Clinical Neurophysiology 129 (2018) 1170–1181

Contents lists available at ScienceDirect

Clinical Neurophysiology

journal homepage: www.elsevier.com/locate/clinph

Sliding window averaging in normal and pathological motor unit action potential trains

Armando Malanda-Trigueros [a,*], Javier Navallas [a], Javier Rodriguez-Falces [a], Ignacio Rodriguez-Carreño [b], Sonia Porta [a], Miguel Fernández-Martínez [a], Luis Gila [c]

[a] Electrical and Electronics Engineering Dept., Public University of Navarre, Spain
[b] Economics Dept., University of Navarre, Spain
[c] Neurophysiology Service, Navarre Hospital Complex, Spain

ARTICLE INFO

Article history:
Accepted 25 February 2018
Available online 19 March 2018

Keywords:
Electromyography
Motor unit action potential
Averaging
Sliding-window
MUAP waveform

HIGHLIGHTS

- A new sliding window algorithm for averaging trains of MUAPs has been tested.
- It performed better than relevant averaging algorithms with normal, myopathic and neurogenic signals.
- The algorithm can be of service for the quantitative analysis of MUAP waveforms.

ABSTRACT

Objective: To evaluate the performance of a recently proposed motor unit action potential (MUAP) averaging method based on a sliding window, and compare it with relevant published methods in normal and pathological muscles.
Methods: Three versions of the method (with different window lengths) were compared to three relevant published methods in terms of signal analysis-based merit figures and MUAP waveform parameters used in the clinical practice. 218 MUAP trains recorded from normal, myopathic, subacute neurogenic and chronic neurogenic muscles were analysed. Percentage scores of the cases in which the methods obtained the best performance or a performance not significantly worse than the best were computed.
Results: For signal processing figures of merit, the three versions of the new method performed better (with scores of 100, 86.6 and 66.7%) than the other three methods (66.7, 25 and 0%, respectively). In terms of MUAP waveform parameters, the new method also performed better (100, 95.8 and 91.7%) than the other methods (83.3, 37.5 and 25%).
Conclusions: For the types of normal and pathological muscle studied, the sliding window approach extracted more accurate and reliable MUAP curves than other existing methods.
Significance: The new method can be of service in quantitative EMG.

1. Introduction

The analysis of motor unit action potential (MUAP) is one of the fundamental tests in routine clinical neurophysiology. Electromyography (EMG) signals are recorded intramuscularly with conventional concentric needle electrodes. These signals usually contain several MUAP trains. Manual, semi or completely automatic techniques (Nandedkar and Barkhaus, 2002; Merletti and Parker, 2004) are used for decompose EMG signals into different MUAP trains. From each MUAP train a representative waveform is formed (Malanda et al., 2015) and characterized with clinically useful

Abbreviations: BPM, best performing method; EA, ensemble averaging; EMG, electromyography; FCA, five-closest averaging; GSMW, gold standard MUAP waveforms; MA, median averaging; MUAP, motor unit action potential; MWP, MUAP waveform parameters; NBP, normalized baseline power; NDEP, normalized differential error power; NEP, normalized error power; SD, standard deviation; SPMF, signal processing merit figures; SWSA, sliding window selective averaging.

* Corresponding author at: Electrical and Electronics Department, Public University of Navarre, Campus de Arrosadía, 31006 Pamplona, Spain. Fax: +34 948 169720.
E-mail address: malanda@unavarra.es (A. Malanda-Trigueros).

https://doi.org/10.1016/j.clinph.2018.02.134
1388-2457/

parameters (Stålberg et al., 1986; Zalewska and Hausmanowa-Petrusewicz, 1995; Nandedkar and Barkhaus, 2002; Kimura, 2002). For diagnostic evaluation, getting a reliable and representative MUAP waveform is thus an essential goal of quantitative EMG.

Noise and artifacts from different sources can distort MUAP waveforms. Averaging methods have been designed for obtaining representative waveforms from MUAP trains. Conventional averaging methods based on the mean of samples are noticeably sensitive to noise and artifacts, particularly, to the interference of potentials from different motor units (Malanda et al., 2016). Furthermore, such methods sometimes perform excessive smoothing of the resulting curves, which can lose morphological details of potentially useful physiological implication. Finally, these methods give rise to amplitude bias when alignment errors are present (Malanda et al., 2015; Sörmo and Lagunas, 2005). Methods based on the median are more robust, but tend to produce ragged waveforms (Malanda et al., 2015).

Robust methods of averaging are especially relevant in automatic extraction of MUAP trains by means of multi-MUP systems (Stålberg et al., 1995; Nandedkar and Barkhaus, 1995), which are designed to reduce the time of MUAP extraction (and consequently patient discomfort), being able to obtain several MUAP trains (generally up to six, in commercial systems) from each point of needle insertion. Patients are asked to perform moderate voluntary muscular contractions, so that several motor units are activated. The presence of different MUAP trains in the recordings leads to frequent superimpositions of potentials, whose waveforms are consequently distorted to some extent, making further demands on the averaging method used to disentangle and extract a representative waveform.

