Maastricht University

A Comparison of Different Methods for Estimating Single-trial P300 Latencies

Citation for published version (APA):

Smulders, F. T. Y., Kenemans, J. L., & Kok, A. (1994). A Comparison of Different Methods for Estimating Single-trial P300 Latencies. *Electroencephalography and Clinical Neurophysiology*, *92*(2), 107-114. https://doi.org/10.1016/0168-5597(94)90051-5

Document status and date:
Published: 01/03/1994

DOI:
10.1016/0168-5597(94)90051-5

Document Version:
Publisher's PDF, also known as Version of record

Document license:
Taverne

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 13 May. 2024

Electroencephalography and clinical Neurophysiology, 92 (1994) 107–114
© 1994 Elsevier Science Ireland Ltd. 0168-5597/94/$07.00

EEP 93625

A comparison of different methods for estimating single-trial P300 latencies

F.T.Y. Smulders, J.L. Kenemans and A. Kok

Department of Psychonomics, University of Amsterdam, Amsterdam (The Netherlands)

(Accepted for publication: 7 October 1993)

Summary Inferences from comparative analyses of reaction time and P300 latency are stronger when the various aspects of the distribution across trials are treated in the same way for both variables. To this end, a number of studies have resorted to estimation of P300 latency at the single-trial level. This report presents a comparative evaluation of two common methods for such single-trial analysis, i.e., peak-picking and template-matching. Both methods were applied to a representative set of real data, comprising different task conditions and two age groups. Relevant scoring parameters were varied: low-pass filter settings (down to 0.94 Hz) for peak-picking, template duration (250–970 msec) and use of covariance vs. correlation for template-matching, and use of a noise-range criterion for both methods. It is concluded that peak-picking with a 3.4 Hz filter, and template-matching using covariance and template duration between 600 and 800 msec, are best in terms of sensitivity and reliability, with peak-picking surpassing template-matching. Also, the marked increase in the number of rejected trials when the noise-range criterion was applied resulted in unwanted modulation of behavioral effects of task conditions and age groups.

Key words: P300; Single-trial ERP; Template-matching; Peak-picking; Mental chronometry

A promise of event-related brain potentials (ERPs) is that they may give insight into the timing of subprocesses that are involved in the reaction process. In the interpretation of effect sizes on their latencies, a comparison is often made with RT. The strongest conclusions can be drawn if the measure of central tendency of the component's latency is sensitive to the attributes of its distribution in a similar manner as is the measure of central tendency of RT (e.g., the mean). In this case, assumptions are met that make available relatively strong methods of inference that were originally developed for RT, e.g., the additive-factor method (Sternberg 1969). Callaway et al. (1984) showed that the peak latency of an ERP component in an averaged time series is not necessarily equal to the mean of the peak latencies in the original epochs. Therefore, the peak latency of the average does not have these ideal properties. The morphology of the averaged ERP reflects not only the average morphology of the single-trial ERP, but also the shape of the latency distribution across trials. It follows that the effect of skewness in the distribution will be a skewing of the averaged component, and the peak latency of the average will tend toward the latency where the largest number of components across trials reach their peak, i.e., the mode of the distribution [1]. The use of the peak latency of the average can be avoided if there is a reliable estimate for a component's latency in each trial. The aim of the current study was to compare the merits of various procedures that have been developed for the estimation of P300 latency in single trials.

Ruchkin and Glaser (1978) described a filter in the time domain for the examination of the P300 on a single-trial basis. The application of this filter may be

---

Correspondence to: Fren Smulders, Department of Psychology, Roetersstraat 15, 1018 WB Amsterdam (The Netherlands).
Tel.: (31) 20-5256123; Fax: (31) 20-5256710; E-mail: fren@uvapsy.psy.uva.nl.

[1] We tested this argument by means of a simulation experiment in which the duration and skewness of the distribution of components were varied. If the duration of the component was at a minimum, i.e., the duration of one sample, the average wave form, by definition equaling the distribution function, peaked at the latency that equaled the mode of the distribution. However, also when the component (a full-period cosine function) duration was longer, resembling a wave form in real data, the peak latency of the average was equivalent to the mode. It was concluded that, given the assumption of invariance of component wave shape and amplitude over the latency distribution, the average will equal the mode of the latencies. In real data, however, these assumptions may be violated. For instance, Roth et al. (1978) found that a selection of relatively long RTs from a data set was associated with smaller P300s than relatively fast response trials. The issue can be studied more directly, however, if there are reliable estimates of P300 latency on every trial, from which the mode of the distribution can simply be computed, and compared to the latency of the average P300.

SSDI 0168-5597(93)E0264-7

followed by peak-picking: P300 is taken as the latency of the maximum amplitude in a window. As an alternative, one can compute the cross-correlation between a sinusoid template and the (filtered) single-trial ERP for all time points in which a P300 is expected. Then the latency of the maximum in the cross-correlation function is taken as the latency of P300 (Fabiani et al. 1987; Gratton et al. 1989). Others prefer the cross-covariance rather than the cross-correlation because the cross-covariance considers not only wave shape but also amplitude (Pfefferbaum and Ford 1988). The same authors proposed to match the template not only in a time window where a P300 is expected (signal range), but also in a window where *no* P300 is expected (noise range). If in the signal range there is not a better fit of the template to the data than in the noise range, the trial is rejected. A potential disadvantage is that the set of rejected trials may not be unbiased with respect to variables that are of interest to the experimenter. Another variable that may be relevant is the duration of the template. Gratton et al. (1989) found that the use of a template that is relatively wide in comparison to the signal is less accurate than the use of a relatively narrow one. On the other hand, a template that is too narrow may match components of too high a frequency.

