Title

The auditory event-related potential is a stable and reliable measure in elderly subjects over a 3 year period

Permalink

https://escholarship.org/uc/item/0wk574wh

Authors

Sandman, Curt A
Patterson, Julie V

Peer reviewed

The auditory event-related potential is a stable and reliable measure in elderly subjects over a 3 year period

Curt A. Sandman[a,*], Julie V. Patterson[b]

[a]*Department of Psychiatry, University of California, Irvine, CA, USA*
[b]*Department of Neurology, University of California, Irvine, CA, USA*

Abstract

Objectives: Valid markers of psychobiological processes, including changes over the lifespan, must be reliable. This study investigated the reliability of the auditory event-related potential (ERP) over a 3 year period.

Methods: Predictable and unpredictable rare tones were embedded in common-to-rare sequences at 3 different ratios (2:3, 2:5 and 2:8). Forty-six older (mean age 72.3 years) volunteers pressed a key to the rare tones, and ERPs (Fz, Cz and Pz) and reaction time (RT) were measured. Reliability across years was assessed using 3 methods: (1) determination of the stability of waveform *components* (P1, N1, P2, N2 and P3); (2) cross-correlation of successive 15 ms epochs of *within-subject* ERPs; and (3) cross-correlation of 15 ms epochs of *between-subject* ERPs.

Results: With all analyses, the ERP was stable. Analysis of the scored components indicated that P3 was especially stable in the unpredictable rare (2:8) condition. Earlier components were equally stable across all conditions. Analysis of 15 ms ERP epochs indicated significant ERP stability 60 ms after stimulation, lasting over 640 ms.

Conclusions: Robust within-subject reliability of the ERP strengthens its potential use for detecting preclinical changes in at-risk elderly populations. © 2000 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Event-related potential; Reliability; Stability; Aging

1. Introduction

An extensive literature assumes that the late positive wave (P3) of the event-related potential (ERP) is a valid index of cognitive activity (Hillyard and Picton, 1987; Segalowitz and Barnes, 1993; Polich and Kok, 1995). The amplitude of P3 reflects brain processes related to decision making (Tueting et al., 1970; Duncan-Johnson and Donchin, 1977; Donchin, 1979, 1981; Johnson, 1986), depth of processing (Wickens et al., 1983; Rugg et al., 1998) and memory (Squires et al., 1976; Donchin and Coles, 1988; Patterson et al., 1991; Pelosi and Blumhardt, 1999). In addition, measures of P3 reflect brain maturation (Courchesne, 1978, 1983; Katsanis et al., 1996; Sangal et al., 1998) and are sensitive to aging (Ford et al., 1979, 1982a,b, 1997; Pfefferbaum et al., 1984a; Patterson et al., 1988; Pratt et al., 1989; Polich, 1996, 1997, 1998; Smulders et al., 1999), and dementia (Goodin et al., 1983; Pfeffer-

baum et al., 1984b; Gordon et al., 1986; Kraiuhin et al., 1986; Patterson et al., 1988; Filipovic and Kostic, 1995; Osawa, 1995; Ford et al., 1997). However, the validity of a test can never exceed the square root of the coefficient of reliability. Thus, sensitive and valid markers of psychobiological processes, including changes over the lifespan, must be reliable.

Auditory event-related potentials (ERPs) in the target detection or 'oddball' task in young, normal volunteers apparently are reliable within sessions (Sklare and Lynn, 1984; Polich, 1986; Fabiani et al., 1987; Kinoshita et al., 1995, 1996; Maeda et al., 1995) over days (Sklare and Lynn, 1984; Fabiani et al., 1987; Kinoshita et al., 1995, 1996; Maeda et al., 1995), months (Lewis, 1984; Karniski and Blair, 1989; Kinoshita et al., 1995, 1996; Maeda et al., 1995), and 1.8 years (Segalowitz and Barnes, 1993). Reliable P3 components of the ERP also have been reported among 36 schizophrenics after 3 months ($r = 0.67$–0.85; Hegerl et al., 1988) and, incidentally, in a group of about 25 elderly subjects across 4 tasks (0.44–0.73; Miller et al., 1987). Stable early ERP components to light flashes have been reported over 10 months (Schellberg et al., 1987) and 5

* Corresponding author. Box 5A, 2501 Harbor Boulevard, Costa Mesa, CA 92626, USA. Tel.: +1-714-957-5435; fax: +1-714-957-5354.
E-mail address: casandma@uci.edu (C.A. Sandman).

years (Rothenberger et al., 1987) in children. The apparent reliability of the ERP suggests that instances of instability may indicate that there are significant sources of error or reflect the presence of *systemic influences* (Fabiani et al., 1987; Kileny and Peters-Kripal, 1987; Karniski and Blair, 1989; Simons and Miles, 1990). For instance, in a multiyear study of ERP reliability (van der Wal and Sandman, 1992), sharp decreases in reliability between successive years predicted death for 7 otherwise healthy elderly subjects. Reliability of the ERP in these subjects was greater than 0.80 in consecutive years except for the year before death.

As reviewed by Simons and Miles (1990) and Segalowitz and Barnes (1993), insufficient attention has been focussed on issues of reliability and stability in ERP research. This is an important issue which has continuing relevance to the use of ERPs in clinical assessment and in the study of trait and state effects (Segalowitz and Barnes, 1993). Additionally, the research has focussed on relatively short time intervals, with the longest being around 14 months (e.g. Sinha et al., 1992), and with few exceptions (Pollack and Schneider, 1992) on younger age groups. At least two general procedures have been applied to test ERP reliability. The most common method compared the scored peaks (especially P3) of the ERP at two different times. This between-subject procedure produces a measure of the tendency for individuals to maintain their rank within in a group across time (Simons and Miles, 1990). The second method cross-correlated individual ERPs recorded at different times to produce a measure of intra-subject reliability (Lewis, 1984; Fabiani et al., 1987; van der Wal and Sandman, 1992). The cross-correlation procedure has an advantage of being independent of inevitable sources of error related to detection of ERP components (Lewis, 1984). In the current study, both of these procedures, in addition to a within-subjects cross-correlation procedure, were used to compute measures of stability and reliability over a 3 year period in an elderly population.

2. Materials and methods

2.1. Subjects

Forty-six subjects (29 women) were selected from 159 elderly individuals (entry age range 60–86 years, mean age 72.3 years) participating in a study on effective aging. All subjects accepted into the program were healthy and living independently. Subjects were administered extensive medical examinations prior to entry into the research program and on a yearly basis thereafter. Each subject was evaluated with the ERP at yearly intervals for 3 consecutive years.