The building of a waveform that serves as a model or template for a set of curves in a repetitive signal is a recursive problem in the field of biomedical signal analysis. Several methods have been put forward for obtaining such templates from biomedical repetitive signals of different kind: EMG (i.e, MUAP analysis) (Stålberg and Antoni, 1983; Nandedkar and Sanders, 1989; Stålberg and Sonoo, 1994; Nandedkar and Barkhaus, 1995), evoked potentials (Hoke et al., 1984; Mühler and von Specht, 1999; Sörmo and Lagunas, 2005; Leonowicz et al., 2005; Rahne et al., 2008) and electrocardiography (Leski, 2002). A comprehensive review of these methods together with a comparative evaluation of a selection of them was recently published (Malanda et al., 2015).

In 2016, the current authors presented a new method for obtaining a representative waveform from a train of MUAPs. Briefly, a window slides along the time axis selecting and averaging the most similar sections of the potential train within its scope. From the obtained pieces of potentials, an assembled potential is generated, that satisfactory represents the waveforms of the MUAP train. This approach was referred to as *Sliding-window selective averaging* (SWSA).

The SWSA approach was compared with a selection of the nine methods evaluated in the previously-conducted comparative study (Malanda et al., 2015) and was found to improve on the performance of the older methods in terms of the criteria of comparison (various signal analysis-based merit figures and MUAP waveform parameters used in the clinical practice). Regarding MUAP waveform parameters, the new algorithm outperformed the other methods evaluated.

The current study extends our previous work to evaluate performance with MUAP recordings from pathological muscles. In the following section, a description of the materials used in the study is given. Next, we briefly describe the SWSA method and the other methods evaluated. Then comes an explanation of the gold standard and the figures of merit used in comparisons. After providing a report of the comparative evaluation results and further discussion of these results, our final conclusions are given.

2. Methods

2.1. Subjects and signals

For this study we made use of the material used in a previous work (Rodríguez-Carreño et al., 2010), with the expressed approval of the Public University of Navarre's Ethical Committee. Specifically, we used 313 EMG signals, between 5 and 6 seconds long, acquired during slight voluntary contractions: 68 signals were from normal muscles, 105 from myopathic muscles, 27 from chronic neurogenic muscles and 72 from subacute neurogenic muscles. The types of muscles and particular neurological diseases related to these signals can be consulted in the previous reference. Details about the recording equipment and acquisition set-up can also be found in that reference.

MUAP trains were extracted from EMG signals using an automatic decomposition procedure (Florestal et al., 2006). The potentials in the MUAP trains consisted of 50 ms-long EMG signal epochs. For all the extracted MUAP trains the potentials were segmented from the EMG signals in such a way that their maximal negative peaks appeared at 40% of the length of the epoch.

Next, the potentials of each MUAP train extracted by the decomposition algorithm were aligned in the time axis by maximum correlation (Campos et al., 2000) and in the amplitude axis by Euclidean distance minimization (Navallas et al., 2006). MUAP trains with an excessively noisy visual appearance or that yielded average waveforms with unrealistic MUAP shapes to the eyes of an expert electromyographer (LG), were considered unacceptable and discarded for subsequent analysis. MUAPs with satellite potentials were also excluded. All the selected MUAP waveforms were well-defined above BL activity and had a rise-time lower than 1 ms (most of them lower than 0.5 ms). Finally, MUAP trains with less than 80 potentials were discarded, as this was set as the minimum MUAP train size for the comparative analysis. A total of 218 MUAP trains were accepted for the study: 37 from normal muscles, 69 from myopathic muscles, 64 from subacute neuropathic muscles and 48 from chronic neurogenic muscles.

2.2. The Sliding-window selective averaging method

The SWSA algorithm starts with the potentials in the MUAP train already aligned in time and amplitude (Fig. 1A). Then a window slides along the MUAP time span delimiting intervals of the set of potentials (Fig. 1A and B). For each time interval, the so-called *median section* is calculated as the median of the samples of all the potentials in the train. The standard deviations of the amplitude samples of the different potentials in the train are also obtained and the minimum value across the time samples in the interval is extracted. Then, the algorithm evaluates potentials with a small Euclidean distance to the *median section*. Potentials with Euclidean distances that are lower than the previously-estimated minimum standard deviation multiplied by a constant parameter (η), are selected and averaged. In this way, the algorithm obtains representative sections for the time intervals under consideration (Fig. 1C). In the final stage, the representative sections obtained as the window slid along the MUAP time span are assembled (Fig. 1D) and averaged to form the final representative waveform. A complete description of the SWSA algorithm can be found in the original article in which it is presented and evaluated (Malanda et al., 2016).