Various parameters of template-matching and peak-picking procedures were manipulated in the analysis of a set of real data. Evaluating the effects of variation in template duration, we adopted a half-sine wave as a template (see also Pfefferbaum and Ford 1988; Strayer and Kramer 1990). We did not include average wave forms for templates (Woody 1967) because the adequacy of the average wave form as a template may well vary between experimental conditions, e.g., because of differences in the amount of latency jitter. Other variables that were manipulated were: inclusion of a noise-range criterion, the use of cross-correlation vs. cross-covariance, and low-pass cutoff frequency (for peak-picking).

To evaluate particular combinations of scoring parameters it was first established that varying the parameters indeed changed the experimental effects on P300 latency. Second, the sensitivity, reliability and bias of each scoring condition were determined for various experimental conditions (see Methods). Since subject variables like age may affect the efficacy of P300 identification procedures (Bashore 1990), two age groups were compared.

Methods

Subjects, stimuli and procedure

A group of young subjects (male university students; N = 13; mean age 21.5 years; range 19–26) received course credits, and a group of elderly (male; recruited via an advertisement; N = 15; mean age 69.8 years; range 65–77) received Dfl. 25.00 for participation. All subjects reported to be healthy, right-handed and to have normal or corrected-to-normal vision. One young and 4 old subjects lacked an identifiable P300 in the average Pz wave form and their data were discarded.

Subjects sat alone in a dimly lit room at a distance of 160 cm from a monitor (Zenith/VGA) attached to a PC-AT. Black-on-white stimuli were the words "LINK" and "RECH" (abbreviations for "LINKS" and "RECHTS", Dutch for "left" and "right"), consisting of square dots ($6 \cdot 6$ pixels) forming the 4 letters, surrounded by a frame consisting of similar dots. Each letter was 7 dots (19 mm) high and 5 dots (14 mm) wide. The frame was 35 mm high and 83 mm wide. Degradation of words was achieved by moving 8 dots from the frame to random locations in the field around each letter. There were 4 degraded versions of each word. Interstimulus-interval varied between 1590 and 2090 msec, stimulus duration was 1000 msec.

There were 4 blocks of 136 trials, corresponding to 2 factors (stimulus quality, S-R compatibility) with 2 levels. In compatible blocks subjects were to respond by pressing one of two buttons in their arm-rests with the index finger corresponding to the meaning of the word. In incompatible blocks they were to respond with the other index finger. The order of blocks was counterbalanced across subjects. Training occurred just before the experimental session, until error percentage was below 10%. Responses were to be made as rapidly as possible without making too many errors.

Physiological recording and analysis

The electroencephalogram (EEG) was recorded from Pz (Jasper 1958) by means of a tin electrode referred to linked earlobes. Tin electrodes were also used to record the vertical (above and below the pupil of one eye) and horizontal (at outer canthi of both eyes) electro-oculogram (EOG). A ground electrode was placed on the forehead. Electrode impedance was kept below 8 kΩ. The amplifiers were set to a time constant of 5 sec and 35 Hz low-pass filtering. 100 Hz sampling started at 490 msec before stimulus onset and lasted for 2560 msec.

Incorrect response trials, trials with an RT deviating more than 2.5 S.D. from the mean, and trials with artifacts (saturation of the AD converter or an EEG amplitude greater than 100 μV) were excluded. Ocular artifact was controlled according to Woestenburg et al. (1983). Pre-stimulus samples served as baseline. All valid trials were entered in two procedures: peak-picking and template-matching.

Before peak-picking, epochs were low-pass filtered with one of the following settings: no filter, 6.2 Hz, 3.4 Hz, 1.8 Hz or 0.94 Hz (−3 dB, low pass, Ruchkin and

Glaser 1978). In the signal condition (S), a trial was rejected only if there was no local maximum in the range where P300 was expected: the signal range (300–1000 msec). In the signal/noise (SN) condition, a trial was also rejected if the amplitude of the largest local maximum in the signal range was not larger than in the range where no P300 was expected: the noise range (1010–1940 msec). The latency of the largest local maximum in the signal range served as P300 latency. Taken together there were 5 (filters) × 2 (S vs. SN) scoring conditions for peak-picking.

Before template-matching, epochs were filtered at 3.4 Hz. The template was a half-period sine wave with a duration varying from 250 to 970 msec in 80 msec steps. For each duration, it was moved across each trial in 10 msec lags in the signal and noise ranges (see above). At each lag the correlation and covariance between epoch and template were computed. In the signal condition a trial was rejected only if the correlation was not larger than zero ($P < 0.05$) at any lag. In the signal/noise condition, a trial was also rejected if the maximum fit (correlation or covariance) in the signal range was not larger than in the noise range. Taken together there were 10 (templates) × 2 (correlation vs. covariance) × 2 (S vs. SN) scoring conditions for template-matching.

