A New Quantitative Measure for Monitoring Somatosensory Evoked Potentials

Richard J. Moulton, Stefan J. Konasiewicz and Paul O'Connor

Abstract: This paper describes the development and testing of a computer algorithm to automate the process of peak identification and somatosensory evoked potential (SSEP) grading. We tested the accuracy of computerized peak detection and evaluated grading schemes using a test set of 60 SSEPs ranked from worst to best by the programmer (RJM) and a blinded grader (PO). The computer algorithm recognized 95% of peaks identified by visual inspection. Twelve percent of peaks identified by the computer were noise. Summed peak to peak amplitude gave the most accurate ranking of SSEPs. Rank correlation between computer and blinded and unblinded expert grading was $r = .82$ for PO, $r = .92$ for RJM, $p < .0001$ for both. Computer and manually summed amplitudes were highly correlated (Pearson $r = .98$, $p < .0001$). Correlation between the 2 expert graders was .86, $p < .0001$. Computer graded SSEPs were significantly related to clinical outcome at 3 months, $p < .0001$. Automatic grading of SSEPs using summed peak to peak amplitude is highly correlated with expert grading. The measure is objective, continuous, and well suited to statistical analysis.

Résumé: Une nouvelle méthode quantitative pour évaluer les potentiels évoques somesthesiques. Dans cet article, nous décrivons le développement et l'évaluation d'un algorithme informatique pour automatiser l'identification des pics et la classification des potentiels évoqués somesthésiques (PES). Nous avons étudié la précision de la détection informatisée des pics et la cote de classification au moyen de 60 PES classifiés du plus mauvais au meilleur par le programmeur (RJM) et un évaluateur travaillant en aveugle (PO). L'algorithme a reconnu 95% des pics identifiés par inspection visuelle. Douze pourcent des pics identifiés par l'ordinateur étaient des artefacts. La sommation de l'amplitude d'un pic à l'autre donnait l'ordre le plus précis des PES. La corrélation entre la classification de l'ordinateur et celle des experts, en aveugle ou non, était de $r = .82$ pour PO, $r = .92$ pour RJM, $p < .0001$ pour les deux. Les amplitudes calculées par ordinateur ou manuellement étaient hautement corrélées (Pearson $r = .98$, $p < .0001$). La corrélation entre les 2 évaluateurs experts était de .86, $p < .0001$. Les PES évalués par ordinateur étaient significativement reliés à l'issue clinique à 3 mois, $p < .0001$. L'évaluation automatique des PES au moyen du calcul de l'amplitude d'un pic à l'autre est hautement corrélée à l'évaluation par des experts. La mesure est objective, continue et se prête bien à l'analyse statistique.

Can. J. Neurol. Sci. 1994; 21: S17-S22

Evoked potential studies have been shown to be well correlated with outcome from severe closed head injury.[1,2,3] Somatosensory evoked potentials (SSEPs) are the single best predictor of outcome.[3] Our own work has shown that it is feasible to monitor SSEPs at hourly intervals for up to several days at a time. With computer automation, a monitoring system can be designed to run without operator intervention for up to 12 hours, with the collection of high quality, artefact-free tracings.[4] In our intitial group of 36 patients, survival and death or vegetative survival at 3 months were correlated to the longest latency peak in the SSEP at the conclusion of monitoring. Approximately 2/3 of patients in whom monitoring was begun within 24 hours of injury, and who died or survived in a vegetative state, lost evoked potential activity progressively from the time of initiation of monitoring. The time course for this deterioration varied from 12 hours to 4-5 days post-injury.[4] We were surprised at the frequency of deteriorating evoked potentials, as this had not been apparent in other studies using single, 2, or 3 isolated evoked potential studies, usually to determine prognosis. Newlon et al. found instances of deteriorating evoked potentials when repeated measurements were conducted, and attributed these changes to "secondary insults" such as raised ICP, delayed hematomas, hypoxia, etc.[5]

The Divisions of Neurosurgery (R.J.M., S.J.K.) and Neurology (P.O.), St. Michael's Hospital, Toronto.
Reprint requests to: Dr. R. Moulton, 38 Shuter Street, Toronto, Ontario, Canada M5B 1A6

