ELSEVIER

International Journal of Psychophysiology 17 (1994) 35–46

INTERNATIONAL JOURNAL OF PSYCHOPHYSIOLOGY

P300 from an auditory oddball task: inter-laboratory consistency

Joel E. Alexander [a], John Polich [a,*], Floyd E. Bloom [a], Lance O. Bauer [b], Samuel Kuperman [c], John Rohrbaugh [d], Sandra Morzorati [e], Sean J. O'Connor [e], Bernice Porjesz [f], Henri Begleiter [f]

[a] *Department of Neuropharmacology, The Scripps Research Institute, La Jolla, CA 92037, USA*
[b] *Department of Psychiatry, University of Connecticut Health Center, Farmington, CT 06030, USA*
[c] *Department of Psychiatry, University of Iowa, Iowa City, IA 52242, USA*
[d] *Department of Psychiatry, Washington University, St. Louis, MO 63108, USA*
[e] *Department of Psychiatry, University of Indiana and VAMC, Indianapolis, IN 46202, USA*
[f] *Department of Psychiatry, State University of New York Health Sciences Center Brooklyn, NY 11203, USA*

(Accepted 22 October 1993)

Abstract

Event-related potentials (ERPs) were recorded from normal subjects for the purpose of evaluating measurement consistency among six laboratories located in different cities within the United States. At each laboratory location 15 male subjects were tested using a simple auditory stimulus discrimination task and identical electrophysiological equipment and recording methods. Assessment of the N1, P2, N2, and P3(00) potentials from both the target and standard stimuli resulted in no reliable differences among laboratories for component amplitudes, latencies, and scalp distributions. Quantitative evaluation of overall waveform and specific component morphology yielded good to excellent agreement across laboratories. The findings suggest that large-scale inter-laboratory human electrophysiological studies are feasible and may prove of value when using ERPs to evaluate cognitive function in humans.

Key words: Event-related potentials; P300; Consistency; Multi-laboratory recording

1. Introduction

When an auditory 'oddball' paradigm is used to elicit the P3(00) event-related brain potential (ERP), subjects are required to discriminate between stimuli that vary on some dimension. This procedure produces robust P3 components with relative ease and good consistency in a wide variety of normal and clinical populations [1,2]. Because the P3 component is thought to reflect the neural activity associated with attentional and memory mechanisms [3,4], it has proven quite useful in the assessment of psychiatric and neurologic disorders (e.g., Refs. 5–8). In particular, some reports have found that P3 amplitude from individuals who have a positive family history for alcoholism may be smaller than that from unaffected control subjects—a finding that suggests that this ERP component may be beneficial in the identification of the genetic predisposition for alcoholism (cf. Refs. 9–13; for a review see Ref.

* Corresponding author. Fax: +1 (619) 5546393.

SSDI 0167-8760(93)E0057-7

14). Thus, the P3 component is being used to guide the search for the underlying genetic factors that contribute to alcoholism.

As part of a large scale, multi-laboratory study in which the P3 and other ERPs are being employed to identify phenotypic markers for alcoholism (Consortium on the Genetics of Alcoholism or COGA), six ERP laboratories have been established in different locations throughout the United States. The COGA laboratories were designed to acquire 19 channels of electroencephalographic (EEG) data using a variety of ERP paradigms. Each laboratory was built using the same amplifiers, computer system, and software so that EEG and ERP data from large numbers of subjects could be obtained in an identical fashion from different areas of the country. The present study was designed to assess the consistency of the ERP data collected using an auditory stimulus discrimination task in different laboratories.

A number of studies have reported reasonable to high intra-subject reliability and inter-subject associations between the P3 and other ERP component measures derived from auditory stimuli [15–21]. However, considerable individual variability for P3 measures has been observed from factors such as recency of food consumption, season of the year, skull thickness, body temperature, cognitive capability, personality, age, and sex among other variables (e.g., Refs. 22–29). Hence, it is important to determine whether or not ERP recordings of subjects drawn from the same population (young adult males) at different laboratory locations produce similar component measures despite such contributions to ERP variation.

2. Methods

2.1. Subjects

A total of 90 young adult males (mean age = 22.6 y, S.D. = 1.8 y) from each of the six laboratories ($n = 15$/laboratory) located in California (CA), Connecticut (CT), Iowa (IA), Indiana (IN), Missouri (MO), and New York (NY) served as subjects and received pecuniary remuneration for their participation. All subjects reported normal hearing, no personal or familial psychiatric or neurologic problems, and were screened for alcohol and drug use, by means of a self-report questionnaire. Of the 90 subjects, 9 were left-handed (10%)—indicating that a normative sample was obtained for this variable.

2.2. Recording conditions and procedure

EEG activity was recorded at 19 electrode sites [30] using an electrode-cap, referred to the nose with a forehead ground, and impedance maintained at 5 kΩ or less. Electro-ocular (EOG) activity was monitored by two electrodes placed at the outer canthus and above the left eye. The filter bandpass was 0.02–50 Hz (3 dB down, 6 dB octave/slope). The EEG was digitized at 3.9 ms/point for 1500 ms, with a 187 ms prestimulus baseline. The same computer was used to average the ERP data on-line, control stimulus presentation, and perform artifact rejection. Trials on which the EEG or EOG exceed ± 73.3 μV were rejected automatically.

ERPs were elicited with 400 auditory stimuli presented binaurally and consisting of 600 Hz and 1600 Hz tones presented at 60 dB SPL (10 ms r/f, 60 ms plateau), with an inter-stimulus interval of 1.5 s. The target tone occurred randomly with a probability of 0.125. Half the subjects had the 600 Hz tone as the target, and half had the 1600 Hz tone as the target. Presentation of stimuli was concluded when 25 target and 75 standard artifact-free stimulus presentations were acquired. Subjects were instructed to press a key pad with their forefinger as rapidly as possible whenever a target tone was detected and to refrain from responding when the standard tone was presented.