For each subject, experimental and scoring condition the mean and standard deviation of P300 latency (across single trials) and the number of trials in which P300 was identified (with regard to selection criteria) were determined. The latter number was taken as a measure of *sensitivity*. The standard deviation was considered as a measure of *reliability* of an estimate. The rationale was that a poor algorithm will tend to distribute the indicated latencies evenly across the scoring window, thereby increasing variance. Finally, the difference between the average RT across trials in which P300 was identified and the average RT across all trials was computed to index the *bias* that resulted from trial selection. The rationale was that if the RTs of selected trials do not form a random sample of the total distribution, the latencies of P300 in the selected epochs will also not form an unbiased sample, assuming there is a correlation across trials between P300 latency and RT (Kutas et al. 1977). Each variable was entered in an ANOVA with age group, experimental manipulations, and parameter settings as independent variables. The effects of filter cut-off frequency (peak-picking) and template duration (template-matching) were evaluated using orthogonal-polynomial contrast vectors. Reported *F* values were significant at 5% with *df* (1, 21).

We shall use the following abbreviations: SQ (stimulus quality), SRC (stimulus-response compatibility), FL (filter – linear trend), FQ (filter – quadratic trend), S/SN (signal criterion only vs. signal and noise criteria); CC (covariance vs. correlation), and TDL (template duration – linear trend) and TDQ (template duration – quadratic trend).

Results

The grand average ERPs in all conditions, for both age groups, are depicted in Fig. 1. P300 can be seen as a pronounced deflection extending from about 300 to 1000 msec. The small ramp-shaped potential at 1200 msec is probably related to the offset of the stimulus at 1000 msec. The effects of task conditions on the la-

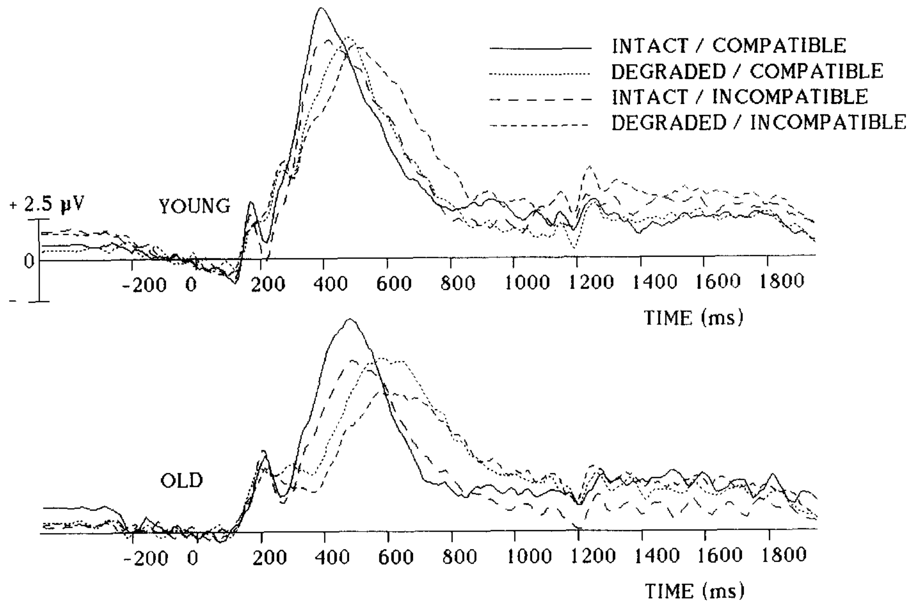

Fig. 1. The averaged ERP on Pz as a function of stimulus quality and S-R compatibility in both age groups.

tency of the average P300 are clearly visible on the maximum and the larger parts of both flanks. P300 appears to be smaller and delayed in the elderly relative to the young group.

Peak-picking

Mean P300 latency. Mean P300 latency is depicted in Fig. 2A. Both task effects, and their interaction with Age, depended on filter length. Relevant F values were 4.3 (SQ × FL), 30.2 (SQ × FQ), 8.6 (Age × SQ × FL), 5.2 (Age × SQ × FQ), 7.2 (SRC × FQ), and 8.0 (Age × SRC × FQ). Fig. 2A shows that task effects were generally largest in the middle frequency range; in older subjects the effects were additionally reduced with high-cutoff settings. Application of the noise-range criterion had no effect on estimated P300 latency.

Number of trials (sensitivity). The numbers of trials with an identifiable P300, as a function of parameter setting, task condition, and age group, are presented in Fig. 2B. With a relatively high filter cutoff frequency (≥ 3.4 Hz) hardly any trial lacked a local maximum in the signal range, which would have prompted rejection. More rigorous filtering led to substantial trial rejection. The number of trials was also reduced in the signal/noise condition, and more so with either no or weak (6.2 Hz), or strong (1.8 Hz, 0.94 Hz) filtering. At a cutoff of 3.4 Hz, the fraction of trials lost due to application of the signal/noise criterion was at a minimum, most notably so with older subjects. Task effects did not depend on scoring parameters. Relevant F values were 38.1 (FL), 456.0 (FQ), 9.2 (Age × FQ), 154.6 (S/SN), 4.5 (FL × S/SN), 123.5 (FQ × S/SN), 11.2 (Age × FL × S/SN), and 4.6 (Age × FQ × S/SN).