As evoked potential measurement evolves from single studies for the purpose of determining prognosis to repetitive studies at short intervals to monitor patient progress or correlate cerebral function with other physiologic parameters (e.g., intracranial pressure (ICP)), some form of data simplification/compression becomes desirable.[6] Statistical comparison of SSEPs mandates some form of data classification or feature extraction. As a result several grading schemes for SSEPs have evolved, based on the central conduction time,[7,8] the number of peaks in the SSEP,[2,3] or a more subjective evaluation of the peak latencies and amplitudes in the SSEP.[9] To be most useful for long-term monitoring (i.e., over several days), an ideal grading scheme for SSEPs should lend itself to computer automation, should include information for the entire duration of the SSEP (250 msec. in our case), should be objective, and should approximate a smooth function over the entire range of SSEPs seen in the severely head-injured population. This would allow trending of SSEP measurements over time and would facilitate statistical comparisons within and among patients.

This paper describes the development and validation of a method of automatically identifying peaks in the SSEP tracing, and automatically grading SSEPs based on the input from the peak detection programme. Summed peak to peak amplitude, beginning at the P_{15}-N_{20} complex is the extracted feature used to trend SSEPs over time and for statistical analysis.

Materials and Methods

All severely head injured patients (GCS $\leqq$ 8) who are not moribund (i.e., GCS = 3 with unreactive pupils) are admitted to our neurosurgical intensive care unit and undergo a standard intensive monitoring and treatment protocol. This involves mechanical ventilation, sedation, pharmacologic paralysis, continuous ICP monitoring, continuous monitoring of SSEPs, brain stem auditory evoked potentials (BAEPs), the electroencephalographic power spectrum, intermittent measurement (12 hourly) of transcranial oxygen extraction (AVDO2), and cerebral blood flow (CBF) measurement using the nitrous oxide clearance technique.

The SSEP monitoring paradigm has been described in detail in previous work.[4] Briefly, median nerve SSEPs are collected automatically at approximately hourly intervals. Seventy-three patients were monitored in this fashion prior to this study. Each SSEP tracing comprises 250 averages of 250 msec. duration. Analysis is carried out on the tracing from the contralateral C3' or C4' electrode positions referenced to linked ears.

A peak detection programme was written for use on a personal computer. The algorithm consists of the iterative determination of maxima and minima (negative and positive peaks) in the digitized evoked potential data array over successively shorter intervals within the array (Figure 1a and b). Once a peak is identified the interval between the newly identified peak and the previous peak is searched for new local minima and maxima. The interval at which the search for peaks is stopped varies according to the position within the SSEP trace, with shorter time periods employed in the early post-stimulus period. The time intervals correspond to frequencies above which one would not reasonably expect non-artefactual activity. This frequency decreases from 1000 Hz. at the start of the SSEP trace to 20 Hz. at the end of the trace. Peaks thus identified are then

