Psychophysiology, 42 (2005), 16–24. Blackwell Publishing Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2005.00264.x

EOG correction: A comparison of four methods

RODNEY J. CROFT,[a,b] JODY S. CHANDLER,[b] ROBERT J. BARRY,[b] NICHOLAS R. COOPER,[c] AND ADAM R. CLARKE[b]

[a]Brain Sciences Institute, Swinburne University of Technology, Melbourne, Australia
[b]Brain & Behaviour Research Institute, University of Wollongong, Wollongong, Australia
[c]Cognitive Neuroscience & Behaviour, Imperial College, Faculty of Medicine, London, UK

Abstract

EOG correction is a class of techniques that account for ocular artifact in the electroencephalogram (EEG) by subtracting electrooculographic data from the EEG. The purpose of this study was to evaluate four of these correction techniques (Verleger, Gasser, & Möcks, 1982 [VGM]; Gratton, Coles, & Donchin, 1983 [GCD]; Semlitsch, Presslich, Schuster, & Anderer, 1986 [SPSA]; Croft & Barry, 2000 [CB]). Blinks, vertical eye movements (VEM), and horizontal eye movements (HEM) from 26 subjects were corrected using these techniques, and eye movement event-related potentials computed to aid validation. HEMs were corrected better by CB, VGM/GCD then SPSA, VEMs by CB, VGM/GCD then SPSA, and blinks by CB, SPSA, GCD and then VGM, with the advantage of CB substantial for blinks ($\eta^2 > .72$), VEMs ($\eta^2 > .60$), and HEMs ($\eta^2 > .27$). It is argued that the CB procedure adequately accounts for ocular artifact in the EEG. Reasons for the limitations of the other procedures are discussed.

Descriptors: Electroencephalogram, EOG, Correction, Ocular artifact, Comparison

Eye movements alter the electromagnetic fields around the eyes and typically introduce artifact into the electroencephalograph (EEG). A number of methods have been proposed to account for ocular artifact, ranging from the fixation/rejection technique (reducing eye movement through instructions to fixate and refrain from blinking, and rejecting data contaminated by eye movement), subtracting ocular artifact from the EEG (EOG correction; e.g. Hillyard & Galambos, 1970) and more recently the separation of ocular and neural potentials through component analyses (e.g., ''Multiple Source Eye Correction''; Lins, Picton, Berg, & Scherg, 1993; ''independent component analysis''; Vigario, 1997). Although there has been considerable debate regarding the best means of accounting for ocular artifact, there is currently no consensus on the issue.

The most popular of these methods for accounting for ocular artifact are the fixation/rejection and EOG correction techniques. However, the fixation/rejection procedure suffers from a number of serious problems. For example, fixation instructions introduce a secondary task that has been shown to affect the CNV (Weerts & Lang, 1973), N1 (Verleger, 1991), and P300 (Ochoa & Polich, 2000; Verleger, 1991) components of the ERP, and rejecting data is not only difficult to accomplish (Verleger, 1993) but it also results in biased sampling due to the link between eye movement and cognition (Anthony, 1985).

Conversely, the main limitation involved with EOG correction is that as well as reducing ocular potentials in the EEG, it also reduces neural potentials (due to neural potentials being recorded in the EOG channels and subtracted from the EEG). This limitation has been viewed by many as strong a priori grounds for not employing EOG correction, even given the limitations of EOG rejection cited above. However, until recently the relevance of this limitation has not been explored, and so the choice to employ EOG rejection rather than correction was based on untested assumption, and the only study that has explored this issue did not support the assumption (Croft & Barry, 2002). Specifically, in terms of the N1P2 complex of the auditory event-related potential, it was found that if we assume an accurate correction coefficient, the maximum attenuation of this complex was 15–22% of the neural potentials at prefrontal sites, or the equivalent of applying the impractical EOG rejection criterion of 4 μV. Particularly as that percentage error will be independent of cognition (dependent on such physical parameters as head and brain morphology), we view this limitation as relatively minor and would thus recommend it over the fixation/rejection procedure and its more pernicious limitations described above (for a review of this issue, see Croft & Barry, 2000a).

With regard to EOG correction, a number of different algorithms have been proposed. What these algorithms have in common is that they all account for ocular artifact by subtracting a proportion (B) of one or more EOG channels from the EEG, where the EOG channels are any recording montages thought to provide reasonable estimates of ocular activity. Bs are estimated separately for each scalp electrode, EOG channel and subject, and although the estimates are essentially the magnitude of an eye movement at a scalp site divided by the magnitude of the same eye movement at an EOG channel, Bs are usually estimated using least-squares regression to minimize the effect of noise on the estimates. Equation 1 provides an example of an EOG

Address reprint requests to: Rodney J. Croft, Brain Sciences Institute, Swinburne University of Technology, Hawthorn 3122, Australia. E-mail: rcroft@bsi.swin.edu.au.

correction algorithm, where estTEEG is the corrected or *estimated true EEG*, MEEG is the measured EEG, *i* is time point, *j* is scalp site, VEOG and HEOG are the vertical and horizontal EOGs, respectively, and *B*1 and *B*2 are the VEOG and HEOG correction coefficients, respectively:

$$\text{estTEEG}ij = \text{MEEG}ij - \text{B1}j * \text{VEOG}ij - \text{B2}j * \text{HEOG}ij \quad (1)$$

There are a number of differences between the EOG-correction algorithms, and although we believe that there is strong theoretical support for some but not other aspects of these algorithms (see Croft & Barry, 2000a), empirical demonstrations have not been strong enough to convince EEG researchers in general (see, e.g., the extensive debate in Brunia et al., 1989, and more recently the related commentaries of Croft, 2000; Picton et al., 2000a; Verleger, 2000). The investigation of four EOG-correction algorithms with contrasting features may offer a means of clarifying this uncertainty (Table 1).