Within-subject standard deviation (reliability). As can be seen in Fig. 3A, within-subject standard deviation (WS-SD) decreased with lower filtering cutoff points, and with application of the signal/noise criterion. In some conditions the decrease as a function of filter setting reached an asymptote in the middle frequency range, or even increased for the lowest cutoff point. This was reflected in interactions between FQ, and S/SN ($F = 6.1$), SQ (9.4), and SRC (9.9), respectively. Main effects of filter setting were reflected in significant F values for FL (99.8) and FQ (10.5); the F value for the S/SN main effect was 82.5. The linear filtering effect depended on Age ($F = 8.1$) and on SRC (6.6; Age × SRC, $F = 15.2$).

Reaction-time difference (bias). Fig. 3B shows mean differences between RT based on all trials and RT based on a selection associated with a particular scoring procedure. With the signal/noise criterion, RT differences depended on SRC (S/SN × SRC, $F = 6.4$) and tended to depend on SQ (S/SN × SQ, $F = 4.0$, $P < 0.058$). Thus, SRC effects on RT were modified when trials were selected with regard to the signal/noise criterion. There was no effect of Age.

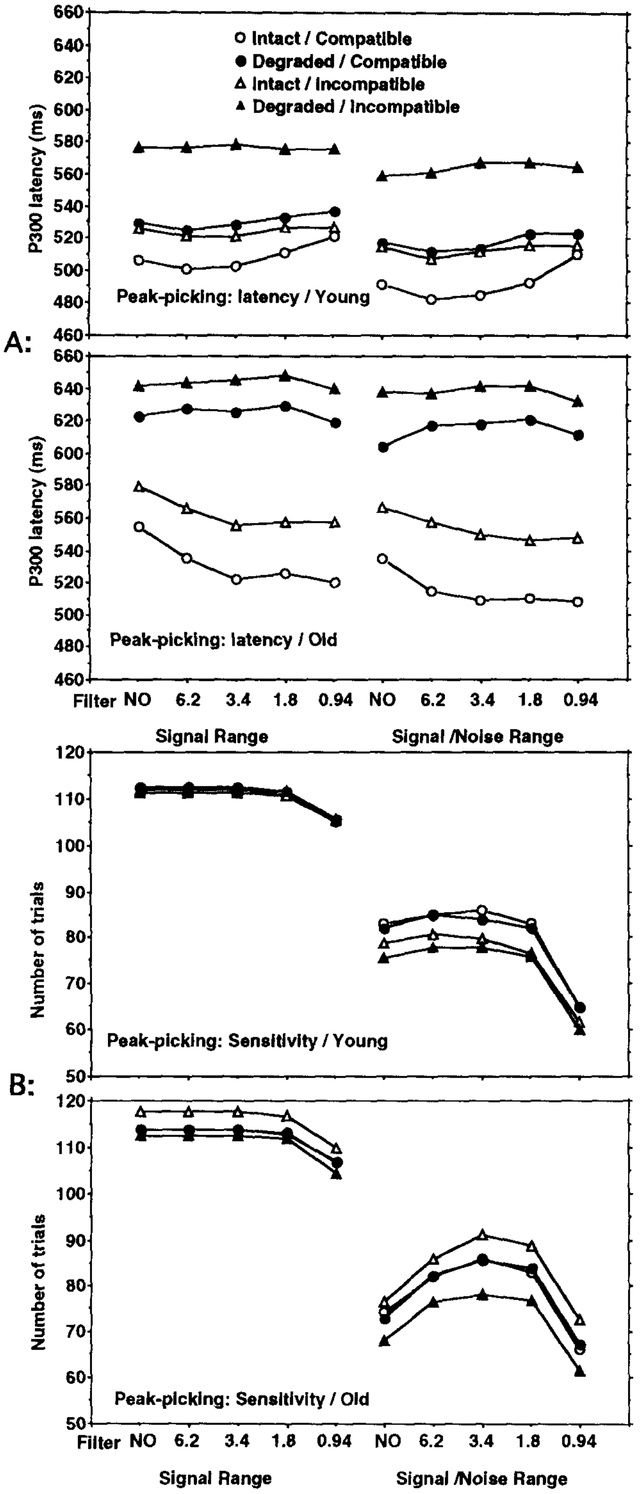

Fig. 2. The mean latency of P300 (A) and the number of trials meeting signal and noise range selection criteria (B) as a function of low-pass filters and signal and noise range selection criteria in different conditions of stimulus quality and S-R compatibility for the two age groups.

Template-matching

Mean P300 latency. Fig. 4A depicts how mean P300 latencies obtained with template-matching procedures

