowitz, G. R. (1982). Electroencephalogram in the septuagenarian:

Studies in a al geriatric population. *Journal of the American Geriatric Society, 3,* 273–275.
Kausler, D. H. (1982). *Experimental psychology and human aging.* New York: John Wiley & Sons.
Kirby, N. H. (1976). Sequential effects in two-choice reaction time: Automatic facilitation or subjective expectancy? *Journal of Experimental Psychology: Human Perception and Performance, 2,* 567–577.
Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science, 197,* 792–795.
Lachman, R., Lachman, J. L., & Butterfield, E. C. (1979). *Cognitive and information processing: An introduction.* Hillsdale, NJ: Erlbaum.
Lai, J., Brown, W. S., Marsh, J. T., & LaRue, A. (1983). Covariation of the P3 latency and Mini-Mental-State scores in geratric patients. *Psychophysiology, 20*(4), 455.
Madden, D. J. (1982). Age differences and similarities in the improvement of controlled search. *Experimental Aging Research, 8,* 91–98.
Madden, D. J., & Nebes, R. D. (1980). Aging and the development of automaticity in visual search. *Developmental Psychology, 16,* 377–384.
Maniscalco, C. I., & DeRosa, D. V. (1983). Memory scanning of young and old adults: The influence of rate of presentation and delay interval on recognition memory performance. *Bulletin of the Psychonomic Society, 21,* 7–10.
Marsh, G. R. (1975). Age differences in evoked potential correlates of a memory scanning process. *Experimental Aging Research, 1,* 3–16.
Marsh, G. R., & Thompson, L. W. (1973). Effects of age on the contingent negative variation in a pitch discrimination task. *Journal of Gerontology, 28,* 56–62.
Marsh, G. R., & Thompson, L. W. (1977). Psychophysiology of aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 219–248). New York: Van Nostrand Reinhold.
McAdam, W., & Robinson, R. A. (1962). Diagnostic and prognostic value of the electroencephalogram in geriatric psychiatry. In T. H. Blumenthal (Ed.), *Medical and Clinical Aspects of Aging* (pp. 557–564). New York: Columbia University Press.
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.
Michaelewski, H. J., Thompson, L. W., Smith, D. B. D., Patterson, J. V., Bowman, T. E., Litzelman, D., & Brent, G. (1980). Age differences in the contingent negative variation (CNV): Reduced frontal activity in the elderly. *Journal of Gerontology, 35,* 542–549.
Milner, B. (1978). Clues to the cerebral organization of memory. In P. A. Buser & A. Rougel-Buser (Eds.), *Cerebral correlates of conscious experiences: INSERM Symposium No. 6* (pp. 139–153). Amsterdam: Elsevier North-Holland.
Mullis, R. J., Holcomb, P. J., Diner, B. C., & 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.
Obrist, W. D. (1979). Electroencephalographic changes in normal aging and dementia. In

F. Hoffmeister & F. Muller (Eds.), *Brain function in old age, Bayer sium VII* (pp. 102–111). Berlin: Springer-Verlag.
Perlmutter, M., & Mitchell, D. B. (1982). The appearance and disappearance of age differences in adult memory. In F. I. M. Craik & S. Trehub, *Aging and cognitive processes.* New York: Plenum.
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. *Electroencephalography & 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. *Electroencephalography & Clinical Neurophysiology, 59,* 104–124.
Picton, T. W., Cerri, A. M., Champagne, S. C., Stuss, D. T., & Nelson, R. F. (1986). The effects of age and task difficulty on the late positive component of the auditory evoked potential. In W. C. McCallum, R. Zappoli, & F. Denoth (Eds.), *Cerebral psychophysiology: Studies in event-related potentials (Electroencephalography and Clinical Neurophysiology, Supplement 38).* Amsterdam: Elsevier Science Publishers.
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.
Plouffe, L., & Stelmack, R. M. (1984). The electrodermal orienting response and memory: An analysis of age differences in picture recall. *Psychophysiology, 21,* 191–198.
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.
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.
Polich, J., Ehlers, C. L., Otis, S., Mandell, A. J., & Bloom, F. E. (1986). P300 latency reflects the degree of cognitive decline in dementia illness. *Electroencephalography and Clinical Neurology.*
Polich, J., Howard, L., & Starr, A. (1985). Effects of age on the P300 component of the event-related potential from auditory stimuli: Peak definition, variation, and measurement. *Journal of Gerontology, 40,* 721–726.
Polich, J., & Starr, A. (1984). Evoked potentials in aging. In M. L. Alpert (Ed.), *Clinical neurology of aging* (pp. 149–177). New York: Oxford University Press.
Rabbitt, P., & Vyas, S. M. (1980). Selective anticipation for events in old age. *Journal of Gerontology, 35,* 913–918.
Ritter, W., Simson, R., Vaughan, H. G., & Friedman, D. (1979). A brain event related to the making of a sensory decision. *Science, 203,* 1358–1361.
Ritter, W., Simson, R., Vaughan, H. G., & Macht, M. (1982). Manipulation of

event-related potential manifestations of information processing stages. *Science, 218*, 909–911.
Ritter, W., Vaughan, H. G., & Simson, R. (1983). On relating event-related potential components to stages of information processing. In A. W. K. Gaillard & W. Ritter (Eds.), *Tutorials in event-related potential research* (pp. 143–158). Amsterdam: North Holland.
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 ERP research: Endogenous components*. Amsterdam: North-Holland.
Salthouse, T. A. (1982). *Adult cognition: An experimental psychology of human aging*. New York: Springer-Verlag.
Salthouse, T. A. (1985a). Speed of behavior and its implications for cognition. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 400–426). New York: Van Nostrand Reinhold.
Salthouse, T. A. (1985b). *A theory of cognitive aging*. Amsterdam: North-Holland.
Salthouse, T. A., & Somberg, B. L. (1982a). Isolating the age deficit in speeded performance. *Journal of Gerontology, 37*, 59–63.
Salthouse, T. A., & Somberg, B. L. (1982b). Skilled performance: The effects of adult age and experience on elementary processes. *Journal of Experimental Psychology: General, 111*, 176–211.
Shepherd, R. J. (1978). *Physical activity and aging*. London: Croom Helm.
Siegel, I. C., Nowlin, J. B., & Blumenthal, J. A. (1980). Health and behavior: Methodological considerations for adult development and aging. In L. Poon (Ed.), *Aging in the 1980s: Psychological issues* (pp. 599–614). Washington, DC: American Psychological Association.
Simons, R. F. (in press). Can the autonomic nervous system contribute to our understanding of event-related potentials?
Slaets, J. P. L., & 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* (pp. 112–125). Berlin: Springer-Verlag.
Spirduso, W. W. (1980). Physical fitness, aging and psychomotor speed: A review. *Journal of Gerontology, 35*, 850–865.
Spirduso, W. W. (1982). Physical fitness in relation to motor aging. In J. A. Mortimer, F. J. Maletta, & G. J. Maletta (Eds.), *The aging motor system* (pp. 120–151). New York: Praeger.
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* (pp. 125–134). Washington, DC: American Psychological Association.
Squires, K., Goodin, D., & Starr, A. (1979). Event related potentials in development, aging and dementia. In D. Lehmann & E. Callaway (Eds.), *Human evoked potentials* (pp. 383–396). New York: Plenum.

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.
Sternberg, S. (1969). The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica, 30*, 276–315.
Sternberg, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology, 27*, 1–32.
Storrie, M. C., & Eisdorfer, C. (1978). Psychophysiological studies in aging: A ten-year review. In M. A. Lipton, A. DiMascio, & K. F. Killam (Eds.), *Psychopharmacology: A generation of progress*. New York: Raven Press.
Strayer, D. L., Wickens, C. D., & Braune, R. (1987). Adult age differences in the speed and capacity of information processing: II. An electrophysiological approach. *Psychology and Aging, 2*, 99–110.
Syndulko, K., Hansch, E. C., Cohen, S. N., Pearce, J. W., Goldberg, Z., Montan, 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* (pp. 279–285). New York: Raven Press.
Tecce, J. J. (1979). A CNV rebound effect. *Electroencephalography and Clinical Neurophysiology, 46*, 546–551.
Tecce, J. J., Boehner, M. B., Cattanach, L., Branconnier, R. J., & Cole, J. O. (1981). Hydergine treatment and brain functioning (CNV rebound) in Alzheimer's patients: Preliminary findings. *Psychopharmacology Bulletin, 173*, 202–206.
Tecce, J. J., Cattanach, L., Yarchik, D. A., Meinbresse, D., & Dessonville, C. L. (1982). CNV rebound and aging. *Electroencephalography and Clinical Neurophysiology, 54*, 175–186.
Thompson, L. W., & Nowlin, J. B. (1973). Relations of increased attention to central autonomic nervous system states. In L. F. Jarvik, C. Eisdorfer, & J. E. Blum (Eds.), *Intellectual functioning in adults* (pp. 107–123). New York: Springer Publishing Co.
Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41*, 67–85.

Article 14.

GENERALIZED IMPLEMENTATION OF AN EYE MOVEMENT CORRECTION PROCEDURE

Gregory A. Miller, Gabriele Gratton, and Cindy M. Yee

University of Illinois at Urbana-Champaign

In press, *Psychophysiology*

Title

Generalized Implementation of An Eye Movement Correction Procedure

Authors

Gregory A. Miller, Gabriele Gratton, and Cindy M. Yee, University of Illinois at Urbana-Champaign

Purpose

Removal of EOG artifact from EEG data off-line, distinguishing blink and saccade artifact.

Description

The eye movement correction procedure algorithm proposed by Gratton and colleagues (Gratton, G., Coles, M.G.H., & Donchin, E., A new method for off-line removal of ocular artifact, Electroencephalography and clinical neurophysiology, 1983, 55, 468-484) has been implemented, with extensions, in standard DEC Fortran IV. Extensions are principally (a) new computational efficiences and (b) suitability for use with both vertical and horizontal EOG channels rather than vertical EOG alone. Single-trial data are screened for blinks and bad data (the latter defined as either a value off-channel or an extended run of identical values, which can occur if a baseline has already been removed from off-channel data). Raw averages are computed for each channel and experimental condition, separately for blink and blink-free data points. These averages are subtracted from the single-trial data. Correlations are then computed between the remainder for each EEG channel and the remainder for the EOG channel, separately for blink and blink-free data points, producing blink and saccade correction factors for each EEG channel. These correction factors are then applied to the original raw averages. Optionally, in an extra pass on the input file, single trials are also corrected, using the same correction factors.

Key Methodological Issues

(1) Artifact estimates are derived from the data the user wishes to correct; not solely from pre-stimulus or baseline data, from instructed eye movements, or from other special recording epochs. (2) Placement of EOG electrodes must be considered carefully, with attention to the specific EEG channels to be corrected. We have used the algorithm only with bipolar, circum-ocular channels, and we recommend strict vertical and horizontal EOG recordings. A monopolar EOG referred to the mastoid, for example, would be vulnerable to EEG contamination. A bipolar EOG placement at 45 degrees from the vertical would confound vertical and horizontal eye movements. (3) Artifact estimates and corrections are performed separately for blinks and saccades. All data points not determined to coincide with blinks are used in computing the saccade correction factor for each channel. (4) Correction factors are computed from all single-trial data following removal of average data for a given trial's experimental condition. That is, the average ERP for all trials in a condition is subtracted from each of the single trials in that condition, for each EOG and EEG channel. The user decides what the set of "conditions" is. In general, each within-subject cell of the experimental design expected to have a different average waveform should be declared as a separate condition. After subraction of the relevant average, all trials contribute to the computation of a single blink and a single saccade correction factor for each channel. (5) Corrections based on multiple channels of EOG are handled in multiple runs of the program. Thus, if non-midline EEG data are recorded, and both vertical and horizontal EOG are available, the user would set up the first run of the program to treat the horizontal channel identically with the EEG channels, removing vertical EOG artifact (blinks and the vertical-axis component of saccades). In a

second run, the residual in the EEG channels would be corrected for correlations with the residual in the horizontal EOG channel (horizontal-axis saccades).

Hardware Characteristics

The program was developed on a DEC PDP-11/73 using single-precision (32-bit) real arithmetic and arrays. Space needs for data arrays can be large for some combinations of number of channels, points per channel, and number of experimental conditions (number of trials has little impact on array usage). One version of the program (7 channels, 100 points, 4 conditions, 4000 trials) requires 108KB. Another version (4 channels, 125 points, 11 conditions, 1500 trials) requires 148KB. For DEC PDP-11 computers, sufficient memory available for virtual arrays is necessary in most cases.

Software Characteristics

The eye movement correction program was originally written in Harris Fortran F66, a proprietary structured language. The program has been re-written in DEC Fortran IV (compiler version 2.6), running under RT-11 or TSX-Plus. It is highly compatible with other Fortran IV and Fortran-77 compilers, as efforts were made in coding to facilitate porting to other compilers.

I/O Characteristics

Standard DEC RT-11 Fortran files are read and written. A user-edited specification file directs program execution, such as whether blink detection and correction is to be done (e.g., it would not be if correction is based on a horizontal EOG channel), and the value beyond which data are to be considered off-channel.

Performance Notes

Runtime depends on the product of number of trials, channels, and points. Execution time for 4000 trials, 7 channels, 100 points, and 4 conditions in unformatted binary files on moderately fast hard disk is 110 minutes for average-ERP correction and optionally another 60 minutes for single-trial correction. For 4 channels, 125 points, and 11 conditions, execution time for 1500 trials is 30 minutes for average-ERP correction and optionally another 30 minutes for single-trial correction.

Algorithm Porting Considerations

(1) Each user should experiment with optimal settings for the blink detection criterion, depending on local amplification. (2) The correction works poorly for EEG channels positioned very close to the EOG site, due to the EEG signal appearing in the EOG channel. Thus, a small amount of true EEG common to both Fpz EEG and supra-orbital EOG recordings will lead the correction algorithm to remove that portion of the EEG from the Fpz channel.

Source Code Porting Considerations

(1) Array space needs can be large. The program runs without modification on a PDP-11/23 with 22-bit addressing (run-time is approximately doubled). The program has been ported successfully to a 256KB PDP-11/34 and to a 256KB PDP-11/73 at other universities. (2) File I/O format must be customized to the user's data files. The program currently uses unformatted binary files for speed and space reasons and utilizes trial identification information stored within the record for each channel.

Program Availability

Source code listing; copy on user-supplied DEC RT-11 9-track magtape at 800 or 1600 BPI; TU58 tape; RK05, RL01, or RL02 disk; dial-up modem; or BITNET. Sample data and output are available.

Documentation

Source code is extensively commented. A user's manual is available.

Support Available

The authors are available for consultation.

Acknowledgements

Extensions and Fortran IV development supported in part by NIMH grant MH39628 to G. A. Miller, and by NIMH grant MH41445, M. Coles, Principal Investigator. Original development supported in part by AFOSR contract # 49620-79-C-0233, monitored by A. Fregly, to E. Donchin.

EPIC

Applications of ERPs to Human Factors

Christopher D. Wickens

In his paper, Dr. Parasuraman has eloquently made the case for the modest, yet significant, contribution that the Event Related Brain Potential (ERP) may make to the discipline of Human Factors, and he has as well provided a nice overview of many of the most promising applications. I applaud both his call for careful theory-based applications, as well as his caution for exercising restraint when describing the real contributions that the measures can make. The objectives of my presentation here will be twofold: (1) to outline what, from my perspective, are the necessary criteria for complete and successful applications of ERPs to human factors systems design and evaluation, and in what way these criteria have been or can be met. (This outline will, I believe, reveal some of the reasons why applications are difficult to achieve and, therefore, why restraint in promises is necessary), (2) to review Dr. Parasuraman's list of candidate applications and to expand upon this list slightly by adding some of my own.

It should be noted in opening that the sequence of steps from the identification of an interesting ERP phenomenon to the successful applications in system design is a long one that will not be easily traversed. Yet, the frustrations inherent in this delay are not unique to the ERP field, as the time lag for applications of basic research in engineering and cognitive psychology is sometimes equally long and tortuous (Adams, 1972; Rouse, 1985). Many lessons learned from the latter are equally relevant to the former, and will be elaborated below. Yet, electrophysiological signals have some unique characteristics not shared by the more conventional manual and vocal performance measures, that dictate special concerns in the course of application. Before this course and the relevant concerns are outlined however, I shall briefly summarize what I see to be the three most relevant, and generic human engineering concerns, for which ERPs may provide general answers. These are:

(1) How to choose which is the "best" system (or system component). Such a choice may involve a simple comparison of performance. But if performance is equal between two systems, or is difficult to measure the choice may require a comparison of the cost of performance in terms of mental workload. And if the system of greater workload is more expensive in dollars to manufacture and operate, then the choice requires a quantitative scale of workload that may be balanced against dollars.

(2) How to predict human error. Will models of human performance be accurate enough to predict when a human operator, interacting with a particular system, may produce an error with major consequences? Can errors of perception, of attention, of memory, and of response be predicted and discriminated?

(3) How to assess the level of skill acquired by an operator through training. If performance reaches asymptote, will the workload continue to be reduced? At what point has training clearly become automated?

These then are the objectives toward which ERP-based applications should be focused.

Criteria for Application

In order to apply ERPs to answer the questions outlined above, the cognitive psychophysiologist must first address the question that will undoubtedly be asked by the HF practitioner: "Why not use more conventional measures of performance -- either manual responses or vocal reports -- to address the same human engineering issues?" The answer to this question is that ERPs can provide information that is either obtained at a greater cost by using performance, or cannot be obtained at all by performance measures. (Hereafter the term "performance" will be used to refer to these conventional behavioral measures, in contrast to the event-related potential measures.) The first issue -- that of relative cost will be deferred until later. However, the second issue will be discussed extensively below, because it is critically important to the issue of applications.

Dissociation and Association

The idea that ERPs can provide information unavailable in performance invokes the concept of dissociation, depicted in Figure 1. Here two forms of variance -- in physiological measures and in performance -- are depicted. These two variances define three regions. In the overlapping region (1), performance and ERPs covary. In the region to the left (Region 2), ERPs show unique variance that is not revealed in performance. This variance provides information to the extent that it correlates with some experimenter-defined manipulation, or subject-reported strategy. In the region to the right (Region 3), ERPs fail to reflect changes shown by performance. Regions 2 and 3 then define dissociation. The existence of Region 2 and, less obviously, Region 3 both provide answers to the question "Why not performance?" But, in order to satisfactorily answer this question, it will be necessary for the cognitive psychophysiologist, working in consort with the Human Factor practitioner, to traverse through the sequence of logical steps shown in Figure 2.

The sequence is entered at point 1, where an important source of dissociation is identified. For example, ERPs may serve as a measure of auditory attention allocation, that is impossible to observe through performance (e.g., Hansen and Hillyard, 1980). The question of which messages may be attended and which ignored can be a critical one in an operational environment such as the aircraft cockpit. At this point, the system designer may rightfully ask if the dissociation, rather than reflecting meaningful unique variance (region 2 or figure 1), or an interpretable lack of effect (region 3), is in fact either noise or insensitivity (a ceiling or floor effect). To convince the practitioner that the ERP measure is meaningful thus requires that the psychophysiologist take a step back, to point 2 of figure 2, and demonstrate an association. That is, provide an experiment in which the ERP correlates directly with performance as in Region 1 of Figure 1. Examples of such studies are those in which P300 is correlated with manipulations of difficulty, which also affects performance (e.g., Isreal, Wickens, Chesney, & Donchin, 1980). These validation studies of course contribute nothing directly to the value of the ERP measures as a system design tool, except to persuade the HF practitioner of the meaningfulness and trustworthiness of the measure, so that a logical case can be established as to why the measure dissociates in some situations (Kahneman, 1982 SPR Address).

Once the HF practitioner has been convinced of the validity of the measure, then potentially valuable information may be gained from either of the two types of dissociation. An example of Type 3 dissociation was provided by an experiment of Isreal, Chesney, Wickens, and Donchin (1980), in which it was observed that the ERP amplitude of the P300 component did not change, even as the demand of a tracking task was varied by changing its bandwidth. Such constancy of the ERP was interpreted in light of a process model of P300. This model posited that the amplitude depended only upon the availability of perceptual/cognitive resources, which were undepleted by the response-load manipulation of tracking task. As a consequence, the unique diagnosticity of the ERP to reveal perceptual/cognitive load was demonstrated. A similar demonstration was provided by Kramer, Wickens, and Donchin (1983), who attributed the lack of reduction of secondary task P300 amplitude with practice on a primary target acquisition task, to the development of response-related strategies. In a more general sense, Type 3 dissociations of P300 and performance can be reasonably and measurably interpreted whenever the manipulations are logically those that involve either the resource demands (Kramer, et al., 1983; Isreal, Chesney, et al., 1980), or the latency (McCarthy & Donchin, 1981) of response-related processes (see Figure 3).

The importance of Type 2 dissociations are easier to justify to the human factors practitioner. This is because of the plethora of perceptual and attentional phenomena directly indexed by ERP components, which are <u>not</u> revealed by performance measures. These dissociations, in fact, increase in their importance in human factors to the extent that automated systems evolve which demand continuous monitoring and understanding with only minimal requirements for the operator to intervene with overt responses (Hart & Sheridan, 1984). Therefore, under such conditions (and their laboratory analogs), performance data are simply <u>not</u> available. Here again however, as in the case of Region 2, the establishment of a meaningful and useful dissociation <u>must</u> be based upon a theory-based or processed-based explanation of why the ERP should be expected to vary in the way that it does. Thus for example, theories of visual or auditory selective attention are necessary to establish the conditions under which attention allocated to a visual or auditory stimulus would be expected to enhance some component of the ERP (Parasuraman, 1985; Hansen and Hillyard, 1980).

"Real World" Complexity

The issue of how to successfully demonstrate a dissociation between performance and ERPs in a manner that will convince the human factors practitioner of the value of the latter, brings up a point addressed directly in Parasuraman's paper -- the extent of "real world complexity" that it is necessary to design into the paradigm. Here the views of the psychophsiologist (generally reflecting those of the cognitive and experimental psychologist) may come into partial conflict with those of the engineering psychologist. The experimental or cognitive psychologist will likely feel that good and careful experimental control of a phenomenon in a laboratory setting is not only necessary, but also sufficient to establish the relevance and validity of that phenomenon as an aid to systems design questions. I agree with this supposition only in part, as qualified by the slightly different perspective of the engineering psychologist. This perspective argues that two further factors must be taken into account beyond careful controls, both of which make it important that the dissociation be demonstrated in environments of greater complexity.

(1) Many Human Factors practitioners whose objective it is to convince will rightly, or wrongly, fail to be convinced unless the dissociation is demonstrated in an environment that has at least some of the trappings of real world complexity. For example, according to this view, a convincing demonstration of the ERP as an aviation workload measure must be provided in the aircraft's simulator, and not just in the tracking laboratory. However,

it is equally important to realize that the increased complexity should not be bought at the price of a shoddy, poorly designed study. Increases in complexity will not compensate for deficiencies in experimental protocol. It is unfortunate for example that many potentially valuable demonstrations have proven to be relatively worthless because small sample sizes have been used to compensate for the increased expenses incurred in conducting hese more complex studies.

(2) More importantly from a fundamental applications standpoint, it is necessary to establish that the variance in the processing phenomenon reflected by the electrophysiological measure, is sufficiently large, relative to the sources of uncontrolled variance that exist in the real world, that the former will have a meaningful impact on system performance. Suppose for example, it is established that, with all variables tightly controlled in the laboratory, a particular ERP component is sensitive to a manipulation of resource allocation. But when some of those controls are released (as they must be outside of the controlled laboratory setting), and small random effects now perturb the measure, the variance in ERPs due to the original phenomenom is now swamped. In such an instance, the true value of the ERP measure to the HF practitioner would be greatly diminished. Yet this diminished worth would never have been revealed until the tight controls were lifted.

This requirement to incorporate some level of complexity in experimental demonstrations of a dissociation should not be viewed pessimistically by the psychophysiologist. In fact, when incorporating real world complexity brings about an increase in overall workload, such an increase may even enhance the prominence of the processing phenomenon under investigation (e.g., resource availability or the criticality of resource allocation).

An example, demonstrating the importance of being able to relax excessive experimental control, involves the concern for control of eye fixation. This concern transcends much of the experimentation with visual ERPs and with non-electrophysiological studies of perception and cognition as well. The view is sometimes expressed that the investigator must know the precise location of fixation in order to draw valid inferences. Yet, conclusions that are drawn only from the eye-controlled scan-free laboratory may be tenuous when applied to the wide-field visual world, and such conclusions legitimately may be distrusted by the human factors practitioner. In fact, such single-minded concerns with fixations are unnecessary. For example, both Wickens, Heffley, Kramer and Donchin (1980), and Kramer, Wickens, and Donchin (1983) have drawn some strong conclusions concerning the allocation of visual attention to a complex "free-scan" display, with many of the uncontrolled aspects of the natural world present. These conclusions gain in their validity precisely because visual fixations were not tightly constrained.

Design Impact

Given that the feasibility of ERP applications has been demonstrated in Step 3 (by iterating back to Step 2 when necessary to convince the skeptic by demonstrating association), the next step (4 in figure 2) is taken when the ERP measure actually impacts on the design of a system. Parasuraman's discussion nicely illustrates the example of Frustrahofer's study in which EEG provided a measure of arousal that dissociated from performance. As another example, the electrophysiological data using steady state ERPs provided by Gomer and Bish (1978) was used to influence the choice of formats for alphanumeric displays.

At Step 4, there are really three levels of impact that ERP measures may have on a design process. First, they may be used to justify a particular design decision because they reinforce behavioral measures, or design considerations, already available. Such, for example, would be the case if an ERP-based workload measure was used to influence the choice of one display format over another, in agreement with ratings obtained by subjective measures. Secondly, ERP measures may provide information unavailable through performance -- the Type 3 dissociation -- although this information may be consistent with the human factors practitioner's "gut beliefs." Here, for example, are ERP measures which may show that a warning light on a multi-element visual display is unattended.

The third level of impact is revealed in Step 5, and is one that would bring greatest satisfaction to the psychophysiologist that electrophysiological measures have truly "come of age." This would describe a situation in which the human factors practitioner made a different design decision as a consequence of the ERP measures than would have been made in their absence. Such a course of action would mean that the implication of the ERP measure for design was pitted against the implications of an alternative design consideration (cost, performance, or the practitioner's "gut belief"), and the ERP measure was victorious.

The last step (6) in the transition to applications, and the ultimate reinforcement for the psychophysiologist, is provided when it is determined that a system, designed with the aid of ERP measures, was actually more effective than one designed without. Unfortunately, such success stories are extremely hard to document, because the "control system-designs" (those systems that are designed by equivalent procedures except for the input provided by the ERP measure), are rarely available to be compared. It is here an act of faith to assume that the control system-design would not have been as effective, had it been completed. It is unfortunate that the absence of such necessary control data to evaluate the effectiveness of design tools is a widespread symptom in human factors, and hampers the evaluation of the usefulness of purely performance-based models and tools as well.

While the path to ultimate success for ERP applications from Step 1 to Step 6 appears long and tortuous, it is important to realize that partial successes may be very much realized at earlier steps along the way. For example, as shown toward the bottom of Figure 2, well designed ERP demonstrations of dissociation can have an impact on the general theory of human performance; ultimately, but with less direct connections, the revised theory may in turn influence system design (though at the time of that influence, the direct contributions of the ERP may not be recognized or acknowledged). For example, the investigations carried out on the ERP as a workload measure by Isreal et al. (1980a, 1980b) have had a direct input toward establishing the stage-of-processing dimension as an important component of the multiple resource model of human time-sharing performance. This model itself, is now becoming a useful tool in system design (North, 1985).

Applications

Parasuraman has nicely described applications in the areas of workload assessment and resource allocation. I would like now to add some further comments to his remarks on the applications of ERPs to these areas of assessment, and also to identify one additional area to which I believe ERP measures have the potential to contribute substantially.

<u>Workload</u>

First, I would like to address the concern that the P300 as a secondary task workload measure may be more obtrusive into primary task performance than is sometimes claimed by proponents of the measure. My answer is that the measure is indeed somewhat obtrusive, and cannot be obtained without imposing some extra cost on the operator's performance (having to count, or keep track of tones). Yet after working in the area of mental workload for over 10 years, I have become increasingly convinced that there is "no free lunch" in workload measures. Those successful measures that are less obtrusive into primary task performance (e.g., heart rate variability, Mulder & Mulder, 1981; pupil diameter, Beatty, 1982), by the very aspect of being less obtrusive on performance and as a result, a less direct measure of cognitive activity, are more susceptible to other influences, and therefore will be less reliable. Similar concerns arise regarding the applications of unobtrusive ERP measures of workload derived from the steady state ERP (Junker, Kenner, & Casey, 1985). The fact that these measures may be obtained by driving the visual field with high frequency luminance functions well above the discriminable critical flicker fusion frequency, is compromised by the apparent inconsistency with which parameters of the SSERP measure respond to imposed manipulations of workload (i.e., The association shown in Step 2 of Figure 2 is lacking).

Most recently, as noted by Parasuraman, research has suggested that ERPs elicited by events embedded naturally within the primary task <u>can</u> directly index the resources allocated to the primary task, and thereby indicate the workload of that task in a way that does not require the instrusive secondary task (Wickens, et al., 1982; Horst, et al., 1985; Kramer, Wickens and Donchin 1985; Sirevaag, et al., in press). This measure at present is limited, however, by the sensitivity of primary task ERPs to other potentially confounding variables. For example, a large ERP elicited by a primary task event may result either because the task was difficult, and more resources were allocated or because the event within the primary task was relatively unexpected. Also of course some primary tasks do not lend themselves naturally to the incorporation of discrete ERP-elicting events.

Resource allocation. As noted above, increasing levels of automation have progressively placed human operators of complex systems in a supervisory role in which large numbers of channels of visual and auditory information must be monitored. There are generally few overt responses made in these environments, but two important covert processes are often required: (1) monitoring, and then detecting and/or recognizing critical events; (2) integrating information to update a mental picture of the system or process under observation. It is of considerable importance for the system designer to attain an accurate model of the distribution, breadth or focusing of attention under these circumstances. To obtain such a representation, three "performance" measures are presently available -- verbal protocols, responses to stimulus events and oculometric measures. Here then, as in Step 1 of Figure 2, it is appropriate to ask what unique variance ERPs have to contribute beyond that available from performance. In short, why not performance?

In answer to this question, it should be noted first that verbal protocols may be unreliable. Secondly, reaction time or response measures cannot be easily collected, because to impose a requirement to respond to each event for which a measure of attention is required, may drastically alter the task characteristics in an unnatural way. Suppose, for example, one wished to infer whether or not the onset of a caution warning light were noticed during the intense involvement with another task. It would not be desirable to instruct the subject to "press the button" when the advisory appears. This alteration of the task would change its natural characteristics and give the processing of the caution light greater priority than it might actually have in the real world. It would be far better here to allow the ERP components to index the degree of processing given to the light occurring in its natural context.

Unlike overt manual or verbal responses, crude measures of visual fixations can be recorded without greatly disrupting the natural information processing characteristics of the task. As long as the inference is made that where the eye is looking is exactly what is processed, then necessary and sufficient information will be available. The problem is this assumption is limited in four respects: (1) Different items or dimensions of information can be either attended or ignored within a single fixation (e.g., Donchin & Cohen, 1967; Kramer, Wickens, & Donchin, 1983). Therefore, fixation can not discriminate attended for unattended dimensions of a single stimulus; nor the allocation of resources between two stimuli within 1˚ of visual angle. (2) Information may be processed and attended outside of the fovea (Posner, 1982). There is, in fact, a substantial literature showing the value of parafoveal motion in tracking performance (Wickens, 1986). The degree to which this processing is taking place however can never be revealed through fixations. (3) Attention may be switched between modalities. It is apparent that fixations provide no clue as to whether auditory information is, or is not, being processed. (4) Long fixations are ambiguous. There may be three reasons why fixation on a particular stimulus is of a long duration. First, a lot of information may be extracted from that stimulus (it is surprising, novel, or complex). Second, the stimulus may be of poor data quality requiring a long fixation to resolve perceptually. Third, attention allocation may be allocated elsewhere producing what Harris and Spady (1985) have referred to as a "blank stare."

All four of these sources of ambiguity may, to some extent, be resolved by recording ERPs elicited by the stimuli in question. For example, a large ERP will discriminate the long fixation resulting from high information, from that resulting from poor data quality or the "blank stare." Therefore, the best "map" of attention to the perceptual world seems to be one that takes advantage of information provided jointly by ERPs and eye fixations.

Information Processing Stages

Studies of latency and amplitude changes have dominated much of the research in ERPs. Yet, in many respects, errors are the most critical variable to be predicted in operational environments, and there is a growing interest in the causes and sources of human errors (Rouse & Rouse, 1983; Norman, 1981; Reason, 1984). One important scheme has associated different classes of errors with different stages of information processing. An important distinction is made between errors of perception or "mistakes," in which a stimulus event is inappropriately categorized, and errors of response selection or "slips" in which an inappropriate action is given to a stimulus event that was correctly categorized.

From a human engineering point of view, the distinction between mistakes and slips is important because different corrective actions should be taken if one or the other form of error is prevalent. Mistakes result because stimulus events are ill-defined, poorly resolved or this categorization is not well learned by the perceiver. A corrective, therefore, would be one that would improve the display format for a particular piece of equipment. In marked contrast, slips occur with highly automated, well learned sequences of behavior, and are often a consequence of a more frequently given response "capturing" the stream of behavior and dominating the appropriate response of lesser frequency (Reason, 1984; Norman, 1981). This will be likely to occur when the level of automaticity is such that careful,-concise monitoring of the selection of responses is not carried out. Corrective system redesign to eliminate slips would involve restructuring the placement of controls, and association of controls to stimuli in a manner designed to prevent their occurance. The importance of ERP measures in discriminating slips from mistakes lies in their insensitivity to response factors (Region 3 in Figure 1, and as shown in Figure 3). P300s produced by mistake will be delayed, probably poorly resolved, and associated with equally long (if not longer) response times. P300s produced by slips may also be longer than those associated with mistakes, but will also be coupled with response times that are quite short, relative to P300 latency (Donchin, et al., in press).

Summary

Human Factors is a blossoming field, with a clear need for new and innovative methodclogies. This neeed is enhanced by the rapid evolution of computerized systems with which humans must interact. Therefore it is with some confidence that one can foresee the application of well-modeled, well understood techniques based upon ERPs, taking place in the future. The many steps in Figure 2 suggest however that these applications may not occur within the immediate future. Yet with the proper selective focus on relevant paradigms, derived by listening to and interpreting the needs of the human factors practitioner, and addressed with concern for real world complexity, there is no reason why the process cannot be accelerated.

Acknowledgments

I would like to thank E. Donchin, A. Kramer, E. Sirevaag for providing helpful comments on earlier drafts of this paper.

FIGURE CAPTIONS

Figure 1: Three Regions of Variance. Region 1: Association. Region 2: Dissassociation (performance insensitivity). Region 3: Dissassociation (ERP insensitivity)

Figure 2: Sequence of steps necessary for complete and satisfactory application of ERP methodology to Human Factors design issues. The text describes the differenct step munbers.

Figure 3: Causes of region 3 dissociation. The figure depict three processing stages at the bottom, only the first two of which reflect P300 latency. The third, response stage does not. This stage also depends upon separate processing resources from those that supply P300-dependent stages (top of figure). Therefore, variation of response load will affect RT and global workload measures, but not P300 latency nor P300 amplitude.

References

Adams, J. (1972. Research and the future of engineering Psychologist. American Psychologist, 27, 615-622.

Beatty, J. (1982). Task evoked pupillary responses, processing load, and the structure of processing resources. Psychological Bulletin, 91, 276-292.

Donchin, E. and Cohen, L. (1967) Average evoked potentials and intramodality selective attention. Electroencephalography and Clinical Neurophysiology, 22, 537-546.

Gomer, F. E. and Bish, K. G. (1978). Evoked potential correlates of display image qualilty. Human Factors, 6, 589-596.

Hansen, J. and Hillyard, S. W. (1980). Endogenous brain potentials associated with selective auditory attention. Elect Except and Clin Neurophysiology, 49, 277-290.

Harris and Spady (1985). Visual Scanning Behavior. Proceedings NAECON. New York: IEEE.

Hart, S. and Sheridan, T. (1984). Pilot workload, performance, and aircraft automotion. Proceedings NATO/AGAARD Proceedings #371, High performance Aircraft. NATO:Newlly

Horst, R. L., Murson, R. C., and Ruchkin, D. S. Processing inferences derived from event related potential measures in a monitoring task. In R. Swezey (Ed.) Proceeding 29th Annual Meeting of the Human Factors Society. Santa Monica. Human Factors.

Isreal, J. B., Chesney, G. L., Wickens, C. D., and 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., and Donchin, E. (1980). The event-related brain potential as an index of display-monitoring workload. Human Factors, 22(2), 212-224.

Junker, A.M., Kenner, K.M., and Casey, E.J. (1985). The effects of task difficulty on the steady state visual evoked response. In G. Lee (Ed.), Proceedings of the 10th Symposium on Psychology in the DOD. United States Air Force Academy.

Kramer, A. F., Wickens, C. D., and Donchin, E. (1983). Analysis of the processing requirements of a complex perceptual-motor task. Human Factors, 25(6), 597-621.

McCarthy, G., and Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. Science, 211, 22-80.

Mulder, G. and Mulder, L. J. (1981). Information processing and cardio vascular control, Psychophysiology, 18, 392-401.

Norman, D. (1981). Categorization of action slips. Psych Review, 88, 1-15.

North, R. (1985). A workload index for iterative construction design. Preocceding NAECON. New York: IEEE.

Parasurman, R. Event related brain potentials and intermodel divided attention.

Posner, M. (1982). Cummulative development of attentional theory. American Psychologist, 37, 168-179.

Reason, J. J. (1984) Lapses of attention in enemy day life. In R. Parasurman and D. Davies. Variaties of Attention, Orlando Florida: Academic Press.

Rouse, W. (1985) On better mouse traps and basic research: getting the applied world to the labatory door. IEEE Transaction on Systems, Man and Cybernetics, SMC-15, 2-8.

Rouse, W. B. and Rouse, S. H. (1983) Analysis and Classification of Human Error. IEEE Transaction on Systems, Man and Cybernetics, SMC-13, 539-549.

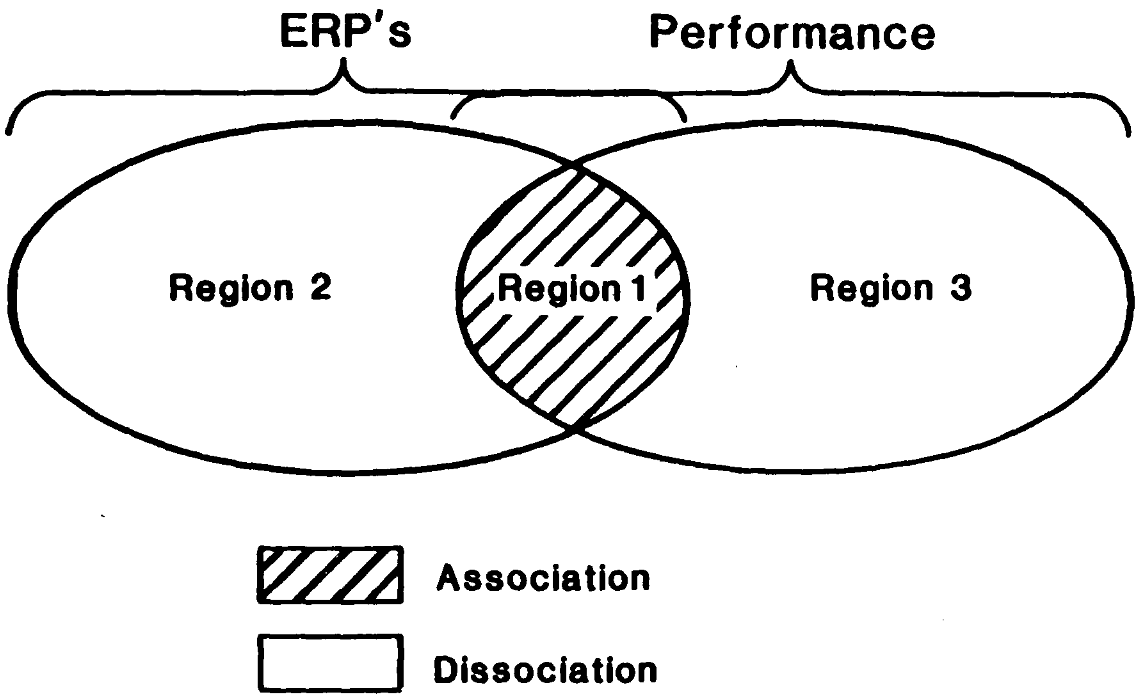
Sirevaag, E.,

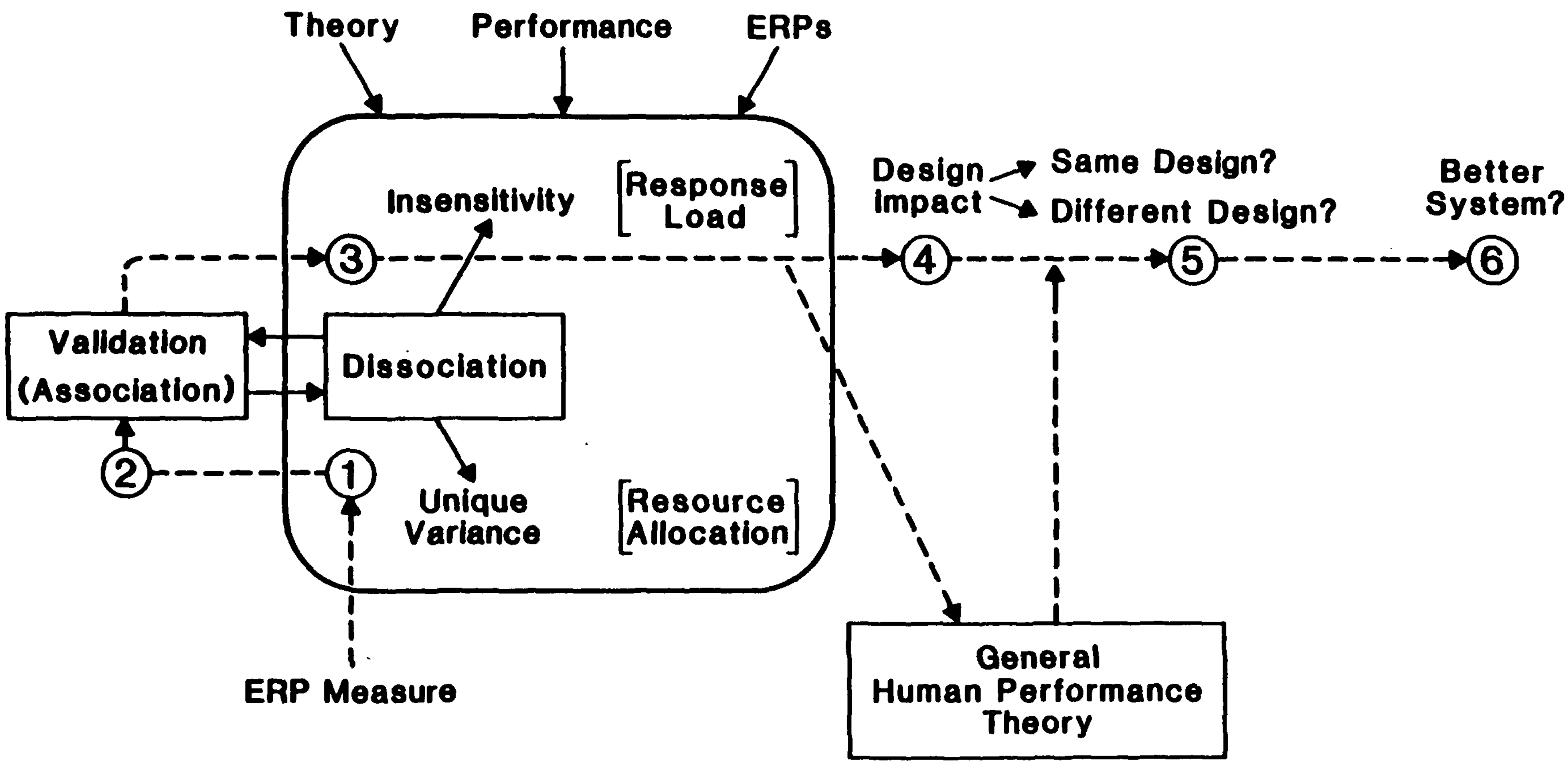
Wickens, C. D. (1986) The effects of control dynamics on manual control performance. In L. Kaufman and K. Boff (Eds.) Handbook of Perception and Performance, Volume 2. John Wiley.

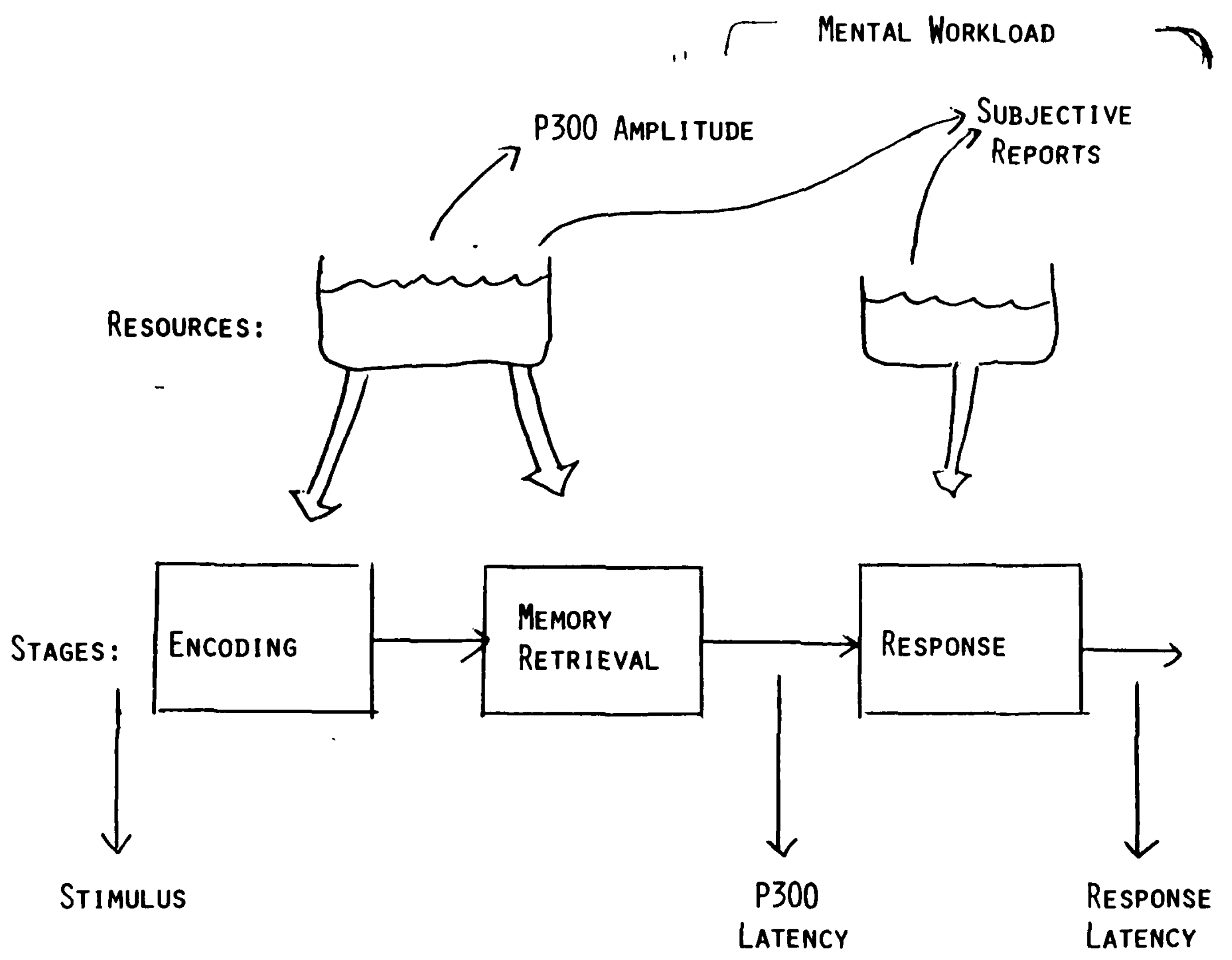
Wickens, C., Kramer, A., Vanasse, L., and Donchin, E. (1983). The performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information processing resources. Science, 221, 1080-1082.

Wickens, C. D., Heffley, E., Kramer, A., And Donchin, E. The event-related brain potential as an index of attention allocation in complex displays. Proceedings of the 24th Annual Meeting of the Human Factors Society. Santa Monica, CA, 1980.

Sources of Variance




Theory
Performance
ERPs
Insensitivity
Response Load
Design Impact
Same Design?
Different Design?
Better System?
3
4
5
6
Validation (Association)
Dissociation
2
1
Unique Variance
Resource Allocation
ERP Measure
General Human Performance Theory


MENTAL WORKLOAD
P300 AMPLITUDE
SUBJECTIVE REPORTS
RESOURCES:
STAGES:
ENCODING
MEMORY RETRIEVAL
RESPONSE
STIMULUS
P300 LATENCY
RESPONSE LATENCY

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 K02-14 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

Descriptors: Event-related brain potentials, P300, memory, mnemonic strategies, von Restorff effect.

Abstract

Subjects were instructed to use either rote or elaborative strategies to memorize words in a von Restorff paradigm. When instructed to use rote strategies, subjects displayed a higher von Restorff effect and a lower recall performance than when instructed to use elaborative strategies. Furthermore, the amplitude of the P300 component of the event-related brain potential predicted subsequent recall only when subjects used rote strategies. When subjects used elaborative strategies, the relationship between P300 amplitude and subsequent recall was not observed.

These results confirm and expand, in a *within subjects* design, the results reported by Karis, Fabiani, and Donchin (1984) who capitalized on different strategies used by *different* subjects. These results also lend support to a three-phase model of the von Restorff effect.

Introduction

Karis, Fabiani, and Donchin (1984) reported that people who used rote rehearsal strategies were more likely to recall stimuli that had elicited large P300s. This relationship between P300 amplitude and subsequent recall was not observed in subjects who used elaborative strategies. The choice of rehearsal strategies in the Karis, Fabiani, and Donchin (1984) experiment was left to the subjects and was ascertained only in a post-experimental debriefing. In the present study the subjects' strategies were manipulated by instructions and each subject performed the task using both rote and elaborative rehearsal strategies. This within-subject design is necessary to confirm the original report by Karis, Fabiani, and Donchin (1984).

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, it 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; Gehring, Gratton, Coles, & Donchin, 1986; Kutas, McCarthy, & Donchin, 1977; Munson, Ruchkin, Ritter, Sutton, & Squires, 1984). For instance, Donchin, et al. (in press), used a choice-reaction time task in which the probability of the stimuli was manipulated and found that subjects were biased in favor of the frequent response, and often erred when a rare stimulus was presented. However, when a large P300 was emitted after an incorrect response to a rare stimulus, the response bias in the following trials was reduced. Similarly, Gehring et al. (1986) observed that the amplitude of P300 after incorrect responses to rare events predicted shifts in the subjects' response strategies.

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)[1] and on variations in inter-stimulus interval (Heffley, 1981; Fitzgerald & Picton, 1981), are consistent with the hypothesis that P300 is the manifestation of a process invoked when there is a need to revise the representations in working memory (or "context updating"; Donchin, 1981)[2]. When a rare event occurs, or an error is made, or conditions change, this information must be incorporated into cognitive schemas. These schemas will govern both the perception and action taken on future trials, and will also affect the recall of the information of the specific trial.

The Role of Novelty in Learning: P300 and Memory

In recent years, the role of novelty in learning has been emphasized 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 theorists appear to 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 in this way can an accurate representation of the environment 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 the 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 von Restorff study. Karis, Fabiani, and Donchin (1984) used the von Restorff paradigm (von Restorff, 1933) to test the hypothesis that P300 amplitude predicts subsequent recall. They presented series of unrelated words and recorded the ERP elicited by each word. After each series the subjects were asked to recall as many words of the series as possible. Most of the series contained a deviant word (an "isolate"), with the isolation 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 according to the magnitude of the 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 (a high von Restorff effect). These subjects reported using primarily rote strategies (e.g., repeating the words). It was only in these subjects that Karis, Fabiani, and Donchin (1984) observed a positive relationship between P300 amplitude and recall. At the other extreme, subjects in group 3 which showed no relationship between P300 and recall, exhibited high overall recall percentages, and no isolation. These subjects reported using complex elaborative strategies (e.g., making up sentences, stories or images). The amplitude of a frontal-positive slow wave"[3] was related to subsequent recall, and the elaborators exhibited more evidence of this component than the rote memorizers. Karis, Fabiani, and Donchin (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, Fabiani, and Donchin (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 rehearsal strategies that influence recall often continue for an extended period. The relationship between P300 and recall will thus depend on the nature of the extended mnemonic processing during rehearsal. The results of the Karis, Fabiani, and Donchin (1984) study are consistent with a three-phase model of the processes that determine the subjects' recall of the words. In the first phase, in which the stimuli are encoded and categorized, the subjects' information processing is independent of their subsequent use of rehearsal strategies. We assume that the process manifested by the P300 is activated by the words that are presented in the distinct, or "isolated," font. Karis, Fabiani, and Donchin 1984) examined the distribution of P300 amplitude in all subjects and found no difference between the different groups of subjects. It would appear that all subjects noticed the isolated words to the same extent and, as a consequence, a P300 was always elicited in response to the isolates. We assume that the representations of the isolated words were "marked" or activated in some fashion, and this is reflected by P300 amplitude. It is at this point that the subjects diverged into groups. In the second phase the subjects engaged in various activities designed to enhance their ability to recall the stimuli. Some subjects opted, when left to their own devices, to adopt rote-rehearsal strategies, while other subjects engaged in various semantic elaboration strategies. When, in the third phase, the subjects were required to recall the words, the rote-memorizers, we proposed, retrieved words on the basis of the initial representations created during phase 1. Some of these representations were marked or activated by their association with P300 and the larger the P300 the stronger the mark or the activation and the more likely the recall. On the other hand, at the moment of retrieval the elaborators relied on the new representations they created during the elaboration process, and these representations did not carry the "marks" associated with the P300; hence the lack of correlation between recall and P300 amplitude in the elaborators.

An Incidental Memory Experiment. The effect of rehearsal strategies on the P300-recall relationship was not anticipated as we designed the study reported by Karis, Fabiani, and Donchin (1984). Thus, no control was attempted over rehearsal strategies. To confirm the results we examined the relationship between P300 and recall in a paradigm where subjects did not have any reason to adopt an elaborative strategy (Fabiani, Karis, & Donchin, 1986). We embedded an incidental memory task in another study (Fabiani, Gratton, Karis, and Donchin, in press; Karis, Coles, & Donchin, 1984), in which subjects were presented with a succession of oddball tasks[4], none of which involved recalling the stimuli. After series in which the subject was instructed to respond differentially to tones, or to count either the letter "H" or "S," we inserted a series composed of randomly mixed male and female names, and instructed the subject 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 elaborative rehearsal strategies to facilitate recall. In this situation, the relationship between P300 and recall could be evaluated in the absence of elaborative processes occurring after P300. Our main prediction was that names that were recalled would elicit larger P300s on their initial presentation than names that were not recalled. Our prediction was confirmed and no consistent differences emerged among subjects, thus indicating that, when elaborative strategies are prevented, the amplitude of P300 predicts subsequent recall.

The Present Study. Although the Fabiani et al. (1986) data are consistent with the predictions derived from the Karis, Fabiani, and Donchin (1984) study, they are based primarily on a negative finding; namely, that individual differences due to rehearsal strategies will be reduced when an incidental memory test is employed, thus allowing for the P300/recall relationship to emerge more consinstently. However, the individual differences observed by Karis, Fabiani, and Donchin (1984) could still be attributed to the subjects' idiosyncrasies rather than to the use of particular mnemonic strategies. Therefore, it is crucial to determine whether recall performance is influenced by the interaction between the updating process (manifested by P300) and mnemonic strategies not only between but also within 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, Fabiani, and Donchin (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 (both sets of instructions are reported in the Appendix). We expected that the same subject would behave like a rote memorizer of the Karis, Fabiani, and Donchin (1984) study when given rote instructions, and like an elaborator when given elaborative instructions.

This study also allowed a test of predictions derived from the "distinctiveness hypothesis," which has been invoked to explain the von Restorff effect (Bird, 1980), as well as the enhanced recall of distinctive items (Fisher, 1981; Hunt and Elliott, 1980; Hunt & Mitchell, 1978; 1982; Schmidt, 1985; Zechmeister, 1972). Distinctive features of an item are those shared by few other items. The distinctive features of a to-be-remembered item, being unique, will be more useful in retrieval than common features, shared by many other items. Thus, the distinctiveness hypothesis focuses on the utility of the encoded information. However, the utility of the encoded information is also relative to the context in which the information is retrieved. As Fisher (1980, p. 310) reports, "a distinctive encoding operation will improve retention to the extent that the encoding context is reinstated at retrieval." In the present experiment, the deviant size of the word can be considered a "distinctive" attribute of the memory representations of the isolates. However, this size attribute should only be useful at retrieval after the subjects have used rote rehearsal strategies. Therefore, we hypothesized that subjects will have a better memory of the word size when they have used rote strategies than when they have 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 which extended over 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. The subjects were grounded via Ag-AgCl Beckman Biopotential electrodes, affixed to the forehead by means of adhesive collars. The same type of electrodes were affixed above and below the right eye to record the vertical electrooculogram (EOG), and on the mastoids. 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 analyzed by using an off-line moving average (1 iteration) corresponding to a 3.14 Hz filter (Ruchkin & Glaser, 1978).

Word Lists

Five word lists were constructed. The first list was used for all the subjects during the first session (list F), and contained 40 series of 15 words. The remaining 4 lists were used for the second and third sessions (a pair per session - Lists A-B and C-D). Each of these lists contained 20 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 anywhere 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).

In list F (the list used for the first session), the first 10 series were used for the "No strategy instruction" condition. Eight of the series contained an isolate. The next 15 series were used for the "Rote strategy" condition, and 11 contained an isolate. Finally, the last 15 series were used for the "Elaborative strategy" condition, and 11 of these also contained an isolate. In lists A, B, C, and D (used for the second and third sessions), 15 of the 20 series contained an isolate. 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 2000 ms 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.

There were three sessions, with 9 to 26 days (median = 17 days) between the first and second session, and 1 to 5 days between the second and third session (median = 2 days). The first session was 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 second and the third. The general procedure and instructions were similar to those used by Karis Fabiani, and Donchin (1984), with the exception of the strategy instructions. As shown in Figure 1, 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).

Insert Figure 1 about here

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. This ensured 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 warning ("ready?") from the experimenter (via an intercom), another series was presented.

In the _first session (baseline and training session)_, 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 series. 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 among words. No strategy instructions were given to the subject before the first group of 10 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 again 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 sessions and subjects. The subjects were given short (approximately five minutes) rest periods after every ten series. At the end of each block 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 the members of each group were run together (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 in manipulating *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, $F(1,25) = 25.67$, $p < 0.0001$.

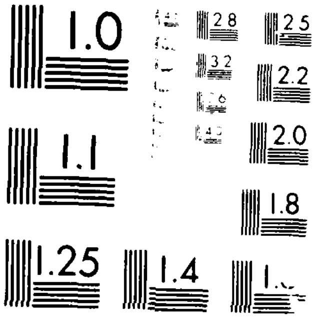
In order to minimize variations from session to session and from subject to subject the instructions were recorded on tape. The subjects were told that they could interrupt the tape if they had questions, and were debriefed at the end of the recording to make sure they understood the instructions correctly. 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. The strategy instructions are presented in the Appendix.

Even though our instructions had proven effective in the pilot study, 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 von Restorff index (a measure of the magnitude of the von Restorff effect). 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 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 the third session (half under rote and half under elaborative instructions), randomly interspersed with an equal number of non-isolated words (half from each strategy instruction). All words in the list were of course printed in the same size font. 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.

AD-A191 244 THE EUENT-RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING A..(U) ILLINOIS UNIU CHAMPAIGN COGNITIUE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 7/10

UNCLASSIFIED 29 FEB 88 CPL-88-1 AFOSR-TR-88-0316 F/G 5/8 NL



MICROCOPY RESOLUTION TEST CHART

Results

Data from the first session (baseline and training sessions) 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 there were no significant order effects (details are presented below). Since the data obtained in these two sessions were comparable, they were combined for all the subsequent analyses.

Analysis of Recall

As in the Karis, Fabiani, and Donchin (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 = the percentage of isolated words recalled (position 6-10) minus the 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 computation 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 1.

Insert Table 1 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, Fabiani, and Donchin (1984) study. On the other hand, the VRI was highly unstable in this condition, perhaps because it was 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 did not reach significance, $F(1,9) = 4.69$, $p = .059$.

Second and Third sessions. 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 2.

Insert Table 2 about here

The two strategy instruction 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 2 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 ERP 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 2a for rote strategies and in figure 2b for elaborative strategies.

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. This was done because there were too few words to plot every position. 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 higher elevation of the curve for the isolated items relative to that for the non-isolates in comparable positions (6 through 10). Clearly, the von Restorff effect is large under rote instructions (Figure 2a), while it is almost absent under elaborative instructions (Figure 2b). Performance, on the other hand, is higher under elaborative than under rote instructions.

Figure 3 compares the results obtained in the present study with the results reported by Karis, Fabiani, and Donchin (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, Fabiani, and Donchin (1984) study are connected by a solid line. The means for the subjects in the present study are connected by a dashed line. Note that the means from the Karis, Fabiani, and Donchin (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 difference between mean values is smaller in the present study than was in the Karis, Fabiani, and Donchin (1984) study. However, it is important to note that, as in that study the subjects chose the strategies they preferred, one might expect more extreme values than in the present study.

From the data reported so far we 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, Fabiani, and Donchin (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 elaborative 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 (48%) to session 3 (53%), $F(1,9) = 7.05$, $p < .03$. 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 combined across the two experimental sessions (session 2 and 3).

Isolation Effect. In order to determine 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), word class (isolates, non-isolates in isolated series and non-isolates in control series) and position (position 6-10 only). The ERP waveforms, averaged over all the subjects 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 and position (isolates, non-isolates in position 6-10, non-isolates in other positions) and subsequent recall (recalled, not-recalled).

ERP waveforms, averaged over all subjects 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; this is most evident at the central and parietal electrodes. Second, there is a difference between isolates recalled and not recalled at the frontal electrode. This is most evident for elaborative instructions, and 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 et al., 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. Next, 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 type), $F(14,112) = 3.99$, $p < .0001$. 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), $F(1,8) = 5.72$, $p < .05$.[5] Note that this effect is present for all the words, not just the isolates. The mean P300 amplitudes (in arbitrary units) are presented in Table 3a for the isolates and in Table 3b for all the words. Note that, even though the amplitude of P300 seems to be slightly larger for the elaborative instruction condition, that result is not significant, ($F(1,8) = 3.17$, $p = .11$).

Insert Table 3 about here

The amplitude of the P300 in the average waveforms for the non isolated words appears quite small (see figure 4). These may be attributed, in part, to subject-to-subject variability in the latency of P300. Therefore, in the analysis presented above, we assessed the amplitude of P300 taking into account the inter-subject latency variability. We also computed average waveforms over all the subjects, where the waveforms from each single subjects were shifted so that the P300 peaks were aligned. These waveforms are shown in Figure 6 for both strategy conditions. An amplitude difference between words recalled and not recalled is noticeable under rote, but not under elaborative instructions.[6]

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 measured the area under the curve in the ERP for Fz for each subject, word type and condition, in a time window ranging from 800 to 1180 ms. An area measure was chosen in this particular case because of 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), $F(1,8) = 57.90$, $p < .001$.

Thus, the analyses of the ERP waveforms confirm our predictions: the isolates elicit larger P300s than the non-isolates in both strategy conditions. In addition, the relationship between P300 amplitude and memory emerges only when rote rehearsal strategies are used. Finally, a frontal-positive slow wave is elicited when the subjects engage in elaborative processing, and the amplitude of this component is also related to <u>subsequent</u> recall.

<u>Analysis of the Size-Recall Test</u>

The size-recall test was devised to test the hypothesis that the subjects would have a better memory of the <u>size</u> 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 instruction 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) - $F(1,8) = 4.81$, $p = .06$.

Discussion and Conclusions

The results of this study confirm the predictions of the Karis, Fabiani, and Donchin (1984) model. Our strategy manipulation was successful in affecting both recall performance (more words were recalled by using elaborative strategies than by using rote) and the VRI (which was higher under rote instructions than under elaborative). The isolates elicited larger P300s than the non-isolates in both strategy conditions, thus reinforcing the claim that the initial processing of the items is the same for all subjects, independently of what type of rehearsal strategy is subsequently used. The relationship between P300 amplitude and recall emerged only when rote rehearsal strategies are used. On the other hand, a frontal-positive slow wave was elicited when the subjects were engaged in elaborative processing, and the amplitude of this component was also related to subsequent recall. It is important to note that the predictions derived from the model were confirmed not only for the isolated words, but for other words as well.

Results of the size-recall test suggest that the word size is a "distinctive" attribute of the memory representation of the isolated word. Subjects were significantly more accurate in identifying the size of the words presented to them under rote than under elaborative instructions. It is plausible that under both strategy instructions the size attribute was "marked" as distinctive when an isolate appeared (a larger P300 to the isolates than to the non-isolates was observed for all the subjects). However, when the words were memorized under rote instructions, the subjects may have also used the size attribute in their retrieval search. When they were instructed to use elaborative strategies, however, they were able to use their stories or images as an aid to retrieval, and did not "rehearse" the size attribute, thus having a poorer memory for it.

Since our original experiment (Karis, Fabiani, & Donchin, 1984) several studies investigating the relationship between ERP components and memory have been reported (e.g., Fabiani, et al., 1986; Johnson, Pfefferbaum, & Kopell, 1985; Neville, Kutas, Chesney, & Schmidt, 1986; Paller, Kutas, & Mayes, 1985). These studies differed in several respects from the Karis, Fabiani, and Donchin (1984) study. In fact, Fabiani et al. (1986) used a name oddball paradigm, followed by an incidental free recall task. Neville et al. (1986) presented the subjects with statements followed by words that could be either semantically congruous or incongruous with the preceding statement, and used an incidental recognition to test memory for the words. Paller et al. (1985) used both incidental recall and incidental recognition to test memory for words that were previously categorized on the basis of their meaning or orthography. Finally, Johnson et al. (1985) used a study-test paradigm, in which each study list was presented to the subjects four times, and was followed each time by a recognition test to assess performance. By and large, these studies have confirmed that a large parietal positivity is related to subsequent memory performance, when conditions similar to those of the Karis, Fabiani, and Donchin (1984) study were employed. When significant variations were introduced in the study, the results were more ambiguous. For instance, investigators usually do not examine the subjects rehearsal and retrieval strategies. Given the very powerful effects that these strategies have on the relationship between P300 and recall, it is evident that studies in which this factors are not given consideration may yield conflicting results.

It is also important to emphasize that we are not reporting a general relationship between P300 and memory. The Karis, Fabiani and Donchin (1984) data are sometimes interpreted as if such a general claim is made. The failure to confirm this putative universal relationship between P300 and memory is then considered relevant to interpretation of the P300 (e.g., Johnson et al., 1985). It may be useful therefore to reiterate that our theoretical stance does not assert that every manipulation known to affect "memory" will also affect P300. Neither are we suggesting that an enhancement of the P300 should accompany each improvement in the subjects' ability to recall, recognize, or otherwise indicate their having information in memory. Such a universalist position would not make much sense on general grounds, and it is clearly inconsistent with the demonstration that P300 is related to recall, in the von Restorff paradigm, only if rote rehearsal strategies are used.

The label "memory" applies to a large variety of processes whereby information about past events is made more or less accessible to retrieval. It is doubtful that all of these instances are the result of the application of a single class of mechanisms. Whether an event will be recalled or not depends on a large number of independent processes, each subject to a plethora of variables.

Of course, speaking of memory, one is speaking about the creation, storage, maintenance and use of representations. The P300, or, more precisely, the process that the P300 manifests, can make contact with memory only through the effect that its invocation may have on the establishment, manipulation, or access to these representations. The specific manner with which such a relationship is established is unknown at this time. The data we report here suggest that the invocation of the P300 process, and the magnitude with which it is invoked, affects in some manner "episodic" representations, that is those representations that preserve the characteristics of the stimuli as physical entities. We infer this from the fact that it is the rote memorizers who are aided by the consequences of the P300, and from the fact that these subjects also recall the physical size with which the word was presented. The specific nature of the way the representations are "marked" or "activated" remains to be determined. Furthermore, we are not even certain that the invocation of the P300 process is the direct causal effect of this change. It is possible that the marking of the representation and the elicitation of the P300 are both the consequences of yet another unspecified process. These remain issues for further research.

It may be useful to assert also a methodological caveat. We suggest that, in considering relationships between P300 and "memory," investigators should choose paradigms in which they narrow the range of variables that can affect recall to variables that are also likely to affect the amplitude of the P300. This, for example, has been to our mind the benefit of the use of the von Restorff paradigm. In previous work (Karis, Bashore, Fabiani, & Donchin, 1984) we attempted to test the hypothesis that P300 amplitude would predict recall by using a general recognition paradigm. The predicted relationship between P300 amplitude and recall was indeed observed, but only weakly. This was so, we believe, because the amount of shared variance between the processes controlling P300 amplitude and those controlling "memory" is relatively small in a global recognition experiment. By introducing the "isolates," Karis, Fabiani, and Donchin (1984) could study the relationships as they operate in the context of stimuli whose recall is largely controlled by the isolation and this very isolation is known to have a strong effect on P300.