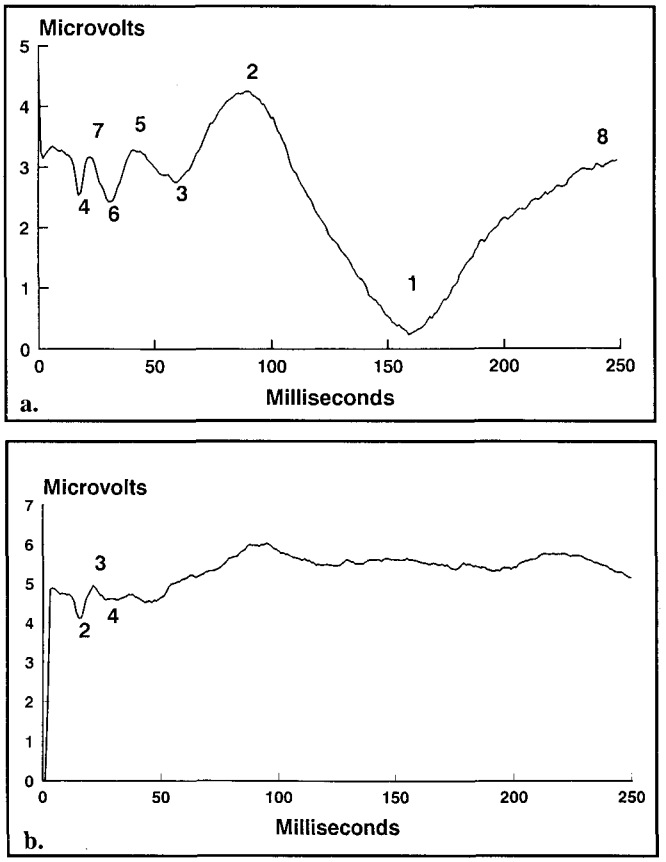

Figure 1: *Peak detection algorithm in intact* ***(a.)*** *and severely attenuated* ***(b.)*** *SSEPs. The numbers opposite the peaks indicate the order in which they were selected by the peak detection algorithm. In Figure* ***1b*** *the first deflection chosen was negative near 100 msec. It was subsequently rejected as noise by the slope and amplitude rules, so that peak 1 does not appear in the final output of the programme. Total amplitude is the sum of the absolute values of (peak 2 - peak 3) + (peak 4 - peak 3). Total amplitude in* ***1a*** *is the sum of the absolute values of (peak 4 - peak 7) + (peak 6 - peak 7) + (peak 6 - peak 5) + ... + (peak 1 - peak 8).*

screened and accepted or rejected based on slope (minimum .08 μV/msec.) and amplitude (minimum .25 μV.) criteria. Finally, no later peaks are selected if the P_{15} - N_{20} complex cannot be identified within the initial portion of the tracing. The frequency, slope, and amplitude criteria were empirically derived and refined in order to minimize errors of peak identification.

In order to test the peak detection and grading algorithm single tracings from 60 different patients were selected. These were not randomly selected, but rather were chosen to represent the full spectrum of activity seen in the course of monitoring these patients (Figure 1a and b). One blinded (PO) and one unblinded grader (RJM) ranked the SSEP tracings from worst to best. The criteria for ranking the peaks included the number, latency, amplitude, and morphology of the peaks in the SSEP. Peaks were initially identified by visual inspection (RJM) and features of interest (see below) were extracted from the peak amplitude and latency data and correlated with the worst to best rankings of both graders using Spearman rank correlation. The feature with the best correlation coefficient was then implemented on the computer with input from the peak detection programme.

https://doi.org/10.1017/S0084255900007567 Published online by Cambridge University Press

The correlation of this measure was checked against the rankings of the blinded and unblinded graders. As a final check on the validity of the method, the results of the peak detection and grading algorithms were applied to the full data set of 73 patients and compared with their Glasgow Outcome Scale[10] at 3 months using univariate ANOVA. Outcome was compressed into 3 categories (good/moderate, severe, and dead/vegetative). All statistical calculations were done using the SAS* software package.

Results

The peak detection programme performed best in high signal to noise conditions. In low signal conditions there was a tendency to misidentify noise as SSEP peaks. Table 1 shows the error counts when peaks identified manually were compared to those identified by the computer. A total of 347 peaks, in 60 SSEP tracings from different patients, were identified and labelled by visual inspection. The computer algorithm made a total of 57 errors (17%) in 30 patients. There were no errors in 30 patients. The algorithm failed to detect 17 peaks (5%) that were identified on visual inspection. Forty peaks (12%) identified by the computer were considered to be noise on visual inspection. These errors tended to occur in patients with severely abnormal evoked potentials. The algorithm was purposely adjusted to err slightly on the side of over-identification of peaks for reasons to be discussed subsequently. Peaks were identified as such on visual inspection if there was a well-defined peak or trough at an appropriate latency in the SSEP. There was clearly some degree of subjectivity in this method, but we felt that it was a fair representation of the "real world" situation in which patients' SSEP morphology is frequently changing, and in which there may already be severe abnormalities in SSEP morphology at the time of baseline measurement. These factors render the use of replicative studies to verify peaks difficult or impossible in this clinical context.