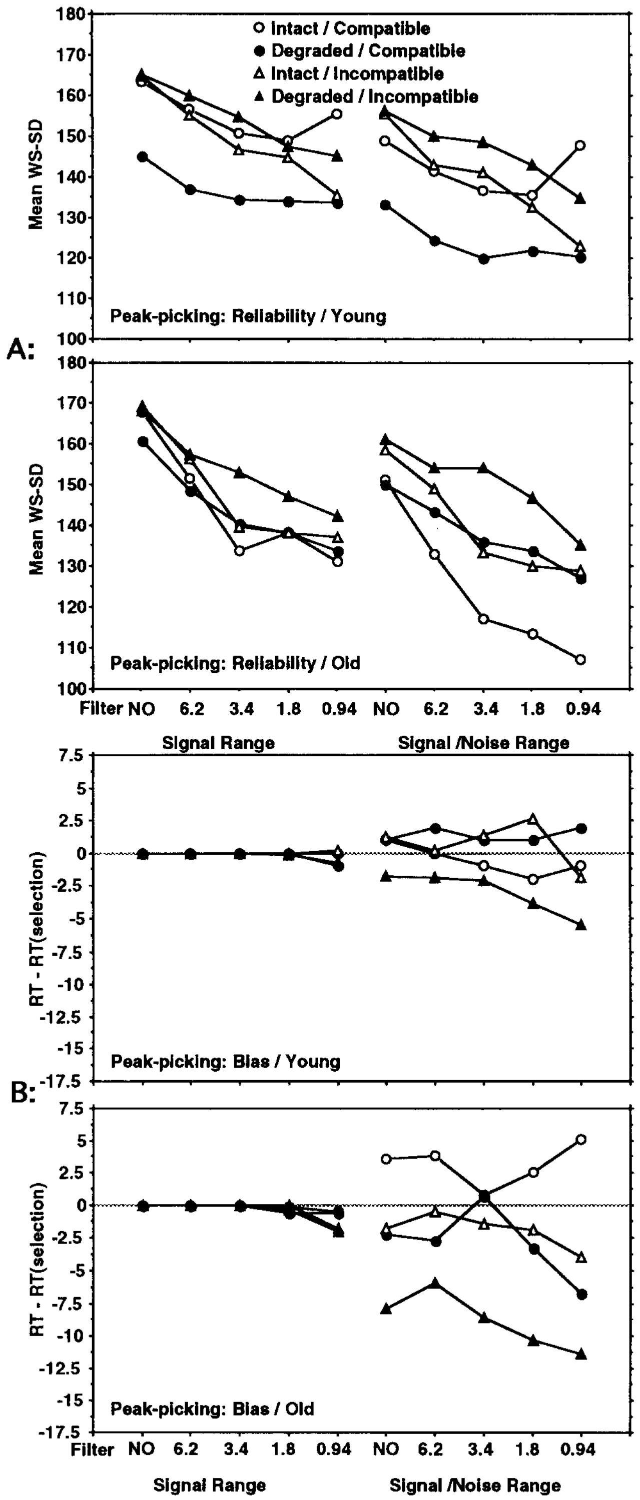

Fig. 3. The mean within-subject standard deviation of P300 latencies (A) and the mean difference between reaction time based on selected trials and reaction time based on all trials (B) as a function of low-pass filters and signal and noise range selection criteria in different conditions of stimulus quality and S-R compatibility for the two age groups.

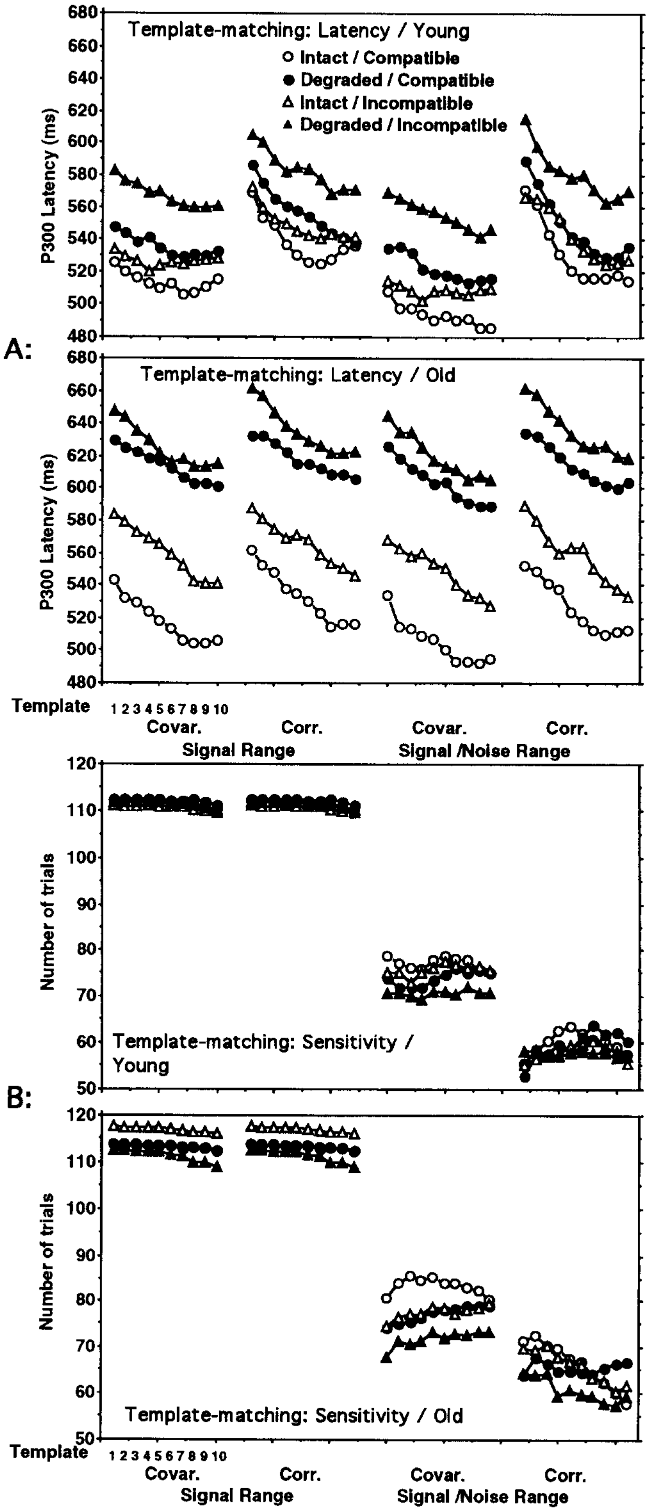

Fig. 4. The mean latency of P300 (A) and the number of trials meeting signal and noise range selection criteria (B) as a function of the use of correlation vs. covariance and signal and noise range selection criteria in different conditions of stimulus quality and S-R compatibility for the two age groups. Template numbers 1–10 refer to sinusoid templates with durations of 250, 330, 410, 490, 570, 650, 730, 810, 890 and 970 msec, respectively.