To evaluate the SWSA algorithm, the parameter η is set to 1.0, while three different values of the window length are considered: L_w = 50, 150 and 250 samples (i.e., 2.5, 7.5 and 12.5 ms, respectively). These parameter values were chosen, on the basis of results in the original study (Malanda et al., 2016), in which L_w was tested

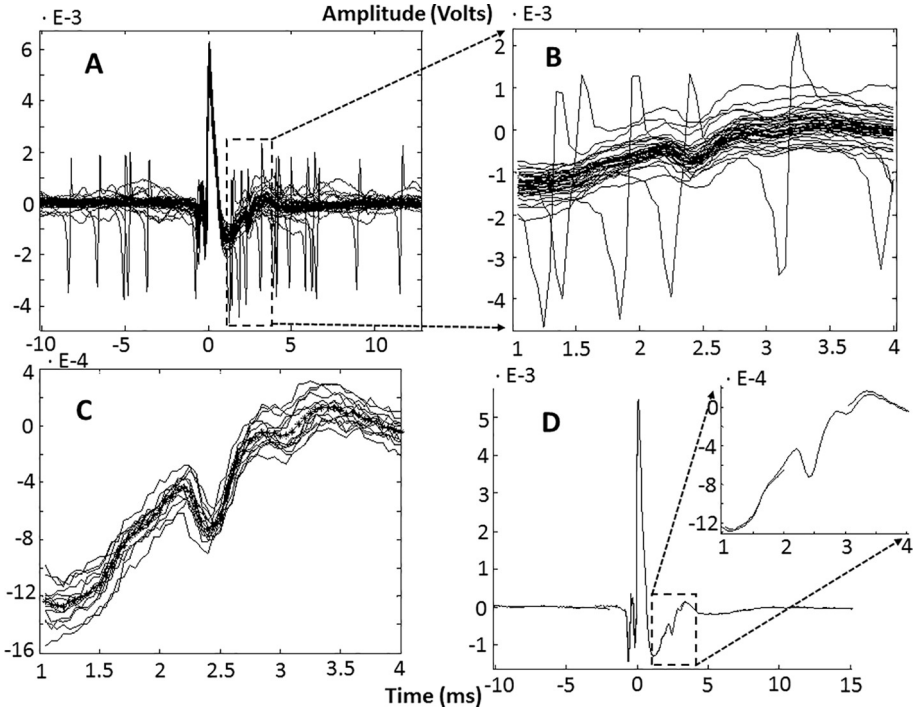

Fig. 1. Different steps in the SWSA method. Potentials of a MUAP train aligned in the time and amplitude axes; a sliding window (box) is used to select sections of these potentials within a time interval (A). Zoomed view of the sections of the potentials delimited by the sliding window (B). Sections selected by the method and the average of these sections (curve of stars) delimited by the sliding window (C). Representative sections for each of the different time intervals assembled to form the final representative waveform; a zoomed view of a time interval delimited by the sliding window is shown in the insert (D). (To facilitate the explanations, the window in the example slides through steps of 40 samples (2 ms) instead of one sample (0.05 ms), as used in the algorithm). (Amplitude and time scales are different in the different subplots).

in the range from 15 to 400 samples (from 0.75 to 20 ms). With these values, performance measurements can be maximized while keeping some variability in the method. The corresponding configurations of the algorithm will be referred hereafter as SWSA 1, SWSA 2 and SWSA 3, respectively.

2.3. Conventional averaging methods

The SWSA was compared with three other methods for obtaining representative curves from MUAP trains:

(a) *Ensemble averaging* (EA)
(b) *Median averaging* (MA) (Antoni, 1983)
(c) *Five-closest averaging* (FCA): Average of the five potentials closest to each other among the train of MUAPs, as measured by the Euclidean distance (Malanda et al., 2015).

These three methods were selected from the nine conventional methods used in the previous work (Malanda et al., 2015). From the results of that work, MA generally performed the best. While not among the best performers, EA was chosen because, being the simplest of the nine methods, it may well serve as a reference point for other comparative studies. FCA generally had intermediate performance, with notably high scores in some scenarios and low in others.

2.4. Gold standard

As in our previous studies (Malanda et al., 2015, 2016), gold standard MUAP waveforms (GSMWs) were used as references to evaluate the representative waveforms output by the different tested methods. For obtaining the GSMWs the potentials in the trains were first time-aligned and MUAP initial and end points were manually marked. Then a subset of potentials with similar shapes within the marked limits was visually selected and, finally, the selected subset was averaged. Manual marking and selection was done by an experienced physician qualified in electromyography. The procedure was carried out with the use of a Matlab graphical user interface especially designed for the task.

2.5. Merit figures

We used two different sets of merit figures, as in our previous works (Malanda et al., 2015, 2016). The first set, the *signal processing merit figures* (SPMFs), measure the similarity between a waveform generated by the tested algorithms and the GSMW. The second set, the *MUAP waveform parameters* (MWPs), are descriptive features typically used for MUAP quantitative characterization; these parameters are related to MU structure and physiology and may be relevant in clinical evaluation. The SPMFs and MWPs are described in detail below.