In order to select a grading algorithm we looked at 4 methods of grading SSEPs based on simple feature extraction: the number of peaks, the latency of the final peak in the SSEP, the central conduction time, and the sum of the absolute values of the peak to peak amplitudes of all the peaks in the SSEP beginning at the P_{15} and extending to the last identified peak in the tracing (Figure 1a and b). If only the P_{15} is present then the summed amplitude is 0. Table 2 shows the correlations between waves ranked by those 4 grading methods, and the same set of waves ranked by the programmer (RJM) and a blinded grader (PO). Summed amplitude gave the highest correlation with expert ranking of the SSEP traces ($r = .97$ for RJM, $r = .84$ for PO, $p < .0001$ for both). Central conduction time performed most poorly ($r = -.62$ for RJM, $r = -.61$ for PO, $p < .0001$ for both).

The summed amplitude algorithm was programmed and the peak detection and summed amplitude grading programs were run on the test data. Table 3 shows the correlations between subjective rankings by the programmer, the blinded expert grader, and between manually and computer summed tracings from the test data. There is excellent correlation between the manually summed and computer summed amplitudes ($r = .98$, $p < .0001$, Figure 2), and between the latter and the unblinded grader/programmer. The correlation is slightly less between the blinded grader and the computer summed amplitudes, but the magnitude of correlation is of the same order as that between the rankings achieved by the blinded and unblinded graders.

As a final check on the validity of the computer peak detection and grading programmes, these were run on the full data set of all waves from 73 patients monitored up until the time of this study and compared with 3 outcome categories based on the Glasgow Outcome Scale at 3 months. Figures 2a and b show the summed amplitude means for each of the clinical outcome groups for the best, worst, and both hemispheres at the start and end of monitoring. These are significantly different at the end of monitoring in each of the outcome categories for the best, worst, and both hemispheres ($F = 11.5$, 18.9, and 15.3 respectively, $p < .0001$ for all three). The differences in total amplitude were much less at the start of monitoring (achieving statistical significance only between the best and worst outcome groups for the worst hemisphere, $p = .032$), confirming our earlier obervation of significant change in evoked potential activity over the course of monitoring in a substantial proportion of patients.[4] The usual duration of monitoring was 3-5 days with a range of 1-10 days. The reasons for this progressive deterioration

Table 1. Peak Detection Programme Accuracy.

	Number of Peaks	Percent	Number of Patients
Missed Peaks	17/347	5	11/60
Noise Identified as a Peak	40/347	12	25/60
No Error			30/60

A total of 347 peaks in 60 SSEP tracings from different patients were identified by visual inspection of the SSEP waveforms.

Table 2. Grading Algorithm Correlations.

	r (RJM)	p value	r (PO)	p
Central Conduction Time	–.62	.0001	–.61	.0001
Number of Peaks	.86	.0001	.79	.0001
Latency of Final Peak	.87	.0001	.80	.0001
Summed Peak to Peak Amplitude	.97	.0001	.84	.0001

Spearman rank correlations were made between tracings ranked subjectively by unblinded (RJM) and blinded graders (PO) and the grading schemes listed above using visually identified peaks as the input to the grading paradigms.

Table 3. Correlations with Computer-Generated Total Amplitude.

	r	p value
Manual vs. Computer Total Amplitude	.98*	.0001
Computer vs. Unblinded Grader	.92	.0001
Computer vs. Blinded Grader (PO)	.82	.0001
Blinded (PO) vs. Unblinded Grader (RJM)	.86	.0001

* Pearson correlation coefficient. All others are Spearmann rank correlations.

*SAS Institute Inc., Box 8000, Cary, North Carolina 27511-8000.

https://doi.org/10.1017/S0084255900007567 Published online by Cambridge University Press