were affected by scoring-parameter settings. First, SQ effects were larger with application of the signal/noise criterion ($F = 10.5$). Second, interactions between SQ and SRC depended both on Age and on the applied measure of fit (covariance vs. correlation); F values were 5.0 (for both SQ × SRC × CC and Age × SQ × SRC × CC). Third, there was an interaction between SQ, CC, template duration (linear), and S/SN ($F = 8.5$). Thus, task and group effects on estimated P300 latency depended on parameter setting for the template-matching procedure.

Number of trials (sensitivity). For nearly all trials the criterion of significance of the maximum correlation between the trial and the template in the signal range was met. Consequently, the loss of trials in the S condition varied between 0.1 and 2.2% (Fig. 4B, left). The inclusion of the noise-range criterion led to a marked increase of rejected trials (Fig. 4B, middle-right), especially when correlation was used instead of covariance (about 33.7–49.9% vs. 26.6–38.8% trials lost). Relevant F values were 489.8 (S/SN), and 66.1 (S/SN × CC). The effect of template duration depended on Age, CC, and S/SN ($F = 8.0$ for Age × S/SN × CC × TDL, and 5.9 for Age × CC × S/SN × TDQ). These effects seem to reflect modest trends of opposite signs across Age × S/SN × CC conditions.

The effects of scoring parameter setting also depended on task conditions. For degraded stimuli, relative to intact, the loss of trials when applying the signal/noise criterion was larger when covariance was used (SQ × CC × S/SN, $F = 4.6$). For incompatible stimuli more trials were lost after application of the SN criterion than for compatible stimuli (SRC × S/SN, $F = 5.3$).

Within-subject standard deviation (reliability). Application of the SN criterion, combined with covariance rather than correlation, resulted in a reduction of the within-subject standard deviation (see Fig. 5A; CC × S/SN, $F = 33.1$). This was especially so with longer template durations (CC × S/SN × TDL, $F = 34.6$), and for older subjects in the intact-compatible condition with all but the shortest template durations (Age × SQ × SRC × CC × S/SN × TDQ, $F = 4.6$). The use of covariance also had a more general effect in reducing standard deviation in intact-compatible stimuli (Age × SQ × SRC × CC, $F = 6.0$). The use of longer template durations led to a decrease in standard deviation, but standard deviations leveled off (signal/noise criterion) or increased again (signal-only criterion) at longer template durations, i.e., 730 or 810 msec. Relevant F values were 33.7 (TDL), 25.2 (TDQ), 6.1 (Age × TDL), 51.5 (S/SN) × TDL), and 34.6 (CC × S/SN × TDL).

Reaction-time difference (bias). Fig. 5B shows mean differences between RT based on all trials and RT based on a selection associated with a particular scoring procedure. RT differences were virtually absent for signal-only conditions. With the signal/noise criterion, there were interactions between CC and SRC (S/SN × CC × SRC, $F = 11.0$), and between CC, Age, and