Verleger, Gasser, and Möcks (1982; VGM) designed an early procedure that differs from a number of others in that it does not differentiate between artifact caused by blinking and vertical non-blink eye movements (VEM). Whether or not these eye movement types affect the EEG differentially (and need to be corrected differently) has been the subject of a great deal of debate. For instance, whereas early research found different *B*s for blinks and VEMs (Corby & Kopell, 1972), a number of research groups failed to find differences between data corrected with and without separate *B*s (Brunia et al., 1989), and further research demonstrated that under certain conditions the distinction disappeared when noise was adequately controlled for (Croft & Barry, 1998a). However, subsequent research demonstrated that the conclusions of Croft and Barry were based on an unrepresentative sample of scalp sites, and, using sensitive protocols, strong evidence for differential blink/VEM propagation was provided using both dipole analysis (Picton et al., 2000b) and more standard EOG correction techniques (Croft & Barry, 2000b). Thus we believe that there is strong evidence that blinks and VEMs propagate and need to be corrected differently, and correspondingly, that there will be limitations to the VGM procedure.

Gratton, Coles, and Donchin (1983; GCD) designed a procedure that has been used extensively in EEG research. It differs from the above VGM procedure primarily in that it corrects blinks and VEMs separately, and it adjusts for event-related potentials (ERPs) in the *B* estimation process. The latter feature is achieved by subtracting the event-related activity in both EEG and EOG channels from the corresponding raw EEG and EOG, prior to *B* estimation. Although this may be an important improvement to the *B* estimation process, it has not been adequately tested, and as this innovation also reduces the power of the ocular artifact signal in both the EOG and EEG channels, there is the possibility that the *B* estimation may suffer due to reduced signal (ocular artifact) to noise (neural activity) ratios (see Croft & Barry, 1998b).

Implemented as the default EOG correction procedure on Neuroscan (a popular data analysis software package), Semlitsch, Presslich, Schuster, and Anderer's (1986; SPSA) correction procedure is also widely used. Its principal feature is that it increases the accuracy of the *B* estimation by calculating *B*s from an eye movement average, which should provide more accurate *B* estimates because it reduces noise in the same way that an ERP does. This procedure also differs from the first two above in that it does not correct horizontal eye movement (HEM) at all, and further, as its *B*s are calculated from blinks only, it should result in poorer correction of VEMs. It also differs from the GCD procedure in that it does not account for ERPs in the *B* estimation process.

A recently developed algorithm is that of Croft and Barry (2000b; CB), which estimates correction coefficients from pre-experiment calibration data. Unlike the GCD procedure that separates blinks and VEMs on the basis of a pattern recognition algorithm, it deals with this distinction by assuming that both eye movement types are composed of a mixture of globe rotation and eyelid movement, and corrects all data points with separate globe rotation and eyelid movement *B*s (using the VEOG and radial EOG [REOG; the mean of voltages immediately above and below the eye] channels, respectively). It also employs the signal-to-noise enhancement of the SPSA procedure; however, it extends this by creating VEM and HEM averages from which globe rotation and horizontal *B*s are calculated, respectively. Another feature of this procedure is that, similar to the GCD procedure, it accounts for ERPs in the averages that the *B*s are calculated from. However, it differs from the GCD procedure in that it does not subtract a portion of the waveform to account for event-related neural potentials, but rather it combines the averages from which the *B*s are calculated in such a way as to make the ocular and neural potentials approximately orthogonal, and thus avoids the problem of reducing the power of the ocular potentials through subtraction (mentioned above).

There are thus four EOG correction methods with a number of different features, and if we can draw out differences in their abilities to correct data we can resolve a number of EOG-correction issues. However, as was shown by the Tilburg symposium on EOG correction (Brunia et al., 1989), where different procedures were employed and yet differences in outcomes not found, it is difficult to demonstrate differences between the methods because it is not clear what a corrected waveform should look like (i.e., we have no TEEG*ij*). Correlation analyses have been employed in an attempt to overcome this difficulty. For example, Gasser, Sroka, and Möcks (1986) assumed that good correction should result in similar broadband power values for different epochs of a subject's resting EEG.

Table 1. *The Correction Features Employed by the Four EOG Correction Algorithms Compared in This Article*

Algorithm	Horizontal EM	ERP adjustment	Separate blink/vertical	Separate eyelid/globe	Signal-to-noise ratio
varVGM	√	×	×	×	×
GCD	√	√	√	×	×
SPSA	×	×	×	×	√
CB	√	√	×	√	√

While improving matters through rigorous quantification, such methods are limited by the correlation that would also result from residual EOG in the EEG channels, and so this approach to validation is not appropriate. Some studies attempt to overcome this difficulty by employing simulations whereby a *model EEG* is created/simulated and thus the adequacy of correction algorithms is assessed by comparing the corrected to the model EEG (e.g., Croft & Barry, 1998b, 2000c; van den Berg-Lenssen, van Gisbergen, & Jervis, 1994). However, such tests are based on a number of assumptions that are themselves in need of determination. For example, if the *model EEG* was created assuming that blinks and VEMs propagate differently and a particular procedure produced poor corrections, the origin of the problem could be either the assumption of different propogations used to create the model EEG or the correction procedure. Hence such studies can only clarify certain aspects of the correction algorithms, and we need to turn to real data for more comprehensive testing.

To overcome the above validation difficulties, what we propose here is to correct eye movements themselves, as these can be manipulated in such a way that correction inaccuracies become obvious. For instance, although it is difficult to determine whether similarities between an "up" eye movement trace in the VEOG and a corrected Fz channel are due to poor correction or genuine similarities between neural and ocular voltages, we can remove the ambiguity by also comparing a corrected "down" eye movement. This is because the up and down movements will result in ocular voltages of opposite polarity, whereas the neural potentials to these movements will be similar. Thus by combining the two eye movements "end to end," we have a scenario where VEOG and the corrected EEG at Fz should not be correlated. This is demonstrated in Figure 1A, where the VEOG and uncorrected EEG traces are of the same polarity during both the up and down eye movements (positively correlated), whereas the uncontaminated EEG is negatively correlated with VEOG during the down and positively correlated with VEOG during the up portion of the figure (leaving the overall uncontaminated Fz and VEOG traces orthogonal). By combining "left" and "right" ocular movements end to end we can similarly test corrections for HEMs. Blink correction is less easily tested because we do not have an "opposite" eye movement to compare a blink to. Thus to test the adequacy of blink corrections we shall employ linear regression techniques to determine the proportion of ocular voltage remaining in the EEG, but shall have to rely on other measures to aid in the validation of these results (i.e., scalp topography). Further, to allow comparison with the GCD procedure (where *B*s are estimated from the data to be corrected), *B*s for all correction procedures will be estimated from the eye movements to be corrected.