3. Results

All analyses of variance employed Greenhouse–Geisser corrections to the degrees of freedom (df) to adjust for violations of the sphericity assumption inherent in repeated measures de-

signs [31]. Only the probability values from the corrected *df* are reported here.

3.1. Behavioral data

Task performance was virtually perfect with fewer than 2% of the target trials missed across subjects and laboratory locations (CA = 1.3%, CT = 0.0%, IA = 0.0%, IN = 2.4%, MO = 0.5%, and NY = 3.2%). Error rate did not differ among laboratory locations, $F(5,84) = 1.2$, $P > 0.25$. Mean reaction time (RT, S.E. = standard error) was reasonably consistent across laboratory locations (CA = 403 ms, S.E. = 9.0 ms; CT = 373 ms, S.E. = 6.0 ms; IA = 362 ms, S.E. = 11.1 ms; IN = 354 ms, S.E. = 8.4 ms; MO = 411 ms, S.E. = 7.1 ms; and NY = 385 ms, S.E. = 9.2 ms) and also did not differ among laboratories, $F(5,84) = 1.3$, $P > 0.25$.

3.2. Component identification

Waveforms for both targets and standards were analyzed at The Scripps Research Institute labo-

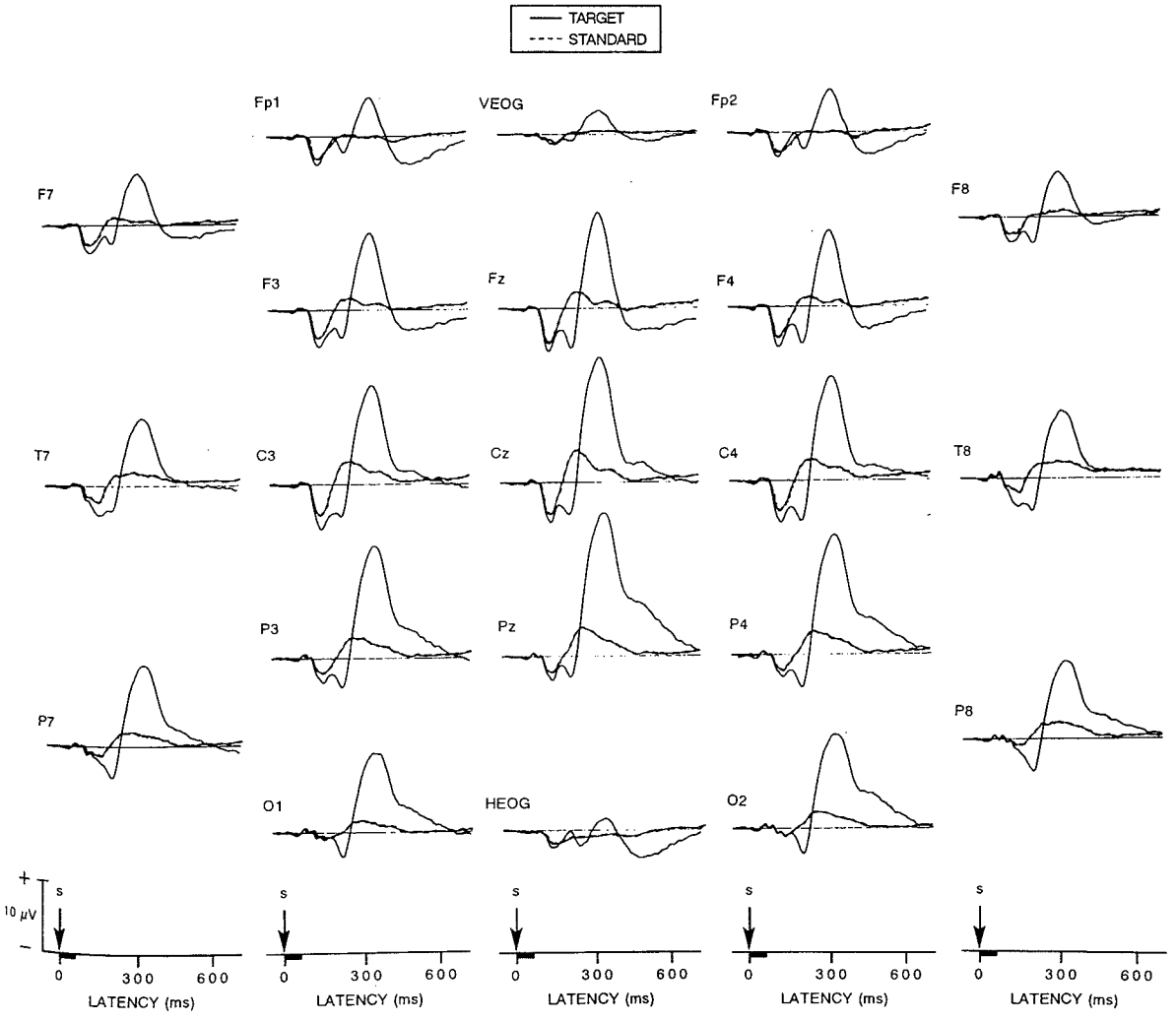
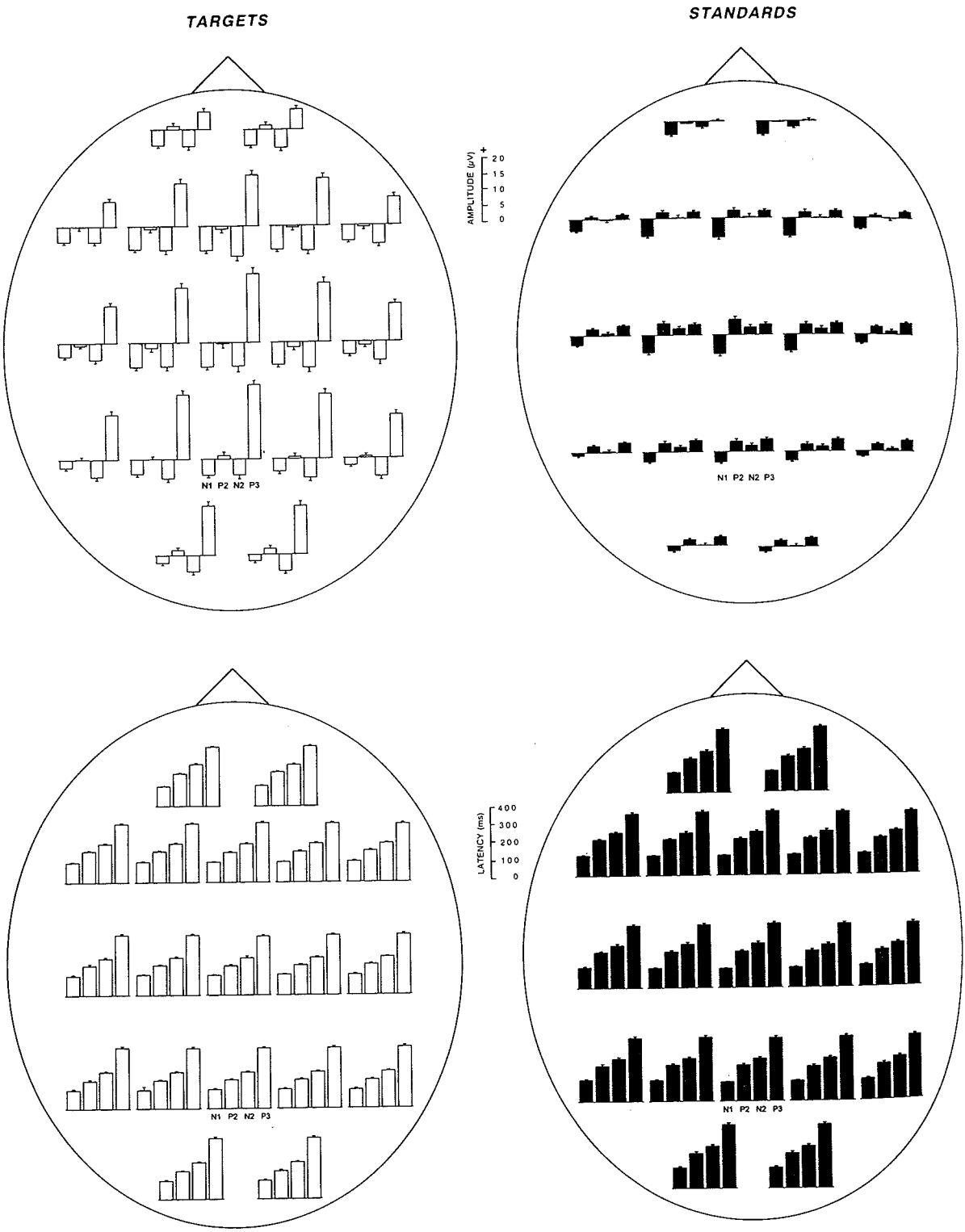