(A) Signal processing merit figures

We calculated the three following merit figures.

Normalized error power (NEP): error power normalized to signal power within MUAP duration:

$$NEP = \frac{\sum_{n_1}^{n_2}[e(n)]^2}{\sum_{n_1}^{n_2}[x(n)]^2} \quad (1)$$

where $x(n)$ is the representative MUAP curve obtained by a given method; $e(n) = x(n)\text{-}g(n)$ is the *error signal* (i.e., the difference between the representative curve and the GSMW); $g(n)$ is the GSMW curve; and n_1 and n_2 are the time samples for the MUAP initial and end marked by the expert.

Normalized baseline power (NBP): signal power outside MUAP duration divided by signal power within MUAP duration:

$$NBP = \frac{n_2 - n_1}{L + n_1 - n_2 - 1} \cdot \frac{\sum_{n=0}^{n_1-1}[x(n)]^2 + \sum_{n_2+1}^{L-1}[x(n)]^2}{\sum_{n=n_1}^{n_2-1}[x(n)]^2} \quad (2)$$

where L is the length of x(n).

Normalized differential error power (NDEP): power of the time derivative of the error signal, normalized to the power of the time derivative of the signal, within MUAP duration:

$$NDEP = \frac{\sum_{n_1}^{n_2}[\delta_e(n)]^2}{\sum_{n_1}^{n_2}[\delta_x(n)]^2} \quad (3)$$

where $\delta_x(n)$ is the time derivative of $x(n)$, estimated by filtering $x(n)$ consecutively with a 4-tap averaging filter, $h_1(n)$, and with a first order differentiator, $h_2(n)$. These filters are defined by their impulse responses:

$$h_1(n) = \begin{array}{l} 0.25 \text{ if } n = 0,\ldots,3, \\ 0 \text{ if } n < 0 \text{ or } n > 3 \end{array} \quad (4)$$

$$h_2(n) = \begin{array}{l} 1 \text{ if } n = 0, \\ -1 \text{ if } n = 1, \\ 0 \text{ if } n < 0 \text{ or } n > 1 \end{array} \quad (5)$$

Similarly, $\delta_e(n)$ is the time derivative of $e(n)$, estimated by filtering $e(n)$ with $h_1(n)$ and $h_2(n)$, sequentially.

NEP tries to measure the similarity of the representative MUAP to the GSMW curve within the time span of the potential: the more similar the two curves, the lower the NEP. NBP is related to baseline amplitude. The lower the NBP, the easier it is to estimate MUAP duration and, consequently, other MUAP parameters. Finally, DEP provides a measure of the overall similarity in shape, mainly the higher frequency components, between the representative MUAP and the GSMW curve. This figure of merit is sensitive to a ragged (noise-like) aspect of the MUAP waveform.

(B) MUAP waveform parameters

We used six parameters used extensively for MUAP waveform quantitative analysis (Stålberg et al., 1996; Zalewska and Hausmanowa-Petrusewicz, 2000): peak to peak amplitude, area, number of phases, number of turns, irregularity coefficient and spike duration. The definition of these parameters are given in the previous work (Malanda et al., 2016). For the number of turns, we used two different thresholds, 25 μV and 50 μV, both of which have been used in previous studies (Pfeiffer and Kunz, 1992; Stewart et al., 1989; Stålberg et al., 1995).

Error measurements for a given averaging algorithm were computed as the differences between the MWPs of the algorithm-generated waveform and those of the GSMW. Relative error measurements were computed by dividing the corresponding error measurements by the MWPs of the GSMW.

For the number of turns, two error measurements were obtained: missed-turn errors and false-turn errors. Missed-turn errors give account of the number of turns that are present in the GSMW but not in an algorithm-generated waveform. On the other hand, false-turn errors counts the turns found in an algorithm-generated waveform that are not present in the GSMW. Temporal coincidence of a turn in the GSMW and in an algorithm-generated waveform was not restricted to a single time sample but to a small time interval. This enforced robustness against waveform variations, misalignments and noise. A 0.25 ms-long interval was used in this study.

2.6. Evaluation tests

To evaluate the significance of the results for each of the cases considered, we first found the method that yielded the least mean square error; this was considered the best performing method (BPM). Then we determined if any other method yielded significantly larger errors than those obtained with the BPM.

For amplitude, area, number of phases, irregularity and spike duration errors and for the three signal processing merit figures, we used the two tailed paired T-test ('ttest' Matlab function), to evaluate if the error distributions of the different methods had significantly different mean than the BPM. We also used the two-sample F-test ('vartest2' Matlab function) to evaluate if error distributions had significantly larger variance than the BPM. Methods with significant difference to the BMP in either of these two tests were considered as significantly different from the BMP.