2.2. Procedure

As described by van der Wal and Sandman (1992), subjects were tested in an electrically-shielded, sound attenuating chamber and reclined in a comfortable chair as electrodes were applied to the scalp. Binaural headphones were placed over the ears and white noise (72 db) from floor speakers masked extraneous noise. Subjects were monitored continuously with video and audio equipment. In a dual rare-event procedure (Sandman et al., 1990), target tones were presented at predictable (fixed, 450 Hz) infrequent intervals, at identical infrequent, but unpredictable (random, 600 Hz) intervals, and at frequent (common, 550 Hz) intervals.

Three versions of the dual rare-event procedure were presented to all subjects in balanced order. The 3 versions varied the probability of the infrequent tone. Examples of the 3 versions are:

2:3 FCCFRCFRRFCRF…
2:5 FCRCRFCCCCFCCRCFCCCCFRCR…
2:8 FCRCCCCCFCCCCCCCFRCRCCCFCRCRCCF…

In these examples, F is the predictable (fixed), R is the unpredictable (random) and C is the frequent (common) tone. The fixed target always was the third (2:3), fifth (2:5) or eighth (2:8) tone. The random target occurred at the same probability as the fixed target, but its position in the sequence was not predictable.

The pure tones were 18 db contrast with background (90 db SPL with white noise background of 72 db SPL measured at the headphone cone: Bruel and Kjaer, Model 2203 sound level meter) and were 100 ms duration with a rise time of 50 ms. Pre-testing at yearly intervals determined that all subjects included in the analysis were able to discriminate the tones. Subjects were given practice to depress a hand-held key each time they heard the random target (matched to the dominant hand) or the fixed target (matched to the non-dominant hand). They were instructed not to respond to common tones and to keep their eyes closed.

2.2.1. EEG procedures

Recordings were made with a Grass polygraph, Model 79, with amplifier settings at 0.30 and 100 Hz. Data were digitized and stored for off-line analysis on a MINC 11/23 computer (digital Equipment Corp). Tones were presented by a microcomputer interfaced with a PDP system. Gold cup electrodes filled with EC-2 creme (Grass) were placed according to the International 10–20 System at Fz, Cz and Pz referenced to linked mastoids. Electrode impedances were matched within 1 kΩ and all were below 10 kΩ.

2.2.2. Event-related potential analysis

The EEG was sampled for 1280 ms at 200 Hz. Baseline was determined by a 280 ms pre-stimulus average. Waveforms for 44 sequences of random, fixed and common tones were collected and averaged. An analog filter conditioned the signal prior to digitization (12 db/octave, 3 db at 100 Hz), minimizing aliasing and phase errors.

The latencies and peak-to-peak amplitudes of the major components were located with an automated system that

placed cursors on a minimum and maximum voltage for waveform troughs and peaks, respectively. The ERP waveforms were evaluated by trained technicians to identify the prominent peaks within specific latency windows (P1, 30–80 ms; N1, 70–200 ms; P2, 130–280 ms; N2, 180–360 ms; P3, 250–600 ms). Spectral interpolation was applied for measurement of the latency of waveform peaks. The waveform was approximated as a sum of sinusoids (the Fourier coefficients) resulting in unlimited temporal resolution and accuracy for frequencies up to 25 Hz (the Nyquist frequency).

2.2.3. Artifact rejection

Trials contaminated by eye movement were automatically rejected by computer software. The efficacy of the rejection system was determined by separately testing subjects asked to make lateral eye movements (eyes closed) to a series of 10 tones in each condition (van der Wal and Sandman, 1992). Eye movement was verified by electrodes attached to the outer canthus and suborbit referenced to linked mastoids. The software system correctly identified 98.4% of the lateral eye movements in the EEG. Eye blinks were detected on 100% of the trials but blinks were infrequent in the closed eyes procedure.

Trial sequences were repeated automatically if reaction times to fixed or random target rare tones exceeded 1500 ms, if subjects pressed a key to the common tone, or because of artifact in the EEG. An artifact was defined as a response contaminated with eye blinks or muscle movement or a response exceeding ± 50 μV. No subject had more than 5 trial sequences repeated and fewer than 2% (less than one trial per subject) of the trials across all subjects were repeated.

2.2.4. Statistical analysis

Reliability over the 3 testing sessions was assessed using 3 methods: (1) determination of the stability of waveform *components* (P1, N1, P2, N2 and P3) identified as the maximum amplitude within a specified time window; (2) cross-correlation of successive 15 ms epochs of *within-subject* ERPs across years; and (3) cross-correlation of 15 ms epochs of *between-subject* ERPs across years.

The stability of waveform components was determined by calculating, for each measured peak of the ERP, Pearson Product Moment correlations between the pairwise comparisons of the 3 testing sessions (year 1 with year 2, year 2 with year 3, and year 1 with year 3) for both latency and amplitude measures.

The *within-subjects* correlation analysis measured the stability of the shape of the intra-individual wave over years. Pearson Product Moment correlations of intra-subject waveforms over years were compared separately for every placement, condition and phase. The correlation coefficient was computed by comparing corresponding points (15 ms epochs) on the waveform for each subject for each year. Correlations were generated for each subject for 3 place-

ments (Fz, Cz and Pz) $\times$ 3 phases (2:3, 2:5 and 2:8) $\times$ 3 conditions (random, fixed and common), and for each comparison (year 1 with year 2, year 2 with year 3, and year 1 with year 3), resulting in 81 correlations per subject per 15 ms epoch. These data were summarized in two ways. First, the total number of significant correlations for each condition was tested against chance. Second, the correlations were converted to z scores, summed and averaged, and then reconverted to r values representing averaged correlations (see Table 2).

The *between-subjects* analysis measured the reliability of the test to produce changes over years. In this analysis, Pearson Product Moment correlations of ERP waveforms across subjects over years were compared separately for every placement and condition. This correlation coefficient was computed by comparing 15 ms epochs across subjects for each year (year 1 with year 2, year 2 with year 3, and year 1 with year 3). Correlations were generated for 3 placements, 3 conditions, and for each year comparison. These correlations were computed for the 2:8 phase only, resulting in 27 total correlations per 15 ms epoch.

3. Results

Fig. 1 shows averaged ERPs for the fixed, random and common conditions at Fz, Cz and Pz grand averaged over subjects for the 2:8 phase. As illustrated in Fig. 1, the random target elicited the largest peak-to-peak amplitude of the P3 wave. The peak-to-peak P3 component was attenuated in the fixed condition and absent in the common condition.

3.1. Between-subject reliability of scored ERP components

3.1.1. Intra- and inter-rater reliability

In order to determine scorer reliability, two trained raters blindly scored the latency and amplitude of P3 waveforms in the random 2:8 phase. Each rater scored the same 46 records on two different occasions without knowledge of previous results, yielding both inter- and intra-rater coefficients. Both intra-rater (0.67–0.86, $P < 0.0001$) and inter-rater (0.67–0.77, $P < 0.001$) reliability estimates of P3 amplitude were highly significant. Rating of P3 latency was less consistent ($r = 0.37$–0.69, $P < 0.01$).