Fig. 1. Grand average event-related potentials from the target and standard stimuli at each electrode position for all subjects ($N = 90$) from the six different laboratory locations.


TARGETS
STANDARDS
AMPLITUDE (μV)
+
20
15
10
5
0
N1 P2 N2 P3
N1 P2 N2 P3
LATENCY (ms)
400
300
200
100
0
N1 P2 N2 P3
N1 P2 N2 P3

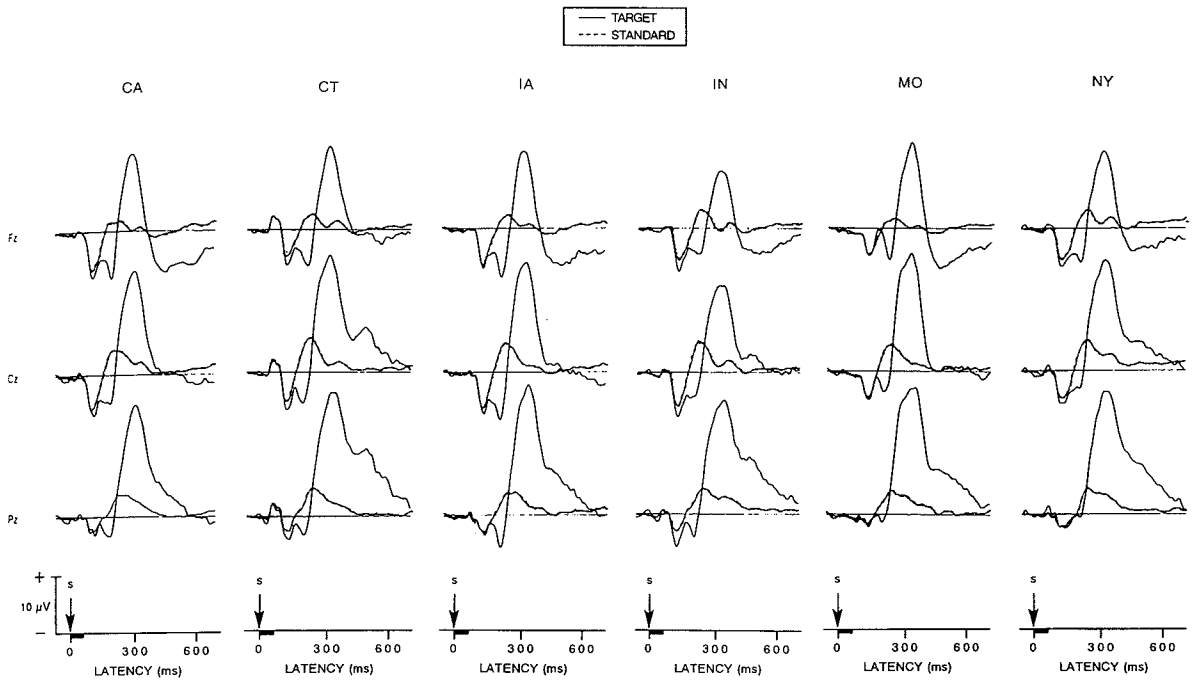


Fig. 3. Grand average event-related potentials from the target and standard stimuli for the midline electrode sites from each laboratory location ($n = 15$). Note that despite relatively minor overall morphological variation among the different laboratory sites, the individual components evidence considerable consistency across laboratories (see Figs. 4, 5, and 7).