As the statistical distributions of the missing-turn and false-turn errors were highly asymmetrical, we used the Wilcoxon sign rank test ('signrank' Matlab function) to make comparisons for these parameters. This test evaluated if the distributions of the errors yielded by each method differed significantly from those of the BPM with respect to the median value. For the three statistical tests used, a P-value of 0.01 or less was considered significant.

The sign of MWP error measurements is not relevant for a performance comparison (i.e., an error measurement taken from a curve provided by a certain method with respect to the GSMW should not be considered either better or worse than another measurement with the same magnitude of error, but of opposite sign). According to this, a method was considered to have significant differences with respect to the GSMW if the corresponding errors of this method were significantly different to the errors of the GSMW, and when changing the sign of the errors of this method they were also significantly different to the errors of the GSMW.

In the first part of the analysis 50 potentials from each MUAP train were used in the tests. In the second part of the analysis we studied the evolution of the final average performance scores as the number of potentials per MUAP train used in the analysis increased from 10 to 80 in steps of 10.

3. Results

Table 1 presents the mean and standard deviation (SD) of the MWP values calculated from the GSMWs obtained from the collections of MUAP trains recorded in the normal and the three types of pathological muscles considered.

Boxplots of the three SPMFs, using 50 potentials per MUAP train, are given in Fig. 2. The median of the distributions, the 25th and 75th percentiles and some of the outliers are shown (see the figure caption). The BPM and the methods that gave significantly larger errors than the BPM are marked in the figure.

EA and FCA presented significantly higher NEP than that of the BPM in all four pathology groups. FCA had the largest average and SD values of all the groups. MA and SWSA 1 presented significantly higher NEP than that of the BPM in the myopathic and subacute neurogenic groups. Of all methods, SWSA 2 and SWSA 3 performed best. SWSA 3 NEP was not significantly larger than that of the BPM in any of the four groups; SWSA 2 NEP was only significantly larger than BPM NEP in the subacute neurogenic group.

FCA had significantly higher NBP than that of the BPM in all four muscle pathology groups: EA NBP was significantly higher than BPM NBP in all pathology groups except that for normal muscles. As with NEP, FCA values were much larger than those of the other methods. SWSA and MA gave the best results.

With respect to NDEP, the results of the statistical analysis were almost identical to the NEP results. The main difference was that

Table 1
Mean and SD (in brackets) of MUAP parameter values obtained from the GSMW.

	Amplitude (mV)	Area (mV·ms)	Irreg. coef.	Num. phases	Num. turns	Spike dur. (ms)
Normal	0.51 (0.35)	0.68 (0.30)	2.60 (0.28)	3.08 (0.80)	2.62 (0.92)	4.10 (1.82)
Myopathic	0.37 (0.31)	0.40 (0.23)	2.81 (0.58)	2.97 (0.97)	2.65 (1.30)	3.76 (1.88)
Subacute neurogenic	0.98 (1.77)	1.31 (2.77)	3.51 (0.88)	4.66 (1.76)	5.30 (2.47)	3.60 (3.45)
Chronic neurogenic	3.82 (3.20)	5.51 (4.74)	3.13 (1.40)	5.27 (3.26)	7.37 (6.08)	5.01 (4.34)