As a final remark, we submit that, with the proper choice of experimental paradigms, P300 and other components of the human ERP can serve as tools in the study of cognitive function. The utility of this family of tools derives from their special ability to allow direct observation of the activity of intermediate information processing states. In the main, the standard methodology of Cognitive Psychology depends on what may be called "final-outcome" studies. That is, what is available for observation are specific behavioral acts, such as button presses or verbal reports. These acts are the observable consequences of a multiplicity of information processing activities (not necessarily serial). A diversity of methods permits inferences regarding the fine structure of the human information processing system from observations of these final outcomes and their distributions. The data presented here illustrate the process whereby the study of human information processing can be augmented by use of observations on such intermediate outputs as the P300. This component serves as an index for activity that takes place during the encoding stages, and this activity definitely affects subsequent recall, as a function of rehearsal strategies. Thus, these data can serve in the development of models of memory.

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, (pp. 47-89) (Vol. 8). New York: Academic Press.

Bird, C. P. (1980). The isolation effect as a function of unique processing orientation. Journal of Experimental Psychology: Human Learning and Memory, 6, 267-275.

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 (pp. 101-110). New York: Academic Press.

Dickinson, A. (1980). Contemporary animal learning theory. New York: Cambridge University Press.

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., 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., & Heffley, E. (1975). Minicomputers in the signal-averaging laboratory. American Psychologist, 38, 449-461.

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 (pp. 349-441). New York: Academic Press.

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). 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.

Fabiani, M., Karis, D., & Donchin, E. (1986). P300 and recall in an incidental memory paradigm. Psychophysiology, 23, 298-308.

Fisher, R. P. (1981). Interaction between encoding distinctiveness and test conditions. Journal of Experimental Psychology: Human Learning and Memory, 7, 306-310.

Fitzgerald, P.G., & Picton, T.W. (1981). Temporal and sequential probability in evoked potential studies. Canadian Journal of Psychology, 35, 188-200.

Gehring, W., Gratton, G., Coles, M.G.H., & Donchin, E. (1986). Response priming and components of the event-related brain potential. Psychophysiology, 23, 437-438.

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. Manuscript 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, MS, 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.

Hunt, R. R., & Elliott, J. M. (1980). The role of nonsemantic information in memory: orthographic distinctiveness effects in retention. *Journal of Experimental Psychology: General*, *109*, 49-74.

Hunt, R. R., & Mitchell, D. B. (1978). Specificity in nonsemantic orienting tasks and distinctive memory traces. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 121-135.

Hunt, R. R., & Mitchell, D. B. (1982). Independent effects of semantic and nonsemantic distinctiveness. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *1*, 81-87.

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., Jr., Pfefferbaum, A., Kopell, B.S. (1985). P300 and long-term memory: Latency predicts recognition performance. Psychophysiology, 22, 497-507.

Karis, D., Bashore, T., Fabiani, M., & Donchin, E. (1982). P300 and memory. Pychophysiology, 19, 328.

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.

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

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimulus with reinforcement. Psychological Review, 82, 276-298.

Munson, R., Ruchkin, D. S., Ritter, W., Sutton, S., & Squires, N. K. (1984). The relation of P3b to prior events and future behavior. Biological Psychology, 19, 1-29.

Neville, H. J., Kutas, M., Chesney, G., Schmidt, A. L. (1986). Event-related brain potentials during initial encoding and recognition memory of congruous and incongruous words. Journal of Memory and Language, 25, 75-92.

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 (Eds.), The orienting reflex in humans. Hillsdale, NJ: Erlbaum.

Paller, K. A., Kutas, M., & Mayes, A. R. (1985). An investigation of neural correlates of memory encoding in man. Psychophysiology, 22, 607.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychological Review, 87, 532-552.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), Conditioning II: Current research and theory (pp. 64-99). New York: Appleton-Century-Crofts.

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). Washington, DC: 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.

Schmidt, S. R. (1985). Encoding and retrieval processes in the memory for conceptually distinctive events. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *11*, 565-578.

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* (pp. 671-704). New York: Basic Books.

Sokolov, E.N. (1975). The neuronal mechanisms of the Orienting Reflex. In E.N. Sokolov and O.S. Vinogradova (Eds.), *Neuronal Mechanisms of the Orienting Reflex*, (pp. 217-235). Hillsdale, NJ: Erlbaum.

Squires, N.K., Squires, K.C., & Hillyard S.A. (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, NJ: Erlbaum.

Von Restorff, H. (1933). Uber die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschung*, *18*, 299-342.

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* (pp. 95-128).

Hillsdale, NJ: Erlbaum.

Wagner, A. R. (1978). Expectancies and the priming of STM. In S H. Hulse, H. Fowler, & W. K. Honig (Eds.), Cognitive processes in animal behavior (pp. 177-210). Hillsdale, NJ: Erlbaum.

Wallace, W.P. (1965). Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. Psychological Bulletin, 63 , 410-424.

Zechmeister, E. B. (1972). Orthographic distinctiveness as a variable in word recognition. American Journal of Psychology, 85, 425-430.

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. For example, with two equiprobable events, A and B, the P300 elicited by the last A in a sequence will be smaller in the sequence AA than BA. Similarly, in third order series, P300 amplitude will decrease from BBA to ABA to BAA to AAA.

2. We use the concept of working memory to emphasizes function (see, for example, Baddeley, 1981, and Baddeley & Hitch, 1974), as opposed to short-term memory, which has often been used to refer to a hypothetical structure.

3. This component was labeled "frontal-positive slow wave" to distinguish it from the more typical slow wave reported by several investigators (see Squires, Squires, & Hillyard, 1975; and Ruchkin & Sutton, 1983). In fact, the typical slow wave is negative frontally, becoming more and more positive as one moves back across the scalp.

4. In the "oddball paradigm" the subject is presented with a sequence of events and a categorization rule is provided that classifies each event into one of two categories. The events are presented in a Bernoulli sequence, and the subject is instructed to respond to each event in some manner that requires attention to the categorization rule. In general, if one of the categories appears more rarely than the other, the stimuli belonging to that category will elicit a P300, whose amplitude will be inversely proportional to probability. However, the rarity is neither necessary, nor sufficient, to elicit the P300. The task relevance of the events is a key element in determining the appearence of P300.

5. 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, 1986) 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.

6. Note that, by shifting the waveforms so that the peaks of the single subjects' P300s were aligned, the possibility of examining other ERP components (not time-locked to P300) was lost.

Table 1

Performance (P) and von Restorff Index (VRI) (Percentages) in Session 1 (Baseline and Training Session).

	Strategy Instructions					
	None		Rote		Elaborative	
S#	P(a)	VRI(b)	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)

(a) P (Performance): % recalled from all positions

(b) VRI (von Restorff index): % isolates recalled minus % non-isolates recalled (position 6-10)

Table 2

Performance (P) and von Restorff Index (VRI) (Percentages) in Session 2 and 3 (Experimental Sessions, Combined Values).

	Strategy Instructions			
	Rote		Elaborative	
S#	P(a)	VRI(b)	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)

(a) P (Performance): % recalled from all positions

(b) VRI (von Restorff index): % isolates recalled minus % non-isolates recalled (position 6-10)

Table 3

P300 Amplitudes (Arbitrary Units) as Measured by a Cross-Covariance Procedure at Pz, for Isolates and for All the Words.

A. ISOLATES

Strategy	Recalled	Not-Recalled
Rote	28.3	18.7
Elaborative	26.6	24.8

B. ALL THE WORDS

Strategy	Recalled	Not-Recalled
Rote	16.0	12.9
Elaborative	16.0	16.4

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, Fabiani, and Donchin (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, Fabiani, and Donchin (1984) study are indicated by a solid line. The means for the subjects in the present study are indicated by a dashed line.

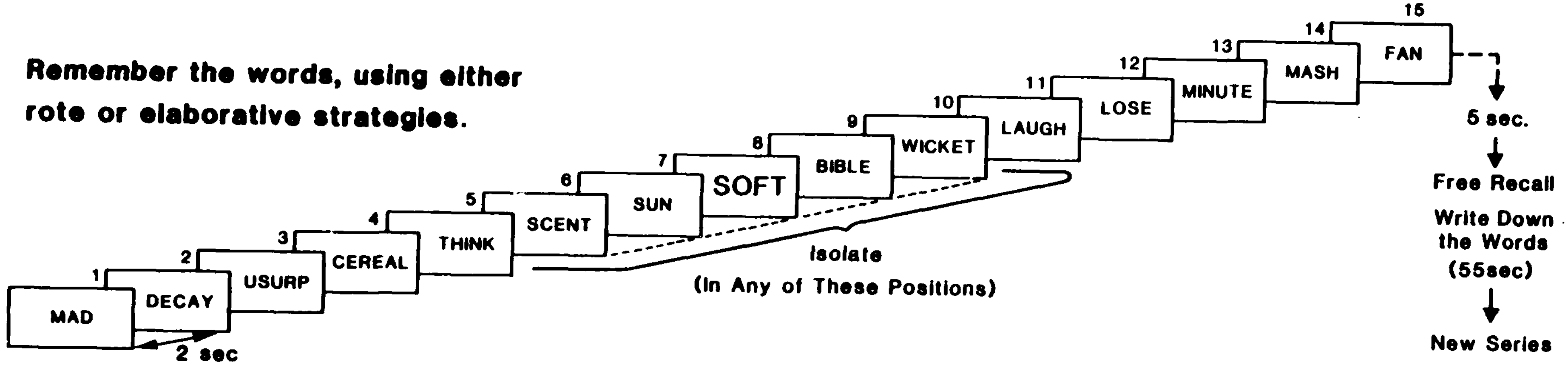
Figure 4. The ERP waveforms, averaged over all subjects at three electrode locations (Fz, Cz, and Pz) for isolates (solid line), non-isolates coming from isolated series (dashed line) and from control series (dotted line). The rote strategy waveforms are plotted in the left panel, the elaborative strategy waveforms in the right panel.

Figure 5. ERP waveforms averaged over all subjects at Fz, Cz and Pz for isolates recalled (solid line) and not recalled (dashed line). The rote strategy waveforms are plotted in the left panel, the elaborative strategy waveforms in the right panel.

Figure 6. ERP waveforms, averaged over all subjects, 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 recalled words by a dashed line. The vertical dashed lines represent the limits of the time window in which P300 amplitude was assessed. The rote strategy waveforms are plotted in the left panel, the elaborative strategy waveforms in the right panel.

Remember the words, using either rote or elaborative strategies.



Session 1

Ⓝ 8 isolated series
2 control series

Ⓡ 11 isolated series
4 control series

Ⓔ 11 isolated series
4 control series

Session 2

Ⓡ 15 isolated series
5 control series

Ⓔ 15 isolated series
5 control series

Session 3

Ⓔ 15 isolated series
5 control series

Ⓡ 15 isolated series
5 control series

Ⓝ = None
Ⓡ = Rote
Ⓔ = Elaborative

FIG.1

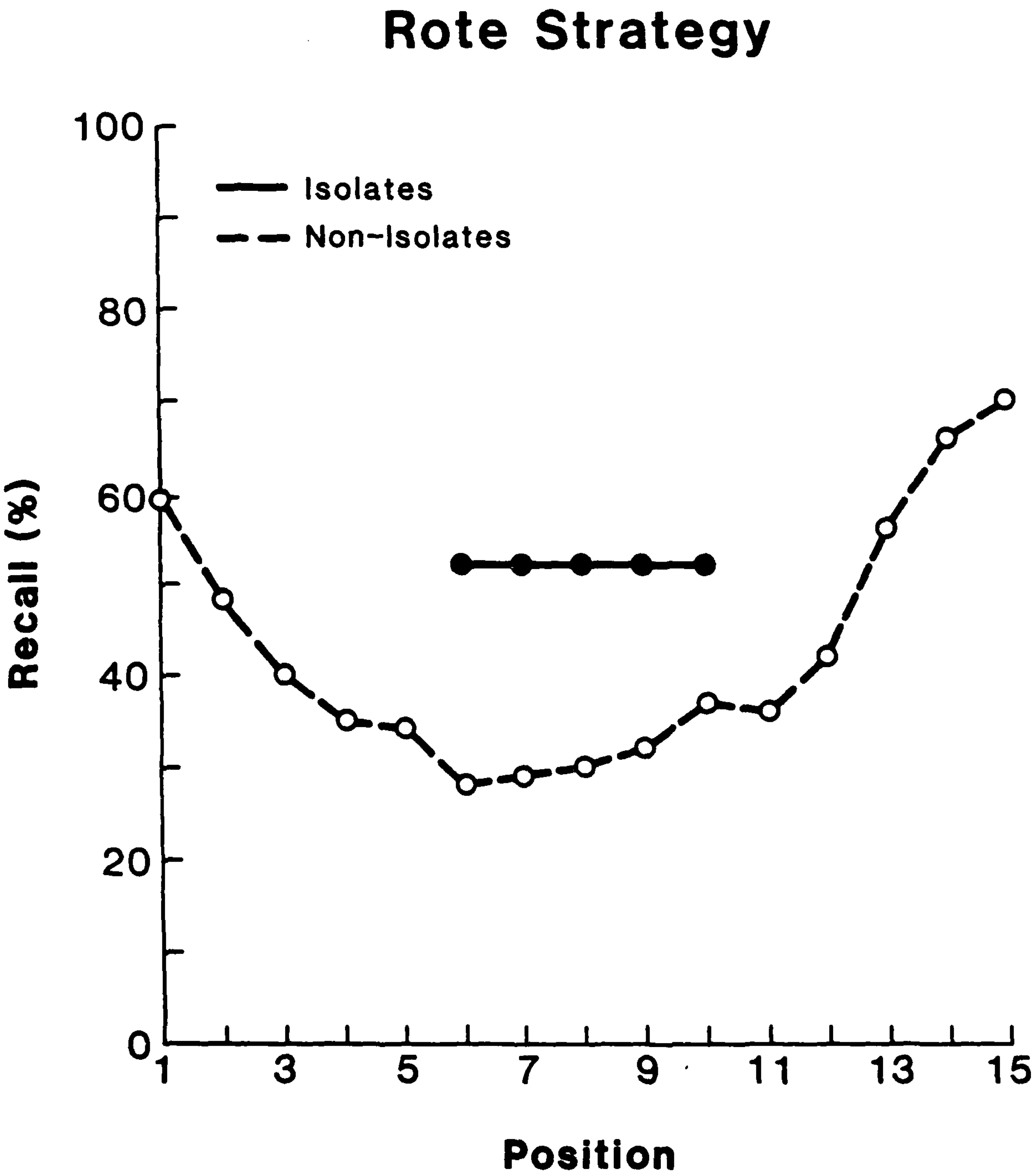

Rote Strategy
100
80
60
40
20
0
Recall (%)
Isolates
Non-Isolates
1
3
5
7
9
11
13
15
Position

FIG. 2a

Elaborative Strategy

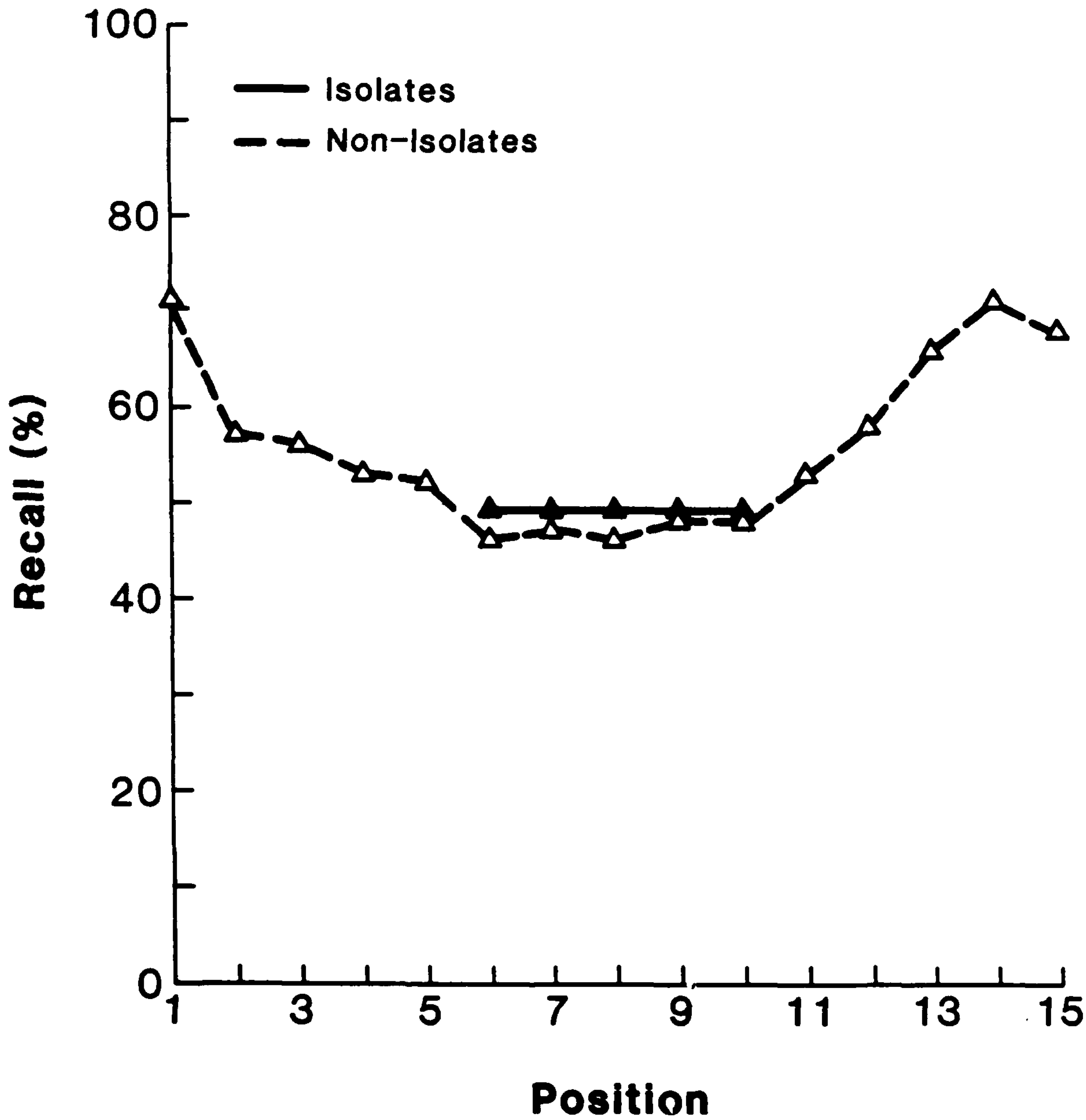
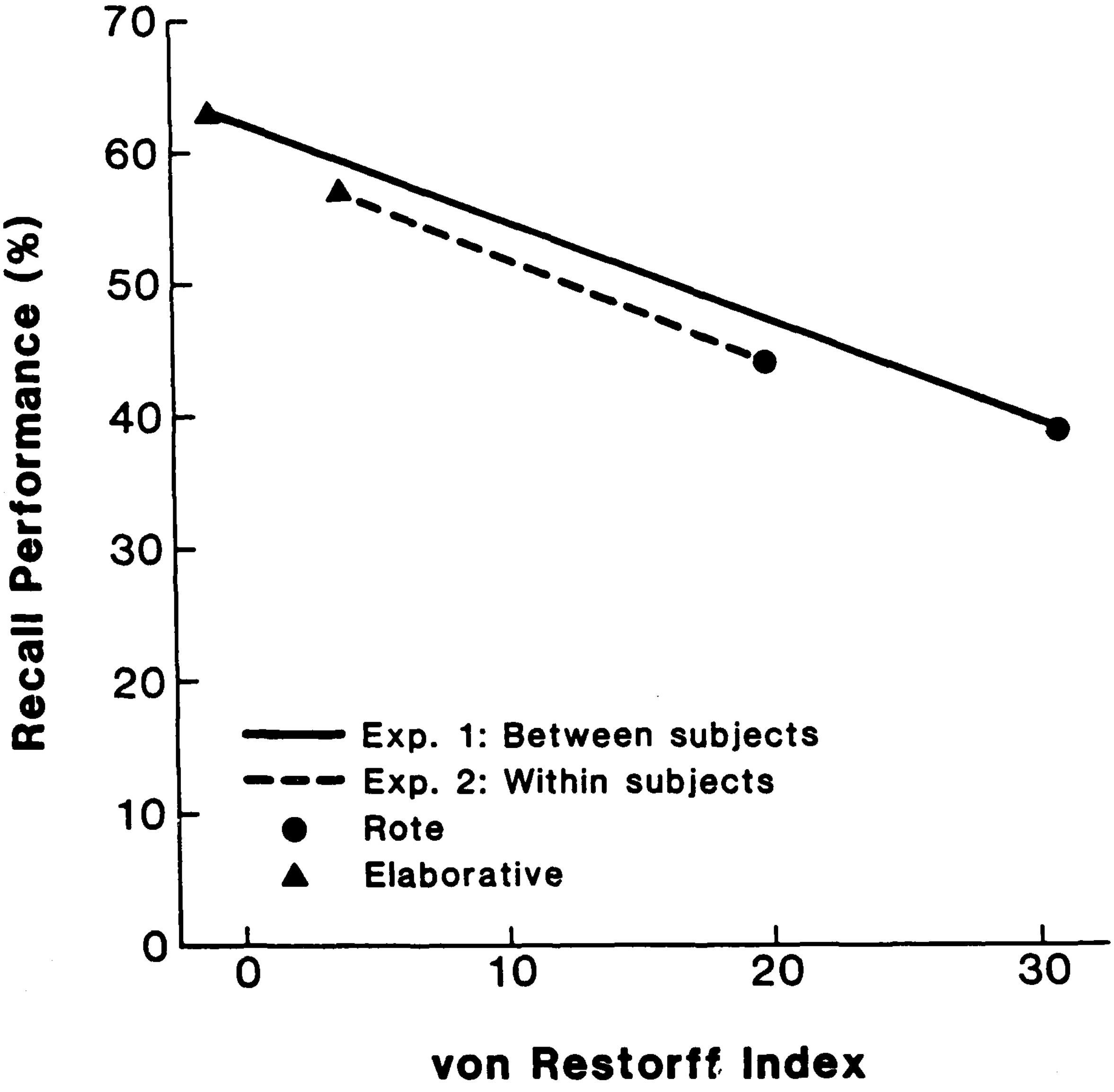


FIG. 2b



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

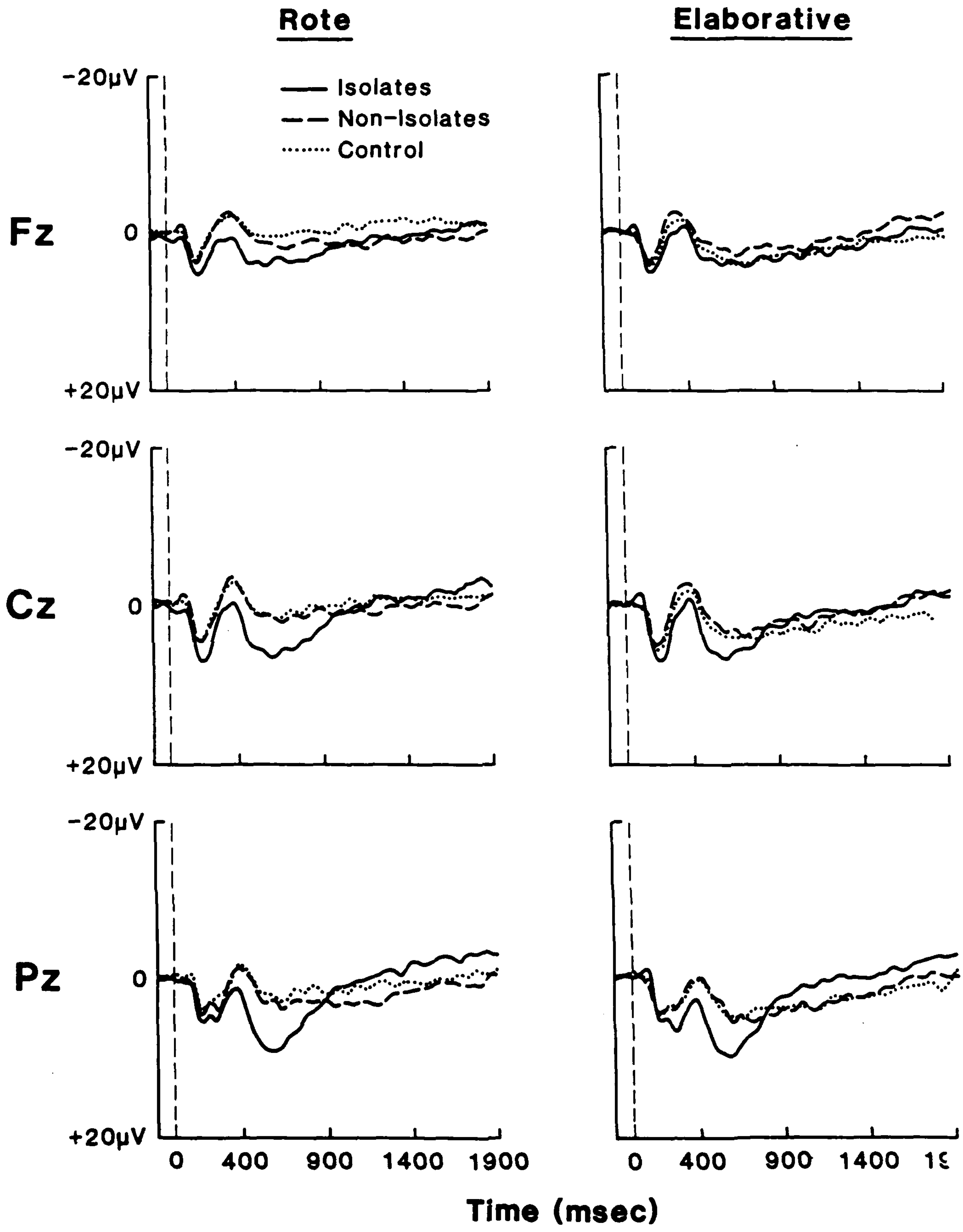
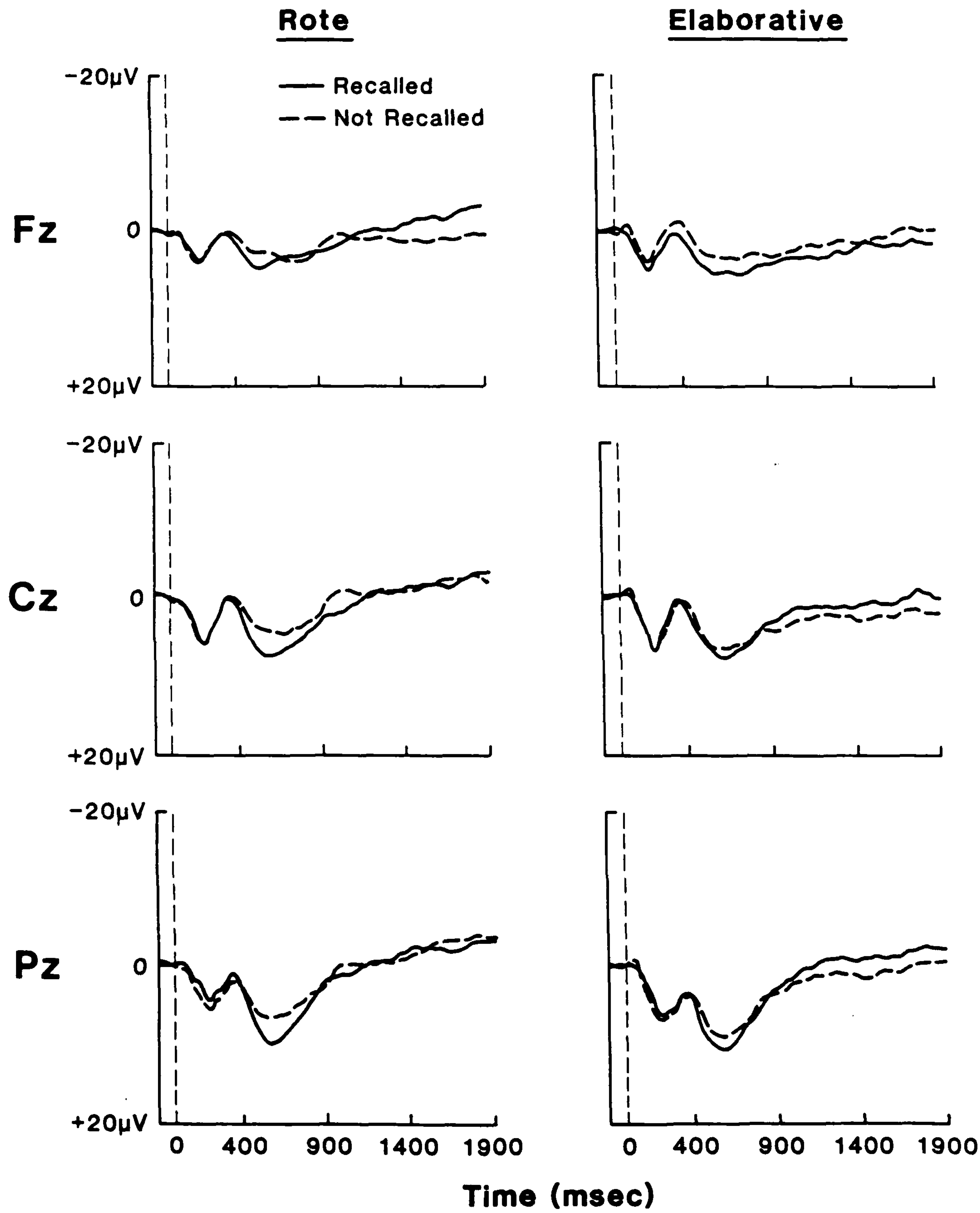

Rote
Elaborative
Isolates
Non-Isolates
Control
-20μV
0
+20μV
Fz
Cz
Pz
0
400
900
1400
1900
Time (msec)

FIG.4



Rote
Elaborative
Recalled
Not Recalled
Fz
Cz
Pz
-20µV
0
+20µV
-20µV
0
+20µV
-20µV
0
+20µV
0
400
900
1400
1900
0
400
900
1400
1900
Time (msec)

FIG.5

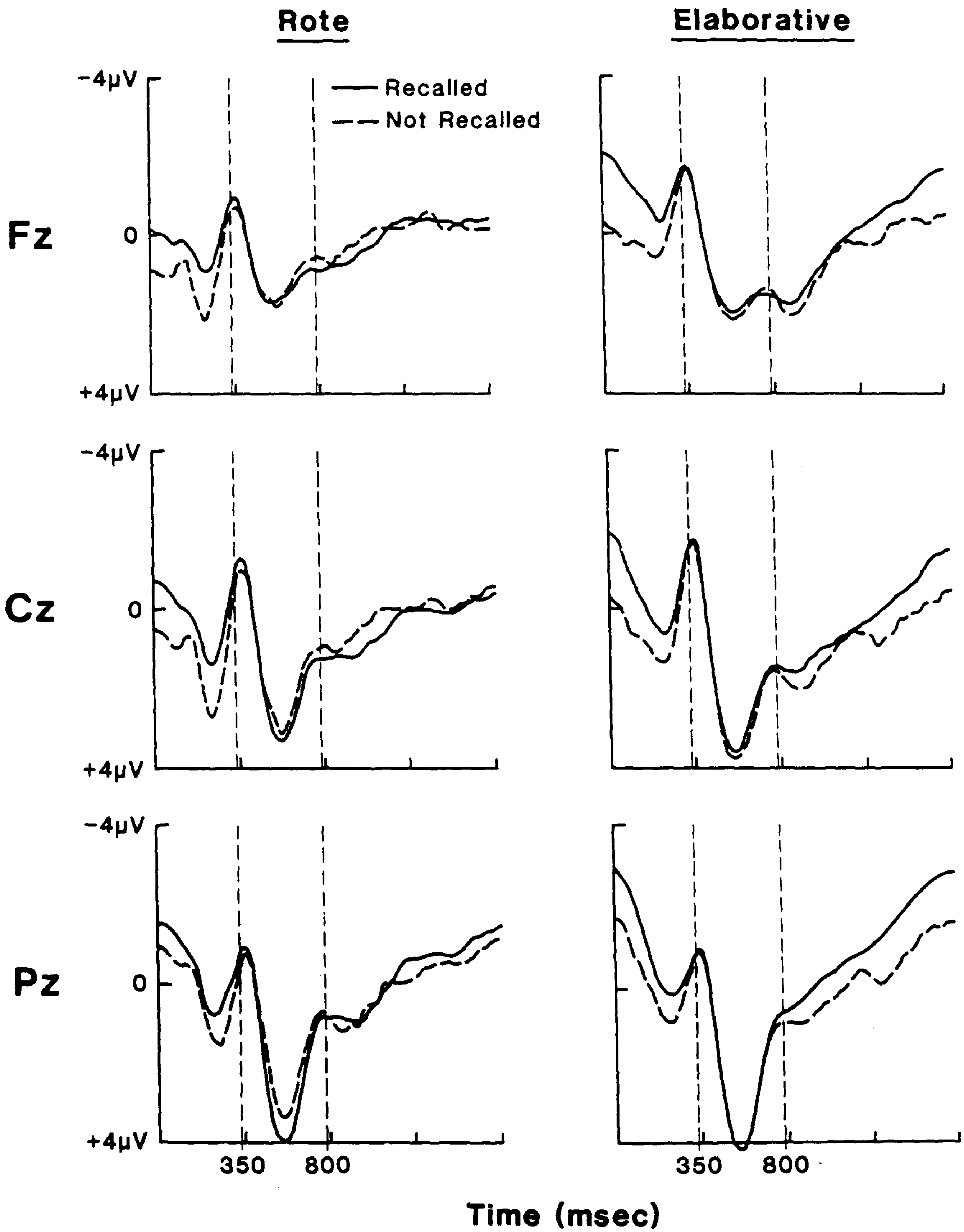

Rote
Elaborative
Recalled
Not Recalled
Fz
Cz
Pz
-4μV
0
+4μV
350
800
Time (msec)

FIG. 6

TALKING OFF THE TOP OF YOUR HEAD: A MENTAL PROSTHESIS UTILIZING EVENT-RELATED BRAIN POTENTIALS (1)

L. A. FARWELL and E. DONCHIN
Dept. of Psychology and Cognitive Psychophysiology Laboratory
University of Illinois at Urbana-Champaign
Champaign, Illinois 61820

Running Title: MENTAL PROSTHESIS

Key words: P300, ERP, event-related potential, prosthesis, self-help device, EEG

Send proofs to:

Dr. Emanuel Donchin, Professor and Head

Dept. of Psychology and Cognitive Psychophysiology Laboratory

University of Illinois at Urbana-Champaign

Champaign, Illinois 61820

SUMMARY

SUMMARY

This paper describes the development and testing of a "mental prosthesis" that provides a means by which a person who is incapable of producing speech, or using any other motor system (i.e., is "locked-in"), can communicate through a computer utilizing the P300 component of the event-related brain potential (ERP). The 26 letters of the alphabet, together with several other symbols and commands, are displayed on a computer screen which serves as the keyboard or prosthetic device. In the developmental studies reported in this paper, undergraduate volunteers took part in two experimental sessions. In the first session (the pilot study/training session) subjects attempted to spell a word and convey it to a voice synthesizer for production. In the second session (the analysis of the operating characteristics of the system) subjects were required simply to attend to individual letters of a word for a specified number of trials while data were recorded for off-line analysis. The purpose of the first session was to determine the feasibility of the mental prosthesis. As subjects viewed the computer screen and attempted to spell the word, elements of the matrix were flashed in a random sequence and the particular flash that invoked a P300 was detected. Convergence on the attended element was achieved following a series of presentations in which those elements of the matrix that did not elicit P300s of sufficient amplitude to define the critical element were eliminated. This procedure was followed until the word was spelled, and the command to route the word to a voice synthesizer was identified. Each subject completed this task successfully. Our aim in the second session was to assess the efficiency of different methods for

identifying the P300 to the attended letter. That is, in the second test session we assessed the operating characteristics of the system as a means of information transfer by analyzing the relative efficacy of various algorithms for detecting P300. Our results in both test sessions suggest that the Mental Prosthesis may prove to be a useful communication device for patients who have no other means of expression.

INTRODUCTION

Several hundred studies have demonstrated that when subjects are presented with a series of stimulus events that can be classified into 2 categories, one of which is rare, task-relevant, or both, these rare/relevant events elicit an event-related brain potential (ERP) with an enhanced positive-going component with a latency of about 300 msec, labelled the P300. (for reviews see Pritchard, 1981; Hillyard and Kutas, 1983; and Donchin et al., 1986.) The amplitude of this component is largest to the task-relevant or target event and varies inversely with the probability of occurrence of that event; that is, the less frequent the target, the larger the P300 amplitude. The elicitation of P300 depends critically, therefore, on the subject's ability to discriminate the events and assign them to the appropriate categories. This experimental arrangement has come to be called the "oddball" paradigm, as the appearance of P300 can be used as an indication that the subject has recognized the occurrence of an event belonging to the rare category.

Note that it is not necessary for the subject to report the occurrence of the a target event through any overt means (e.g., button press). Indeed, in the classic oddball task, the subject is often required only to maintain a running mental count of number of occurrences of the target. Thus, the appearance of the P300 signals the subject's recognition of the rare, task-relevant events without recourse to verbal or motor means of communication. This attribute of the P300 suggests that it may be possible to develop a "Mental Prosthesis" utilizing the oddball paradigm that would permit persons who, as a result of neurological injury or disease, are incapable of any overt means of communication. In this paper we describe such a Mental Prosthesis and provide a detailed evaluation of the utility of the P300 as a communication channel.

Prostheses are generally used to execute the function of a damaged or dysfunctional motor system by employing a substitute motor system. In recent years, refinements in prosthetic devices have been made to aid in communication by employing various motor systems. For example, quadriplegic patients with good control of the neck muscles can activate buttons with a rod attached to the forehead. Increasingly sophisticated prostheses have been developed to allow individuals with more severe disabilities to communicate through the extension of remaining motor functions. Recent developments include a typewriter that can be operated by means of a light beam directed by head movements, (Soede et al. 1974); a typewriter system controlled by a dental palate key operated with Morse code (Saarnio, 1974); a typewriter operated by EMG signals (Torok, 1974); a switch that can be operated by slight hand or toe movements (Hammond, 1974); a variety of switches that can be interfaced with communication boards, such as a switch that can be operated by interrupting a beam of light with any part of the body that can be moved (e.g., the tongue) (Hardiman et al., 1979); several efficient systems based on connecting a joystick with a keyboard (Shwedyk and Gordon, 1977; Vasa and Lywood, 1975); and a portable system capable of producing a liquid crystal display of one of a number of pre-programmed messages that can be attached to an electric wheelchair and controlled by the motion of a joystick (Jardine et al., 1983).

It is not unusual for the the oculomotor system to remain functional when other voluntary motor systems are damaged quite severely. Numerous devices have been developed to take advantage of this by detecting eye position as a means of communication. A variety of means have been utilized to detect eye posttion, including reflection via mirrors and refraction via prisms (Kate et al. 1984), focusing an image of the eye and eyelids on a set of photocells (Wardell, 1977), measuring corneal reflection of infrared light (Rinard and Rugg, 1976; Rubin and Stark, 1984), and visual evoked potentials (Sutter, 1983).

All of the systems described above substitute one motor system for another. In some patients, however, no functional voluntary motor systems remain sufficiently intact to permit their utilization for communication, even though the patients retain sensory and cognitive abilities. For such individuals, the only option available for communication would be a "mental prosthesis" that utilizes non-motor manifestations of mental activity for communication. The system we describe here allows the subject to push a metaphorical switch by focusing attention on one of a series of stimulus events. The discrimination between the event on which the subject is focusing and the other events in the series carries the information that the subject is communicating. By detecting which of the events in the series generates a P300, the appropriate computer-implemented algorithm can identify the message the subject is trying to communicate and send it for him or her (Farwell et al., 1986; Donchin, 1987).

The system works as follows: a 6-by-6 matrix containing the letters of the alphabet and a few one-word commands (see Figure 1) is displayed on a computer-controlled CRT screen. The "stimulus events" that occur in the test consist of intensifications of either a row or a column of the matrix. The subject is attempting, at any instant, to communicate the contents of one cell in the matrix. As the subject focuses attention on that cell, the column and the row containing the cell become "relevant" events. There are twelve possible events (6 rows and 6 columns) only 2 of which are relevant. These events are therefore both task-relevant and rare. Thus, any flash that contains the cell on which the subject is focusing should elicit a P300. The amplitude of the P300 following each flash is assessed, and the attended cell is identified as the cell at the intersection of the row and column that elicit the largest P300s.

We report here a study in which 4 healthy volunteers used the system to communicate a 5-letter word to a computer. The primary purpose in this study is to determine the number of trials and the rate of event presentation that are required to achieve a specified level of accuracy in communication. In a clinical situation, of course, each choice distinguished by the system could be used to communicate an entire word, phrase, or sentence, rather than simply an individual letter.

METHODS

Subjects

Four healthy subjects, 3 females and one male, whose ages ranged from 20 to 36 years, participated in the study.

Data Acquisition and Analysis

The electroencephalogram (EEG) was recorded from Ag-AgCl Beckman Biopotential electrodes placed at the Pz (parietal) site (10-20 International system), referred to linked mastoids. This electrode site was chosen because it is where the largest amplitude P300 is recorded in young adults (Pritchard, 1981; Kutas and Hillyard, 1983; Donchin et al., 1986). Electro-oculogram (EOG) was recorded from sub- and supraorbital electrodes (above and below the right eye). The subjects were grounded at the forehead. Electrode impedance did not exceed 5 kilohm. Brain electrical activity was amplified by Grass model 12 amplifiers with low- and high-pass filters set at half-amplitude frequencies of 35 and 0.01 Hz, respectively. These signals were digitized at a rate of 50 samples per second. Data were analyzed in real time in the pilot/training session and off-line in the assessment session, both of which are described in detail below.

Pilot/Training Session

The purpose of this session was twofold: 1) to determine if it is possible to utilize the P300 as a communication channel that would have sufficient speed and accuracy to be of practical use, by employing a real-time data analysis procedure; and 2) to familiarize the subjects with the apparatus and procedures of the Mental Prosthesis and give them some practice in utilizing the system.

Insert Figure 1 About Here

Subjects were presented with a 6-by-6 matrix whose cells contained the letters of the alphabet as well as several one-word commands for controlling the system (Figure 1). The matrices were displayed on a computer-controlled CRT. In each "trial," each of the the 6 rows of the matrix, or each of the 6 columns, were intensified for a period of 100 msec., with an interval of 500 msec from the beginning of the intensification of each row/column and of the subsequent row/column to be intensified. The rows were selected for intensification in a random order, and then the columns were intensified in a similar manner.

Subjects were instructed to attend to a given letter and to keep a running mental count of the the number of times it flashed. They completed 5 blocks of 120 trials each under these conditions. The fifth block of trials for each subject in the pilot/training session provided an average waveform for the P300 to a rare, task-relevant event that was used as a template in subsequent the signal-detection block described below.

In the sixth block of trials, we used a real-time signal-detection algorithm to allow subjects to communicate using the Mental Prosthesis. Their task was to have the computer generate the word "BRAIN." We employed a covariance signal-detection algorithm that is described in detail below, as is the stimulus presentation procedure we used.

Trials with muscle or EOG artifact were eliminated. Values for EOG artifact rejection were determined on the basis of peak-to-peak EOG response amplitude, and of muscle artifact rejection on the basis of mean absolute deviation of the Pz channel.

Subjects selected each of the letters in the word "BRAIN" in turn, and silently counted the flashes of the row or column containing the letter until the system displayed the letter it had selected in a specified position on the screen (see Figure 1). After the letters spelling the word "BRAIN" had been displayed, the subject selected the "TALK" command, and the word was sounded by means of a Votrax speech synthesizer. In a few cases, an incorrect letter selection was made, and the subject used the BKSP (backspace) command to correct the error.

The system recognized the selections on the basis of a comparison of the covariances of the ERPs following each flash with the template described above. (The covariance signal-detection algorithm is described in more detail below.) All of the subjects completed the task successfully. That is, they were able to use the Mental Prosthesis to spell the owrd and to transmit it to the voice synthesizer for production.

In the pilot/training session, our concern was to determine whether the system could be employed by the subjects and to give them some training on the system. Consequently, ERP data were analyzed in real time, but were not recorded on any storage medium. Information on the time required for each selection was not retained.

Analysis of the Operating Characteristics of the System

The pilot/training session demonstrated that the P300 elicited in these circumstances can be used as a switch subjects can employ to communicate a choice of one out of 36 items. The process depends, of course, on the presentation of many stimuli and on the assessment of the P300 following each flash. The effectiveness of this procedure depends on the degree to which the message can be communicated with a small number of trials using an efficient, cost-effective, on-line detector of the P300. We recalled the same 4 subjects and ran 10 additional blocks to assess the the relationship between the accuracy with which the cell selected by the subject was identified and the number of trials used in the detection. Four different methods for detecting the P300 were used. The data were recorded on disk for off-line analysis.

In an attempt to improve the efficiency of the system, we compared 2 different intervals between row/column flashes. Half of the blocks were run with a 125 msec delay between the onset of the intensification of a given row or column and the onset of the intensification of the next row or column to be flashed (inter-stimulus interval or ISI), and half with a 500 msec ISI. As before, each row was intensified for 100 msec, and then the each of the columns was intensified similarly. Figure 2 illustrates the time course of events in the blocks using the ISI of 125 msec.

Insert Figure 2 About Here

The relevant data consisted of the EEG digitized for 600 msec after the onset of each flash. The EEG was digitized continuously from 20 msec prior to the first flash in each trial to 600 msec after the sixth flash. The subsequent trial began approximately 620 msec after the sixth flash. The inter-trial intervals (ITI) measured from the beginning of one trial to the beginning of the next trial, then, were 1245 and 3120 msec for the 125 and 500 msec ISIs respectively. Note the distinction between inter-stimulus interval (ISI)--the time from the onset of the flash of one row or column to the onset of the flash of the next row or column--and inter-trial interval (ITI)--the time from the onset of one trial (6 row or column flashes) to the onset of the next trial.

Each block consisted of 30 trials. Five blocks at 125 msec ISI were followed by 5 blocks at 500 msec ISI for 2 of the subjects, and the order was reversed for the other 2 subjects.

Subjects were instructed to keep a running mental count of the flashes of the letter "B" until the "CHOOSE ONE LETTER OR COMMAND" instruction (see Figure 1) was turned off for 500 msec and then turned back on. They were then to count the letter "R" until the same signal appeared, then the letter "A", then "I", and then "N". Trials with muscle or EOG artifact were eliminated in the same manner as in the pilot/training session. After each 30 uncontaminated trials were accumulated, the "CHOOSE..." instruction turned off for 500 msec, and then a new block began with the next letter to be attended. After 5 blocks at one ISI, the subjects received a short break, and then the next series was run.

Thus, in effect, the subject spelled the word "BRAIN" in each series of 5 blocks, with approximately 30 trials for each letter. (When a trial was rejected for artifact, it was not recorded for inclusion in the signal-detection computations, and an additional trial was presented. Therefore, the total number of trials in each block viewed by the subject was sometimes a few more than 30, but the number of trials recorded was 30 in every case.) Note that each trial contained 6 distinct events, namely the flashes of each of the rows or columns, only one of which was task-relevant.

Data Analysis

In analyzing the data, we sought to determine how many trials were required to detect the letter on which the subject was focusing at different levels of accuracy, for each of 4 different detection methods.

For analytic purposes, each trial was divided into 6 data windows or subtrials, each consisting of the data for 600 msec after onset of the flash of a row or column (see Figure 2). Thus, since the ISI (i.e., the time between flashes) was less than 600 msec, these subtrials contained overlapping data. For each subtrial we computed a score that measured the magnitude of the P300 in the epoch following the presentation of the row or the column.

Four different algorithms were used to compute the scores: (a) stepwise linear discriminant analysis (SWDA), (b) peak picking, (c) area, and (d) covariance. We will briefly describe each of these algorithms. The interested reader can find detailed discussions of these procedures in Donchin and Heffley (1975) and Coles et al. (1986).

A. Stepwise discriminant analysis. SWDA is a classification procedure. In the present case, a score was computed that reflects the "distance" between each epoch and the mean of a group of trials known to include a P300, as well as the distance from the mean of a group that does not include a P300. This measurement was performed by applying a discriminant function to the data from the epoch. That function was developed on the basis of a "training set" of trials whose group membership was known. The ERPs we recorded while the subject was focusing on the first 2 letters ("B" and "R") served as the training set for our analysis. The remaining ERPs provided the "analysis set." We used the training set data to compute discriminant weights that distinguished between the attended subtrials (600 msec following the flash of a row or column containing the attended cell) and the unattended subtrials (600 msec following the flash of a row or column not containing the attended cell). These weights were applied to individual subtrials in the analysis set and summed across trials in order to identify the attended cell of the matrix.

B. Peak picking. The amplitude of P300 was defined as the difference between the lowest negative point prior to the P300 window (defined as the time range within which the average attended waveform in the training set for each subject was positive) and the highest positive point in the P300 window. The window for the P300 ranged typically between 220 and 500 msec.

C. Area. The "area" of P300 was calculated as the sum of the data points in the P300 window (as defined above).

D. Covariance. A P300 template was computed as the average of the attended subtrials in the training set for each subject. P300 scores in the analysis set were derived by computing the covariance of each subtrial with this template. The covariance was computed using all of the points in the 600-msec epoch.

The values attained from the above analyses were then used to determine the letter upon which the subject was focusing attention. Row and column scores given by the respective algorithms were summed to compute a unique score for each cell in each pair of trials (one trial in which rows were flashed and one trial in which columns were flashed). For example, the score for "B," which is located in the first column and the second row (see

Figure 1), was the sum of the score for the first column and the score for the second row. (By "first" and "second" here we refer to the spatial position in the matrix, and not the temporal position in the sequence of row or column flashes. Since flashes were in random order, the "first" column would be flashed first only approximately one-sixth of the time.)

The scores computed for each letter were summed across trials to determine which cell was identified as the cell selected by the subject. Each test could yield one correct response or one of 35 possible errors. The test was considered a "hit" if the algorithm yielded the largest total score, summed across trials, for the letter on which the subject was focusing. For example, if the subject was attending to the letter "B" and 6 trials were being considered in the analysis, a correct response would be achieved if the total of the 6 "B" scores--the scores for the rows and columns containing "B"--was greater than the total of the 6 scores for any other cell in the matrix.

RESULTS

The principal aim in this phase of the study was to derermine the speed with which the letter on which the subject is focusing can be determined, given the detection technique employed for analyzing the trials. A rather distinct P300 is elicited by the correct letter, as can be seen in Figure 3. This figure presents ERP responses to intensifications of attended or correct letters and of unattended letters, averaged across all trials for each subject.

Insert Figure 3 About Here

If the technique required the presentation of 30 trials for correct detection, however, it would be quite limited--the rate of information transmission would be slower than one character per one and one-half minutes at an ISI of 500 msec. We hypothesized that the transmission rate could be speeded, however, by shortening the ISI and overlapping the data-collection epochs. We knew from previous research that the rate could be increased by taking advantage of the fact that the P300 can be detected using a substantially smaller number of trials (Squires and Donchin, 1976). We examined, therefore, the accuracy of detection of the attended letter as a function of the number of trials at each ISI for each of the 4 detection algorithms. Detection accuracy was estimated by means of an iterative sampling technique akin to bootstrapping (Efron, 1979). Bootstrapping provides an estimate of a parameter in the absence of adequate data on its sampling distribution by obtaining many random sub-samples from the available data and computing the parameter afresh for each of these sub-samples. The distribution of these values approximates the actual distribution.

We randomly chose 1000 sets of 2 trials, 1000 sets of 4 trials, and so on up to 1000 sets of 40 trials from the analysis set. (Recall that the analysis set consisted of 30 trials for each of 3 letters, a total of 90 trials.) The sampling was with replacement. We applied the 4 signal-detection algorithms, computed scores for each of the 36 stimuli, and determined how many times out of 1000 the stimulus that the subject was attending had the highest score with each algorithm at each number of trials considered in the analysis. This provided an estimate of the percent of correct identifications of the chosen stimulus out of the 36 presented, as a function of the number of trials considered in the analysis. By multiplying by the inter-trial interval, we obtained an estimate of the accuracy of each algorithm as a function of time.

Insert Figure 4 About Here

Figure 4 illustrates the iterative-sampling analysis procedure. This figure presents the results of only one iteration, in each of 2 specific cases. For illustrative purposes, we have included one set of data where the discrimination between the attended and unattended stimuli is quite clear, and another set where the discrimination is much less clear. In these figures we plot the scores assigned to each of the 36 letters by the application of one detection technique: a) SWDA scores for subject 3 in the 500 msec condition, and b) covariance scores for subject 2 in the 500 msec ISI condition. The scores we plotted for each of the sample sizes assigned to the correct character can be distinguished from scores assigned to characters sharing either a row or a column with the correct character as well as from entirely incorrect characters.

Thus, the 36 scores at each sample size fall into 3 groups: 1) one "attended" letter (solid line); 2) 10 unattended letters that share a row or a column with the attended letter--and therefore are flashed at the same time as the attended letter either when flashed with their row or with their column (dotted and chained lines); and 3) 25 other unattended letters--which never are flashed at the same time as the attended letter (dashed lines).

Note that the highest score was generally obtained for the attended cell; and, as might be predicted, the scores for unattended cells that were in the same column or the same row with the attended cell were higher, in general, than the scores for the other unattended cells. The decision algorithm is, of course, "correct" whenever the attended letter is assigned the highest score. It can be seen that this is generally the case. It can also be seen that there is considerable individual variance. It is evident in Figure 4 that as the number of trials increases the expected reduction in signal-to-noise ratio yields greater accuracy: the larger the number of trials, the more often the correct letter is assigned the highest score, and the larger, in general, the difference between the score assigned to the correct letter and the scores assigned to the incorrect letters.

For each of the 4 subjects we obtained 1000 data sets similar to those illustrated in Figure 4, and analyzed each data set with each of 4 detection algorithms. It is the total data set (including all 1000 of these smaller data sets) that was analyzed as we assessed the accuracy of this communication channel as a function of its speed.

As we noted, the analysis procedure was repeated 1000 times for each combination of sample size (2-40 trials) and algorithm (SWDA, peak picking, area, and covariance), at each ISI (125 and 500 msec). In each of the 1000 iterations, for each sample size, we picked a new random sample of the trials for inclusion in the analysis. For a particular iteration, at any given sample size, each analysis procedure provided either a correct or an incorrect determination. We tallied the number of correct determinations for each sample size, for each subject, at each ISI, employing each analysis algorithm.

Insert Figure 5 About Here

In Figure 5 we plot the proportion of correct decisions out of of the 1000 iterations of the procedure at each sample size. To allow comparison of the data obtained with the 2 ISIs and to facilitate the evaluation of the speed of the system, the percent of correct identifications of the attended letter is plotted against the time required to present a given number of trials, rather than against the number of trials. The interval therefore is equal to the number of trials times the inter-trial interval. As can be seen in Figure 5, there are considerable individual differences in the subjects' ability to use the system, as well as in the relative effectiveness of the different detection algorithms. Moreover, different algorithms were more effective for different subjects. All of the subjects, however, were able to achieve a high level of accuracy in communicating their choices to the system at a speed of some seconds per choice.

Insert Table I about here

Table I presents speed and accuracy figures for the fastest algorithm for each subject at each ISI. When the subjects' optimal ISI and signal-detection algorithm were used, the mean time required to achieve 80% accuracy of determination of the one stimulus out of 36 that the subject was attending was 20.9 seconds (Table Ia). For 95% accuracy, the mean time required was 26.0 seconds (Table Ib). A choice of one out of 36 contains 5.2 bits of information, so the speed at 95% accuracy was 0.20 bits per second, or 12.0 bits per minute. By using the "BKSP" (backspace) command (see Figure 1) with the same speed and accuracy, a subject could correct errors and achieve over 99.9% accuracy with a speed of 0.18 bits per second, or 10.8 bits per minute.

SWDA and peak picking proved to be the most efficient algorithms. At 125 msec ISI, SWDA was the fastest algorithm to reach both 80% and 95% accuracy in 3 out of 4 cases. At 500 msec ISI, peak picking was fastest to reach both 80% and 95% accuracy in 3 out of 4 cases. (A possible explanation of this difference in algorithm effectiveness as a function of ISI is discussed below.) When considering the 4 subjects, 2 ISIs, and 2 accuracy criteria (80% and 95%), SWDA yielded the fastest times to reach the accuracy criterion in 8 cases out of 16, and peak picking in 6 cases. Area and covariance were each fastest in one case.

Insert Table II about here

Table II shows the times taken by each of the 4 algorithms to reach 80%

and 95% accuracy, for each subject at each ISI. As shown in Table II and Figure 5, different signal-detection algorithms were more effective for different subjects. This is a result of differences in the characteristic ERPs for different subjects and differences in the information utilized by the algorithms.

Discussion

This study addressed two distinct questions. We sought to determine if it is indeed the case that the P300 can be employed as a switch by means of which the subject can toggle a choice. This question is clearly answered in the affirmative. Indeed, the specific arrangement we used to present choices to the subject amplifies the power of the the P300 to act as a binary switch, as the series of choices allows for the reliable identification of one choice among 36 distinct objects. In principle, this method can be used in a manner that would allow for a choice among more items, as the number of rows and columns can be increased. However, such an increase would entail a cost in that the total number of flashes required for each choice would be increased. The optimal size of the matrix remains a matter for further investigation.

The answer to this first question was not entirely surprising. There is by now an extensive literature that establishes the reliability with which the P300 is elicited by rare, task-relevant events within the framework of the oddball paradigm. It is quite clear that almost any arrangement that would impose a categorization on a series of events, however abstract the categorization, can be used to elicit sizeable P300s provided the two categories are presented in a Bernoulli sequence, that the stimuli play an important role in the subject's information processing, and that one of the categories occurs with a somewhat lower frequency (See

Fabiani et al., in press, for a discussion of the varieties of the oddball paradigm). Note that even though improbable events do tend to elicit a P300 with ease, rarity is neither a necessary nor a sufficient condition for the elicitation of the P300. In any event, our data do confirm that the P300 can be used as a communication channel by taking advantage how it responds to task-relevant events in the oddball paradigm, as used in the arrangement described above.

There is, however, a second question whose answer was by no means self-evident. The utility of communication channel based on the P300 depends, as do all communication channels, on the signal-to-noise ratio. It is evident that the P300 on which this channel is based is buried in the "polyneural roar of the EEG," to use Ross Adey's felicitous phrase. The detection and measurement of the P300, as is true for other ERP components, requires signal averaging. Thus, it was conceivable that while the P300 can, in principle, serve as a switch, its reliability under the signal-to-noise conditions which it presents would have been quite impractical for actual use. Our main purpose in this study, then, was to examine the operating characteristics of the communication channel.

The prime task of the channel is to communicate the choice the subject has made among the 36 options. Thus, the performance index for the channel is the accuracy with which this choice is communicated, as a function of the speed with which the channel operates. The speed is controlled by the rate at which the stimuli are presented. The accuracy is controlled by the efficiency of the signal-to-noise reduction achieved by the detection algorithms. It is for this reason that we used as independent variables the inter-stimulus interval within each trial and the various detection procedures.

The conclusions are quite clear. The channel can operate reasonably well at the speed of 12 bits per minute. A character, chosen from among 36 items, can be detected with 95% accuracy within 26 seconds.

The inter-stimulus interval proves to be an important variable. To obtain accurate discrimination of the attended stimulus, a certain signal-to-noise reduction is required. This can be achieved by increasing the interval between stimuli from 125 to 500 msec, allowing for a better definition of the P300. Alternately, the signal-to-noise reduction can be achieved by an increase in the number of trials. Which of these methods is more effective depends on the subject and the signal detection algorithm.

The various tables and figures reviewed above show that different detection methods varied in their effectiveness when applied to the data of the different subjects. The differences in effectiveness are due to an interaction between the nature of the procedures and the specific attributes of the subject's data. It is useful to consider the differences among the detection algorithms.

Comparison of the different algorithms

Covariance computes, essentially, how similar the individual ERPs are to a template consisting of the average waveform for the attended cell in the training set. All time points are included, and each point is weighted according to the mean amplitude of that point in the training set.

SWDA involves much more extensive computations on the training set data than covariance, but it is in general more efficient because it gives greater weight to time points that were more effective in distinguishing between attended and unattended cells in the training set.

The primary weakness of both SWDA and covariance is sensitivity to latency variability. If an ERP component, such as P300, appears in a given trial with much longer or shorter latency than the modal latency in the training set, then the discriminant weights (or, similarly, the weights in the covariance algorithm) will not be applied to the points that best characterize the P300, and accuracy will be lost. Latency jitter during the training set, also, will add noise to the system and result in less effective weights.

Peak picking, on the other hand, is highly insensitive to latency variability. The P300 peak can be located anywhere in a relatively wide time window. All of the information contained in the other points, however, is lost by this procedure. Moreover, at a short ISI, insensitivity to latency variability becomes a weakness instead of a strength. Since the peak picking algorithm locates a maximum value at any point within a considerable range, it is susceptible to falsely attributing a P300 peak generated by a previous or subsequent flash to the stimulus being considered. This fact undoubtedly accounts for a large part of the interaction between algorithm and ISI shown in Tables I and II, where peak was the most accurate algorithm at 500 msec ISI and the one of the least accurate at 125 msec.

The area analysis algorithm, like the covariance algorithm, considers all of the points in a broad range, but it is a purely additive, rather than multiplicative, procedure, and does not use a training set. Therefore it misses some information contained in a consistent, distinctive ERP shape and time course, but also avoids some of the noise introduced into SWDA and covariance by variability in the time course and shape of ERPs. It takes advantage of information contained in a broad, flat ERP that is lost in the peak picking algorithm, but by the same token is influenced by noise at points at a distance from the peak.

Because of these differences, different algorithms are more effective in different cases. For a subject whose P300s have a distinct peak with considerable latency variability, peak picking is likely to be the most efficient algorithm, at least when a relatively long ISI is used. For a subject whose ERPs have any distinctive shape and little latency variability, SWDA is likely to be the most efficient. For a subject whose P300s tend to be broad and flat, without much of a peak and with considerable latency variability, area is likely to be the most efficient. In a clinical application, data such as the data reported here could be collected and analyzed, and the optimal algorithm and timing parameters for the individual could be determined and utilized for real-time signal detection.

Neither of the ISIs was clearly superior. In several cases, shortening the ISI from 500 to 125 msec resulted in a nearly equal trade-off between shorter trials and a greater number of trials required to reach a given accuracy criterion. The signal-to-noise ratio decreased with shorter trials because there was more overlap between the data contained in successive subtrials (Figure 2), and consequently more contamination in the 600 msec analysis epoch after each flash of a row or column with ERP components related to the flashes of other rows or columns.

Conclusions, applications, and potential improvements

The above differences notwithstanding, the general conclusion is sustained by the data. It is quite possible to use the P300 as an effective communication switch, and the communication channel can be organized so that the choices can be communicated using a relatively small number of trials.

We can assume for the rest of this discussion that the characters can be communicated with reliability at the rate of one character every 26 seconds, or 2.3 characters per minute.

This is, of course, rather a low rate for a communication channel. Even a slow typist can type 150 characters per minute. Voice communication is even faster. However, it is equally clear that when no other channel is available because the skeleto-musculature is completely disabled, the ability to communicate even at the rate of a few characters per minute would be most welcome.

The utility of this approach for cases of such severe disability demands further research. We have recently initiated a series of studies to determine the feasibility of utilizing this technuque with patients who are incapable of speech or other motor system output.

In addition to clinical applications in cases of severe, permanent disability, there may be other cases in which a very short term disability may require a dependence on such a prosthesis. Examples of this might be burn victims and individuals in the early stages of recovery from severe trauma.

The value of the P300 channel may be further enhanced if the procedure is used as a method for choosing from a menu of commands rather than as a method for spelling words. The elements in the matrix may well be words such as "Nurse," "Water," "Pain," or "Dinner." Each of these choices may in turn call for another menu. In such a paradigm the rate of communication would be enormously amplified, even though the domain of the communication would be constricted. Furthermore, the communication speed we have assessed in this study examined the channel without any attempt to benefit from a number of obvious procedures for accelerating the communication. As a computing device must be a part of the system, it is relatively trivial to incorporate in the channel the known constraints of the language. With each letter presented the number of actual options is reduced, as combinations of characters appear with quite uneven probabilities in English. The system may be allowed to "guess" so that, for example, having detected a "TH" pair it can be relatively sure the following character would be one of the vowels.

It may also be possible to enhance the speed of the system by incorporating additional components of the ERP. If, for example, we were to present the rows and columns in a regular sequence, one would expect to see a CNV develop as the time for the appearance of the correct column, or row, neared. The relative effectiveness of a random presentation utilizing the P300 solely and a presentation that capitalized on both a CNV and the P300 is a matter for further research.

The procedures we describe in this paper and the data we adduce serve to illustrate the feasibility, and the limitations, of the "biocybernetic" concept. The term "biocybernetics" has been used to describe an attempt sponsored during the 1970's by DARPA to develop a "biocybernetic" channel. That channel was to enhance the communication between people and machines by adding channels of communication that employed psychophysiological means. Several approaches were proposed (see Gomer et al., 1979). There were several attempts to use the ERP as a switch. Vidal and his associates have, for example, used the differences between the responses to different checkerboards which flashed on different parts of the screen to create an EEG-driven joystick that controlled the movements of a displayed "mouse" (Hickman and Vidal, 1976). Donchin and his colleagues developed within the framework of the biocybernetic program, and in subsequent work, the use of the P300 as an index of mental workload (see Donchin et al., 1986 for a review). In the assessment of workload, however, the P300 is used as a metric rather than as a switch.

A caveat may be in order. The biocybernetic concept has often been mistaken, especially in the popular press, as an attempt to use the computer to "read the mind" of a subject. Here, too, it is possible to be deceived by the appearance of a subject "writing to the screen" or "speaking through the computer" with the brain waves. One may be tempted to see this as a direct communication between the computer and the mind that somehow by-passes the control people have over the inner workings of their minds. Such an innovation would be treated with dismay by many, and with glee by some. We emphasize that this paper does not report the development of a means by which one can eavesdrop on the mind (see Donchin, in press).

The procedure we describe above accomplishes no more than to provide the subject with a switch that can be wielded at the subject's discretion. The recording would be of no use whatsoever if the subject chose to ignore our instructions and to focus attention elsewhere in the environment. Furthermore, the probes we attach to the head record signals that can be interpreted solely within the framework of the stimulus arrangement we have provided. Thus, there is no more "mind reading" in the procedures we describe than there is when a person is handed a pencil and asked to record impressions. The contents, and the reliability, of the information obtained will depend to a degree on the sharpness of the pencil; but the subject's willingness to report and the accuracy of these reports will be the factors that ultimately determine the utility of the communication. We report here that the P300 can serve as a pencil, and that the pencil is actually rather sharp. The mind, however, retains control over the use of the pencil.

ACKNOWLEDGMENTS

The authors would like to thank Brian Foote for assistance with computer programming; Michael Anderson for technical assistance; Michelle Roser, Amy Pena, and Susan Farrell for their assistance in implementing this research; and Gregory Miller for his helpful comments. We are also extremely grateful to Ted Bashore, of the Medical College of Pennsylvania, with whom we are now collaborating in this project, for his very detailed review of this manuscript, and his helpful support of this project. Michael Coles, Art Kramer and other colleagues at the CPL have been very supportive as usual. We thank Barbara Mullins for her help with the manuscript. Supported in part by AFOSR Grant #F49620-85-C-0041, monitored by Al Fregly.

REFERENCES

Coles, M.G.H., Gratton, G., Kramer, A., and Miller, G. Principles of signal acquisition and analysis. In M.G.H. Coles, E. Donchin, and S.W. Porges, Psychophysiology: Systems, Processes, and Applications. New York: Guilford Press, 1986.

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

Donchin, E. Can the mind be read in the brain waves? In F. Farley and C.H. Null (Eds.), Science Seminars, Federation of Behavioral, Psychological, and Cognitive Sciences. Washington, D.C., in press.

Donchin, E., Karis, D., Bashore, T.R., Coles, M.G.H., and Gratton, G. Cognitive psychophysiology and human information processing. In M.G.H. Coles, E. Donchin, and S.W. Porges, Psychophysiology: Systems, Processes, and Applications. Guilford Press, New York, 1986: 244-267.

Donchin, E. and Heffley, E. Minicomputers in the signal-averaging laboratory. American Psychologist, 1976, 13: 299-312.

Donchin, E., Kramer, A.F., and Wickens, C. Applications of event-related potentials to problems of engineering psychology. In M.G.H. Coles, E. Donchin, and S. Porges (Eds.), Psychophysiology: Systems, Processes, and Applications. Guilford Press, New York, 1986: 702-718.

Donchin, E., Ritter, W., and McCallum, C. Cognitive psychophysiology: the endogenous components of the ERP. In E. Callaway, P. Tueting, and S. Koslow (Eds.), Brain event-related potentials in man. Academic Press, New York, 1978: 424-437.

Efron, B. Bootstrap methods: another look at the jackknife. The Annals of Statistics, 1979, 7: 1-26.

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

Farwell, L.A., Donchin, E., and Kramer, A.F. Talking heads: a Mental Prosthesis for communicating with event-related brain potentials of the EEG Psychophysiology, 1986: 24: 434.

Gomer, F.E., Beidman, L.R., and Levine, S.H. The application of biocybernetic techniques to enhance pilot performance during tactical missions. McDonnell Douglas Astronautics Company-St. Louis division Technical Report MDC E2046, 1979.

Gratton, G., Coles, M., and Donchin, E. A New Method for Off-Line Removal of Ocular Artifact. Electroenceph. clin. Neurophysiol., 1983: 55, 468-484.

Hammond, P. A Detector for Very Small Movements of Limbs and Digits. In K. Copeland (Ed.), Aids for the Severely Handicapped, 1974: 72-76.

Hardiman, C.H., Holbrook, A., and Hedrick, D.L. Nonverbal Communication Systems for the Severely Handicapped Geriatric Patients. The Gerontologist, 1979: 19, 96-101.