ratory through the use of computer-assisted scoring. Amplitudes and latencies of the N1, P2, N2, and P3 components at each electrode site were identified by locating the most negative or positive component within the latency windows of 75–125, 100–200, 140–300, and 275–400 ms, respectively. Amplitude was measured relative to the prestimulus baseline, with peak latency defined as the time point of maximum positive or negative amplitude within the latency window. The ERP data were analyzed after signal averaging as described above, with no other filtering or smoothing algorithms applied.

The grand average ($N = 90$) ERP waveforms for the target and standard stimuli at each electrode position are illustrated in Fig. 1. The mean amplitudes and latencies (+2 standard errors) for the N1, P2, N2 and the P3 components are presented in Fig. 2. As is apparent from these figures, clear and robust ERP components were obtained at all electrode sites.

3.3. Inter-laboratory differences

The grand average ERPs from each laboratory location ($n = 15$) for the target and standard waveforms for Fz, Cz, and Pz electrode sites are illustrated in Fig. 3. The mean amplitude and latency (+1 standard error) for the N1, P2, N2, (Cz electrode) and P3 (Pz electrode) component values obtained from the six laboratory locations are presented in Fig. 4.

A two-factor mixed design (laboratory location × electrode) analysis of variance was performed

Fig. 2. Mean (+2 standard errors) amplitudes and latencies for the N1, P2, N2 and P3 components from the target and standard stimuli for all subjects. These data are presented to illustrate the quantitative information derived from analysis of the individual ERPs that were used to construct Fig. 1.

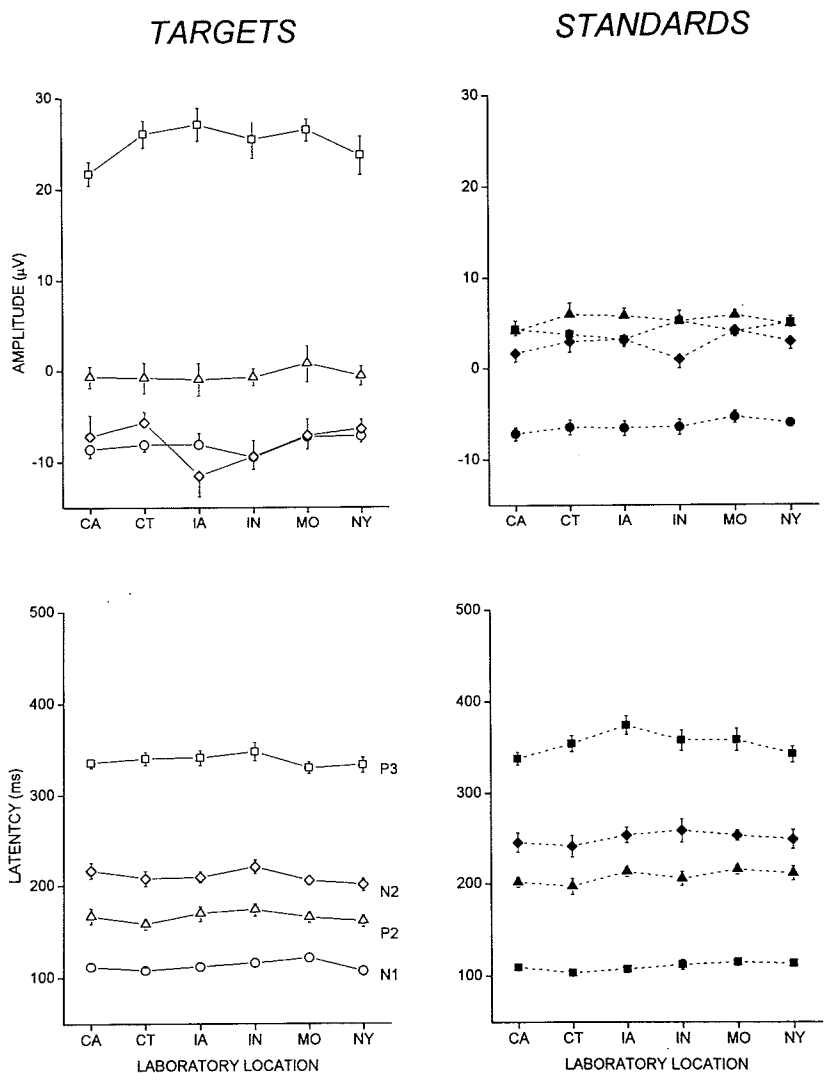


Fig. 4. Mean (+1 standard error) amplitude and latencies for the N1, P2, N2 (Cz electrode) and P3 (Pz electrode) components as a function of laboratory location.

on the amplitude and latency data from the Fz, Cz, and Pz electrodes for each component elicited by the target and standard stimuli. Laboratory location ($df = 5, 84$) did not affect P3 amplitude or latency for either the target or standard stimuli ($P > 0.15$ in all cases), nor did laboratory location interact ($df = 10, 168$) with electrode position for any of these measures ($P > 0.05$ in all cases). Similarly, laboratory location did not affect the amplitudes or latencies of the N1, P2, or N2 components for either the target or standard stimuli ($P > 0.10$ in all cases), nor did laboratory location interact with electrode position ($P > 0.20$ in all cases). Thus, no statistically reliable differences among laboratory locations were obtained for any of the components measured at the midline electrode sites.