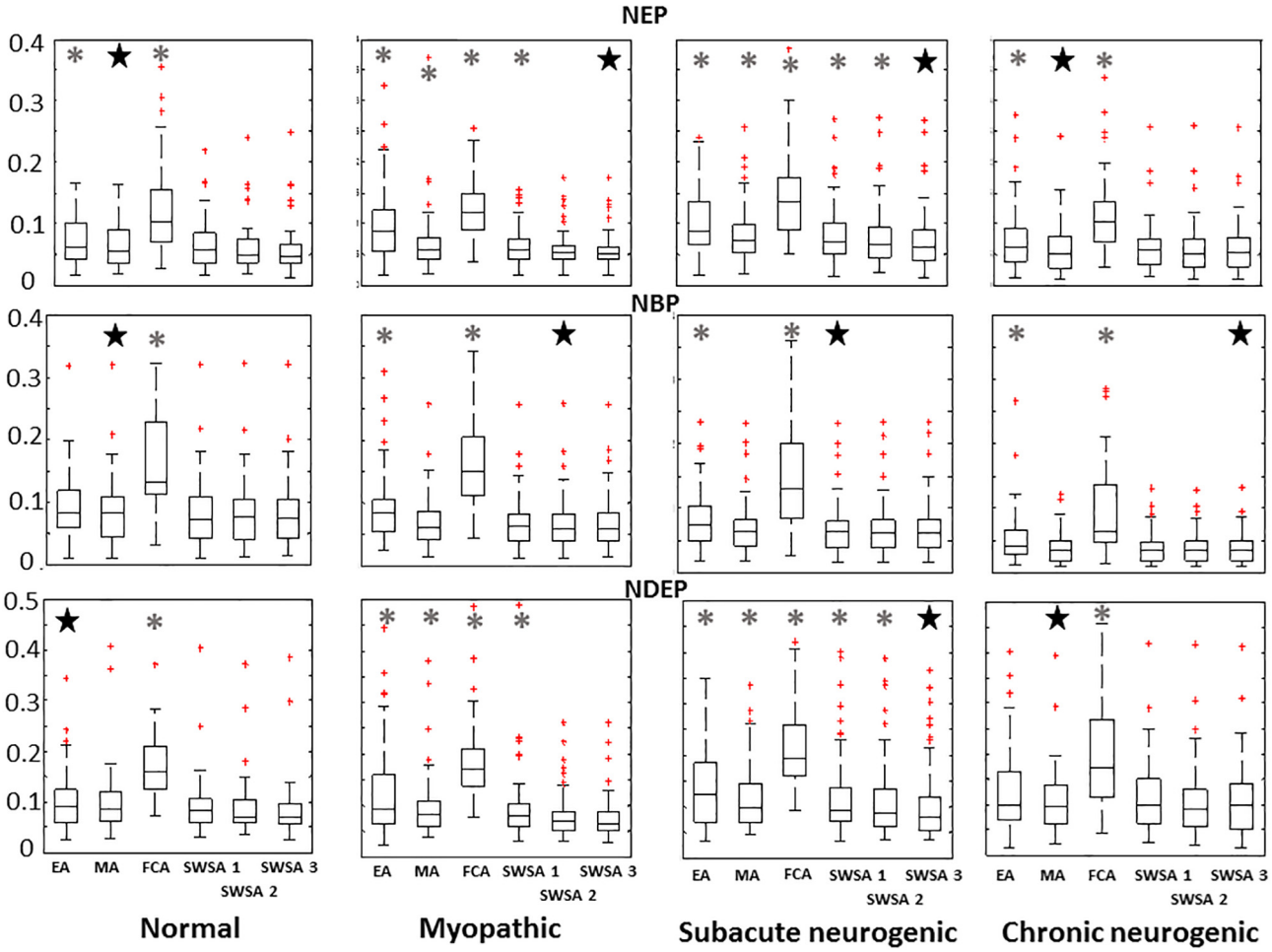

Fig. 2. Boxplots of the three SPMF for the four studied muscle pathology groups. On each panel, boxplots of the six evaluated methods are pictured (from left to right: EA, MA, FCA, SWSA 1, SWSA 2 and SWSA 3). On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually. Plots have been zoomed to better view the central parts of the distribution and some outliers may not appear. The best performing method is marked with a black star, while methods with an error that is significantly larger are marked with grey asterisks.

EA NDEP was only significantly larger than the BMP NDEP for the myopathic and subacute neurogenic groups.

Tables 2 and 3 show the percentage of scenarios (i.e. combinations of muscle pathology group and specific merit figures) in which a method's SPMF error was not significantly larger than that of the BPM, for the case of 50 potentials per MUAP train. So the larger this percentage (*score*), the better the performance. Table 2 shows these percentages for each of the muscle pathology groups (pooling the three SPMFs). Table 3 shows percentages by SPMF (pooling the muscle pathology groups).

The best performing method was SWSA 3, as SPMF errors were not significantly larger than those of the BPM in any of the

Table 2
Percentages of scenarios where SPMF error was not significantly larger than that of the BPM, by muscle pathology grouping.

	EA	MA	FCA	SWSA 1	SWSA 2	SWSA 3	Mean
Normal	66.67	100	0	100	100	100	77.78
Myopathic	0	33.33	0	33.33	100	100	44.44
Subacute neurogenic	0	33.33	0	33.33	33.33	100	33.33
Chronic neurogenic	33.33	100	0	100	100	100	72.22
Mean	25.0	66.67	0	66.67	83.33	100	56.94

Table 3
Percentages of scenarios where SPMF error was not significantly larger than that of the BPM, by SPMF grouping.

	EA	MA	FCA	SWSA 1	SWSA 2	SWSA 3	Mean
NEP	0	50	0	50	75	100	45.83
NBP	25	100	0	100	100	100	70.83
NDEP	50	50	0	50	75	100	54.17
Mean	25	66.67	0	66.67	83.33	100	56.94

scenarios considered. The second best method was SWSA 2, which only had significantly larger SPMF errors than those of the BPM in the case of NEP and NDEP in the subacute neurogenic group; yielding an overall error percentage score of 83.33%. Next were SWSA 1 and MA, with SPMF errors greater than those of the BPM for NEP and NDEP in the myopathic and the subacute neurogenic groups; the overall score was 66.67%. FCA had significantly larger errors than the BPM in all the cases.