Based on the earlier discussion of the strengths and weaknesses of the four correction procedures, a number of predictions can be made. With regard to the correction of HEMs, the SPSA procedure should perform worst because it makes no attempt to correct this type of artifact, the VGM and GCD procedures should both correct well, and differ only insofar as the ERP adjustment of GCD improves/impairs correction, and the CB procedure should perform best because it enhances signal-to-noise ratios and also controls for forward propagation of neural potentials in the *B* estimation process without minimizing ocular potentials. With regard to the correction of VEMs, the SPSA procedure should again perform worst because it estimates *B*s from blinks only, and as argued above, there is good reason to

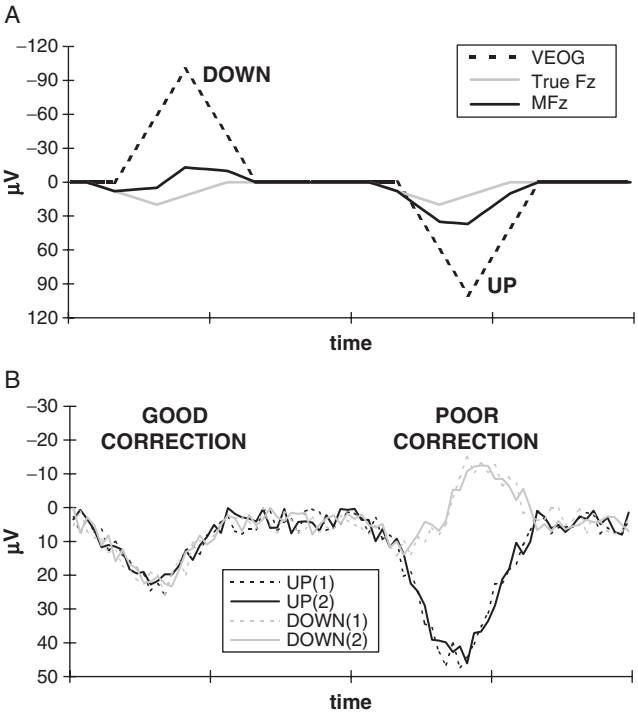

Figure 1. A: VEOG and neural potentials at Fz contaminated (MFz) and uncontaminated (True Fz) by ocular artifact are shown for hypothetical "down" and "up" eye movements. Whereas the VEOG and MFz waveforms are of opposite polarity during the up and down movements, True Fz is similar in both and uncorrelated overall with VEOG. For instance, in this example, the correlation between VEOG and MFz is .84, and between VEOG and True Fz it is .00. B: Hypothetical EEG corresponding to up and down eye movements that have been EOG corrected are shown. As can be seen on the left, where correction is good the EEG for different epochs and types will look very similar. However, as can be seen on the right, a poor correction will still result in EEG from similar eye movements that appear similar, even though they do not accurately represent neural activity. This is why a dissociation of eye movement types needs to be built into validation procedures that reply on split-half reliabilities.

believe that blinks and VEMs propagate differently. The VGM procedure should perform better than the SPSA procedure because it estimates *B*s from a mixture of blinks and VEMs, resulting in *B*s somewhere between those of blinks and VEMs. The GCD procedure should perform better than VGM because it estimates separate *B*s for blinks and VEMs, and the CB procedure should provide the best correction because it corrects separately for the eyelid and globe-rotation aspects of VEMs, it enhances signal-to-noise ratios through averaging, and it also accounts for forward propagation of neural potentials without reducing ocular potentials. With regard to blink correction, it is predicted that the VGM correction will perform worst, because as argued above, it obtains *B*s that fall somewhere between the propagation values of blinks and VEMs and thus does not provide an accurate correction of the blinks themselves. The GCD procedure should perform better because it estimates blink and VEM *B*s separately, but it should not perform as well as the SPSA procedure because it does not maximize the signal-to-noise ratios as does the SPSA procedure (and there may be a problem with reducing the blink signal in the *B* estimation process due to the ERP subtraction technique). The CB procedure should

perform better than the others in this condition because it can account for both the eyelid movement and globe-rotation aspects of the blinks, it employs a similar signal-to-noise ratio enhancement to that of SPSA, and it accounts for ERPs without minimizing ocular potentials in the *B* estimation process.

Methods

Participants

Ten women (9 right- and 1 left-handed) and 16 men (13 right- and 3 left-handed) aged 21 to 48 (mean = 29.3) years participated in the study and were paid for their time. Each gave written informed consent and was free to withdraw from the study at any time without penalty.

Procedure

Upon arrival at the laboratory, participants completed consent forms and had EEG recording apparatus attached. They were then seated in an armchair in an electrically shielded sound-attenuated booth, and verbally instructed to keep their head and body reasonably still and to perform the EOG calibration task (described below). This task lasted 4 min, 10 s.

EOG Calibration Task

There were two parts to this task. In both, the participant was positioned such that their eyes were approximately 40 cm from and at the same height as the center of a 30-cm (diagonal) computer monitor. The plane of the monitor screen was perpendicular to both the floor and the sagittal plane of the subject, and this monitor was used to display stimuli for both parts of this task. The first part of the task involved the sequential presentation of 0.8-s-duration 2.5 × 1.5 cm dark red rectangles (SOA = 1.0 s) in one of four locations (center top, a; center bottom, b; left middle, c; and right middle, d). The presentation order was pseudorandom, with a total of 50 eye movements for each of a to b (down), b to a (up), c to d (right), and d to c (left). The participant was cued to follow the rectangles around the screen via written instructions on the screen preceding this part of the task. The second part of the task involved the alternate presentations of 0.8-s-duration low intensity dark-red and high intensity royal blue rectangles (2.5 × 1.5 cm) at the center of the screen (SOA = 1.0 s). The participant was cued, via written instructions on the screen immediately preceding this part of the task, to blink whenever the central square changed.

Data Acquisition

EEG was recorded from 19 scalp sites (Fp1, Fp2, F7, F8, F3, Fz, F4, C3, Cz, C4, T3, T4, P3, Pz, P4, T5, T6, O1, O2) plus the right ear, and referenced to the left ear using tin electrodes. Participants were grounded midway between Fz and Fpz. EOG was recorded above (E1) and below (E3) the left eye and from the outer canthi of the left (E5) and right (E6) eyes. VEOG was computed as E1 − E3, HEOG as E5 − E6, and REOG as (E1 + E3)/2. A gain of 2500 was used for each channel with a bandpass of 0.05 to 100 Hz. Impedances were below 5 kΩ at the start of the recording and data were digitized at 500 Hz.