Hickman, R.J. and Vidal, J.J. Biocybernetic control in man-machine interaction. Annual Technical Report 1975-1976, Report # UCLA-ENG-7851 for Advanced Research Projects Agency. University of California, Los Angeles, 1976.

Hillyard, S. A. & Kutas, M. Electrophysiology of cognitive processing. In M. R. Rosenzweig & L. W. Porter (Eds.), Annual Review of Psychology, 1983:

34, 33-61.

Jardine, P., Jerman, S. Seddon, N. and Rees, J.A. A Communication Aid for the Severely Handicapped. Journal of Biomedical Engineering, 1984: 6, 230-236.

Kate, J.H. ten, Duyvis, J. Donker, and Le Poole, J.B. Optical communication aids in rehabilitation. Medical Progress through Technology, 1984: 10, 109-123.

Pritchard, W.S. Psychophysiology of P300. Psychological Bulletin, 1981: 89, 506-540.

Rinard, G. and Rugg, D. An Ocular Control Device for Use by the Severely Handicapped. 1976 Conference on Systems and Devices for the Disabled, Boston, Mass., June 1976, 76-79.

Rubin, R., and Stark, L. A Preliminary Investigation of the Information Carrying Potential of Voluntary Eye Movements. Computers in Biology and Medicine, 1984: 14, 303-313.

Saarnio, I. Typewriter Control by Dental Palate Key. In K. Copeland (Ed.), Aids for the Severely Handicapped, 1974: 59-76.

Squires, K.C. and Donchin, E. Beyond Averaging: The Use of Discriminant Functions to Recognize Event Related Potentials Elicited by Single Auditory Stimuli. Electroenceph. clinical Neurophysiol., 1976: 41, 449-459.

Sutter, Erich E. An oculo-encephalographic communication system. Proceedings of the Sixth Annual Conference on Rehabilitation Engineering, San Diego, 1983: 171-173

Shwedyk, E., and Gordon, J. Communication Aid for Nonvocal Handicapped People. Medical and Biological Engineering and Computing, 1977: 15, 189-194.

Soede, M., Stassen, H.G., Van Lunteren, A., and Luitse, W. A Lightspot-Operated Typewriter (LOT). In K. Copeland (Ed.), Aids for the Severely Handicapped, 1974: 42-52.

Squires, K.C., and Donchin, E. Beyond averaging: The use of discriminant functions to recognize event related potentials elicited by single auditory stimuli. Electroenceph. clin. Neurophysiol., 1976: 41, 449-459.

Torok, Z. A Typewriter Operated by Electromyographic Potentials (GMMI). In K. Copeland (Ed.), Aids for the Severely Handicapped, 1974: 77-82.

Vasa, J.J. and Lywood, D.W. High-speed Communication Aid for Quadriplegics. Medical and Biological Engineering, 1976: July, 445-449.

Wardell, R.W. An Eye Position Controlled Communication Aid for the Severely Disabled. - The Eyewriter. Department of Systems Design, University of Waterloo, Waterloo, Ontario, Canada, 1977.

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

FIGURE CAPTIONS

Figure 1

CRT display used in the Mental Prosthesis. The rows and columns of the matrix were flashed alternately. The letters selected by the subject ("B-R-A-I-N") were displayed at the top of the screen in the pilot study.

Figure 2

Time course of events in the blocks using the ISI of 125 msec. Six columns (or 6 rows) were intensified ("flash") in a random sequence for 100 msec, at 125 msec intervals (ISI). EEG was recorded from 20 msec prior to the first flash until 600 msec after the last flash. Each flash served as the onset of a 600 msec analysis epoch ("subtrial"). A trial comprised 6 flashes--one flash of each row or one flash of each column--plus the associated data collection time, a total of 1245 msec.

Figure 3

Average waveforms for attended vs. unattended cells for each of the 4 subjects. A) 500 msec ISI; B) 125 msec ISI.

Figure 4

Examples illustrating the comparison of scores generated by the analysis algorithms, with different numbers of trials included in the analysis. This figure provides an example of the data that went into the speed/accuracy calculations, and is provided for the sake of illustration. In this figure we plot A) SWDA scores for subject 3 in the 500 msec condition and B) Covariance scores for subject 2 in the 500 msec ISI condition. Each line represents the score obtained for one letter, as sample size was varied systematically from 2 to 40 trials per sample.

Scores fall into 3 groups: 1) The "attended" subtrials (solid line), which begin with a flash of the attended letter (and its row or column); 2) unattended subtrials that begin with the flash of a letter that shares a row or column with the attended letter (dotted and chained lines)--of course, the row or column containing this unattended letter flashes as well; and 3) other unattended subtrials (dashed lines). A correct identification was made whenever the attended stimulus received the highest score.

Note that this figure includes data from only one iteration of the analysis procedure, i.e., it represents only one random sample of each size and therefore includes only one score for each letter at each sample size. Speed/accuracy results (see Figure 5) were obtained by repeating this sampling and analysis procedure 1000 times for each sample size and tallying the proportion of iterations in which the attended cell resulted in the highest score at each sample size.

It can be seen that, in these particular samples, the attended cell did generally achieve the highest score. It can also be seen that there was considerable individual variance.

Figure 5

Graphs of the accuracy of each of the 4 algorithms in identifying the attended stimulus, as a function of the number of trials considered in the analysis. Each of the graphs presents the increase in the accuracy of the algorithm as sample size increases. To facilitate the comparison of the 125 msec and 500 msec ISIs and the analysis of the speed of the system, the number of trials has been transformed to the time required to present the trials. Time is the product of the number of trials considered and the duration of a single trial (ITI). Accuracy is the percent of correct identifications of the attended cell out of 1000 iterations of one of the 4 algorithms.

TABLE I

Times required to obtain a) 80% and b) 95% accuracy, by subject and ISI, using the fastest algorithm in each case. Accuracy is the percent of correct identifications of the attended cell out of 1000 iterations of the signal-detection algorithm. Time is the product of the number of trials considered and the duration of a single trial (ITI). (Times have been interpolated when the least number of trials required to reach 80% or 95% accuracy not only reached but in fact exceeded that level.) Times for the fastest combination of an algorithm and an ISI for each subject are followed by 2 asterisks (**).

TABLE II

Times required to obtain a) 80% and b) 95% accuracy, by subject and ISI for the 4 signal-detection algorithms. Accuracy is the percent of correct determinations of the attended cell out of 1000 iterations of the algorithm. Time is the product of the number of trials considered and the duration of a single trial (ITI). (Times have been interpolated when the least number of trials required to reach 80% or 95% accuracy not only reached but in fact exceeded that level.) Times for the fastest algorithm at each ISI for each subject are followed by an asterisk (*). Times for the fastest combination of an algorithm and an ISI for each subject are followed by 2 asterisks (**). Cases where the algorithm did not result in 80% (a) or 95% (b) accuracy after 80 trials are indicated by "X."

FOOTNOTES

1. The research at the Cognitive Psychology Laboratory described was supported in part by AFOSR contract #F49620-79-C-0233. The support of Al Fregly is appreciated.

CRT Display Used in the Mental Prosthesis

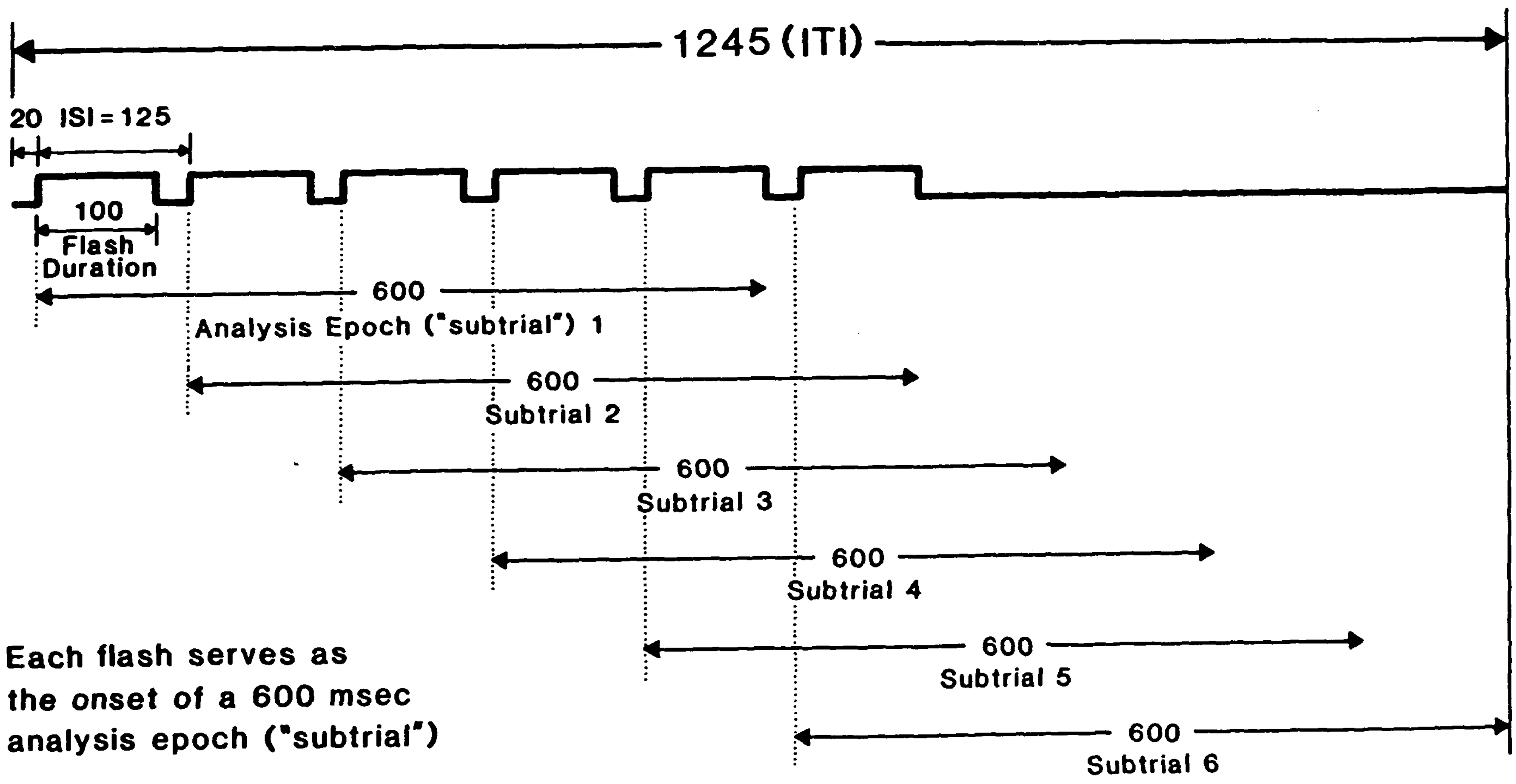
MESSAGE

BRAIN

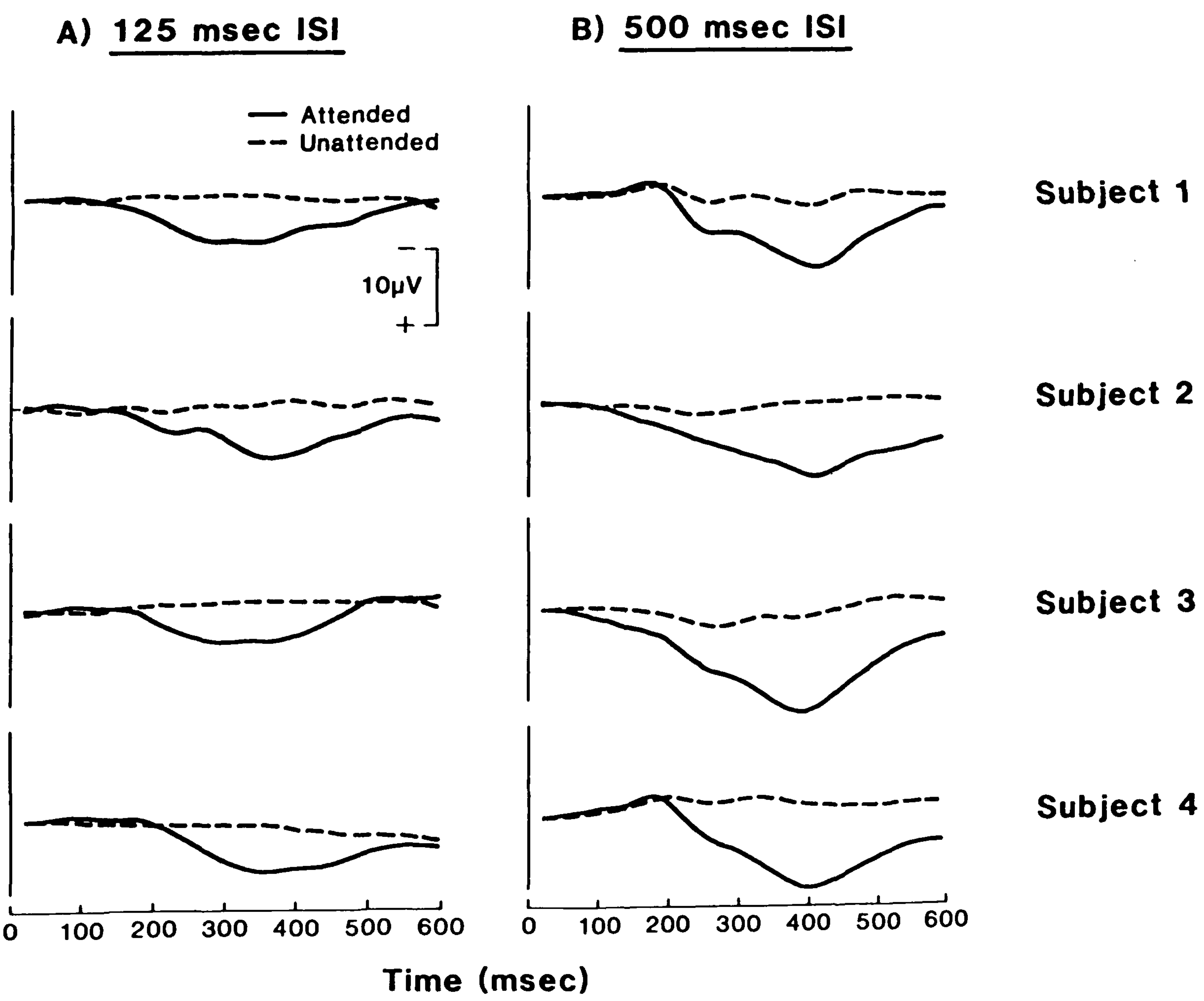
Choose one letter or command

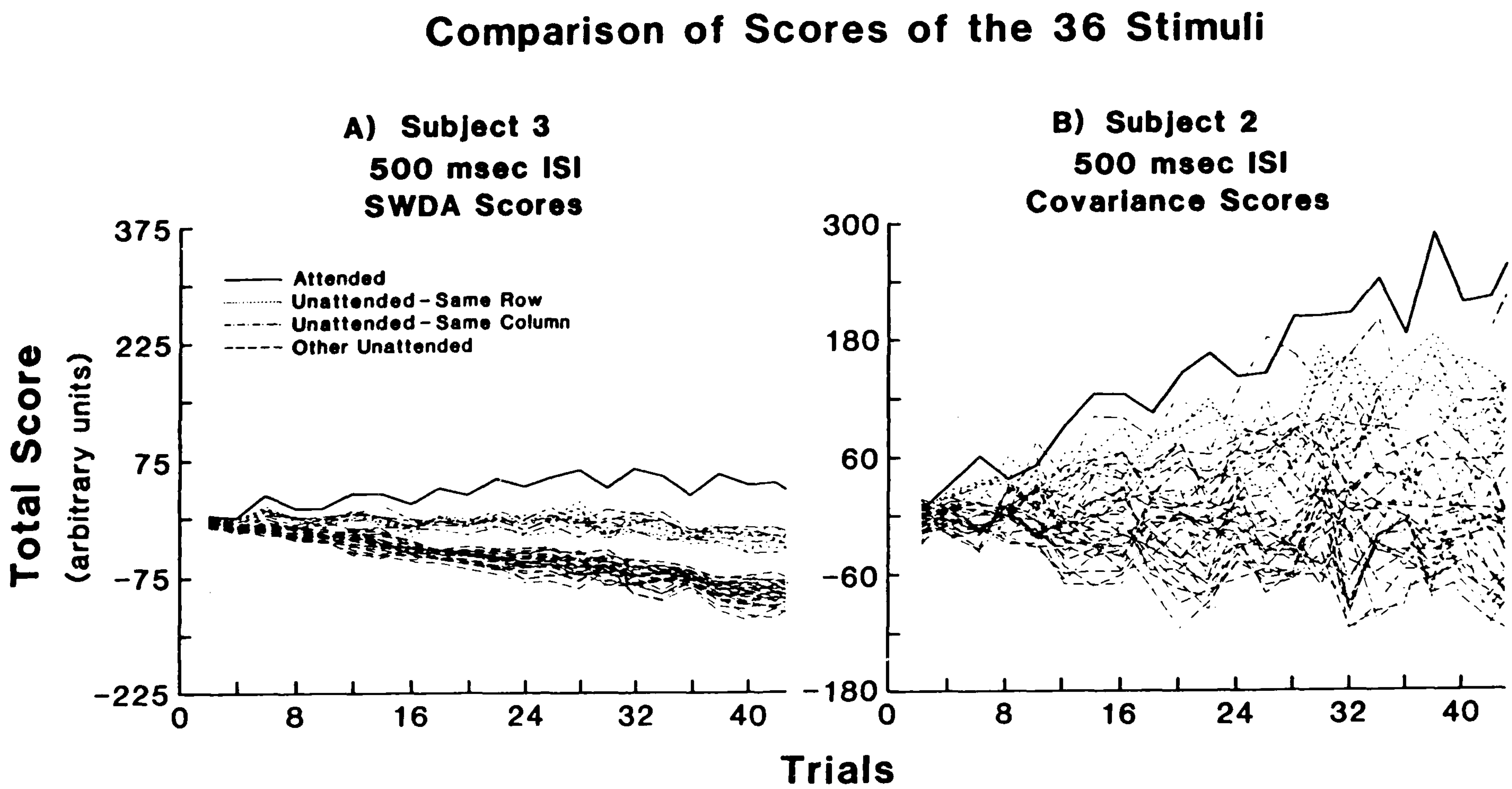
A	G	M	S	Y	*
B	H	N	T	Z	*
C	I	O	U	*	TALK
D	J	P	V	FLN	SPAC
E	K	Q	W	*	BKSP
F	L	R	X	SPL	QUIT

Time Course of Events, 125 msec ISI



ERPs for Attended and Unattended Cells




Comparison of Scores of the 36 Stimuli
A) Subject 3
500 msec ISI
SWDA Scores
B) Subject 2
500 msec ISI
Covariance Scores
Total Score
(arbitrary units)
Attended
Unattended - Same Row
Unattended - Same Column
Other Unattended
375
225
75
-75
-225
300
180
60
-60
-180
0
8
16
24
32
40
Trials

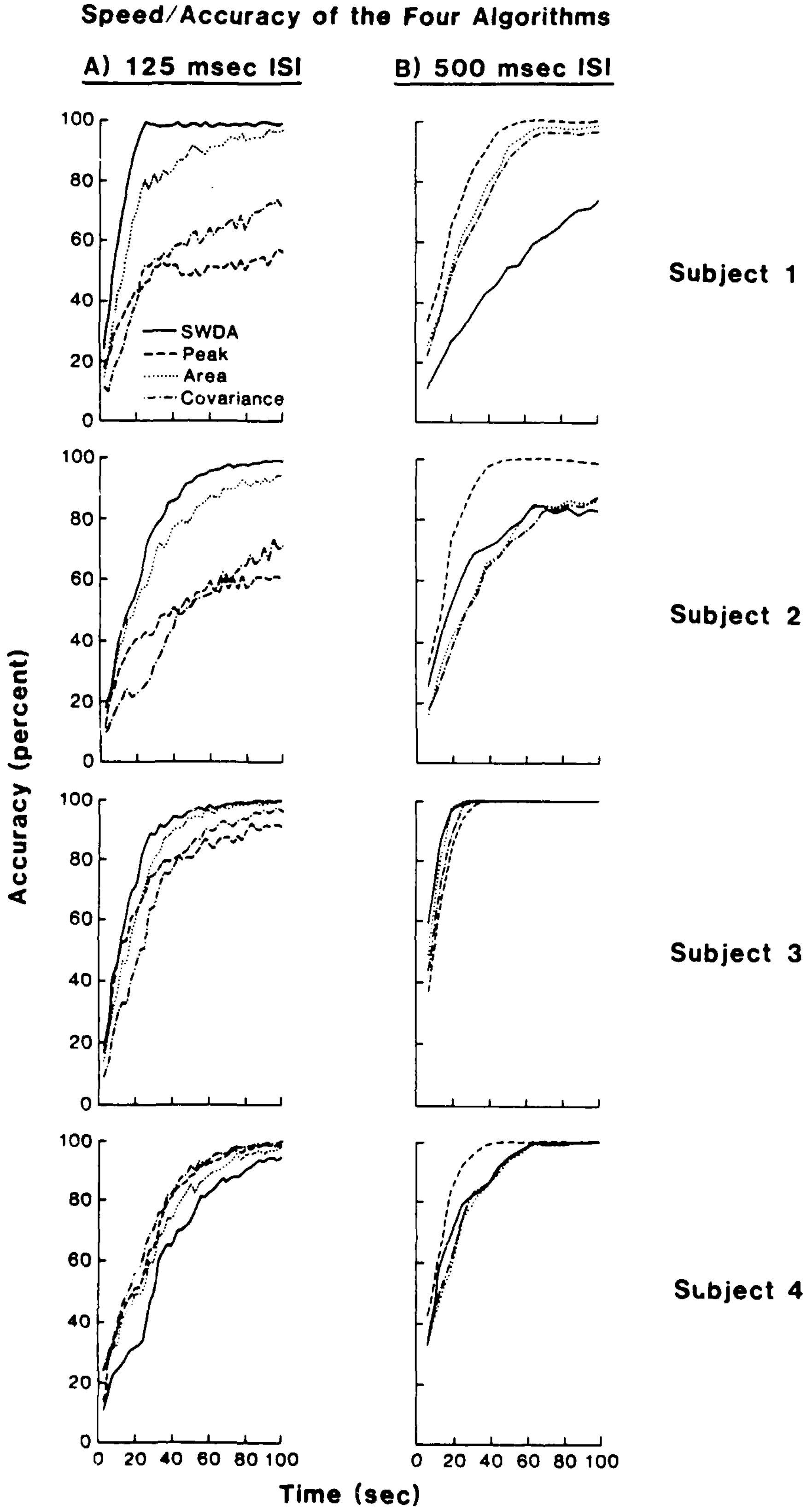

Speed/Accuracy of the Four Algorithms
A) 125 msec ISI
B) 500 msec ISI
Subject 1
Subject 2
Subject 3
Subject 4
SWDA
Peak
Area
Covariance
Accuracy (percent)
Time (sec)
0
20
40
60
80
100

TABLE I
SPEED/ACCURACY: FASTEST ALGORITHMS

	A) 80% Accuracy				B) 95% Accuracy			
ISI (msec)	125		500		125		500	
Subj. #	Time (sec)	Algo.	Time	Algorithm	Time	Algorithm	Time	Algorithm
1	15.7**	SWDA	28.2	Peak	21.6**	SWDA	42.5	Peak
2	33.4	SWDA	23.3**	Peak	57.5	SWDA	35.5**	Peak
3	22.3	SWDA	11.1**	SWDA	46.4	SWDA	17.6**	SWDA
4	36.7	Cov	17.7**	Peak	64.0	Area	29.3**	Peak
Mean	27.0			20.1		47.4		31.2
Mean time to 80% accuracy, fastest ISI and Algorithm for each subject				20.9	Mean time to 95% accuracy, fastest ISI and Algorithm for each subject			26.0

Times required to obtain a) 80% and b) 95% accuracy, by subject and ISI, using the fastest algorithm in each case. Accuracy is the percent of correct identifications of the attended cell out of 1000 iterations of the signal-detection algorithm. Time is the product of the number of trials considered and the duration of a single trial (ITI). (Times have been interpolated when the least number of trials required to reach 80% or 95% accuracy not only reached but in fact exceeded that level.) Times for the fastest combination of an algorithm and an ISI for each subject are followed by two asterisks (**).

TABLE II
SPEED/ACCURACY: 4 ALGORITHMS

A) 80% Accuracy			B) 95% Accuracy	
PEAK PICKING				
Subj. #	Time to 80% Accuracy		Time to 95% Accuracy	
	125 msec ISI	500 msec ISI	125 msec ISI	500 msec ISI
1	X	28.2*	X	42.5*
2	X	23.3**	X	35.5**
3	39.8	17.3	X	26.0
4	38.8	17.7**	70.4	29.3**
SWDA				
Subj. #	Time to 80% Accuracy		Time to 95% Accuracy	
	125 msec ISI	500 msec ISI	125 msec ISI	500 msec ISI
1	15.7**	114.8	21.6**	202.8
2	33.4*	56.9	57.5*	X
3	22.3*	11.1**	46.4	17.6**
4	54.4	26.7	X	49.5
AREA				
Subj.#	Time to 80% Accuracy		Time to 95% Accuracy	
	125 msec ISI	500 msec ISI	125 msec ISI	500 msec ISI
1	29.1	39.9	76.7	59.3
2	49.0	56.6	X	X
3	29.3	12.6	55.8	17.9
4	45.5	44.9	82.2	52.9
COVARIANCE				
Subj. #	Time to 80% Accuracy		Time to 95% Accuracy	
	125 msec ISI	500 msec ISI	125 msec ISI	500 msec ISI
1	X	42.9	X	62.4
2	X	X	X	X
3	41.8	15.5	82.2	22.6
4	36.7	28.6	64.0	52.0

Times required to obtain a) 80% and b) 95% accuracy, by subject and ISI for the four signal-detection algorithms. Accuracy is the percent of correct determinations of the attended cell out of 1000 iterations of the algorithm. Time is the product of the number of trials considered and the duration of a single trial (ITI). (Times have been interpolated when the least number of trials required to reach 80% or 95% accuracy not only reached but in fact exceeded that level.) Times for the fastest algorithm at each ISI for each subject are followed by an asterisk (*). Times for the fastest combination of an algorithm and an ISI for each subject are followed by two asterisks (**). Cases where the algorithm did not result in 80% (a) or 95% (b) accuracy after 80 trials are indicated by "X."

Appendix A

Talking Off the Top of Your Head:
Toward a Mental Prosthesis Utilizing Event-Related Brain Potentials

Lawrence A. Farwell and Emanuel Donchin

TALKING OFF THE TOP OF YOUR HEAD: TOWARD A MENTAL PROSTHESIS UTILIZING EVENT-RELATED BRAIN POTENTIALS (1)

L. A. FARWELL and E. DONCHIN
Dept. of Psychology and Cognitive Psychophysiology Laboratory
University of Illinois at Urbana-Champaign
Champaign, Illinois 61820

In press, Electroencephalography and clinical Neurophysiology

A procedure for using multi-electrode information in the analysis of components of the Event-Related Potentials: Vector Filter

Gabriele Gratton, Michael G. H. Coles, and Emanuel Donchin

Cognitive Psychophysiology Laboratory

University of Illinois at Urbana-Champaign

Acknowledgements

This research was supported in part by a grant from the Air Force Office of Scientific Research, contract #F49620-83-0144, Dr. Al Fregly, Project Director, and by a grant from the National Institute of Mental Health, MH41445. Portions of this study were presented at the 23rd Annual Meeting of the Society for Psychophysiological Research, Asilomar, California, September 26-29, 1983, at the 3rd International Conference on Cognitive Neuroscience, Bristol, England, September 17-20, 1984, and at the 4th International Conference on Cognitive Neuroscience, Dourdan, France, June 15-19, 1987. The comments of Monica Fabiani, Demetrios Karis, Art Kramer, Greg Miller, Dave Strayer, Earl Hunt, Chris Wood and an anonymous reviewer on earlier versions of this paper are greatly appreciated.

Requests for reprints should be sent to: Gabriele Gratton, University of Illinois, Psychology Department, 603 E. Daniel, Champaign (IL) 61820, U.S.A.

Running title: Vector filter

Abstract

This paper presents a procedure, Vector filter, that decomposes the event-related brain potential into components on the basis of scalp distribution. It is assumed that the voltage values observed at several electrode sites at any point in time is given by the linear combination of a set of components and background noise, and that the scalp distribution of each component is invariant and known. Each component's scalp distribution is expressed by a set of weights, one for each electrode. The amplitude of the component, at any point in time, is then derived using a least squares criterion. We review two problems in the use of the procedure: the selection of the set of components and the derivation of the scalp distribution of each of the components. Two application of the procedure are discussed: identification of the component structure of an event-related potential waveform, and filtering of a waveform for a particular component. Unlike other component decomposition procedures, Vector filter can be applied when the latency of a component varies as a function of trial, condition, or subject population.

DESCRIPTORS: Multivariate Analysis of ERPs, Scalp Distribution, Vector filter.

Event-related brain potentials (ERPs) are usually described in terms of underlying components. These components are assumed to vary in amplitude and latency, as a function of experimental manipulations, individual differences, and other variables. Furthermore, several components can be active at the same moment in time, so that the voltage observed at any time point and at any electrode may be the result of the activity of several overlapping components. This of course complicates the analysis of ERPs, since it is often difficult to decide to what extent the voltage observed at a particular point in time at a particular electrode can be attributed to one rather than another component.

Several procedures have been used to disentangle overlapping components. Most of these procedures, such as Principal Component Analysis (for reviews see Coles, Gratton, Kramer, & Miller, 1986; Donchin & Heffley, 1978) or waveform subtraction, assume that the latency of the components is invariant across different conditions. Principal Component Analysis assumes that the timecourse of each component is constant across a set of data. The subtraction procedure assumes that the only difference between two waveforms is the presence of one, target component. Thus, both the timecourse and the amplitude of the other, "background" components, that are common to both waveforms, are assumed to be fixed. However, it is evident that there are circumstances in which the assumption of the invariance of component latency is inappropriate. This is especially the case with respect to late ERP components (such as P300, N400, slow wave, etc.).

The Vector filter procedure,[1] to be described in this report, relies on the distribution of the voltages across the scalp (the "scalp distribution") as the anchor point for the decomposition of an ERP into its components. As we shall see, the Vector filter procedure provides for variability in the latency of a component.

It is widely believed that scalp distribution provides at least one method to distinguish among components (e.g., Donchin, 1978; Donchin, Ritter, & McCallum, 1978; Kutas & Hillyard, 1982; Naatanen & Picton, 1987; Picton, Woods, Stuss, & Campbell, 1978). A corollary of this belief is that components are characterized by particular scalp distributions.2 This view is consistent with the assumption that each ERP component represents the activation of a particular neuronal ensemble (or set of ensembles). While the determination of these ensembles from scalp distribution information is an extremely difficult task (and one which we are not attempting here), "it is somewhat simpler just to differentiate components of the evoked potential on the basis of their scalp distribution. Scalp-recorded events with different voltage distributions must derive from different sources ..." (Picton et al., 1978, p. 525).

Vector filter provides a way of exploiting this view in order to distinguish among the contribution of several components to an obtained ERP. As its name implies, "Vector filter" is a signal enhancing procedure. The procedure filters for a particular scalp distribution, which is believed to characterize the signal of interest (the component). In general, most techniques for analyzing components can be conceptualized as linear filters, that weight temporal, frequency, or spatial information in order to enhance the components of interest. For frequency filters, different weights are assigned to different frequencies. Thus, a low-pass filter of 10 Hz, assigns a high weight to frequencies below 10 Hz, and a low weight to higher frequencies. (These weights are generally selected a priori.) For temporal filters, different weights are assigned to different time points in the ERP waveform. For example, for the P300 component, the weights assigned to voltage values for time points between, say, 300 and 500 ms, are higher than those for other time points. (Note that the weights can be determined in advance or by using statistical techniques, such as Principal Component Analysis.) In the case of spatial filters, different weights are assigned to voltages recorded from different electrode locations. For example, when we select one particular electrode location, we assign a weight of 1 to that electrode, and a weight of 0 to all the other electrodes. Vector filter is a type of spatial filter that allows one to select any set of weights for different electrodes.

Method

The model

The basic model used by Vector filter assumes that the voltage values observed at a particular moment in time at different electrode locations are given by the linear combination of a series of components and noise. It is assumed that, across timepoints and repetitions, the scalp distribution of each component is constant, while its amplitude is variable. The scalp distribution of a component can be described by a set of weights, one for each scalp electrode. These ideas can be formalized by the following expression (1):

(1) Volt(ij) = Sum(k) c(ik)W(jk) + Err(ij)

Where Volt(ij) is the voltage recorded at the electrode (j) at timepoint (i), c(ik) is the amplitude of component (k) at timepoint (i), W(jk) is the weight of component (k) at electrode (j), and Err(ij) is the noise at the electrode (j) at timepoint (i). If we define spatial filters in terms of particular sets of weights, then the filters are identified with the scalp distribution associated with each component.

The set of weights for each component are assumed to be known. (In later sections of this paper we will review procedures for their estimation.) Note that the assumption of linearity implies that the scalp distribution (ratio among values at different electrodes) remains constant as the amplitude of the component changes. Support for this view is provided by McCarthy and Wood (1985). These authors argue that when there is a change in the activity of a source underlying a component, there should be a proportional change in the scalp recorded activity at each electrode.

Derivation of filtered waveforms

The procedure comprises three critical elements: (a) an obtained ERP waveform for each of several electrodes (the data set) for which we want to know the relative contribution (amplitude) of the constituent components at each time point; (b) a set of spatial filters, each defining a component, which take the form of vectors of weights, one per electrode, that specify the extent to which a component is active at the electrode -- these weights thus describe the scalp distribution of each component; and (c) an output waveform (amplitude by time function) for each component that describes its amplitude at each time point in the obtained ERP waveforms. Note that the procedure results in the transformation of the input data set (one ERP waveform for each electrode) into an output set of waveforms (one for each component).

This tranformation is accomplished in the following way. First, for each timepoint, we determine the combination of the components which provides the optimal approximation to the observed data. This is accomplished by estimating a particular "beta coefficient" (labelled c(ik) in equation (1)) for each component. The beta coefficients are taken as estimates of the amplitude of the components at that particular point in time. The procedure is then repeated for each time point in the epoch, so that, for each component, we get amplitude by time functions, depicting their time courses.

The crucial step in the procedure, therefore, is the estimation of the beta coefficients to assign to each component to produce the optimal approximation to the observed data. This is accomplished by applying a least squares procedure (see Table 1) in which we determine the coefficients that should be given to each component to minimize the error in the description of voltage values observed at all the electrodes for that timepoint. Note that the components are used as predictors, and the observed data (values at the different electrodes) as the criterion.

Insert Table 1 about here

Table 1 shows how the observed values at the different electrode locations (1 to N) at a particular time point can be considered as a "vector" of criterion values (Volt (1) to Volt (N) in Table 1), that are optimally approximated by the linear combination of a set of components (A to Z), each of which is characterized by a "vector" of values (e.g., A(1) to A(N)), one for each electrode location. Both vectors can be visually represented as "profiles" of values at different locations. Thus, we can have a profile describing the data observed at a particular time point, and a profile describing each of the components. The "free" parameters in the system of equations presented in Table 1 are the beta coefficients of the components at that particular time point (c(A) to c(Z)). Note that these values are fixed across all the electrode locations. We take the beta coefficients as representing the amplitude of each component for that time point. The estimation of the beta coefficients (amplitudes of the components) is repeated for each time point in the ERP waveform. When these coefficients are plotted over time, the result is a set of amplitude by time functions, depicting the time course of each component. These amplitude by time functions can also be considered as the outputs of a set of spatial filters, with each filter "tuned" to the scalp distribution of a particular component.

Characteristics of the procedure

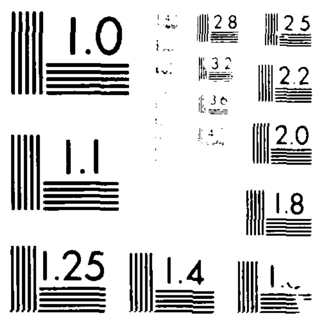
Vector filter is based on a multiple regression (for each time point) in which the set of predictors is given by the components, the criterion is given by the voltage values observed at the different electrode locations, and the regression coefficients are used as estimates of the amplitude of each component at each data point. Thus, several properties of multiple regression apply.

a. Test of the model. It is possible to assess the degree to which the particular set of components that are chosen actually account for the variance present in the data. This is accomplished by computing the multiple correlation coefficient. This statistic can be used to determine the value of the model in accounting for the observed data.

b. Number of components. Two considerations govern the choice of the number of components to be used. First, a large proportion of the variance in scalp distribution should be explained. Second, the amplitude estimates for each component should be reliable. These two criteria leads to opposing strategies. The more components that are used, the more variance will be explained, but the lower will be the reliability of the amplitude estimates.

AD-A191 244 THE EUENT-RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING A..(U) ILLINOIS UNIU CHAMPAIGN COGNITIUE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 8/10

UNCLASSIFIED 29 FEB 88 CPL-88-1 AFOSR-TR-88-0316 F/G 5/8 NL


1.0
2.8
2.5
3.2
2.2
3.6
2.0
1.1
1.8
1.25
1.4

To resolve this dilemma, we note that the reliability of the amplitude estimates will depend on the degree to which a particular set of components accounts for more variance than is accounted for by an equal number of randomly selected components (see McNemar, 1969, and other specialized texts for tests of significance of this difference). Note that the proportion of variance explained by a set of components chosen at random is equal, as an average, to $(n-1)/(N-1)$, where n is the number of components, and N is the number of electrodes (see McNemar, 1969, p. 203).

As an example of the application of these concepts, consider the case in which we want to account for 80% of variance in scalp distribution. What is the maximum number of components that can be included in the set given that we need to obtain reliable coefficients with N electrodes? Based on procedures described in McNemar (1969), this number can be computed by, first, establishing the critical F (for a multiple R of .9), and, second, determining the ratio between the number of components and the number of electrodes to achieve this critical F. It turns out that this ratio is such that the number of components should be smaller than $(N+1)/2$. Thus, as a general rule of thumb, it can be expected that low reliability will obtain whenever the number of components is more than half the number of electrodes.

c. Interrelationship between components. Since the components are defined in terms of scalp distribution, it is clear that two components with the same scalp distribution cannot be entered simultaneously in the procedure. Even when two components have very similar (although not equivalent) scalp distributions, the amplitude estimates for each of the two components will not be reliable. In fact, small fluctuations in the observed scalp distribution may result in large variations in the amplitude estimates of the two components. (This problem is also known as the "problem of multicollinearity" in multiple regression analysis.) Therefore it is not advisable to use components with very similar scalp distributions. The similarity in scalp distribution between two components can be measured by using (2):

(2) R(12) = Sum(j) A(j)*B(j) / Sqrt((Sum(j) A(j)**2)*(Sum(j) B(j)**2)))

where R(12) is a measure of the similarity between the scalp distributions of any pair of components (A and B) for which the weight at electrode (j) are A(j) and B(j), respectively. In no case should two components with an R(12) higher than 0.8 be entered simultaneously (see Lewis-Beck, 1980, p. 60; this and other specialized texts also describe more sophisticated methods of detecting and dealing with multicollinearity).

Although the problem of component similarity may appear to limit the applicability of the procedure, it can be dealt with in a number of ways, each of which increases the discriminability between the scalp distributions of components. One approach is to increase the number of electrodes and to choose electrode locations that maximize the discrimination between components. Another approach is to adopt a Laplacian reference system (Hjorth, 1979). This procedure, which is technically equivalent to computing the current source flow, tends to produce localized components. Localized components will usually have quite distinguishable scalp distributions, and very low R(12)'s.

d. Vector filter and other linear transformations. As we have seen, the Vector filter procedure is based on a linear model. One consequence of this fact is that it can be performed before or after other linear manipulations of the ERP data. For instance, applying Vector filter to single trial data, and then averaging the vector output waveforms, will produce results that are equal to those obtained by averaging the single trial input data and then performing the Vector filter operation on the average waveforms. This property makes Vector filter suited for the analysis of single trials.

Applying Vector filter

In this section we consider two critical issues in the application of Vector filter. First, we discuss the problem of selecting the appropriate set of components, and we examine the consequences of failure to select the appropriate set. Second, we review approaches to the problem of specifying the scalp distribution of each component in the set, and present a practical example.

Selecting the set of components

The Vector filter procedure is based on the assumption that a given set of components, with known scalp distribution, is responsible for the observed ERP. How is the set of components selected?

The answer to this question depends in part on the aims that underlie the application of Vector filter. If the purpose is to isolate the part of the ERP waveform that contains the components of interest, then it is sufficient to merely use these components as the set. In this case, the application of Vector filter corresponds to the application of other kinds of signal-enhancing procedures. For instance, frequency filters are used to isolate activity in the ERP waveform that has a particular frequency range. Note that, in the case of frequency filters, it is assumed that the component of interest has a specifiable frequency range. Similarly, in the case of Vector filter, we assume that the component of interest has a specifiable scalp distribution. Note that we do not need to know the scalp distribution (or frequency characteristics) of all the components present in the ERP waveform.

On the other hand, when the aim of the procedure is to decompose a particular ERP waveform into its component structure, then more sophisticated selection procedures are required. These procedures can be based on a "step-wise" approach, in which the goodness of fit provided by a particular set of components can be estimated, following addition or subtraction of particular components. In this case, a set of theoretically interesting components can be selected first, and the fit of the model based on this set to the ERP can be assessed for each data point, or across a number of data points. Then, depending on the results of the goodness-of-fit test, new components may be added or old components may be deleted. A criterion for halting this step-wise operation may be the determination of the simplest model (i.e., based on the least number of components) that produce a satisfactory fit to the data (with a probability of rejection of less than .05).

These procedures for selecting the set of components are predicated on a priori knowledge of the possible components present in the ERP waveform. In most ERP experiments, the investigator approaches the data with some hypotheses about the components that should be elicited in the particular experimental paradigm. Thus, it is not unreasonable to expect that the set of components can be specified in advance of the application of Vector filter. However, it is always possible that the investigator is mistaken and fails to include components that are really present. What are the consequences of this error?

In the rest of this section we review these consequences by applying

Vector filter to simulated data, for which we know the set of components that are present. The simulated waveforms were obtained by adding components with fixed scalp distributions to each other. The distributions are described by a "profile" of weights (one weight for each electrode - see upper row of Figure 1). The amplitude of each input component varied over time (see second row of Figure 1). To obtain the ERP waveform for each electrode (Fz, Cz, and Pz), the amplitude of each component for each time point is multiplied by the weight of that component at each electrode, and the amplitudes for different components are then summed (see third row of Figure 1).3

Insert Figure 1 About Here

Applying the Vector filter operation to the ERP waveform (third row) yields the waveforms shown in the lower row. These waveforms represent the time course of the two components retrieved by the procedure when (a) all the components present in the data are considered, and (b) the weights for each component are the same as those of the simulated components (shown in the upper panel). Note that, not surprisingly, Vector filter perfectly retrieves the amplitude and time course of each original component.

In the next two simulations we examine cases in which the set of components used for Vector filter is a subset of the input set -- that is, we have failed to include some of the components present in the data. In most cases this will result in some error in the retrieval of the original components. An example is given by the second simulation, illustrated in Figure 2.

Insert Figure 2 About Here

In this simulation, we used the same simulated ERP waveform used in simulation 1. However, only Component A was entered in the Vector filter procedure, while Component B was ignored. The error is illustrated in the lower panel, where the dotted line indicates the "input" Component A, while the solid line depicts the Component A as retrieved by the Vector filter procedure.

Simulation 3 illustrates a case in which ignoring one component does not result in a retrieval error for the other components. In this simulation, we have added a third component (Component C) to the set of input components. This simulation is illustrated in Figure 3.

Insert Figure 3 About Here

Again, the first and second rows depict the scalp distribution and time course of the input components, the third row, the composite ERP waveform, and the bottom row, the outputs of the Vector filter procedure, when only two components (A and B) were included in the set. In this case, failure to include a component (C) that was present in the original ERP waveform, did not lead to any error in the retrieval of components A and B.

In general, the error introduced by ignoring one or more of the components depends on the amplitude of the ignored components and on the degree of similarity between the component that is ignored and those that are entered in the analysis. In simulation 2, the component ignored (B) has a scalp distribution that is quite similar to that of the component entered in the analysis (A), R(12)=0.38. In simulation 3, the component ignored (C) has a scalp distribution that is very dissimilar to that of the components entered in the analysis (A and B): both the R(12)'s for A and C and for B and C were equal to 0.

Taken together, simulations 2 and 3 reveal that inaccurate specification of the set of components will have minimal consequences under some conditions. Specifically, when the omitted components are either very small or have scalp distributions that are very dissimilar from those of the components that are included in the set.

In practical cases, the estimation of the dissimilarity between included and omitted components requires the specification of the scalp distribution of the omitted component. If one can specify the distribution of a component, why not include it in the set? As we noted above, inclusion of a large number of components in the set may lead to unreliability in the estimation of component amplitudes. This will occur when the number of components in the set approaches the number of electrodes. If one cannot specify the distribution of an omitted component, how can one estimate its similarity to the included components? Clearly, estimation of similarity in this case is impossible. However, one can take steps to reduce the likelihood of similarity by (a) choosing electrode locations that maximally specify components of interest, and/or (b) using the Laplacian derivation or similar procedures to localize the scalp distribution of components.

Estimating component distribution

As we have noted, a fundamental requirement for the use of Vector filter is specification of the distribution of the components to be included in the set.

There are several ways in which the scalp distribution of a particular component can be estimated. Although each of these ways, taken alone, may not always be satisfactory, the convergence of several methods can be taken as strong evidence in favor of the hypothesis that a particular scalp distribution can represent the component. The methods to estimate the scalp distribution of a component can be divided into three categories: biophysical methods, statistical methods, and experimental methods.

a. Biophysical methods. These methods rely on the fact that knowledge of the source of a component can be used to predict its scalp distribution. For example, knowledge of the functional anatomy of the motor cortex enabled Brunia (1980) to account for the different distributions of potentials that precedes foot and finger movements. At present, knowledge of the sources of most ERP components is limited. However, as such knowledge becomes available, so biophysical methods may be used to provide descriptions of the scalp distribution of a component.

b. Statistical methods. A second class of methods includes those that are based on some statistical or mathematical properties of the data. For instance, by analyzing the pattern of variance and covariance among electrodes at a particular time point, Lehmann and Skrandies (1980) and Wood, McCarthy, and Darcey (1986) were able to derive a set of components to describe the scalp distribution at that time point. Each of the components could be described in terms of its own scalp distribution. To the extent that this scalp distribution can be reliably observed, it can be used to define the scalp distribution of that particular component.

Another statistical procedure to determine an optimal Vector filter is to select that set of weights that best discriminates between two data sets, one of which is believed to contain the component of interest. For example, discriminant analysis can be used to derive the set of scalp distribution weights that best discriminate between the single trial ERPs to expected and unexpected events. If one believes that these two classes of events elicit P300s of different magnitudes (and that this is the only difference between the two classes), then it is possible to use the set of weights to characterize the P300 component. This may be particularly useful when few electrodes are used such that only one filter can be applied.

Information about the constancy of a particular scalp distribution over time can also be used to arrive at a definition of components. This method is based on the assumption that, if a number of consecutive time points share a similar scalp distribution, they can be attributed to the same underlying component. Thus, Lehmann and Skrandies (1986) divide the observed ERP waveform into segments, such that the time points within each segment are associated with a "constant" scalp distribution. Each segment is then defined as a component. Constancy occurs when, for two timepoints, the maximum and minimum voltages occur at the same (or adjacent) electrodes. Constancy could also be defined in terms of the magnitude of the R(12)'s (correlation) between two distributions, as described above.

Other techniques are available that rely on more restrictive models of ERPs, and which may allow one to derive analytically the scalp distribution of the ERP components. For instance, Mocks (1987) recently proposed a model that assumes that the ERP is the result of the linear combination of a limited set of components whose time course and scalp distribution are fixed, although not known. When these assumptions are met, it is possible to estimate the time course and scalp distribution of each component from the data. Although this procedure may be promising in some cases, it is not applicable when the time course of a component varies.

Scherg and von Cramon (1985) proposed a mixed biophysical/analytical procedure to disentangle the component structure of the ERP. Their procedure uses both temporal and spatial information. First, each component is assigned a specific waveshape, although some flexibility in latency is allowed. Initial hypotheses are made about the nature, locus and orientation of the source of each component, and their consequent scalp distribution. The ERP predicted by this model is then computed, and the discrepancy between the predicted and observed ERP is measured. The parameters (latency, source localization and source orientation of each component) are then adjusted to reduce the discrepancy. The procedure is repeated until a convergence toward a particular solution is reached. This approach appears to provide a promising way of integrating biophysical and statistical methods. However, like other procedures, it is based on a number of assumptions, including the number, time course, and nature of the components.

c. Experimental methods. These methods are predicated on the belief that there is some experimental setting that can be used to provide a "pure" instance of a particular component. For example, the ERP elicited by the omission of an expected stimulus is thought to represent a P300 that is uncontaminated by the superimposition of exogenous components (Simson, Vaughan, & Ritter, 1976). Alternatively, P300 can be defined in terms of that aspect of an ERP waveform that is probability dependent (Squires & Donchin, 1976). In the case of Processing Negativity, the component is identified by subtraction of two waveforms, those for attended and unattended stimuli (Naatanen, 1982). The difficulty with these procedures is that it must be asssumed that the derived waveforms contain one and only one component. If this condition is satisfied, then the scalp distribution of the pure component can be used.

d. Evaluation of the methods for estimating scalp distribution. In summary, several procedures can be used to identify the scalp distribution of an ERP component. For theoretical reasons, biophysical methods are clearly the most appropriate. However, since knowledge of the sources of ERP components is limited at the present time, one must resort to other techniques. Although none of these appears to be applicable in all cases, the convergence of several techniques towards a similar solution may be considered as a strong evidence for that solution. Such convergence is illustrated in the next section.

e. A practical example: The scalp distribution of P300. In this section we provide an example of how different procedures used to estimate the scalp distribution of P300 converge on a common solution. The data from this example come from a study by Miller et al. (1987). During this study, 10 subjects were run in four tasks. In the first task (AUDITORY COUNT 50/50), the subjects had to count one of two tones, where both tones had a probability of .50. In the second task (AUDITORY RT 20/80), the subjects had to give a speeded response to one of two tones, with probabilities of .80 and .20 respectively. In the third task (AUDITORY OMITTED STIMULUS 10/90), the subjects were presented with a series of tones; on 10% of the trials the tone was omitted, and the subjects had to count the number of omissions. In the fourth task (VISUAL NAMES 20/80), the subjects were presented with male names (p=.20) and female names (p=.80) and had to count the number of male names. ERPs were recorded from Fz, Cz, and Pz, referenced to linked mastoids. The grand-average waveforms for each electrode, task, and stimulus type are presented in Figure 4.

Insert Figure 4 About Here

A "P300" is clearly noticeable for all the rare or target stimuli. For the purposes of this demonstration we derived four definitions of P300, one for each task. Thus P300 was defined as:

(a) the difference (at each electrode) between the amplitude at the moment of the maximum positive peak after 300 ms for the "target" and "non-target" trials in the AUDITORY COUNT 50/50 task ("target effect/auditory");

(b) the difference (for each electrode) between the amplitude at the moment of the most positive peak after 300 ms for the "rare" and "frequent" trials in the AUDITORY RT 20/80 task ("probability effect/auditory");

(c) the amplitude (at each electrode) at the moment of the most positive peak after 300 ms for the "omitted" stimulus trials in the AUDITORY OMITTED STIMULUS 10/90 task ("omitted stimulus/auditory");

(d) the difference (for each electrode) between the amplitude at the moment of the most positive peak after 300 ms for the "rare" and "frequent" trials in the VISUAL NAME 20/80 task ("target and probability effects/visual").

This set of independent procedures yields different definitions of P300 scalp distribution. They cover different modalities, latencies, probabilities, task requirements, and included different sets of possible overlapping components. Note that all these procedures belong to the class of "experimental methods" as described above.

The mean P300 amplitudes at each electrode for each of the four definitions are presented in Table 2. In this table, the average of the four definitions of P300 for each electrode is also shown.

Insert Table 2 about here

To determine whether these four definitions of P300 result in a consistent scalp distribution, we computed the R(12) between each pair of definitions, and between each definition and the average of the four definitions. The resulting values of R(12) are shown in Table 3.

Insert Table 3 about here

It is clear from Table 3 that the different definitions yield very similar scalp distributions, and that each distribution is very similar to the average distribution. Thus, it appears that the four procedures used to estimate the scalp distribution of P300 converge on the same solution, represented by the average definition.

Discussion

In this paper we have seen how one can proceed from the assumption that ERP components are characterized by particular scalp distributions, to the identification and quantification of ERP components in an observed waveform. We noted that this assumption is held by many researchers in the field, and that it appears to be based on the belief that two different scalp distributions cannot be produced by the same neural generator.

Although this assumption is widespread, it has generally been used only as a qualitative guide to component identification. The assumption leads to the selection of particular recording locations. For instance, studies of movement related potentials generally include lateral electrode placements, while those of P300 include midline montages. The assumption also often leads to the selection of one particular scalp site for the analysis, where the component is supposed to be maximally evident.

The procedure we have proposed (Vector filter) allows the investigator to take advantage of the quantitative implications of the assumption. Vector filter is based on a model that assumes that the observed ERP is given by the linear combination of a set of components. Note that the following assumptions are made: (a) the set of components is limited, (b) the scalp distribution of each component is known and fixed, and (c) the components combine in a linear fashion (without interaction) in determining the composite ERP. We have shown how investigators can choose the set of components, and we have discussed the consequences of choosing the incorrect set. We have also considered ways in which the scalp distribution of a particular component can be established, and we have illustrated how four different ways of defining the scalp distribution of P300 converge. We have not addressed the validity of the assumption of linear combination. However, it should be noted that this assumption is common to most decomposition techniques (cf. PCA, waveform subtraction).

There are two ways in which Vector filter can be useful. First, Vector filter can provide a decomposition of the ERP into components. This use of the procedure requires that the three assumptions reviewed in the previous paragraph apply. Violation of these assumptions will lead to a misrepresentation of the component structure. Note that one can test the degree to which the chosen set of components account for the observed data using statistical tests. If these tests reveal that the model satisfactorily accounts for the data, then the model and its assumptions are supported. Note however that "a good fit" does not exclude the possibility that other models might apply.

A second use of Vector filter is to enhance the signal (component of interest) with respect to other aspects of the ERP waveform. In this case, the procedure is used to "filter for" the scalp distribution of a particular component, or to "filter out" the contribution of components or noise with different scalp distributions. In this case, it is not necessary for the assumptions that underlie the procedure to be completely valid. Just as with frequency filtering it is not necessary to know the exact frequency of the signal of interest (and of the noise), so with Vector filter it is not necessary to know the exact scalp distribution of the component of interest (and of the noise). (An investigation of the effects of imprecision in the estimate of component distribution is presented in a companion paper, Gratton, Kramer, Coles, & Donchin, submitted.)

Both these uses of Vector filter represent solutions to the problem of isolating overlapping components. The advantage of the procedure over other methods (e.g., PCA, waveform subtraction) is that it does not require that the latency and timecourse of the components are constant over different trials, subjects, or experimental conditions. Thus, Vector filter appears to be particularly suitable for studies that involve latency variability.

We have evaluated the use of Vector filter to improve the signal-to-noise ratio of the P300 component in two studies. In the first study (Gratton et al., submitted) we simulated complex ERPs by adding several different components and noise to a P300. In this study we found that Vector filter helps in the detection of P300, by improving the estimation of P300 latency by approximately 25%. In a second study (Fabiani, Gratton, Karis, & Donchin, in press) we measured the reliability of P300 amplitude and latency estimates taken with several different procedures on observed data. In these studies we found that the best results are obtained when Vector filter is used in association with time-domain (peak-picking and cross-correlation) procedures for the identification of P300.

The examples of the use of Vector filter reviewed above show how scalp distribution information can be used in the analysis of ERP waveforms, despite the limited knowledge available at present about the sources of components. It is clear, however, that as this knowledge accumulates, so the procedure will become more and more reliable.

Footnotes

1. The term "Vector filter" was originally used to refer to the particular geometrical properties of this procedure when a multivariate description of ERP scalp distribution data is used. However, in the present paper, we adopted an alternative approach to the description of the procedure. We believe that this non-geometric approach is easier both for us to present and for the reader to follow. However, for historical reasons, we have chosen to maintain the label "Vector filter."

2. For example, the Contingent Negative Variation (CNV) is described as "largest at the vertex ... Amplitudes also show a progressive diminution anteriorly and posteriorly from central regions ..." (Donchin et al., 1978, pp. 357-358). The P300 "is characterized by its scalp distribution as it tends to be larger at central and parietal electrodes" (Donchin et al., 1978, p. 356). The N400 "has a marked posterior distribution over the scalp, with a slight but consistent right hemispheric predominance ..." (Kutas, Lindamood, & Hillyard, 1984, p. 219). Naatanen and Picton (1987) use scalp distribution and other procedures to distinguish six components of the auditory N1.

3. In this example we used three electrodes (Fz, Cz, and Pz), but the number of electrode locations used can be any number larger than 1.

References

Brunia, C. H. M. (1980). What is wrong with legs in motor preparation? In H. H. Kornhuber & L. Deecke (Eds.), Motivation, motor and sensory processes of the brain. Progress in Brain Research, Vol. 54 (pp. 232-236). Amsterdam: Elsevier/North Holland.

Coles, M. G. H., Gratton, G., Kramer, A. F., & Miller, G. A. (1986). Principles of signal acquisition and analysis. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), Psychophysiology: Systems, Processes, and Applications (pp. 183-221). New York, NY: Guilford.

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 Potentials research (pp. 501-510) (EPA-600/9-77-043). US Govt. Printing Office.

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) (EPA-600/9-77-043). US Govt. Printing Office.

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 (pp. 349-411). New York: Academic Press.

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, R. Jennings, and M.G.H. Coles (Eds.), Advances in Psychophysiology, vol. II. Greenwich, CT: JAI.

Gratton, G., Kramer, A.F., Coles, M.G.H., & Donchin, E. (submitted). Simulation studies of latency measures of components of the Event-Related brain Potential. Psychophysiology.

Hjorth, B. (1979). Multichannel EEG preprocessing: analogue matrix operations in the study of local effects. Pharmacopsychiatry, 12, 111-118.

Kutas, M., & Hillyard, S. A. (1982). The lateral distribution of event-related potentials during sentence processing. Neuropsychologia, 20, 579-590.

Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum & J. Requin (Eds.), Preparatory states and processes (pp. 217-237). Hillsdale, NJ: Erlbaum.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. Electroencephalography and clinical Neurophysiology, 48, 609-621.

Lehmann, D., & Skrandies, W. (1986). Segmentation of evoked potentials based on spatial field configuration in multichannel recordings. In W.C. McCallum, R. Zappoli, & F. Denoth (Eds.), Cerebral Psychophysiology: Studies in event-related potentials (EEG Suppl. 38) (pp. 27-29). Amsterdam: Elsevier.

Lewis-Beck, M.S. (1980). Applied regression: An introduction. Beverly Hills: Sage.

McCarthy, G., & Wood, C.C. (1985). Scalp distribution of event-related potentials: An ambiguity associated with analysis of variance models. Electroencephalography and Clinical Neurophysiology, 62, 203-208.

McNemar, Q. (1969). Psychological statistics. New York: Wiley.

Miller, G.A., Bashore, T.B., Marshall, N., Gratton, G., & Donchin, E. (1987). Study of variation of event-related brain potentials with aging. Unpublished manuscript.

Mocks, J. (1987, June). A new method of ERPs analysis. Paper presented at the IV Conference on Cognitive Neuroscience, Dourdan, France.

Naatanen, R. (1982). Processing negativity-evoked potential reflection of selective attention. Psychological Bulletin, 92, 605-640.

Naatanen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and analysis of the component structure. Psychophysiology, 24, 375-425.

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.

Scherg, M., & Von Cramon, D. (1985). Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. Electroencephalography and Clinical Neurophysiology, 62, 32-44.

Simson, R., Vaughan, Jr., H.G., & Ritter, W. (1976). The scalp topography of potentials associated with missing visual stimuli. Electroencephalography and clinical Neurophysiology, 40, 33-42.

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.

Wood, C.C., McCarthy, G., & Darcey, J. (1986). Principal component analysis and dipole localization models applied to surface potentials generated by multiple dipole sources. Electroencephalography anmd clinical Neurophysiology, 64, 80).

Table 1. Expanded Model

Voltages observed at each electrode		Contribution of each component (A,B,...,Z) to the voltage observed at each electrode, given by the product of the coefficients (c) of the component and of its weight for that electrode (A, B, ... Z)	Error term for each electrode
Volt (1)	=	c(A) x A(1) + c(B) x B(1) + ... + c(Z) x Z(1) +	Err(1)
Volt (2)	=	c(A) x A(2) + c(B) x B(2) + ... + c(Z) x Z(2) +	Err(2)
Volt (3)	=	c(A) x A(3) + c(B) x B(3) + ... + c(Z) x Z(3) +	Err(3)
...	=	... ...	...
Volt (N)	=	c(A) x A(N) + c(B) x B(N) + ... + c(Z) x Z(N) +	Err(N)

Notes: Volt(1 to N) are the observed values at electrode (1 to N);

A(1 to N), B(1 to N), etc. are the weights of the spatial filters, corresponding to components (A to Z), at electrode (1 to N) - these weights are assumed to be known;

c(A to Z) are the coefficients that are estimated by the least squares procedure that minimizes the sum of the squared errors (Err(1 to N)) for each electrode.

Table 2. Mean P300 amplitude at each electrode (microvolts)

Definition	Fz	Cz	Pz
Target Effect/Auditory	0.1	0.8	0.7
Probability Effect/Auditory	2.8	5.1	4.3
Omitted Stimulus/Auditory	-1.5	4.1	6.9
Target & Prob. Effects/Visual	1.2	2.6	3.1
Average	0.65	3.15	3.75

Table 3. R(12s) of the scalp distributions of different definitions of P300

Definition	TE/A	PE/A	OS/A	TPE/V
Target Effect/Auditory				
Probability Effect/Auditory	.957			
Omitted Stimulus/Auditory	.918	.795		
Target & Prob. Effects/Visual	.972	.990	.907	
Average	.986	.956	.945	.993

Figure Legends

Fig 1. First row: Scalp distribution of two hypothetical ERP components (Components A and B). Second row: Time course of Components A and B. Third row: Composite ERP waveform given by the sum of Components A and B. Fourth row: Time course of Components A and B as retrieved by the Vector filter procedure.

Fig 2. First row: Scalp distribution of two hypothetical ERP components (Components A and B). Second row: Time course of Components A and B. Third row: Composite ERP waveform given by the sum of Components A and B. Fourth row: Time course of Component A as retrieved by the Vector filter procedure when only Component A is filtered for.

Fig 3. First row: Scalp distribution of three hypothetical ERP components (Components A, B, and C). Second row: Time course of Components A, B, and C. Third row: Composite ERP waveform given by the sum of Components A, B, and C. Fourth row: Time course of Components A and B as retrieved by the Vector filter procedure.

Fig 4. From Miller et al. (in preparation). Grand average waveforms at Fz (solid), Cz (dashed), and Pz (dotted) for each task and stimulus (n=10).