3.1.2. Reliability of P3 amplitude

Fig. 2 presents reliability coefficients calculated for 1 and 2 year intervals for the amplitude of the ERP recorded from Fz, Cz and Pz during the fixed and random conditions. Highly reliable P3 amplitudes were apparent especially in the random condition for all 3 rare stimulus ratios for the 1 and 2 year intervals. For each interval, the reliability of P3 in the random condition increased as the probability of a rare target decreased (i.e. changed from 2:3 to 2:8). Inspection of the correlations revealed the opposite pattern in the fixed condition because the least frequent
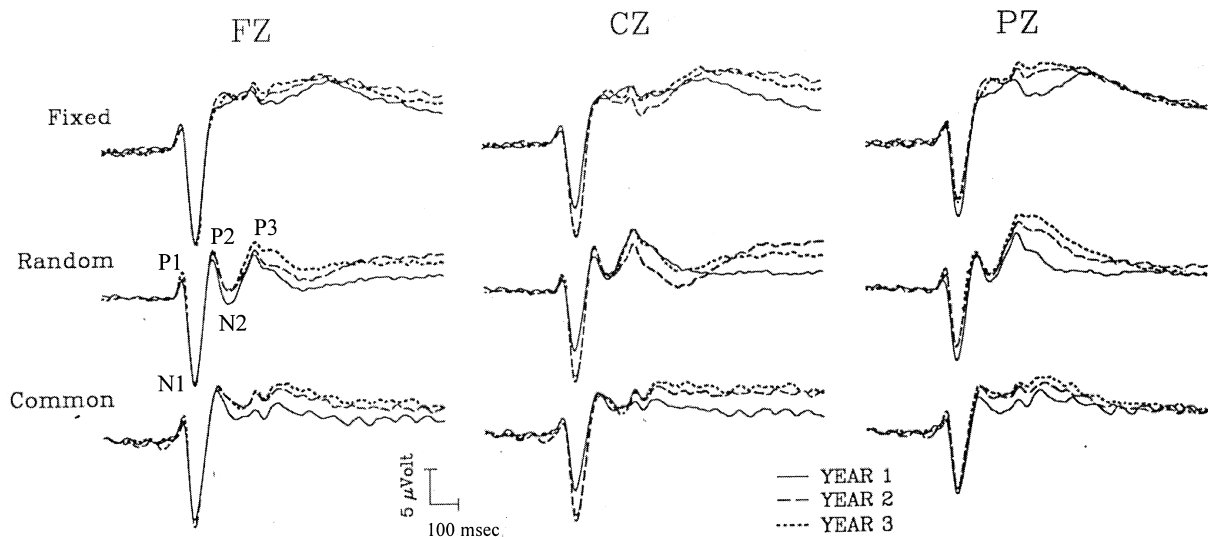
Fig. 1. Group averaged event-related potentials averaged over Fz, Cz and Pz in the 2:8 stimulus phase for the fixed, random and common conditions for 3 consecutive years. The P1, N1, P2, N2 and P3 components are indicated for the random condition at Fz.

targets resulted in the lowest correlations (poorest reliability).

3.1.3. Reliability of P3 latency

Fig. 2 also shows the reliability coefficients calculated for 1 and 2 year intervals for the latency of the ERP recorded from Fz, Cz and Pz during the fixed and random conditions. The latency of P3 was not reliable over 1 or 2 year intervals in any condition, stimulus probability or electrode location. Although several reliability coefficients for other components reached statistical significance for latency, the variance shared between test intervals for P3 latency rarely exceeded 25% and was universally much lower than for amplitude.

3.1.4. Reliability of early compared to late ERP components

Although the focus of the study was on the reliability of the P3 wave, the reliability of amplitude for the other major ERP components (P1, N1, P2 and N2) also was examined and is presented in Fig. 2. The reliability of N1 and P2 amplitudes was significant for both testing intervals, all placements, conditions and stimulus probabilities. The amplitude of N2 was less consistent, achieving the highest reliability in the 2:5 condition at Cz.

The differential reliability of N1 and P3 to the fixed and random rare targets can be seen in Fig. 3. The reliability of N1 is virtually identical for both targets, regardless of stimulus probability. Reliability of P3 is dependent on stimulus probability (it increases as stimulus probability decreases) and predictability (reliability is very high for unpredictable targets but low for predictable targets). The differential effect of the fixed and random target conditions on the N1 and P3 components was tested in an ANOVA (component × target type).

The component × target type interaction was significant for each probability (2:3, $F_{1,8} = 21.40$, $P < 0.002$; 2:5, $F_{1,8} = 15.42$, $P < 0.004$; 2:8, $F_{1,8} = 80.44$, $P < 0.00001$).

3.1.5. Stability of reaction time (RT)

The reliability of RT over both 1 and 2 year intervals was statistically significant (Table 1). However, the correlation coefficients were lower than for P3, especially comparisons with the first year. For all 3 phases, in both target conditions, reliability between years 2 and 3 was highest. Unlike measures of P3, reliability of RT was comparable in both the fixed and random target conditions.