Data Analysis

Following re-referencing to linked ears, data were epoched −200 to 800 ms post stimulus, baseline corrected, and then each of the ocular artifact methods was applied to the data. For each resulting data set, data were baseline corrected and averages created for each of the five eye movement types. The "up" average was then appended to the "down," and the "left" to the "right" to form "vertical" and "horizontal" averages, respectively. There were thus three types of averages—"vertical," "horizontal," and "blink."

Validation Measures

Two dependent variables were employed to test the validity of the four correction methods. These variables are not independent, but rather probe different aspects of the data that highlight the relative adequacy of the corrections. The first is based on the assumption that the EOG and EEG channels are *relatively* uncorrelated when viewed over the whole up–down or left–right averages (see introduction and Figure 1A for an explanation), and we shall refer to this as the "regression validation." This involved employing least-squares linear regressions to determine the magnitude of the relation between the corrected data and the EOG channels, where corrected EEG was the criterion variable, HEOG the predictor variable for the horizontal averages, and VEOG the predictor variable for the vertical and blink averages. These were calculated separately for each method, site, and subject, and the resultant unstandardized beta coefficients converted to absolute value to ensure that errors of opposite polarity were not obscured in the statistical analyses (abs(*B*)).

Although it may appear circular to employ least-squares regression to test the result of algorithms that are essentially least-squares regressions, it should be noted that the corrections and validations apply to different subsets of data. For example, whereas the VGM procedure estimates correction coefficients with least-squares regression over the combined blink, vertical, and horizontal eye movement data, the validation regression is over *either* the horizontal, vertical, *or* blink data. In the case of VGM, such a distinction may provide, for example, a test of the view that blinks and saccades can be corrected with the same coefficients (Verleger, 2000). That is, although there may not be a relation between the corrected EEG and the VEOG channel across both eye movement types, the over- and undercorrection that might occur in that instance (e.g., as per the views of Croft, 2000, and Picton et al., 2000a) would be apparent in the blink and saccade data, respectively, when assessed separately.

The second form of validation employed a split-half method to determine whether corrected waveforms are internally consistent (as an ERP composed of half of the vertical eye movements should resemble an ERP composed of the other half when the split is random). However, a simple split-half test will not be adequate because consistent correction errors will result in similar waveforms (particularly as correction errors may be larger in magnitude than the ERPs themselves), and so a variant must be employed. Specifically, if we append an up to a down eye movement (U/D), and append a down to an up eye movement (D/U), then as discussed above we would expect the underlying neural potentials to be similar for the U/D and D/U waveforms, whereas we would expect the EOG channel to be of opposite polarity. This dissociation of neural and ocular potentials can be seen in Figure 1B. Therefore, by randomly splitting the eye movements in such a way that we have an U/D average based on half the vertical eye movements and a D/U average based on the remaining half, we can determine how effectively the ocular artifact has been removed by determining how similar are the U/D and D/U waveforms.

To determine how similar the waveforms are we first compute a difference wave (for each data point, the U/D voltage minus the

D/U voltage), and then the standard deviation of the data points in the difference wave over the entire epoch. Further to the "split-half" aspect described above, this varies from the "regression" validation technique in that it does not involve the EOG channels in the determination of correction adequacy. We refer to this variant of the split-half reliability method as "standard deviation" validation (STDEV). It is similarly employed for HEM data, but is not appropriate for blink data as there is no blink-opposite to juxtapose with the blink data.

Correction Algorithms

Verleger et al. (1982). A variant of the Verleger et al. (1982) procedure was employed (hereafter varVGM), whereby VEOG and HEOG *B*s were estimated using least-squares multiple regression from the above calibration data (where the EEG channel is the criterion and VEOG and HEOG the predictor variables). The use of the calibration data results in high EOG power and thus accommodates Verleger et al.'s observation that such influences as forward propagation may distort *B*s where EOG power is low. The addition of the HEOG channel is as recommended by Verleger et al. and it has been implemented as a multiple regression because this has been found to reduce distortions to *B*s that may result from VEOG/HEOG covariance (Croft & Barry, 2000c). Further, this procedure results in common *B*s for VEMS and blinks, which is a mark of the Verleger et al. procedure.

Gratton et al. (1983). We employed the 27/09/2001 version of the Gratton et al. (1983) procedure that is distributed through Neuro Scan Labs, without modification.

Semlitsch et al. (1986). We employed the Scan 4.2 version (default ocular artifact correction method) of the SPSA (1986) procedure that is distributed through Neuro Scan Labs, without modification.

Croft & Barry (2000b). We employed the RAAA of Croft and Barry (2000b).

Statistical Analysis

For each of the regression and standard deviation validation techniques (dependent variables *absolute beta* and *standard deviation*, respectively), data were grouped into nine scalp regions: frontal left (mean [Fp1, F7, F3], frontal midline (Fz), frontal right (mean [Fp2, F8, F4]), central left (mean [C3, T3]), central midline (Cz), central right (mean [C4, T4]), posterior left (mean [P3, O1, T5]), posterior midline (Pz), and posterior right (mean [P4, O2, T6]). For each of the eye movement types, repeated-measures ANOVAs tested for relations between the methods (four levels; varVGM, GCD, SPSA, CB), sagittality (frontal, central, posterior; for blink and VEM corrections only), and laterality (left, midline, right; for HEM corrections only), employing the Huynh–Feldt correction for sphericity where appropriate. Repeated-measures contrasts were employed for follow-up analyses, testing for differences between the most accurate and second most accurate procedures, the second and third most accurate, and the third and least accurate procedures, as well as interactions between these differences and topography (polynomial contrasts; sagittality for blink and VEM correction, and laterality for HEM correction). Where differences between two methods were not "significant," methods were still ordered according to relative accuracy. Type I error was controlled in the follow-up analyses using the Bonferroni adjustment ($\alpha = .0167$). Effect sizes will be reported, where significant, for the main effect of method only.