Boxplots of the error measurements of the nine MWPs, using 50 potentials per MUAP train, are given in Fig. 3. The median of the distributions, the 25th and 75th percentiles and some of the outliers are shown (see the figure caption). The best performing method and methods that had a MWP distribution that was significantly different from that of the BPM are marked in the figure.

The amplitude measurement had negative bias (average error of negative sign) with EA for all the tested groups. When applied to MUAP trains recorded from normal and both neurogenic muscles, FCA had errors significantly larger than those of the BPM. Similarly, for chronic neurogenic muscles, EA had significant amplitude errors.

Area measurements obtained by all the evaluated methods had positive bias (larger values on average than those in the GSMW) in normal, myopatic and subacute neurogenic groups. With normal muscles, EA, MA, SWSA 1 and SWSA 2 all exhibited significantly different area measurements relative to the SWSA 3, which was the BPM. With pathological muscles, EA estimates were significantly worse than those of the BPM.

Regarding number of phases, for the normal muscle group, MA, SWSA 2 and SWSA 3 curves had no errors at all, FCA had large errors (mean and SD of 9.10% and 22.12%, respectively) and EA had smaller errors (5.40% and 18.45%, respectively). Applied to the myopathic and chronic neurogenic groups, FCA errors were

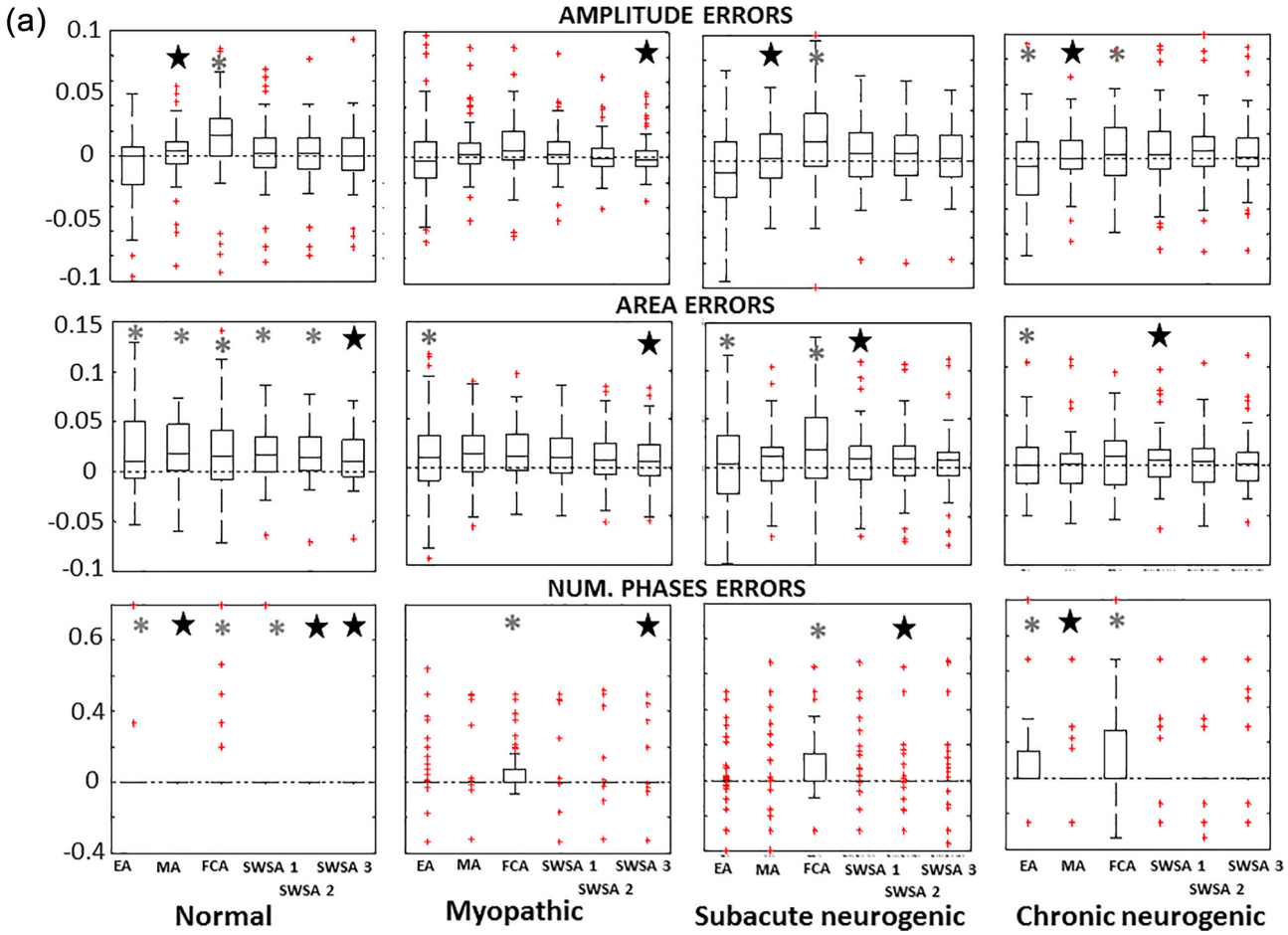

Fig. 3. Boxplots of the MWPs for the four studied muscle pathology groups. On each panel, boxplots of the six evaluated methods are pictured (from left to right: EA, MA, FCA, SWSA 1, SWSA 2 and SWSA 3). On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually. Plots have been zoomed to better view the central parts of the distribution and some outliers may not appear. The best performing method is marked with a black star, while methods with an error that is significantly larger are marked with grey asterisks.