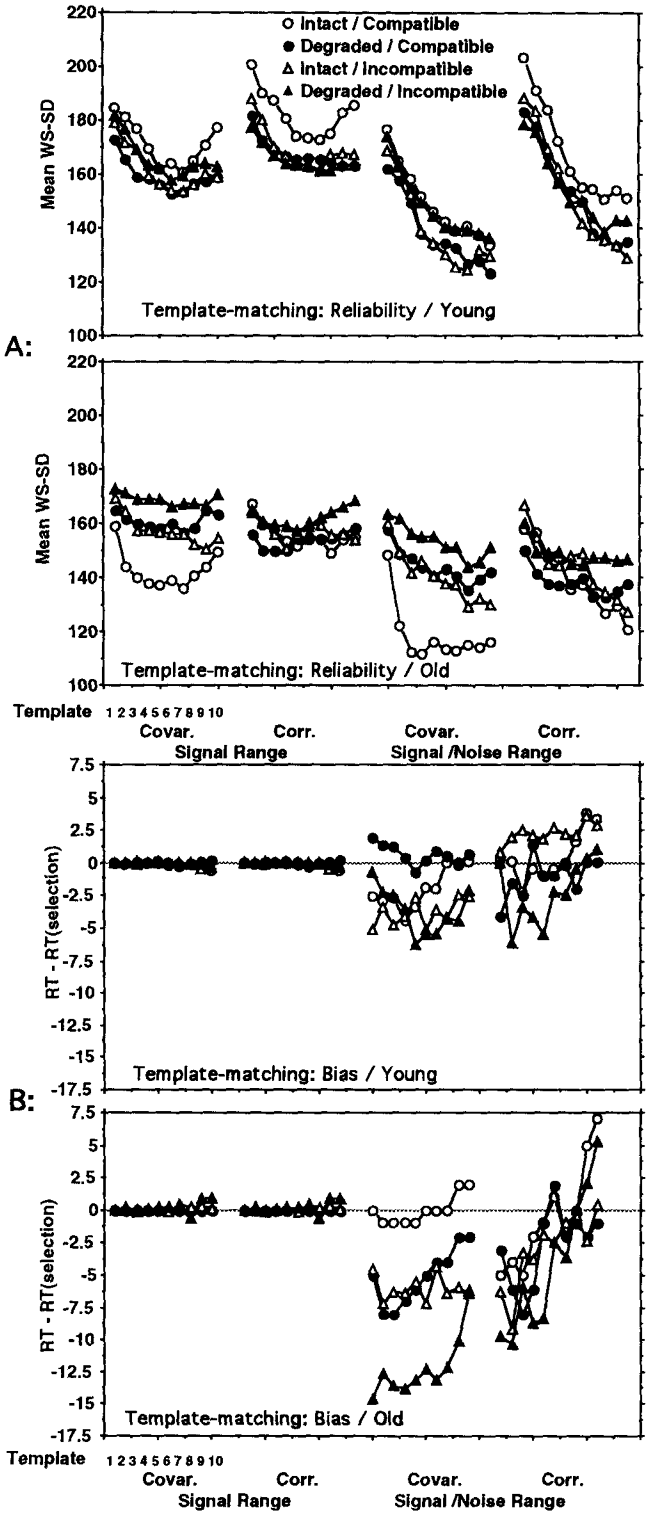

Fig. 5. The mean within-subject standard deviation of P300 latencies (A) and the mean difference between reaction time based on selected trials and reaction time based on all trials (B) as a function of the use of correlation vs. covariance, various templates and signal and noise range selection criteria in different conditions of stimulus quality and S-R compatibility for the two age groups.

SQ (S/SN × CC × Age × SQ, $F = 16.9$). That is, bias in RT was not distributed evenly across groups and task conditions, especially when covariance was used, and with older subjects.

Discussion

For both peak-picking and template-matching, the pattern of effects on mean single-trial P300 latency depended on the scoring parameters. Thus, for both methods, we may ask which particular scoring parameters are most adequate in terms of three criteria: sensitivity, reliability and bias.

Peak-picking

As might be expected, sensitivity of the peak-picking method was severely reduced by application of the signal/noise criterion (see Fig. 2B). Without this criterion, low-pass filter settings of 6.2, 3.4, and 1.8 Hz or no filter appeared more adequate than 0.94 Hz. With the signal/noise criterion, 3.4 Hz was the most adequate cut-off frequency, across task conditions and age groups. It should be noted that the number of trials meeting the signal/noise criterion is also of importance in the choice of parameter settings in the signal condition, as it can be considered a measure of the extent to which P300 detections in the signal range are truly (as opposed to false) positive, i.e., their specificity.

Both application of the signal/noise criterion and lowering of the filter cut-off frequency increased reliability of peak-picking (see Fig. 3A). The filter effect seemed to be stronger with older subjects and in compatible conditions. In both age groups the filter effect reached an asymptote around 3.4 Hz in about half of the conditions. A further lowering of the cut-off frequency had the effect of both decreasing sensitivity and increasing reliability in a number of conditions. A related result for simulated data was reported by Gratton et al. (1989). These authors found that with increasing filtering detection accuracy also increased, even when the signal itself became distorted.

The reduction in the number of trials through application of the signal/noise criterion was associated with a significant bias with respect to S-R compatibility effects on RT (see Fig. 3B). This modulation of task effects argues against the use of the noise range criterion.

Template-matching

Sensitivity of the template-matching procedure was also severely reduced by application of the signal/noise criterion (see Fig. 4B). Combined with the correlation criterion, the number of selected trials came close to 50%, which would indicate equal probabilities of large positive waves being detected in latency ranges with and without a P300. The effects of template duration were modest and mixed, having different directions in different combinations of group and task.

Effects of template duration on reliability of the template-matching procedure were less ambiguous. With increasing duration reliability either increased exponentially, reaching an asymptote for intermediate durations, or was largest for intermediate durations (i.e., 570–810 msec, see Fig. 5A). Reliability was also increased by the combined use of signal/noise criterion and covariance, especially for young subjects; for the elderly this was more dependent on specific task conditions.

As with peak-picking, application of the signal/noise criterion introduced a significant bias in RT. This appeared to be mainly so when using covariance for older subjects, in the more difficult task conditions (see Fig. 5B). Thus, the difference in mean RT of selections and mean RT of all trials depended both on age group and on task condition, indicating that in the covariance/noise criterion condition age and task effects were modulated, relative to the unbiased set of trials in the signal-only conditions. As in the case of peak-picking, this argues against the use of the signal/noise criterion. Furthermore, the use of covariance, rather than correlation, increased both sensitivity and reliability. Reliability was generally at maximum with template durations of 650 or 730 msec.

Peak-picking versus template-matching

Our choices then, are peak-picking with 3.4 Hz low-pass filtering, or template-matching using covariance and template durations between 600 and 800 msec. In both cases the use of a noise-range criterion should be avoided, or its selection effects on the pattern of behavioral results evaluated. Comparing the two optimal methods, the signal/noise data (Figs. 2B and 4B) indicate higher sensitivity for peak-picking (3.4 Hz filter) than for template-matching (covariance, 730 msec duration), in both age groups. As noted earlier, the relative sensitivity in the signal/noise condition is also important to the choice of parameters in the signal condition. As to reliability, comparison of Figs. 3A and 5A (signal range) also favors peak-picking (3.4 Hz) over template-matching (covariance, 730 msec). Bias (Figs. 3B and 5B) in the signal-only conditions was negligible for both methods.