Results

Regression Validation

VEOG data. There was a main effect of method, $F(3,75) = 100.094$, $\varepsilon = .65$, $p < .001$, sagittality, $F(2,50) = 10.845$, $\varepsilon = .71$, $p < .001$, and a Method × Sagittality interaction, $F(6,150) = 6.646$, $\varepsilon = .41$, $p < .001$. Follow-up analyses revealed that CB (first) had lower abs(B)s for VEOG data than GCD (second), $F(1,25) = 99.539$, $p < .001$, $\eta^2 = .80$, and that this difference was greater at posterior sites (relative to frontal), $F(1,25) = 15.524$, $p < .001$. No differences were found between the GCD and varVGM (third) abs(B)s, $F(1,25) = 0.905$, $p = .350$, and this did not interact with sagittality, $F(1,25) < 1.131$, $p > .298$. The varVGM had lower abs(B)s than SPSA (fourth), $F(1,25) = 12.745$, $p < .001$; $\eta^2 = .34$, and there was a trend for this difference to be greater at fronto-posterior sites (relative to central), $F(1,25) = 6.555$, $p < .05$.

HEOG data. There was a main effect of method, $F(3,75) = 328.362$, $\varepsilon = .79$, $p < .001$, laterality, $F(2,50) = 33.191$, $\varepsilon = .81$, $p < .001$, and a Method × Laterality interaction, $F(6,150) = 14.817$, $\varepsilon = .75$, $p < .001$. Follow-up analyses found that CB (first) had lower abs(B)s for HEOG data than varVGM (second), $F(1,25) = 17.405$, $p < .001$, $\eta^2 = .41$, and this difference did not interact with laterality, $F(1,25) < 0.427$, $p > .520$. No differences were found between the VGM and GCD (third) abs(B)s, $F(1,25) = 0.125$, $p = .727$, and this did not interact with laterality, $F(1,25) < 1.978$, $p > .172$. GCD had lower abs(B)s than SPSA (fourth), $F(1,25) = 821.801$, $p < .001$, $\eta^2 = .97$, and this difference was greater at right relative to left hemisphere sites, $F(1,25) = 28.349$, $p < .001$.

Blink data. There was a main effect of method, $F(3,75) = 79.025$, $\varepsilon = .62$, $p < .001$), sagittality, $F(2,50) = 6.060$, $\varepsilon = .82$, $p < .01$, and a Method × Sagittality interaction, $F(6,150) = 5.395$, $\varepsilon = .48$, $p < .01$. Follow-up analyses revealed that CB (first) had lower abs(B)s for blink data than SPSA (second), $F(1,25) = 66.402$, $p < .001$, $\eta^2 = .73$, and that this difference was greater at fronto-posterior sites (relative to central), $F(1,25) = 7.406$, $p < .01$. No overall differences between SPSA and GCD (third) abs(B)s were found, $F(1,25) = 0.097$, $p = .758$, although relative to the GCD, the SPSA procedure had lower abs(B)s at fronto-posterior relative to central sites (where they did not differ), $F(1,25) = 4.321$, $p < .05$. The GCD had lower abs(B)s than varVGM (fourth), $F(1,25) = 43.767$, $p < .001$, $\eta^2 = .64$, but this did not interact with sagittality, $F(1,25) < 2.667$, $p > .115$.

Standard Deviation Validation

VEOG data. There was a main effect of method, $F(3,75) = 39.476$, $\varepsilon = .59$, $p < .001$, sagittality, $F(2,50) = 8.396$, $\varepsilon = .76$, $p < .01$, and a Method × Sagittality interaction, $F(6,150) = 4.120$, $\varepsilon = .33$, $p < .05$. Follow-up analyses found that CB (first) had lower STDEVs for up/down data than GCD (second), $F(1,25) = 37.747$, $p < .001$, $\eta^2 = .60$, and that this difference was greater at fronto-posterior sites (relative to central), $F(1,25) = 14.793$, $p < .001$. No overall difference was found between the GCD and varVGM (third) STDEVs, $F(1,25) = 0.878$,

$p = .358$. However, whereas varVGM had similar STDEVs over the scalp, GCD had lower STDEVs over central relative to fronto-posterior sites, $F(1,25) = 10.749$, $p < .01$. varVGM had lower STDEVs than SPSA (fourth), $F(1,25) = 13.209$, $p < .001$, $\eta^2 = .35$, and this difference was greater at fronto-posterior sites (relative to central) $F(1,25] = 7.323$, $p < .05$.

HEOG data. There was a main effect of method, $F(1.2, 30.1) = 200.059$, $\varepsilon = .40$, $p < .001$, laterality, $F(2,50) = 39.914$, $\varepsilon = .89$, $p < .001$, and a Method × Laterality interaction, $F(1.9, 48.7) = 43.141$, $\varepsilon = .33$, $p < .001$. Follow-up analyses revealed that CB (first) had lower STDEVs for left/right data than GCD (second), $F(1,25) = 8.904$, $p < .01$, $\eta^2 = .26$, and that this difference was larger at lateral sites (relative to midline), $F(1,25) = 21.702$, $p < .001$. No differences were found between the GCD and varVGM (third) corrections, $F(1,25) = 1.164$, $p = .291$, and there was no Method × Laterality interaction, $F(1,25) = 0.360$, $p = .554$. The varVGM procedure had lower STDEVs than the SPSA (fourth) procedure overall, $F(1,25) = 215.389$, $p < .001$, $\eta^2 = .90$, with this difference larger at lateral sites (relative to midline), $F(1,25) = 57.786$, $p < .001$.

Discussion

We will discuss the results from the "regression" and "standard deviation" validity tests together as they were essentially the same. With regard to the correction of HEMs, as predicted, the SPSA procedure resulted in the poorest correction as it does not employ a designated channel for HEM correction (see Figures 2 and 3). The failing of the SPSA correction was more pronounced at lateral sites, which is consistent with lateral sites being the principal region of HEM contamination. The varVGM and GCD procedures performed similarly, and as the primary difference between these (in terms of HEM correction) is that the GCD procedure subtracts an ERP component prior to *B* estimation, this result suggests that this innovation does not appreciably affect the EOG correction process. The CB produced the best HEM correction, with its advantage over the varVGM and GCD procedures probably due to its increased signal-to-noise ratio. The standard deviation test found that this advantage was largest over lateral and frontal sites, which is where HEMs have their largest effect, but also at fronto-posterior relative to central sites.