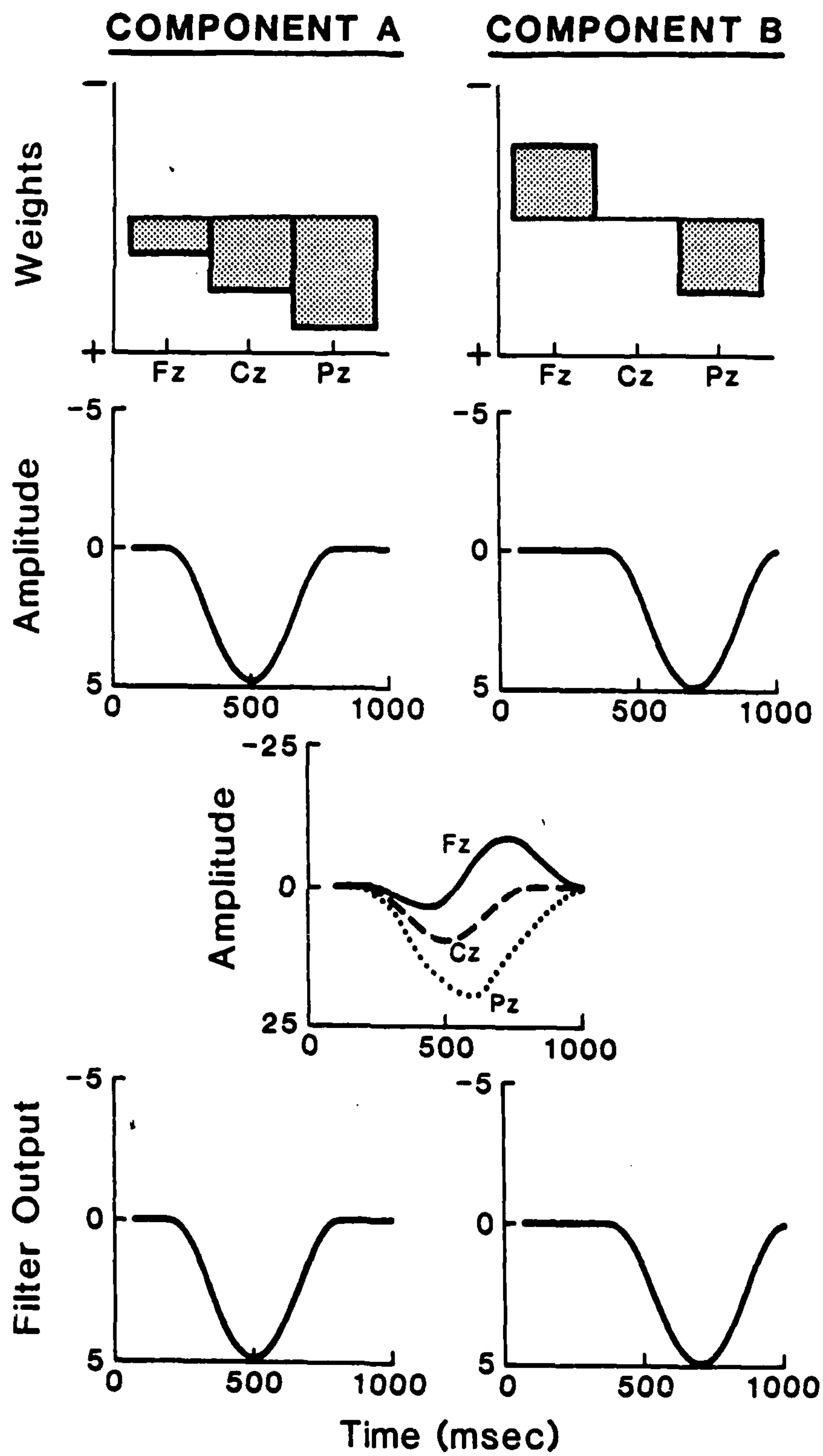
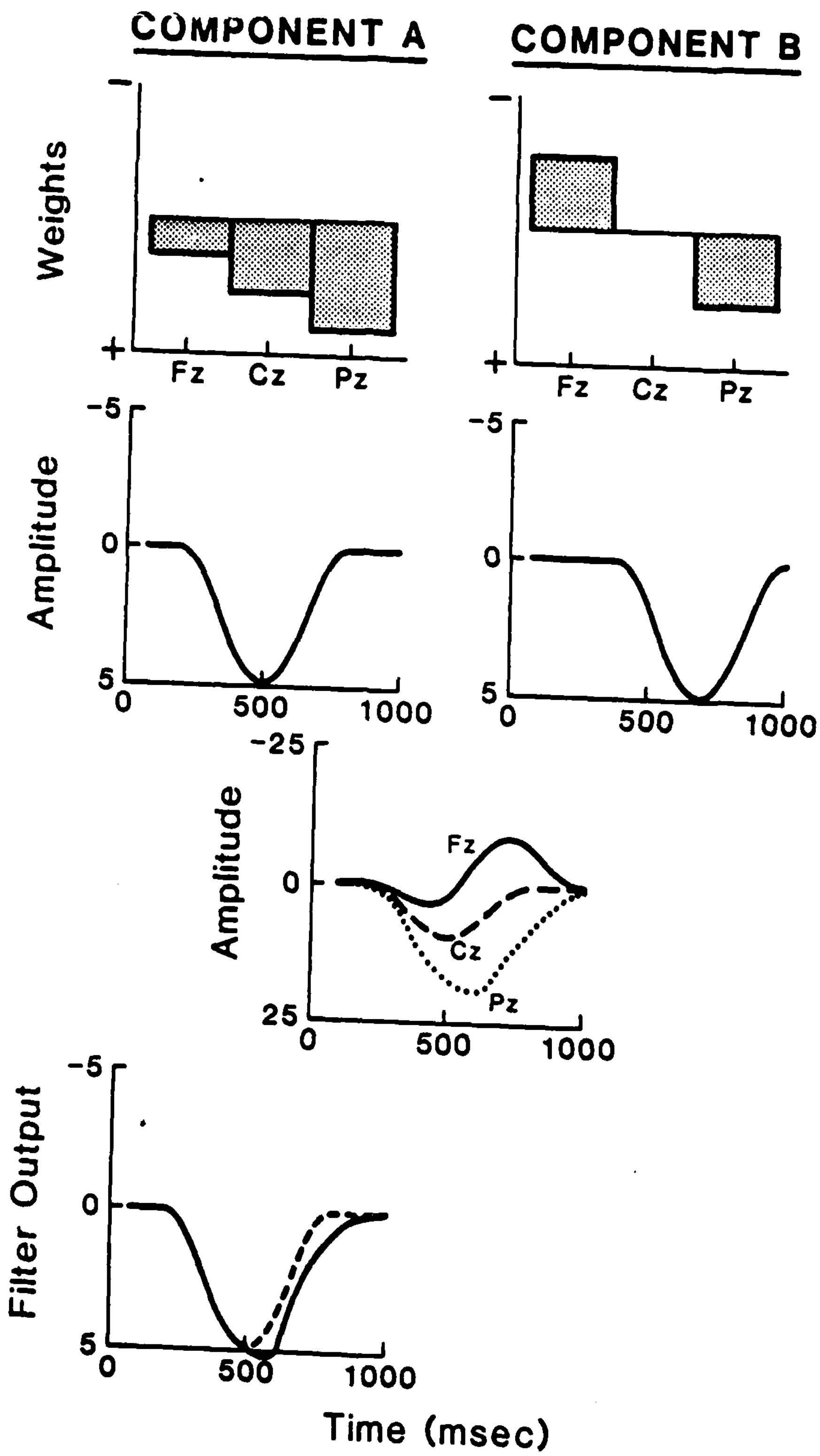

COMPONENT A
COMPONENT B
Weights
Fz
Cz
Pz
Amplitude
Filter Output
Time (msec)
-5
0
5
500
1000
-25
25

Fig. 1



COMPONENT A
COMPONENT B
Weights
−
+
Fz
Cz
Pz
Amplitude
−5
0
5
0
500
1000
−25
Fz
Cz
Pz
25
Filter Output
Time (msec)

Fig. 2

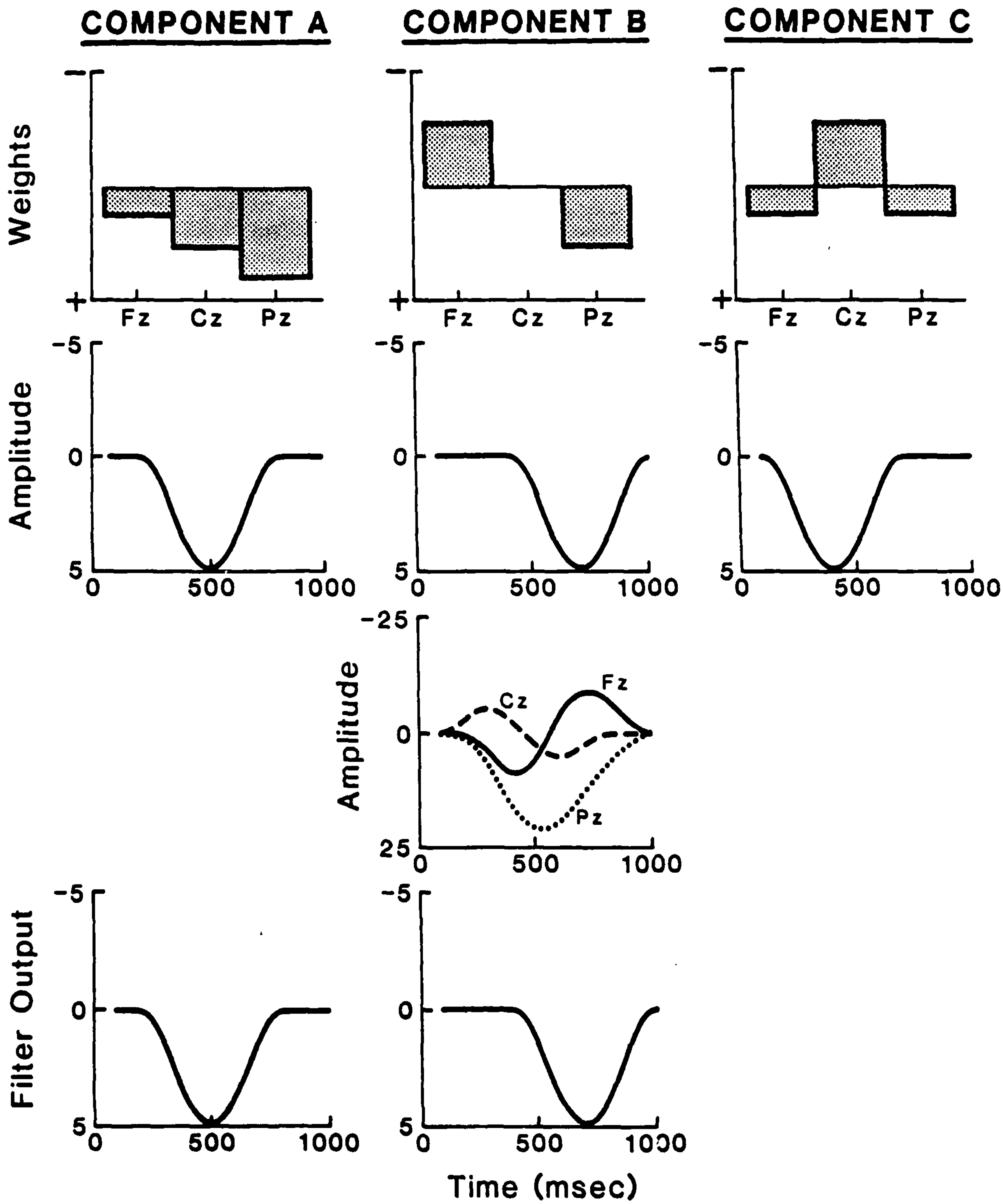
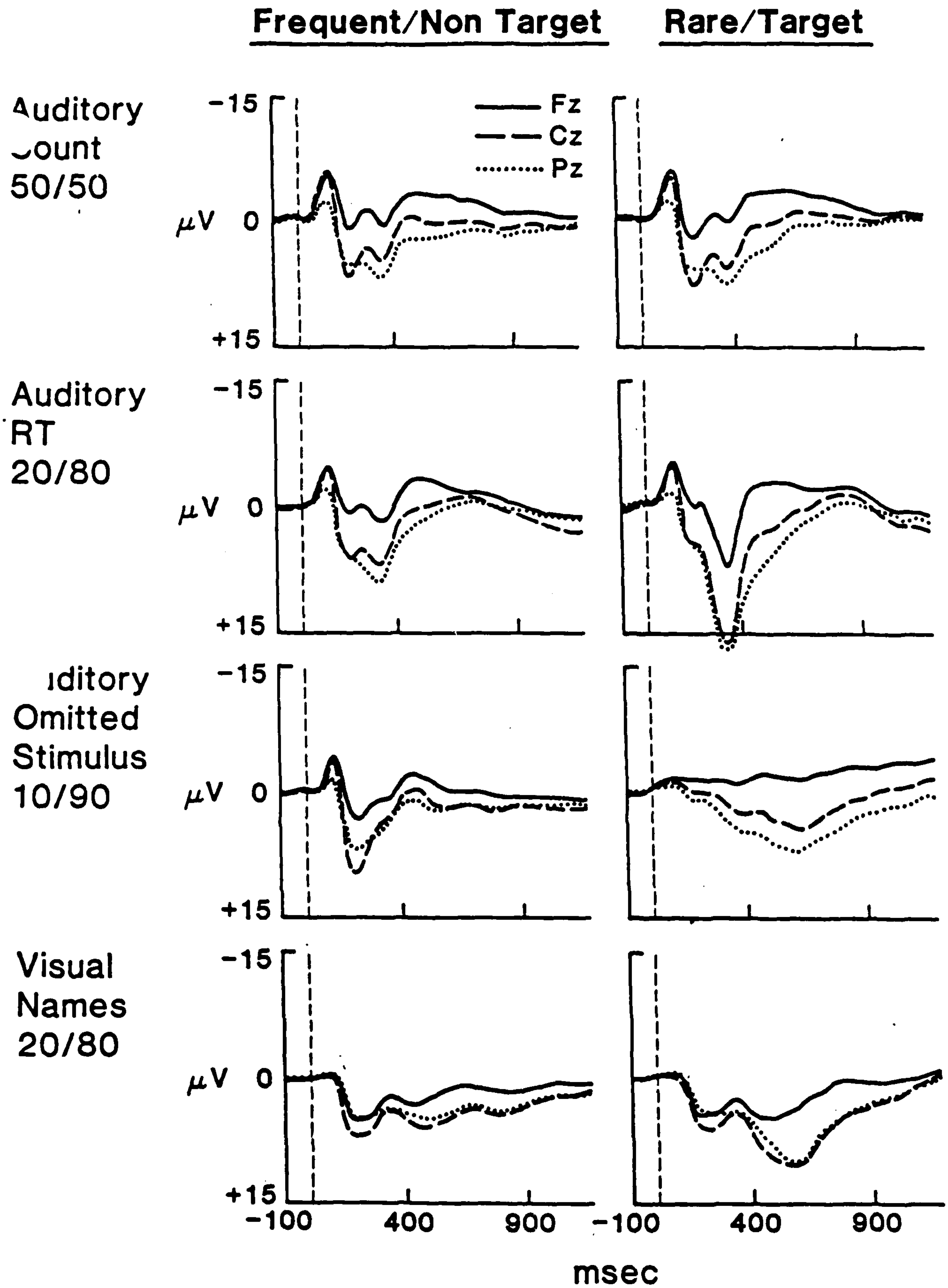

COMPONENT A
COMPONENT B
COMPONENT C
Weights
Fz
Cz
Pz
Amplitude
-5
0
5
500
1000
-25
25
Filter Output
Time (msec)

Fig. 3



Frequent/Non Target
Rare/Target
Fz
Cz
Pz
Auditory Count 50/50
Auditory RT 20/80
Auditory Omitted Stimulus 10/90
Visual Names 20/80
μV
−15
0
+15
−100
400
900
msec

Fig. 4

Simulation Studies of Latency Measures of Components of the Event-Related Brain Potential

Gabriele Gratton, Arthur F. Kramer, Michael G.H. Coles, and Emanuel Donchin

Cognitive Psychophysiology Laboratory
University of Illinois, Champaign, Illinois

Acknowledgements

This research was supported in part by a grant from the Air Force Office of Scientific Research, contract #F49620-83-0144, with Dr. Al Fregly as Project Director, and by a grant from the National Institute of Mental Health, MH41445. 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.

We wish to thank Monica Fabiani, Demetrios Karis, Greg Miller, Chris Wood, and two anonymous reviewer for their helpful comments on earlier versions of this paper.

Send reprint requests to: Gabriele Gratton, University of Illinois, Psychology Department, 603 E. Daniel, Champaign, Illinois, 61820.

Running title: Simulation studies of latency measures

Abstract

We compared the accuracy of P300 latency estimates obtained with different procedures under several simulated signal and noise conditions. Both preparatory and signal detection techniques were used. Preparatory techniques included frequency filters and spatial filters (single electrode selection and Vector filter). Signal detection techniques included peak-picking, cross-correlation, and Woody filter. Accuracy in the latency estimation increased exponentially as a function of the signal-to-noise ratio. Both Woody filter and cross-correlation provided better estimates than peak-picking, although this advantage was reduced by frequency filtering. For all signal detection techniques, Vector filter provided better estimates than single electrode selection. Large component overlap impaired the accuracy of the estimates obtained with single electrode 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.

Descriptors: P300 Latency Measures, Simulations, Frequency Filters, Spatial Filters, Vector Filter, Peak-picking, Cross-correlation, Woody Filter.

In this paper we compare several of the techniques that are commonly used for the measurement of the latency of components of the Event-Related Brain Potential (ERP). In this paper we will focus on one particular component, the P300. Theoretical speculation (e.g., Coles & Gratton, 1985; Coles, Gratton, Bashore, Eriksen, & Donchin, 1985; Donchin, 1981) and empirical evidence (e.g., Duncan-Johnson & Donchin, 1982; Kutas, McCarthy, & Donchin, 1977; McCarthy & Donchin, 1981; Ragot, 1984) suggest the utility of the latency of P300 in the study of cognition. However, since ERP components are estimated from data containing substantial amounts of noise, the measures of latency necessarily yield approximations to the "true" latency of the components. In the studies reported here, we simulated ERP data so that known ERP components were embedded in noise of different amounts and qualities. The degree to which different procedures accurately estimated the known parameters of the ERP components was then 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 to be used (see Fabiani, Gratton, Karis, & Donchin, in press, for an overview of different procedures used for the definition of one particular ERP component - the 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 our studies.

A procedure designed to estimate the latency of an ERP component should take into account the characteristics of both the signal (i.e., the ERP component of interest) and the noise (i.e., all the other electrical activity recorded by the scalp electrode). From the point of view of the ERP researcher, there are at least two categories of noise. The electrodes record, in addition to the component of interest, the background EEG (random noise) and the activity of concurrent ERP components (systematic noise).1 (Systematic noise refers to the fact that this type of noise does not average to zero across a large number of trials.) Previous studies that have simulated ERPs have focussed on the effect of random noise (intended to simulate background EEG) on the accuracy of latency estimates (see Pfefferbaum, 1983; Woody, 1967). However, a thorough evaluation of latency estimation procedures can be obtained only when the effects of both random and systematic noise are examined.

The term "background EEG" refers to 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 (i.e., not white noise), 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 the detection of a component when there is only noise present (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 filter," Wiener, 1949). In our studies, we simulated the background noise by using the deviation of single-trial ERPs from the average ERP.

The second source of noise is provided by those ERP components that temporally and spatially overlap the target component. In this case, since ERP components are believed to be invariant across trials, the average value of the noise over a large number of trials represents an estimate 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, 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 also 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 (see a related discussion about Wiener filter in Wastell, 1981).

In our studies, 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, 1986). 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 shifts. Such filters may be used during the preparation of the ERP data for signal detection techniques. In our studies, 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 multielectrode information for improving signal detection). In general, researchers have simply selected one electrode location to use for further analysis. (We will label this procedure "single electrode selection"). One of the goals of this paper 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 filter), described by Gratton, Coles, and Donchin (submitted).2 Vector filter assumes that scalp distribution is a defining characteristic of an ERP component. Therefore information about 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 Vector filter with that of those obtained by selecting a single electrode. Vector filter can be thought of as a linear filter where the weights for each electrode are chosen to optimize the detection of the target component. Single electrode selection can be thought of as a linear filter with a weight of 1 assigned to the selected electrode, and a weight of 0 assigned to all the other electrodes.

The purpose 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 et al., 1986). 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 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 consider a component as that segment of a 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. Component latency is defined as the "lag" between the template and the waveform necessary to produce the maximum resemblance. Woody (1967) proposed a particular variant of the cross-correlation procedure, where the template is "adapted" to the average waveform through several iterations (Woody filter).3 Since filtering affects the signal-to-noise ratio, its impact on different signal detection procedures will also be considered in the present paper.

There were three phases to our studies. 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 estimates was computed, and the merits of each procedure were evaluated.

We should note the differences between the basic design (labelled Simulation 1) and the control conditions we ran to investigate particular problems (labelled Simulation 2, 3, 4, 5, and 6).

Simulation 1: Basic Design

In this section we discuss the results obtained from the basic design. We shall present 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 Vector filter, on the accuracy of the latency estimates in later sections.

Method

The basic design consisted of a factorial combination of several simulated conditions and analytic procedures. The simulated conditions yielded 220 different conditions (11 P300 amplitude levels x 20 overlapping component conditions). For each condition we obtained 100 repetitions by adding background EEG noise (to be described in a subsequent section), sampled at random and with reselection from a set of 100 background noise trials (note that the relationship between waveform and recording electrode - Fz, Cz, and Pz - was maintained). This procedure was repeated five times, once for each of the five filtering conditions (four filter settings and one control condition where no filter was used). This produced a total of 110,000 waveforms for each of the three electrodes. Each of these sets of waveforms was filtered spatially using either single electrode selection or Vector filter. 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.

Note that evaluation of the different signal-detection algorithms and of the different spatial filtering techniques was based on their performance for the same set of waveforms. However, comparisons among different frequency filtering procedures, overlap conditions, and P300 amplitudes were based on different waveforms, obtained by sampling independently from the set of background noise trials. A list of the experimental conditions for the basic design is shown in Table 1.

Insert Table 1 about here

Model. Our simulations were based on the following model:

Volt(it) = cP300(t) x P300(i) + SUM (cOVER(jt) x OVER(ij)) + N(it)

where:

Volt(it) is the potential recorded at electrode i at time t;

cP300(t) is the amplitude of the P300 component at time t;

P300(i) is the weight of the P300 component at electrode i;

cOVER(jt) is the amplitude of the overlapping component j at time t;

OVER(ij) is the weight of the overlapping component j at electrode i;

N(it) is the background EEG noise at 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 used in the Vector filter procedure (Gratton et al., submitted, and see Footnote 2).

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, 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 vary, 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" 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 as described in the ERP literature (see Donchin, Ritter, & McCallum, 1978). The parameters of N100 and P200 were not varied systematically, since they do not overlap temporally with P300 in our data set, and, therefore, should not affect the P300 latency estimates. The amplitude and latency of N200 and Slow Wave were systematically varied to simulate the different degrees of overlap with P300 that may be present in observed 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 2.

Insert Table 2 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 0.1 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. We chose to label this value "signal-to-noise ratio." (Strictly speaking the value should be referred to as the "signal-to-background-noise ratio". For reasons discussed earlier, we do not include systematic noise, or component overlap, in this estimate.) This ratio varied systematically from 0 to 5, in half unit increments. Both N200 and Slow Wave partially overlapped with P300. N200 had a frontally maximum scalp distribution that was clearly different from that of P300. The scalp distribution of the Slow Wave was maximally positive at Pz, and was therefore more similar to that of P300. These manipulations allowed us to study the impact of different levels of component overlap (using 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 one set 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 Slow Wave manipulations, and four repetitions of the control condition (where the amplitudes of the overlapping component were zero), 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 noise. Estimates of background EEG activity were obtained from a set of 100 trials recorded from an individual subject in an oddball experiment (Fabiani et al., in press).4 The background EEG was obtained by subtracting the appropriate average from each single-trial record. A power spectral analysis of these waveforms revealed maxima in the following bands: 0-4 Hz for all electrodes, 8-10 Hz for the parietal electrode, and 14-16 Hz for all electrodes.

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 (perhaps due to sampling variation), and smaller variance during the pre-stimulus period. Under these conditions, 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 (Simulation 2).

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. In particular the relationship among electrodes is altered, double peaks are noticeable, etc.

Off-line Frequency Filters. The study included a comparison among four off-line, low-pass frequency filtering procedures, and a no filtering condition. All the filters were based on a moving average (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), and two iteration levels (1 and 2 iterations) were used. Moving average filters cannot be perfectly described in terms of the half-amplitude cut-off frequencies (Ruchkin & Glaser, 1978). However, a running average of 7 points corresponds roughly to a -3 dB cut-off point at 6.29 Hz when one iteration is used, and at 4.43 Hz when two iterations are used. For 13 consecutive points the corresponding values are 3.38 Hz and 2.38 Hz, respectively.

We should emphasize that the comparison between filtering procedures described above does not exhaust all off-line frequency filters available to the investigator. However, Simulation 5 considers the impact of a wider range of less commonly used filters (up to 1.42 Hz).

Spatial Filters. Two types of spatial filters were compared: single electrode selection and Vector filter.

a. Single electrode 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.

b. Vector Filter. This procedure, described by Gratton et al.

(submitted), 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.

In the present case, we used a Vector filter characterized by the following weights: 0.15 for Fz, -0.53 for Cz, and 0.83 for Pz. The choice of these weights was based on a previous study in which they were found to produce on optimal discrimination between two groups of rare and frequent trials. As we have noted elsewhere (Gratton et al., submitted), component weights can be estimated by using discriminant analysis to determine the optimal discrimination between a set of waveforms in which the component of interest is present and another set in which it is absent (or smaller). Since the P300 component is larger for rare than for frequent stimuli (e.g., Duncan-Johnson & Donchin, 1977), the weights derived from the discriminant function procedure should represent the P300 component. Note that the set of weights does not correspond to the typical P300 distribution. This is because the background noise is not "random." Further discussion of this point is provided in this Simulation 6, in which we varied parametrically the weights of Vector filter 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.

In this paper we adopted a procedure which allowed us to evaluate both cross-correlation and Woody filter. We used a template equivalent to the P300 component that we used to generate the simulated waveforms for the first iteration of the Woody filter procedure. Thus, this 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.

For each signal detection procedure, the temporal window began 300 msec post-stimulus and ended 800 msec post-stimulus (250 msec before and 250 msec after the P300 peak). The duration of the template used for the cross-correlation algorithm was 500 msec.

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 the latency estimates;

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, because we assumed that the variability in MSE obtained with different procedures was proportional to its absolute value.

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.5

Results

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 Figure 2.

Insert Figure 2 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 log MSE. 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 2. 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 the latter 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 Vector filter as a spatial filtering technique reduced the error in latency estimation in comparison with single electrode selection (Pz). This advantage is evident at intermediate and high signal-to-noise ratios. At low or intermediate signal-to-noise ratios (0.5 to 2.0), the advantage of Vector filter is comparable to that of cross-correlation. However, the advantage of Vector filter rarely reaches the 50% level and is usually about 25%. The advantages of Vector filter and of cross-correlation appear to be independent. Fourth, low-pass frequency filters produced marked improvements in the accuracy of latency estimation. The largest improvement was obtained with a 2.38 Hz low-pass filter. Low-pass filters with a higher cut-off 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 Vector filter was unaffected, and in fact the smallest MSE values were obtained by the joint use of frequency filters, Vector filter, 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 3.

Insert Figure 3 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 become 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 applied to Pz waveforms, and for all procedures at a signal-to-noise ratio of 1, where the skewed distribution indicates the presence of systematic error.

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 4.

Insert Figure 4 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. In fact, at a signal-to-noise ratio of 5, the amplitude of P300 at Pz is 500 units, while with Vector filter it is 260 units -- a reduction to 52% of Pz amplitude that can be predicted on the basis of the weights used. This interaction of the performance of latency adjustment procedures with signal-to-noise ratio 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 Vector filter. This effect is confounded in part with the overall reduction in amplitude produced by Vector filter. 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. Log 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 filters) are shown in Figure 5.

Insert Figure 5 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, Slow Wave amplitude = 100 units, Slow Wave latency = 1280 msec), "large N200 overlap" (N200 amplitude = 100 units, N200 latency = 400 msec, Slow Wave amplitude = 100 units, Slow Wave latency = 1280 msec), and "large Slow Wave overlap" (N200 amplitude = 100 units, N200 latency = 300 msec, Slow Wave amplitude = 200 units, Slow Wave latency = 1000 msec).

Inspection of this figure reveals that component overlap impaired the accuracy obtained with each procedure to a different degree. The relative advantage of procedures based on Vector filter in comparison to those based on Pz channel selection interacted with the degree and type of component overlap. In particular, the advantage was smaller for the "large Slow Wave overlap" condition than for the other conditions. We should note here that the scalp distribution of Slow Wave was close to that of P300, while the scalp distribution of N200 was quite different. Thus, the improvement obtained with Vector filter is not affected by overlapping components with a scalp distribution different from that of P300, but is affected by an overlapping component with a scalp distribution similar to that of P300. However, even in the worst case (large Slow Wave 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.

Discussion

The following conclusions can be drawn from the data presented in the Results section:

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 the same as the target component).

4. Frequency filters improve markedly the accuracy of estimates obtained with peak-picking and, to a lesser extent, with cross-correlation.
5. Vector filter yields estimates that are more accurate than single electrode selection (Pz).
6. Overlapping components impair the accuracy of estimates of single electrode selection, while the latency estimates of Vector filtered data are impaired only if the scalp distribution is similar to that of P300 (e.g. Slow Wave).
7. The accuracy improvements obtained with cross-correlation, Vector filter, and increases in signal-to-noise ratio appear to be independent. The improvements in accuracy obtained with Vector filter and frequency filters are also independent.
8. The accuracy improvements obtained with cross-correlation and frequency filtering are <u>not</u> additive -- that is, the combined use of both these procedures does not produce more accurate estimates 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.

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, five additional analyses were performed. These analyses explored the effects of five variables on the accuracy of latency estimates: (a) standardizing background EEG noise, (b) varying the duration of P300, (c) varying the latency of P300, (d) varying the parameters of the frequency filter, and (e) varying the parameters used for Vector filter.

Simulation 2: 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 might affect the veridicality of our simulation procedure.

Method

To investigate further the effects of the standardization procedure, we ran part of the basic design (Simulation 1) on non-standardized waveforms. The replication was exact, apart from the absence of frequency filtering. 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.

Results

Some of the results obtained with non-standardized noise are presented in Figure 6.

Insert Figure 6 About Here

Discussion

Comparison of the accuracy of latency estimation with standardized (Figure 5, upper left panel) and non-standardized background noise (Figure 6) indicates that the standardization procedure did not significantly affect the results. All the findings were replicated. Thus, the standardization procedure did not impair the veridicality of the simulation.

Simulation 3: Effect of P300 duration.

Since the template we used for the cross-correlation technique was the same as the target component itself, the detection of P300 obtained with this algorithm may be more accurate than that which could be obtained with observed data, when the "true" P300 waveshape is not perfectly known. Therefore, Simulation 1 does not really address the question of the advantage of iterating with Woody filter. In Simulation 3, we considered cases in which we do not have a good representation of the P300 waveshape.

Method

To this end, 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.

Results

Log 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 Vector filter) and four signal detection algorithms (peak-picking, cross-correlation, Woody filter with two iterations, and Woody filter with three iterations) are shown in Figure 7.

Insert Figure 7 About Here

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 Vector filter) 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.

Discussion

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 the template for the first iteration of Woody filter may be appropriate in cases in which the duration of the target component is not known.

Simulation 4: Effect of variability of P300 latency.

The second and third iterations of the Woody filter procedure use the average waveform as the template. If the latency of the component varies from trial to trial, then the average will not provide an accurate representation of the ERP. The rationale for iteration is that a more accurate template will be obtained by aligning single trials on the basis of single trial latency estimates, and then deriving a new average. Of course, iterations will only be useful when the waveshape of the target component is not known. In Simulation 4, we investigated the effects of latency jitter and of lack of knowledge about the waveshape of the target component.

Method

The latency of P300 was varied by sampling from a distribution with a mean of 550 msec and a standard deviation of 83 msec, and the duration of P300 was varied, with three levels, 300, 500, and 700 msec. The duration of the template for cross-correlation (and first iteration of Woody filter) was fixed at 500 msec. The other parameters of this simulation are reported in Table 3.

Results

The results of this analysis are presented in Table 3.

Insert Table 3 About Here

Discussion

These results indicate that, in conditions of latency jitter, the advantage of Vector filter over single electrode selection remains, while the advantage of cross-correlation over peak-picking depends 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 that are 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.

Simulation 5: Effect of variation in frequency filter parameters.

The data presented in Simulation 1 indicates that the use of a low-pass frequency filter leads to an improvement in the estimation of P300 latency. The lower the cut-off point, the greater was the improvement. Of course, at some point, extremely low cut-offs will completely attenuate the P300 and therefore lead to poor detection. To determine which cut-off level leads to the greatest accuracy in the estimation of P300 latency, we ran another simulation with a wider range of cut-off settings.

Method

Five cut-offs points (-3 dB) were used: 6.29 Hz, 3.38 Hz, 2.32 Hz, 1.76 Hz, and 1.42 Hz. Another condition was included in which no off-line filter was used. Each of these frequency filter conditions was applied to 150 trials generated by adding different samples of EEG noise to target and overlapping components. The parameters of these components were the same as in Simulation 4 and are shown in Table 4.

Results

MSE data for the different frequency filters, for Vector filter and single electrode selection, and for peak-picking and cross-correlation, are reported in Table 4.

Insert Table 4 about here

Discussion

The results confirm that frequency filters improve the accuracy of the P300 latency estimates. This is particularly true for latencies estimated with peak-picking, while the gain for estimates obtained with cross-correlation was quite small. The most accurate estimates were obtained at a cut-off point of 1.76 Hz. Lower and higher cut-offs were associated with inferior estimates, although those obtained for adjacent cut-offs were almost as good as those for 1.76 Hz. This suggests that the setting of the low-pass filter does not need to be very precise.

Simulation 6: Effect of variation of Vector filter weights.

The weights for Vector filter used in the basic design were chosen to discriminate between the target component (P300) and various sources of noise. The scalp distribution we filtered was not that of the target component (in our case, P300). Recall that the weights of this filter were chosen on the basis of their ability to discriminate between sets of waveforms associated with rare and frequent stimuli. However, other weights could have been chosen, and some of them might have produced superior results than the set selected. To evaluate the consequences of the choice of Vector filter weights, we varied them systematically and studied their impact on the accuracy of latency estimation.

Method

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, and no frequency filtering was used. The overlapping components included N200 (amplitude = 50 units; latency = 300 ms) and Slow Wave (amplitude = 100 units; latency = 800 ms). Following the nomenclature used previously, this condition may be described as "moderate Slow Wave overlap." P300 latency was estimated using cross-correlation (the length of the template was equal to that of P300). This procedure was selected because it yielded the best results in the basic simulation.

The 47 sets of weights of Vector filter used in this simulation are presented in Table 5.

Insert Table 5 about here

Note that 45 of these sets can be organized in 9 groups of five, such that each group represents an "individual" profile, whose mean is equal to 0.00, 0.15, 0.29, 0.41, and 0.50. This variation in mean was chosen to encompass possible variations in the sensitivity of the filter to variation in polarity. The nine profiles (labelled with the letters from A to I) were selected so as to cover a variety of filters, all satisfying the criterion of having the parietal electrode weighted equally or more positively than the other electrodes. Note that the filter having the profile labelled C and the mean of the weights equal to 0.15 is that used for all the previous simulations. The last two sets of weights characterize, respectively, a filter that is equivalent to the selection of the Pz electrode, and a filter equivalent to computing the average across electrodes. The selection of the Pz electrode can also be considered equivalent to a filter having a profile E, and a mean of the weights equal to 0.33. The average of the electrodes correspond to a filter with all the weights equal to 0.58. Note that, for each set, the sum of squares of the weights is equal to 1.00. Thus, as the profiles change and as their mean value changes, the units of the output of the Vector filters do not change. Note, finally, that the "input" scalp distribution of P300 (in this study) was equivalent to a set of weights with profile G, and mean equal to 0.54. This set of weights is also shown in Table 5. Some of the profiles are shown graphically in Figure 8.

Insert Figure 8 About Here

Results

The effect of varying the weights of Vector filter was evaluated by comparing the accuracy obtained with each of the 47 sets of weights with the accuracy obtained with single electrode selection (Pz). Thus, the MSE associated with the single electrode selection (Pz) was divided by that observed for each set of weights to yield a relative performance measure for each filter. Note that a value above 1 indicates that the MSE obtained with a given set of weights is smaller than that obtained with single electrode selection (and, therefore, that the latency estimate is more accurate), while a value below 1 indicates a higher MSE. In general, the higher the relative performance measure, the better the filter. These relative performance values are shown in Table 6.

Insert Table 6 about here

Discussion

Table 6 shows that the optimal set of Vector filter weights corresponds to the profile C with a mean value of 0.15. This is in fact the set of weights for Vector filter we used in the basic design. With this setting the gain in the accuracy of latency estimates relative to single channel selection is more than 20%. Thus, the parameters we chose were optimal. Note that this set of weights was originally selected on the basis of observed data. The fact that these weights yield the best performance in our simulation supports the procedure (discriminant analysis between a set of rare and frequent trials) we used to derive this set of weights.

Although the set of weights used in the basic design yields the most accurate latency estimates, other sets of weights also produce improvement in comparison to single electrode (Pz) selection. In general, these sets share the following characteristics: (a) the weights are ordered with maximum values at Pz, followed by Fz, and by Cz; (b) the mean of the weights is positive; and (c) the weight for Cz is negative.

Note that several sets of weights yield accuracy that is worse than single channel selection. This is not surprising. In fact, these sets of weights correspond to scalp distributions which do not enhance the discrimination between signal and noise. Rather, they enhance the noise or reduce the signal.

The results of this simulation indicate that filtering for a scalp distribution that is different from that of the target component can produce even better results than filtering for the target component itself (equivalent to profile G with mean of 0.54). This is likely to occur when the scalp distribution of the target component and that of the noise overlap. It is important to note that although the average background noise value is equal to 0, it can be characterized on single trials by large positive and negative peaks. To the extent that the scalp distribution of the positive peaks is similar to that of the target component, these peaks may be misidentified as the target component. For instance, a positive noise peak may be characterized by large voltage values at Pz and Cz, and thus resemble the scalp distribution of P300. In our case, high positive values at Pz tend to co-occur with high positive values at Cz (the correlation between noise at these two electrodes is above 0.80). For this reason, a larger reduction of noise and consequently an improvement of the signal-to-noise ratio, is obtained by filtering for a scalp distribution that maximizes the discrimination between the P300 and the positive noise peaks. This is achieved when different polarity values are assigned to the weights for Pz and Cz. In general, it is advisable to use a set of weights which dissociates the activity of different electrodes with strong noise coherence. A detailed discussion of the selection of weights is given in Gratton et al. (submitted).

General Discussion

The results presented in this paper 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 most important 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 high levels of signal-to-noise ratio tend to produce the most accurate estimates at low levels. Thus, while knowledge of the signal-to-noise ratio may be critical for estimating the power of the procedure, it is irrelevant for the choice of the algorithm for latency estimation.

The use of spatial information for enhancing the detection of ERP components can be very useful, particularly when information about the scalp distribution of the target component is available. In general, Vector filter produced more accurate latency estimates than single electrode selection at Pz, or than the average across electrodes. This effect was most evident when overlapping components were present and when these components had a scalp distribution which was different from the P300 (N200). However, the choice of the weights of Vector filter is also important. These weights should be such that the discrimination between signal and noise is enhanced, rather than merely mirroring the spatial distribution of the signal. In our case, we used a discriminant function analysis between two sets of observed waveforms from rare and frequent trials to select the set of weights for P300. This procedure yields weights which dissociate the parietal and the central electrode, while still emphasizing the positive trend across electrodes. These weights fared quite well when compared to the selection of the Pz electrode, or to other sets of weights that emphasize the positive, parietal distribution of P300.

Frequency filtering produces improvement in the accuracy of latency estimation. The advantage is particularly evident when peak-picking is used as the signal detection procedure. The accuracy obtained with cross-correlation does not seem to be much improved by the use of low-pass frequency filters, possibly because this procedure, in contrast with peak-picking, is already based on several time-points. Our investigation showed that the optimal frequency filter for P300 has a cut-off at 1.76 Hz. Further filtering produces impairment of signal detection because the signal itself is degraded. Of course, the signal is also degraded with a cut-off of 1.76 Hz, but the noise is presumably degraded to a greater extent. It should be noted that our findings relate only to filters based on moving averages, and that we did not compare the effect of filters with other band-pass characteristics, or the effect of high-pass filters.

The choice of the signal detection algorithm may also affect the accuracy of latency estimates. 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 these studies 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.

Footnotes

1. Other sources of noise, such as those related to eye-movements, electromyographic activity, and electrocardiographic activity, will not be considered here.

2. Gratton et al. (submitted) described a procedure that allows the investigator to decompose the ERP observed at several electrode locations into the contribution of components charaterized by different scalp distributions. The procedure is comprised of three critical elements, (a) an obtained ERP waveform for each of several electrodes, for which we want to know the time course of the constituent components; (b) a set of characteristic weights (one for each electrode) for each component -- these weights describe the characteristic scalp distribution of the component; and (c) an output waveform (amplitude x time function) for each component that describes its timecourse in the obtained waveforms. This output waveform can be interpreted as a result of the application of a specific spatial filter that emphasizes the contribution of a component with a particular scalp distribution. The amplitudes of each component at each time point are derived by using a least squares criterion. In most of the simulation studies presented in this paper, only one set of weights is used. However, in the last simulation, the use of other sets of weights will be examined.

3. The "Woody filter" has been applied to ERPs (e.g., Kutas et al., 1977) to overcome the problem of latency jitter. The procedure was used to obtain estimates of the waveshape of the target component 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 iterations of the Woody procedure 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.

4. In this experiment, the subject was presented with one of two tones on any given trial. The tone probabilities were 0.2 and 0.8. 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 (1983). Separate averages were obtained for frequent and rare trials and for each electrode.

5. 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) that is 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 the 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, 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.

References

Coles, M.G.H., & Gratton, G. (1985). Psychophysiology and contemporary models of human information processing. In D. Papakostopoulos, S. Butler, and I. Martin, (Eds.), Clinical and Experimental Neuropsychophysiology (pp. 594-614). Beckenham, England: Croom Helm.

Coles, M.G.H., Gratton, G., Bashore, T., 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., Kramer, A.F., & Miller, G.A. (1986). Principles of signal acquisition and analysis. In M.G.H. Coles, E. Donchin, and S.W. Porges (Eds.), Psychophysiology: Systems, Processes, and Applications (pp. 183-221). New York: Guilford.

Derbyshire, A.J., Driessen, G.J., & Palmer, C.W. (1967). Technical advances in the analysis of single acoustically evoked potential. Electroencephalography and clinical Neurophysiology, 22, 467-481.

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

Donchin, E., Ritter, W., & McCallum, C. (1978). Cognitive psychophysiology: the endogenous components of the ERP. In E. Callaway, P. Tueting, and S.H. Koslow (Eds.) Event-Related brain Potentials in man. 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., & Donchin, E. (1979). The time constant in P300 recording. Psychophysiology, 16, 53-55.

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). The definition, identification, and reliability of measurement of the P300 component of the Event-related Brain Potential. In P.K. Ackles, J.R. Jennings, and M.G.H. Coles (Eds.), Advances in psychophysiology Vol. II. Greenwich, CT: JAI

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. (Submitted). A procedure for using multi-electrode information in the analysis of components of the event-related potentials: Vector filter. Psychophysiology.

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

McCarthy, G., & Donchin, E. (1978). Brain potentials associated with structural and functional visual matching. Neuropsychologia, 18, 571-585.

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 neuroelectrical signals from multiple data channels by linear filter method. Electroencephalography and clinical Neurophysiology, 38, 191-198.

Palmer, C.W., Derbyshire, A.J., & Lee, A.W. (1966). A method for analyzing individual cortical responses to auditory stimuli. Electroencephalography and clinical Neurophysiology, 29, 204-206.

Pfefferbaum, A. (1983). 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. (1984). Perceptual and motor space representation: An event-related potential study. Psychophysiology, 21, 159-170.

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). US Government Printing Office.

Ungar, P., & Basar, E. (1976). Comparison of Wiener filtering and selective averaging of evoked potentials. Electroencephalography and clinical Neurophysiology, 40, 516-520.

Wastell, D.G. (1977). Statistical detection of individual evoked responses: An evaluation of Woody's adaptive filter. Electroencephalography and clinical Neurophysiology, 42, 835-839.

Wastell, D.G. (1981). When Wiener filtering is less than optimal: An illustrative application to the brainstem evoked potential. Electroencephalography and clinical Neurophysiology, 51, 678-682.

Wiener, N. (1949). Extrapolation, interpolation and smoothing of stationary time series. New York: Wiley.

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

Table 1. Basic Design

(A) Trials per condition (100)

(B) Signal detection algorithms (4)

peak-picking
cross-correlation
Woody filter with two iterations
Woody filter with three iterations

(C) Scalp distribution filtering techniques (2)

single electrode selection (Pz)
Vector filter

(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
Slow Wave:	100, 200

(F) Latency levels of overlapping components (4)

N200:	300, 400
Slow Wave:	800, 1080

(G) Frequency filter conditions (5)

no filter
6.29 Hz (7 points, 1 iteration)
4.43 Hz (7 points, 2 iterations)
3.38 Hz (13 points, 1 iteration)
2.38 Hz (13 points, 2 iterations)

(H) Dependent Variables (2)

root mean square error (MSE)
number of trials required to obtain 3 msec error

Table 2

Parameters of ERP Components

Components	Amplitude (units)	Latency (msec)	Duration (msec)	Scalp Distribution Weights Fz	Cz	Pz
N100	100	100	100	-0.6	-1.8	-0.6
P200	150	250	150	1.0	1.0	0.5
N200	50/100	300/400	250	-1.2	-0.8	-0.4
Slow Wave	100/200	800/1080	800	-0.4	0.0	0.4
P300	0 to 500	550	500	0.4	0.8	1.2

Table 3

MSE in the case of Latency Jitter

Measures	P300 duration 300 msec	500 msec	700 msec
Peak-picking			
Pz selection	40	55	75
Vector filter	24	49	68
Cross-correlation			
Pz selection	61	47	75
Vector filter	26	35	67
Woody filter (2 iterations)			
Pz selection	56	41	74
Vector filter	20	37	66
Woody filter (3 iterations)			
Pz selection	49	42	74
Vector filter	20	43	67

Component 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

Slow Wave amplitude = 100 units Slow Wave latency = 800 msec

Table 4

MSE for different low-pass cut-offs

-3 dB Cut-offs	Peak-Picking Pz	Peak-Picking Vector filter	Cross-Correlation Pz	Cross-Correlation Vector filter
No filter	50	47	31	34
6.29 Hz	47	44	31	30
3.38 Hz	40	38	42	32
2.32 Hz	46	39	49	37
1.76 Hz	38	38	32	26
1.42 Hz	44	42	43	36

Component parameters:

P300 amplitude = 250 units (signal-to-noise ratio = 2.5)

N100 amplitude = 100 units

P200 amplitude = 150 units

N200 amplitude = 50 units

Slow Wave amplitude = 100 units

N100 latency = 100 msec

P200 latency = 250 msec

N200 latency = 300 msec

Slow Wave latency = 800 msec

Table 5. Sets of Vector filter weights used for Simulation 6.

Mean of the weights		A	B	C	D	E	F	G	H	I
Mean = 0.00	Fz	0.41	0.21	0.00	-0.21	-0.41	-0.58	-0.71	-0.79	-0.82
	Cz	-0.82	-0.79	-0.71	-0.58	-0.41	-0.21	0.00	0.21	0.41
	Pz	0.41	0.58	0.71	0.79	0.82	0.79	0.71	0.58	0.41
Mean = 0.15	Fz	0.54	0.35	0.15	-0.05	-0.24	-0.41	-0.53	-0.61	-0.64
	Cz	-0.64	-0.61	-0.53	-0.41	-0.24	-0.05	0.15	0.35	0.54
	Pz	0.54	0.71	0.83	0.91	0.94	0.91	0.83	0.71	0.54
Mean = 0.29	Fz	0.64	0.47	0.29	0.11	-0.06	-0.21	-0.32	-0.39	-0.42
	Cz	-0.42	-0.39	-0.32	-0.21	-0.06	0.11	0.29	0.47	0.64
	Pz	0.64	0.79	0.90	0.97	1.00	0.97	0.90	0.79	0.64
Mean = 0.41	Fz	0.70	0.56	0.41	0.26	0.12	0.00	-0.09	-0.15	-0.17
	Cz	-0.17	-0.15	-0.09	0.00	0.12	0.26	0.41	0.56	0.70
	Pz	0.70	0.82	0.91	0.97	0.99	0.97	0.91	0.82	0.70
Mean = 0.50	Fz	0.70	0.61	0.50	0.39	0.30	0.21	0.15	0.11	0.09
	Cz	0.09	0.11	0.15	0.21	0.30	0.39	0.50	0.61	0.70
	Pz	0.70	0.79	0.85	0.89	0.91	0.89	0.85	0.79	0.70

Selection of Pz electrode*:	Fz	0.00
	Cz	0.00
	Pz	1.00
Average across electrodes**:	Fz	0.58
	Cz	0.58
	Pz	0.58
Scalp distribution of input P300***:	Fz	0.27
	Cz	0.54
	Pz	0.80

* This is equivalent to profile E with a mean of 0.33.

** The weights within each set were chosen with the constraint that the sum of their squared values had to be equal to 1.

*** This is equivalent to profile G with a mean of 0.54.

Table 6. Ratio between MSE for single channel selection (Pz) and the MSE for each of the Vector filters.

Mean	Profiles								
	A	B	C	D	E	F	G	H	I
0.00	0.18	0.34	0.38	0.61	0.43	0.51	0.38	0.45	0.48
0.15	0.82	1.13	1.22	1.13	1.08	0.90	0.82	0.79	0.76
0.29	0.93	1.03	1.11	1.07	1.00	1.00	0.89	0.91	0.84
0.41	0.58	1.04	1.00	1.04	1.00	0.92	0.96	0.93	0.96
0.50	0.70	0.93	1.00	1.05	0.97	1.00	0.90	0.93	0.90
Selection of Pz electrode*:					1.00				
Average across electrodes:					0.67				

* This is equivalent to profile E with a mean value of 0.33

Figure Legends

Figure 1. Examples of simulated single-trial waveforms. Waveforms without EEG noise are shown in the upper panel, waveforms with EEG noise are displayed in the lower panel. The amplitude is expressed in arbitrary units (simulated P300 amplitude = 250 units).

Figure 2. 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 3. Histograms of latency estimates for four different signal-to-noise ratios for two detection algorithms and two spatial filtering techniques. The vertical lines indicate the latency of the simulated P300.

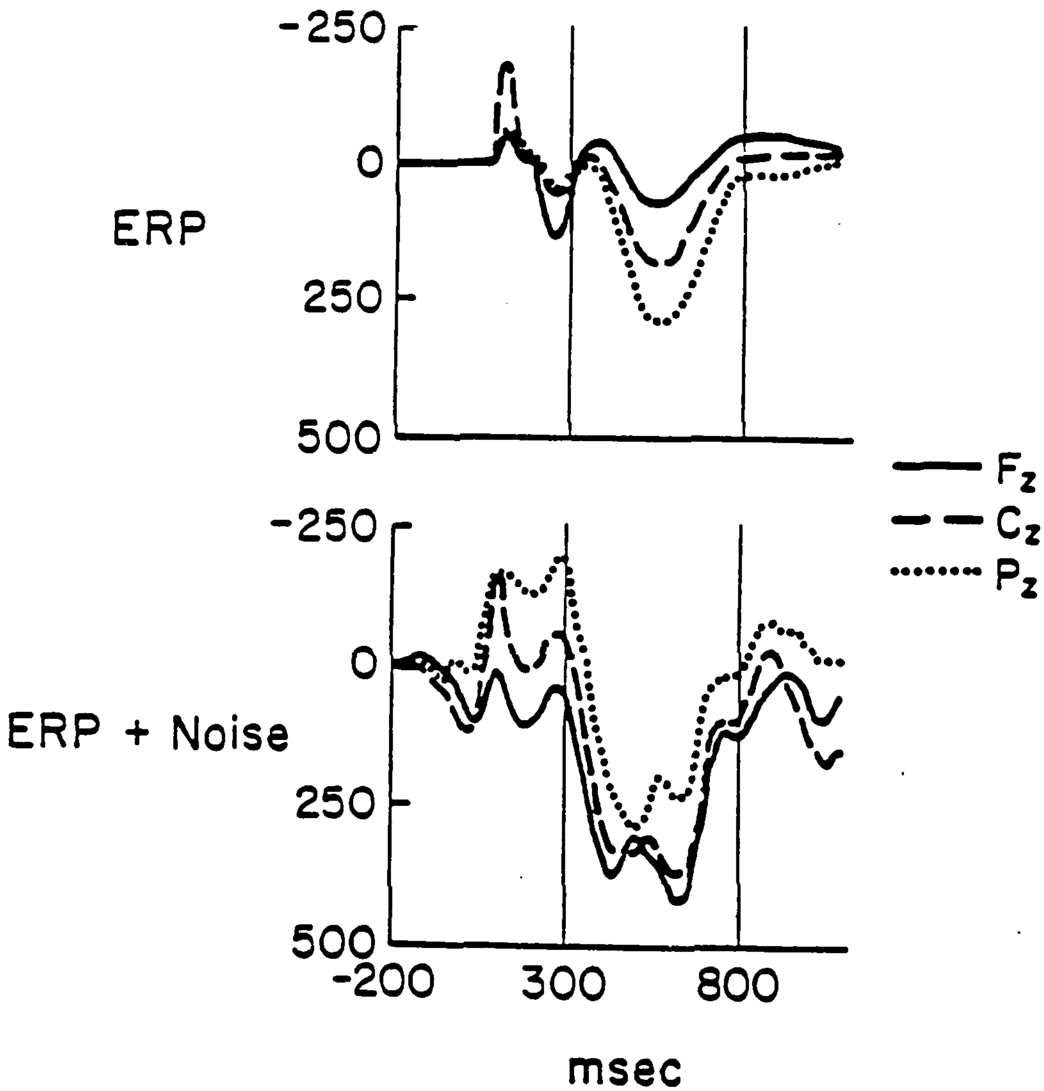
Figure 4. Latency-adjusted average waveforms over 100 trials. P300 peak latency was computed with cross-correlation (upper panels), Woody 2-iterations (middle panels), Woody 3-iterations (lower panels). The left column shows waveforms obtained with a signal-to-noise ratio of 0 (no P300 was present), the right column shows 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 5. 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. The results for two detection algorithms and two spatial filtering techniques are shown for each condition.

Figure 6. 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. The results for two detection algorithms and two spatial filtering techniques are shown.

Figure 7. Effect of P300 duration on the accuracy of latency estimation with peak-picking, cross-correlation and Woody filter with two and three iterations.

Figure 8. Schematic representation of some of the sets of the Vector filter weights (profiles) used for Simulation 6.



-250
0
ERP
250
500
-250
0
ERP + Noise
250
500
-200
300
800
msec
F z
C z
P z

Fig. 1

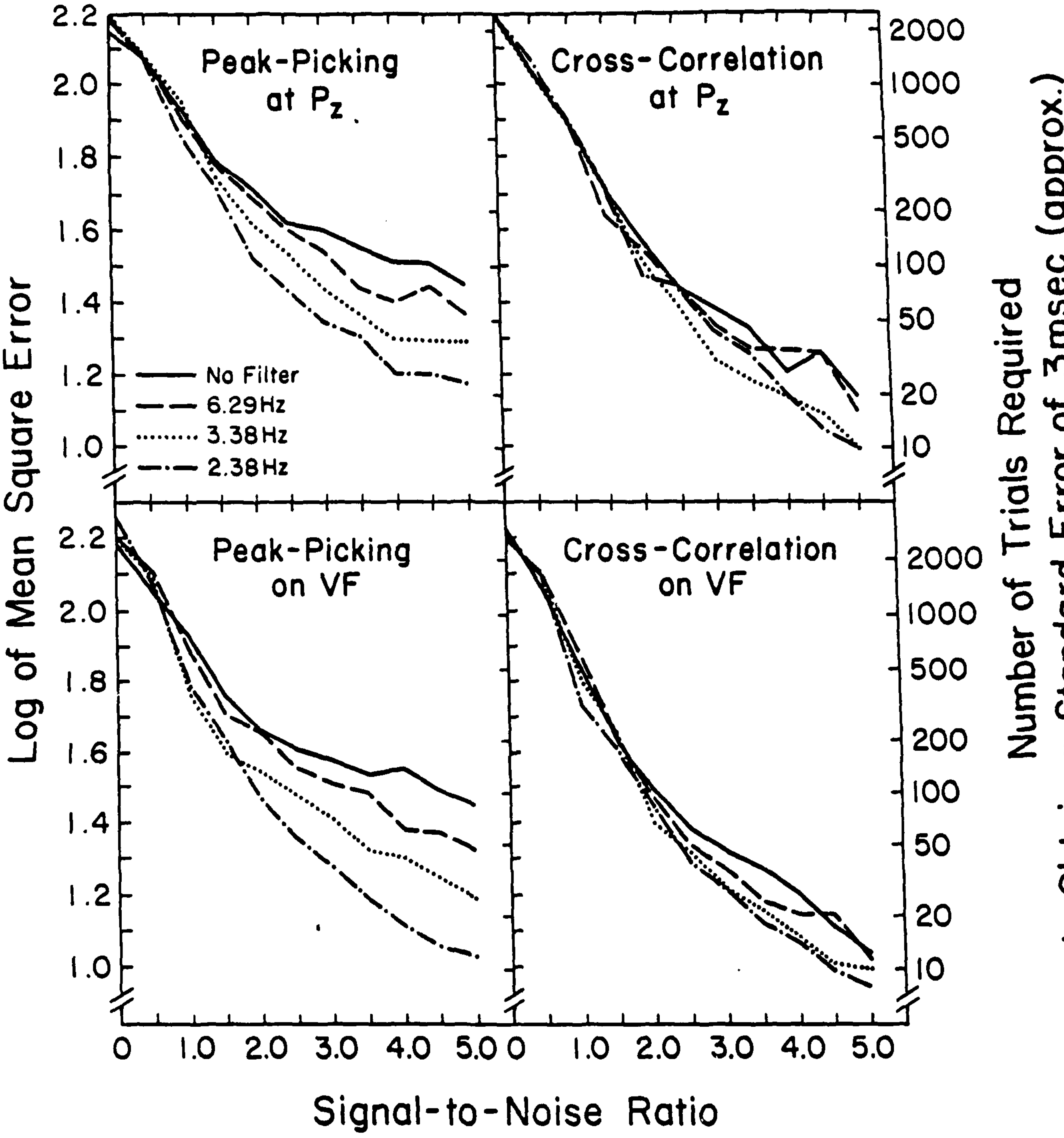

Peak-Picking at P_z
Cross-Correlation at P_z
Peak-Picking on VF
Cross-Correlation on VF
No Filter
6.29Hz
3.38Hz
2.38Hz
Log of Mean Square Error
Number of Trials Required to Obtain a Standard Error of 3msec (approx.)
Signal-to-Noise Ratio
2.2
2.0
1.8
1.6
1.4
1.2
1.0
2000
1000
500
200
100
50
20
10
0
1.0
2.0
3.0
4.0
5.0

Fig. 2

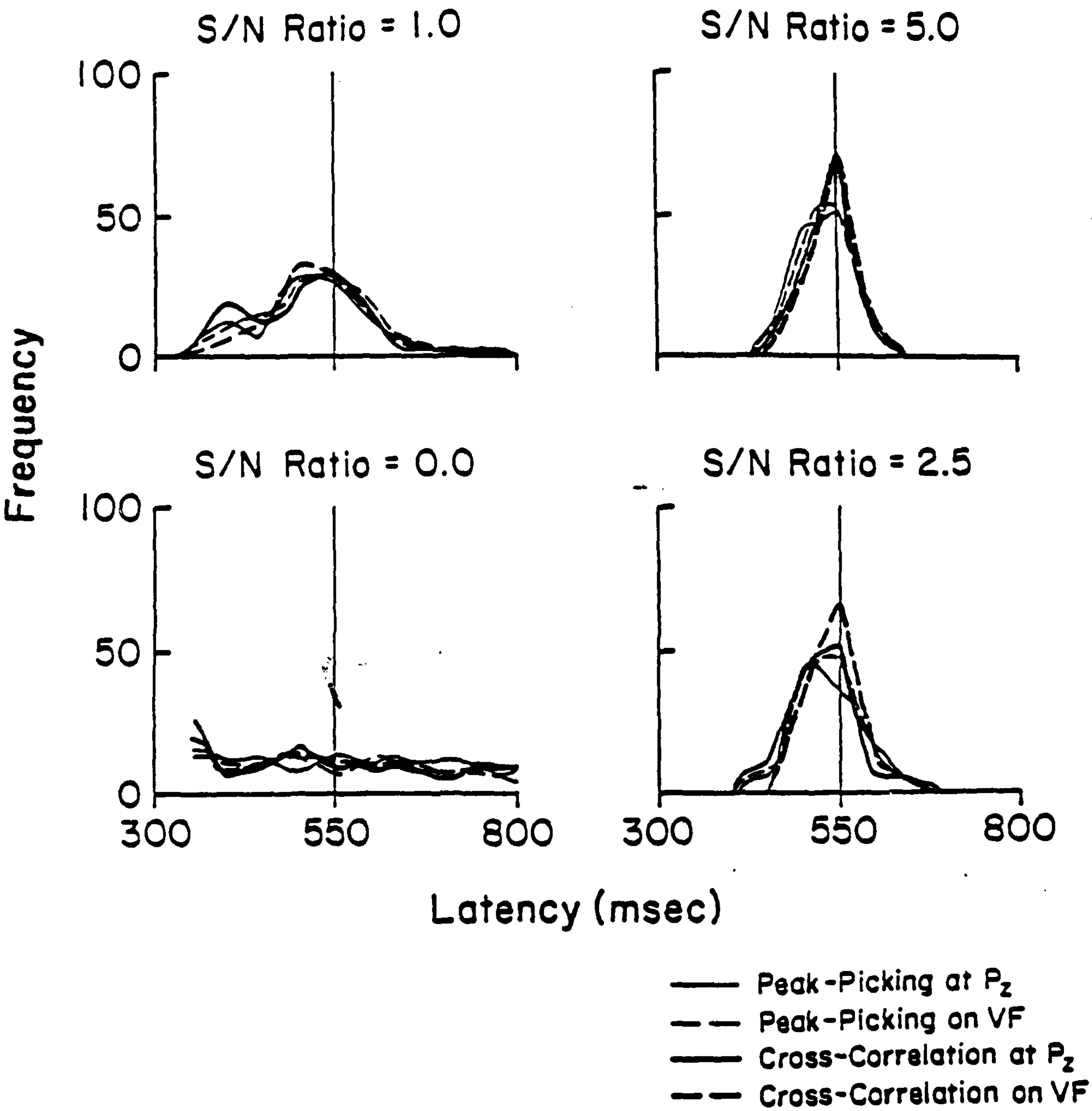
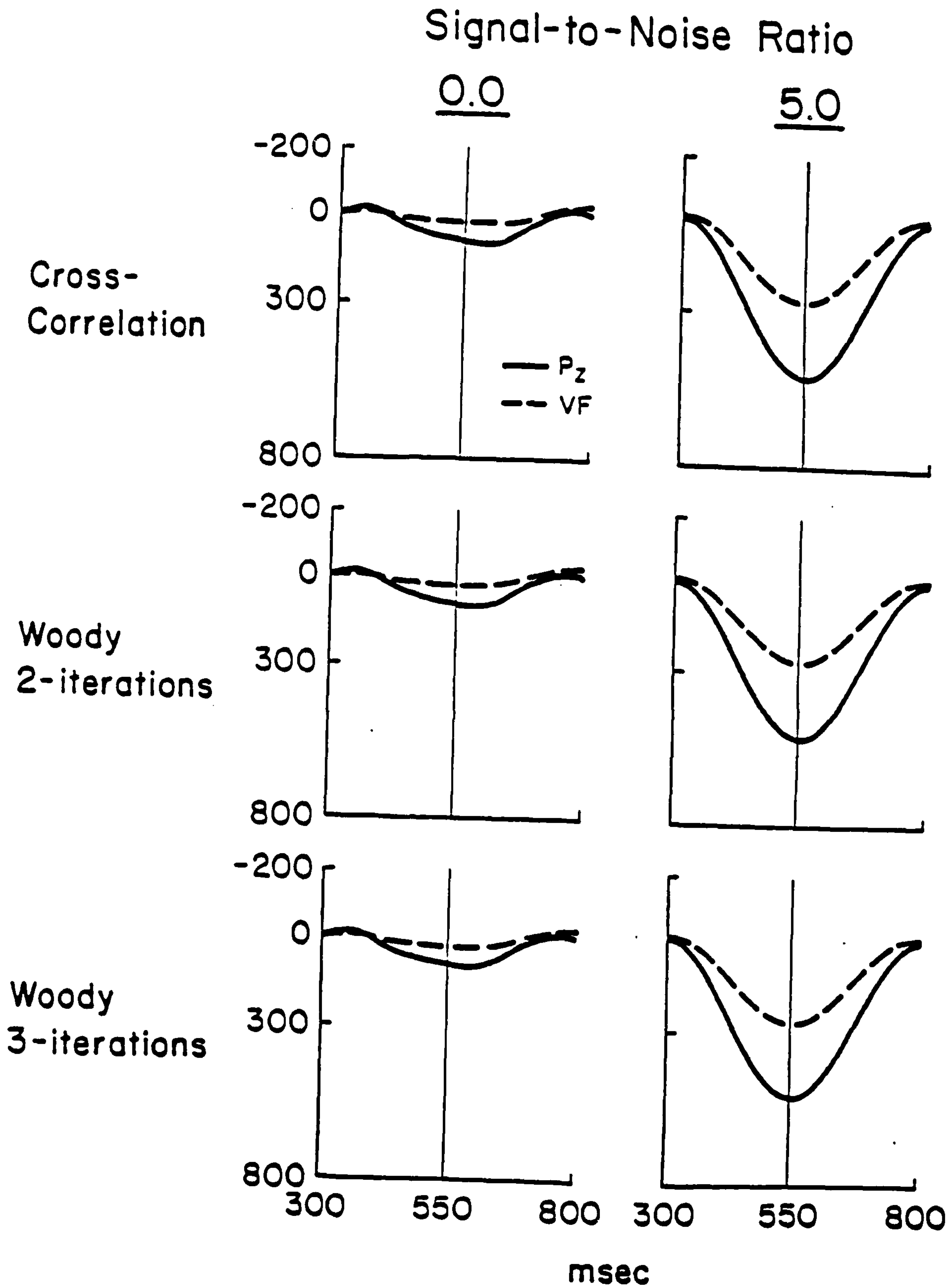
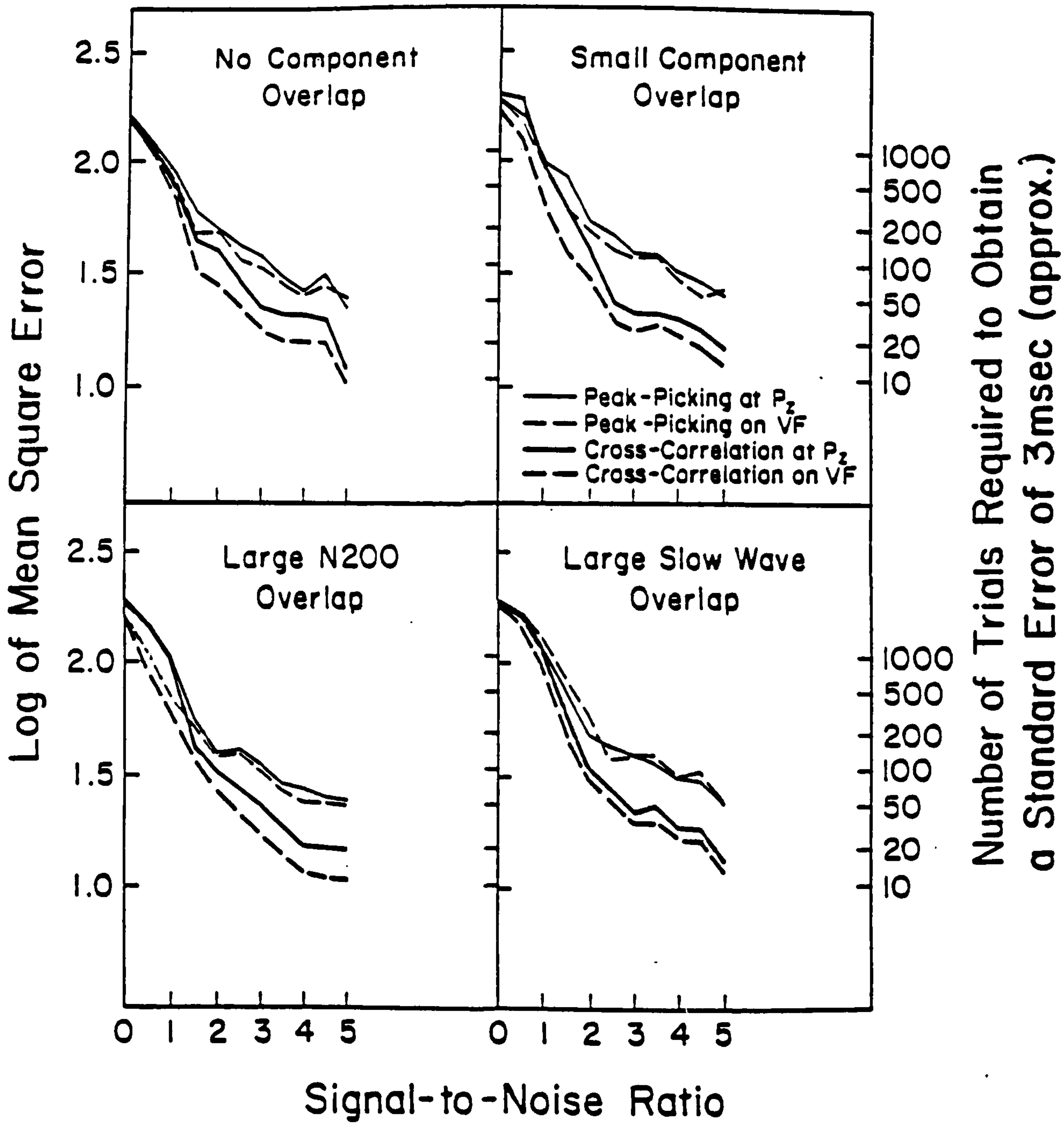

S/N Ratio = 1.0
S/N Ratio = 5.0
S/N Ratio = 0.0
S/N Ratio = 2.5
Frequency
100
50
0
300
550
800
Latency (msec)
Peak-Picking at P_z
Peak-Picking on VF
Cross-Correlation at P_z
Cross-Correlation on VF

Fig. 3



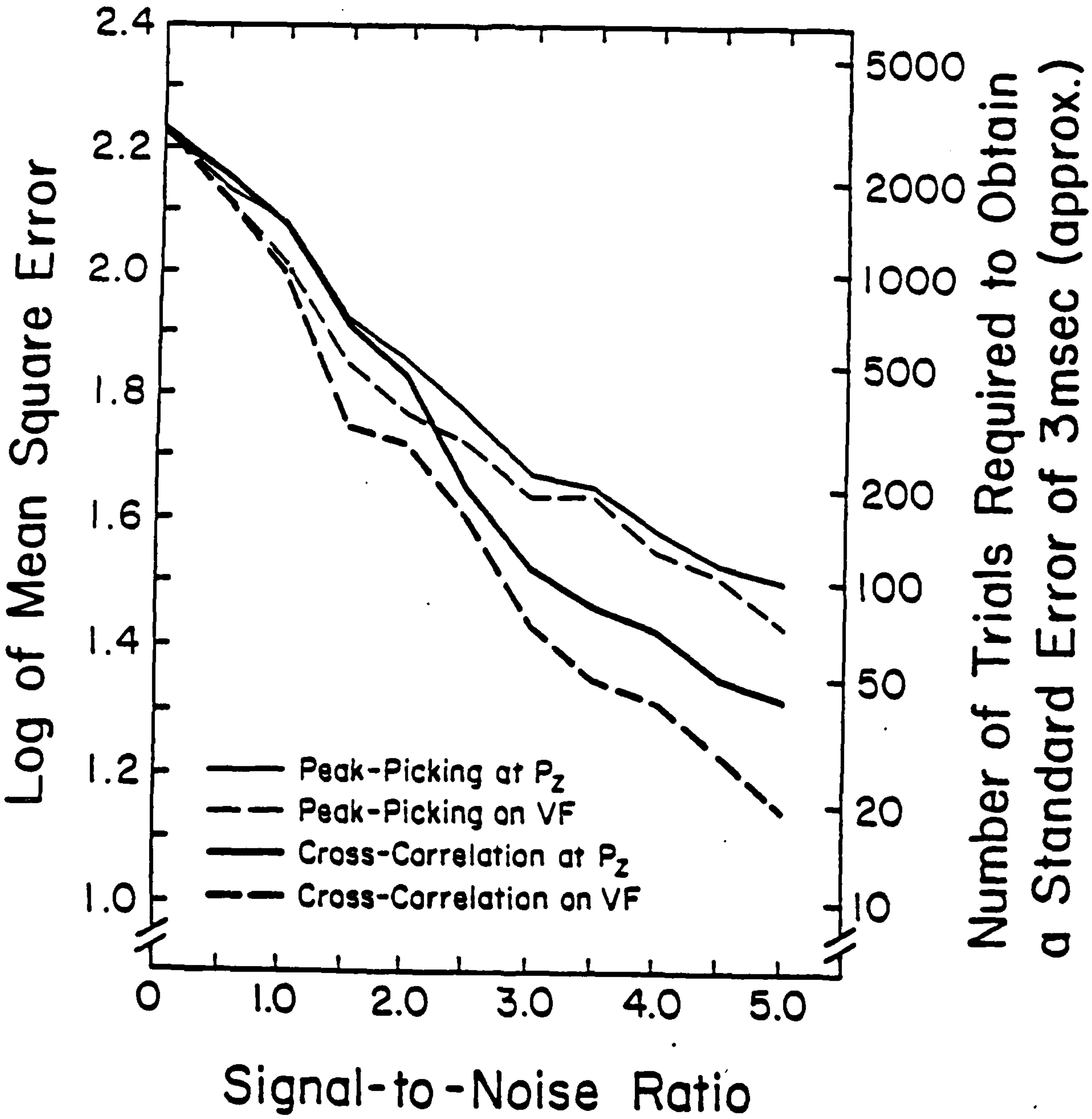
Signal-to-Noise Ratio
0.0
5.0
Cross-
Correlation
Woody
2-iterations
Woody
3-iterations
-200
0
300
800
P_z
VF
300
550
800
msec

Fig. 4



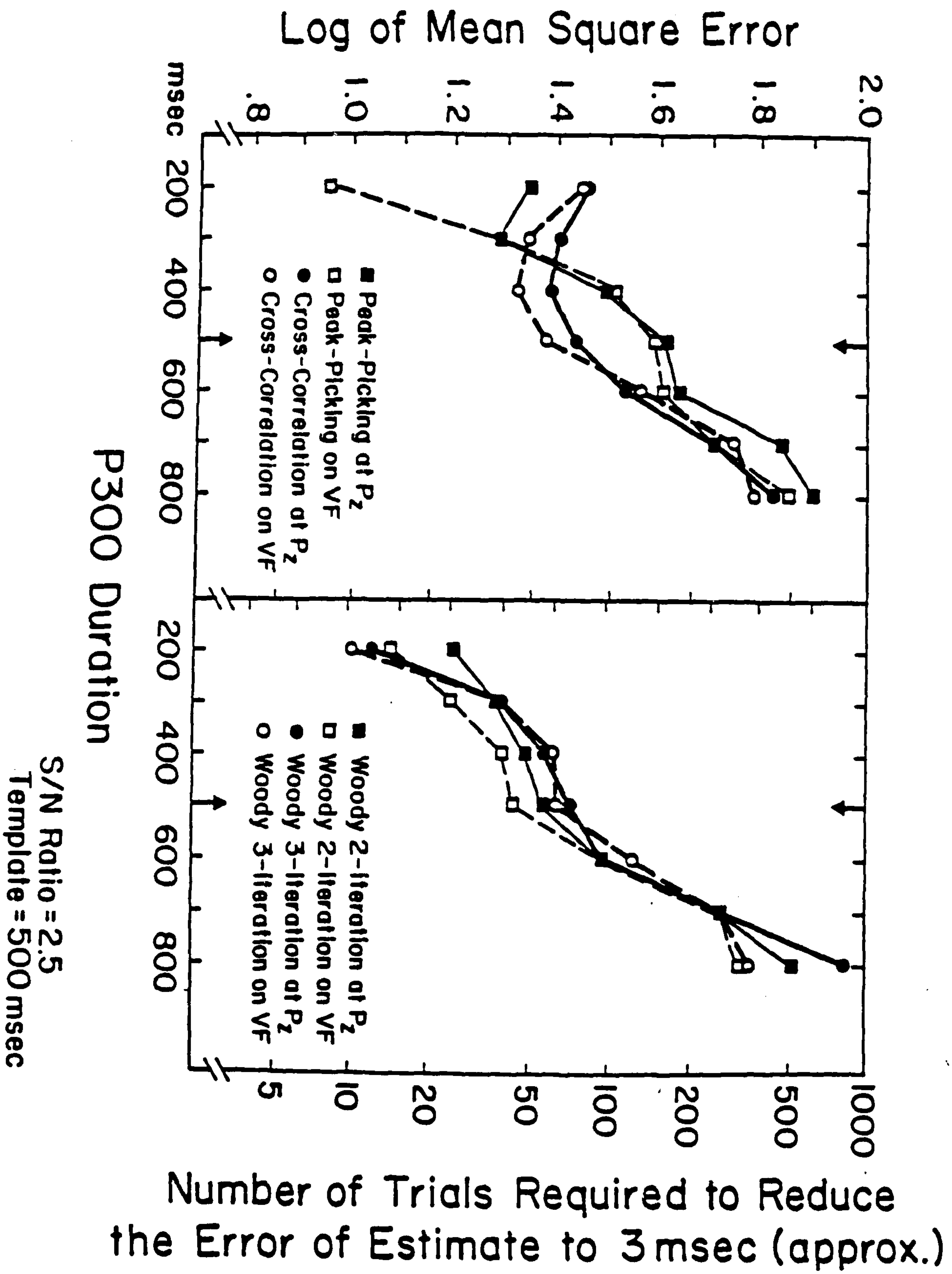
No Component Overlap
Small Component Overlap
Large N200 Overlap
Large Slow Wave Overlap
Log of Mean Square Error
2.5
2.0
1.5
1.0
Number of Trials Required to Obtain a Standard Error of 3msec (approx.)
1000
500
200
100
50
20
10
Peak-Picking at Pz
Peak-Picking on VF
Cross-Correlation at Pz
Cross-Correlation on VF
0 1 2 3 4 5
Signal-to-Noise Ratio

Fig. 5



Log of Mean Square Error
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0
1.0
2.0
3.0
4.0
5.0
Signal-to-Noise Ratio
Number of Trials Required to Obtain a Standard Error of 3msec (approx.)
5000
2000
1000
500
200
100
50
20
10
Peak-Picking at Pz
Peak-Picking on VF
Cross-Correlation at Pz
Cross-Correlation on VF

Fig. 6



Log of Mean Square Error
2.0
1.8
1.6
1.4
1.2
1.0
.8
msec 200 400 600 800
200 400 600 800
■ Peak-Picking at P_z
□ Peak-Picking on VF
● Cross-Correlation at P_z
○ Cross-Correlation on VF
■ Woody 2-iteration at P_z
□ Woody 2-iteration on VF
● Woody 3-iteration at P_z
○ Woody 3-iteration on VF
P300 Duration
S/N Ratio = 2.5
Template = 500 msec
1000
500
200
100
50
20
10
5
Number of Trials Required to Reduce the Error of Estimate to 3 msec (approx.)