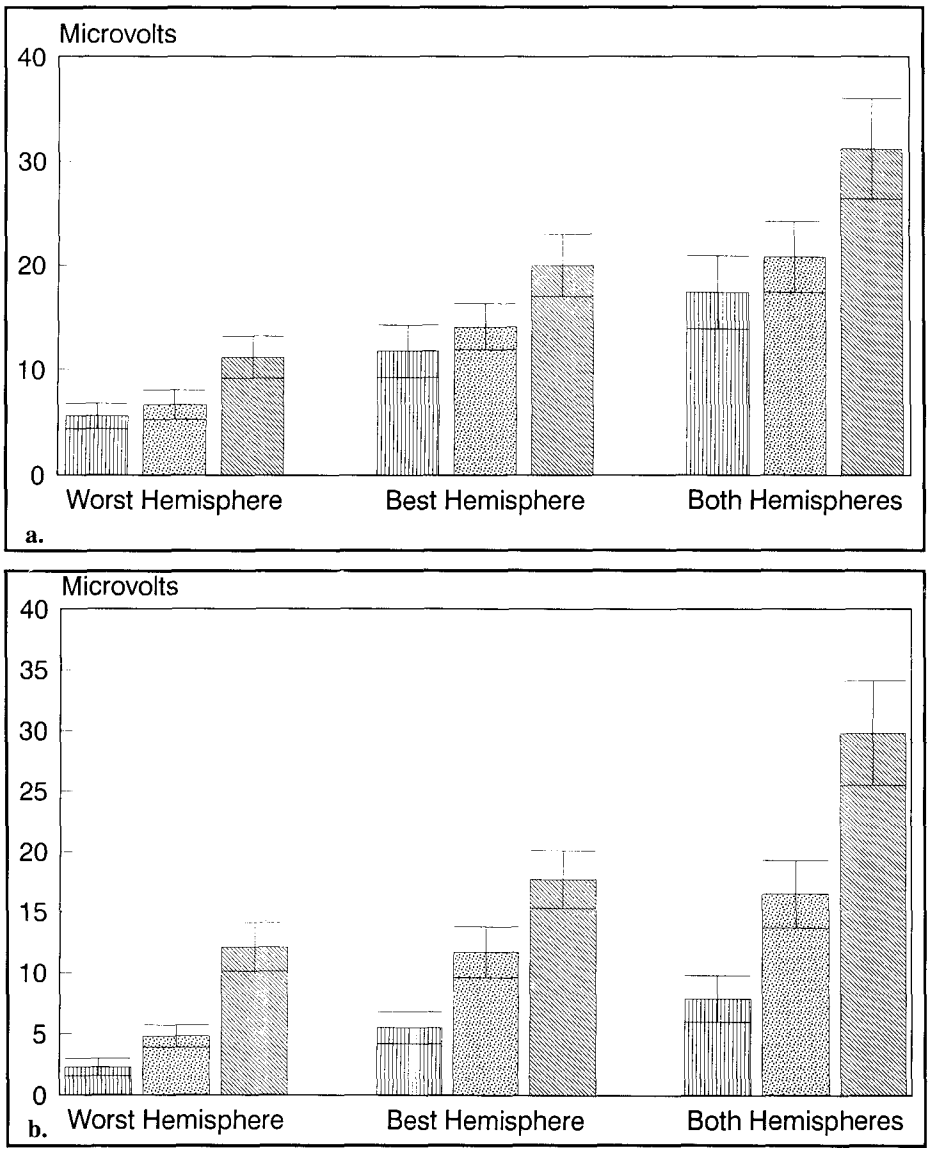

Figure 2: *Summed amplitudes for the start* ***(a)*** *and end of monitoring* ***(b)*** *in 3 clinical outcome groups: dead/vegetative (vertical stripe), severely disabled (stippled), and good outcome/moderate disability (diagonal stripe). The error bars indicate the standard errors of the means. At the end of monitoring there are significant ($p < .0001$) differences between all 3 outcome groups in the best, worst, and both hemispheres. At the start of monitoring there is a much smaller difference ($p < .02$) in the worst hemisphere only between patients with good outcome/moderate disability and vegetative/dead patients.*

in SSEP activity in the early hours after severe head injury, including the role of raised intracranial pressure, are the subject of ongoing research.

Discussion

Monitoring of SSEPs and BAEPs over periods of hours to days has been shown to be technically feasible and to produce clinically useful data.[4,11] The large volumes of data thus generated were the impetus for the development of an automated system of SSEP analysis. Automated and semi-automated peak detection and grading schemes have been previously implemented using BAEPs.[11,12] Our monitoring efforts are focussed largely on SSEPs because of their greater utility in head injury.[3,7] Various schemes have been employed to grade SSEPs, thereby simplifying the process of making within and between patient comparisons. For the most part these rely on subjective rating of the number, amplitude, morphology, and latency of peaks in the SSEP tracing.[9,13] Greenberg et al.,[9] and later Newlon[14] employed a four point grading scheme. In earlier work we found that the longest latency activity present in the SSEP correlated well with outcome.[4] Other workers had described similar findings previously.[2] Lindsay et al. have described a simple grading scheme based on the number of peaks in the SSEP.[3] The latter techniques are considerably more objective and simpler to implement than more complex and subjective grading schemes. There is likely significant intercorrelation between the various features used to grade SSEPs (peak latency, amplitude, and morphology), thereby explaining the clinical validity of all these approaches.