Similar analyses (laboratory location × electrode) were performed on the amplitude and latency data from the remaining 16 electrodes (Fp1, Fp2, F3, F4, F7, F8, C3, C4, P3, P4, T7, T8, P7, P8, O1, O2) for the P3 and N1, P2, N2

components (the results from each laboratory for these electrode sites are not illustrated). Except for N1 latency from the target stimuli ($P < 0.05$, with no specific laboratory location effects found using Newman–Keuls post-hoc comparisons), laboratory location ($df = 5, 84$) again demonstrated no significant main effects for amplitude or latency from any of these components for the target or standard stimuli ($P > 0.05$ in all cases). In addition, none of the interactions between laboratory location and electrode ($df = 75, 1260$) were significant for either the amplitude or latency data from the target or standard stimuli for all components ($P > 0.05$ in all cases). Hence, it

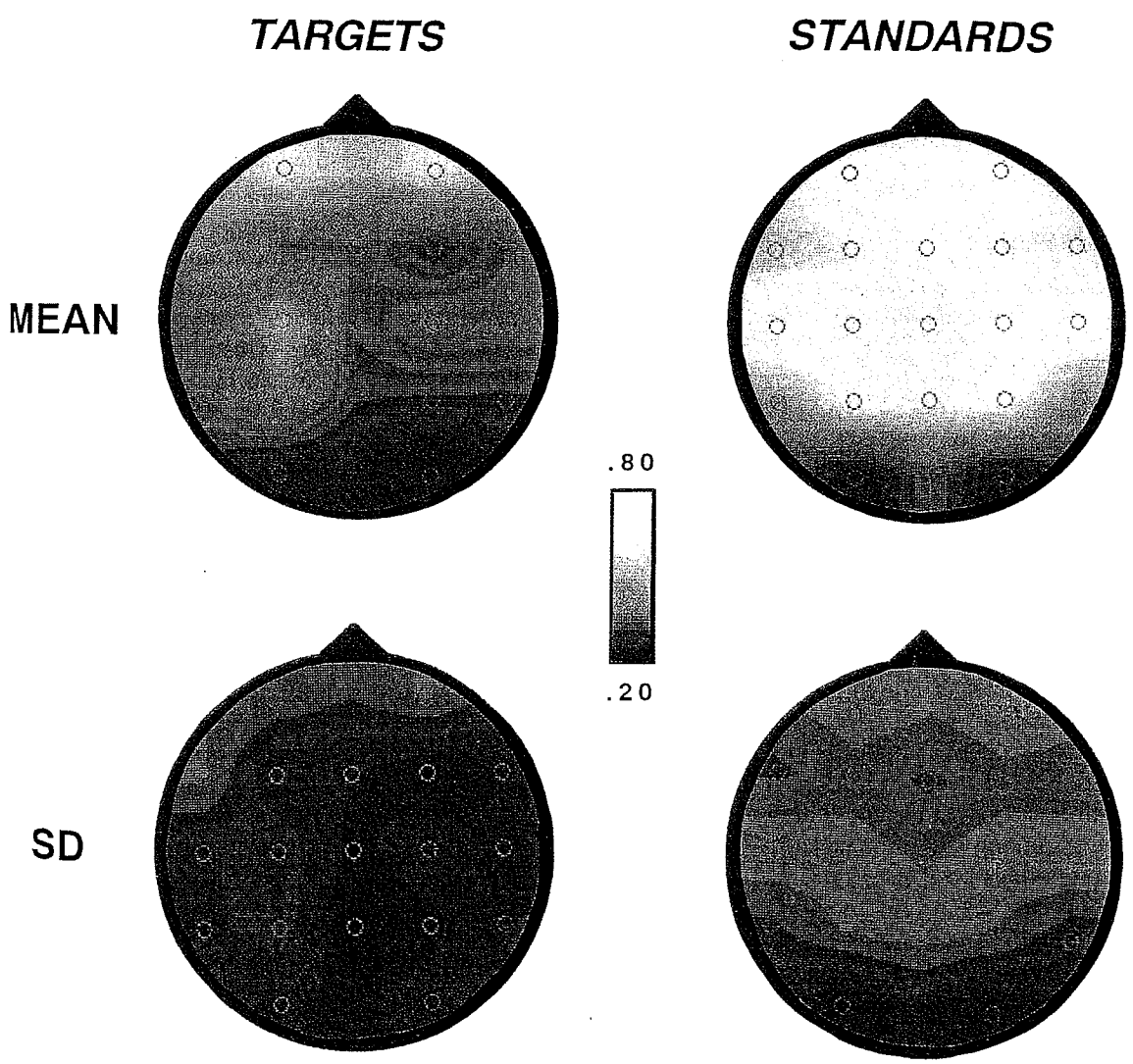


Fig. 5. Mean and standard deviation (S.D.) correlation values at each electrode site from the target and standard stimuli for the inter-laboratory comparisons of the average ERP waveforms. These values were obtained by correlating the grand average waveforms from each laboratory location on a point-by-point basis across laboratories thereby producing the overall mean and standard deviation correlational values for each electrode (see text for details).