Fig. 7

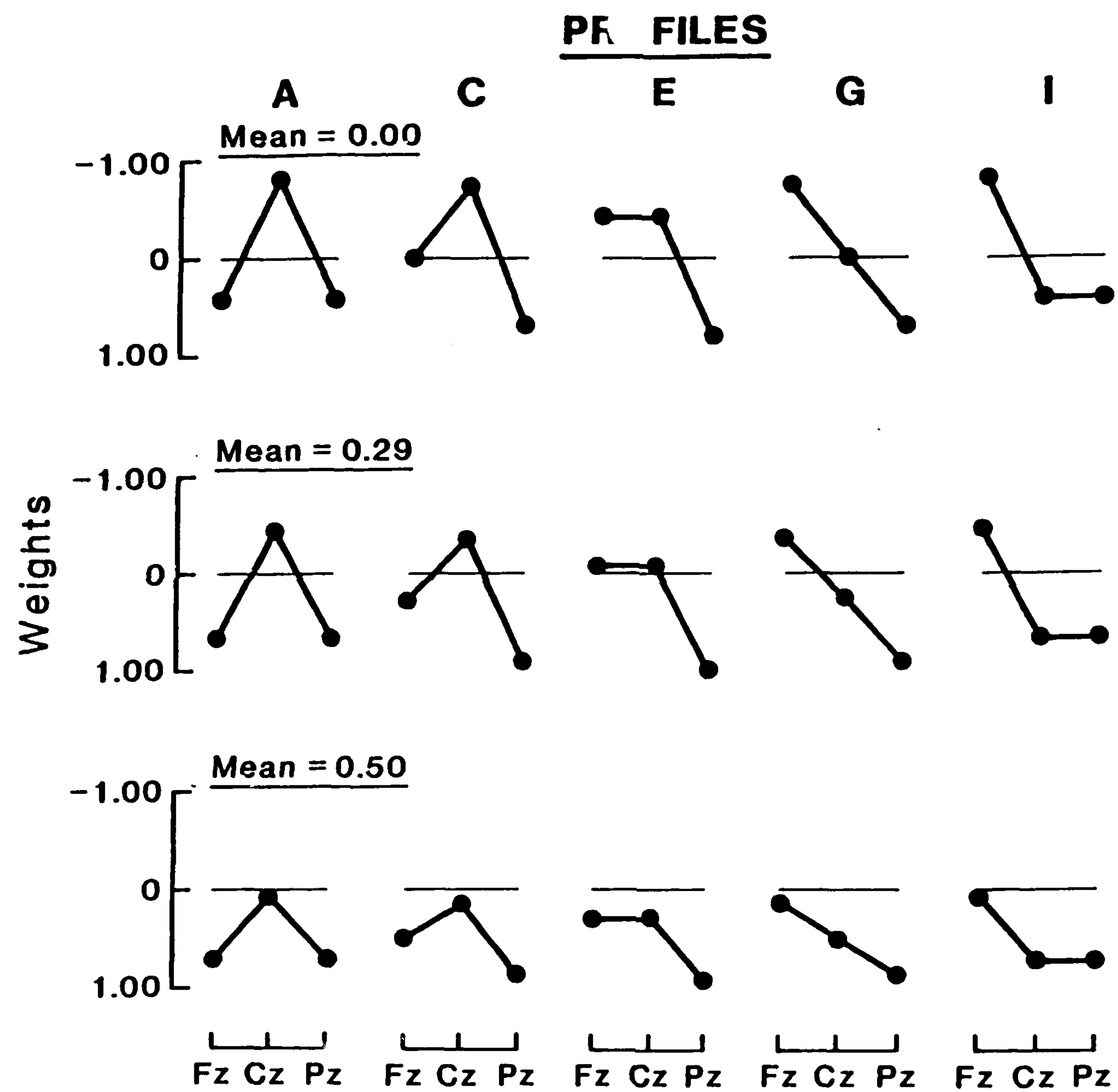

PR FILES
A
C
E
G
I
Mean = 0.00
Mean = 0.29
Mean = 0.50
-1.00
0
1.00
Weights
Fz Cz Pz

Fig. 8

Resource Reciprocity: An Event-Related Brain Potentials Analysis

Erik J. Sirevaag, Arthur F. 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 in which 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 (the primary task) performed alone and with a concurrent auditory discrimination task (the secondary task). Primary 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. Increased primary task difficulty, reflected in increased RMS error scores, was associated with decreased secondary task P300 amplitudes and increased primary task P300 amplitudes. The increases in primary task P300 amplitudes were complementary to the decrements obtained for the secondary task, supporting the hypothesis of reciprocity between primary and secondary task P300 amplitudes across several different manipulations of primary task difficulty.

KEYWORDS: EVENT-RELATED BRAIN POTENTIALS, P300, RESOURCES, MENTAL WORKLOAD, DUAL-TASK PERFORMANCE

Introduction

This study is concerned with examining the assumptions which underlie the use of the P300 component of the Event-Related Brain Potential (ERP) as a measure of mental workload. The ERP represents electroencephalographic (EEG) activity which is time-locked to an event (Donchin et al. 1978). By averaging over records which follow repetitions of the event, the contribution to the average ERP of background activity unrelated to the processing of the event (noise) diminishes while the contribution of the time-locked activity (signal) 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 et al. 1965).

The proposal that P300 amplitude can serve a role in the measurement of mental workload derived from observations that the amplitude of the P300 is severely reduced when a subject's attention is directed away from the task in which the eliciting stimuli are embedded (Johnson and Donchin 1978, 1982; Duncan-Johnson and Donchin 1977, 1978). An important question, however, was whether P300 amplitude would reflect graded changes in attention that are characteristics of different levels of workload.

A systematic relationship between P300 amplitude and mental workload was demonstrated in a series of studies that employed dual-task techniques (Brown 1978; Ogden et al. 1979). In these studies, subjects performed two tasks concurrently (see Donchin et al. 1986 for an in depth review). One task was designated as primary, the other as secondary. Primary tasks included both system monitoring and manual control. The ERP eliciting secondary tasks involved either visual or auditory discriminations (Isreal et al. 1980a; Isreal et al. 1980b; Natani and Gomer, 1981; Kramer et al.

1983). Results of these studies indicated that the amplitudes of P300s elicited by secondary task stimuli decreased as the demands placed on subjects by the primary tasks increased. Furthermore, this relationship between P300 amplitude and task demands occured only when these demands were perceptual/cognitive in nature. Motor demands did not systematically affect the amplitude of the P300 component.

The results of these studies were interpreted within the framework of Multiple Resource theory (Kinsbourne and Hicks 1978; Navon and Gopher 1979; Sanders 1979; Wickens 1980; Friedman and Polson 1981). These models propose that dual task performance can be conceptualized in terms of a number of different resources that are limited in quantity. Tasks that require the same types of resources will be timeshared more poorly than tasks that require different resources. The allocation of this limited processing commodity to the performance of a given task is determined by the motivation and skill of the operator and the demands of the task. Workload, therefore, is viewed as a hypothetical construct reflecting the interaction between task demands and operator attributes (Gopher and Donchin 1986).

When an operator is in a demanding multi-task situation, and the tasks require the same resources, tradeoffs in performance will be observed. If the operator is instructed to optimize performance on one task, fewer resources will be available for the other tasks. Furthermore, increases in primary task difficulty entail the allocation of a larger share of the resources to the primary task. Inevitably, these resources are no longer available to the secondary tasks and performance on the secondary tasks deteriorates (Navon and Gopher 1979; Norman and Bobrow 1975).

While the results of the P300 studies are consistent with the Multiple Resource model, there was an important implication that had not been adequately examined. 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 and secondary task events, there should be a reciprocal relationship between primary and secondary task P300 amplitudes.

The amplitude reciprocity hypothesis was explicitly tested by Wickens et àl. (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 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 indicated that P300s associated with the step changes increased in amplitude with increased primary task difficulty; while secondary task P300 amplitudes decreased in a complementary manner.

While these data did confirm the basic prediction of the resource reciprocity hypothesis, we considered it necessary to test this hypothesis with ERPs that were elicited by primary and secondary task events recorded within the same block of trials. In other words, the case for amplitude reciprocity would be stronger if a reciprocal relationship between concurrently recorded primary and secondary task ERPs is found. Furthermore, no study to date has examined the relationship between task demands and P300 amplitude at the level of the single subject. We report here a study that tests the P300 amplitude resource-reciprocity hypothesis using concurrently recorded primary and secondary task ERPs both across tasks and within single subjects.

Previous research has indicated that while P300 amplitude is sensitive to increases in the system order of a tracking task (Kramer et al. 1983), manipulations of the number of dimensions in which a subject is required to track produce no changes in secondary task P300 amplitude (Wickens et al. 1977). Given that the P300 is sensitive to perceptual/cognitive demands and is relatively unaffected by motor demands (Isreal et al. 1980a), these results imply that the system order manipulation places a heavy load on perceptual/cognitive processing while the dimensionality manipulation loaded primarily on response processing. Therefore, an orthogonal manipulation of dimensionality and system order should provide conditions with varying degrees of primary and secondary task resource competition.

In the present study, a step tracking task was developed in which subjects performed 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 combination of independent variables allows for the replication of past findings and at the same time introduces a previously unexplored condition (two dimensional tracking in a second order system). Because the present study contains four levels of the primary task, each of which should place different demands upon the operator, it provides a unique opportunity to examine the sensitivity of both primary and secondary task P300 amplitudes to graded changes in workload. Thus, the present study examines whether the reciprocity of P300 amplitude can be demonstrated both across tasks and within subjects in a study involving orthogonal combinations of dependent variables and concurrently recorded primary and secondary task ERPs.

Method

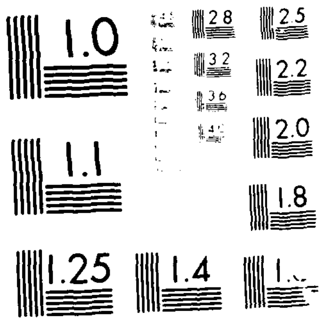
Subjects Forty dextral males between the ages of 18 and 25 were paid for their participation in this study. None of the subjects had any previous experience with the step tracking task. All subjects had normal hearing and normal or corrected to normal vision.

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 occurred 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:

AD-A191 244 THE EVENT-RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING A..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 9/10

UNCLASSIFIED 29 FEB 88 CPL-88-1 AFOSR-TR-88-0316 F/G 5/8 NL


1.0
2.8
2.5
3.2
2.2
3.6
2.0
1.1
1.8
1.25
1.4

$$X(T)=[(1-A)\int U(T)\,dt]+[(A)\iint U(T)\,dt]$$

where U=stick position
T=time
A=contribution of the second order system

The value of "A" was varied to create velocity and acceleration conditions. When A=0, the system was a pure first-order system in which movements of the stick increased or decreased the velocity with which the cursor moved. This will be referred to as the velocity condition. When A=95, the system included a second-order component. That is, the joystick controlled the acceleration of the cursor. In this "acceleration" condition, it was 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. On any particular trial a high (1400 Hz) or low (1200 Hz) pitch tone could occur with an equal probability. The subjects task was to count covertly the total number of one category of the tones. 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 for the tones (1280 msec) and the step changes (1280 msec) 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 Beckman Biopotential electrodes were attached to the forehead and used as grounds. The scalp and mastoid electrodes were Burden Ag-AgCl electrodes affixed witn collodion. The vertical electrooculogram (EOG) was recorded from Beckman electrodes affixed with adhesive collars above and below the subject's right eye. All electrode impedances were below 10 kohms.

The EEG and EOG were amplified by Van Gogh model 50000 amplifiers with a 10 sec time constant and a low-pass filter of 35 Hz, 3 db/octave rolloff. The recording epoch for both the EEG and EOG was 1280 msec beginning 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, 0 db at 14.27 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 the eye-movement correction algorithm developed by Gratton et al. (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 dimensionality 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 eight experimental conditions. Each condition lasted approximately 5 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 task difficulty on single and dual-task performance. Before receiving practice in the step tracking task, all subjects performed three blocks of the auditory discrimination task to familiarize them with the stimuli and procedure. 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.

This study was part of a much larger research effort that examined individual differences related to the performance of perceptual/motor tasks. For this reason, it was necessary to insure that subjects were treated identically. This required us to run the subjects through the experimental conditions in a fixed order rather than in a counterbalanced fashion. The order of the experimental blocks is presented in table 1.

<Insert table 1 about here>

Note that single task blocks always preceded dual-task blocks, and easier tracking conditions preceded difficult tracking conditions. While this order may have diminished differences between single and dual-task performance as a result of practice, this is not the comparison of primary concern in this study. With respect to the comparison that is of interest (ie between dual-task blocks differing in dimensionality and system order) any learning due to practice effects should improve performance during the more difficult conditions, rather than enhance performance decrements due to increased primary task difficulty. Thus, this conservative design will tend to diminish differences between between easy and difficult tracking conditions due to practice effects rather than amplify them.

Another potential confound introduced by the fixed presentation order is a possible fatigue effect. Because the difficult conditions were presented later in the session, any performance decrements could be the result of increased fatigue. We consider this explanation unlikely given that the experimental session rarely lasted for more than one hour and subjects were given the opportunity to take breaks if so desired. Furthermore, it is difficult to see how fatigue could artifactually produce reciprocity. Although one could argue that secondary task P300 amplitude decrements were due to fatigue, this does not account for concomitant increases in the amplitude of primary task P300s.

Tracking accuracy data were collected 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. Accuracy in the auditory discrimination task was assessed by comparing the subject's count of target tones with the actual number of tones presented.

Results and Discussion

The primary and secondary task 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. The RMS error measures will be used to define the difficulty level of the experimental conditions. 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. The RMS data will also be examined to determine whether subjects did, in fact, protect the level of performance on the primary task even as the task demands increased. This is an important observation for secondary task performance decrements are difficult to intrepret if subjects do not perform the primary task comparably during both single and dual-task conditions.

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 secondary task events. Recall that with increases in the difficulty of the primary tracking task we predicted that (a) increases in primary task P300s would be observed, and (b) decreases in secondary task P300 amplitudes would be obtained. Thus, the P300s associated with the primary task step changes and the secondary task tones will be evaluated to test this prediction. The reciprocity of primary and secondary task P300 amplitudes will be analyzed both across and within individual subjects.

Performance data The average root mean square (RMS) error for an experimental condition reflects the average distance between the cursor and the target square. Low values of the RMS error scores, therefore, reflect increased tracking accuracy. To facilitate the comparison of performance and ERP data, the RMS data were range-corrected according to the following transformation:

$$X(T) = 100 * \frac{X(I) - X(MIN)}{X(RNG)}$$

where, X(T)= transformed score;
X(I)= score obtained in a given condition;
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. The same pattern of results was also obtained with the uncorrected scores indicating that our results were not artifactually produced by the range-correction procedure. Indeed, differences between single and dual-task RMS error scores involved less than two percent of the total variance for both the corrected and uncorrected measures.

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 [$F(1,39)=321.81$; $p<.01$], and as the control order was increased from a velocity to an acceleration system [$F(1,39)=2246.03$; $p<.01$].

The effect of system order was consistently larger than the effect of dimensionality. This is illustrated by the significant interaction between system order and dimensionality [$F(1,39)=132.17$; $p<.01$]. Tukey tests (Tukey 1977) performed on pairwise comparisons indicate that order significantly affected performance in both one [$F(1,39)=45.30$; $p<.01$] and two dimensions [$F(1,39)=61.40$; $p<.01$]. Similarly, the effect of dimensionality was significant for both velocity [$F(1,39)=202.21$; $p<.01$], and acceleration [$F(1,39)=4196.28$; $p<.01$] 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 [$F(1,39)=9.32$; $p<.004$], the magnitude of this effect was quite 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 in 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 data confirm that subjects protected their performance on the primary task, for there was no significant increase in RMS 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 the different secondary task conditions.

The grand average target ERPs for each of the dual-task conditions are displayed in fig. 1. As predicted, the one-dimensional velocity condition was associated with the largest secondary task P300, and the smallest secondary task P300 was elicited during the most difficult two-dimensional acceleration condition.

<Insert fig. 1 about here>

A Principle 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 et al. 1986, 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 fig. 2.

<Insert fig. 2 about here>

It has been suggested (Donchin et al. 1978; Donchin et al. 1986) 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 P300 component (it has the correct latency and scalp distribution and displays a significant target effect [$\underline{F}(1,39)=8.84$; $p<.005$].

Having established component 5 as representative of the P300 we will now examine the effects of the experimental manipulations of dimensionality and system order 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 acceleration tracking conditions than when the tracking task required velocity control [$F(1,39)=21.49$; $p<.0001$).

The component scores for this factor at Cz were used to provide numerical estimates of P300 amplitude for each individual 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 [$F(1,39)=13.15$; $p<.0008$] and system order [$F(1,39)=18.27$; $p<.0001$] significantly reduced secondary task P300 amplitude with no sign⁻ricant interaction [$F(1,39)=0.42$; $p=0.52$]. Given the RMS error results, 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 (Isreal et al. 1980a; Kramer et al. 1983; Natani and Gomer, 1981; Kramer et al. 1987).

Primary Task ERPs The ERPs elicited by the step changes in the various dual-task tracking conditions are displayed in fig. 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 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 also 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.

Fig. 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 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 fig. 5.

<Insert fig. 5 about here>

Component 1 is active in the appropriate latency range, and with the correct 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 than Fz or Pz [F(2,78)= 7.13; p<.01; and F(2,78)=28.13; p<.01, respectively].

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 various 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. An 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 x 3 Electrodes), was submitted to a PCA in which 5 factors were extracted and Varimax rotated. The component structure associated with this PCA is presented in fig. 6.

<Insert fig. 6 about here>

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 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 much 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 the P300 for the auditory task. The procedure described above was applied to obtain numerical estimates of P300 amplitude. Thus, the relevant 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 [$F(1,39)=35.58$; $p<.0001$]. This effect did not interact with the dimensionality manipulation [$F(1,39)=1.56$; $p>.20$]. The overall effect of increasing the number of dimensions to be tracked did not reach statistical significance [$F(1,39)=3.88$; $p=.056$]. Analysis of the simple main effects indicated that, although P300 amplitude did not increase as a function of increased dimensionality for velocity systems [$F(1,39)=0.38$; $p>.50$], it did increase as dimensionality increased for acceleration systems [$F(1,39)=5.52$; $p<.025$]. 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 order interaction was not significant [$F(1,39)=0.33$; $p=.57$]. 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 An important goal of this study was to assess the degree to which P300 amplitudes associated with the two tasks would be reciprocal. In fig. 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.

<Insert fig. 7 about here>

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 fig. 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 within the subjects (indicating the presence of instances of both under and over reciprocity) the obtained value of 0.04 for the mean slope did not differ significantly from the predicted value of 0 ($t=0.11, p>0.10$); and the mean intercept value of 92.99 did not differ significantly from the predicted value of 100 ($t=1.57$, $p>0.05$). An additional within subjects analysis was also performed to test the reciprocity hypothesis. When secondary task P300 amplitude was plotted as a function of primary task P300 amplitude, a negative slope for this function was obtained for 29 out of the 40 subjects. For the 11 subjects with positive slopes, the value of the slope was generally close to 0 indicating that their P300s did not vary systematically as a function of the experimental manipulations. Thus, evidence in support of the reciprocity theory was obtained both across and within subjects.

<Insert table 5 about here>

Conclusions

This experiment confirms the existence of a reciprocal relationship between the amplitudes of the P300s associated with two concurrently performed tasks. The prediction of reciprocity derives from a large body of evidence which has indicated that variations in P300 amplitude are sensitive to the manner in which subjects allocate processing resources between two tasks under dual-task conditions. In other words, P300 amplitude has emerged as a psychophysiological metric of the resource tradeoffs that are presumed to underlie the concept of mental workload (Kahneman 1973; Navon and Gopher 1979; Wickens 1980, 1984; Gopher and Donchin 1986).

The RMS error data confirm that the orthogonal manipulation of system order and dimensionality employed in this study successfully produced a wide variability in performance within which to assess the reciprocity of primary and secondary task P300 amplitudes. Because the difficulty of the secondary task was held constant during all the step-tracking conditions, the model of resource reciprocity upon which this experiment is based predicts that as the 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 complexity 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 fig. 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 unobtrusive fashion in many different situations because 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 and there are no structural effects to impede the performance of the primary task. The RMS error data from this experiment confirm that, indeed, a secondary auditory discrimination task can be imposed in a dual-task setting with minimal 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.

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 as opposed to response-related processing (Isreal et al. 1980a; Isreal et al. 1980b). 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 to moment basis.

In addition to validating the prediction of P300 amplitude reciprocity, this experiment produced a number of other important results. 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 by 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 the discrepant results. 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 employed. Thus, although this manipulation produced a significant change in workload, the magnitude of this change was quite small. It is conceivable that the increased power resulting from an experiment involving 40 subjects allowed us to detect this small effect where the earlier study by Wickens et al. failed. Furthermore, this study provides the first evidence that an increase in dimenionality entails a significant drain on perceptual/cognitive resources during acceleration tracking (a previously unexplored condition).

In conclusion, using concurrently recorded ERPs associated with primary and secondary tasks, this experiment demonstrates 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 interpreted 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

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., 1986. 'Principles of signal acquisition and analysis'. In: M.G.H. Coles, E. Donchin, and S. Porges (eds.), Psychophysiology: Systems, Processes, and Applications. New York: Guilford Press, pp. 183-221.

Donchin, E. & Heffley, E., 1975. Minicomputers in the signal averaging laboratory. American Psychologist, 30, 299-312.

Donchin, E. & Herning, R.I., 1975. A simulations 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., and Wickens, C., 1986. 'Applications of brain event-related potentials to problems in engineering psychology'. In: M.G.H. Coles, E. Donchin, and S. Porges (eds.), Psychophysiology: Systems, Processes, and Applications. New York: Guilford Press, pp. 702-718.

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.

Freidman, A. and Polson, M.C., 1981. The hemispheres as independent resource systems: Limited capacity processing and cerebral specialization. Journal of Experimental Psychology: Human Perception and Performance, 7, 1030-1058.

Gopher, D. & Donchin, E., 1986. 'What is workload -- An introduction'. In: L. Kaufman and K. Boff (eds.), Handbook of perception and human performance. New York: John Wiley, pp. 41.1-41.49.

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.

Isreal, J.B., Chesney, G.L., Wickens, C.D., & Donchin, E., 1980a. 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., 1980b. 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.

Kinsbourne, M., & Hicks, R.E., 1978. 'Functional cerebral space: A model for overflow, transfer and interference effects in human performance: A tutorial review'. In: J. Requin (ed.), Attention and Performance VII. Hillsdale, New Jersey: Lawerence Erlbaum, pp. 345-362.

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., Sirevaag, E. & Braune, R., 1987. A psychophysiological assessment of operator workload during simulated flight missions. Human Factors, 29(2).

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

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.

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

Sanders, A.F., 1979. 'Some remarks on mental load'. In: N. Moray (Ed.), Mental workload: Its theory and measurement. New York: Plenum Press, pp. 41-78.

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

Tukey, J.W., 19[illegible]. Exploratory Data Analysis. Reading, Mass.: Addison Wesley.

Wickens, [illegible], 198[illegible]. 'The structure of processing resources'. In: R.S. Nickerson (Ed.), Attention and Performance VIII. Hillsdale, N.J.: Lawrence Erlbaum.

Wickens, [illegible], 1984. Engineering Psychology and Human Performance. Columbus, Ohio: Charles Merrill Publishing Company.

Wickens, [illegible], Israel, 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.

Table 1
Order of task presentation

	VELOCITY 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
RMS error and counting 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

SUBJECT	SLOPE	INTERCEPT	SUBJECT	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

Mean slope = 0.04 Std. error = 0.09
Mean intercept = 92.99 Std. error = 4.47

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 +/- 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.

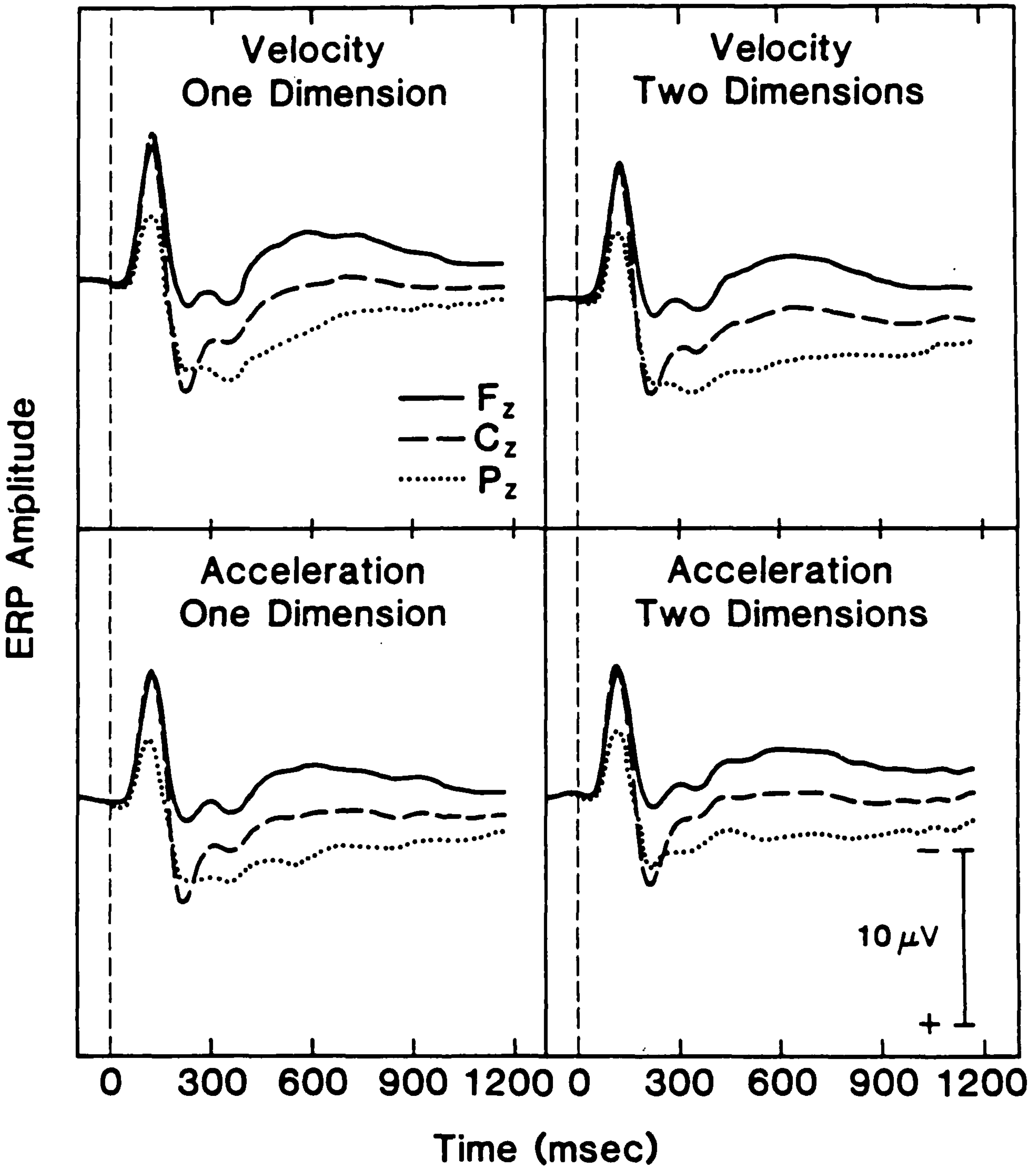
Figure 3 The scalp distribution of the grand average ERPs elicited by the primary task step changes for the different step tracking conditions.

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.

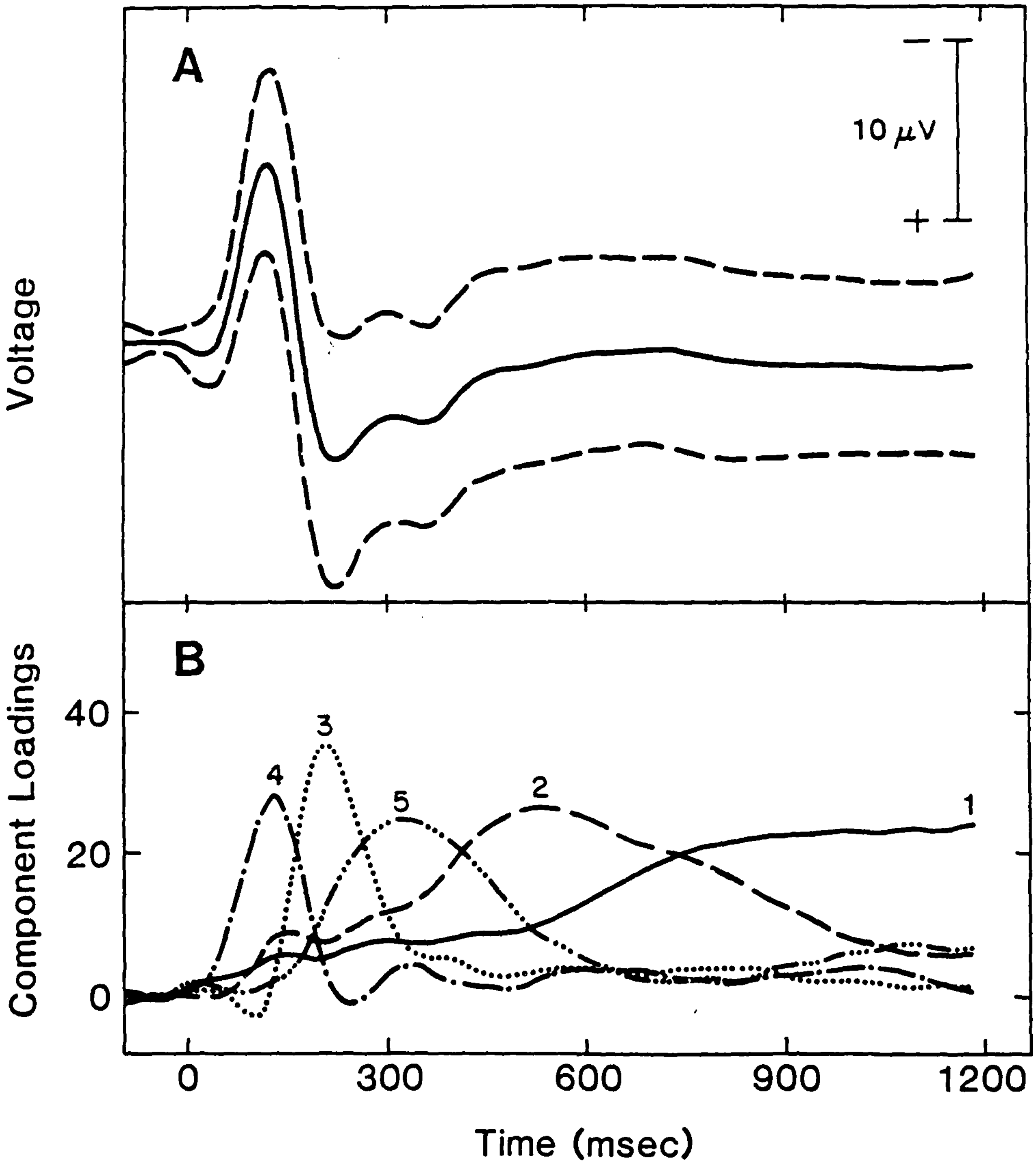
Figure 5 Panel A) Grand average waveform with +/- one standard deviation unit for the primary task waveforms. Panel B) Component loadings for the first 4 components extracted from a PCA of the primary task ERPs.

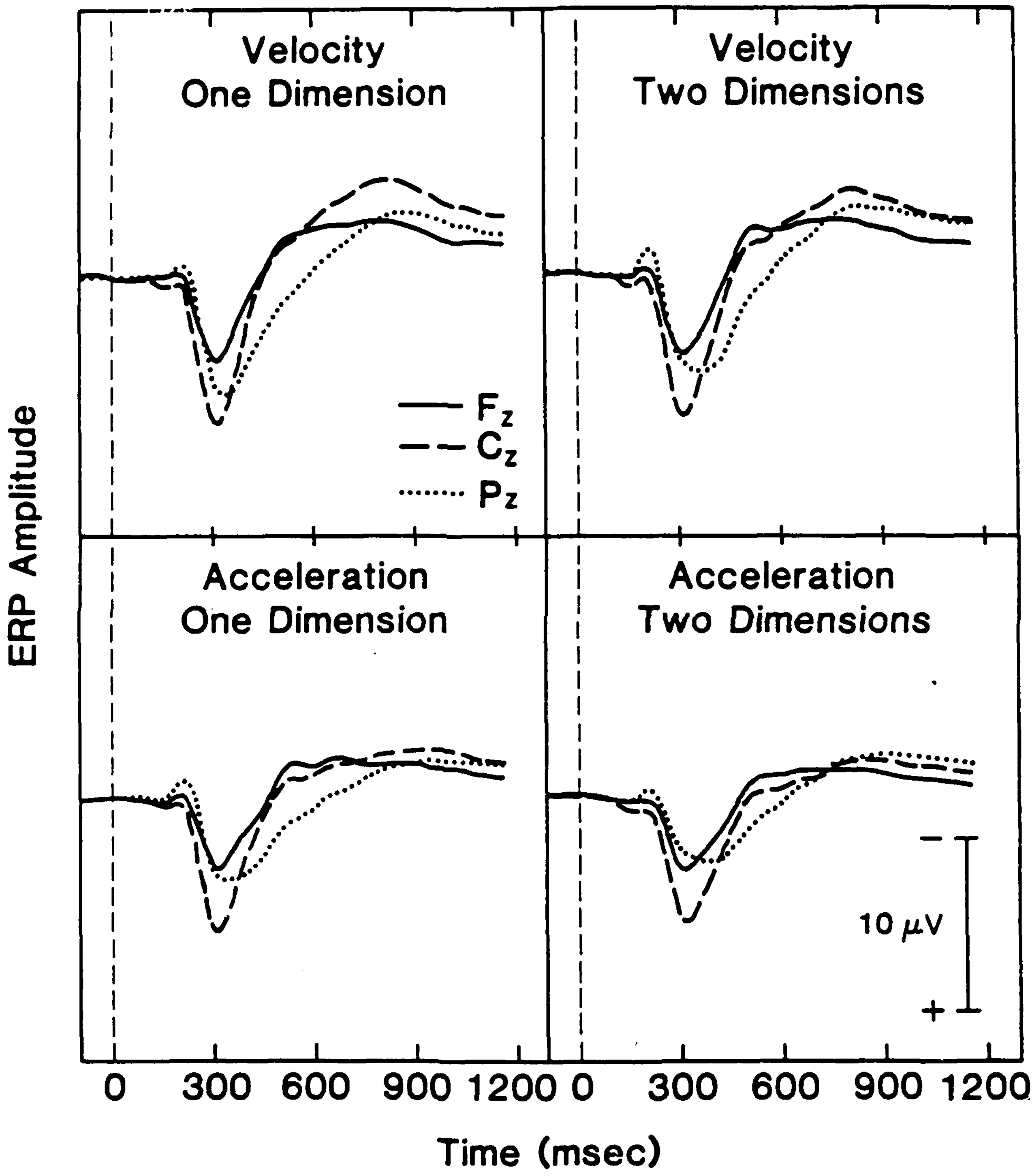
Figure 6 Panel A) Grand average waveform with +/- one standard deviation unit for the combined primary and secondary task ERPs. Panel B) Component loadings for the first 5 components extracted from a PCA of combined primary and secondary task ERPs.

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.

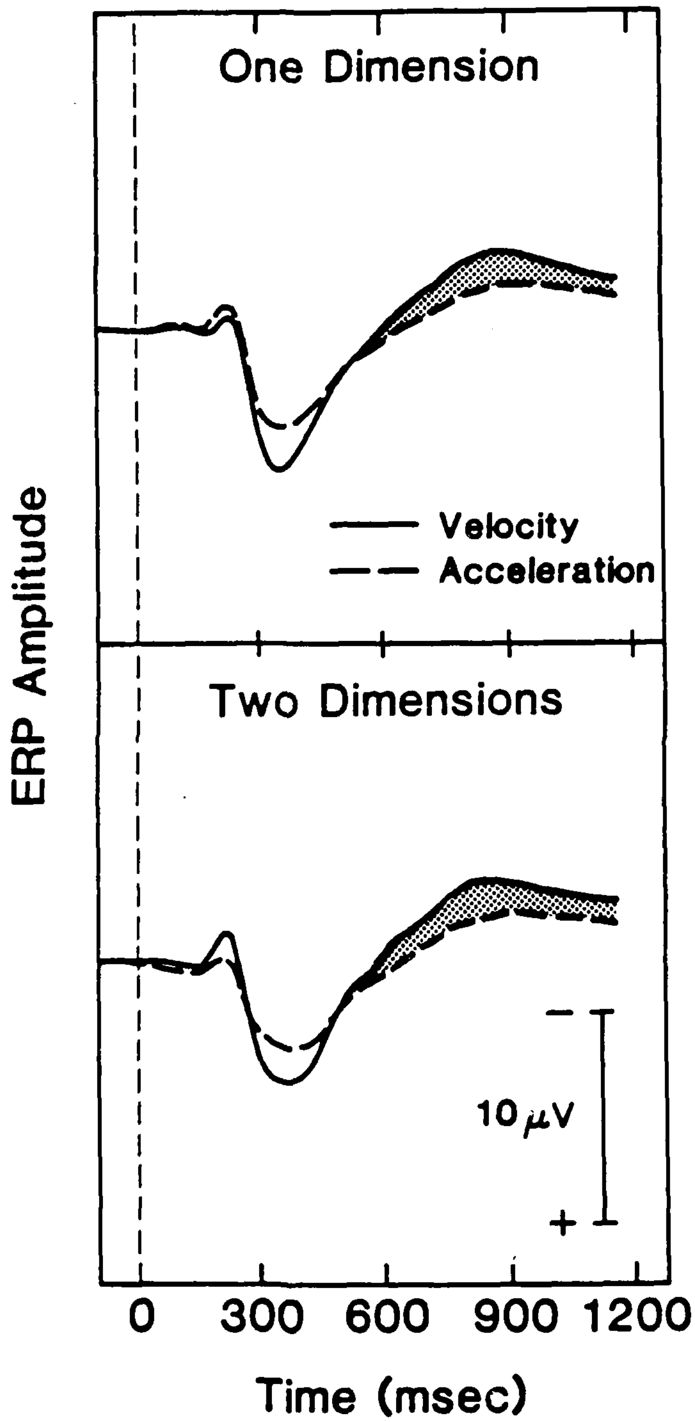


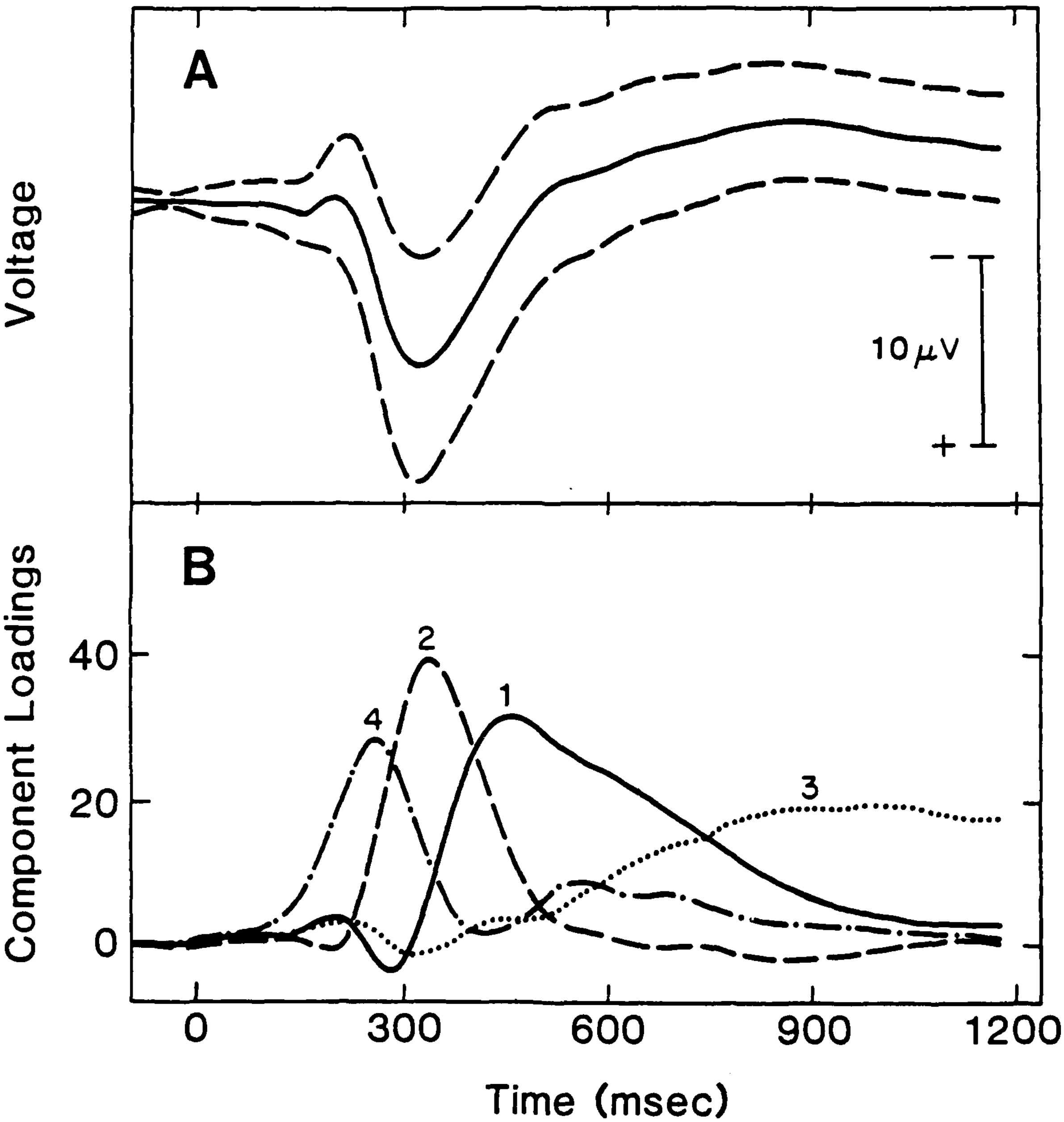
Velocity
One Dimension
Velocity
Two Dimensions
Acceleration
One Dimension
Acceleration
Two Dimensions
F_z
C_z
P_z
ERP Amplitude
10 μV
−
+
0
300
600
900
1200
Time (msec)

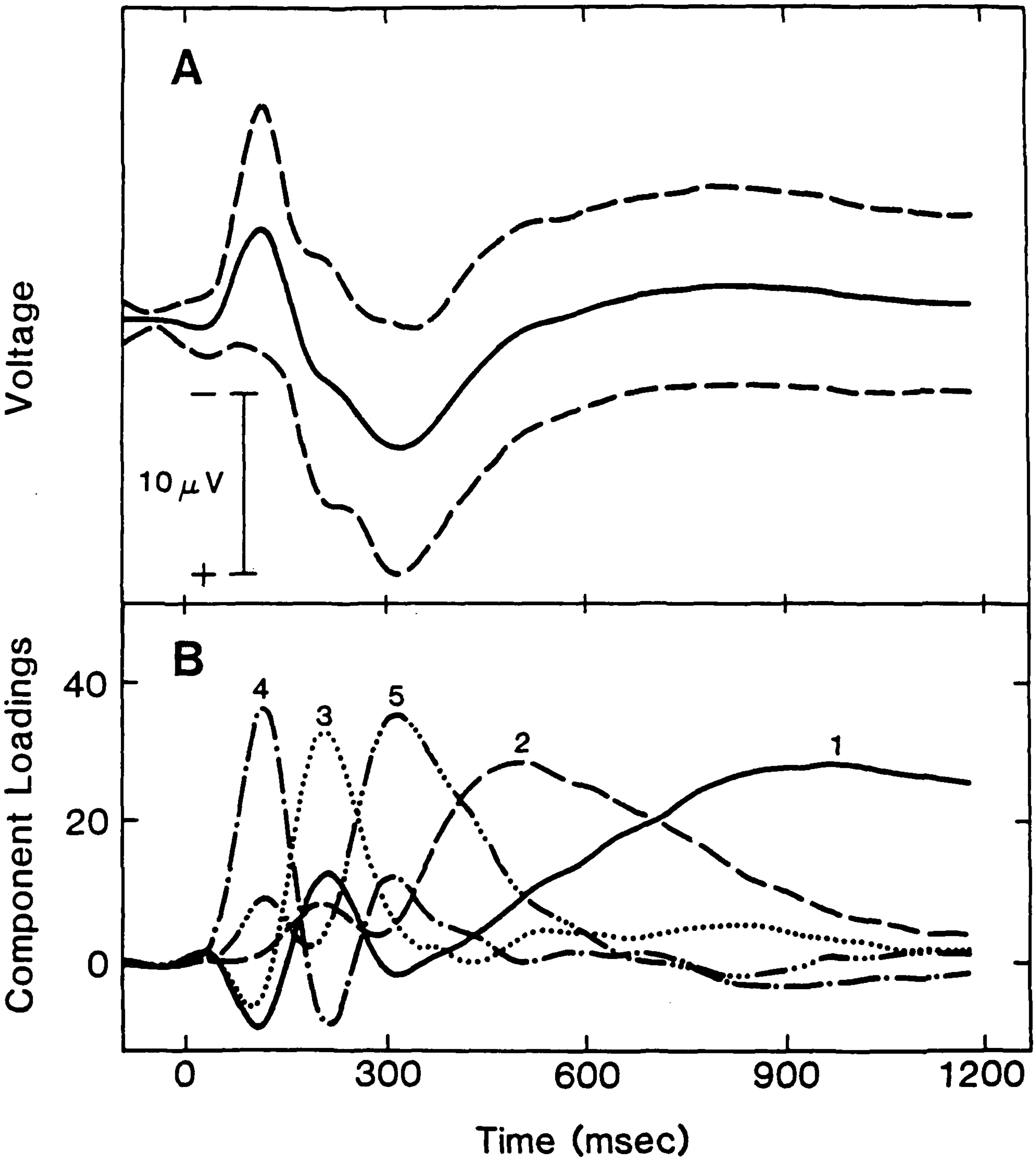

A
10 μV
-
+
Voltage
B
Component Loadings
40
20
0
4
3
5
2
1
0
300
600
900
1200
Time (msec)



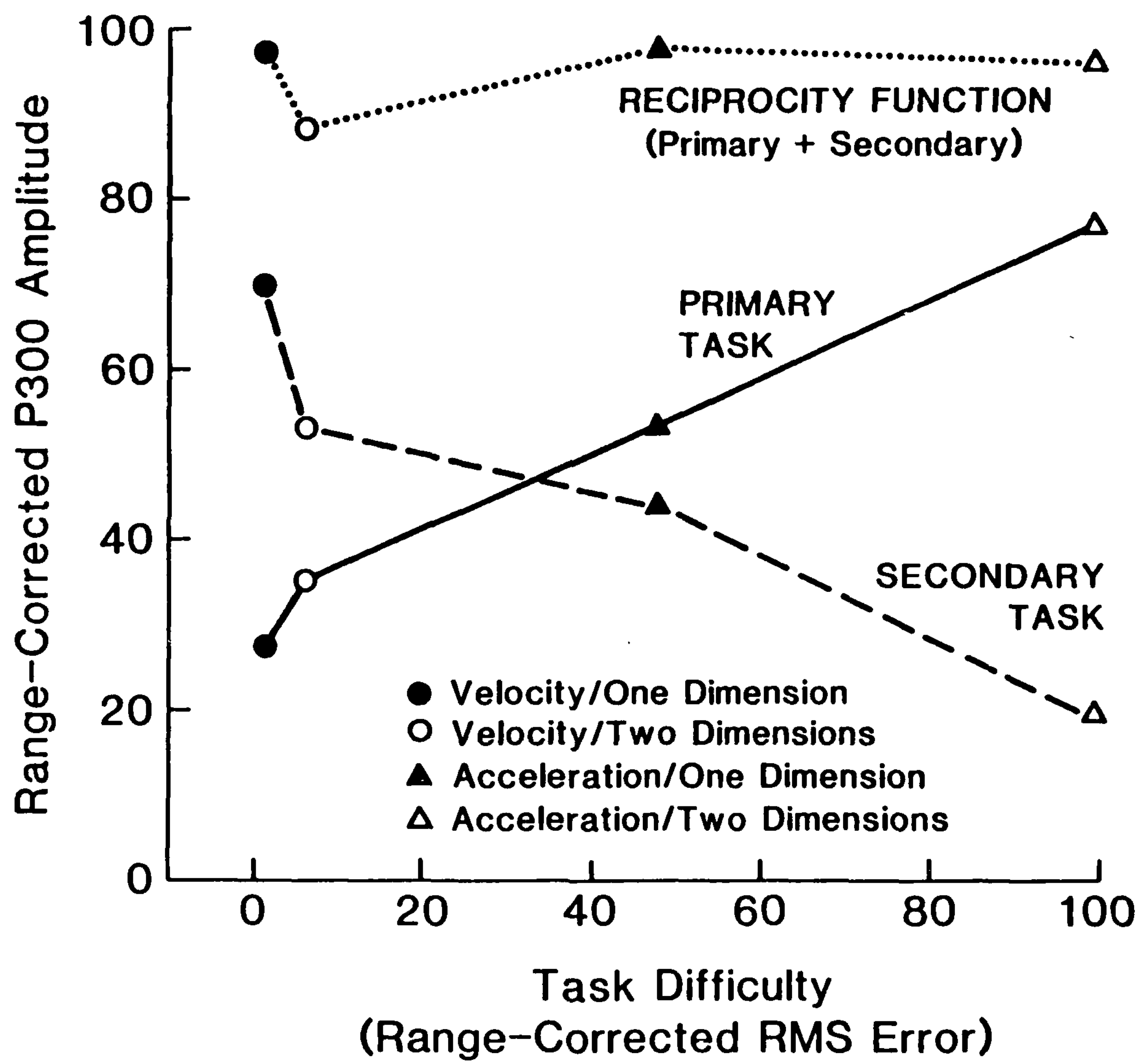
Velocity
One Dimension
Velocity
Two Dimensions
Acceleration
One Dimension
Acceleration
Two Dimensions
F_z
C_z
P_z
ERP Amplitude
10 μV
−
+
0
300
600
900
1200
Time (msec)


One Dimension
Velocity
Acceleration
Two Dimensions
ERP Amplitude
10 μV
0
300
600
900
1200
Time (msec)


A
Voltage
−
10 μV
+
B
Component Loadings
40
20
0
4
2
1
3
0
300
600
900
1200
Time (msec)



A
Voltage
10 μV
B
Component Loadings
40
20
0
4
3
5
2
1
0
300
600
900
1200
Time (msec)


Range-Corrected P300 Amplitude
100
80
60
40
20
0
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)

Article 21.

An animal model for the P300 component of the event-related potential in humans

N. Stolar, S. Sparenborg, E. Donchin and M. Gabriel

Department of Psychology
University of Illinois
603 E. Daniel Street
Champaign, IL 61820

Submitted for publication, 6/87

Abstract

This study assesses the use of a conditioning paradigm with rabbits as a model system for identifying neural substrates of the late positive (P300) component of the human event-related potential. Multi-unit and macropotential responses in five limbic system structures (the anterior and posterior cingulate cortices [ACC and PCC], the dentate gyrus [DG], and the anterior ventral and medial dorsal thalamic nuclei [AVN and MDN]) were recorded during the acquisition and performance of locomotory conditioned responses. The behavior, performed in an activity wheel in response to a 0.5 sec. tone conditional stimulus (CS+), prevented the occurrence of a shock unconditional stimulus (US) scheduled 5 sec. after CS+ onset. The rabbits also learned to ignore a different tone (CS-), not predictive of the US. Training was given daily (120 trials, 60 with each stimulus, in a random order), until discrimination between the CSs reached criterion. The intertrial interval (ITI) varied from 5 to 25 sec. in a random sequence. After criterion (Experiment 1), standard asymmetric probability (AP) sessions were given in which the CS+/CS- proportions were .2/.8 or .8/.2. The standard AP sessions were the same as conditioning sessions except for the probability manipulation. In addition, nonstandard AP sessions involving a brief (1 sec) ITI, no US, and behavioral response prevention, were presented before training and at several behaviorally-defined stages of acquisition. A significant discriminative response, i.e., a greater unit discharge to the CS+ than to the CS-, developed in all regions during the course of behavioral acquisition. Generally, the magnitude of the unit responses elicited by rare CSs exceeded that elicited by frequent CSs and by equiprobable CSs. During the standard AP sessions, the ACC and the DG exhibited these effects in trained rabbits independently of CS relevance. The minimal PCC, AVN and MDN responses to the CS- in trained rabbits was increased in the standard AP sessions when the CS- was presented rarely, but the ample response to the CS+ in these structures was not altered, or it was reduced by rare presentation. Response amplitudes in all areas were reduced during the nonstandard AP sessions, compared to standard sessions, but in all areas the responses elicited by the rare CSs exceeded those elicited by the equiprobable and frequent CSs in these sessions. In Exp 2, standard AP sessions were presented to trained rabbits with lesions of the subicular complex of the hippocampal formation, in order to study the effects of interrupting the hippocampal projection to the PCC and AVN. The unit and macropotential discriminative responses in the AVN were enhanced, and the macropotential responses to the rare CS+ in the PCC and AVN were enhanced in rabbits with lesions, relative to controls. These data are discussed in relation to current models of P300 and limbic system functions, and implications for the localization of the P300 in humans are considered.

Event-related potentials (ERPs) can be extracted from the ongoing EEG activity that is recorded from the scalp in human subjects. These potentials constitute that portion of the EEG that is time-locked to the eliciting event (6). Considerable evidence has accumulated in the past two decades indicating that specific components of ERPs can be recorded reliably under fairly well-defined experimental circumstances. Some of these components have been studied in sufficient detail to permit development of theories about the functional relevance of the intracranial generators (12,14,16,30,41). However, there is as yet little empirical information about the nature of this intracranial activity, its sites of origin or its functional relevance. We report here an investigation of multi-unit activity and intracranial macropotentials recorded in limbic system structures in awake, behaving rabbits in response to stimuli that varied in terms of their relative frequency of presentation (i.e., probability) and behavioral relevance. The results suggest that the activity elicited by these stimuli is analogous to the P300, one of the well-defined ERP components.

The P300 was first described by Sutton (63). It appears as a wave of positive polarity with a latency of at least 300 msec in response to events that, in Sutton's phrase, "resolved the subject's uncertainty". Subsequent work has shown that the P300 amplitude is sensitive to both the behavioral relevance of the eliciting stimuli (i.e., whether or not the stimuli call for action by the subject) and the probability of the stimuli. The relationship between these two variables remains somewhat controversial. However, it is clear that events must have behavioral relevance if they are to elicit the P300, and that under normal circumstances the lower the probability of the event, the larger will be the P300 that it elicits.

The data supporting these conclusions have been reviewed by Pritchard (54). The same data also form the basis for several theoretical interpretations of the P300. Particularly useful in subsuming many of the empirical findings is the theoretical account proposed by Donchin et al.(12,14), stating that the P300 is a manifestation of activity invoked when there is a need to update templates in the subject's working memory. Yet, alternative theories of P300 are also consistent with the data (9,10,56,57), and there is no present means to discriminate decisively among these alternatives.

The reasons for this are both technical and theoretical. There is some controversy regarding the number of "components" that are in fact active at any given latency (57). Furthermore, it is possible that different constellations of intracranial activities yield similar configurations of activity at the surface of the skull (69). If this is true, the theory of the P300 will remain indeterminate unless information about intracranial activity is available. Thus, one reason to investigate the intracranial origins of the P300 has to do with the measurement of this potential: such information can indicate whether the P300 reflects a collection of functionally disparate informational processes in the brain, or a process that is unitary. A second reason to carry out such investigations is that knowedge from neuroscience about the behavioral functions of the involved neural systems can facilitate, and may decisively establish, the functional interpretation of the P300.

The need to elucidate the neural origins of such components as the P300 is being addressed by three main avenues of research. When clinical circumstances have permitted it, investigators have replicated ERP results derived from scalp recordings, using intracranial electrodes in humans (33,40,42,43,60,61,71,72,73,74). In addition, several groups are attempting to record the magnetic equivalent of the human P300 (34,48,55). This research will become increasingly important as the technology for the recording of magnetic fields develops, thus permitting more powerful inferences about intracranial sources. The use of animal models to study the intracranial origins of the P300 is an obvious approach to the problem, assuming that animals exhibit phenomena analogous to the P300 in humans. Indeed, studies of ERPs in a variety of non-human species have indicated that a wave very similar to the P300 can be recorded with eliciting conditions similar to those used for study of the P300 in humans (1,4,19,27,31,35,36,37,39,40,44,45,46,47,51,52,59,67) A particular advantage of this approach is the opportunity for localizing the recorded potentials. Just as for surface ERPs, the interpretation of intracranial ERPs can be confounded by voltage signals originating at loci distant from the recording site. However, this problem can be addressed in the cerebral cortex by obtaining laminar ERP profiles that permit the derivation of the signal sources (69). Moreover, the sites of origin of single- and multi-unit activity are easily determined as this activity originates in the immediate vicinity of the recording probe (3).

In the present paper we report the first application of multi-unit recording in behaving animals to the problem of localizing brain activities analogous to the P300. The data were obtained from rabbits, while they were engaged in a discriminative avoidance task. In this task the rabbits learn to respond when they hear a shock-predictive tone (CS+) and to ignore a different tone (CS-), not predictive of shock. Several considerations led us to choose this particular model. First, past work has demonstrated that event-related single- and multi-unit activity can be recorded from cortical and thalamic structures of the limbic system during acquisition and performance in this paradigm. This activity is similar in certain respects to the P300 (27). The structures of interest were the anterior and posterior cingulate cortices (ACC and PCC), the medial dorsal and anterior ventral nuclei of the thalamus (the MDN and AVN), and certain areas of the hippocampal formation. Of particular interest was the sensitivity of the neurons in these structures to variations of the behavioral relevance of the eliciting stimuli (20,22,25,50). That is, distinctive neuronal discharges developed in response to CS+ presentations, relative to the responses to the CS-, as the rabbits apprehended the different associative significances of these stimuli. Sensitivity to the behavioral relevance of eliciting stimuli is, as noted above, a hallmark of the P300. Furthermore, the theory of the limbic system's role in discriminative conditioning that is based on the data from rabbits (23,29) is intriguingly compatible with the mnemonic interpretation of the functional significance of the human P300 (12,14).

Our past work has demonstrated robust unit sensitivity in the limbic structures to the behavioral relevance of the CSs. In the first experiment (Exp 1), we attempt to extend the analogy between the unit activity and the the P300 by examining the degree to which the same structures exhibit sensitivity to the manipulation of the probability of the stimuli.

A rather substantial body of evidence supports the idea that the hippocampal formation is involved importantly in the processing of rare and/or behaviorally relevant stimuli (8,29,32,53,64). This viewpoint, as well as the rather striking potentials recorded in the hippocampus in human patients responding to rare and relevant stimuli (33,43) suggest that the hippocampal formation may be involved in the generation of the P300. Indeed, it is possible that the hippocampus may be a primary site of neural sensitivity to stimulus relevance and probability, and that the exhibition of this sensitivity in nonhippocampal areas may be owed to influences projected from the hippocampus. Experiment 2 provides data relevant to this hypothesis, in the form of macropotentials and unit activity recorded from the PCC and AVN, both major projection targets of the hippocampal formation (65). These records were obtained in intact rabbits and in rabbits previously given bilateral electrolytic lesions in the subicular complex, the region of the hippocampal formation in which efferent projections to the

PCC and AVN originate (65). The results indicate that the responses to stimulus-relevance and probability in the PCC and AVN are not hindered, indeed certain aspects of the response are enhanced, in rabbits with disrupted hippocampal afferents.

METHODS

Subjects, Surgical Procedures, Electrodes and Target Areas

The subjects were male, New Zealand White rabbits, weighing 1.5-2.0 kg at the time of their delivery to the laboratory. Seventeen and 29 rabbits served as subjects in Experiments 1 and 2, respectively. They were maintained on ad libitum water and food throughout the experiments. Following a minimum period of 48 hr, surgical anaesthesia was induced by injection of 12.5 mg/kg of chlorpromazine in solution (25 mg/ml), followed after ten minutes by 25.0 mg/kg of sodium pentobarbital in solution (50 mg/ml), into the marginal vein of the pinna. The rabbits were placed in a headclamp and the skull was exposed and prepared for stereotaxic implantation of the electrodes. Supplemental injections of sodium pentobarbital in solution (0.07 mg/kg) were given at 0.5-hr intervals to maintain anesthesia.

The electrodes were made from stainless-steel insect pins insulated with epoxylite (outside diameter: 0.5 - 0.8 mm). Recording surfaces were formed by removing insulation from the tip (tip lengths: 20-60 microns; impedance: 0.25 - 2 megohms). Neuronal activity was monitored acoustically and with an oscilloscope during electrode lowering. Two stainless-steel machine-screw (256 x 1/8) electrodes for recording epidural EEG were threaded into the skull flush with the skull undersurface. A stainless-steel screw threaded into the frontal sinus served as the reference electrode. Once lowered, the electrodes and a miniature multipin connector to which they were pre-soldered, were affixed to the skull with dental acrylic.

The targets of the intracranial electrodes were: the anterior cingulate cortex (ACC: Brodmann's Area 24), the posterior cingulate cortex (PCC: Areas 29b and 29c), the dentate gyrus (DG), the mediodorsal thalamic nucleus (MDN) and the anteroventral thalamic nucleus (AVN). Four of these sites were chosen for each rabbit.

The stereotaxic coordinates, based on the atlas of Fifková and Marsala (5), were AP -4.0, L 0.9-1.5, DV 2.0-3.0 for the ACC; AP 4.0, L 0.8, DV 1.0-2.0 and AP 9.0, L 2.0, DV 4.0, respectively, for the dorsomedial and ventrolateral PCC; AP 6.0, L 6.5, DV 3.5-4.0 for the DG; AP 4.5, L 1.2, DV 8.5 for the MDN; and AP 1.8, L 2.3, DV 7.5-8.0 for the AVN. The data for the two PCC subfields were combined. The surface electrodes were placed 1 mm from the midline, 5.0-6.0 mm anterior to bregma, overlying Area 8, or 5.0-9.0 mm posterior to bregma, overlying Area 29d. These sites are referred to respectively as the anterior surface (AS) and the posterior surface (PS).

Procedures of Behavioral Training.

After a minimum of seven days following surgery, each rabbit received avoidance conditioning in an activity wheel, a replica of the apparatus of Brogden and Culler (2). The wheel was located in a shielding chamber housed in a room adjacent to the one containing the computer that controlled data collection. An exhaust fan and a speaker produced a masking noise of 70 dB re 20 N/m. The conditional stimuli were pure tones (1 or 8 kHz, 85 dB re 20 N/m, 500 ms in duration, rise time=3 ms) played through a speaker directly above the wheel. The unconditional stimulus (US) was a constant AC current (1.5-2.5 ma) delivered through the grid floor of the wheel. A response was defined as any wheel rotation exceeding 2 degrees. The US current was established for each rabbit at the outset of conditioning, as the minimum value needed to reliably elicit responses.

During training, onset of the positive conditional stimulus (CS+) was followed 5 seconds later by US onset. The CS+ and US were terminated by responses. A response after CS+ onset and prior to US onset prevented the US. The maximum duration of the US, given failure of response, was one second. Presentations of the negative conditional stimulus (CS-; at the frequency [1 or 8 kHz] not used for the CS+) were randomly interspersed with the CS+ presentations. The CS- was also response-terminated, but it was not followed by the US. The interval between the end of a trial (defined as offset of the CS or of wheel rotation when locomotion occurred) and the onset of a new trial (CS onset) was 10, 15, 20 or 25 sec. These values occurred in a random sequence. Responses during the interval reset it.

The sequence of CS+ and CS- for a given training session was one of three pseudorandom sequences designed to minimize the subject's ability to predict the next CS, and to minimize the likelihood that the subject would be influenced by higher-ordered sequential conditional probabilities. The sequences were selected on a rotating basis throughout training for a given subject.

Each subject received training (120 trials daily, 60 with the CS+ and 60 with the CS-) until the proportion of avoidance responses (i.e., responses to the CS+ before US onset) exceeded the proportion of responses to the CS- by .60 or more in any 60-trial block. This performance had to occur in two consecutive sessions. This criterion yields an asymptote of discriminative behavior not exceeded with further training.

Pretraining. Before conditioning, two pretraining (PT) sessions were given. In the first, the tones later used as CSs were presented alone, and in the second, the tones and the US were presented in a noncontigent (explicitly unpaired) fashion. The frequency and temporal distribution of the US during PT were identical to the average values obtained during the first session of conditioning in a sample of 100 rabbits. The only deviation from this procedure was that in PT, the US was never presented in the interval from three seconds before to three seconds after a CS. The PT sessions served to provide a baseline of neuronal and behavioral responses to the CSs, for comparison with the responses that developed during training.

Standard Asymmetric Probability Sessions. Following criterion attainment, the rabbits received two standard asymmetric probability (AP) sessions. In the first of these, ten of the rabbits received the CS+ and CS- in .2 and .8 of the trials, respectively. These proportions were interchanged in the second session. Seven additional rabbits received these treatments in the opposite order. Following the two standard AP sessions, all of these rabbits received a standard training session in which the CSs were presented equally often.

Nonstandard AP Sessions. These sessions involved 200 CS presentations with a brief (1-sec.) interstimulus interval, the prevention of wheel movement, and no US. None of the rabbits attempted to locomote in the wheel after the first few trials. This procedure was designed to approximate the "oddball" paradigm used in studies of ERPs with human subjects (13).