3.2. Between-subject reliability of ERP epochs across years

3.2.1. Years 1 and 2

Correlations of 15 ms epochs on the ERP waveform across subjects are plotted in Fig. 4 for each electrode site and for the fixed, random and common conditions for the 2:8 phase. Over Fz, the correlations for the random target were significant after 96 ms and reached highly significant levels ($r > 0.80$) between 430 and 450 ms. For the fixed target, significant reliability was reached by 60 ms and maximal reliability was evident around the N1 (140–190 ms) component. The common tone also was highly reliable but the values were slightly lower. The pattern was similar for Pz waveforms. However, significant values for the random target were not achieved until 128 ms and the correlations were lower than for Fz. The correlations reached maximal levels around the N1 and P3 complexes, dropped, and then increased ($r > 0.50$) after 700 ms. The pattern at Pz for the fixed tone was identical, but slightly lower than at Fz. Reliability of the common tone was somewhat less over
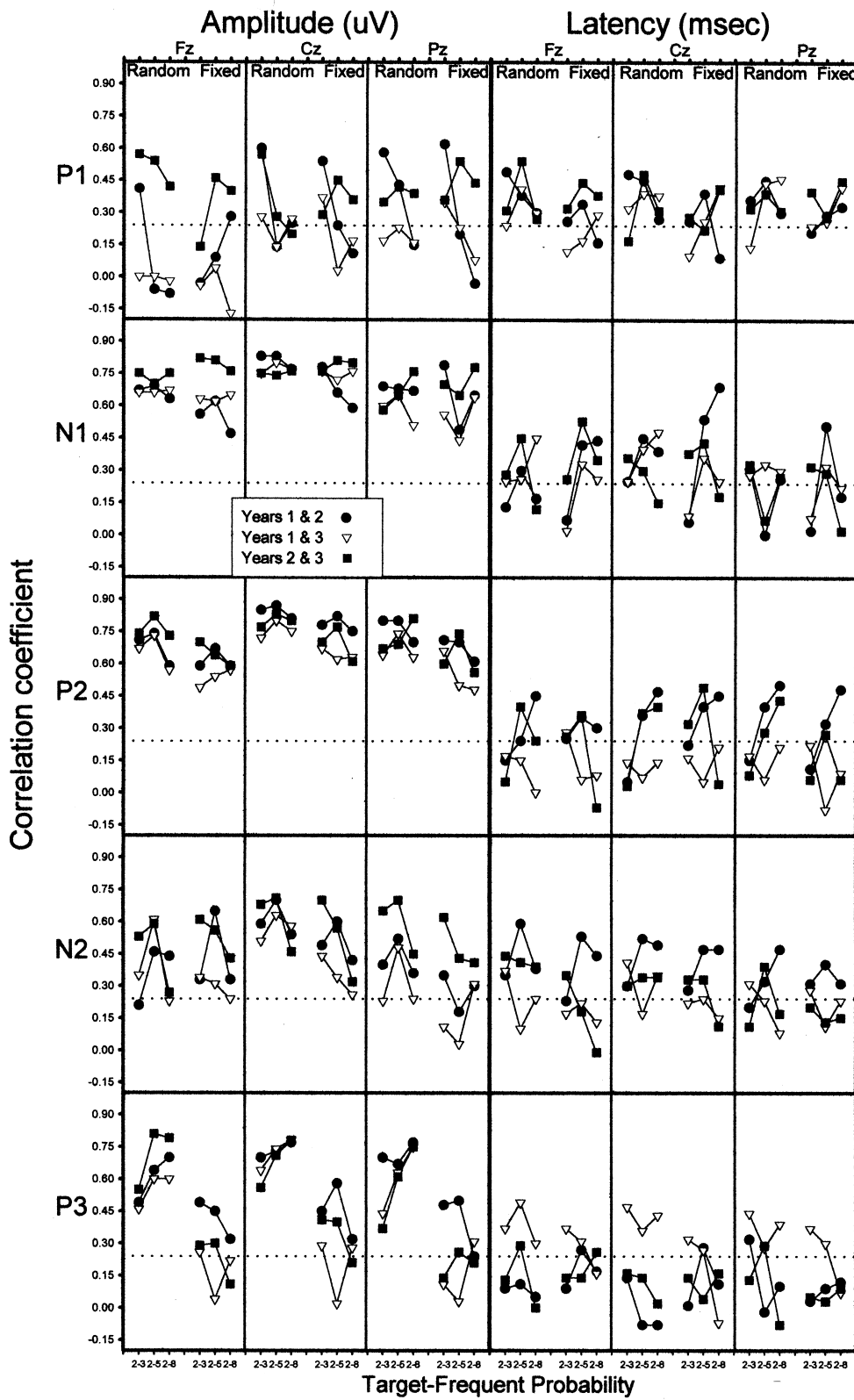
Fig. 2. Average correlations between each year for the fixed and random conditions, for each phase (2:3, 2:5 and 2:8), and for each electrode location (Fz, Cz and Pz) for P1, N1, P2, N2 and P3 amplitude (left column) and latency (right column). The dotted horizontal line indicates the $P = 0.05$ significance level.
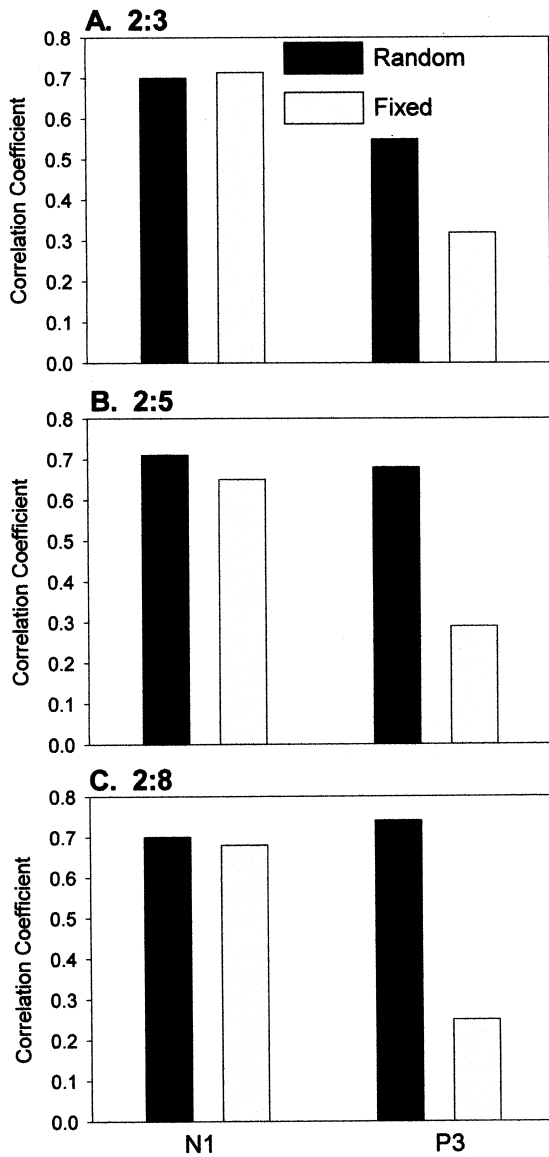
Fig. 3. Comparison of average correlation coefficients for the N1 and P3 components in the random and fixed conditions for the 2:3 phase (A), 2:5 phase (B), and 2:8 phase (C) pooled over both testing intervals.

Pz, especially around the N1-P2 transition. Similar patterns characterized the Cz placement.

3.2.2. Years 2 and 3

Virtually the same pattern of stability existed between years 2 and 3 as with years 1 and 2. The earliest evidence of stability was to the fixed target (60 ms) compared to stability by 90 ms for the random target and the common tone. For all 3 placements, but especially for Pz and Fz, early (90–160 ms) latency waves in the fixed condition were highly reliable. Response to the random target also was highly significant early (130–160 ms), middle (200–270 ms) and late (384–512 ms) at all placements. Although responses to the common tone at all placements was signif-

icant, the magnitude of correlation was lower than the target tones (Fig. 4).