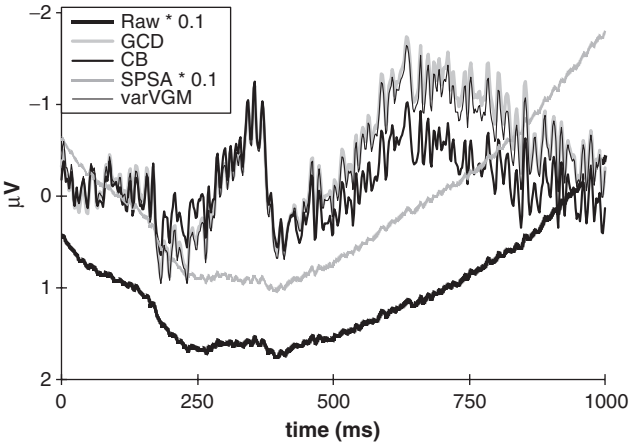

Figure 3. Grand-average raw and the corrected waveforms at F7 for each of the four methods are shown for "left" eye movements. Note that the raw and SPSA traces have been multiplied by 0.1 to allow comparison with the corrected waveforms.

As predicted, VEMs were not corrected as well by the SPSA procedure as the other procedures, particularly at fronto-posterior sites where differences between blink and VEM propagation become more apparent (Croft & Barry, 2000b; Picton et al., 2000b). This is probably because its correction coefficients are based on blinks only (see Figures 4 and 5). The varVGM procedure based its *B*s on a mixture of blinks and VEMs and hence it corrected better than the SPSA procedure. However, although the VGM procedure calculated *B*s from a mixture of blinks and VEMs, it corrected VEMs just as well as the GCD procedure that calculated *B*s from VEMs only. The CB procedure corrected better than the GCD and varVGM procedures, with this advantage greater at frontal and lateral sites. The advantage of this CB can be seen in Figures 4 and 5 where the CB-corrected event-related neural potentials are similar for up and down eye movements, whereas the other procedures' event-related neural

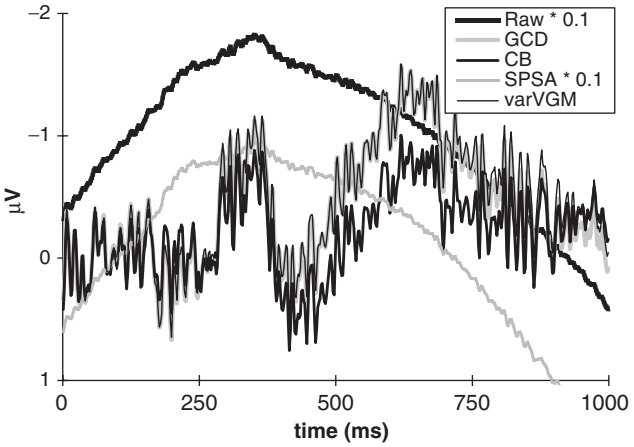

Figure 2. Grand-average raw and corrected waveforms at F7 for each of the four methods are shown for "right" eye movements. Note that the raw and SPSA traces have been multiplied by 0.1 to allow comparison with the corrected waveforms.

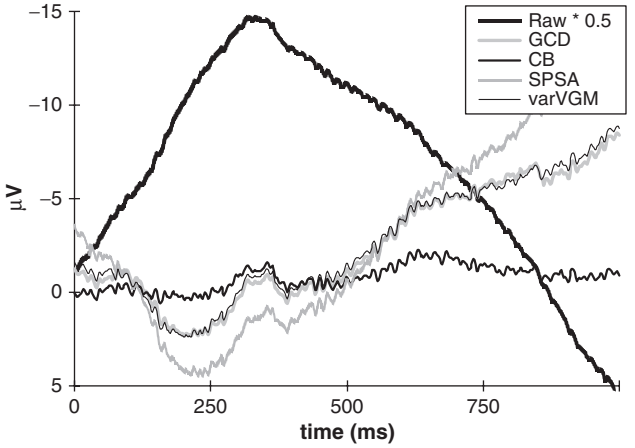

Figure 4. Grand-average raw and corrected waveforms at Fp1 for each of the four methods are shown for "down" eye movements. Note that the raw trace has been halved to allow comparison with the corrected waveforms.

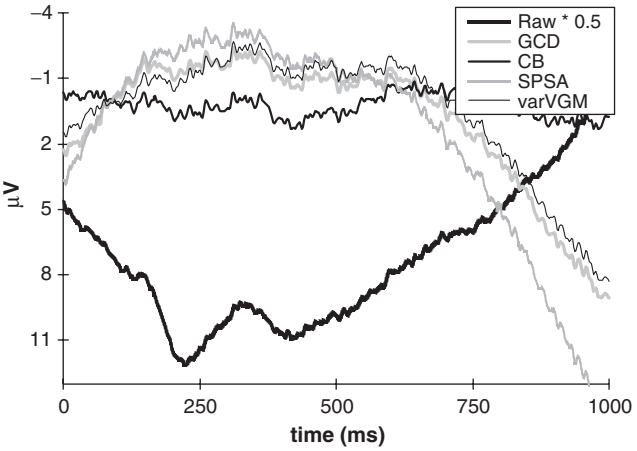

Figure 5. Grand-average raw and corrected waveforms at Fp1 for each of the four methods are shown for "up" eye movements. Note that the raw trace has been halved to allow comparison with the corrected waveforms.