The nonstandard AP sessions were given 2-5 min. after the end of the PT sessions and after training sessions representing certain landmarks of behavioral acquisition. These were the first session of conditioning, the session in which a significant behavioral discrimination first occurred, and the second of the two sessions in which the acquisition criterion was met. The session in which the first significant behavioral discrimination occurred was defined as the session in which the proportion of avoidance responses to the CS+ exceeded that to the CS- by .25 or more. This value approximates the minimum required to produce a significant chi-square (p=.05) for a difference between correlated proportions (66).

A subset of the rabbits (N=9) received only one nonstandard AP session at each training stage. In these sessions the only asymmetric probability treatment was the rare presentation (i.e., in .2 of the trials) of the CS+. The remaining rabbits (N=8) received two nonstandard AP sessions, one with rare and the other with frequent CS+ presentations in a counterbalanced order. The duration of the nonstandard AP sessions was about 10 min.

Subicular lesions. The rabbits with subicular lesions (N=10) and controls (N=18) received standard conditioning and overtraining as described above, but nonstandard AP sessions were not given to these rabbits. The data reported here were recorded during the third postcriterial overtraining session and during a single standard AP session with rare CS+ presentation (CS+/CS- proportions = .2/.8). The lesions were made at two subicular loci, one anteromedial and another posterolateral, along the septotemporal axis of the hippocampal formation. The coordinates (anteromedial site: P 6.00, L 2.50, V 5.00; posterolateral site: P 7.50, L 4.25, V 4.00) were derived from the atlas of Fifkova and Marsala (5). The lesioning electrodes were made with stainless-steel insect pins coated with epoxylite. The insulation was removed 0.50-0.75 mm. from the pointed ends of the pins. A current of 1.5 ma was delivered at each site for 30 or 45 sec.

Recording and Analysis of Neural Activity.

Throughout training, the neural records were fed into a field effect transistor (FET) which served as a high impedence source-follower located about 2.5 cm from the recording site. The FET outputs were split, one limb entering single-ended preamplifiers with gain (4000) and bandwidth (1/2 amplitude cutoffs at 0.1 and 20 Hz) suitable for ERP signals. The other limb entered preamplifiers with gain (10,000) and bandwidth (1/2 amplitude cutoffs at 600-10,000 Hz) suitable for recording unit action potentials. The outputs of the unit preamplifiers were fed through active bandpass filters (1/2 amplitude cutoffs at 600 and 10,000 hz, rolloff=18 db/octave) to remove any residual slow potential activity. Outputs of the band-pass filters were fed in parallel into Schmitt triggers and into a circuit for full-wave rectification and RC integration (3). The Schmitt trigger outputs measured the firing frequency of the largest three or four spikes on each record, whereas the integrated activity measured energy fluctuations of the entire record, including activity below the triggering thresholds. The rise and fall time constants of the integrators were 15 and 75 ms, respectively.

Each Schmitt trigger produced a digital pulse each time a unit spike exceeded the preset voltage. The output pulses were fed into an LSI 11/23 computer programmed to process the neural data and to control the behavioral experiment, on-line. The triggering levels were set independently, under computer control, such that the mean rate of pulses fell within limits of 110-190 per second. Preceding each trial, the trigger setting was adjusted upward or downward automatically to compensate for occasional shifts in baseline activity.

The macropotential and unit records were digitized and the Schmitt trigger pulses were counted for each 10 ms interval, for a total duration of 1.0 second, 300 ms before and 700 ms after CS onset.

Several procedures were used to ensure that neural and non-neural (e.g., electromyographic) electrical activity resulting from movement did not influence the data. First, the mean latency of the avoidance response over all training sessions was greater than three seconds, but the analysis involved only neuronal results in the initial 400 ms after CS onset. Infrequently, responses occurred during the first 500 ms after CS onset. Such trials were discarded and repeated. When non-locomotor movements (e.g., licking, sneezing, grooming) occurred during the 300 ms pre-CS period or in the initial 400 ms period following CS onset, the artifact generated was detected by a Schmitt trigger, resulting in automatic rejection and repetition of the trial. In addition, the criterion for setting trigger levels relative to the neuronal records resulted in a constant baseline rate of Schmitt trigger spikes in the range of 110-190 per second. With this criterion, the expected number of counts per interval in the quiet rabbit was about 1.5. Neuronal bursts produced routinely counts of 5 or 6, but counts of 8 or more were rare unless artifacts occurred. Trials in which six or more of the 120 pre-CS 10-ms intervals (30 intervals from each of the four channels) had counts of 8 or more were automatically discarded and repeated. Finally, the experimenter rejected trials with a single keyboard command if any sound of movement was detected on the audio monitor immediately before or after CS onset. Previously, electrical effects from several electrodes located inadvertently in nonneuronal regions of the intracranial cavity (e.g., the midline, the cerebral aqueduct) were analyzed. Significant and systematic "response" activity to the tone stimuli for records of this kind was not found in any stage of conditioning.

The activity sampled on each trial was stored on digital magnetic tape for subsequent processing. Histograms indicating the average CS+ and CS- elicited firing frequencies, integrated activity, and macropotentials were displayed continuously on a VT100 graphics terminal as they were being formed during the sessions. The offline analysis of the taped data involved the computation of standard scores (z-scores) for each of the 70 10-ms intervals after CS onset. Each score was obtained by subtracting the mean of the 30 pre-CS intervals from the value in each of the 70 post-CS intervals and dividing the difference by the standard deviation of the pre-CS intervals.

Because the number of sessions required to attain the behavioral criterion varied among the subjects, the statistical analysis focused on three behaviorally-defined stages of training common to all subjects: a) A pretraining stage including the combined data of the two pretraining sessions; b) An intermediate stage, including the first session of training and the session of the first significant behavioral discrimination; and c) An asymptotic training stage, including the data obtained during the two criterial sessions and the single postcriterial session with symmetric CS probability. The neural data of the separate sessions within each of these stages were pooled by averaging the z-scores. Data from the nonstandard AP sessions that were given immediately after the standard training sessions were also pooled in this way.

Data from the standard training sessions were submitted to analysis of variance, with orthogonal factors of stage (3 levels: pretraining, intermediate training, and asymptotic training), CS type (two levels: CS+ and CS-) and interval after CS onset (40 levels: each 10-ms interval). Data from the nonstandard AP sessions were submitted to analysis of variance, with orthogonal factors of CS type (two levels: CS+ and CS-) and interval after CS onset (40 levels: each 10-ms interval). Separate analyses were performed for the data at each training stage. The data from the single overtraining session and the two standard AP sessions were submitted to an analysis of variance with orthogonal factors of CS type, interval after CS onset and session (3 levels: rare, frequent and equiprobable CS+).

Given a significant overall F, individual comparisons of mean response magnitudes were carried out ($p<.05$) using Fisher's protected least significant difference test (68).

<u>Histological Identification of Recording Sites and Lesions</u>.

After testing, each rabbit received an overdose of sodium pentobarbital followed by transcardial perfusion with normal saline and 10% formalin. Each brain was frozen and sectioned at 40 microns. The sections containing the electrode tracks and lesions were photographed while still wet (21). After photography, the sections were stained for Nissl and myelin using formol-thionin (17).

The numbers of unit and macropotential records used in the analyses are shown in Table 1. These numbers represent the final results of the histological analysis and additional screening which resulted in the exclusion of macropotential and/or unit records in certain sessions due to poor quality, technical problems, and, in one case, failure to meet the behavioral criterion of learning. The fact that subsets of the rabbits were assigned to limited sets of the behavioral treatments, as detailed in the description of the training procedures, also influenced the numbers of records in the analyses.

The dorsal subicular complex lesions are illustrated in Figure 1. Additional description of the lesions, and data yielded by the rabbits with lesions during standard training sessions, are reported elsewhere (26,28).

RESULTS

A. OVERVIEW

In all recording sites, rare presentation of the CS- increased the magnitude of the unit response in trained rabbits, relative to the responses that occurred when the CS- was presented more frequently than, or as frequently as, the CS+ (Figure 2). In addition, in the ACC, rare presentation of the CS+ increased the magnitude of the unit response relative to the other conditions of CS presentation (Figure 3). The response increments to the rare stimuli were present at latencies as brief as 70 ms after CS onset, attaining peak amplitudes from 90-300 ms after CS onset, depending on the recording site (Figure 7).

The foregoing results were obtained from trained rabbits during the standard AP sessions that were identical in all respects to standard avoidance conditioning sessions except for the CS probability manipulation. A somewhat different pattern of results was obtained during the nonstandard

AP sessions that were intended to approximate the "oddball" paradigm found effective in eliciting P300 in humans. These sessions involved high rate (1/sec) CS presentations with no US, and prevention of behavioral responses. The results obtained in these sessions indicated that in trained rabbits, a greater average unit discharge occurred in response to the the rare CS+ than to the frequent CS-. In all but one site (the AVN), the rare CS- and the frequent CS+ elicited statistically equivalent, minimal average unit responses (Figure 4). The AVN showed a greater response to the CS+, even when it was presented frequently during these sessions. Finally, the unit responses of both of the thalamic nuclei exhibited a greater discharge to the rare than to the frequent tones, during nonstandard AP sessions given before conditioning.

With some exceptions, results similar to those just described for the neuronal activity measure were also obtained for the intracranial and epidural macropotentials. These data support the conclusion that the neuronal activity in all of the monitored limbic localities exhibits both of the response features, sensitivity to stimulus relevance and probability, that characterize the P300.

Lesions of the subicular complex of the hippocampal formation enhanced the discriminative unit response (i.e., the "target effect") recorded from the AVN, and they attenuated the response in the PCC. However, these lesions did not obviously alter the responses exhibited in the PCC and AVN during the manipulation of CS probability. This outcome is in conformity with other data indicating that extracranial macropotentials exhibit normal or enhanced sensitivity to rare stimuli in monkeys or humans with damaged hippocampi (38,39,51,72). Thus, the integrity of the hippocampal formation does not appear to be required for the exhibition of this sensitivity. Indeed, the hippocampal projections appear from the present data, and from other results (26,28), to be involved in limiting the relevance- and probability-driven activities in these regions.

B. DETAILED RESULTS OF EXP 1: MANIPULATION OF STIMULUS RELEVANCE AND PROBABILITY

1. Behavior

Behavioral acquisition replicated that observed in past studies (20,29). The mean number of training sessions prior to criterion was 4.6. During the final training stage (consisting of the combined data from the criterial and final overtraining sessions, in which asymptotic performance occurred), the rabbits avoided the US on an average of 77% of the trials, and they responded to the CS- on fewer than 9% of the trials. There were no significant effects of the probability manipulations on performance during overtraining.

2. Neural Changes During Acquisition

Overview. As in past studies, the unit responses to the CSs increased in magnitude and became discriminative during conditioning. These effects in the ACC, PCC, MDN and AVN represent replication of past findings with this paradigm. The changes in the DG are reported here for the first time. The amplitudes of certain components of the surface and intracranial macropotentials increased during training, but target effects developed only minimally in the macropotentials, and primarily in the epidural recording sites.

Unit activity. The firing frequency and integrated unit activity yielded qualitatively similar outcomes in this study but the integrated activity measure was less variable and more sensitive to the experimental manipulations than the spike frequency measure. Thus, only the results obtained with the integrated unit activity measure are presented. The full set of data are available upon request (62).

As in past studies, the histogram profiles of the PCC and AVN were triphasic in form, consisting of peak excitatory discharges at 20-40 and 80-250 ms, and an inhibitory pause at 40-100 ms after CS onset (Figure 5, rows 2 and 5). Also, as before, the profiles of the ACC and the MDN were essentially biphasic, exhibiting a rising excitatory discharge that reached a peak value at approximately 100-150 ms. This level was maintained in the MDN throughout the remainder of the sampling interval (Figure 5, row 4), whereas the activity in the ACC declined gradually after 150 ms (Figure 5, row 1). An excitatory discharge with a 200-300 ms peak and subsequent, gradual decline was recorded in the DG, an area not previously studied in this paradigm (Figure 5, row 3).

The analyses indicated that a discriminative unit response developed during training in all recording sites. That is, the average response to the CS+ exceeded the response to the CS- for the data of the asymptotic training stage, but no such difference occurred during pretraining (Figure 5). This conclusion is based on the results of individual comparisons of the means comprising the interaction of the training stage, stimulus and interval factors. The individual comparisons indicated that in each of four sites (PCC, DG, MDN, AVN), the CS+ elicited significantly greater average unit responses than the CS- in a majority of the 10-ms poststimulus intervals from 100-400 ms after CS onset during the asymptotic training stage, but not during the pretraining stage (Table 2). In addition, the magnitude of the average CS+ elicited response in the asymptotic stage exceeded the response recorded during pretraining at several intervals in all of these areas (Table 2).

The training-related development of increased unit response magnitudes during conditioning was indicated for the ACC by significant two-way interactions of stimulus with interval, and stage with interval ($F[39/429]=3.36$; $p<.0001$; $F[78/858]=1.60$, $p=.0011$, respectively). The individual comparisons based on the first of these interactions showed a significantly greater discharge to the CS+ than to the CS- at 240-270 ms for the profiles pooled across all training stages. Since there was no interaction of stimulus and training stage, these results do not provide a direct demonstration of development of a discriminative response during training. However, the occurrence of a significant overall discriminative response, and the absence of any discriminative effect during pretraining (Figure 5, upper left panel) support the conclusion that differentiation of CS+ and CS- developed with training, in replication of several past reports. The comparisons based on the interaction of training stage and interval demonstrated significant overall (nondiscriminative) increases in unit activity in virtually all poststimulus intervals after 100 ms during the asymptotic training stage, relative to pretraining.

Macropotential profiles. The waveshapes of the average intracranial macropotentials varied with the recording site, but certain shared features were also exhibited. Thus, in all but the PCC, the intracranial profiles showed a wave of negative polarity with a peak value at 100-200 ms after CS onset. Within this same latency range the surface recordings exhibited a large positive deflection (Figure 6). In addition, all but the PCC profile exhibited one or two positive deflections from 150-400 ms, latencies dominated by negativity in the surface macropotential records. Finally, a majority of the intracranial records exhibited rather sharp positive potentials at brief latency (30-70 ms), but small negative deflections were observed in the posterior surface recordings within this interval. Thus, the profiles of the surface and intracranial potentials were approximate mirror images. The only exception, the PCC profiles, exhibited a negative potential after 150 ms, and a sharp positive potential at 75 ms, in advance of the initial positive potentials in the other sites.

The overall polarity opposition between the surface and intracranial profiles is in accord with the traditional assumption that surface records are driven primarily by voltages that are the distal sources of intracranial voltage sinks, the primary sites of neural activation.

Macropotential amplitude changes during acquisition. The amplitudes of the intracranial and surface macropotentials increased during the course of behavioral acquisition, as indicated by comparison of the profiles obtained during pretraining (Figure 6, left column), with those exhibited during the asymptotic training stage (Figure 6, right column). This conclusion is based on marginally significant main effects of training stage for the ACC (p=.0661), DG (p=.0413) and MDN (p=.0582). The increased amplitudes were indicated, in the case of the AS, PS, DG and PCC, by the occurrence of significant interactions of the stage and interval, and of the stage, interval and stimulus factors (Table 3). Individual comparisons indicated that the AS and PS positive deflections from 75-200 ms, and the negative deflections from 300 to 400 ms, increased in the asymptotic training stage relative to the pretraining stage. Similar findings were obtained for the PCC. However, in this instance, the peak of the sharp positive deflection occurred at a briefer latency than in the surface records.

The principal amplitude increase in the DG occurred in relation to the profile elicited by the CS+, which was significantly reduced relative to the CS- profile during pretraining, but which underwent an enlargement during the asymptotic training stage. These changes occurred from 80-170 ms after CS onset (Figure 6, row 5, and Table 3).

Macropotential discriminative responses during acquisition. The interactions of the training stage and stimulus factors by interval indicated that discriminative macropotentials developed during training, in the AS, PS, PCC and DG (Table 3). During pretraining, the anterior surface (AS) macropotential profile elicited by the CS+ exhibited no differences relative to the CS- elicited profile. In the case of the PS profile, and for the PCC, the CS- elicited more ample potentials than did the CS+.

During the asymptotic training stage, however, the amplitudes of the AS and PS positive potentials elicited by the CS+ exceeded significantly those elicited by the CS-. The discriminative responses in these instances occurred from 60-200 ms for the AS, and from 110-180 ms after CS onset for PS (Figure 6, rows 1 and 2). In the PCC, the potential elicited by the CS+ exceeded that elicited by the CS- in the brief-latency positive wave (from 60-110 ms), and at greater latencies (160-360 ms, Figure 6, row 3). Discriminative macropotential development in the DG occurred in the form of an increase during asymptotic training in the potential elicited by the CS+. This potential had been significantly reduced relative to that elicited by the CS- during pretraining (80-170 ms, Figure 6, row 5). However, the potential elicited by the CS+ exceeded that elicited by the CS- in the trained rabbits (200-210 ms, Figure 6, right column, row 5).

Discriminative macropotential development in the AVN (Figure 6, row 6) was suggested by a significant interaction of the stimulus and interval factors ($F[39/156]=1.87$, $p=.0040$). However, the individual comparisons were not significant.

In summary, robust increases in macropotential amplitude occurred during the asymptotic training stage in all of the sites from which recordings were made, excepting the AVN. Even in the case of the AVN, however, the nonsignificant differences suggested a training-induced increase in macropotential amplitude, a result that has occurred in past studies (28) and that would have been significant in this experiment given a larger sample of AVN records.

In contrast to the robust amplitude increases in all of the macropotential records, and in contrast to the development of discriminative unit responses in all of the areas, discriminative macropotential responses developed only in the epidural macropotential records, and in the PCC and DG. Large macropotential amplitude increases and minimal or nonexistent discriminative macropotential effects have been replicated for ACC, PCC and AVN since the present observations were made (28).

3. Manipulation of CS probability during overtraining: The standard AP sessions.

Prologue. Each rabbit received two AP treatments, during the first and second postcriterial (overtraining) sessions, in which the proportions of CS+ to CS- presentation were either .2/.8, or .8/.2. The third overtraining session, in which the stimuli were presented with equal frequency, provided data for comparison with the two standard AP sessions. The order of the two AP sessions was counterbalanced. With this arrangement, the effects of the probability manipulation were assessed by the interaction of sessions and stimuli.

Unit responses to the CS-. An enhanced unit response to the rare CS- was indicated by the significant interactions of the session and stimulus factors. The three-way interaction of these factors with the interval factor was significant in the analyses of ACC, DG and AVN unit activity. Individual comparisons revealed a greater neuronal response to the CS- when it was presented rarely than when it was presented frequently and equally as often as the CS+. This occurred in a majority of 10 ms intervals from 90 to 400 ms after CS onset (Figure 7, Table 4).

The increased unit response of the PCC to the rarely presented CS-, relative to the symmetrically and frequently presented CS- (Figure 7, right panel, second row) was indicated by a significant interaction of the session and stimulus factors, without significant participation of the interval factor (Table 4). Rare presentation of the CS- increased the average standard score to a value of 7.13 from values of 2.90 and 2.55, in the symmetrically and frequently presented CS- conditions respectively; however, the individual comparisons for these effects were not significant. Similarly, the average standard score associated with rarely presented CS-in the MDN was 11.92, compared to 6.84 and 5.52 for the symmetrically and frequently presented CS-, respectively (Figure 7, right panel, fourth row). However, the probability manipulation did not yield significant effects in this site, again perhaps due to the relatively small sample of MDN records.

Unit responses to the CS+. Rare CS+ presentation did not augment the already ample responses to the symmetrically and frequently presented CS+ in the PCC, AVN and MDN. However, the rare CS+ did increase significantly the activity elicited in the ACC and DG. Individual comparisons following a significant three-way interaction in the analysis of the ACC data indicated that the response to the rarely presented CS+ exceeded the responses to the frequently and symmetrically presented CS+s from 50 to 400 ms after CS onset (Table 4 and Figure 7, upper left panel). Similarly, the unit responses in the DG to both the rarely and symmetrically presented CS+s exceeded the responses to the frequently presented CS+ in several 10 ms intervals from 170 to 390 ms (Table 4 and Figure 7, left panel, third row). Finally and somewhat paradoxically, the responses of the AVN to both the rare and the frequent CSs were reduced significantly relative to the response to the symmetrically presented CS+ at intervals from 170-300 ms (Figure 7, bottom left panel). A similar but nonsignificant reduction was also exhibited by the MDN (Figure 7, left panel, fourth row).

Macropotentials. In parallel with the unit results, both of the rare CSs elicited DG macropotential responses of greater magnitude than those elicited by the frequently and symmetrically presented CSs (Figure 8, row 3). These results were indicated by a significant interaction of stimulus and session ($F[2/8] = 5.29$, $p = .0344$). However, no individual comparisons attained acceptable significance levels for this effect, and none of the remaining intracranial records yielded significant macropotential changes in response to the standard AP conditions.

In contrast, robust effects of the probability manipulation occurred in the anterior and posterior surface (AS and PS) macropotentials, as indicated by significant 3-way interactions of stage, stimulus and interval (Table 5). The AS activity elicited by the rare and equal CS+ exeeded that elicited by the frequent CS+ at several intervals, from 60 to 140 ms after CS onset. These intervals correspond to the large positive peak in the AS record (Figure 8, upper left panels; Table 5). The rare and equal potentials in this instance were of nearly identical amplitude in these intervals. The PS potential elicited by the rare CS+ exceeded significantly the potentials elicited by both the equal and frequent CS+, again during the large positive wave from 60 to 200 ms (Figure 8, upper right panels).

The rare CS- also elicited significantly greater positive potentials than the frequent and equal CS- in the AS and PS, but this effect was of a small amplitude and it occurred in a rather limited range of intervals (Table 5).

4. Nonstandard AP sessions

Unit activity in the asymptotic training stage. The average unit responses were attenuated under all conditions of stimulation during the nonstandard AP sessions, relative to the responses in the standard AP sessions, perhaps due to the contextual alteration represented by the novel procedures, or to the refractoriness of the unit responses produced by the brief interstimulus interval. This attenuation notwithstanding, clear effects of the manipulation of stimulus probability and relevance were observed during these session.

The rare CS+ elicited significantly greater excitatory unit responses than the frequent CS- during the asymptotic training stage in all of the intracranial regions (Figure 9, third column). In all sites but the AVN, the rare CS- and the frequent CS+ elicited statistically equivalent, minimal responses (Figure 9, lower right panel). In the exceptional case, the AVN, the CS+ elicited a significantly greater response than the CS- regardless of the presentation frequency (Figure 9, upper right panel). A similar pattern occurred—numerically, but not significantly, in the other thalamic nucleus, the MDN. Thus, the combination of rareness and task relevance increased the neuronal responses in all areas.

These results were indicated by significant interactions of the stimulus and interval factors from separate analyses performed on the rare CS+ and the rare CS- sessions during the asymptotic training stage (Table 6). Separate analyses were performed because about half of the animals had only the rare CS+ session and the other half had both rare CS+ and rare CS- sessions. The effects of rare CS+ presentations in asymptotically trained rabbits occurred in the 170 to 400 ms range in all sites except the DG. In this site, the effect consisted of a greater inhibition of unit firing in response to the rare CS+ than to the frequent CS- from 100-140 ms, and a greater excitation to the rare CS+ than to the frequent CS- from 240-400 ms.

Macropotentials. The general pattern of the unit response during the nonstandard AP sessions (i.e., an enhancement of the response to the rare CS+) was also observed in certain macropotential records, including the two surface records, the DG and the MDN records (Figure 10). Although not significant, this pattern was also suggested by the average macropotentials in the AVN. In these records, the significant interactions and individual comparisons (Table 7) indicated that the macropotentials elicited by the CS+ in the rare CS+ sessions were of significantly greater amplitude than those elicited by the frequent CS-, but there were no significant amplitude differences in the sessions with the rare CS- and frequent CS+.

Activity during the pretraining stage. The basic effect of the probability manipulation during pretraining was the enhancement of thalamic neuronal activity to rarely presented stimuli (Figure 9, upper left panels). This effect occurred in the AVN, as indicated by significant interactions of the stimulus and interval factors, when the rare stimulus was the prospective CS+ ($F[39/195]=2.67$, $p<.0001$), and when it was the prospective CS- ($F[39/117]=5.67$, $p<.0001$). In addition, the effect occurred in the MDN when the rare stimulus was the prospective CS- ($F[39/117]=3.03$, $p<.0001$). Also, the unit response in the MDN to the rare prospective CS+ exceeded that to the frequent prospective CS- at every interval, but due to the small sample in this case, this difference was not significant in the analysis of variance. Finally, in contrast to the significant effects of the probability manipulation during pretraining in the thalamic structures, no significant effects occurred in the cortical structures (Figure 9, lower left panels). The manipulation of stimulus probability during the nonstandard AP treatments did not significantly affect the intracranial macropotentials in any of the recording sites during pretraining.

C. DETAILED RESULTS OF EXP 2: SUBICULAR LESIONS AND ACTIVITY IN THE PCC AND AVN.

1. Unit Activity.

In a recent study, we reported that subicular lesions reduced the magnitude of the average CS elicited unit response in the PCC, and they increased the magnitude of the overall and the discriminative unit response in the AVN (26,28). The reduced PCC response occurred in the early training stages, from pretraining to the session of criterion attainment; the unit response in the PCC increased gradually during training, such that the difference between the records in rabbits with lesions and in controls was nonsignificant during and after criterion attainment. The lesion-induced enhancement of the overall and discriminative unit response in the AVN reached its peak magnitude during criterion attainment and postcriterial overtraining. These results, in combination with the lesion-induced loss of the PCC unit response, form part of the empirical basis for a theoretical model of limbic system functional interactions during learning (29).

Here we present the data from the standard AP session (given immediately after overtraining) in which the CS+ was presented rarely, and the CS- frequently. These data are relevant to the possible role of hippocampal projections in governing the PCC and AVN neuronal response to the CS probability manipulation. Figure 11 shows the PCC unit responses in intact controls (upper left column) and in rabbits with lesions (upper right column) during the the standard AP session. The figure shows a reduced PCC unit response in the rabbits with lesions relative to controls. This effect was not significant, although as noted, the PCC responses in rabbits with lesions were significantly reduced in the earlier training stages.

The enhancement of the overall and discriminative AVN unit response in rabbits with lesions, present during overtraining, also occurred during the standard AP session as shown in the lower half of Figure 11. This conclusion was indicated by a significant interaction of the lesion, stimulus and interval factors ($\underline{F}[39/546]=2.77$, $\underline{p}<.0001$) in the analysis of the standard AP session data. Individual comparisons indicated no significant differences between the lesion and control unit responses to the CS-. However, the CS+ in rabbits with lesions elicited a significantly greater average unit response than the CS+ in controls from 160-400 ms after CS onset. Thus, the magnitude of the neuronal response to the rare CS+ in the AVN in rabbits with lesions was significantly greater than its magnitude in control rabbits. The fact that the response to the rare CS+ and frequent CS- presentation in the rabbits with lesions was not greater than the response to the equiprobably presented CS+ in the preceding overtraining session indicates that the response to the CS+ was already enhanced during this session, relative to the control response, and it was not further incremented by the rare presentation of the CS+ in the standard AP session. This result is perhaps to be expected from the data of Exp 1, indicating that AVN neurons in intact rabbits did not show a specific response to the rarely presented CS+ during overtraining.

2. Macropotentials

The overall macropotential amplitude and the magnitude of the discriminative macropotential response in the PCC were enhanced in the rabbits with lesions, relative to controls, during overtraining (Figure 12). However, no further alteration occurred during the following session in which the CS+ was presented rarely. The effects of the lesions during overtraining were indicated by a significant interaction of the lesion, stimulus and interval factors ($F[39/1326]=1.81$, $p<.0019$). Individual comparisons demonstrated enhanced responses in the brief-latency positive component from 60-80 ms after CS onset, and in the longer latency negative component from 170 to 400 ms. That is, the positive wave elicited in response to the CS+ was significantly greater than the positive wave elicited by the CS-, and the negative wave elicited by the CS+ was significantly greater than the negative wave elicited by the CS- (compare the profiles in the left and right upper panels in Figure 12). Both lesion and control rabbits exhibited enhanced average macropotentials in the PCC in response to the rare CS+, relative to the potentials elicited by the frequently presented CS- during the standard AP session that followed overtraining. This outcome was indicated by a significant interaction of stimulus and interval ($F[39/858]=8.79$; $p<.0001$). Again the enhancement occurred in the brief-latency positive deflection and in the longer latency negative deflection. However, the absence of any significant influence of the lesion factor in this analysis suggested that the presentation of the rare and relevant CS+ yielded a differential response in controls that was of the same magnitude, statistically, as the differential response in the rabbits with lesions (compare the profiles in the left and right lower panels in Figure 12).

Thus far, the data suggest that subicular lesions alter the overall and discriminative response magnitude, but not the response to rare stimulus presentation. The only disclaimer to this conclusion was yielded by the AVN macropotential data (Figure 12). Here, the subicular lesions increased the overall macropotential response of the AVN during overtraining as indicated by a significant interaction of the lesion and interval factors ($\underline{F}$[39/780]=5.03, $\underline{p}$<.0001, upper panels of Figure 13). Individual comparisons showed the average macropotential response to be significantly more positive 70 ms after CS onset, and significantly more negative from 150 ms to the end of the analysis period, in the rabbits with lesions, relative to controls. The absence of an effect of the stimulus factor in this analysis suggested that the lesions did not alter the magnitude of the discriminative response. Yet, both the overall and discriminative response were increased by the lesions during the standard AP session, as indicated by an interaction of lesion stimulus and interval ($\underline{F}$[39/546]=7.16, $\underline{p}$<.0001; lower panels, Figure 13). Individual comparisons indicated that there occurred a dynamic negative macropotential response to the rare CS+, and no response to the frequent CS- in the rabbits with lesions (lower right panel, Figure 13). In contrast, the AVNs in controls exhibited no trace of a discriminative macropotential response during the standard AP session (lower left panel, Figure 13). These data represent the only instance in this study of an altered neural sensitivity to the manipulation of CS probability in rabbits with lesions. The observed alteration is an unexpected enhancement of the probability response.

To summarize, the present results are in conformity with recent findings indicating that hippocampal projections to the AVN are involved in limiting the discriminative unit and macropotential responses in this structure, as well as the discriminative macropotential response in the PCC. The hippocampal projections may also limit the AVN macropotential response to rare CS+ presentation in intact animals. These data contradict the hypothesis that the neural response in the PCC and AVN to rare and relevant

CS presentations require intact projections from the hippocampus.

DISCUSSION

This investigation is directed toward establishing whether the study of neural activity during avoidance conditioning in rabbits will aid the search for the neural origins of ERPs in humans. The criterion that we apply to this question is the degree to which neural activity in rabbits exhibits functional properties that are analogous to ERPs in humans. It should be stressed that at this time we are looking for analogy in terms of the way in which the activity in people and in rabbits is altered by manipulation of the eliciting conditions. Whether the activities are also homologous (i.e., identical in terms of their evolutionary origins) remains to be seen.

Past studies of limbic neural activity, and the present data, demonstrate target effects, i.e., the development during training, of different neural responses to the positive CS, relative to responses elicited by the negative CS. Here we demonstrate in addition that the unit and macropotential responses in these structures are significantly altered by the manipulation of CS probability. Since target and probability effects are hallmarks of the P300, these data satisfy the primary criteria for the analogy that we are seeking. Furthermore, the fact that the target and probability effects occurred in terms of the firing of neurons localizes these effects to the monitored structures. The involved structures thus become candidates for neural generation of the human potentials.

The principal reason for seeking an animal model for the P300 is not only to demonstrate and localize activity analogous to the P300. Indeed, that objective is important primarily as a stepping stone to the programmatic invasive investigation of the neural causal antecedents and consequences of the analogous activity. This consideration brings to light an additional use of animal models: to provide a mapping of the central neural structures that exhibit ERP functional properties, and to document regional differences in that expression. If sensitivity to stimulus relevance and probability were to be exhibited identically in every structure, under all conditions of training and testing, the conduct of programmatic invasive research would not be greatly facilitated. More useful would be preliminary findings indicating the differential expression of phenomena analogous to the P300 in various areas of the brain. This differential expression was found in the present study: the effects of the probability manipulation depended on the recording site and the testing procedure. When the probability of the CSs was manipulated during standard AP sessions in fully trained rabbits, rare presentation of the CS- enhanced the unit discharge to that stimulus in all of the recording sites. However, rare presentation enhanced the response to the CS+ in only two of the sites, the ACC and the DG. One of the remaining sites (the PCC) showed no change in response to rare CS+ presentation, and the other two (the AVN and the MDN) showed suppression of unit activity in response to the rare CS+.

Past results provide some clues to the possible reasons for these regional differences. For example, consider the suppression of firing in the AVN and MDN in response to rare CS+ presentation (Figure 7). In the AVN, the rare and the frequent CS+s, both unusual stimuli, reduced the response relative to the response evoked by the CS+ when it and the CS- were equally probable. The response reduction in the MDN occurred relative to the response to the frequently presented CS+ and relative to the CS+ presented equally as often as the CS-. These reductions of the thalamic response can be viewed as being compatible with a recent theoretical model, which holds that the cingulate cortices exert a limiting influence on the thalamic nuclei with which they are interconnected (29). One of the factors that invokes this limiting influence is the occurrence of unexpected events (such as altered stimulus incidence), and the result of limiting is a prevention of pre-programmed behaviors. This prevention is one of several processes that are invoked, hypothetically, to promote behavioral quiescence and the direction of attention to the novel events. Numerous findings provide the inferential basis of this model, including the recent data indicating that the target effect in the AVN, and behavioral responding, are enhanced in rabbits with lesions in the PCC and also the hippocampus, the sole sources of cortical input to the AVN (26,28). The enhancement of behavioral responding appeared as a hyper-response during extinction training, when rabbits with intact corticothalamic connections exhibited suppressed behavioral and AVN responses. Experiments to determine whether the MDN response is enhanced in rabbits with damage in the reciprocally interconnected ACC are currently in progress. In any event, the basic premise that arises from these data is that the limiting process may be responsible for the absence of unit response increments in the AVN and the MDN under conditions of rare CS+ presentation.

The situation is somewhat more complex with regard to the PCC. The layer VI (corticothalamic) neurons in this region should, hypothetically, be the origins of the limiting influence. Thus, they should be activated by the presentation of unexpected events. However, the PCC is also directly activated by afferents from the AVN, the region that is limited when unexpected events occur. Since the AVN is somewhat suppressed by virtue of the limiting effect, the net effect in the PCC could be no significant change in the response due to the alteration of CS+ probability. This is indeed the outcome that we observed. Nevertheless, the nonsignificant increase of the PCC unit response to the rare and frequent CSs, relative to the equal CS, are in accord with the idea that this region is the site of the neuronal activity that limits the firing of the AVN.

The unit responses to the CS- underwent a gradual reduction, as the target effect developed during training. However, the reduction was reversed and the response to the CS- dramatically enhanced in all of the monitored areas, during the sessions in which the CS- was presented rarely. Yet, in no case did the frequent presentation of the CS- also lead to an enhancement of the evoked response. We can say with some confidence that the enhancement of the response to the CS- in the thalamic sites was not due to a lessening of the limiting process. The PCC and subicular lesions that yielded the phenomenon (enhancement of the AVN response) giving rise to the concept of a limiting process, did not increase the AVN response to the CS-, they only increased the response to the CS+ (op. cit.). These results, and the absence of any response enhancement due to frequent CS- presentation, suggest that the increased response to the rare CS- was not due to a lessening of the limiting influence from the limbic cortex. At present we do not know the origin of this effect.

In contrast to the AVN and the MDN, the ACC and the DG have been observed in past studies to exhibit target effects predominantly in the early stages of acquisition and, in the case of the DG and other hippocampal areas, in response to the alteration of familiar training contingencies (29). These sites exhibit a lessening of the target responses during the asymptotic stages of training (25,29,50). The early and temporary nature of the responses in these structures suggests that these structures are involved in processes that are important during the initial stages of acquisition, and during adaptation to the occurrence of novel task contingencies, but that they are not particulary important in relation to performance of the well-practiced habit. Thus, although we cannot at present provide a mechanistic account of their functions, it is perhaps not surprising that rare presentation enhanced the response to both CS+ and CS- in these structures. Indeed, the available data suggest that these structures should be quite sensitive to novel events, as they appear to be involved in the neural reprogramming that is invoked whenever the subject is faced with the need to adapt to novel task contingencies.

Different effects of the probability manipulation occurred during the nonstandard AP sessions, in which the tones were presented at a 1/sec. rate, with response prevention and no US. In these sessions, the neuronal responses in all of the recording sites were attenuated relative to the response magnitudes during the standard training and AP test sessions. Nevertheless, the combination of rareness and behavioral relevance was associated with significantly greater unit and/or macropotential responses than any of the remaining three combinations of stimulus properties. Perhaps only the rare, relevant stimuli had sufficient eliciting capacity to override the attenuating effects of the unusual and unexpected properties of the nonstandard sessions. Thus, the different effects of the nonstandard AP treatment, relative to the standard AP treatment, may be attributable solely to the disruptive effects of the nonstandard treatments, rather than some special variety of information processing that is invoked by one or more of the parameters of the nonstandard sessions.

The present results indicate that the effects of relevance and probability manipulation are found ubiquitously throughout the limbic telencephalon and diencephalon. Whereas there is no reason to assume that all of these limbic sites contribute to the P300 in humans, these results add plausibility to the hypothesis that the ERPs recorded from the human scalp are manifestations of information processing activities that are rather widely distributed in the brain, rather than processes originating in a single structure, such as the hippocampus. The widespread nature of these effects in the rabbit is in agreement with suggestions from other recent studies in animal and human subjects (43,51,71). Nevertheless, it is quite possible that only a very limited portion of the widespread intracranial activity has the necessary biophysical properties that would allow its activity has the necessary biophysical properties that would allow its fields to be recorded on the scalp. The resolution of this issue depends on the concurrent study of multiple intracranial locations and the scalp sites. Yet, the present and past work with the rabbit preparation do seem to provide some important information in regard to this issue.

It is now well-known that the cingulate cortical excitatory discharges that occur from 100-400 ms after CS onset are critically dependent on afferents from the limbic thalamic nuclei: lesions in these nuclei eliminate all CS elicited excitatory and discriminative cingulate cortical activity after 50 ms following CS onset (7,24). In the present study, clear neuronal sensitivity to stimulus probability and relevance occurred in the thalamic nuclei, and these effects also occurred in the cingulate cortical areas driven by the limbic thalamic afferents. Also, as we have seen from past work and from the data of the present Experiment 2, the thalamic manifestations of sensitivity to stimulus relevance and probability are not dependent on intact connections from the cortical areas (26,28). These considerations, in concert, raise the possibility that the sensitivities to stimulus probability and stimulus relevance manifested by surface recordings made on the human scalp may be a product of processes of the limbic and other nonspecific nuclei of the thalamus. It is after all the case that the robust volleys of neuronal action potentials and ERPs recorded at relatively brief latency from the cortex and from the scalp reflect in large part the massive driving of cortical cells by thalamocortical axons. Thus, it seems quite plausible to consider the possibility of a thalmic origin for the P300. Of course, this view does not deny the possibility that the P300 recorded from the scalp, although driven largely by thalamic cells, can nevertheless by modulated importantly by cortico-cortical influences, such as those that originate in the subiculum of the hippocampal formation. In fact, the model of limbic participation in the avoidance task that we have mentioned previously associates the occurrence of rare stimuli with the elicitation of a limiting process, represented by cingulate cortical unit activity that is thalamically driven, but that also requires an input to cingulate cortex from the hippocampus. Indeed, the macropotential responses in the hippocampal dentate gyrus elicited by rare CS presentations exhibited greater amplitude and duration than the potentials in any of the other studied areas, and the latency of the response to the rare CS+ was briefer than in the other areas.

The assertion that responses analogous to the P300 may depend on the output of the hippocampal formation may seem to be at odds with other data (51), and with the results of the present Exp 2, suggesting that neural responses to stimulus relevance and probability manipulations are retained and possibly enhanced in animals with hippocampal damage. However, in our view, these propositions are not necessarily incompatible. The P300 in intact humans may represent primarily the evocation of the limiting process brought about by the interaction of thalamic and hippocampal influences within the cortical areas that receive hippocampal projections. In animals with hippocampal damage, the activity analogous to the P300 may reflect primarily thalamocortical neural volleys in these same cortical areas in response to rare and/or relevant stimuli. The fact that these responses are enhanced in the brain-damaged animals may indeed be due to the absence of the limiting process, the very process that is responsible for the effect in the intact human.

These suggestions must of course be tempered by the observation that the morphology of the ERPs recorded from the brain surface in the rabbit, which is sensitive to event probability, is quite distinct from the morphology of the human P300. The potentials in the rabbit are negative-going in the very same latency range in which the potentials in humans are positive. A morphology similar to the P300 appears in the rabbit much earlier, from 100 to 200 ms after stimulus onset. At present, it is impossible to state which of the potentials recorded in rabbits, if any, is the better analog of the human P300.

In this connection it should be noted that there is no reason to believe that the morphology of the waves should be similar in two such different species. It is possible that the negative potential at 300 ms is analogous to the positive potential at that latency in humans and that the difference in polarity is the result of relatively trivial difference in the orientation of the electrical sources, or some other trivial difference between humans and animals. On the other hand, it is possible that the neural activity that gives rise to the P300 in humans may originate in a system, such as the limbic system, that begins to exhibit sensitivity to stimulus probability and relevance long before 300 ms. These manifestations may not appear on the human scalp until 300 ms has elapsed because of various fortuitous voltage cancellation effects that occur in the earlier latencies. Such effects are in fact likely given the prevalence of several cascaded excitatory and inhibitory discharges that characterize stimulus processing in a variety of neural structures in the brief latency domain. In this connection it is relevant to note that the occurrence of probability and stimulus relevance effects at latencies considerably under 300 ms has been found in studies of human and animal intracranial potentials (4,46,46,71). These data favor the second interpretation outlined above. However this issue is ultimately resolved, our data clearly establish that the rabbit can be used as a subject in the pursuit of a model of the P300. This is welcome news because the rabbit affords a most convenient preparation for investigating the neuroanatomy and functional neurophysiology of the phenomena analogous to the P300.

REFERENCES

1 Arthur, D.L. and Starr, A., Task-relevant late positive component of the auditory event-related potential in monkey resembles P300 in humans, Science, 223 (1984) 186-188.

2 Brogden, W.J. and Culler, F.A., A device for motor conditioning of small animals, Science, 83 (1936) 269.

3 Buchwald, J., Holstein, S.B. and Weber, D.S., Multiple-unit recording: Technique, interpretation, and experimental applications. In R. Thompson and M. Patterson (Eds.), Bioelectric Recording Techniques: Part A, Cellular Processes and Brain Potentials, Academic, New York, 1973, pp. 202-242.

4 Buchwald, J.S. and Squires, N., Endogenous auditory potentials in the cat: A P300 model. In C.D. Woody (Ed.), Conditioning, Plenum, New York, 1982, pp. 503-515.

5 Bures, J., Petran, M. and Zachar, J., Electrophysiological Methods in Biological Research, Academic, New York, 1967.

6 Callaway, E., Tueting, P. and Koslow, S.H., Event-related Brain Potentials in Man, Academic, New York, 1978.

7 Colletier, P. Tenner, J. Sparenborg, S.P. and Gabriel, M., Separate and combined medial and anterior thalamic lesions and cingulate cortical neuronal correlates of avoidance learning in rabbits, Society for Neuroscience Abstracts, 11 (1985) 836.

8 Deadwyler, S.A., West, M.O., Christian, E.P., Hampson, R.E. and Foster, T.C., Sequence-related changes in sensory-evoked potentials in the dentate gyrus: A mechanism for item-specific short-term information storage in the hippocampus, Behavioral and Neural Biology, 44 (1985) 201-212.

9 Desmedt, J.E., Scalp-recorded cerebral event-related potentials in man as point of entry into the analysis of cognitive processing. In F.O. Schmitt (Ed.), The Organization of the Cerebral Cortex, MIT, Cambridge, 1981, pp. 441-473.

10 Desmedt, J.E. and Debecker, J., Wave form and neural mechanism of the decision P350 elicited without pre-stimulus CNV or readiness potential in random sequences of near-threshold auditory clicks and finger stimuli, Electroencephalography and Clinical Neurophysiology, 47 (1979) 648-670.

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

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

13 Donchin, E. and Isreal, J.B., Event-related potentials and psychological theory. In H.H. Kornhuber and L. Deecke (Eds.), Motivation, Motor and Sensory Processes of the Brain: Electrical Potentials, Behavior and Clinical Use. Progress in Brain Research, Elsevier-North Holland, Amsterdam, 1981, pp. 697-715.

14 Donchin, E., Karis, D., Bashore, T.R., Coles, M.G.H. and Gratton, G., Cognitive psychophysiology and human information processing. In M.G.H. Coles, E. Donchin and S. Porges (Eds.), Psychophysiology: Systems, Processes, and Applications, Guilford, New York, 1986, pp. 244-267.

15 Donchin, E., Otto, D.A., Gerbrandt, L.K. and Pribram, K.H., While a monkey waits: Electrocortical events recorded during the foreperiod of a reaction time study, Electroencephalography and Clinical Neurophysiology, 31 (1971) 115-127.

16 Donchin, E., Ritter, W. and McCullum, C., Cognitive psychophysiology: The endogenous components of the ERP. In E. Calloway, P. Tueting and S.H. Koslow (Eds.), Event-related Brain Potentials in Man, Academic, New York, 1978, pp. 349-441.

17 Donovick, P.J., A metachromatic stain for neural tissue, Stain Technology, 49 (1974) 49-51.

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

19 Farley, G.R. and Starr, A., Middle and long latency auditory evoked potentials in cats. I. Component definition and dependence on behavioral factors, Hearing Research, 10 (1983) 117-138.

20 Foster, K., Orona, E., Lambert, R.W. and Gabriel, M., Early and late acquisition of discriminative neuronal activity during differential conditioning in rabbits: Specificity within the laminae of cingulate cortex and the anteroventral thalamus, Journal of Comparative and Physiological Psychology, 94 (1980) 1069-1086.

21 Fox, C.A. and Eichman, J., A rapid method for locating intracerebral electrode tracks, Stain Technology, 34 (1959) 39-42.

22 Gabriel, M., Foster, K. and Orona, E., Interaction of the laminae of cingulate cortex and the anteroventral thalamus during behavioral learning, Science, 208 (1980a) 1050-1052.

23 Gabriel, M., Foster, K., Orona, E., Saltwick, S. and Stanton, M., Neuronal activity of the cingulate cortex, anteroventral thalamus and hippocampal formation in discriminative conditioning: Encoding and

extraction of the significance of conditional stimuli. In J. Sprague and A.N. Epstein (Eds.), Progress in Physiological Psychology and Psychobiology, Vol. 9, Academic, New York, 1980b, pp. 125-231.

24 Gabriel, M., Lambert, R.W., Foster, K., Orona, E., Sparenborg, S. and Maiorca, R.R., Anterior thalamic lesions and neuronal activity in the cingulate and retrosplenial cortices during discriminative avoidance behavior in rabbits, Behavioral Neuroscience, 97 (1983a) 675-696.

25 Gabriel, M. and Orona, E., Parallel and serial processes of the prefrontal and cingulate cortical systems during behavioral learning, Brain Research Bulletin, 8 (1982) 781-785.

26 Gabriel, M. and Sparenborg, S., Anterior thalamic discriminative neuronal responses enhanced during learning in rabbits with subicular and cingulate cortical lesions, Brain Research, 384 (1986a) 195-198.

27 Gabriel, M., Sparenborg, S.P. and Donchin, E., Macropotentials recorded from the cingulate cortex and anterior thalamus in rabbits during the "oddball" paradigm used to elicit P300 in humans, Society for Neuroscience Abstracts, 9 (1983b) 1200.

28 Gabriel, M., Sparenborg, S. and Stolar, N., Hippocampal control of cingulate cortical and anterior thalamic information processing during learning in rabbits, Experimental Brain Research, 1987, (in press).

29 Gabriel, M., Sparenborg, S. and Stolar, N., An executive function of the hippocampus: Pathway selection for thalamic neuronal significance code. In R.L. Isaacson and Pribram K (Eds.), The Hippocampus: Volume IV, Plenum, New York, 1986b, pp. 1-39.

30 Gaillard A.W.K. and Ritter, W. (Eds.), Tutorials in ERP Research: Endogenous Components, North Holland, Amsterdam, 1983.

31 Glover, A.A., Onofrj, M.C., Ghilardi, M.F. and Bodis-Wollner, I., P300-like potentials in the normal monkey using classical conditioning and an auditory "oddball" paradigm, Electroencephalography and Clinical Neurophysiology, 65 (1986) 231-235.

32 Gray, J.A., The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-hippocampal System, Oxford University, Oxford, 1982.

33 Halgren, E., Squires, N.K., Wilson, C.L., Rohrbaugh, S.W., Babb, T.L. and Crandall, P.H., Endogenous potentials generated in the human hippocampal formation and amygdala by infrequent events, Science, 210 (1980) 803-805.

34 Hari, R., Aittoniemi, K., Jarvinen, M.-L., Katila, T. and Varpula, T., Auditory evoked transient and sustained magnetic fields of the human brain: Localization of neural generators, Experimental Brain Research, 40 (1980) 237-240.

35 Harrison, J and Buchwald, J., Cat "P300": Effects of association cortex ablation, Society for Neuroscience Abstracts, 12 (1986) 1131.

36 Harrison, J., Buchwald, J. and Kaga, K., Cat P300 present after primary auditory cortex ablation, Electroencephalography and Clinical Neurophysiology, 63 (1986) 180-187.

37 Harrison, J., Buchwald, J., Kaga, K., Woolf, N. and Butcher, L., Cat "P300": Effects of septal lesions, Society for Neuroscience Abstracts, 11 (1985) 1234.

38 Johnson, R. and Fedio, P., ERP and P300 activity in patients following unilateral temporal lobectomy, Society for Neuroscience, 10 (1984) 847.

39 Kaga, K., Buchwald, J. and Harrison, J., Cat "P300": Effects of

hippocampal lesions, Society for Neuroscience Abstracts, 11 (1985) 1234.

40 Katayama, Y., Tsukiyama, T and Tsubokawa, T., Thalamic negativity associated with the endogenous late positive component of cerebral evoked potentials (P300): Recordings using discriminative aversive conditioning in humans and cats, Brain Research Bulletin, 14 (1985) 223-236.

41 Kutas, M. and Hillyard, S., Reading senseless sentences: Brain potentials reflect semantic incongruity, Science, 207 (1980) 203-205.

42 McCarthy, G. and Wood, C.C., Intracranial recordings of endogenous ERPs in humans, Electroencephalography and Clinical Neurophysiology, 1987, in press.

43 McCarthy, G., Wood, C.C., Allison, T., Goff, W.R., Williamson, P.D. and Spencer, D.D., Intracranial recordings of event-related potentials in humans engaged in cognitive tasks, Society for Neuroscience Abstracts, 8 (1982) 976.

44 Neville, H.J. and Foote, S.L., Auditory event-related potentials in the squirrel monkey: Parallels to human late wave responses, Brain Research, 298 (1984) 107-116.

45 O'Brien, J.H., P300 wave elicited by a stimulus-change paradigm in acutely prepared rats, Physiology and Behavior, 28 (1982) 711-713.

46 O'Connor, T. and Starr, A., Generator of P300 in the cat, Society for Neuroscience Abstracts, 9 (1983) 1195.

47 O'Connor, T. and Starr, A., Intracranial potentials correlated with an event-related potential, P300, in the cat, Brain Research, 339 (1985) 27-38.

48 Okada, Y.C., Kaufman, L. and Williamson, S.J., The hippocampal formation

as a source of the slow endogenous potentials, Electroencephalography and Clinical Neurophysiology, 55 (1983) 417-426.

49 Olton, D.S., Becker, J.T. and Handelmann, G.E., Hippocampus, space, and memory, Behavioral and Brain Sciences, 2 (1979) 313-365.

50 Orona, E. and Gabriel, M., Multiple-unit activity of the prefrontal cortex and the mediodorsal thalamic nucleus during acquisition of discriminative avoidance behavior in rabbits, Brain Research, 263 (1983) 295-312.

51 Paller, K.A., Zola-Morgan, S., Squire, L.R. and Hillyard, S.A., Monkeys with lesions of hippocampus and amygdala exhibit event-related brain potentials that resemble the human P300 wave, Society for Neuroscience Abstracts, 10 (1984) 849.

52 Pineda, J.A., Foote, S.L. and Neville, H.J., Effects of noradrenergic locus coeruleus lesions on squirrel monkey event-related potentials, Society for Neuroscience Abstracts, 12 (1986) 1136.

53 Pribram, K.H. and McGuinness, D., Arousal, activation, and effort in the control of attention, Psychological Review, 82 (1975) 116-149.

54 Pritchard, W.S., Psychophysiology of P300, Psychological Bulletin, 89 (1981) 506-540.

55 Richer, F., Johnson, R.A. and Beatty, J., Sources of late components of the brain magnetic response, Society for Neuroscience Abstract, 9 (1983) 656.

56 Rosler, F., Endogenous ERPs and cognition: Probes, prospects and pitfalls in matching pieces of the mind-body puzzle. In A.W.K. Gaillard and W. Ritter (Eds.), Tutorials in ERP Research: Endogenous Components, North

Holland, Amsterdam, 1983, pp. 9-36.

57 Ruchkin, D.S. and Sutton, S., Positive slow wave and P300: Association and disassociation. In A.W.K. Gaillard and W. Ritter (Eds.), Tutorials in ERP Research: Endogenous Components, North Holland, Amsterdam, 1983, pp. 233-250.

58 Shek, J.W., Wen, G.Y., and Wisniewski, H.M., Atlas of the Rabbit Brain and Spinal Cord, Karger, Basel, 1986.

59 Squires, N.K. and Buchwald, J.S., Endogenous auditory evoked potentials in the cat, Society for Neuroscience Abstracts, 7 (1981) 452.

60 Squires, N.K., Halgren, E., Wilson, C. and Crandall, P., Human endogenous limbic potentials: cross-modality and depth/surface comparisons in epileptic subjects. In A.W.K. Gaillard and W. Ritter (Eds.), Tutorials in ERP Research: Endogenous Components, North Holland, Amsterdam, 1983, pp. 217-232.

61 Stapleton, J.M., Derrick, B.E., Crandall, P.H. and Halgren, E., Endogenous limbic potentials in humans: Component and site specificity, Society for Neuroscience Abstracts, 9 (1983) 1199.

62 Stolar, N., Effects of target and probability on P300-like activity recorded from limbic structures in rabbits. Unpublished thesis, University of Illinois, Urbana, Ill., 1987.

63 Sutton, S., Braren, M., Zubin, J. and John, E.R., Evoked potential correlates of stimulus uncertainty, Science, 150 (1965) 1187-1188.

64 Vinogradova, O.S., Functional organization of the limbic system in the process of registration of information: Facts and hypotheses. In R.L. Isaacson and K.H. Pribram (Eds.), The Hippocampus: Vol. II, Plenum, New

York, 1975, pp. 1-99.

65 Vogt, B.A., Sikes, R.W., Swadlow, H.A. and Weyand, T.G., Rabbit cingulate cortex: Cytoarchitecture, physiological border with visual cortex, and afferent cortical connections of visual, motor, postsubicular, and intracingulate origin, Journal of Comparative Neurology, 248 (1986) 74-94.

66 Walker, H.M. and Lev, J., Statistical Inference, Holt, New York, 1953.

67 Wilder, M.B., Farley, G.R. and Starr, A., Endogenous late positive component of the evoked potential in cats corresponding to P300 in humans, Science, 211 (1981) 605-607.

68 Winer, B.J., Statistical Principles in Experimental Design, McGraw-Hill, New York, 1962.

69 Wood, C.C. and Allison, T., Interpretation of evoked potentials: A neurophysiological perspective, Canadian Journal of Psychology/Review of Canadian Psychology, 35 (1981) 113-135.

70 Wood, C.C., Allison, T., Goff, W.R., Williamson, P.D. and Spencer, D.D., On the neural origin of P300 in man, Progress in Brain Research, 54 (1980) 51-56.

71 Wood, C.C. and McCarthy, G., A possible frontal lobe contribution to scalp P300, Society for Neuroscience Abstracts, 11 (1985) 879.

72 Wood, C.C., McCarthy, G., Allison, T., Goff, W.R., Williamson, P.D. and Spencer, D.D., Endogenous event-related potentials following temporal lobe excisions in humans, Society for Neuroscience Abstracts, 8 (1982) 976.

73 Wood, C.C., McCarthy, G., Squires, N.K., Vaughan, H.G., Woods, D.L. and McCallum, W.C., Anatomical and physiological substrates of event-related potentials: Two case studies. In R. Karrer, J. Cohen and P. Tueting (Eds.),

Brain and Information: Event-related Potentials, Annals of the New York Academy of Science, 425 (1984) 681-721.

74 Yingling, C.D. and Hosobuchi, Y., A subcortical correlate of P300 in man, Electroencephalography and Clinical Neurophysiology, 59 (1984) 72-76.

FIGURE CAPTIONS

Figure 1. A set of six coronal sections of the rabbit brain illustrating the minimal (dark areas) and maximal (hatched areas) extents of the subicular lesions. (AP = anterior/posterior location of the depicted sections in millimeters.) The depictions are derived from the atlas of Fifkova & Marsala, presented in Bures et al., 1967.

Figure 2. Average of the unit responses to the CS- during the standard AP (asymmetric probability) sessions. The responses in each recording site were elicited in trained rabbits during separate sessions in which the CS- was presented in .2, .5 or .8 of the trials. The plotted values were obtained by averaging the response magnitudes, in the form of z-scores, of a set of 10 consecutive 10-msec intervals containing the peak response of each of the sites. The 10-msec interval between 80 and 400 msec in which the z-score was the first, after CS onset, to fall within the 95% confidence interval of the largest z-score for a given profile was selected as the initial interval of the window. The recording sites are indicated by the arrows referencing an artistic depiction of a sagittal section of the rabbit brain at approximately 1.5 mm from the midline. The depiction is based on the atlas of Shek, Wen and Wisniewski, 1986.

Figure 3. Average of the unit responses to the CS+ during the standard AP (asymmetric probability) sessions. The responses in each recording site were elicited in trained rabbits during separate sessions in which the CS+ was presented in .2, .5 or .8 of the trials. The artistic depiction of the sagittal brain section and the method for obtaining plotted values are described in the legend of Figure 2.

Figure 4. The average unit responses during the non-standard AP sessions that followed the standard overtraining sessions. The reponses were

elicited following the presentation of the CS+ and the CS- in asymmetric proportions (.2/.8: [Rare CS+/Frequent CS-], or, .8/.2: [Frequent CS+/Rare CS-].) The artistic depiction of the sagittal brain section and the method for obtaining the plotted values are described in the legend of Figure 2.

Figure 5. Average unit responses in the form of z-scores derived from the intracranial sites during pretraining (left column) and during the asymptotic training stage ("trained"; right column.) The responses were elicited following presentations of the CS+ (solid line) and the CS- (dashed line) in equal (.5/.5) proportions. (Data are shown for each of the recording sites in 40 consecutive 10-msec intervals.)

Figure 6. Average macropotential responses in the form of z-scores derived from the records of the two brain (epidural) surface sites (two upper rows) and four intracranial sites during pretraining (left column) and during the asymptotic training stage ("Trained"; right column). The responses were elicited following presentations of the CS+ (solid line) and CS- (dashed line) presented in symmetric (.5/.5) proportions. Data are shown for six recording site in 40 consecutive 10-msec intervals. Positive voltage responses are indicated by downward deflections. The average macropotentials obtained from the medial dorsal nucleus, very similar to those shown in the figure for the anterior ventral nucleus, were thus omitted for brevity.

Figure 7. Average unit responses in the form of z-scores derived from the five intracranial records during the standard AP (asymmetric probability) sessions. Data are shown for each recording site in 40 consecutive 10-msec intervals. The responses were elicited by the CS+ (left column) and CS- (right column) in symmetric (.5/.5) and asymmetric (.2/.8 and .8/.2) proportions. The responses elicited by the rare, equal and frequent CSs are indicated by the solid, dashed and dotted lines respectively.

Figure 8. Average macropotential responses in the form of z-scores derived from the two surface records (top row) and from the five intracranial recording sites during the standard AP (asymmetric probability) sessions. Data are shown for each recording site in 40 consecutive 10-msec intervals. The responses were elicited by the CS+ (first and third columns) and CS- (second and fourth columns) in symmetric (.5/.5) and asymmetric (.2/.8 and .8/. 2) proportions. The responses elicited by the rare, equal and frequent CSs are indicated by the solid, dashed and dotted lines respectively.

Figure 9. Average unit responses in the form of z-scores derived from records of the anterior ventral thalamic nucleus and the anterior cingulate cortex during the non-standard AP training procedure which followed pretraining and asymptotic training ("trained"). The responses were elicited by the CS+ (solid lines) and CS- (dashed lines) presented in asymmetric proportions (.2/.8: [Rare CS+] or .8/.2: [Rare CS-]). The data of the posterior cingulate cortex and medial dorsal nucleus, very similar respectively to the data of the anterior cingulate cortex and anterior ventral thalamic nucleus in the figure, were thus omitted for brevity.

Figure 10. Average macropotential responses in the form of z-scores derived from the two surface records (top row) and from the five intracranial recording sites during the nonstandard AP (asymmetric probability) sessions. Data are shown for each recording site in 40 consecutive 10-msec intervals. The responses were elicited by the CS+ (solid lines) and CS- (dashed lines) in asymmetric proportions. These proportions were .2/.8 (rare CS+; left columns) and .8/.2 (rare CS-; right columns).

Figure 11. Average unit responses in the form of z-scores derived from the records of the posterior cingulate cortex (upper row) and the anterior ventral thalamic nucleus (lower row) in controls (left column), and in

rabbits with subicular lesions (right column). The responses were elicited by the rarely presented CS+ (solid bars) and by the frequently presented CS- (open bars) during the standard AP session presented on the day after the final (third) overtraining session in EXP 2.

Figure 12. Average macropotential responses in the form of z-scores derived from the records of the posterior cingulate cortex in controls (left column) and in rabbits with subicular lesions (right column) during overtraining (upper row) and during the standard AP session (lower row) given on the next day. The responses were elicited by the CS+ (solid lines) and CS- (dashed lines) in symmetric (.5/.5) and asymmetric (.2/.8) proportions.

Figure 13. Average macropotential responses in the form of z-scores derived from records of the anterior ventral thalamic nucleus in controls (left column) and in rabbits with lesions of the dorsal subciular complex (right column). The data were obtained during overtraining (upper row) and a subsequent standard AP session (lower row). The responses were elicited following presentation of the CS+ (solid lines) and CS- (dashed line) in symmetric (.5/.5) and asymmetric (.2/.8) proportions.

TABLE CAPTIONS

Table 1. The number of unit and macropotential records are given for each brain area and training stage. In the case of the nonstandard AP (asymmetric probability) sessions (right portion of table), the number of unit records equaled the number of macropotential records within each cell, except in the two cells in which two entries are given. In these cells, the first number represents the unit records and the second, the macropotentials.

Table 2. UNIT ACTIVITY DURING STANDARD TRAINING: Probability levels, F ratios, and statistically significant ($p < .05$) individual comparisons derived from analyses of the interactions of training stage X interval and training stage X stimulus X interval in the unit records of five recording sites during standard training. The intervals (in milliseconds) in which responses were significantly greater during the asymptotic training stage than during pretraining are listed in the third column. Intervals containing greater responses to the CS+ than to the CS- are listed in the fifth and sixth columns. (N = number of records, NS = not significant, * = the response to the CS- exceeded the response to the CS+; see Table 1 for the recording site labels.)

Table 3. INTRACRANIAL MACROPOTENTIALS DURING STANDARD TRAINING: Probability levels, F ratios, and statistically significant ($p < .05$) individual comparisons derived from analyses of the interactions of training stage X interval, and training stage X stimulus X interval in the macropotential records of the PCC, DG, AS and PS during standard training. The intervals (in milliseconds) in which the potentials were significantly increased during the asymptotic training stage relative to the pretraining stage, are given in the third column. Intervals in which the potentials elicited by

the CS+ were significantly greater than the potentials elicited by the CS- are given in the fifth and sixth columns. (*=the potentials elicited by the CS- exceeded those elicited by the CS+; N=number of records; NS=not significant.)

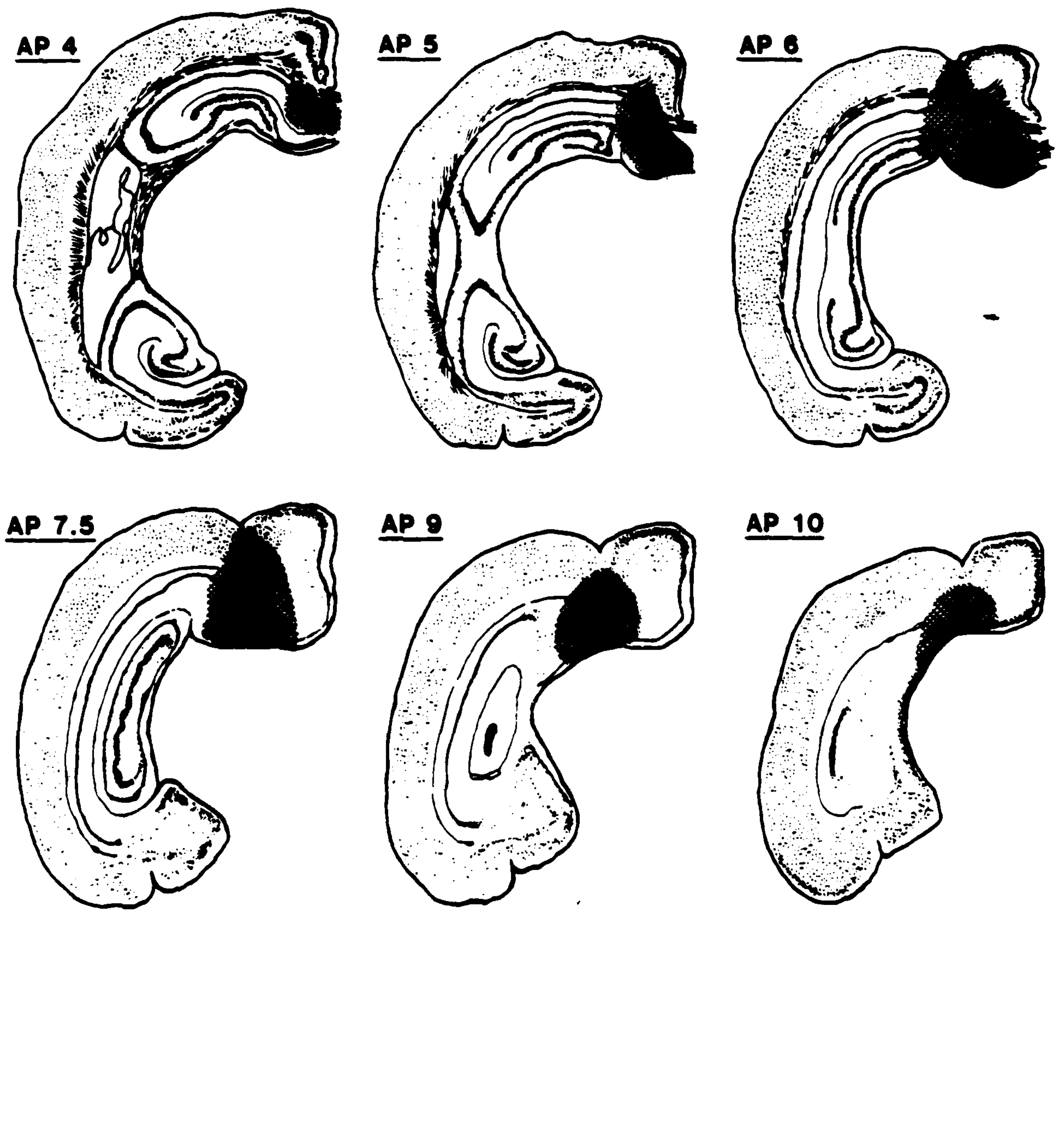
Table 4. UNIT ACTIVITY DURING THE STANDARD AP SESSIONS: Probability levels, *F* ratios, and statistically significant ($p < .05$) individual comparisons derived from analyses of the interactions of training stage X stimulus, and training stage X stimulus X interval, in the unit records of the five recording sites during the standard AP session. The intervals (in milliseconds) in which responses were significantly altered by the probability manipulation are listed for each comparison. Significantly greater responses occurred in the listed intervals in response to CSs presented at the frequency indicated by the first of the two proportions (to the left of the slashes), than to the CSs presented at the frequency indicated by the second of the two proportions (to the right of the slashes) given as the column headings. For example, the first entry in the fourth column of the table indicates that the CS+ presented on .2 of the trials elicited significantly greater responses than the CS+ presented on .8 of the trials from 50 to 400 ms after CS onset. Asterisks indicate a reversal of this relation. That is the CSs presented at the frequency indicated by the second of the two proportions pair elicited greater responses than CSs presented at the frequency indicated by the first proportion of the pair. (N=number of records; NS=not significant.)