Some attention should be given to the difference in results between single-trial methods and P300 latency estimation in the average (see Introduction; footnote 1). Fig. 6 shows P300 latency in the averaged wave form after 3.4 Hz low-pass filtering, defined as the latency of the maximum amplitude between 350 and 850 msec post stimulus. The same figure also shows the average means and modes of the single-trial P300

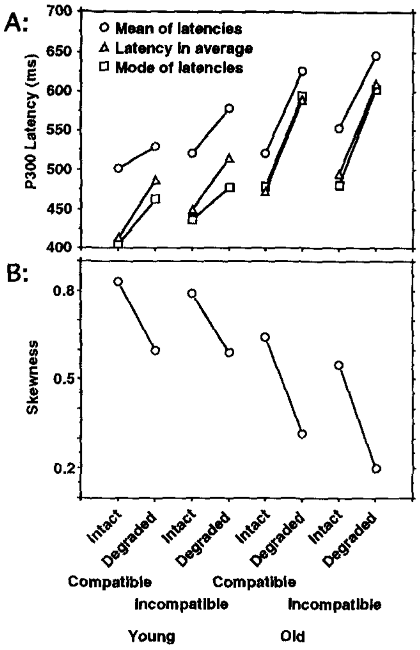

Fig. 6. A: the mean and mode of the P300 latency distribution and the latency of the average P300 in different conditions of stimulus quality and S-R compatibility for optimal template-matching and peak-picking for the two age groups. B: skewness of the P300 latency distribution in the same conditions.

latencies. Average wave form latencies were generally closer to the mode of the single-trial distribution than to the mean. This is in line with the notion that the peak latency of an average wave form tends toward the mode of the distribution of single-trial latencies. It can also be seen that the effect of stimulus quality (intact vs. degraded) was larger for average wave form latencies (and modes) than for means. This indicates more skewed distributions for intact than for degraded stimuli, which is confirmed by the actual skewness values (computed as 3 (mean − median)/S.D., Ratcliff 1979).

In conclusion, it was possible to distinguish between the merits of different methods for scoring P300 latency on single trials in a set of real data. Peak-picking appeared to be more adequate than computationally more elaborate template-matching. However, this method may still not be as good as one would wish: both the power to discriminate between epochs with and without a P300 and the minimization of within-subject variation may be subject to future improvements.

We wish to thank Martin Elton for his comments on an earlier version of the manuscript, and Wouter Schmidt for designing the stimuli.

References

Bashore, T.R. Age-related changes in mental processing revealed by analyses of event-related potentials. In: J.W. Rohrbaugh, R. Parasuraman and R. Johnson, Jr. (Eds.), Event-Related Brain Potentials: Basic Issues and Applications. Oxford Univ. Press, New York, 1990: 242–275.

Callaway, E., Halliday, R., Naylor, R. and Thouvenin, D. The latency of the average is not the average of the latencies. Psychophysiology, 1984, 21: 571.

Fabiani, M., Gratton, G., Karis, D. and Donchin, E. Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. In: P. Ackles, J.R. Jennings and M.G.H. Coles (Eds.), Advances in Psychophysiology, Vol. 2. JAI Press, Greenwich, CT, 1987: 1–78.

Gratton, G., Kramer, A.F., Coles, M.G.H. and Donchin, E. Simulation studies of latency measures of components of the event-related brain potential. Psychophysiology, 1989, 26: 233–248.

Jasper, H.H. The ten–twenty electrode system of the International Federation. Electroenceph. clin. Neurophysiol., 1958, 10: 371–375.

Kutas, M., McCarthy, G. and Donchin, E. Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. Science, 1977, 197: 792–795.

Pfefferbaum, A. and Ford, J.M. ERPs to stimuli requiring response production and inhibition: effects of age, probability and visual noise. Electroenceph. clin. Neurophysiol., 1988, 71: 55–63.

Ratcliff, R. Group reaction time distributions and an analysis of distribution statistics. Psychol. Bull., 1979, 86: 446–461.

Roth, W.T., Ford, J.M. and Kopell, B.S. Long-latency evoked potentials and reaction time. Psychophysiology, 1978, 15: 17–23.

Ruchkin, D.S. and Glaser, E.M. Simple digital filters for examining CNV and P300 on a single trial basis. In: D.A. Otto (Ed.), Multidisciplinary Perspectives on Event-Related Brain Potential Research. U.S. Government Printing Office, Washington, DC, 1978: 579–581.

Sternberg, S. The discovery of processing stages: extensions of Donders' method. Acta Psychol. (Amst.), 1969, 30: 276–315.

Strayer, D.L. and Kramer, A.F. Attentional requirements of automatic and controlled processing. J. Exp. Psychol.: Learn., Mem. Cogn., 1990, 16: 67–82.

Woestenburg, J.C., Verbaten, M.N. and Slangen, J.L. The removal of eye-movement artefact from the EEG by regression analysis in the frequency domain. Biol. Psychol., 1983, 16: 127–147.

Woody, C.D. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Med. Biol. Eng., 1967, 5: 539–553.