Reliability and agreement of manual and automatic methods for single trial detection and estimation of somatosensory evoked potential features

J. A. Biurrun Manresa, F. G. Arguissain, D. E. Medina Redondo, C. D. Mørch and O. K. Andersen

SEVERAL methods have been proposed in the past for automatic detection and estimation of somatosensory evoked potential (SEP) features [1]. However, the agreement (presence/absence of SEP features) and reliability (measurement error) of these methods compared to the current gold standard (manual detection) has not been properly established. Thus, the aim of this study is to assess the agreement and reliability of manual and automatic single trial feature detection methods.

SEPs were recorded from the vertex of 15 healthy volunteers, in response to electrical stimulation of the plantar nerve. A total of 120 trials were recorded for each subject. Selected features (amplitudes and latencies from N1, N2 and P2 peaks) were extracted from each trial, using three methods: manual detection, an automatic algorithm based on the derivative of the signal and classification of features using fuzzy logic (Fuzzy-Deriv), and another automatic algorithm based on wavelet filtering and multiple linear regression (WF-MLR) [1]. Agreement was assessed as overall agreement (p_o), agreement by chance (p_e), positive agreement (p_{pos}), negative agreement (p_{neg}) and Cohen's κ [2]. Absolute and relative reliability were assessed by the coefficient of variation (CV) and the intraclass correlation coefficient (ICC), respectively [3].

Results are shown in Table 1. In general, there is low agreement between methods (low κ), since most of the agreement appears to happen by chance (p_e is very close to p_o). Regarding reliability, there is a wide variation across the different features and methods, although in general variation is smaller in latencies compared to amplitudes.

In conclusion, these results indicate that the interpretation of SEP features must be performed cautiously, since the outcome features (peak amplitude and latencies) are strongly dependent on the detection method used.

Agreement	p_o (%)	p_e (%)	p_{pos} (%)	p_{neg} (%)	κ
Manual vs Fuzzy-Deriv	85.7 (23.5 - 99.2)	78.2 (25.7 - 99.2)	92.3 (0 - 99.6)	42.1 (0 - 87.7)	0.2 (-0.1 - 0.8)
Manual vs WF-MLR	88.3 (1.7 - 100)	88.3 (1.7 - 100)	92.8 (0 - 99.6)	0 (0 - 100)	0 (-0.3 - 0.1)
Fuzzy-Deriv vs WF-MLR	93.3 (0 - 100)	93.3 (0 - 100)	85.0 (0 - 99.6)	0 (0 - 100)	0 (-0.2 - 0.1)

Absolute reliability (CV)	N1 amp (%)	N1 lat (%)	N2 amp (%)	N2 lat (%)	P2 amp (%)	P2 lat (%)
Manual vs Fuzzy-Deriv	2.3 (0.3 - 51.5)	2.9 (0.7 - 39.1)	1.4 (0.2 - 5.0)	2.4 (0.5 - 4.2)	3.1 (0.6 - 27.4)	2.7 (0.7 - 16.5)
Manual vs WF-MLR	33.1 (13.6 - 733.0)	8.9 (3.0 - 38.8)	22.0 (12.2 - 39.0)	7.2 (3.8 - 33.6)	15.2 (3.0 - 107.4)	6.2 (2.9 - 18.2)
Fuzzy-Deriv vs WF-MLR	66.8 (13.8 - 1612.1)	10.9 (4.0 - 40.1)	23.9 (6.3 - 434.9)	11.5 (4.5 - 34.3)	15.0 (1.8 - 59.7)	6.4 (2.9 - 15.5)

Relative Reliability (ICC)	N1 amp	N1 lat	N2 amp	N2 lat	P2 amp	P2 lat
Manual vs Fuzzy-Deriv	0.98 (0.15 - 1)	0.50 (0 - 0.93)	0.99 (0.87 - 1)	0.75 (0.40 - 1)	0.97 (0.46 - 1)	0.62 (0.02 - 0.88)
Manual vs WF-MLR	0.72 (0.04 - 0.93)	0.26 (0 - 0.60)	0.62 (0.28 - 0.87)	0.16 (0 - 0.45)	0.78 (0.02 - 0.87)	0.20 (0- 0.52)
Fuzzy-Deriv vs WF-MLR	0.78 (0.02 - 0.92)	0.14 (0 - 0.50)	0.67 (0.08 - 0.92)	0.03 (0 - 0.42)	0.78 (0 - 0.87)	0.23 (0 - 0.53)

Table 1: Agreement and reliability among the three methods. Median values (range) across subjects are reported.

REFERENCES

[1] Hu L, Mouraux A, Hu Y, Iannetti GD, "A novel approach for enhancing the signal-to-noise ratio and detecting automatically event-related potentials (ERPs) in single trials", Neuroimage 2010;50(1):99-111.

[2] Kundel HL, Polansky M, "Measurement of observer agreement", Radiology 2003;228(2):303-8.

[3] Atkinson G, Nevill AM, "Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine", Sports Med 1998, 26(4):217-238.

This work was supported by The Danish Council for Independent Research | Technology and Production Sciences (FTP), grant number: 11-116749.

José A. Biurrun Manresa, Federico G. Arguissain, Carsten D. Mørch and Ole K. Andersen are with the Center for Sensory-Motor Interaction, Dept. of Health Science and Technology, Aalborg University, 9220 Aalborg Øst, Denmark (phone: +45 9940 8715; fax: +45 9815 4008; e-mails: jbiurrun@hst.aau.dk, fga@hst.aau.dk, cdahl@hst.aau.dk, oka@hst.aau.dk).

Enzo D. Medina Redondo is with the Dept. of Informatics, Faculty of Engineering, National University of Entre Ríos, 3101 Oro Verde, Entre Ríos, Argentina (e-mail: dmedina@bioingenieria.edu.ar).