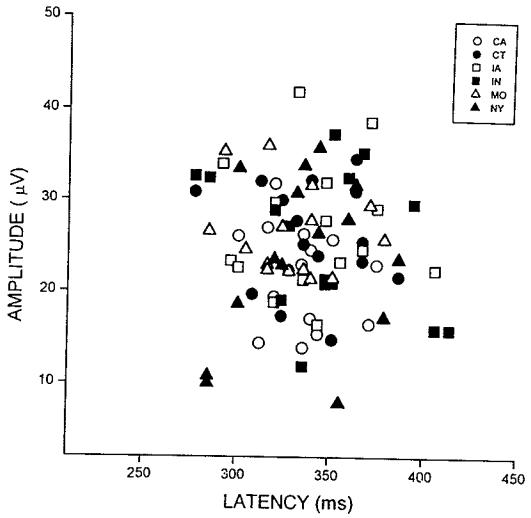


Fig. 6. P3 amplitude and latency values (Pz electrode) from all subjects recorded at each laboratory location ($N = 90$). Note that each laboratory demonstrated similar inter-subject amplitude and latency variability, since data points from each laboratory location occur over the range of possible amplitude and latency values.

is reasonable to conclude that the few inter-laboratory differences observed for these electrode sites were more spurious than real.

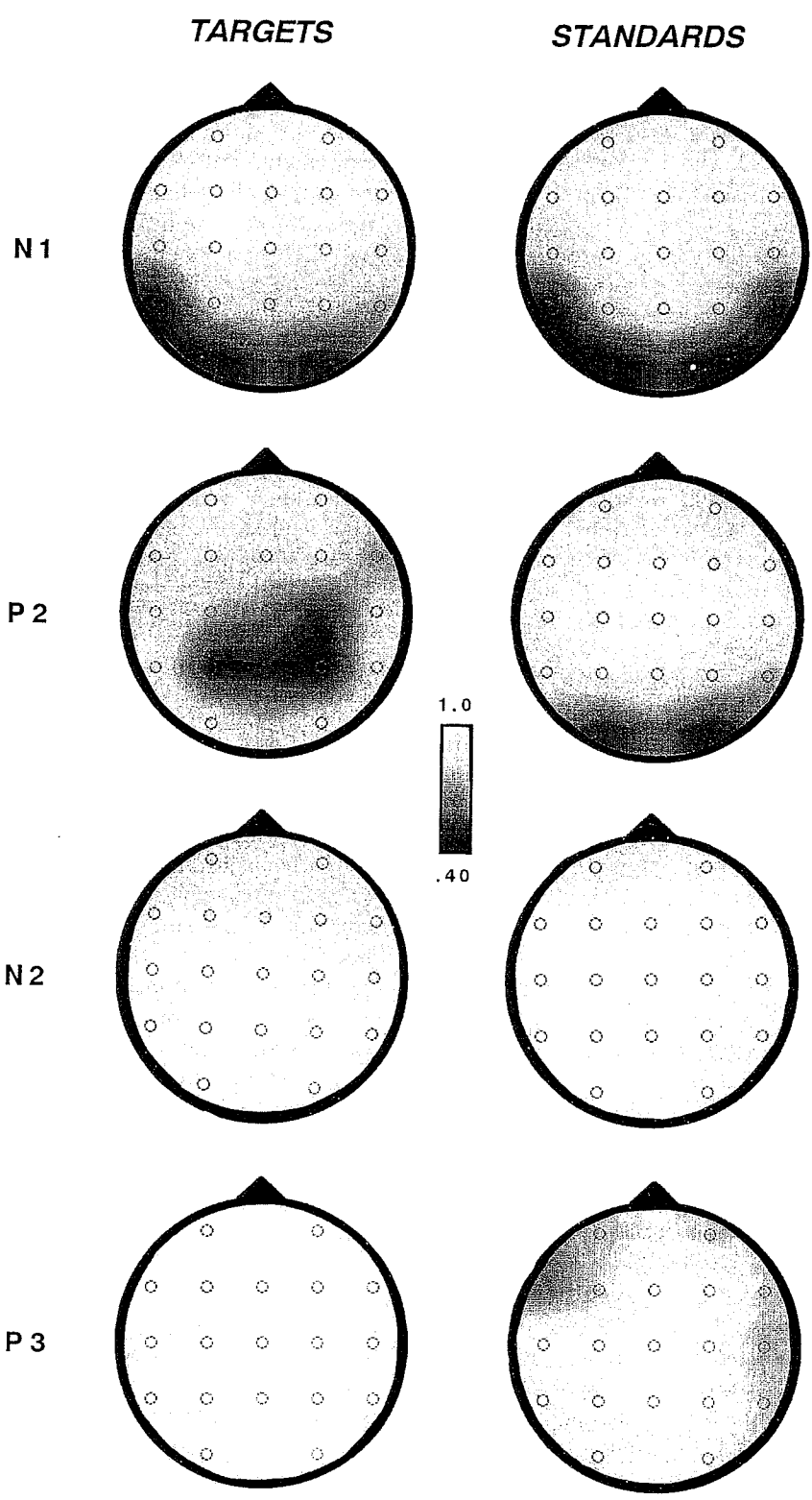
3.4. Inter-laboratory agreement

Waveform consistency. As might be expected from the waveform data illustrated in Fig. 3, the various laboratory locations produced very similar ERPs across subjects with some inter-laboratory variation observed for various portions of the waveforms. Consistency of overall waveform morphology among laboratory locations was assessed by computing the Pearson product–moment correlation coefficients for the voltage values ($N = 336$ points) of the grand average waveforms from each laboratory at a given electrode between all possible pairs of laboratory locations (e.g., CA vs. CT, CA vs. IA, etc.) for a total of 15 correlations from each electrode. It should be noted expressly that the correlations were computed using the waveform voltage values from the grand averages for each of the six laboratories. Although not a typical application, this procedure provides an appropriate and direct method of quantifying waveform agreement among laboratories.