https://doi.org/10.1017/S0084255900007567 Published online by Cambridge University Press

However, none of these techniques approximates a continuous function, particularly the subjective grading schemes that employ a relatively small number of grades. Changes in peak amplitudes short of complete disappearance may be of physiologic significance, but cannot be expected to be reliably shown by any of these methods. Hume and Cant,[8] and later Lindsay et al.[7] described the use of central conduction time to determine prognosis following severe head injury. Although central conduction time has the advantage of being a continuous, objective measure with relatively little within and between patient variablity, it may not be sufficiently sensitive to the earliest changes in the SSEP, i.e., those that tend to occur in intermediate and long latency peaks (> 30 msec. post stimulus) and initially consist of changes in amplitude only. Our data show that central conduction time correlated least well with expert ranking of a test set of evoked potentials, presumably in part due to lack of sensitivity to changes in intermediate and long latency peaks. The particular SSEP measurement paradigm (250 msec. recording duration) may have adversely affected accuracy to some extent, but was probably not the major reason for the poor correlation. Compared with the measures listed above, we found that the summed peak to peak amplitude in the SSEP correlated best with SSEPs ranked by expert graders. We did not test the various subjective grading schemes described in the literature as these are extremely difficult to automate compared with simpler feature extraction techniques.

Grading SSEPs becomes more difficult when one is relying on a computer for identification of peak amplitudes and latencies to be used as input to the grading algorithm. Misinterpretation of noise as signal is bound to occur to some extent, particularly in conditions of severely attenuated signal (i.e., a patient with very low amplitude or absent SSEPs). Grading schemes which rely very heavily on the identification of certain peaks (subjective grading schemes, central conduction time) are bound to be more prone to error in this circumstance, as are peak counts and grading based on the last peak in the tracing. Summing the peak to peak amplitude over the entire 250 msec. trace has the advantages of not being dependent on identification of particular peaks, and of minimizing the effects of small amounts of noise. Noise is generally of low amplitude compared to genuine peaks, and therefore contributes relatively little to total amplitude. The algorithm was purposely adjusted to miss fewer real peaks, in favour of the inclusion of some low amplitude noise. This approach is borne out by a comparison of the raw error counts with the correlation between amplitudes summed manually and by computer. Although there is an overall error rate of 17% with respect to identifying particular peaks, primarily as a result of identifying noise as SSEP peaks, the impact of this on the summed amplitude is negligible, as shown by the nearly perfect correlation between manually and computer summed amplitudes (r = .98). Excellent correlation was seen between computer summed amplitude and peaks ranked by blinded and unblinded experts. Although there was some discrepancy in the correlations between the unblinded and blinded graders and computer grading (r = .92 vs. r = .82), the latter correlation coefficient was of approximately the same magnitude as the correlation between the blinded and unblinded graders (r = .86).

The summed amplitude algorithm relies heavily on amplitude components from later peaks in the SSEP. The increased weighting of these peaks may be justified in light of recent work showing that the presence of long latency activity (i.e., N_{70}) in the SSEP was the best predictor of ***good quality*** survival following cardiac arrest.[15] It is also true that the later peaks are the most variable peaks both within and between patients.[16] Further work will be necessary to characterize the random variability over time, and to characterize the degree of change typically associated with clinically meaningful change. As a first approximation, a fall in amplitude summed from both hemispheres of 20% or greater from the baseline value, was associated with severe disability or worse outcome at 3 months in our patients. The use of summed amplitude to reduce the volume of data and to facilitate the trending of data over long periods of time (typically 2 - 10 days) necessarily involves some loss of information, as does any data simplification/compression technique. Raw tracings are stored on disk at the time of recording so that no data is lost, and individual patients' tracings can be reviewed and analyzed with respect to the presence or absence of particular peaks. The automated grading system was devised primarily to allow objective statistical analysis of trends over time, within and among patients, and for correlation of SSEP data with other monitored parameters, a number of which are the subjects ongoing studies in our unit, particularly in the realms of cerebral blood flow and metabolism, and ICP management.