3.2.3. Years 1 and 3

Although the majority of correlations were highly significant ($P < 0.01$), the pattern was different and the degree of association was less than correlations between ERP responses in contiguous years. Maximal stability for the response to the fixed target was apparent over the N1 complex, at all electrode locations, as with previous analyses. Responses to the random target remained stable, especially beyond 384 ms. Conversely, the reliability of the common tone dropped below $P < 0.01$ after 384 ms at all 3 locations (Fig. 4).

3.3. Within-subject reliability of ERP epochs across years

For the analysis of the 2:8 phase, 1145 correlations of a total 1242 possible (91%) were significant at the $P < 0.01$ level. Binomial distribution indicated that this was a highly significant outcome. Responses to all tones were equally reliable (random, 378 of 414 (91%) significant at $P < 0.01$; fixed, 385 of 414 (93%); and common, 382 of 414 (92%)). Inspection of reliability between year 1 and years 2 and 3 showed that correlations were lower than estimates between years 2 and 3. In order to determine if the age of the subject was related to the magnitude of the reliability coefficients, pairwise correlations between age and the relationship between years 2 and 3 were computed for the 2:8 phase and the random condition for Pz. The correlation was not significant ($r = 0.04$).

3.3.1. Years 1 and 2

For the 2:8 phase across electrode placements, 373 of 414 (91%) of the correlations for waveform shape were significant at $P < 0.01$. A high proportion of the correlations over years were significant for the random (121 of 138 (88%)), fixed (128 of 138 (93%)) and common (124 of 138 (90%)) tones. Lower reliability was associated with posterior place-

Table 1
Correlations of reaction time between years[a]

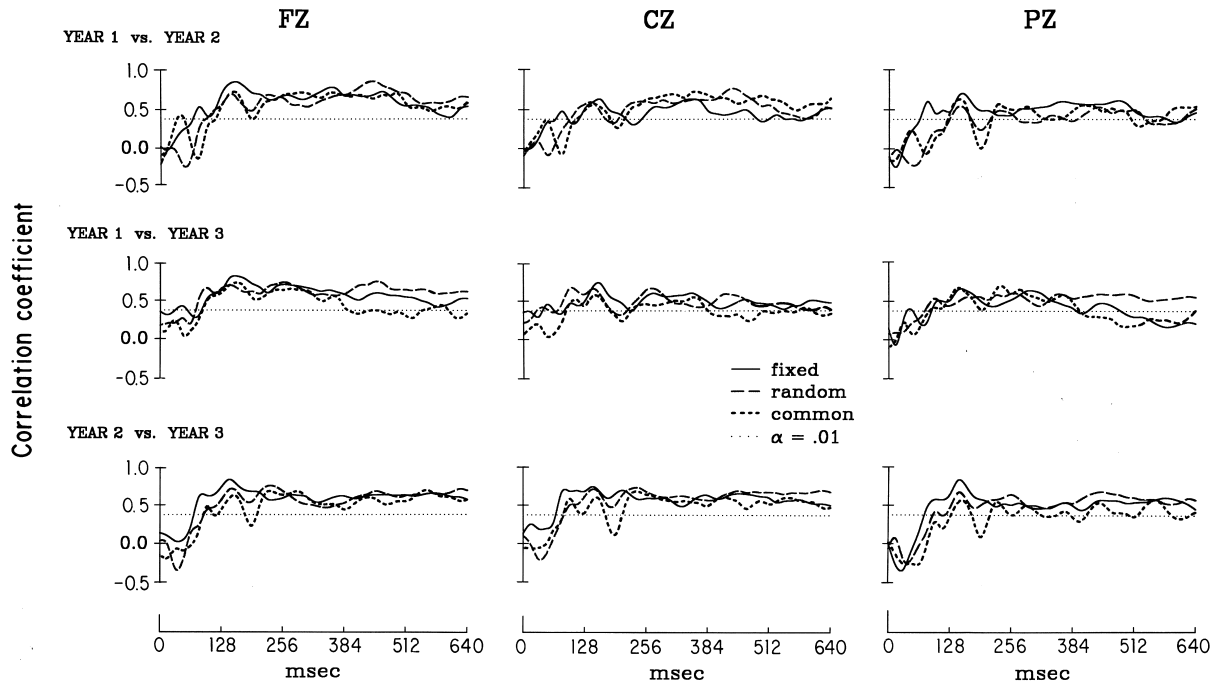|	Fixed condition		Random condition	
	Year 2	Year 3	Year 2	Year 3
Phase 2:3				
Year 1	0.50	0.41	0.44	0.52
Year 2	–	0.57	–	0.72
Phase 2:5				
Year 1	0.44	0.50	0.49	0.71
Year 2	–	0.73	–	0.72
Phase 2:8				
Year 1	0.25	0.32	0.38	0.55
Year 2	–	0.63	–	0.70

[a] $P = 0.05$, $r = 0.24$.

Fig. 4. Average bi-yearly correlation coefficients for the fixed, random, and common conditions for each electrode placement in the 2:8 phase. Coefficients were computed between-subjects in 15 ms epochs. The dotted horizontal line indicates the $P = 0.01$ significance level.

ments for the fixed and random targets, but not the common tone. Inspection of the correlations for individual subjects showed that the source of ERP instability was a combination of 3 subjects with low reliability over all placements and 8 subjects with one unreliable electrode placement.

3.3.2. Years 2 and 3

In the last 2 years of the study, 402 of 414 (97%) correlations were significant at $P < 0.01$. Of the 12 non-significant correlations, 9 were shared by two subjects.

3.3.3. Years 1 and 3

Of the 414 possible correlations between years 1 and 3, 374 (90%) were significant at $P < 0.01$. Stability was comparable for the fixed (123 of 138 (89%)), random (125 of 138 (91%)) and common (126 of 138 (91%)) tones. As in the previous analysis, lower stability was associated with posterior placements.

3.3.4. Averaged intra-individual waveform correlations

Table 2 presents the averaged correlations between years 1 and 2, 2 and 3, and 1 and 3 for the 2:8 target probability condition. The reliability at all placements and conditions for all tones were highly significant. Inspection of the correlations showed that reliability coefficients were highest between contiguous years, especially years 2 and 3. However, the overall stability of the waveforms was remarkably high, as was apparent from the shape of the averaged waveform over years (Figs. 1 and 4). In Fig. 5, individual ERPs over Pz for all 3 years are illustrated for one

subject with highly reliable ($r > 0.95$) waveforms (Fig. 5A), one subject with moderately reliable ($r = 0.68$) waveforms (Fig. 5B), and one subject with unreliable ($r > 0.04$) waveforms (Fig. 5C).