Table 5. SURFACE MACROPOTENTIALS DURING THE STANDARD AP SESSIONS: Probability levels, *F* ratios, and statistically significant ($p < .05$) individual comparisons derived from analyses of the training stage X stimulus X interval interaction in the macropotential records of the surface recordings during the standard AP sessions. The intervals (in milliseconds)

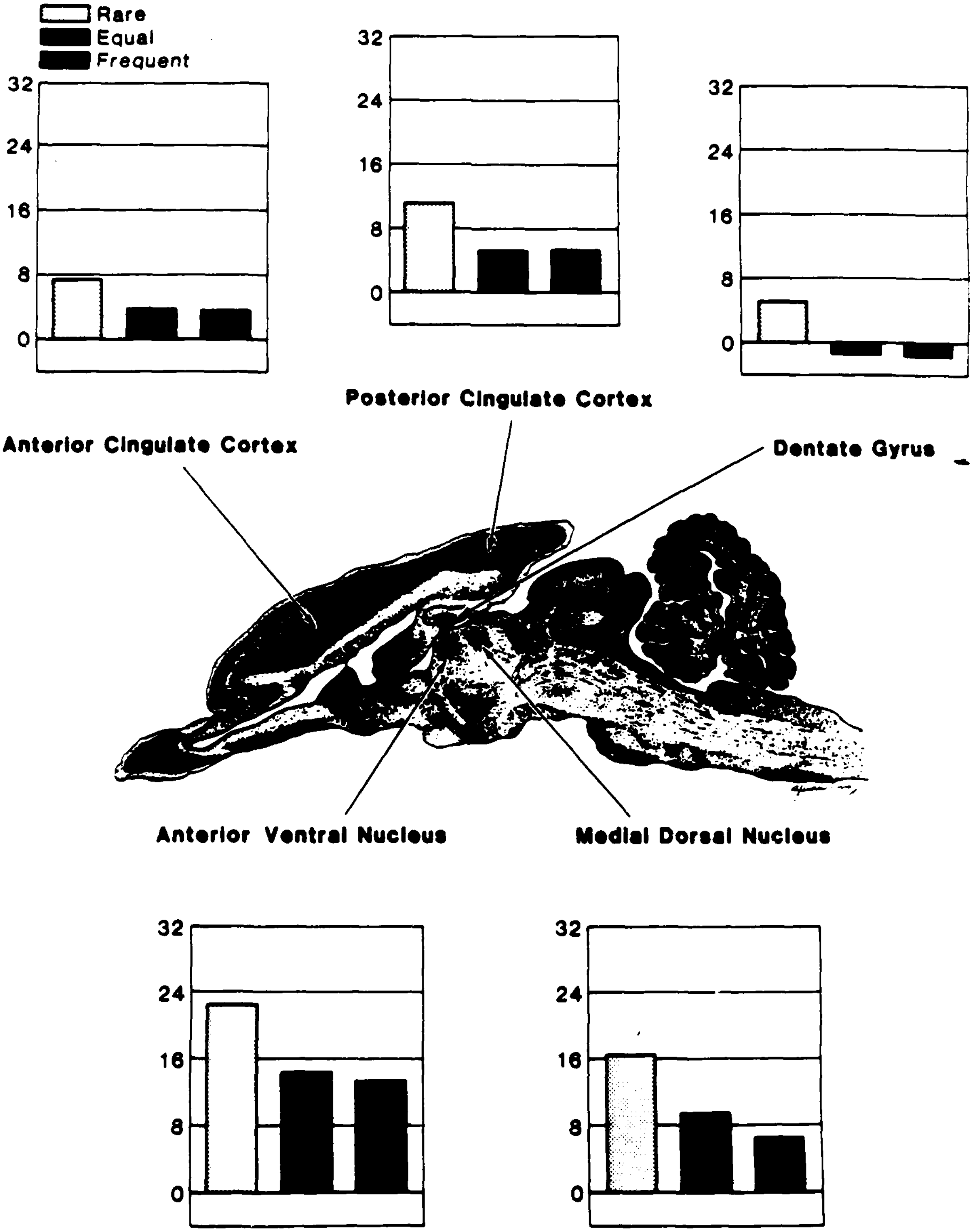
in which the potentials were significantly altered by the probability manipulation are listed for each comparison. See caption of Table 5 for a description of the abbreviations and symbols.

Table 6. UNIT ACTIVITY DURING THE NON-STANDARD AP SESSIONS: Probability levels, *F* ratios, and statistically significant ($p < .05$) individual comparisons derived from analyses of the stimulus X interval interaction in the unit records during the non-standard AP sessions. Data are presented separately for sessions in which the rare stimulus was the CS+, and for sessions in which the rare stimulus was the CS-. The intervals (in milliseconds) in which responses were significantly greater to the CS+ than to the CS- are listed in the third and sixth columns. (*=the response to the CS- exceeded the response to the CS+; N=number of records; NS=not significant.)

Table 7. MACROPOTENTIALS DURING THE NONSTANDARD AP SESSIONS: Probability levels, *F* ratios, and statistically significant ($p < .05$) individual comparisons derived from analyses of the stimulus type by interval interaction in the macropotential records of the MDN and the surface recordings during the asymptotic training stage of the non-standard AP training procedure. Data are presented separately as sessions in which the rare CS was the CS+ or the CS-. The intervals (in milliseconds) in which responses were significantly greater to the CS+ than to the CS- are listed (third column). (N=number of records; NS=not significant.)

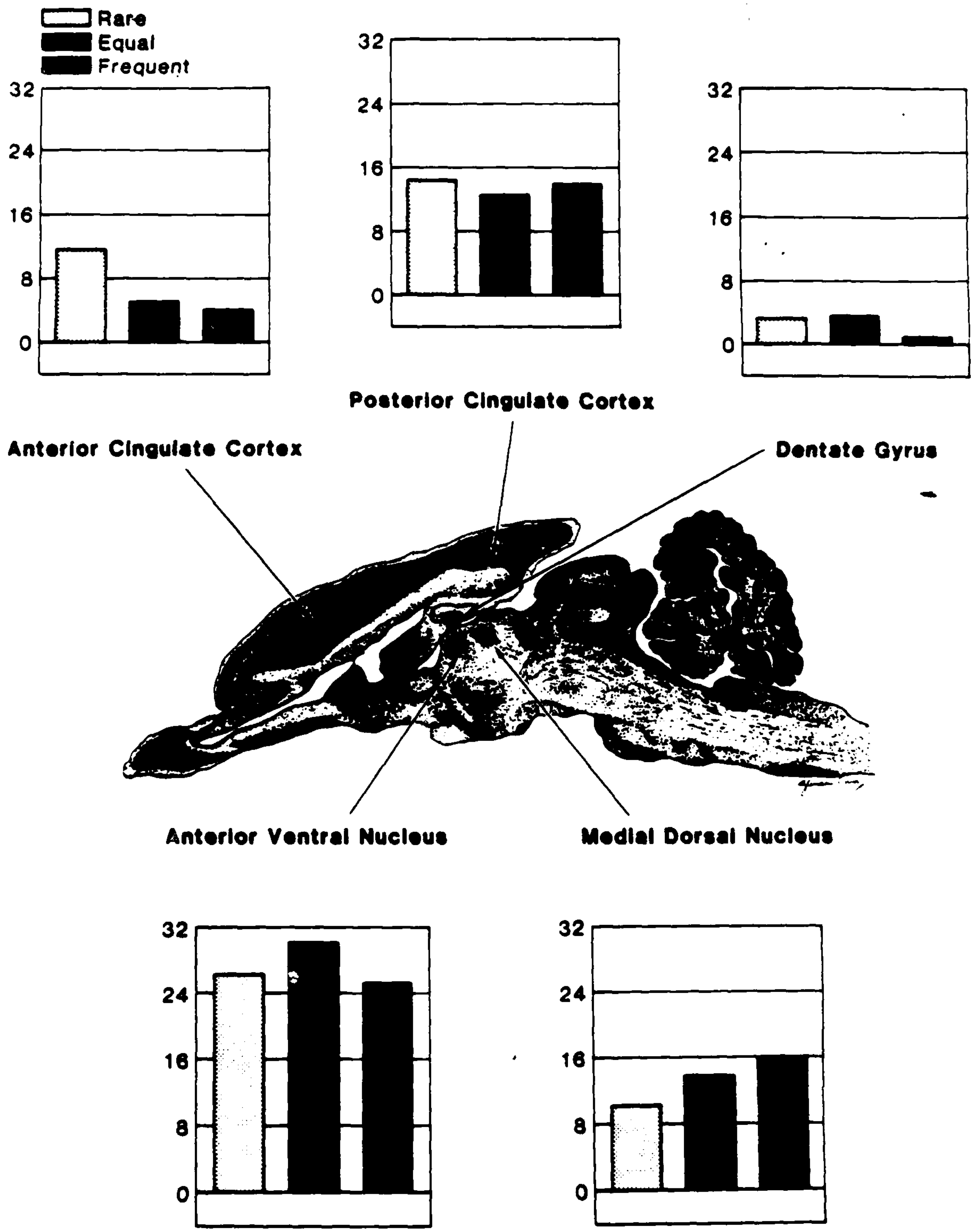

AP 4
AP 5
AP 6
AP 7.5
AP 9
AP 10

CS-

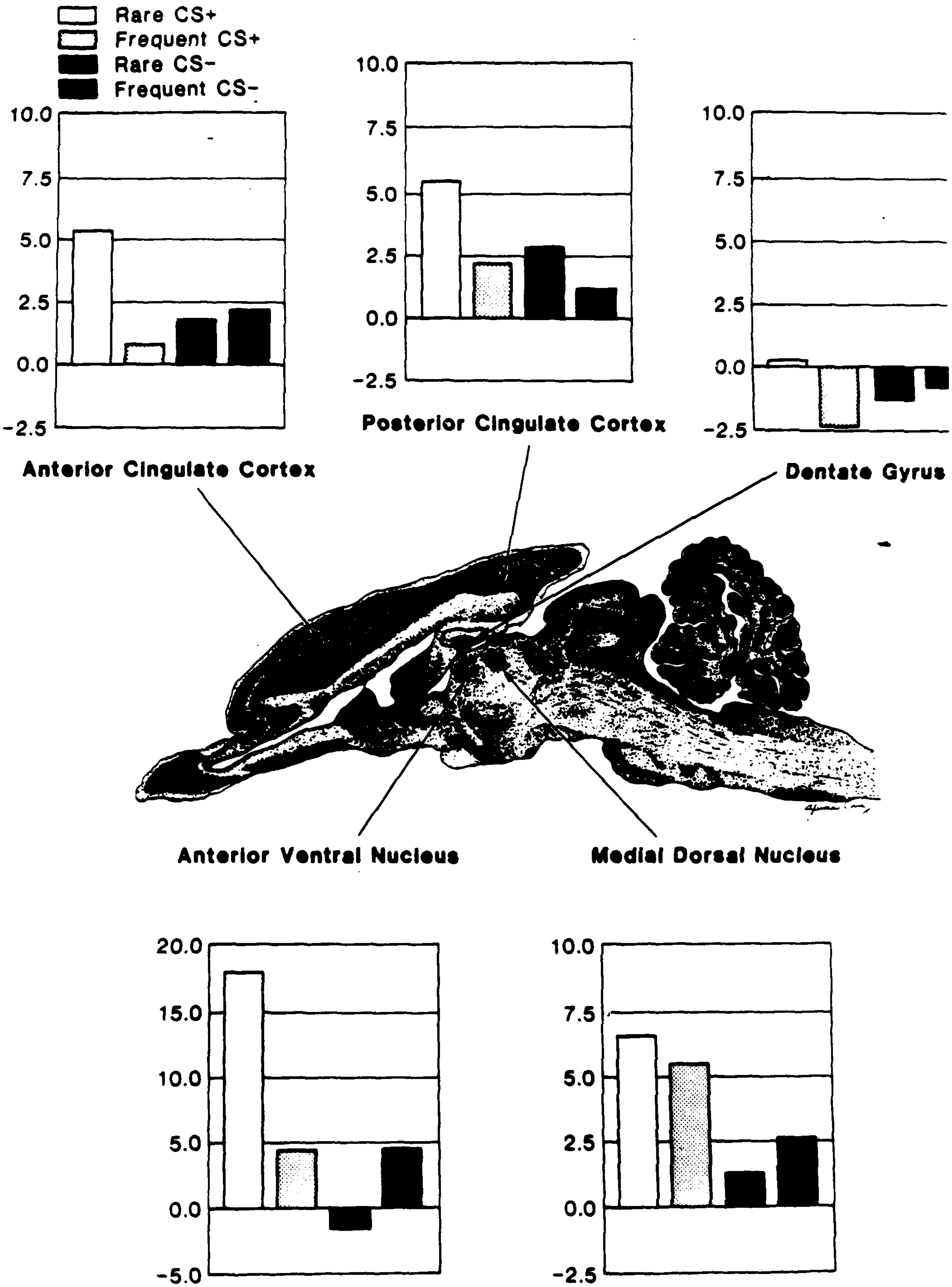


Rare
Equal
Frequent
32
24
16
8
0
Anterior Cingulate Cortex
Posterior Cingulate Cortex
Dentate Gyrus
Anterior Ventral Nucleus
Medial Dorsal Nucleus

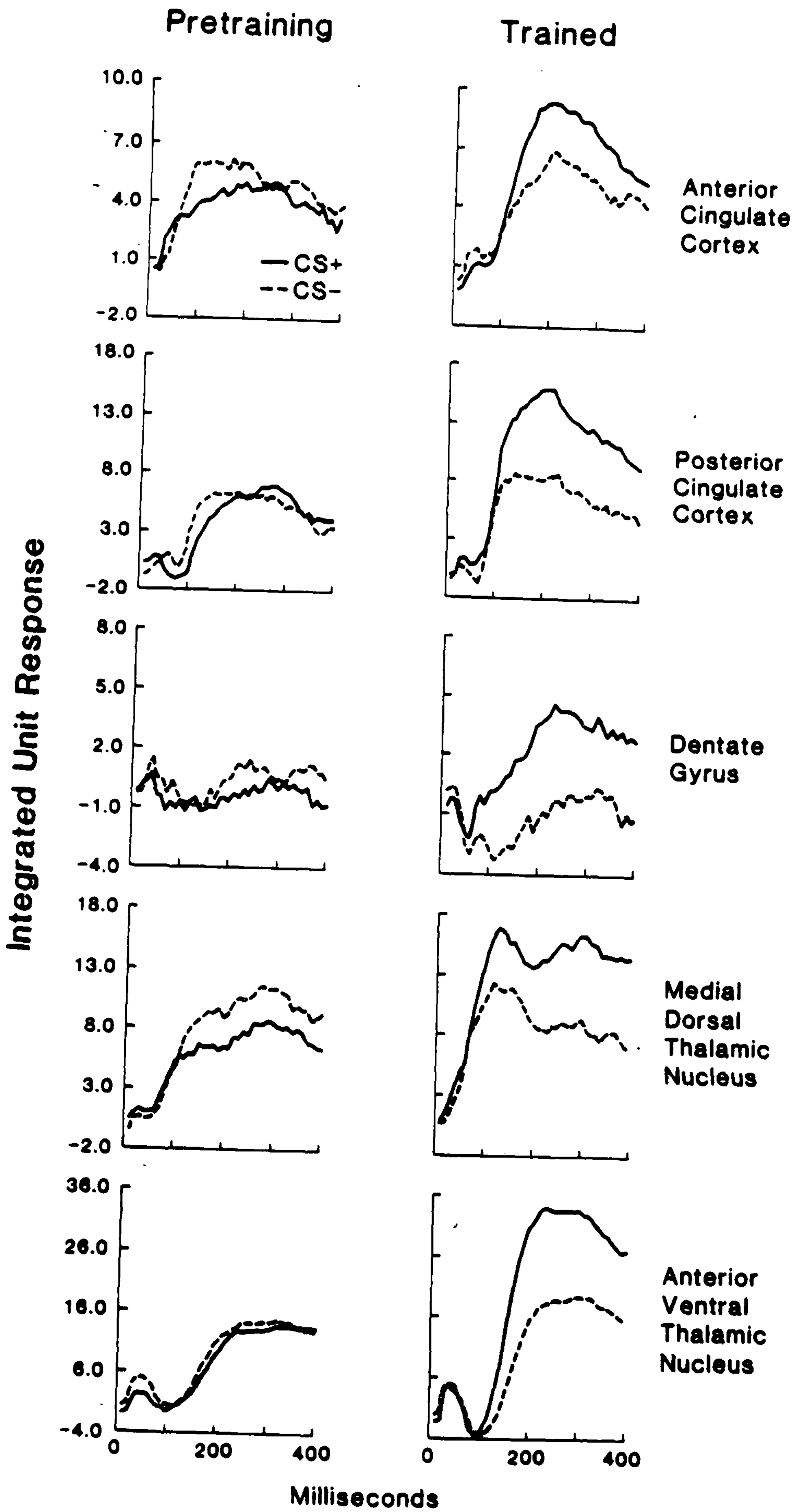
CS +



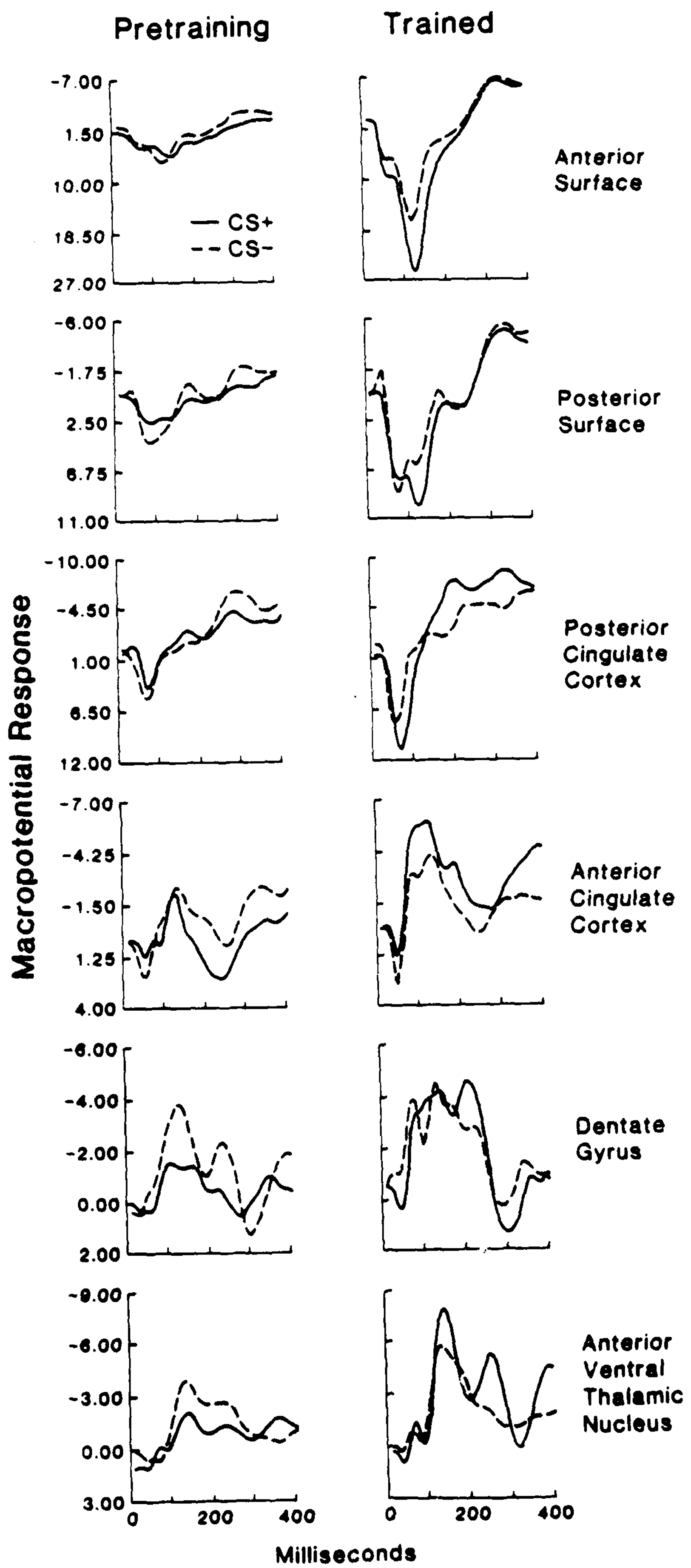
Rare
Equal
Frequent
32
24
16
8
0
Posterior Cingulate Cortex
Anterior Cingulate Cortex
Dentate Gyrus
Anterior Ventral Nucleus
Medial Dorsal Nucleus

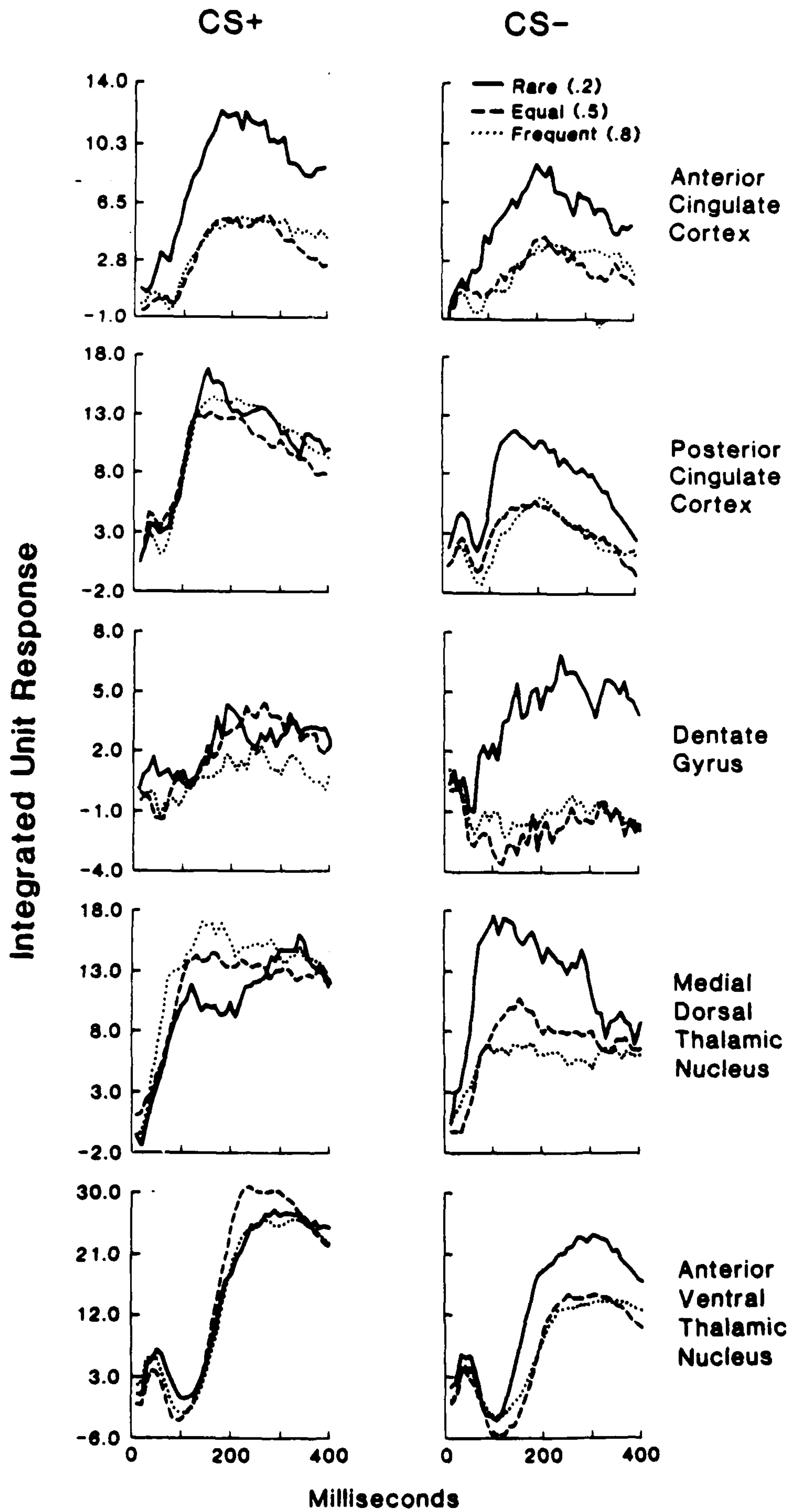


Rare CS+
Frequent CS+
Rare CS-
Frequent CS-
10.0
7.5
5.0
2.5
0.0
-2.5
Anterior Cingulate Cortex
Posterior Cingulate Cortex
Dentate Gyrus
Anterior Ventral Nucleus
Medial Dorsal Nucleus
20.0
15.0
10.0
5.0
0.0
-5.0

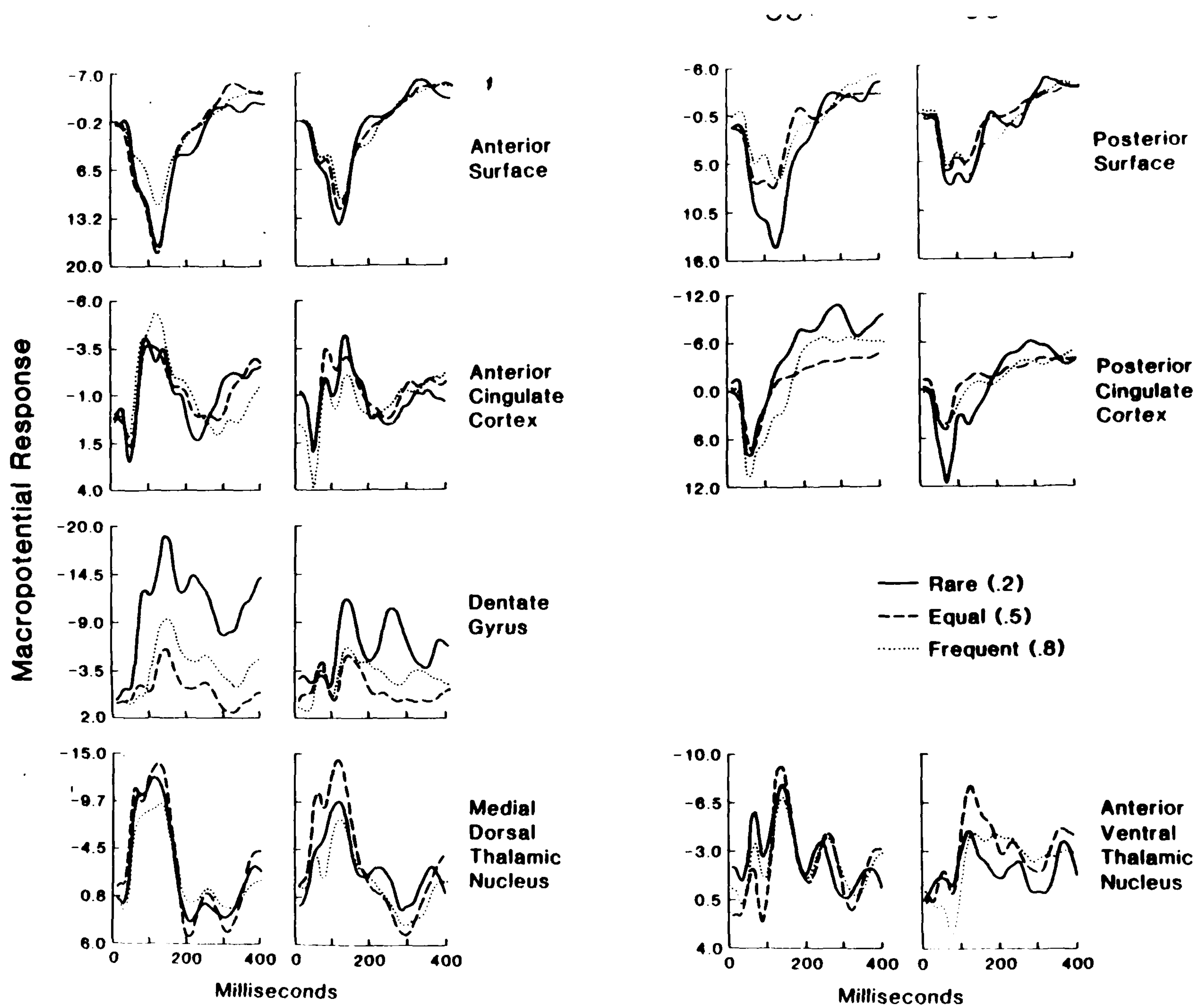


Pretraining
Trained
10.0
7.0
4.0
1.0
-2.0
—CS+
--CS-
Anterior Cingulate Cortex
18.0
13.0
8.0
3.0
-2.0
Posterior Cingulate Cortex
8.0
5.0
2.0
-1.0
-4.0
Dentate Gyrus
18.0
13.0
8.0
3.0
-2.0
Medial Dorsal Thalamic Nucleus
36.0
26.0
16.0
6.0
-4.0
Anterior Ventral Thalamic Nucleus
0
200
400
0
200
400
Integrated Unit Response
Milliseconds


Pretraining
Trained
Anterior Surface
Posterior Surface
Posterior Cingulate Cortex
Anterior Cingulate Cortex
Dentate Gyrus
Anterior Ventral Thalamic Nucleus
— CS+
-- CS-
Macropotential Response
Milliseconds
-7.00
1.50
10.00
18.50
27.00
-6.00
-1.75
2.50
6.75
11.00
-10.00
-4.50
1.00
6.50
12.00
-7.00
-4.25
-1.50
1.25
4.00
-6.00
-4.00
-2.00
0.00
2.00
-9.00
-6.00
-3.00
0.00
3.00
0
200
400



CS+
CS-
Rare (.2)
Equal (.5)
Frequent (.8)
Anterior Cingulate Cortex
Posterior Cingulate Cortex
Dentate Gyrus
Medial Dorsal Thalamic Nucleus
Anterior Ventral Thalamic Nucleus
Integrated Unit Response
Milliseconds
14.0
10.3
6.5
2.8
-1.0
18.0
13.0
8.0
3.0
-2.0
8.0
5.0
2.0
-1.0
-4.0
18.0
13.0
8.0
3.0
-2.0
30.0
21.0
12.0
3.0
-6.0
0
200
400
0
200
400


Macropotential Response
Anterior Surface
Posterior Surface
Anterior Cingulate Cortex
Posterior Cingulate Cortex
Dentate Gyrus
Medial Dorsal Thalamic Nucleus
Anterior Ventral Thalamic Nucleus
Rare (.2)
Equal (.5)
Frequent (.8)
Milliseconds
-7.0
-0.2
6.5
13.2
20.0
-6.0
-0.5
5.0
10.5
16.0
-6.0
-3.5
-1.0
1.5
4.0
-12.0
-6.0
0.0
6.0
12.0
-20.0
-14.5
-9.0
-3.5
2.0
-15.0
-9.7
-4.5
0.8
6.0
-10.0
-6.5
-3.0
0.5
4.0
0
200
400

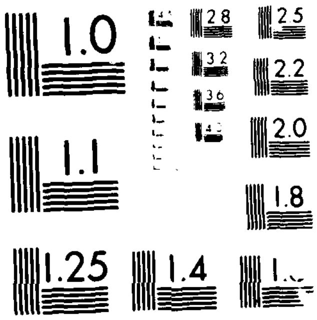
AD-A191 244 THE EVENT-RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING A..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 10/10

UNCLASSIFIED 29 FEB 88 CPL-88-1 AFOSR-TR-88-0316 F/G 5/8 NL

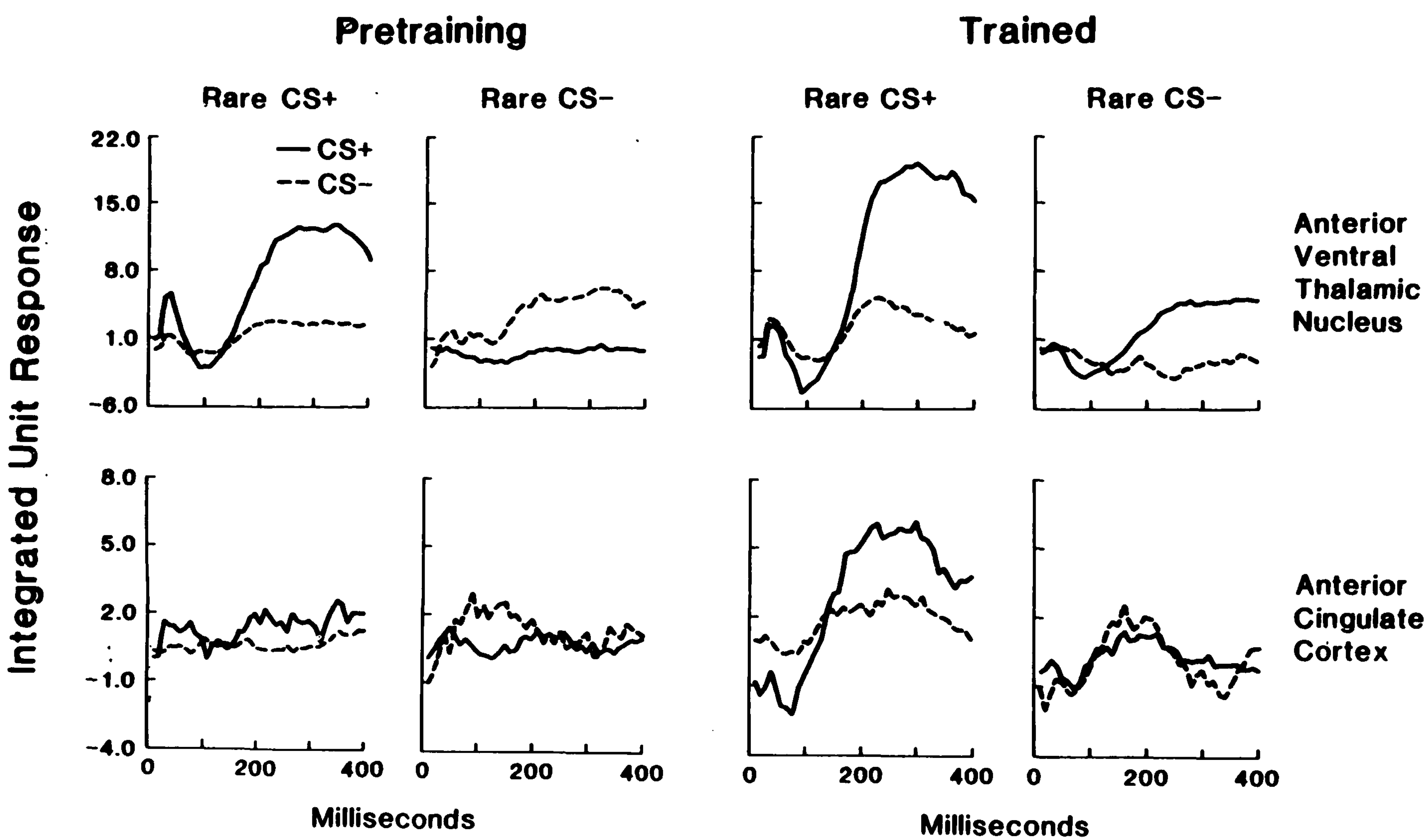
END

DATE FILMED

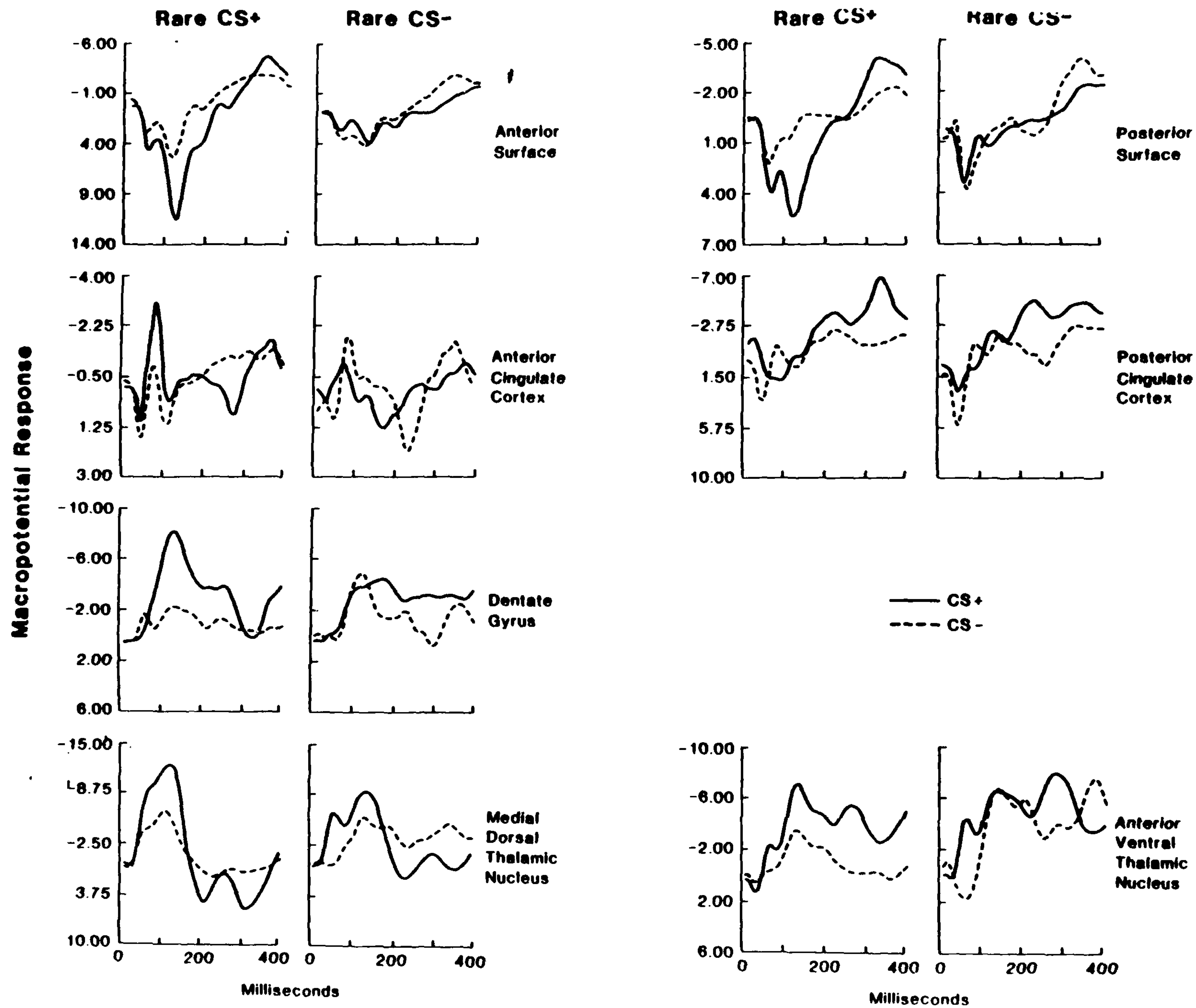
88

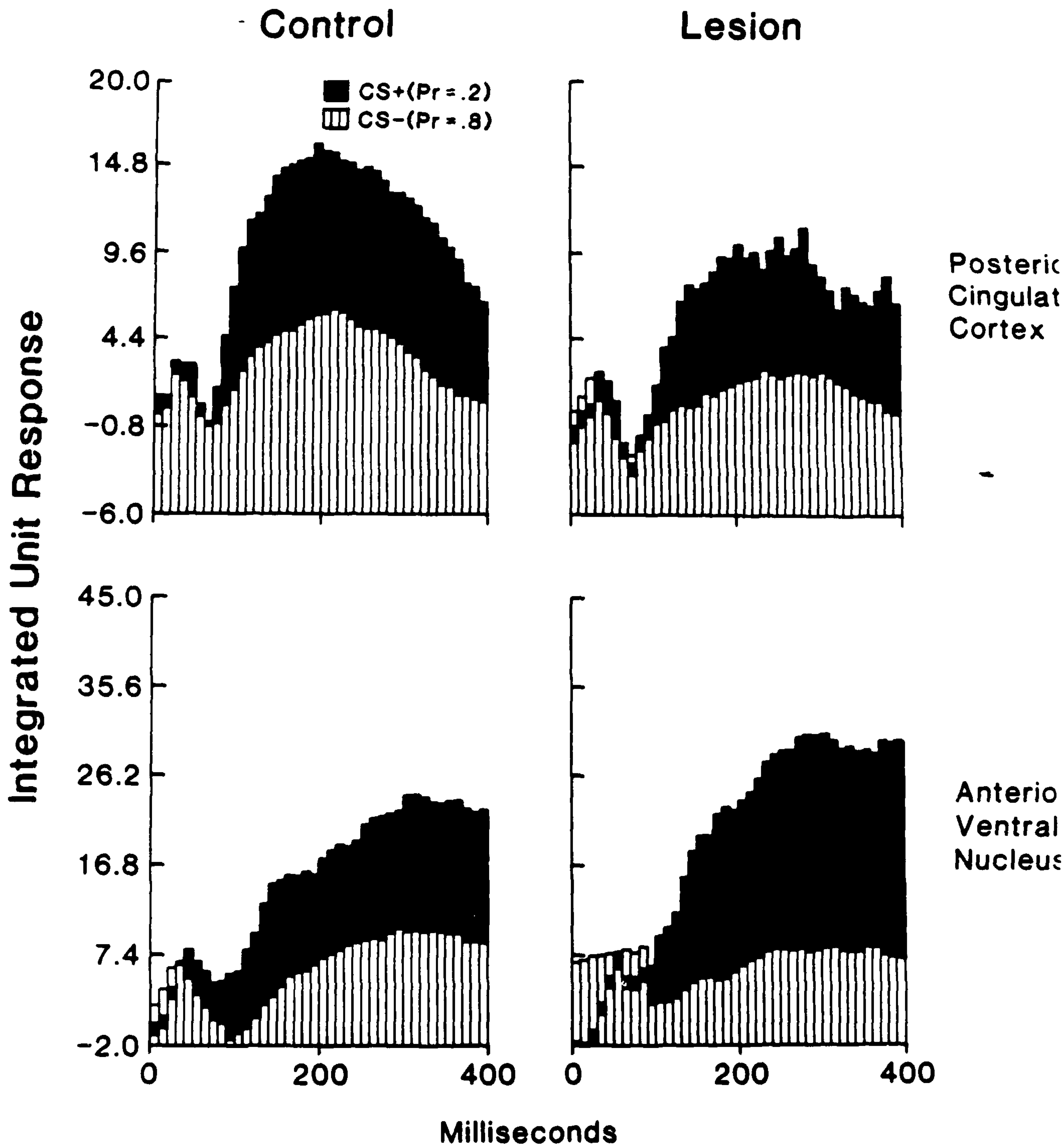


MICROCOPY RESOLUTION TEST CHART

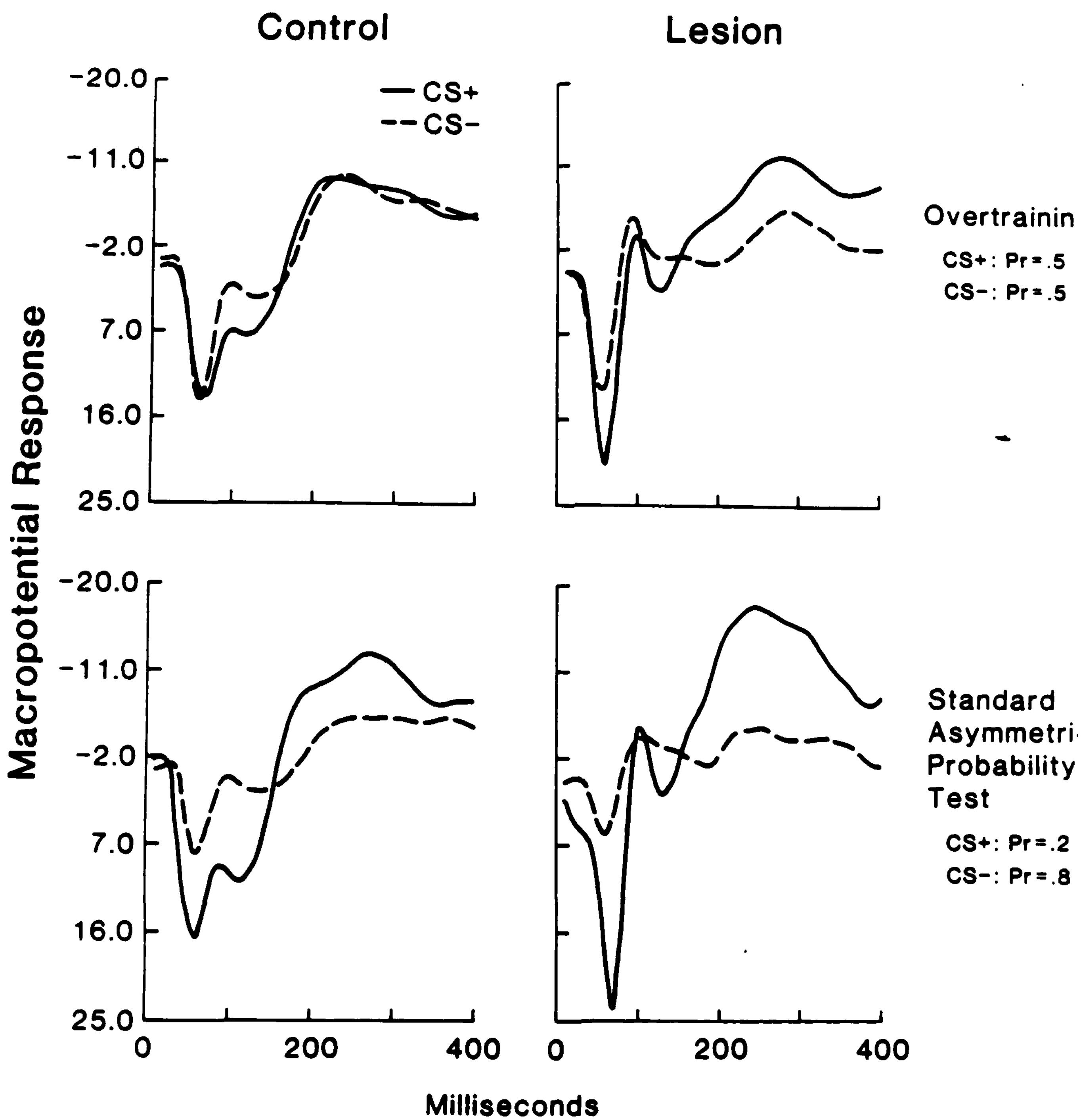


Pretraining
Trained
Rare CS+
Rare CS-
Rare CS+
Rare CS-
CS+
CS-
Integrated Unit Response
22.0
15.0
8.0
1.0
-6.0
8.0
5.0
2.0
-1.0
-4.0
0
200
400
Milliseconds
Anterior
Ventral
Thalamic
Nucleus
Anterior
Cingulate
Cortex

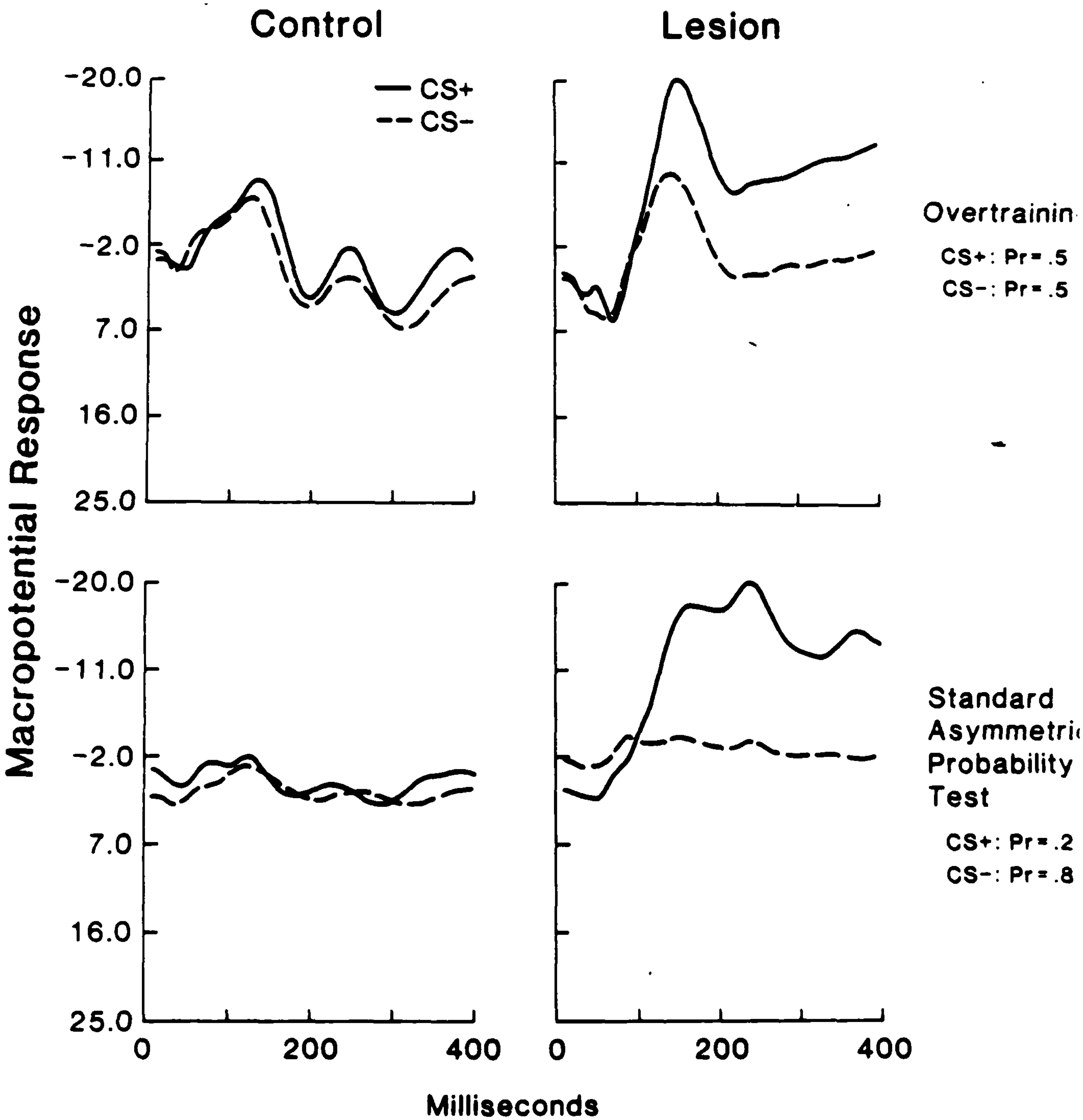

Rare CS+
Rare CS-
Rare CS+
Rare CS-
Anterior Surface
Anterior Cingulate Cortex
Dentate Gyrus
Medial Dorsal Thalamic Nucleus
Posterior Surface
Posterior Cingulate Cortex
Anterior Ventral Thalamic Nucleus
CS+
CS-
Macropotential Response
Milliseconds
Milliseconds



Control
Lesion
CS+(Pr=.2)
CS−(Pr=.8)
Integrated Unit Response
20.0
14.8
9.6
4.4
−0.8
−6.0
45.0
35.6
26.2
16.8
7.4
−2.0
0
200
400
Posteric
Cingulat
Cortex
Anterio
Ventral
Nucleus
Milliseconds



Control
Lesion
CS+
CS-
Macropotential Response
-20.0
-11.0
-2.0
7.0
16.0
25.0
0
200
400
Milliseconds
Overtrainin
CS+: Pr=.5
CS-: Pr=.5
Standard
Asymmetri
Probability
Test
CS+: Pr=.2
CS-: Pr=.8



Control
Lesion
CS+
CS-
-20.0
-11.0
-2.0
7.0
16.0
25.0
Overtrainin
CS+: Pr= .5
CS-: Pr= .5
Standard
Asymmetri
Probability
Test
CS+: Pr=.2
CS-: Pr=.8
Macropotential Response
0
200
400
Milliseconds

Stages of Training / Neural Structures	Standard Training Sessions		Overtraining AP Sessions		Non-Standard AP Sessions*			
					Pretraining		Trained	
	Unit Activity	Macropotential Activity	Unit Activity	Macropotential Activity	Rare CS+	Rare CS-	Rare CS+	Rare CS-
Anterior Cingulate Cortex (ACC)	12	11	11	11	13	8	11; 10	7
Posterior Cingulate Cortex (PCC)	7	7	12	11	8	7	7	6
Dentate Gyrus (DG)	7	6	7	5	8	5	7	5;4
Medial Dorsal Thalamic Nucleus (MDN)	4	5	4	3	5	4	4	3
Anterior Ventral Thalamic Nucleus (AVN)	6	3	6	4	6	4	6	4
Anterior Surface (AS)	--	10	--	13	12	8	10	9
Posterior Surface (PS)	--	12	--	14	13	9	11	11

	N	Stage by Interval	Intervals with Significant Differences	Stage by Stimulus by Interval	Intervals with Significant Differences	
					Pretraining	Trained
ACC	12	P=.0011 F=1.60	140-290	NS	-	-
PCC	7	P<.0001 F=4.05	90-400	P=.0002 F=1.78	100-150* 250,370,380	110-400
DG	7	NS	-	P<.0001 F=1.92	370*	80-400
MDN	4	NS	-	P<.0001 F=1.93	310*	100-400
AVN	6	P<.0001 F=3.45	160-400	P<.0001 F=5.98	30-50*,70*	120-400

	N	Stage by Interval	Intervals with Significant Differences	Stage by Stimulus by Interval	Intervals with Significant Differences	
					Pretraining	Trained
PCC	7	NS	-	P=.0014 F=1.62	40-60*, 260-330* 370-390*	60-110, 160-360
DG	6	NS	-	P<.0001 F=2.43	80-170*, 320,330	50*,200 210
AS	10	P<.0001 F=8.08	60,80-160, 300-400	P<.0001 F=2.89	NS	60-200
PS	12	P<.0001 F=2.86	70,90,110-140, 320-360	P=.0018 F=1.42	80*,90* 310-340*	40*,50* 110-180

	N	Stage by Stimulus	Stage by Stimulus by Interval		Intervals with Significant Differences		
					.2/.8	.2/.5	.5/.8
ACC	11	P=.0001 F=14.45	P<.0001 F=3.18	CS+ CS-	50-400 60-350, 380-400	40-400 80-400	360*,390-400* 310*,320*,340*
PCC	12	P=.0183 F=4.82	NS		-	-	-
DG	7	P=.0004 F=16.10	P<.0001 F=2.56	CS+ CS-	170,190-210, 350-390 70-400	40,60 70-400	230,270-300, 350-370 100*,110*
MDN	4	NS	NS		-	-	-
AVN	6	NS	P<.0001 F=2.24	CS+ CS-	NS 140-400	60-90 180*, 200-260* 120-400	170-300 -

	$\underline{N}$	Stage by Stimulus by Interval		Intervals with Significant Differences		
				.2/.8	.2/.5	.5/.8
AS	13	P<.0001 F=2.43	CS+	70-140,210 220	40*,50*, 200-220, 310-350*	60-140
			CS-	80-120, 180-200*	90-100	NS
PS	14	P<.0001 F=2.26	CS+	50-160,190, 200	70-210	80-100, 170-180*
			CS-	90-100, 170-190*	NS	170-200*

	N	Pretraining: Stimulus by Interval	Pretraining: Intervals with Significant Differences	N	Trained: Stimulus by Interval	Trained: Intervals with Significant Differences
ACC:						
Rare CS+	13	NS	-	11	P<.0001 F=3.95	20*,30*,60-80*, 170-330,350,380-4
Rare CS-	8	NS	-	7	NS	-
PCC:						
Rare CS+	8	P=.0019 F=1.89	NS	7	P=.0010 F=1.99	170-390
Rare CS-	7	NS	-	6	P<.0001 F=2.42	NS
DG:						
Rare CS+	8	NS	-	7	P<.0001 F=2.83	100-120*,140*,24 250,320,350,380-
Rare CS-	5	NS	-	5	NS	-
MDN:						
Rare CS+	5	NS	-	4	P=.0037 F=1.93	140-400
Rare CS-	4	P<.0001 F=3.03	90-400*	3	NS	-
AVN:						
Rare CS+	6	P<.0001 F=2.67	200-400	6	P<.0001 F=19.13	90*,190-400
Rare CS-	4	P<.0001 F=5.67	10,50*,80-110*, 130-400*	4	P<.0001 F=3.36	210-400

	$\underline{N}$	Stimulus by Interval	Intervals with Significant Differences
DG:			
Rare CS+	7	p=.0002	90-230,260,390,400
Rare CS-	4	F=2.16	
MDN:			
Rare CS+	4	P<.0001 F=6.63	60-160,200-220, 300-350
Rare CS-	3	NS	-
AS:			
Rare CS+	10	p<.0001 F=4.52	60,70,100-200
Rare CS-	9	NS	-
PS:			
Rare CS+	11	p<.0001 F=3.63	100-170
Rare CS-	11	NS	-

P300 Latency and Task Requirements

Stephen Jenkins, Gabriele Gratton,
Michael G. H. Coles, & Emanuel Donchin

Cognitive Psychophysiology Laboratory
University of Illinois

We have argued elsewhere (Kutas et al, 1977; McCarthy & Donchin, 1981; Magliero et al, 1984) that P300 latency is relatively independent of response-related processes and that it can be used as a measure of stimulus evaluation time.

In the present experiment, we sought to determine the degree to which P300 latency was influenced by task-oriented decision processes. Eight subjects performed letter-discrimination and same-different judgement tasks in separate sessions. In both tasks, one of four visual arrays (HHHHH, SSHSS, SSSSS, and HHSHH) was presented on each of 1200 trials. In the letter-discrimination task, subjects were required to execute left- or right-hand responses as a function of the center letter in the array. In the same-different judgment task, they were required to indicate whether all the letters in the array were the same, or whether one was different, by responding with their left- or right-hands.

In the letter-discrimination task, reaction times were longer when the central letter in the array was different from the surrounding letters. In the same-different judgment task, reaction times were the same for all arrays. However, P300 latency was longer when the central letter was different from the surrounding letters in both letter-discrimination and same-different judgment tasks.

These results reveal that measures of reaction time and P300 latency can be dissociated and support the claim that P300 latency is independent of task-oriented decision processes.

N100 AND P300 TUNING EFFECTS DURING AN ATTENTION SWITCHING TASK.

Erik Sirevaag* and Arthur Kramer* (SPON: L. Malmgren).
Dept. Psych., University of Illinois, Urbana-Champaign, Il 61820.

The question of central importance to the study of selective attention concerns how, given the continuous bombardment of information to which our central nervous system is subject, are particularly important sensations, memories and representations selected for further processing? Dichotic listening tasks designed to simulate the "cocktail party effect" -- provide convincing evidence of the ability of individuals to selectively attend to one speaker and ignore a chorus of other voices. However, Lewis (1970, JEP, 85, 225-228) found that if subjects repeat a random string of words presented to one ear, it takes them longer if a semantically related word is presented simultaneously to the opposite ear. This contextual interference represents a failure of selective attention. Treisman, Squire, and Green (1974, Mem. and Cog., 2(4), 641-646) found that the effect of context is restricted to items in early list positions which suggests that attentional selectivity builds up (becomes better tuned) with successive stimulus presentations.

Investigations of the electro-physiological concomitants of selective attention have indicated that specific components of the Event-Related-Potential (ERP) are related to selection processes. A negative potential peaking approximately 100 msec post-stimulus is sensitive to stimulus set selections based on simple stimulus features (location, pitch, color, etc). This potential is presumed to reflect the overlap between the sensory N100 component and an attention related sustained negativity labelled the processing negativity (PN). A subsequent component, the P300, is influenced by a number of response set factors such as instructions, strategies, and complex as well as simple stimulus attributes. The P300 has been related to the further hierarchical evaluation of attended stimuli.

Two separate sets of studies have explored the nature of perceptual tuning effects upon components of the ERP. In the auditory modality, Donald and Young (1982, Exp. Brain Res., 46, 357-367) found that the selectivity of the P300 emerged immediately with the first few stimuli in a block and did not change over succeeding trials in a sustained attention paradigm. PN tuning, on the other hand, did not emerge until after the presentation of several trials and resulted from a decrement in the PN associated with the rejected channel over time. In the visual modality, Hillyard, Munte, and Neville (1985, Atten. and Perf. XI, Hillsdale: Erlbaum, 63-84) have found that the PN selectivity was present for the first stimuli in a sequence and remained unchanged with successive stimuli despite a general decline in PN amplitude over time.

The purpose of the present study was to examine the buildup of auditory perceptual tuning in a paradigm requiring subjects to switch attention to different input sources more rapidly than was required in previous studies.

Methods

Fig 1. Task Description. Nine subjects (7 male and 2 female) participated in the experiment. An arrow presented visually before each sequence of twelve auditory stimuli indicated which ear was to be attended to. Prior to the experiment a target frequency was designated for each subject. If a tone was presented to the attended ear and was of the target frequency a button press response was required. Subjects performed 50 blocks of these 13 trial sequences followed by a 5 min rest period. In a single session, 7 groups of 50 blocks were performed. Behavioral data (RT and accuracy) and electroencephalographic (EEG) data from three electrode positions (Fz, Cz, and Pz) were recorded in each block. The amplitudes and latencies of a variety of ERP components were analyzed as a function of the ear, frequency, and target/non-target classification of the eliciting tones. The ERP components and behavioral data were also examined as a function of the time point of the stimulus following a switch in attentional locus (Time 1, Time 2, and Time 3). In the subsequent figures the following convention is used to designate the four possible combinations of attended vs. ignored frequency and location attributes of the tone stimuli:

F+ L+ : Attended Frequency in the Attended Location
F+ L- : Attended Frequency in the Ignored Location
F- L+ : Ignored Frequency in the Attended Location
F- L- : Ignored Frequency in the Ignored Location

Results

Fig. 2 Mean Reaction Time as a Function of Time Point. Mean RT did not vary as a function of time point [$F(2,16)=0.14$; $p=0.87$].

Fig. 3 False Alarm Rate as a Function of Time Point. The number of false alarms did not vary as a function of time point [$F(2,16) = 0.90$; $p=0.43$.]

Fig. 4 Miss Rate as a Function of Time Point. Significantly more misses occurred for stimuli at Time 1 than for stimuli at either Time 2 or Time 3 [$F(2,16)=10.07$; $p=0.002$].

Fig. 5 Sensitivity as a Function of Time Point. Significantly lower sensitivity scores were associated with stimuli at Time 1 than at Time 3 [$F(2,16)=4.91$; $p=0.02$].

Fig. 6 Grand Average ERPs at Fz as a Function of Stimulus Type. Prominent negativities occurred with a mean latency of 110 msecs. At each of the three time points, more negativity was associated with stimuli of the attended frequency $F(1,8)=54.65$; $p=.008$. The ear of presentation had no effect $F(1,8)=0.15$; $p=0.71$.

Fig. 7 Grand Average ERPs at Fz as a Function of Time Point. While less negativity was associated with tones of the non-target frequencies at later points in time, target frequency evoked negativities remained unchanged at all three time points $F(2,16)=6.25$; $p=0.01$.

Fig. 8 PN Amplitude as a Function of Time Point. Average of the single subject estimates of PN amplitude 100 msec post-stimulus at Fz illustrating the increased separation between the target and non-target frequency channels.

Fig. 9 Grand Average ERPs at Pz as a Function of Time Point. Large P300s were associated only with stimuli matching on both the target frequency and location dimensions $F(1,8)=36.48$; $p=0.0003$. The amplitude of the P300 to these tones increased as a function of time point $F(2,16)=12.18$; $p=0.0006$.

Fig. 10 P300 Amplitude as a Function of Time Point. Average of the single subject estimates of P300 amplitude at Pz illustrating the increased separation between the tones matching both target attributes and the other tone categories.

Conclusions

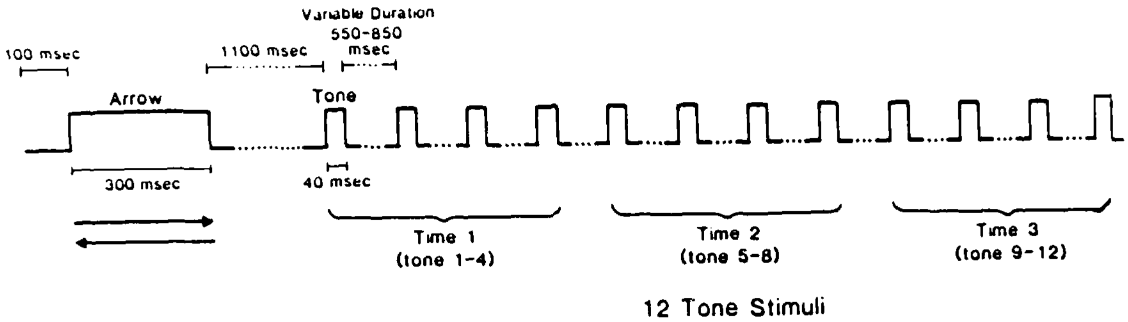
1. The behavioral data suggest that subject sensitivity increased with successive stimulus presentations within a given attentional set. The fact that RT and false alarm rate were not affected by time indicates that the sensitivity changes cannot be attributed to either speed-accuracy tradeoffs or differential response biases.

2. The PN data are consistent with the hypothesis that increased perceptual tuning occurred within the 12 tone sequence and that this was accomplished by a reduction in the amplitude of the PN associated with non-target tone frequencies regardless of location. However, a number of alternative explanations exist. Although extensive bootstrapping has indicated that the PN differences are probably not artifactually produced by differences in refractory periods or number of trials per average, it is possible that the sensitivity of the PN to frequency rather than ear is due to such factors as probability, the nature of the dimension that switches, etc. Further work is needed to clarify the precise nature of the observed PN effect.

3. The P300 data provide convincing evidence of a P300 tuning effect. In contrast with the PN effect, P300 tuning is accomplished by an increase in the amplitude of P300s associated with target stimuli.

4. The data support the hypothesis that there is a phasic build-up of attentional processes during tasks requiring sustained attention when subjects are required to modify their attentional set after every few trials. The data are also consistent with hierarchical models of stimulus selection.

:. 1



Task: Respond to (1000 or 1400Hz) tones in (left or right) ear.