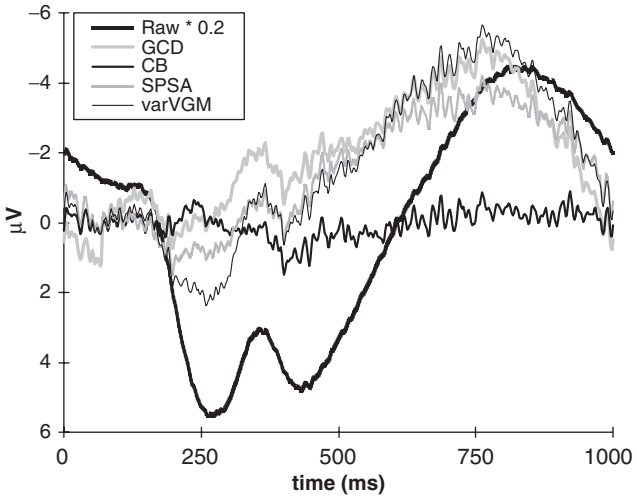

Figure 6. Grand-average raw and the corrected waveforms at Fp1 for each of the four methods are shown for "blinks." Note that the raw trace has been multiplied by 0.2 to allow comparison with the corrected waveforms.

potentials are visibly dependent on the eye movements themselves. Although there are a number of variables that differentiate the CB from the GCD and varVGM procedures, we believe that the signal-to-noise ratio enhancement and the globe rotation/eyelid movement distinction are the primary reasons for the advantage of the CB. It is possible that the CB method of accounting for ERPs in the *B* estimation process may also have improved correction.

In relation to the blink data, the varVGM procedure, the only one that did not calculate correction coefficients for blinks separately, corrected the blink data least well. There was little difference between the SPSA and GCD corrections, and, probably due to the blink correction *B*s being based on blinks only, they both performed better than the varVGM procedure. The signal-to-noise ratio enhancement of the SPSA procedure resulted in slightly better corrections frontally (trend level; $p = .048$), but demonstrated no benefit at other sites. Thus, although it has been shown that *B*s become more accurate when calculated from averaged rather than raw eye movement data (Croft & Barry, 1998a, 1998b), for blinks this improvement may only be important frontally in routine EEG studies. The CB procedure performed substantially better than the SPSA and GCD procedures, particularly at fronto-posterior sites. The only difference between the SPSA and CB procedures (in terms of the blink correction) is that the latter method treats blinks as being composed of both eyelid movement and globe rotation components, and uses different *B*s to correct each of these within each blink. That the advantage of the CB was so large is consistent with dipole analysis findings of different generators for blink and non-blink vertical eye movements (Picton et al., 2000b), and provides indirect support for the separation of eyelid and rotational globe movement in EOG correction methods (Croft, 2000).

The validation measure employed here for the blink data was less thorough than that used for the VEM and HEM data. However, as can be seen in Figure 6, the results are consistent with an intuitive notion of "good correction," with the poorer corrections resulting in waveforms that resembled the blink artifact itself (as can be seen in the raw trace). Another means of providing validation of the procedures is by considering the scalp topographies of corrected blink waveforms. Figure 7 shows such scalp topographies at three time points (corresponding to the three largest EOG deflections in the blink grand-average waveform, at 71 ms, 331 ms, and 651 ms). Similar to the uncorrected waveforms (scaled here by a factor of 0.1 to allow comparison), the varVGM-corrected waveform had the strongest sagittal gradients, the SPSA and GCD corrections had strong sagittal gradients at two of these time points, and the CB corrections showed little resemblance to the uncorrected topography. These illustrations provide strong face validity for the above blink-correction outcomes.

It should be noted that we have not compared EOG-correction methods that employ different correction coefficients for different eye movement frequencies (e.g., Gasser, Sroka, & Möcks, 1985; Whitton, Lue, & Moldofsky, 1978; Woestenburg, Verbaten, & Slangen, 1983). These were excluded because we

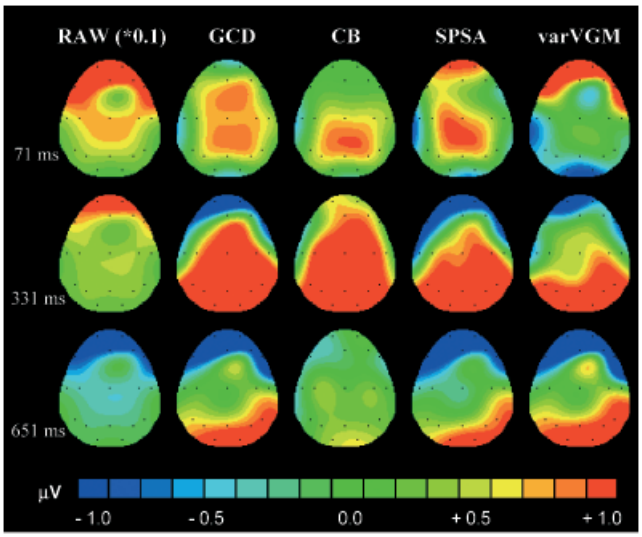

Figure 7. Scalp topographies at three time points of the grand-average blink waveforms are shown, for the raw (scaled by 0.1 to allow comparison) and GCD, CB, SPSA, and varVGM corrected data. The time points were chosen as the three largest deflections of the VEOG channel.

consider that the evidence for frequency-dependent eye movement propagation is not strong (see Croft & Barry, 1998b), and as a thorough test of these procedures is beyond the scope of the present article. A more important omission is that of component analysis solutions to the problem of ocular artifact. Although this class of methods offers a strong means of accounting for ocular artifact and certain of these methods have proven successful empirically (e.g., Lins et al., 1993), their consideration is beyond the scope of the present article. This article thus should be considered as a comparative test of certain EOG correction methods only, and it does not purport to offer comparisons relative to either frequency domain approaches or component analysis solutions. It may be noted, however, that the strength of the CB procedure demonstrated in this article appears in part to be due to the use of separate correction coefficients for eyelid movement and globe rotation. This is consistent with component analysis demonstrations of multiple ocular-artifact sources (Picton et al., 2000b), but it is not as comprehensive as these because the CB procedure does not account for extra nonocular artifact sources, such as those resulting from muscle activity.

It is also important to note that because of the way that the CB algorithm is implemented, it may not be appropriate for certain data sets. That is, as it assumes that an eye movement calibration has been performed (where subjects are instructed to follow a visual cue around a screen; see Croft & Barry, 2000a) and also that radial EOG data has been recorded (usually defined as the average of channels above and below the eye or eyes; e.g., Elbert, Lutzenberger, Rockstroh, & Birbaumer, 1985), where these conditions have not been met the CB cannot be employed. In such cases it is recommended that out of the algorithms tested, the Gratton et al. (1983) should be employed as although it was not as successful as the CB, it performed second best on all three eye-movement types, whereas the SPSA was limited with respect to non-blink vertical and horizontal eye movement correction, and the varVGM was limited with respect to blink correction.