4. Discussion

The purpose of this study was to explore the stability of the ERP, especially the P3 component, in an elderly group over a 3 year period. To our knowledge this is the first

Table 2
Averaged correlations of ERP waveforms between each year for the 2:8 condition ($N = 46$)[a]

	Years 1 and 2	Years 1 and 3	Years 2 and 3
Fixed			
Fz	0.86	0.88	0.88
Cz	0.80	0.83	0.85
Pz	0.81	0.80	0.91
Random			
Fz	0.84	0.83	0.86
Cz	0.78	0.77	0.83
Pz	0.78	0.78	0.89
Common			
Fz	0.86	0.83	0.88
Cz	0.81	0.77	0.85
Pz	0.77	0.75	0.85

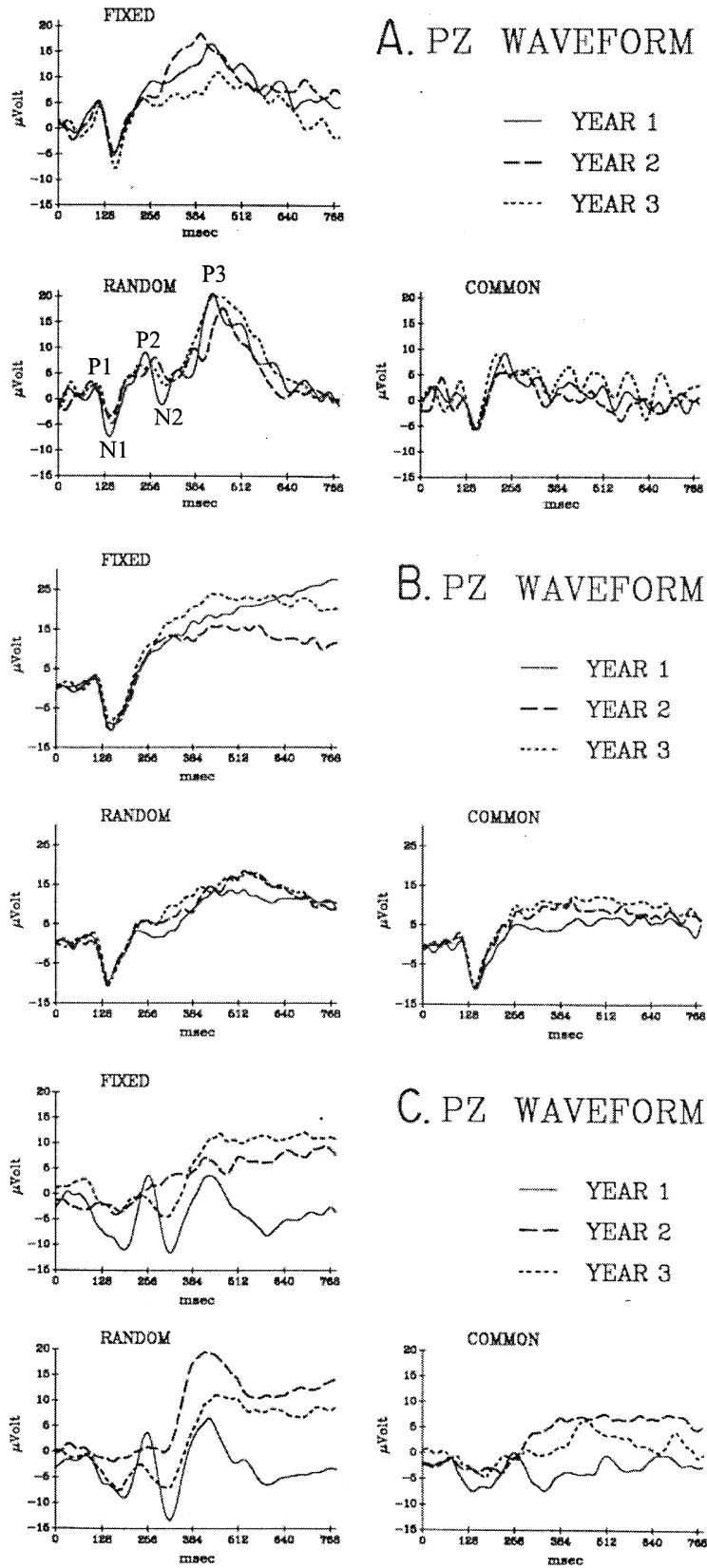[a] Correlations were converted to z scores; $P < 0.01 = 0.37$.

Fig. 5. Representative waveforms for (A) a subject with high correspondence ($r > 0.90$) among years, (B) a subject with moderate ($r = 0.70$–0.80) correspondence among years, and (C) a subject with low correspondence ($r < 0.20$) among years. The P1, N1, P2, N2 and P3 components are indicated for the random condition for (A).

formal, *multi-year* evaluation of reliability of the ERP in an elderly population. This is surprising because the ERP, and especially P3, is often proposed as a valid index of information processing and brain maturity (Pfefferbaum et al., 1984a; Rosler et al., 1986; Hillyard and Picton, 1987; Donchin and Coles, 1988). The stability of the ERP was determined over a 3 year period with 3 different methods: (1) *between-subject* reliability of ERP components; (2) *between-subject* reliability of discrete 15 ms epochs of the waveform; and (3) *within-subject* reliability of discrete 15 ms epochs of the waveform. For each of the 3 methods, the ERP waveform, including P3, was highly reliable over years, especially for the amplitude measures. These results provide support for the use of P3 as a valid measure of cognitive activity in the elderly.

4.1. Component reliability

The reliability of ERP components is initially dependent upon their reliable identification. Kramer (1985) concluded that trained technicians were capable of reliably recovering the structure of simulated ERPs. In the current study, both inter- and intra-rater reliability estimates showed that two trained scorers showed significant agreement in their scoring of P3 amplitude, and that each rater was consistent across two scoring sessions. Our finding that ratings of P3 latency were less consistent than amplitude agrees with previous findings (Sklare and Lynn, 1984; Polich, 1986; Fabiani et al., 1987; Segalowitz and Barnes, 1993; Kinoshita et al., 1995), and probably reflects the variability of P3 over the different experimental conditions, and the long duration of this component. The lower amplitude P3 elicited by the fixed condition and as a function of the probability of the target, for example, would make the selection of the appropriate point to determine its latency more difficult.

Although the reliability of each of the major components of ERP (P1, N1, P2, N2 and P3) was evaluated in this study, the procedures were designed to elicit P3. All 3 rare-event phases resulted in stable P3 components over 3 years, but the 2:8 phase produced the largest and most reliable P3. The amplitude of the late P3 component in this phase was as stable over years as the earlier components, N1 and P2. Typically, the late components are more variable than early ones because they are controlled more by endogenous factors, such as evaluation of the environment and decision making, than by exogenous factors, such as the physical features of external stimulation (Donchin, 1979, 1981; Hillyard and Picton, 1987; Donchin and Coles, 1988). The stability of the P3 response observed in this study argues for its validity as a neurological marker of individual differences in cognition. Our finding that P3 was more reliable for rare events argues for the use of this methodology when P3 is used to study trait and state effects.