Computer peak detection and SSEP grading based on a summed peak to peak amplitude algorithm is well correlated with subjective expert grading and clinical outcome in a group of severely head-injured patients. The technique allows rapid analysis of the large amounts of data that are typical in a long-term monitoring application. Trends can rapidly be discerned and the technique is well-suited to statistical analysis.

Acknowlegdements

This work was supported in part by a grant from the St. Michael's Hospital Research Society.

References

1. Rappaport M, Hopkins HK, Hall K, Belleza T. Evoked potentials and head injury. 2. Clinical applications. Clin Electroencephalogr 1981; 12: 167-176.
2. De La Torre JC, Trimble JL, Beard RT, Hanlon K, Surgeon JW. Somatosensory evoked potentials for the prognosis of coma in humans. Exp Neurol 1978; 60: 304-317.
3. Lindsay KW, Carlin J, Kennedy I, et al. Evoked potentials in severe head injury – analysis and relation to outcome. J Neurol Neurosurg Psychiatry 1981; 44: 796-802.
4. Moulton R, Kresta P, Ramirez M, Tucker W. Continuous automated monitoring of somatosensory evoked potentials in posttraumatic coma. J Trauma 1991; 31: 676-685.
5. Newlon PG, Greenberg RP, Hyatt MS, Enas G, Becker DP. The dynamics of neuronal dysfunction and recovery following severe head injury assessed with serial multimodality evoked potentials. J Neurosurg 1982; 57: 168-177.
6. Bertrand O, Bohorquez J, Pernier J. Technical requirements for evoked potential monitoring in the Intensive Care Unit. *In*: Rossini PM, Maugiere F, eds. New Trends and Advanced Techniques in Clinical Neurophysiology (EEG Suppl. 41). Elsevier Science Publishers, 1990: 51-70.
7. Lindsay KW, Pasaoglu A, Hirst D, et al. Somatosensory and auditory brain stem conduction after head injury: a comparison with clinical features in prediction of outcome. Neurosurgery 1990; 26: 278-285.
8. Hume A, Cant BR. Central somatosensory conduction time after head injury. Ann Neurol 1981; 10: 411-419.

9. Greenberg RP, Mayer DJ, Becker DP, Miller JD. Evaluation of brain function in severe human head trauma with multimodality evoked potentials. Part 1: Evoked brain-injury potentials, methods, and analysis. J Neurosurg 1977; 47: 150-162.
10. Jennett B, Bond M. Assessment of outcome after severe brain damage: a practical scale. Lancet 1975; 1: 480-484.
11. Bertrand O, Garcia-Larria L, Artu F, Maugiere F, Pernier J. Brainstem monitoring. I. A system for high-rate sequential BAEP recording and feature extraction. Electroencephalogr Clin Neurophysiol 1987; 68: 433-445.
12. Pool KD, Finitzo T. Evaluation of a computer-automated program for clinical assessment of the auditory brainstem response. Ear Hear 1989; 10: 304-310.
13. Rappaport M, Hall K, Hopkins K, and Belleza T. Evoked potentials and Head Injury. 1. Rating of Evoked Potential Abnormality. Clin Electroencephalogr 1981; 12: 154-166.
14. Newlon PG. Utility of multimodality evoked potentials in cerebral injury. *In*: Aminoff MJ, ed. Neurologic Clinics. (Vol 3. No. 3). Philadelphia: W. B. Saunders, 1985: 675-686.
15. Madl C, Grimm G, Kramer L, et al. Early prediction of individual outcome after cardiopulmonary resuscitation. Lancet 1993; 341: 855-858.
16. Yamada T. The anatomic and physiologic bases of median nerve somatosensory evoked potentials. *In*: Gilmore R, ed. Neurologic Clinics (Vol. 6 No. 4). Philadelphia: W. B. Saunders, 1988: 705-733.

https://doi.org/10.1017/S0084255900007567 Published online by Cambridge University Press