The size of the correlations ranged from 0.20 to 0.80 and, given the large number of points, were all highly significant ($P < 0.001$). The mean of the 15 correlation coefficients obtained from all of the laboratory location comparisons then was computed for each electrode to obtain a general index of the amount of agreement among laboratories at a specific scalp site. The mean and standard deviation (S.D.) for the correlation values from each electrode site for the target and standard stimuli are illustrated in Fig. 5. The overall correlation across laboratory locations was somewhat variable among electrode positions and decreased appreciably from the anterior to posterior electrode sites (top) for ERPs from both the target (mean correlation = 0.38) and standard (mean correlation = 0.57) stimuli. However, as indicated by the distribution of the standard deviation values for the mean correlations (bottom), the inter-laboratory variability of the correlations was relatively small across electrode sites for both the target (mean S.D. = 0.23) and standard (mean S.D. = 0.28) stimulus waveforms. Thus, the inter-laboratory agreement for overall waveform morphology was moderate to good and consistent in its scalp variation among laboratory locations.

To determine whether some of this variability might be caused by outlier values from a specific laboratory location, each component's amplitude and latency from each electrode site were plotted

Fig. 7. Mean correlation values at each electrode site from the target and standard stimuli for the inter-laboratory comparisons of the N1, P2, N2, and P3 components. These values were obtained by correlating the component values from the average waveforms on a point-by-point basis across laboratories thereby producing the overall mean and standard deviation correlational values for each electrode (see text for details).


TARGETS
STANDARDS
N1
P2
N2
P3
1.0
.40

against each other in the form of a scattergram (cf. Refs. 20 and 27), with each laboratory's data portrayed by a different symbol. To illustrate this procedure, the P3 data from the target stimuli at the Pz electrode site are presented in Fig. 6. When the N1, P2, N2, and P3 components from all of the electrode sites for both the target and standard stimuli were assessed in the same manner, the six different laboratories produced component values that were similarly variable across subjects. Given this outcome, the inter-subject component variation contributing to the patterns of inter-laboratory waveform correlations was consistently variable across laboratories.

Component consistency. To assess the degree of inter-laboratory agreement for the N1, P2, N2, and P3 components, the correlational procedures used to quantify the overall waveform morphological consistency were applied to just the voltage values that defined each component. These voltage values were based on the latency windows used to identify each component for the target and standard stimuli (N1: 75–125; P2: 100–200; N2: 140–300; P3: 275–400 ms). Hence, the degree of consistency among laboratory locations for each specific component was computed by calculating the correlations using the voltage values within each component latency window.

The mean correlations from each electrode for each component are presented in Fig. 7. Note that the scale for the Pearson r values ranges from 0.40 to 1.0 and is appreciably higher than the overall waveform correlations, since the degree of inter-laboratory consistency for specific components was generally quite strong across electrodes (N1 mean correlation: target = 0.84, standard = 0.82; P2 mean correlation: target = 0:79, standard = 0.89; N2 mean correlation: target = 0.94, standard = 0.97; P3 mean correlation: target = 0.96, standard = 0.83). Further, the P3 component from the target stimuli demonstrated almost perfect inter-laboratory consistency across all electrode locations. Because P3 measurements are the primary dependent variable in many basic and applied ERP studies, the high degree of consistency among the ERPs obtained from different laboratories is encouraging.

4. Discussion

The present study assessed the degree of ERP consistency with an auditory oddball paradigm from six different laboratories using identical equipment, software, and methodology. Virtually no statistically reliable differences among laboratories for the amplitude or latency of the individual ERP components were obtained. Moreover, moderate to good consistency among laboratories for overall waveform morphology was found, with excellent consistency obtained for the P3 component. Although there was some variability in scalp distribution patterns for the inter-laboratory waveform correlations, these effects were reasonably uniform, lacking any conspicuous amplitude and latency outlier values. However, despite the high degree of consistency observed among laboratory locations, considerable variation in ERP component values was obtained (see Fig. 6), such that the lack of significant differences among the laboratory locations may have occurred because the relatively large sample sizes employed (n = 15 in each group) were still not sufficient to detect a Type II error.

The present findings indicate that individual variation in ERP measures can restrict somewhat the morphological agreement for the ERP waveforms from different samples, even though good to high intra-subject temporal reliability of ERP component measures from auditory stimuli have been reported [15–18,20,21]. This outcome may result from the influence of intra- and extra-subject factors that conspire to affect waveform morphology variation (cf. Refs. 22–29).

In contrast to the inter-laboratory variability observed for overall waveform morphology, there was generally good agreement observed among laboratories across electrode sites (see Fig. 7). The inter-laboratory consistency was especially strong for the P3 component from the target stimuli. Because the P3 has been found to be sensitive for the assessment of psychiatric and neurologic disorders (e.g., cf. Refs. 5–8), the high degree of consistency observed in the present study supports the utility of multi-laboratory data acquisition projects and implies that ERPs are a generalizable neuroelectric measurement tool.

Thus, it is not unreasonable to conclude that multi-laboratory ERP recordings are feasible and may prove highly useful in the assessment of mental processes related to normal and clinical studies of cognitive function.

Acknowledgements

This national collaborative study is supported by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) by U.S.P.H.S. grants NIAAA U10AA08401, U10AA08402, U10AA08403. The first author, J.E.A., was supported by NIAAA Training Grant AA07456–10. This paper is publication NP7958 from The Scripps Research Institute. We thank George Collins for his superlative help.

The Consortium on the Genetics of Alcoholism (H. Begleiter, SUNY HSCB, Principal Investigator, T. Reich, Washington University, Co-Principal Investigator) includes six different centers where data collection takes place. The six sites and Principal Investigator and Co-Investigators are: Indiana University (J. Nurnberger, Jr., P.M. Conneally); University of Iowa (R. Crow, S. Kuperman); University of California at San Diego and The Scripps Research Institute (M. Schuckit, F.E. Bloom); University of Connecticut (V. Hesselbrock); State University of New York, Health Sciences Center at Brooklyn (H. Begleiter, B. Porjesz); Washington University in St. Louis (T. Reich, C.R. Cloninger).

References

[1] Picton, T.W. (1992) The P300 wave of the human event-related potential. *J. Clin. Neurophysiol.*, 9: 456–479.