Similarly, relative to the random condition, the stability of P3 declined in the fixed condition for all phases, probably

reflecting the smaller and less detectable P3 responses. Larger amplitude P3 responses have previously been observed to be unpredictable as compared to predictable stimuli (Hillyard and Picton, 1987; Sandman et al., 1990), and the greater reliability of the ERP elicited by rare or target tones compared with predictable or frequent tones may result from the more favorable signal-to-noise ratio of the larger amplitude components (Fabiani et al., 1987). The fixed condition is certainly not inherently unreliable, however, because the exogenous early components remained very stable (e.g. Fig. 2). Another possibility is that subjects profited from repeated exposures to the predictable fixed condition resulting in diminished P3 amplitudes across years. This possibility is supported by the results for RT because, despite increasing age through the course of the study, RT decreased over years. Measures of RT also were stable over years, but not as stable as ERP amplitudes. The stability of the ERP amplitude measures over 3 years provides support for their utility in studies of individual differences in attention and cognition.

4.2. Waveform stability

The advantage of waveform analysis was that it removed decision errors related to interpreting and scoring of components. In the first analysis, the stability of the ERP between subjects was examined every 15 ms. The results of this analysis indicated the extent that an amplitude occurring at a discrete 15 ms interval in 1 year predicted amplitude at the same interval in a subsequent year. Significant correlations determined that within the 3 year period, subjects retained their relative position at each 15 ms interval. Remarkable consistency was detected for conditions, phases and placements as early as 60 ms after stimulation and was sustained for 700–800 ms.

In the second analysis of waveform stability, within-subject correlations of waveforms were computed. A significant correlation in this analysis indicated that subjects retained the *shape* of their waveforms over years. This analysis was sensitive to variation over years in individual subjects, and as such may be most useful as a marker of change. The results indicated that most subjects had very significant correspondence between waveforms over years. Average correlations were greater than 0.80, supporting the shape of the ERP as a very reliable measure of individual differences. Because non-linear change over time, or low reliability over repeated testing, has been proposed as an index of cognitive decline, dementia, and even death in elderly subjects (Rosen et al., 1986; Storandt et al., 1986; Wilson and Kaszniak, 1986; van der Wal and Sandman, 1992), longitudinal assessment of waveform stability also may be a sensitive indicator of early stages of insidious processes.

The ability of the ERP to provide information about brain function independently of a motor response makes it especially useful in the assessment of dementing processes,

since some individuals are not able to provide a behavioral response such as reaction time. The findings of this study that P3 is a reliable index of information processing in the target detection task support its use as a valid measure of brain function in aging and age-related disease processes.

Acknowledgements

References

Courchesne E. Changes in P3 waves with event repetition: long-term effects on scalp distribution and amplitude. Electroenceph clin Neurophysiol 1978;45:754–766.

Courchesne E. Cognitive components of the event-related brain potential: changes associated with development. In: Gaillard AW, Ritter W, editors. Tutorials in event-related potential research: endogenous components, Amsterdam: Elsevier, 1983. pp. 329–344.

Donchin E. Event-related brain potentials: a tool in the study of human information processing. In: Begleiter H, editor. Evoked brain potentials and behavior, New York: Plenum Press, 1979. pp. 13–88.

Donchin E. Surprise!...Surprise? Psychophysiology 1981;18:493–513.

Donchin E, Coles MG. Precommentary: is the P300 component a manifestation of context updating? Behav Brain Sci 1988;11:357–374.

Duncan-Johnson CC, Donchin E. On quantifying surprise: the variation of event-related potentials with subjective probability. Psychophysiology 1977;14:456–467.

Fabiani M, Gratton G, Karis D, Donchin E. Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. In: Ackles PK, Jennings JR, Coles MGH, editors. Advances in psychophysiology, 2. London: JAI Press, 1987. pp. 1–78.

Filipovic SR, Kostic VS. Utility of auditory P300 in detection of presenile dementia. J Neurol Sci 1995;131:150–155.

Ford JM, Hopkins WF, Roth WT, Pfefferbaum A, Kopell BS. Age effects on event-related potentials in a selective attention task. J Gerontol 1979;34:388–395.

Ford JM, Duncan-Johnson CC, Pfefferbaum A, Kopell BS. Expectancy for events in old age: stimulus sequence effects on P300 and reaction time. J Gerontol 1982a;37:696–704.

Ford JM, Pfefferbaum A, Tinklenberg JR, Kopell BS. Effects of perceptual and cognitive difficulty on P3 and RT in young and old adults. Electroenceph clin Neurophysiol 1982b;54:311–321.

Ford JM, Roth WT, Isaacks BG, Tinklenberg JR, Yesavage J, Pfefferbaum A. Automatic and effortful processing in aging and dementia: event-related brain potentials. Neurobiol Aging 1997;18:169–180.

Goodin DS, Starr A, Chippendale T, Squires KC. Sequential changes in the P3 component of the auditory evoked potential in confusional states and dementing illnesses. Neurology 1983;33:1215–1218.

Gordon E, Kraiuhin C, Stanfield P, Meares R, Howson A. The prediction of normal P3 latency and the diagnosis of dementia. Neuropsychologia 1986;24:823–830.

Hegerl U, Gaebel W, Gutzman H, Ulrich G. Auditory evoked potentials as possible predictors of outcome in schizophrenic outpatients. Int J Psychophysiol 1988;6:207–214.

Hillyard SA, Picton TW. Electrophysiology of cognition. In: Plum F, editor. Handbook of physiology – the nervous system, V. New York: Oxford University Press, 1987. pp. 513–570.

Johnson Jr R. A triarchic model of P300 amplitude. Psychophysiology 1986;23:367–384.

Karniski W, Blair RC. Topographical and temporal stability of the P300. Electroenceph clin Neurophysiol 1989;72:373–383.

Katsanis J, Iacono WG, McGue MK. The association between P300 and age from preadolescence to early adulthood. Int J Psychophysiol 1996;24:213–221.

Kileny PR, Peters-Kripal J. Test-retest variability of auditory event related potentials. Ear Hear 1987;8:110–114.

Kinoshita S, Maeda H, Nakamura J, Kodama E, Morita K. Reliability of the probability effect on event-related potentials during repeated testing. Kurume Med J 1995;42:199–210.

Kinoshita S, Inoue M, Maeda H, Nakamura J, Morita K. Long-term patterns of change in ERPs across repeated measurements. Physiol Behav 1996;60:1087–1092.

Kraiuhin C, Gordon E, Meares R, Howson A. Psychometrics and event-related potentials in the diagnosis of dementia. J Gerontol 1986;41:154–162.