It is concluded that, although each of the ocular artifact correction methods resulted in cleaner data than would be obtained by not correcting at all, there were important differences between the outcomes of the different methods. Specifically, it was found that blinks need to be corrected with *B*s calculated from data containing eyelid movement (such as blinks), that VEMs should be corrected with *B*s calculated from data containing globe rotation (such as VEMs), that HEMs need to be corrected with *B*s calculated from HEMs, that signal-to-noise ratio enhancement in the *B* estimation process improves correction at frontal sites (and possibly more globally), and that separating globe rotation and eyelid movement is preferable to separating VEM and blinks in the correction process. Further, we demonstrated that the CB, which incorporates all of these advantages, corrected HEMs, VEMs, and blinks better than each of the other methods, and it is recommended that this procedure be employed in preference to the VGM, GCD, and SPSA procedures.

REFERENCES

Anthony, B. J. (1985). In the blink of an eye: Implications of reflex modification for information processing. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds). *Advances in psychophysiology* (Vol. 1. pp. 167–218). Greenwich, CT: JAI Press.

Brunia, C. H. M., Möcks, J., van den Berg-Lenssen, M. M. C., Coelho, M., Coles, M. G. H., & Elbert, T., et al. (1989). Correcting ocular artifacts in the EEG: A comparison of several methods. *Journal of Psychophysiology*, *3*, 1–50.

Corby, J. C., & Kopell, B. S. (1972). Differential contributions of blinks and vertical eye movements as artifacts in EEG recording. *Psychophysiology*, *9*, 640–644.

Croft, R. J. (2000). The differential correction of eyelid-movement and globe-rotation artifact from the EEG (reply to Verleger). *Journal of Psychophysiology*, *14*, 207–209.

Croft, R. J., & Barry, R. J. (1998a). EOG correction: A new aligned-artifact average solution. *Electroencephalography and Clinical Neurophysiology*, *107*, 395–401.

Croft, R. J., & Barry, R. J. (1998b). EOG correction: A new perspective. *Electroencephalography and Clinical Neurophysiology*, *107*, 387–394.

Croft, R. J., & Barry, R. J. (2000a). Removal of ocular artifact from the EEG: A review. *Neurophysiologie Clinique/Clinical Neurophysiology*, *30*, 5–19.

Croft, R. J., & Barry, R. J. (2000b). EOG correction of blinks with saccade coefficients: A test and revision of the aligned-artifact average solution. *Clinical Neurophysiology*, *3*, 444–455.

Croft, R. J., & Barry, R. J. (2000c). EOG correction: Which regression should we use? *Psychophysiology*, *37*, 123–125.

Croft, R. J., & Barry, R. J. (2002). Issues relating to the subtraction phase in EOG artefact correction of the EEG. *International Journal of Psychophysiology*, *44*, 187–195.

Elbert, T., Lutzenberger, W., Rockstroh, B., & Birbaumer, N. (1985). Removal of ocular artifacts from the EEG—A biophysical approach to the EOG. *Electroencephalography and Clinical Neurophysiology*, *60*, 455–463.

Gasser, T., Sroka, L., & Möcks, J. (1985). The transfer of EOG activity into the EEG for eyes open and closed. *Electroencephalography and Clinical Neurophysiology*, *61*, 181–193.

Gasser, T., Sroka, L., & Möcks, J. (1986). The correction of EOG artifacts by frequency dependent and independent methods. *Psychophysiology*, *23*, 704–712.

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for the off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–484.

Hillyard, S. A., & Galambos, R. (1970). Eye movement artifact in the CNV. *Electroencephalography and Clinical Neurophysiology*, *28*, 173–182.

Lins, O. G., Picton, T. W., Berg, P., & Scherg, M. (1993). Ocular artifacts in EEG and event related potentials: I. Scalp topography. *Brain Topography*, *6*, 51–63.

Ochoa, C. J., & Polich, J. (2000). P300 and blink instructions. *Clinical Neurophysiology*, *111*, 93–98.

Picton, T. W., van Roon, P., Armilio, M. L., Berg, P., Ille, N., & Scherg, M. (2000a). Blinks, saccades, extraocular muscles and visual evoked potentials (reply to Verleger). *Journal of Psychophysiology*, *14*, 210–217.

Picton, T. W., van Roon, P., Armilio, M. L., Berg, P., Ille, N., & Scherg, M. (2000b). The correction of ocular artifacts: A topographic perspective. *Clinical Neurophysiology*, *111*, 53–65.

Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300. *Psychophysiology*, *23*, 695–703.

van den Berg-Lenssen, M. M. C., van Gisbergen, J. A. M., & Jervis, B. W. (1994). Comparison of two methods for correcting ocular artifacts in EEGs. *Medical & Biological Engineering & Computing*, *32*, 501–511.

Verleger, R. (1991). The instruction to refrain from blinking affects auditory P3 and N1 amplitudes. *Electroencephalography and Clinical Neurophysiology*, *78*, 240–251.

Verleger, R. (1993). Valid identification of blink artifacts: Are they larger than 50 μV in EEG records? *Electroencephalography and Clinical Neurophysiology*, *87*, 354–363.

Verleger, R. (2000). Should we use different estimates of correcting EEG artifacts produced by blinks and saccades? *Journal of Psychophysiology*, *14*, 204–206.

Verleger, R., Gasser, T., & Möcks, J. (1982). Correction of EOG artifacts in event-related potentials of the EEG: Aspects of reliability and validity. *Psychophysiology*, *19*, 472–480.

Vigario, R. N. (1997). Extraction of ocular artifacts from EEG using independent component analysis. *Electroencephalography and Clinical Neurophysiology*, *103*, 395–404.

Weerts, T. C., & Lang, P. J. (1973). The effects of eye fixation and stimulus response location on the contingent negative variation (CNV). *Biological Psychology*, *1*, 1–19.

Whitton, J. L., Lue, F., & Moldofsky (1978). H. A spectral method for removing eye movement artifacts from the EEG. *Electroencephalography and Clinical Neurophysiology*, *44*, 735–741.

Woestenburg, J. C., Verbaten, M. N., & Slangen, J. L. (1983). The removal of eye-movement related artifact from the EEG by regression analysis in the frequency domain. *Biological Psychology*, *16*, 127–147.

(Received April 8, 2002; Accepted September 5, 2003)