[2] Polich, J. (1993) P300 in clinical applications: Meaning, method, and measurement. In E. Niedermeyer and F. Lopes da Silva (Eds.), *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, 3rd edn.*, William and William, Baltimore, MD, pp. 1005–1018.

[3] Donchin, E., Karis, D., Bashore, T.R., Coles, M.G.H. and Gratton, G. (1986) Cognitive psychophysiology and human information processing. In M.G.H. Coles, E. Donchin and S.W. Porges (Eds.), *Psychophysiology: Systems, Processes, and Applications*, Guilford Press, New York, pp. 244–267.

[4] Pritchard, W.S. (1981) Psychophysiology of the P300. *Psychol. Bull.*, 89: 506–540.

[5] Pfefferbaum, A., Ford, J.M., White, P. and Roth, W.T. (1989) P3 in schizophrenia is affected by stimulus modality, response requirements, medication status, and negative symptoms. *Electroencephalogr. Clin. Neurophysiol.*, 46: 1035–1044.

[6] Polich, J. (1991) P300 in the evaluation of aging and dementia. In C.H.M. Brunia, G. Mulder and M.N. Verbaten (Eds.), *Event-related Brain Research (EEG Suppl. 42)*, Elsevier, Amsterdam, pp. 304–323.

[7] Polich, J., Ladish, C. and Bloom, F.E. (1990) P300 assessment of early Alzheimer's disease. *Electroencephalogr. Clin. Neurophysiol.*, 77: 179–189.

[8] Pritchard, W.S. (1986) Cognitive event-related potential correlates of schizophrenia. *Psychol. Bull.*, 100: 43–66.

[9] Begleiter, H., Porjesz, B., Bihari, B. and Kissin, B. (1984) Event-related brain potentials in boys at risk for alcoholism. *Science*, 225: 1493–95.

[10] Hill, S.Y. and Steinhauer, S.R. (1993) Assessment of prepubertal and postpubertal boys and girls at risk for developing alcoholism with P300 from a visual discrimination task. *J. Stud. Alcohol.*, 54, 350–358.

[11] O'Conner, S., Hesselbrock, V. and Tasman, T. (1986) Correlates of increased risk for alcoholism in young men. *Prog. Neuropsychopharm. Biol. Psychiatry*, 10: 211–218.

[12] Polich, J. and Bloom, F.E. (1988) Event-related brain potentials in individuals at high and low risk for developing alcoholism: Failure to replicate. *Alcoholism Clin. Exp. Res.*, 12: 368–373.

[13] Porjesz, B., Begleiter, H., Bihari, B. and Kissin, B. (1987) Event-related brain potentials to high incentive stimuli in abstinent alcoholics. *Alcohol*, 4: 283–287.

[14] Polich, J., Pollock, V. and Bloom, F.E. (1994) Meta-analysis of P300 amplitude from males at risk for alcoholism. *Psychol. Bull.*, 115: 55–73.

[15] Fabiani, M., Gratton, G., Karis, D. and Donchin, E. (1987) 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. 2*, JAI Press, Greenwich, CT, pp. 1–78.

[16] Karniski, W. and Blair, R.C. (1989) Topographical and temporal stability of the P300. *Electroencephalogr. Clin. Neurophysiol.*, 72: 373–383.

[17] Kileny, P.R. and Kripal, J.P. (1987) Test-retest variability of auditory event-related potentials. *Ear Hearing*, 8: 110–114.

[18] Lewis, G.W. (1984) Temporal stability of multichannel, multimodal ERP recordings. *Int. J. Neurosci.*, 25: 131–144.

[19] Michalewski, H.J., Prasher, D.K. and Starr, A. (1986) Latency variability and temporal interrelationships of the auditory event-related potentials (N1, P2, N2, and P3) in normal subjects. *Electroencephalogr. Clin. Neurophysiol.*, 65: 59–71.

[20] Polich, J. (1986) Normal variation of P300 from auditory

stimuli. *Electroencephalogr. Clin. Neurophysiol.*, 65: 236–240.

[21] Zamrini, E.Y., Meador, K.J., Thompson, W.O. and Lee, G.P. (1991) Reproducibility of P3. *Int. J. Neurosci.*, 61: 113–120.

[22] DiTraglia, G.M. and Polich, J. (1990) P300 and introverted/extroverted personality types. *Psychophysiology*, 8: 177–184.

[23] Geisler, M.W. and Polich, J. (1990) P300 and time-of-day: Circadian rhythms, food intake, and body temperature. *Biol. Psychol.*, 31: 117–136.

[24] Geisler, M.W. and Polich, J. (1992) P300, food consumption, and memory performance. *Psychophysiology*, 29: 76–85.

[25] Geisler, M.W. and Polich, J. (1992) P300 and individual differences: Morning/evening activity preference, food, and time-of-day. *Psychophysiology*, 29: 86–94.

[26] Pfefferbaum, A. and Rosenbloom, M. (1989) Skull thickness influences P3 amplitude. *Psychopharmacol. Bull.*, 25 493–496.

[27] Polich, J. (1992) On the correlation between P300 amplitude and latency. *Bull. Psychon. Soc.*., 30: 5–8.

[28] Polich, J. and Geisler, M.W. (1991) P300 seasonal variation. *Biol. Psychol.*, 32: 173–179.

[29] Polich, J. and Martin, S. (1992) P300, cognitive capability and personality: A correlational study of university undergraduates. *Pers. Individ. Diff.*, 13: 533–543.

[30] Scharbrough, F., Ghatrian, G.-E., Lesser, R.P., Luders, H., Nuwer, M. and Picton, T.W. (1990) *Guidelines for Standard Electrode Position Nomenclature*. American Electroencephalography Society, Bloomfield, CT.

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

[32] Donchin, E. and Coles, M.G.H. (1988) Is the P300 component a manifestation of context updating? *Behav. Brain Sci.*, 11: 357–374.