Kramer AF. The interpretation of the component structure of event-related brain potentials: an analysis of expert judgments. Psychophysiology 1985;22:334–344.

Lewis GW. Temporal stability of multichannel, multimodal ERP recordings. Int J Neurosci 1984;25:131–144.

Maeda H, Morita K, Nakamura J, Masayuki I, Kinoshita S, Kodama E, Maki S, Nakazawa Y. Reliability of the task-related component (P3b) of P3 event-related potentials. Psychiatry Clin Neurosci 1995;49:281–286.

Miller GA, Bashore TR, Farwell LA, Donchin E. Research in geriatric psychophysiology. In: Schaie KW, editor. Annual review of gerontology and geriatrics, New York: Springer, 1987. pp. 1–27.

Osawa M. P300 and neuropsychological function in normal aging and dementia. Electroenceph clin Neurophysiol 1995;44(Suppl):369–376.

Patterson JV, Michalewski HJ, Starr A. Latency variability of the components of auditory event-related potentials to infrequent stimuli in aging, Alzheimer-type dementia, and depression. Electroenceph clin Neurophysiol 1988;71:450–460.

Patterson JV, Pratt H, Starr A. Event-related potential correlates of the serial position effect in short-term memory. Electroenceph clin Neurophysiol 1991;78:424–437.

Pelosi L, Blumhardt LD. Effects of age on working memory: an event-related potential study. Brain Res Cog Brain Res 1999;7:321–334.

Pfefferbaum A, Ford JM, Wenegrat BG, Roth WT, Kopell BS. Clinical application of P3 component of event-related potentials. I. Normal aging. Electroenceph clin Neurophysiol 1984a;59:85–103.

Pfefferbaum A, Wenegrat BG, Ford JM, Roth WT, Kopell BS. Clinical application of the P3 component of event-related potentials. II. Dementia, depression and schizophrenia. Electroenceph clin Neurophysiol 1984b;59:104–124.

Polich J. Normal variation of P300 from auditory stimuli. Electroenceph clin Neurophysiol 1986;65:236–240.

Polich J. Meta-analysis of P300 normative aging studies. Psychophysiology 1996;33:334–353.

Polich J. EEG and ERP assessment of normal aging. Electroenceph clin Neurophysiol 1997;104:244–256.

Polich J. P300 clinical utility and control of variability. J Clin Neurophysiol 1998;15:14–33.

Polich J, Kok A. Cognitive and biological determinants of P300: an integrative review. Biol Psychol 1995;41:103–146.

Pollack VE, Schneider LS. Reliability of late positive component activity (P3) in healthy elderly adults. J Gerontol 1992;47:M88–M92.

Pratt H, Michalewski HJ, Patterson JV, Starr A. Brain potentials in a memory-scanning task. II. Effects of aging on potentials to the probes. Electroenceph clin Neurophysiol 1989;72:507–517.

Rosen WG, Mohs RC, Davis KL. Longitudinal changes: cognitive, behavioral, and affective patterns in Alzheimer's disease. In: Poon LW, editor. Handbook for clinical memory assessment of older adults, Washington, DC: American Psychological Association, 1986. pp. 294–301.

Rosler F, Sutton S, Johnson Jr. R, Mulder G, Fabiani M, Plooij-Van Gorsel E, Roth W. Endogenous ERP components and cognitive constructs. A

review. In: McCallum WC, Zappoli R, Denoth F, editors. Cerebral psychophysiology: studies in event related potentials, Amsterdam: Elsevier, 1986. pp. 51–92.

Rothenberger A, Woerner W, Blanz B. Test-retest reliability of flash-evoked potentials in a field sample: a 5 year follow-up in schoolchildren with and without psychiatric disturbance. Electroenceph clin Neurophysiol 1987;40(Suppl):624–628.

Rugg MD, Walla P, Schloerscheidt AM, Fletcher PC, Frith CD, Dolan RJ. Neural correlates of depth of processing effects on recollection: evidence from brain potentials and positron emission tomography. Exp Brain Res 1998;123:18–23.

Sandman CA, Donnelly JF, O'Halloran JP, Isenhart R. Age-related changes in P3 amplitude as a function of predictable and unpredictable rare events. Int J Neurosci 1990;52:201–211.

Sangal RB, Sangal JM, Belisle C. P300 latency and age: a quadratic regression explains their relationship from age 5 to 85. Clin Electroenceph 1998;29:1–6.

Schellberg D, Glasser T, Kohler W. The intra-individual reproducibility of flash evoked potentials in a sample of children. Int J Psychophysiol 1987;5:135–143.

Segalowitz SJ, Barnes KL. The reliability of ERP components in the auditory oddball paradigm. Psychophysiology 1993;30:451–459.

Simons RF, Miles MA. Nonfamilial strategies for the identification of subjects at risk for severe psychopathology: issues of reliability in the assessment of event-related potential and other marker variables. In: Rohrbaugh JW, Parasuraman R, editors. Event-related brain potentials: basic issues and applications, New York: Oxford University Press, 1990. pp. 343–363.

Sinha R, Bernardy N, Parsons OA. Long-term test-retest reliability of event-related potentials in normals and alcoholics. Biol Psychiatry 1992;32:992–1003.

Sklare DA, Lynn GE. Latency of the P3 event-related potential: normative aspects and within-subject variability. Electroenceph clin Neurophysiol 1984;59:420–424.

Smulders FT, Kenemans JL, Schmidt WF, Kok A. Effects of task complexity in young and old adults: reaction time and P3 are not always dissociated. Psychophysiology 1999;36:118–125.

Squires KC, Wickens C, Squires NK, Donchin E. Sequential dependencies of the waveform of the event-related potential. Science 1976;193:1142–1146.

Storandt M, Botwinick J, Danziger WL. Longitudinal changes: patients with mild SDAT and matched healthy controls. In: Poon LW, editor. Handbook for clinical memory assessment of older adults, Washington, DC: American Psychological Association, 1986. pp. 277–284.

Tueting P, Sutton S, Zubin J. Quantitative evoked potential correlates of the probability of events. Psychophysiology 1970;7:385–394.

van der Wal EA, Sandman CA. Evidence for terminal decline in the event-related potential of the brain. Electroenceph clin Neurophysiol 1992;83:211–216.

Wickens C, Kramer A, Vanasse L, Donchin E. Performance of concurrent tasks: a psychological analysis of the reciprocity of information-processing resources. Science 1983;221:1080–1082.

Wilson RS, Kaszniak AW. Longitudinal changes: progressive idiopathic dementia. In: Poon LW, editor. Handbook for clinical memory assessment of older adults, Washington, DC: American Psychological Association, 1986. pp. 285–293.