Fig. 2

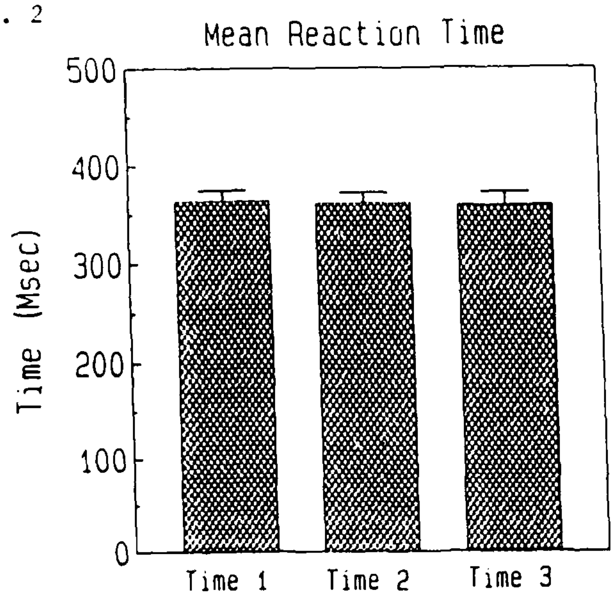


Fig. 3

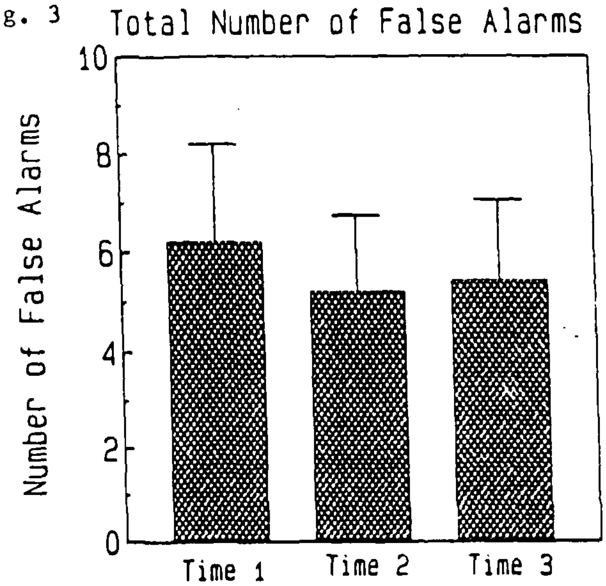


Fig. 4

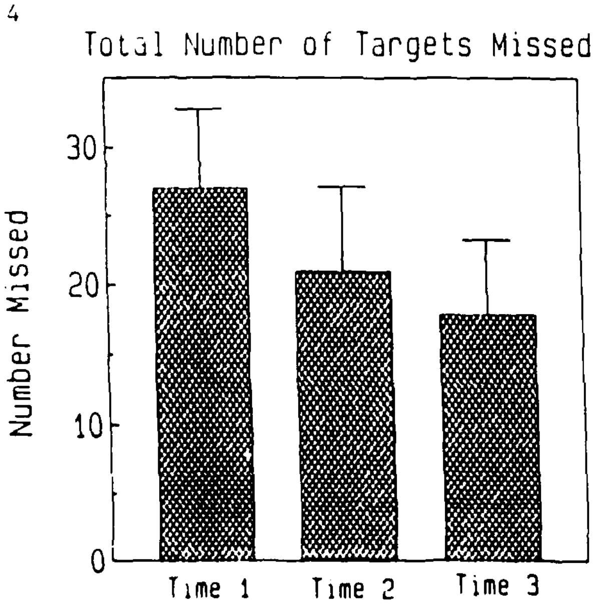


Fig. 5

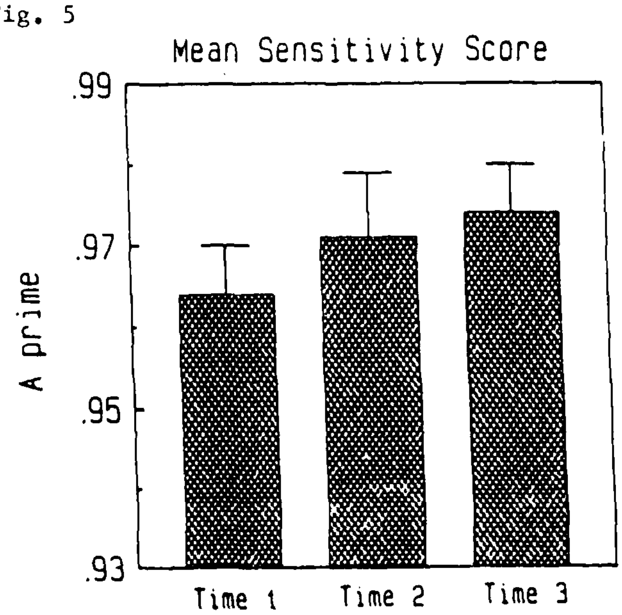


Fig. 6

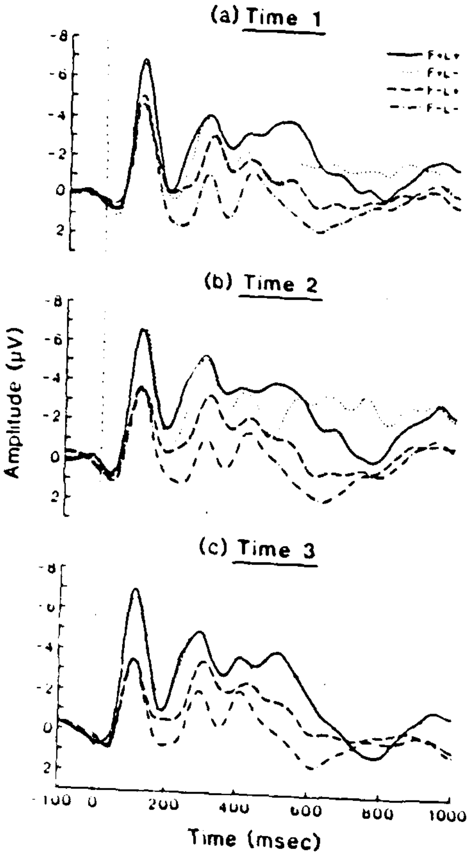


Fig. 7

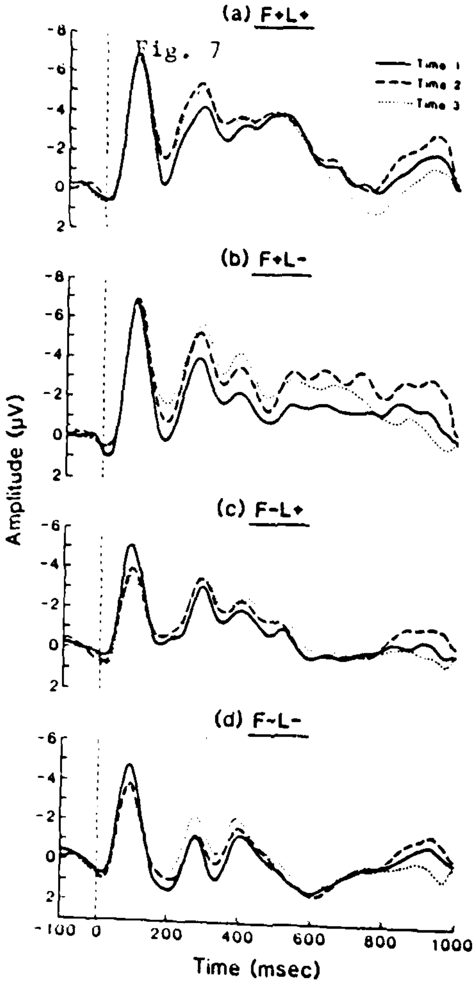


Fig. 9

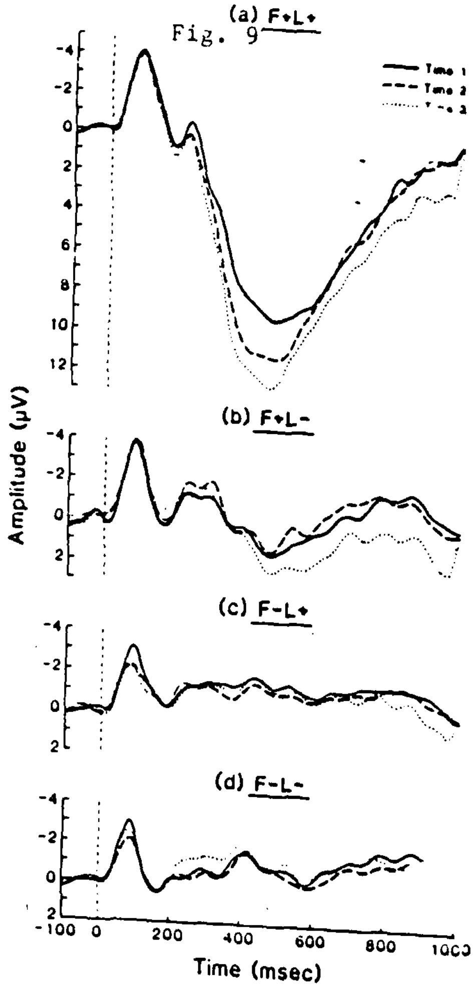


Fig. 8

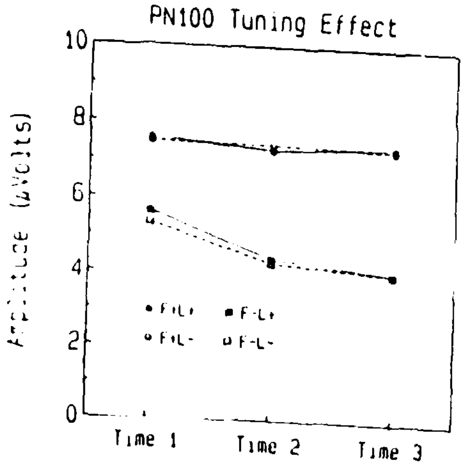
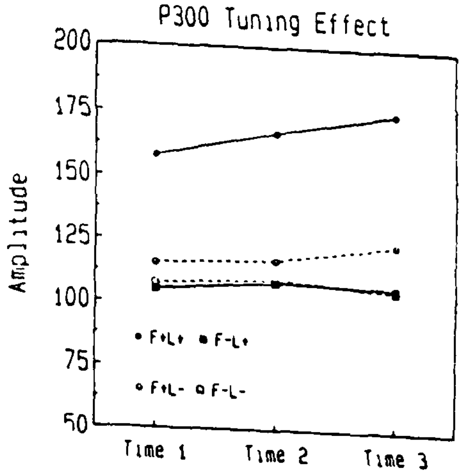


Fig. 10



An Evaluation of Age Differences in the Development of Automaticity

William J. Gehring, David L. Strayer, Arthur F. Kramer,
Emanuel Donchin, and Gregory A. Miller
Cognitive Psychophysiology Laboratory
University of Illinois
Champaign, Illinois

The present study examines the development of automaticity, as reflected in overt performance and in the P300 component of the event-related brain potential, in healthy elderly and young adults. The subjects practiced consistently (CM) and variably (VM) mapped versions of the Sternberg memory search task. Four elderly (mean age = 74) and two young (mean age = 22) adults performed the task for four consecutive days; performance from the first and last days was compared.

Reaction time (RT) increased with memory load for both groups, although the effect was greater for the elderly subjects. The effect of memory load was greater under VM conditions than under CM conditions for young subjects, but for elderly subjects the effect did not differ between conditions. Overall, RT performance improved over the four sessions, but RTs decreased more for young subjects than for the older subjects. Furthermore, the effect of memory load decreased with practice only for the young subjects. Taken together, these results suggest that the young subjects developed automatic processing more rapidly than the elderly subjects.

An examination of detection sensitivity revealed that, under CM conditions, A' was very high, and was unaffected by both memory load and training for young subjects. For elderly subjects, however, A' decreased with memory load early in training; after consistent practice, A' was relatively unaffected by memory load. Further, by session 4, detection sensitivity was higher for elderly subjects than for young subjects.

In VM training, A' decreased as a function of memory load for all subjects and all conditions, but was lower for the elderly than for the young subjects. By session 4, A' was less affected by memory load for the young than for the elderly subjects.

Examination of B'', a measure of response criterion, revealed that younger subjects' performance became less conservative with CM practice. Elderly subjects, however, did not decrease their response criterion with such practice, although A' increased significantly. With VM training, young subjects did not change their response criterion, whereas elderly subjects became more conservative. These results suggest that elderly subjects did not vary their response criterion on the basis of improved sensitivity. It has been previously suggested that such a conservative strategy decreases the rate at which automaticity develops (Schneider & Fisk, 1982).

In CM conditions, the effects of memory load on P300 latency were equivalent for young and elderly subjects early in training. Following consistent practice, however, P300 latency was insensitive to memory load for young subjects, but remained affected by memory load for elderly subjects. In addition, practice in CM conditions resulted in a greater overall reduction in P300 latency for the young. These results are consistent with the pattern of RT results and suggest that the stimulus evaluation process became automated more rapidly for the young.

In contrast, P300 latency increased as a function of memory load in VM conditions for young subjects across the four sessions. For the elderly subjects, however, P300 latency was insensitive to memory load throughout the experiment. Moreover, P300 latency for older subjects increased with training. These results are paradoxical, given the pattern of RT results. Similar findings of P300 slope diminishing with age have been reported by Pfefferbaum (1980) and Strayer, Wickens, and Braune (1985). One possible explanation for these results is that P300 was smaller and more variable for the elderly subjects. Indeed, P300 latency distributions for memory load 2 and memory load 4 were rectangular and overlapping for the elderly. Given the reliability of our measure of P300 latency (Gratton, 1984) these results suggest that the latency of the process manifested by P300 was more variable in the elderly.

The flat slopes observed following CM practice in the young should be contrasted with the flat slopes observed in the elderly in the VM conditions. The distributions of the young subjects were always Gaussian, but overlapped following consistent practice, yielding a flat slope. The distributions of the elderly, in contrast, were rectangular, and thus the overlap of these distributions--and the resulting flat slope--was due to the high variability in the P300s of the elderly.

RT distributions were analogous to the P300 latency distributions: following practice in any condition, the variance of the distribution decreased for all subjects. This effect was more pronounced in CM conditions. Distributions for the elderly, however, were more variable than for the young subjects, and the effect of practice on the variance was less for the elderly. These results further support the notion that the P300 distributions were rectangular due to variability in stimulus evaluation time.

In sum, it appears that the performance of the elderly was characterized by conservative response strategies and by highly variable stimulus evaluation and response processes. For the elderly subjects, performance improved following CM practice; nevertheless, they failed to attain automatic processing over 5000 trials, in contrast to the young subjects. It remains to be determined if automaticity would develop in the elderly with additional training.

Article 25.

P300 Operating Characteristics: Performance/ERP Analysis of Dual-Task Demands and Automaticity

Arthur F. Kramer and David L. Strayer

Department of Psychology
University of Illinois
Champaign, Illinois

Fourth International Conference on Cognitive Neuroscience
June 14-19, 1987
Paris-Dourdan, France

The present study examines the attentional requirements of automatic and controlled processing. The amplitude of the P300 component of the ERP was used as a metric of the attentional resources invested in a pair of tasks. Subjects performed two tasks (a Sternberg Memory Search task and a Recognition Running Memory task) both separately and together. Two stimulus-response mapping conditions were employed: consistent mapping (CM) and varied mapping (VM). Processing priority was manipulated between the two tasks by instructions. Subjects received extensive training (>24,000 trials) prior to the experiment.

It was predicted that if automatic processing demands attentional resources, then CM events should elicit large P300s. However, since the resources utilized during the automatic attention response are assumed to be allocated in an all-or-none fashion it was hypothesized that the amplitude of the P300's elicited in the CM conditions would not tradeoff with priority and should not show a dual task decrement. In contrast, if automatic processing does not utilize or demand attentional resources, there should be little or no P300 activity elicited by CM events. However, subjects may allocate spare capacity to the automatic processing even though unnecessary. In this case, P300 amplitude should vary as a function of processing priority. As the concurrent task priority increases, the spare capacity remaining to be allocated to automatic processing should diminish, resulting in a graded effect on P300 amplitude.

Methods

Five subjects practiced two tasks, both separately and together for twelve two and one-half hour sessions prior to performing in the experimental conditions. This amount of practice was necessary to ensure that the subjects had developed automaticity in the CM conditions. The tasks included a Sternberg Memory Search paradigm and a Recognition Running Memory task. The Sternberg paradigm was run with two memory load conditions (set size 1 and 4) and two response mapping (CM and VM) conditions. The Recognition Running Memory task was performed by comparing the item on trial n to the item on trial n-2. Subjects made a CRT response on each trial for each of the tasks. Task priority was also manipulated in the dual task conditions such that subjects either concentrated on one task, gave equal priority to both tasks, or maximized their performance on the other task (priority conditions: 100/0, 90/10, 50/50, 10/90, 0/100, where the first number refers to the priority of the recognition running memory task and the second refers to the Sternberg task).

ERP Recording

EEG was recorded from three midline sites (Fz, Cz, and Pz, according to the International 10-20 system) and referred to linked mastoids. Two ground electrodes were positioned on the left side of the forehead. EOG electrodes were placed above and below the right eye. Electrode impedances did not exceed 10 Kohms.

The EEG and EOG were amplified with Van Gogh model 50000 amplifiers (time constant 10 sec and upper half amplitude of 35 Hz). Both EEG and EOG were sampled for 1300 msec, beginning 100 msec prior to stimulus onset. The data were digitized every 10 msec. The ERP's were digitally filtered off-line (-3Db at 8.8 Hz, 0 Db at 29 Hz) prior to statistical analysis. EOG artifacts were also corrected off-line (see Gratton, Coles and Donchin, 1983).

Figure 1. A schematic representation of the temporal structure of the Sternberg paradigm (top), the recognition running memory task (middle), and the combination of the two tasks in the dual task conditions (bottom).

Figure 2. Single and dual task Sternberg and Recognition Running Memory RT's for correct responses. On the left side of the figure the RT's are plotted for both tasks in each of the priority conditions. The right side of the figure displays the Performance Operating Characteristics (POC) for both CM and VM conditions. Least squares polynomial regression lines are fitted to the data for the CM and VM set size 1 and 4 conditions. Several effects are noteworthy. First, the RT's obtained in the CM conditions indicated that we were successful in training the subjects to automaticity; the RT's were unaffected by memory load or dual task demands (e.g., all points in the upper right hand portion of the POC). Second, even though subjects received the same amount of practice in the VM condition as they did in the CM condition they did not develop automaticity. RT's increased as a function of memory load and showed large dual task tradeoffs as a function of processing priority for the larger memory load conditions.

Figure 3. Single and dual task Sternberg and Recognition Running Memory A´s for correct responses (A´ is a nonparametric measure of senstiviity, d´). On the left side of the figure the A´s are plotted for both tasks in each of the priority conditions. The right side of the figure displays the POC´s for both CM and VM conditions. The pattern of results obtained for the A´ measure are consistent with that obtained for RT. CM performance is unaffected by memory load or dual task demands while A´ decreases as a function of memory load in the VM condition. A´ also varied as a function of dual task priority in the set size 4 VM conditions.

Figure 4. Single and dual task Sternberg and Recognition Running memory Pz overplots for both CM and VM conditions. Perusal of the grand averages reveals the classic N200/P300 complex. It is evident from the waveforms that the P300 is less variable in the CM Sternberg conditions than it is in the VM conditions. This observation is consistent with the reduced variance in both P300 latency and RT in the CM conditions. It is also clear from the waveforms that the priority effect shown in the RT and A´ POC´s is obtained for P300 amplitude. The amplitude of the P300 varies with processing priority in the VM but not in the CM conditions.

Figure 5. Grand average P300 amplitudes at Pz for single and dual task Sternberg and Recognition Running Memory tasks. The amplitude and latency of the P300's were obtained from a single trial cross-correlation procedure. On the left side of the figure the P300's are plotted for both tasks in each of the priority conditions. The right side of the figure displays the POC's for both CM and VM conditions. The pattern of results obtained for the P300 measures are consistent with the grand average waveforms. In CM conditions, large P300s were elicited by all events and P300 amplitude was uninfluenced by processing priority. Furthermore, dual task P300 amplitude was equivalent to single task P300 amplitude. In VM conditions, P300 amplitude varied as a function of processing priority. The greater the attention allocated to the task, the larger the P300s elicited by these events. Reciprocity was found between the two tasks under VM load 4 conditions.

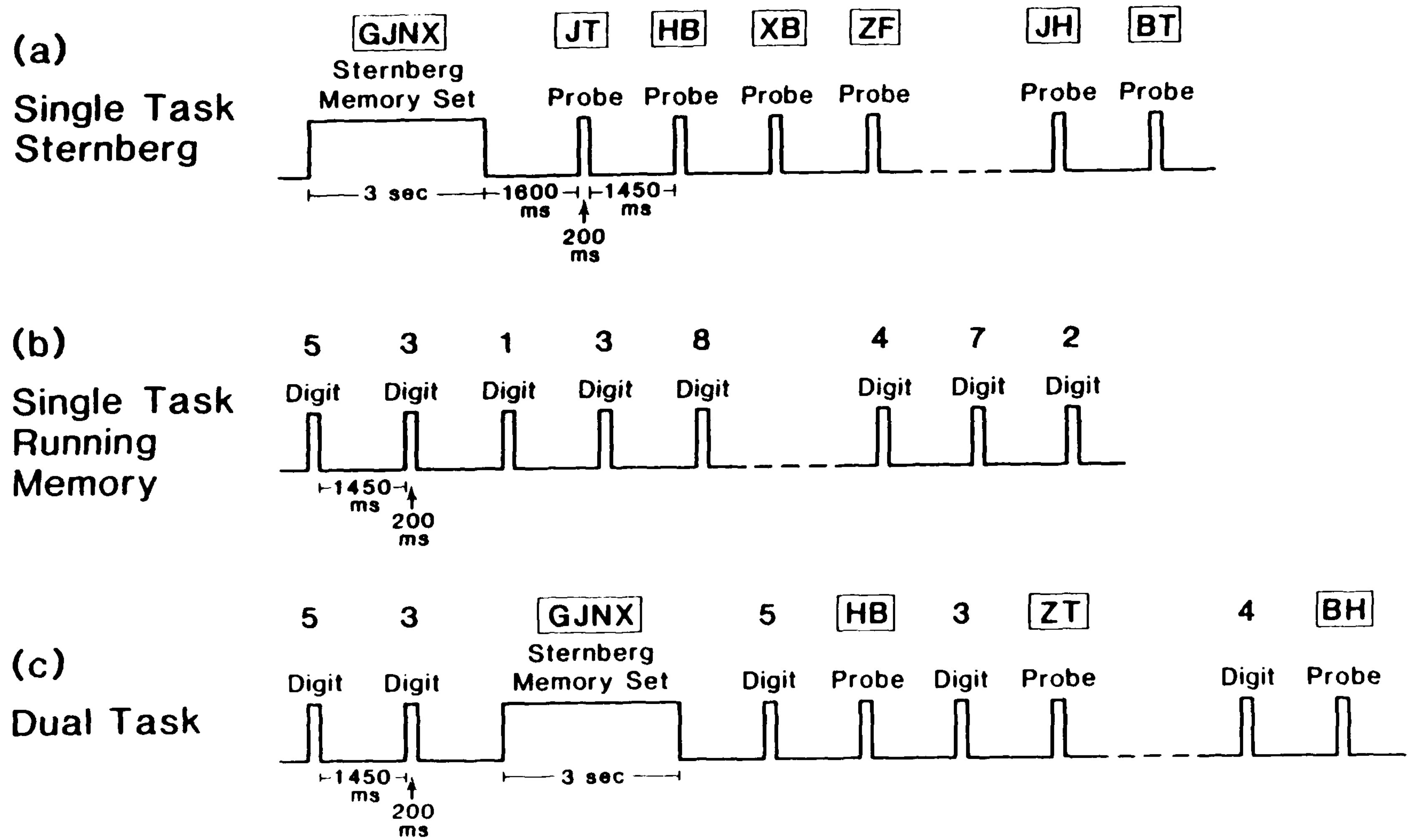
Figure 6. Grand average P300 latencies at Pz for single and dual task Sternberg and Recognition Running Memory tasks. On the left side of the figure the P300's are plotted for both tasks in each of the priority conditions. The right side of the figure displays the POC's for both CM and VM conditions. The pattern of results obtained for the P300 latency measure mimicked the RT findings. P300 latency increased as a function of set size and processing priority in the VM conditions. In the CM conditions, P300 latency was insensitive to memory load and dual task priority instructions. Given that P300 latency is sensitive to stimulus evaluation processes, but not response selection processes, these results suggest that stimulus encoding and memory comparison operations can be performed in parallel with the information processing activities of concurrent tasks after substantial practice in a consistently mapped environment.

Figure 7. RT/P300 latency ratios for single and dual task Sternberg and Recognition Running Memory tasks. On the left side of the figure the ratios are plotted for both tasks in each of the priority conditions. The right side of the figure displays the POC's for both CM and VM conditions. A ratio greater than 1 indicates that the peak P300 latency occurred prior to the overt response while a ratio of less than 1 indicates that the P300 followed the overt response. In CM conditions, P300 latency was relatively constant and did not vary as a function of priority. In addition, reaction time preceded P300 latency in all CM conditions and the RT/P300 ratio did not vary as a function of priority. In VM conditions, P300 latency increased with memory load and varied as a function of priority. Furthermore, P300 latency preceded RT, and the RT/P300 ratio varied as a function of priority. Thus, as attention was withdrawn from the task, the RT/P300 ratio increased in the VM conditions.

Conclusions

Taken together, these results support the hypothesis that attentional resources are allocated to automatic processing. When a CM target is presented, attention is automatically allocated to the task. However, consistent with the hypothesis that the automatic attention response occurs in an all-or-none fashion, P300 amplitude did not vary as a function of priority in the CM conditions. On the other hand, the tradeoff in resources between the two tasks in the VM conditions was reflected in the P300's. The amplitude of the P300 varied in a reciprocal fashion with large P300's being elicited in the task that was emphasized.

A second purpose of the experiment was to examine differences in automatic and controlled processing from a chronometric perspective. Earlier research (e.g., van Dellen et al., 1984; Strayer and Kramer, 1986) suggested that perceptual processing becomes more efficient during the development of automaticity. If this proposition is correct, then P300 latency should be uninfluenced by changes in priority under practiced CM conditions and the relative timing of reaction time and P300 latency should not change as a function of priority. These predictions were supported by the RT/P300 ratio data. The results suggest that an efficient information extraction process emerges following consistent practice. This may be likened to the tuning of a perceptual filter and may correspond to the "pop out" effect, where CM targets appear to jump out of the display.



(a)
Single Task
Sternberg
GJNX
Sternberg
Memory Set
JT
HB
XB
ZF
JH
BT
Probe
Probe
Probe
Probe
Probe
Probe
3 sec
1600
ms
1450
ms
200
ms
(b)
Single Task
Running
Memory
5
3
1
3
8
4
7
2
Digit
Digit
Digit
Digit
Digit
Digit
Digit
Digit
1450
ms
200
ms
(c)
Dual Task
5
3
GJNX
Sternberg
Memory Set
5
HB
3
ZT
4
BH
Digit
Digit
Digit
Probe
Digit
Probe
Digit
Probe
1450
ms
200
ms
3 sec

LEGEND FOR FIGURES 2, 3, 5, 6 & 7

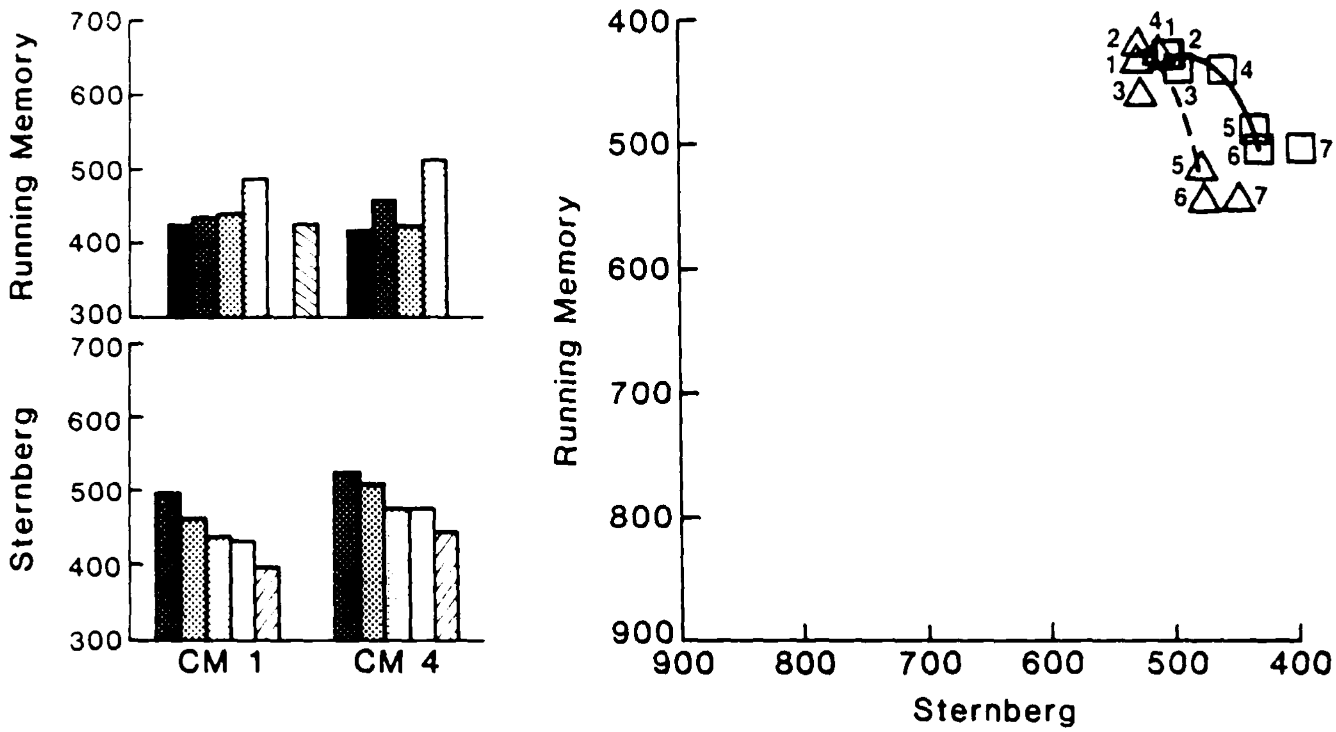
	Running Memory	Sternberg
1	Single Task	–
2	100	0
3	90	10
4	50	50
5	10	90
6	0	100
7	–	Single Task

Session 1 □ ————

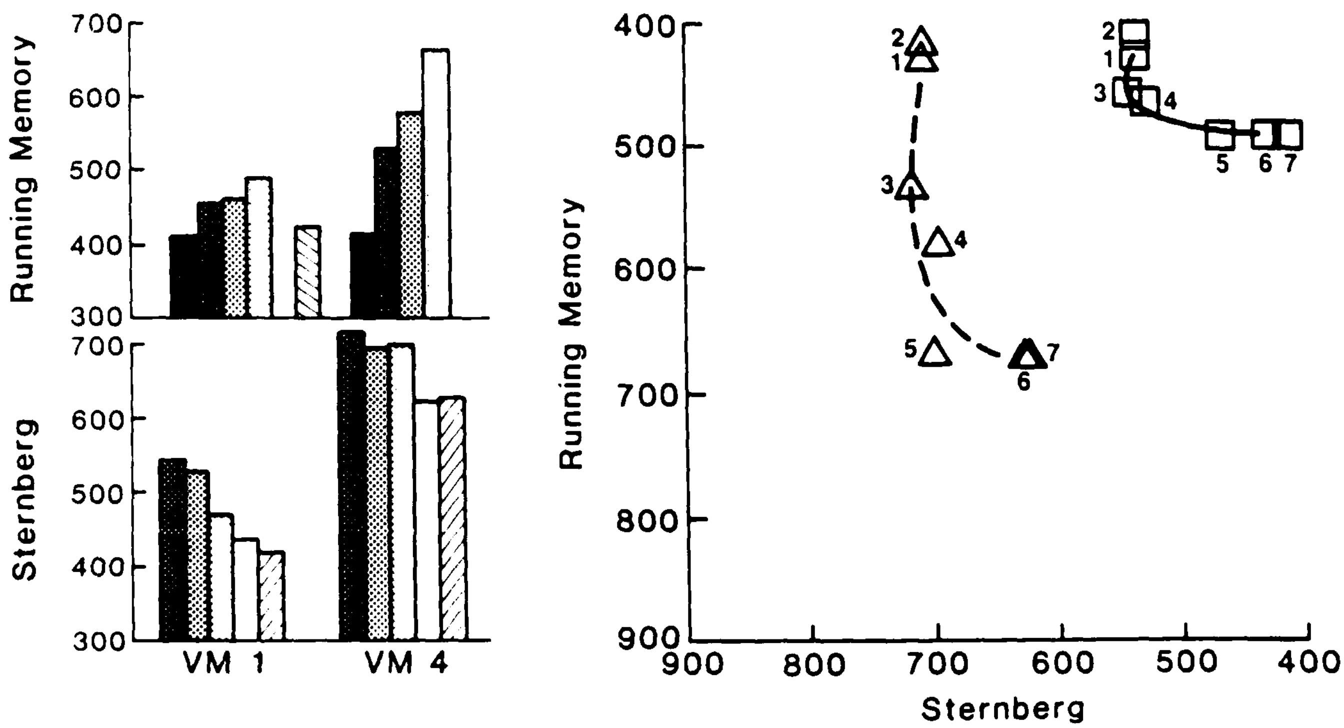
Session 4 △ - - - -

REACTION TIME

Consistent Mapping

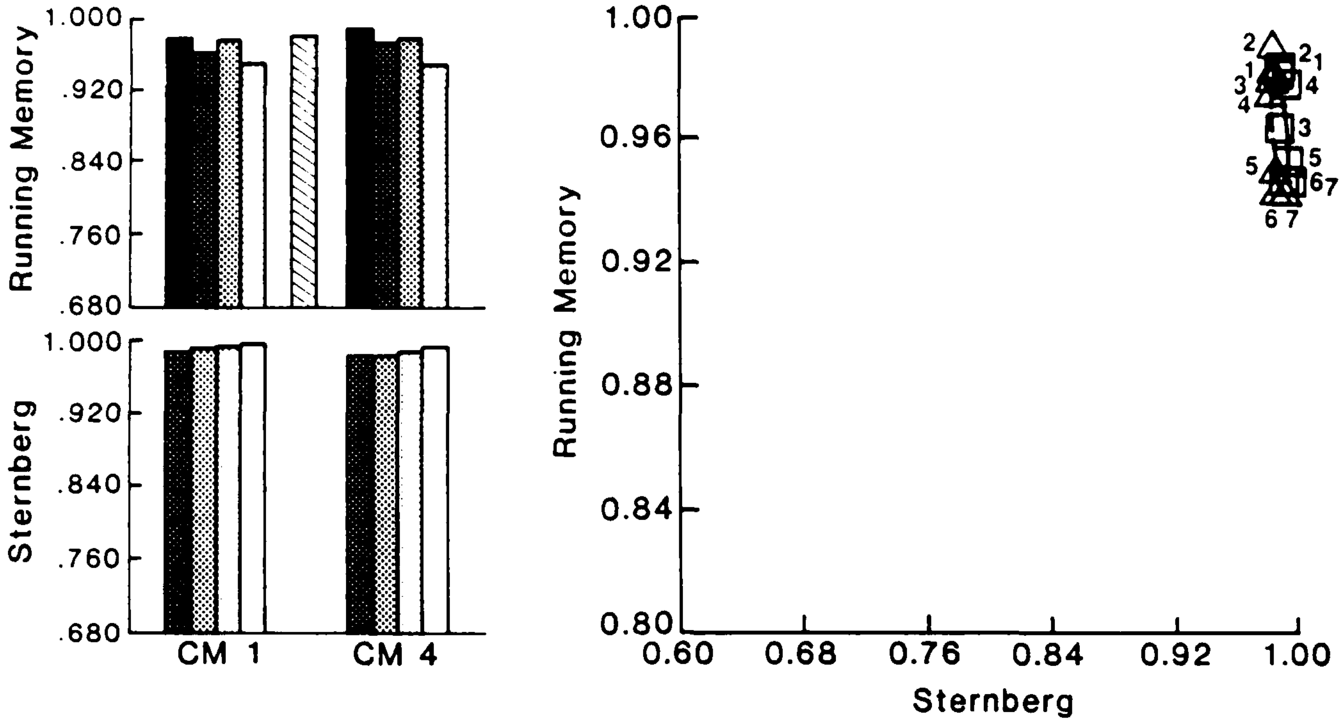


Variable Mapping

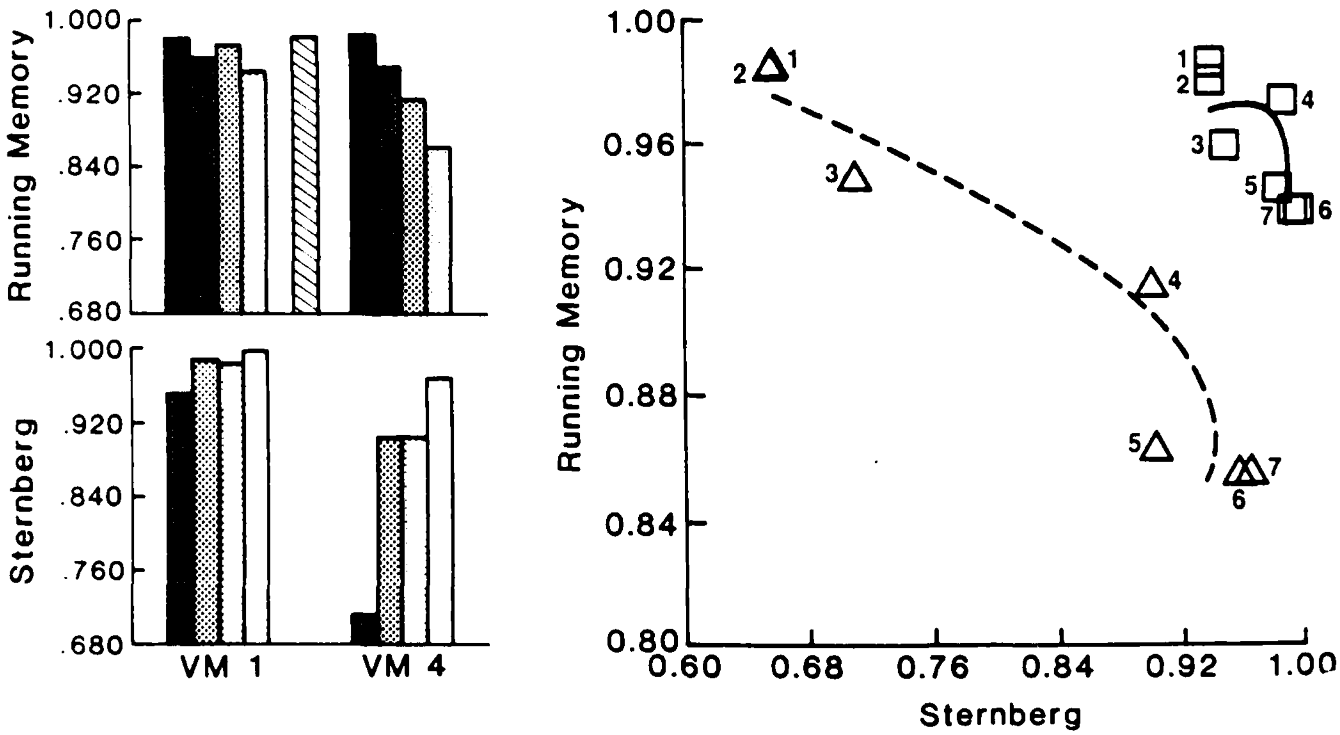


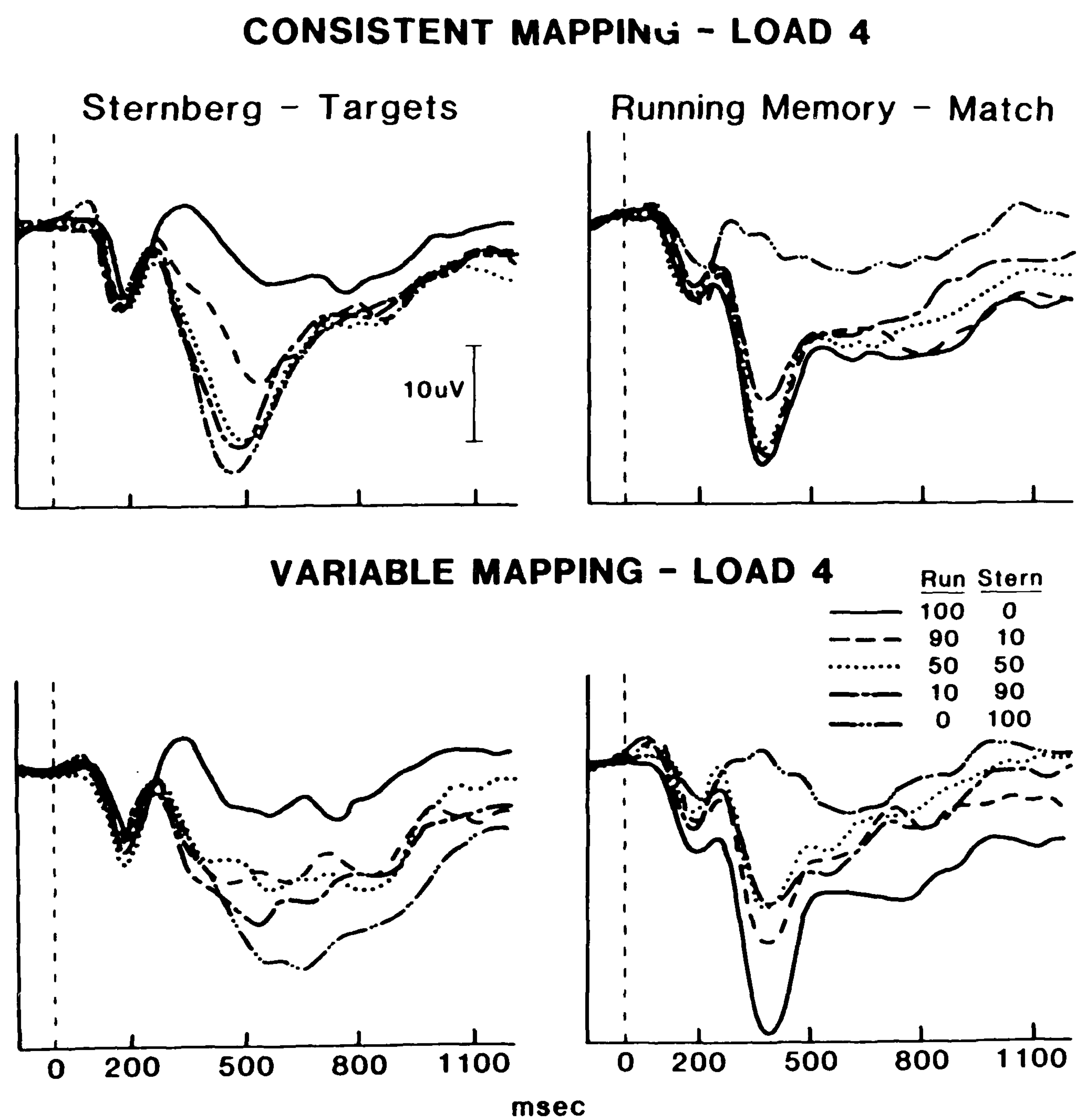
A'

Consistent Mapping



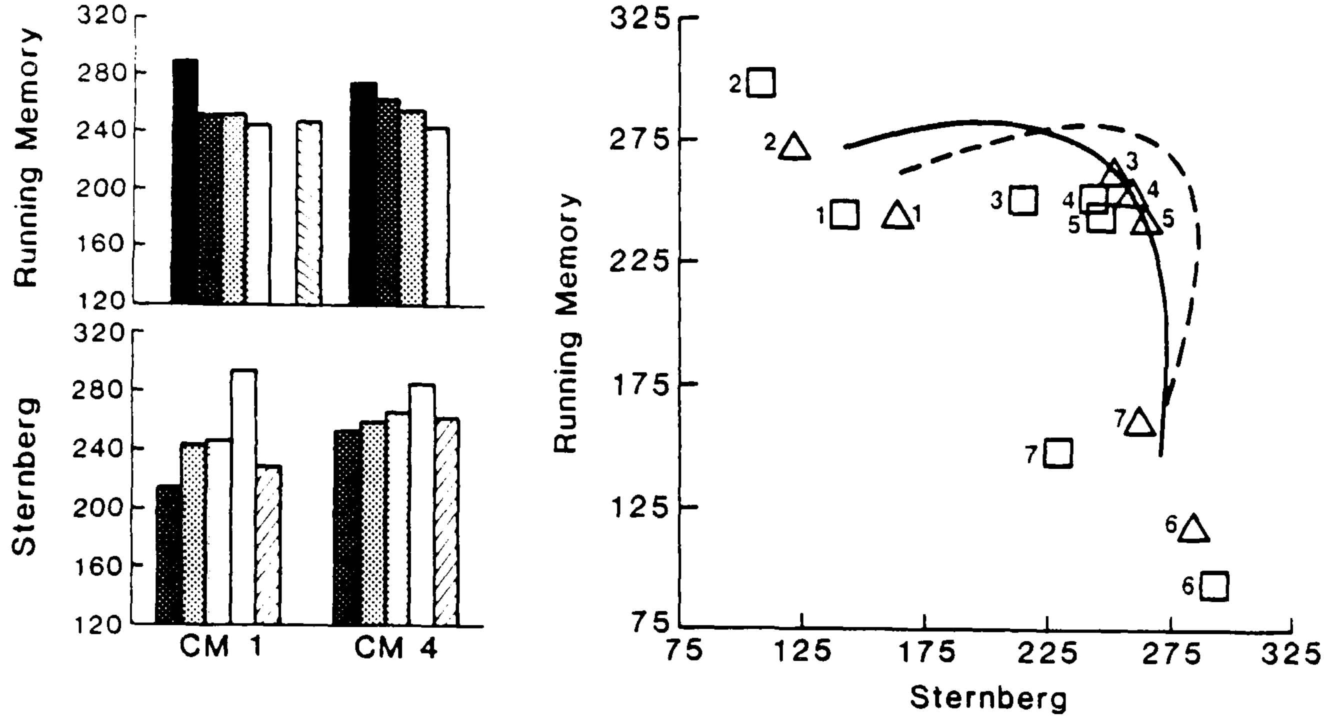
Variable Mapping



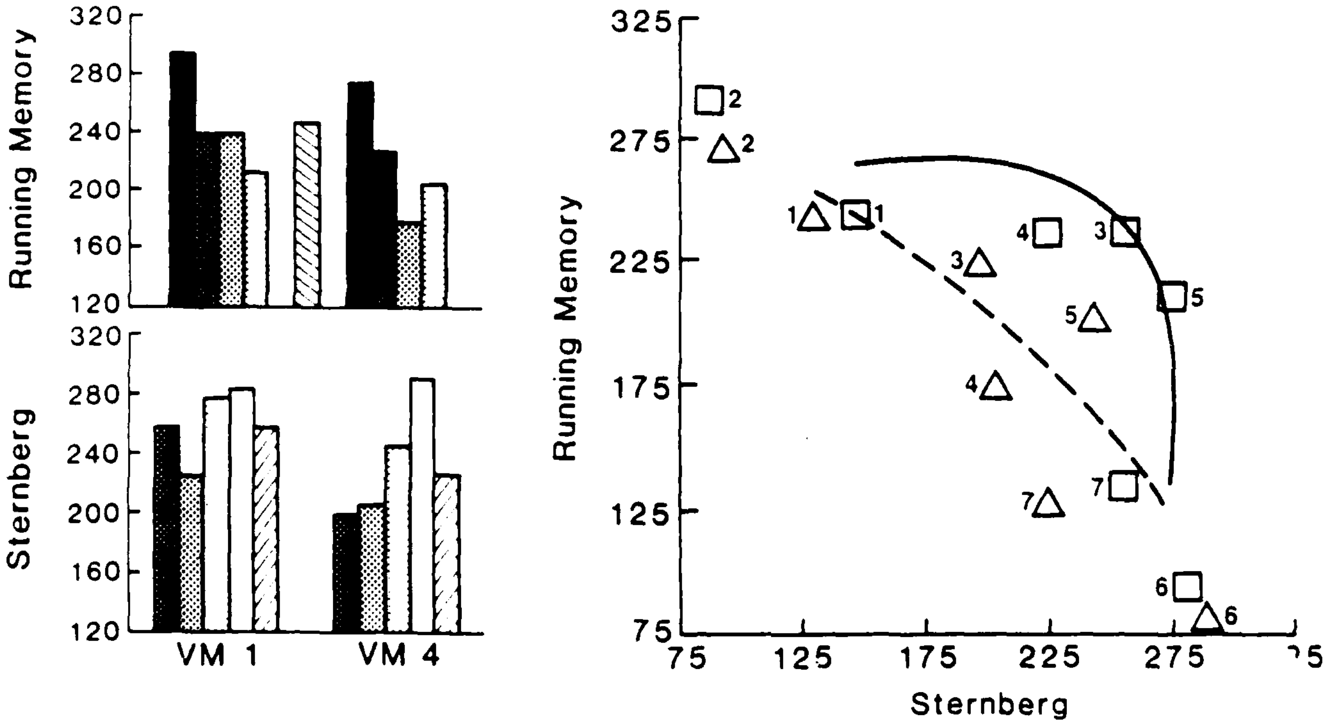

CONSISTENT MAPPING - LOAD 4
Sternberg - Targets
Running Memory - Match
10uV
VARIABLE MAPPING - LOAD 4
Run Stern
100 0
90 10
50 50
10 90
0 100
0 200 500 800 1100
0 200 500 800 1100
msec

P300 AMPLITUDE

Consistent Mapping

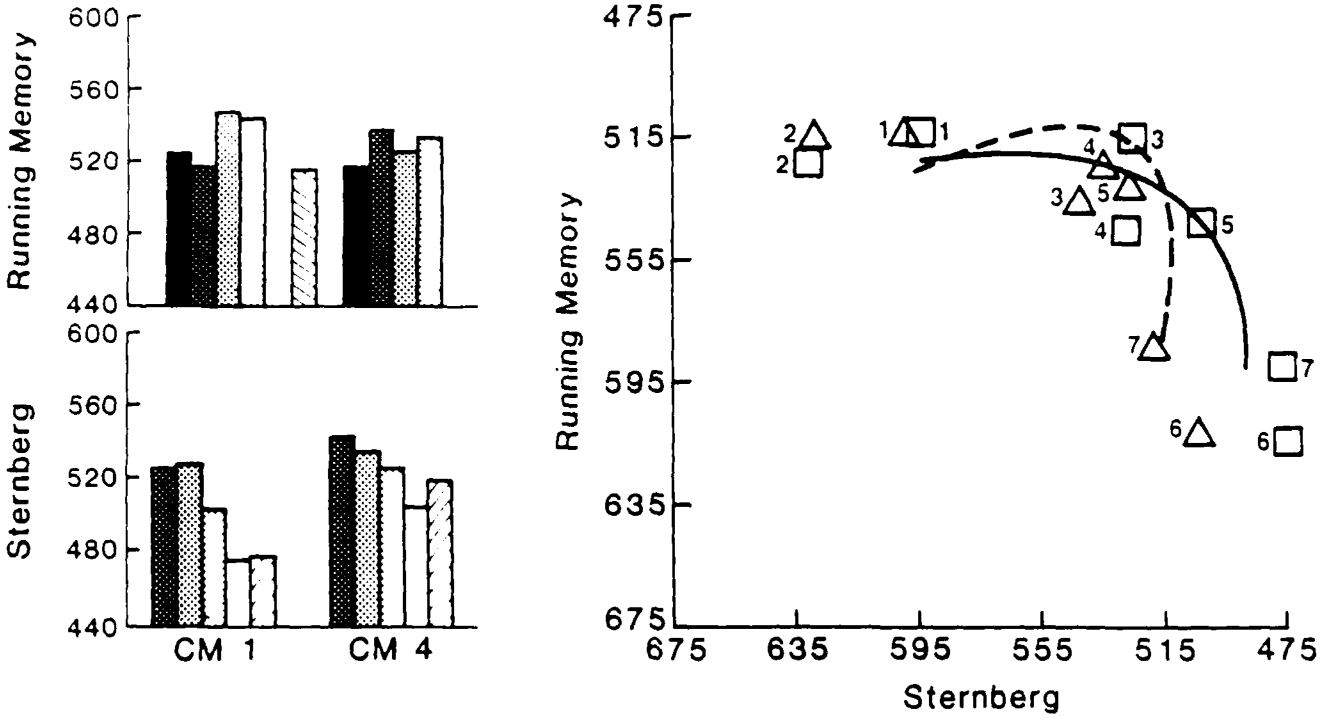


Variable Mapping

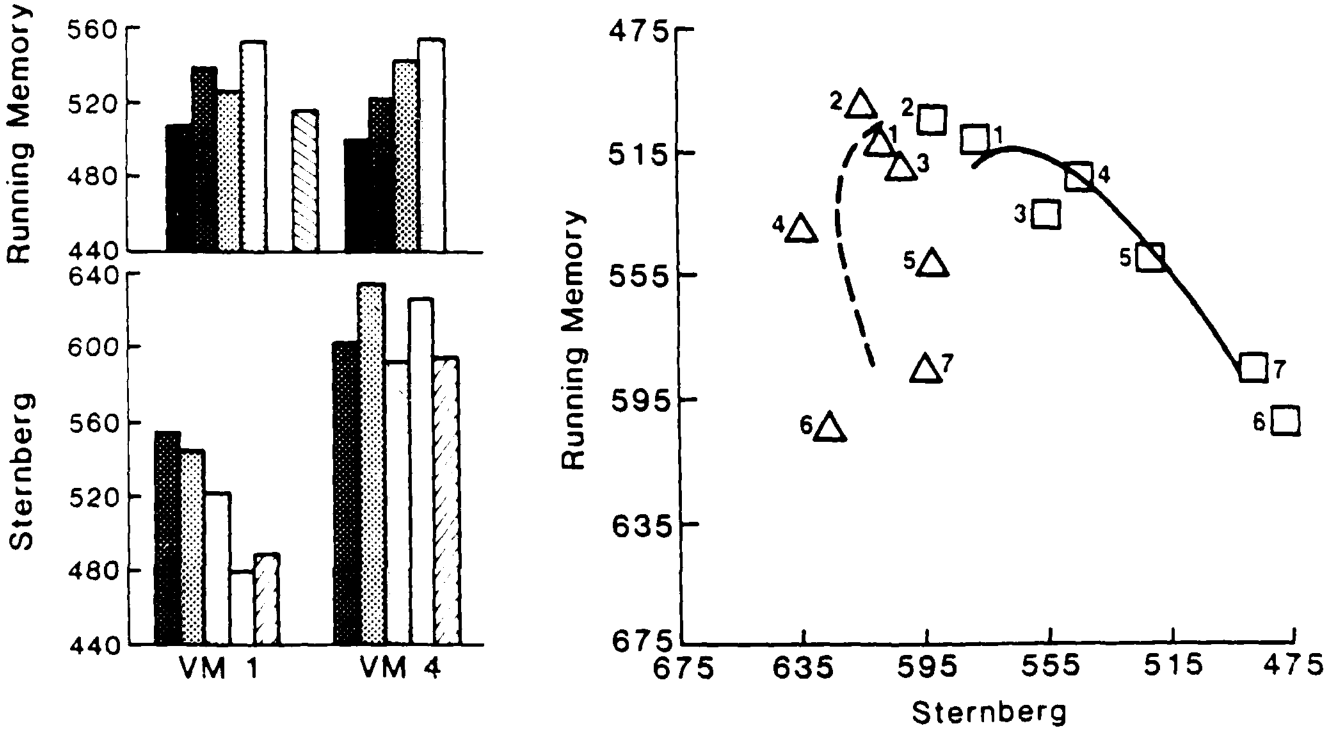


P300 LATENCY

Consistent Mapping

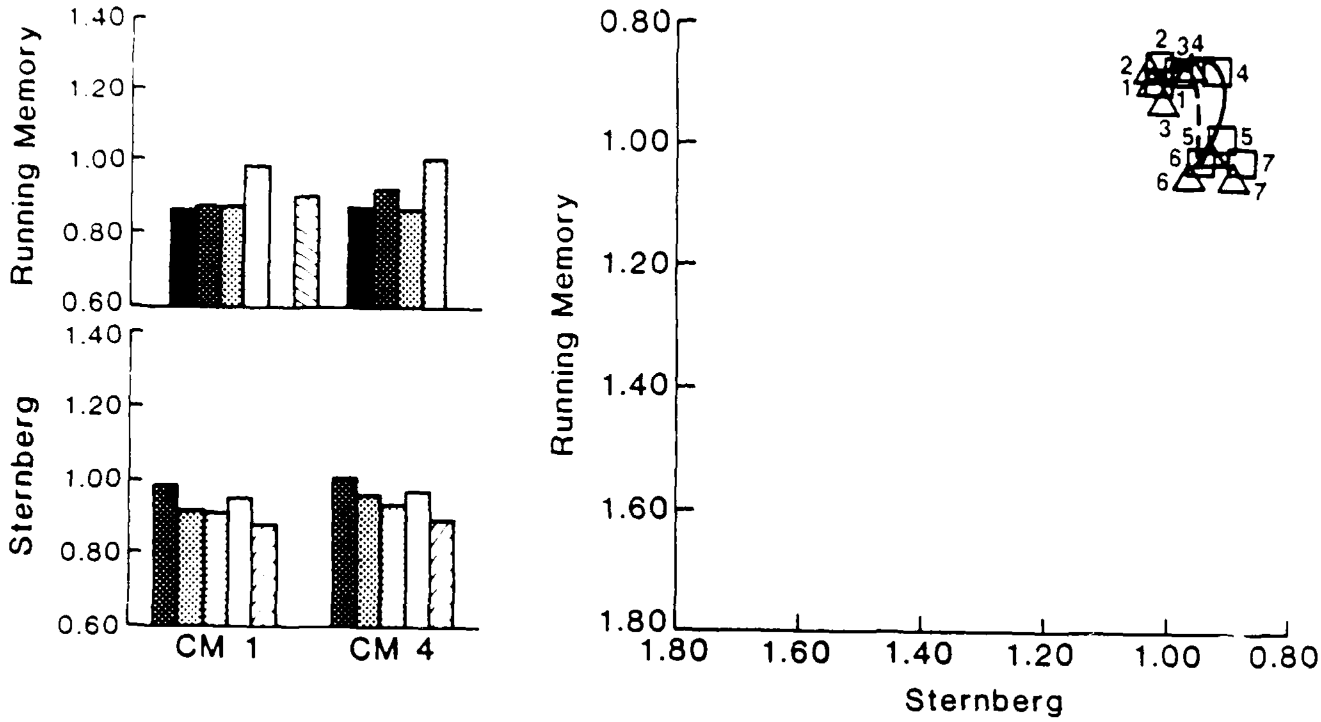


Variable Mapping

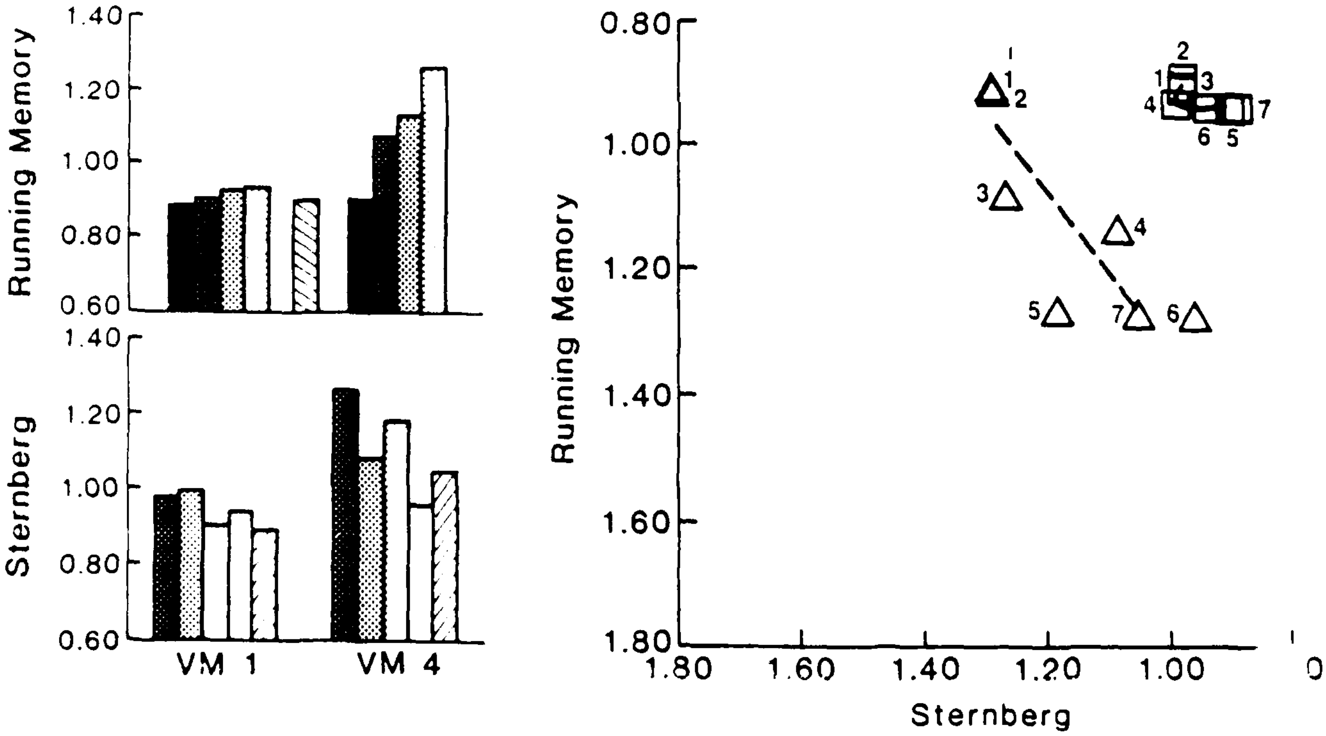


RT/P300 RATIO

Consistent Mapping



Variable Mapping



Adult Age Differences in the Development of Automaticity: A Psychophysiological Assessment

David L. Strayer, William J. Gehring, Arthur F. Kramer, and Gregory A. Miller

University of Illinois at Urbana-Champaign

Abstract

This study provides a fine-grained analysis of the age-related differences in the acquisition of automatic processing. Following consistent mapping (CM) training, young subjects develop automatic processing. While elderly subjects improve under CM training conditions, this is not due to the development of automatic processing, but to a reduction in response-related processing. Older subjects adopted a more conservative response bias than the young, which may interfere with the development of automaticity.

One of the most ubiquitous age-related changes in performance is the slowing of behavior with senescence. However, automatic skills which are acquired in young adulthood appear to be resistant to these decrements (Casey et al., 1987). Several investigators have sought to determine if the elderly can develop automatic processing skills and thereby eliminate the age-related decrements in performance. Studies which have addressed this issue have found that the elderly generally do not attain the same level of asymptotic performance in tasks that require automatic processing, nor do they improve at the same rate as young adults (e.g., Madden & Nebes, 1980; Nobel et al., 1964; Plude et al., 1983).

The purpose of this study is to provide a fine-grained analysis of the age-related differences in the development of automaticity. We will employ the converging methodologies of additive factors logic and the P300 component of the Event-Related Brain Potential (ERP) to localize the changes in information processing with practice and S-R mapping. Several lines of converging evidence suggest that the latency of the P300 component is sensitive to stimulus evaluation processes, but relatively insensitive to response-related processes (e.g., Magliero et al., 1984). Since a major portion of the slowing with age has been localized to response-related processing (Strayer et al. 1987), we seek to determine the extent to which automatic processing can bypass this information processing bottleneck. We further ask whether the stimulus evaluation processes can be automated in the elderly.

Method

Subjects

Eight young (mean age = 20.6, sd=1.5) and eight elderly (mean age = 73.1, sd=6.7) subjects participated in the experiment. All subjects had normal or corrected-to-normal vision and were in good health. Subjects were paid for their participation.

Procedure

Subjects performed a variant of the Sternberg memory search task (1966) under both consistent mapping (CM) and varied mapping (VM) conditions. Each block of trials consisted of a memory set presented for 3 seconds, followed by 30 probe trials. On each trial two probe letters were simultaneously presented within 1.5 degrees of visual angle. Subjects were instructed to press one button if either of the probes was a member of the memory set and press another button if neither probe was a member of the memory set. The experiment was conducted in four sessions (days), resulting in 5760 CM trials and 5760 VM trials.

ERP Recordings

In sessions one and four, EEG was recorded from Fz, Cz, and Pz sites. EEG and EOG were sampled every 10 msec for 1300 msec, beginning 100 msec prior to stimulus onset. EOG artifacts were corrected off-line. Single trial estimates of P300 latency were derived using a peak picking algorithm at the Pz electrode within a window from 300 to 1150 msec. Average ERPs were generated for each experimental condition. Each subject contributed a maximum of 360 trials to each average.

Results

Our analyses will focus on the first and last sessions of practice. The design is a 2 (age: young vs old) X 2 (session: first vs last) X 2 (mapping: CM vs VM) X 2 (set size: 2 vs 4) X 2 (response type: target vs distractor) split-plot factorial. The results will be hierarchically organized. Within each dependent measure we will examine the effects of practice and S-R mapping, followed by an analysis of age-related differences in these effects. For all analyses, a significance level of .05 is adopted.

Reaction Time

Table 1 presents mean reaction time to targets as a function of age, session, mapping, and set size. The linear regression slopes for memory set size are also presented.

Table 1

	S-R Mapping:	CM			VM		
	Memory Set Size:	2	4	Slope	2	4	Slope
Young	Session 1	469	533	31.9	494	630	68.2
	Session 4	414	434	9.9	454	572	58.8
Old	Session 1	652	737	42.8	685	804	59.2
	Session 4	558	597	19.2	619	752	66.6

Reaction time decreased with practice $F(1,14)=19.1$, was shorter for CM than for VM conditions $F(1,14)=114.9$, and increased as a function of set size at a greater rate in VM conditions than in CM conditions $F(1,14)=61.0$. The difference between CM and VM conditions increased from session 1 to session 4, $F(1,14)=17.0$, as did the CM and VM difference as a function of memory set size, $F(1,14)=10.3$. In addition, reaction time was shorter for targets than for distractors, $F(1,14)=113.9$.

Elderly responded more slowly than young, $F(1,14)=19.5$. Both groups improved with practice. The absolute level of reduction in the memory slope for CM conditions was equivalent; however, the ratio of CM slopes in session 1 vs session 4 revealed a greater proportional reduction for the young (3.22) than for the elderly (2.23). Furthermore, the ratio of VM to CM slopes in session 4 was larger for the young (5.94) than for the elderly (3.47). This reflects differences in asymptotic levels of performance. The session 4 CM slopes of the elderly were twice the slopes of the young.

Detection Sensitivity

The non-parametric measure A' was adopted to determine changes in detection sensitivity. A' ranges from .50 for chance accuracy to 1.0 for perfect detection accuracy. A' was quite high throughout the experiment, ranging from .86 to .99. A' decreased as a function of set size, $F(1,14)=94.9$, and was larger for CM than VM conditions, $F(1,14)=100.0$; however, the decrease in A' as a function of set size was larger for VM than CM conditions, $F(1,14)=41.7$. A' also increased from session 1 to session 4, $F(1,14)=30.3$, and this was more evident for VM than CM conditions, $F(1,14)=4.45$. This interaction is probably due to a ceiling effect for CM conditions, since initial A' values were quite high.

A' increased from session 1 to session 4 more for the elderly than for the young, $F(1,14)=7.7$, reflecting poorer detection sensitivity for the elderly early in training and greater detection sensitivity for the elderly following training. Furthermore, elderly were more affected by S-R mapping, $F(1,14)=6.5$, and by set size, $F(1,14)=7.3$, than the young. These latter effects are heavily influenced by VM performance in session 1, where memory set size produced its greatest effect on the elderly.

Response Bias

The non-parametric measure B'' was used to assess subjects' response bias. Larger values of B'' reflect a more conservative response bias. In session 1, subjects adopted a more conservative response bias for CM than VM conditions; however, in session 4 the subjects responded more conservatively in VM conditions, $F(1,14)=15.5$. In addition, B'' decreased with increasing memory set size for VM conditions, but not for CM conditions, $F(1,14)=12.9$, indicating that subjects adopted a more risky response strategy to compensate for the more difficult VM condition.

Elderly responded more conservatively than young, $F(1,14)=6.2$. Furthermore, young subjects became less conservative with practice, but elderly subjects became more conservative following practice, $F(1,14)=4.53$. Elderly also tended to respond more conservatively in CM conditions than VM conditions, while young subjects tended to respond more conservatively in VM conditions than in CM conditions, $F(1,14)=6.6$. This was coupled with a tendency of the elderly to become more conservative as memory set size increased in CM conditions, $F(1,14)=9.2$.

P300 Latency

Table 2 presents mean P300 latency to targets as a function of age, session, mapping, and set size. The linear regression slopes for memory set size are also presented.

Table 2

	S-R Mapping:	CM			VM		
	Memory Set Size:	2	4	Slope	2	4	Slope
Young	Session 1	590	654	32.1	614	697	41.5
	Session 4	610	621	5.6	611	707	47.8
Old	Session 1	708	738	14.7	708	754	23.0
	Session 4	631	654	11.6	678	738	29.6

P300 latency increased with set size, $F(1,14)=80.1$; however, the effect was greater for VM than CM conditions, $F(1,14)=13.2$, particularly in session 4, $F(1,14)=6.1$. P300 latency was shorter for CM than VM conditions, $F(1,14)=23.8$, and this was more pronounced in session 4, $F(1,14)=16.1$. In addition, P300 latency was shorter for targets than distractors, $F(1,14)=93.7$.

P300 latency was shorter for young than elderly, $F(1,14)=8.5$. However, the effect of memory set size was greater for the young than for the elderly, $F(1,14)=8.3$. Furthermore, there were differential effects of mapping, response type, and session for young and elderly, $F(1,14)=8.5$, and mapping and set size, $F(1,14)=5.0$. For the young, the effect of memory set size produced equivalent effects on P300 latency for CM and VM conditions in session 1, but in session 4 the memory set size effect was substantially reduced in the CM condition. In contrast, the effects of memory set size were relatively constant across session for both CM and VM conditions for the elderly. The improvement with practice can be illustrated by comparing the ratio of the CM slopes in session 1 and 4 for the two age groups. The CM session1/session4 ratio for young was 5.73 and 1.27 for elderly.

RT/P300 Ratio

A single-trial ratio of RT to P300 latency was calculated to determine the proportion of stimulus evaluation accomplished at the moment of response in each condition. A ratio of 1.0 indicates that the RT response and peak of the P300 co-occurred. Ratios less than 1.0 indicate that the response preceeded P300 latency and ratios greater than 1.0 indicate that the response followed P300 latency. Previous research suggested that a large portion of the age-related slowing is due to response-related processing (Strayer et al., 1987). This analysis was conducted to determine if post-stimulus evaluation processing was reduced following consistent practice in the elderly.

The RT/P300 ratio decreased with practice, $F(1,14)=18.8$, and increased with set size, $F(1,14)=53.2$. This latter effect was more evident in session 1, $F(1,14)=6.7$. Further, the RT/P300 ratio was larger for VM than CM conditions, $F(1,14)=41.0$, and this was more pronounced in set size 4, $F(1,14)=20.5$.

The RT/P300 ratio was larger for elderly than young, $F(1,14)=10.1$. The average RT/P300 ratio was 0.89 for the young and 1.12 for the elderly. This implies that elderly engaged in more post-stimulus evaluation processing prior to their response than the young. Age did not enter into any interactions, suggesting that the age differences in post-stimulus evaluation were not modified by practice or S-R mapping.

Discussion

Both young and elderly improved with practice. This improvement was observed as decreases in reaction time and increases in response accuracy. However, elderly responded more slowly and more accurately than young following practice. This suggests that the elderly were trading response speed for accuracy, adopting a more conservative response bias than the young.

Reductions in the reaction time memory set size slope were apparent for both age groups, but the young improved more rapidly and achieved a lower asymptote. Reductions in the P300 latency memory set size slopes were apparent only for the young. This suggests that the stimulus evaluation processes become automated for the young, but not for the elderly. The improvement in reaction time performance for the elderly was attributed to a reduction in response-related processing which was apparent as a trimming of long latency responses. This pattern of data suggests that the improvements in performance of the elderly are not the result of automatic processing, but rather are due to a reduction in post-stimulus evaluation processing.

One possible interpretation for why the elderly do not acquire automatic processing may be their conservative response bias. It has been suggested that such a conservative strategy interferes with the development of automaticity (Shiffrin et al., 1984). One prediction from this interpretation is that if elderly subjects adopted a less conservative response bias, then their acquisition rates and asymptotes should be similar to young. Further, if young subjects adopt a more conservative response bias, then the rate of improvement and the asymptote should be similar to that of the elderly.

Thus age differences in the development of automaticity appear to be the result of strategic changes in information processing.

The research was supported by NASA Ames Research Center under contract number 1-5-25-477 with Dr. Michael Vidulich as technical monitor and by the Air Force Office of Scientific Research under contract number F49620-79-C-0233 monitored by Dr. Al Fregly.

References

Casey, B.J., Rogers, W.A., & Fisk, A.D. (1987). Cognitive skill acquisition: A developmental approach. Proceedings of the 31st Annual Meeting of the Human Factors Society. New York, New York.

Madden, D.J., & Nebes, R.D. (1980). Aging and the development of automaticity in visual search. Developmental Psychology, 16, 377-384.

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.

Nobel, C.E., Baker, B., & Jones, T.A. (1964). Age and sex parameters in psychomotor learning. Perceptual and Motor Skills, 19,935-945.

Plude, D.J., Kaye, D.B., Hoyer, W.J., Post, T.A., Saynisch, M.J., & Hahn, M.V. (1983). Aging and visual search under consistent and varied mapping. Developmental Psychology, 19, 508-512.

Shiffrin, R.M., Dumais, S.T., & Schneider, W. (1984). Characteristics of automatism. In J. Long and A. Baddeley (Eds.), Attention and Performance IX. Erlbaum: Hillsdale, New Jersey.

Sternberg, S. (1966). High speed scanning in human memory. Science, 153, 652-654.

Strayer, D.L., Wickens, C.D., & Braune, R. (1987). Adult age differences in speed and capacity of information processing: 2. An electrophysiological approach. Psychology and Aging, 2, 99-110.

DAT

FILM