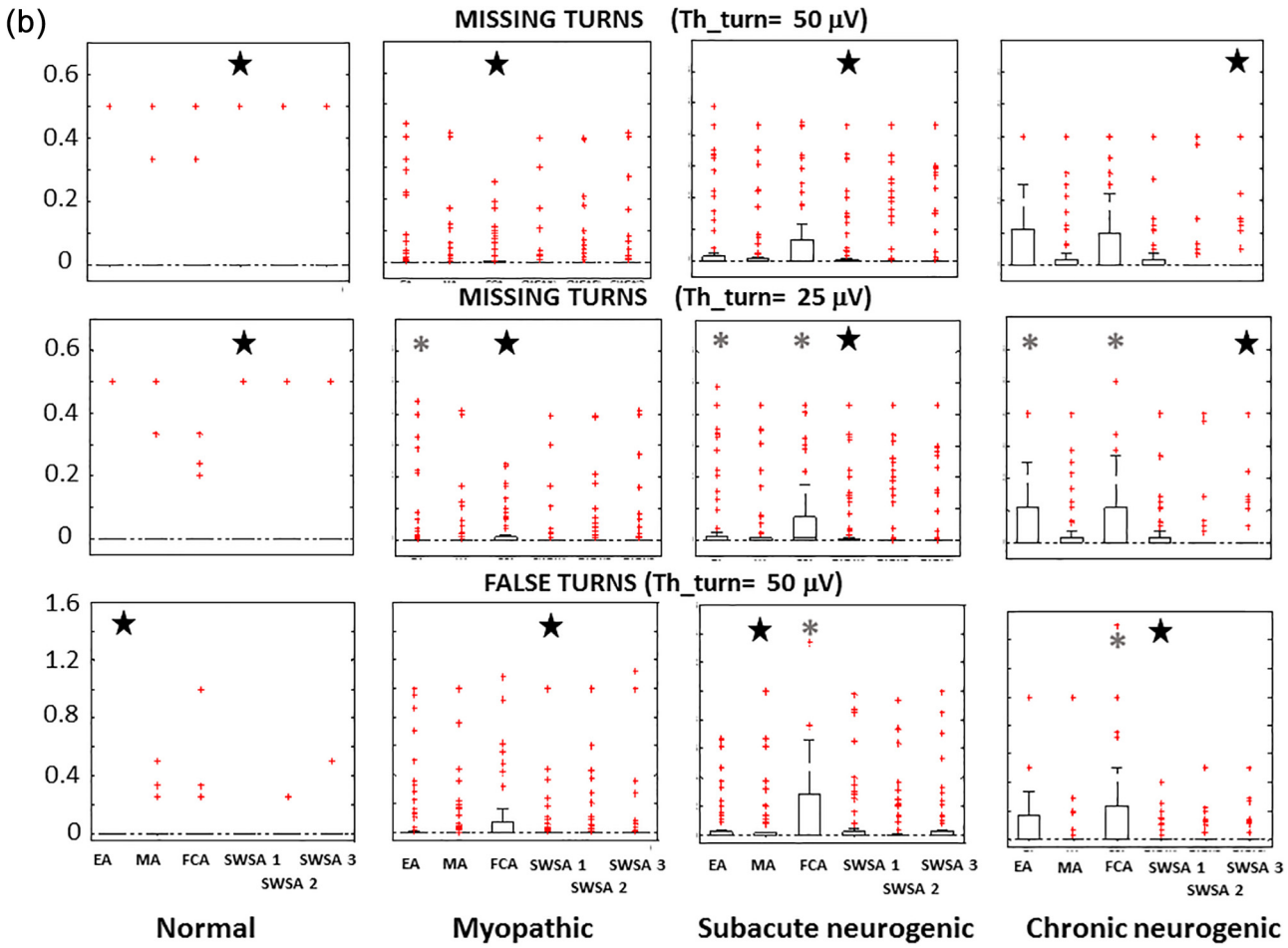

Fig. 3 (*continued*)

significantly (and indeed markedly) larger than those of the BPM (means 5.44% and 13.92%; and SD 17.32% and 28.63%, respectively). With chronic neurogenic muscles, EA also gave estimates with large errors (mean 13.74% and SD 32.07%).

Irregularity coefficient errors were significant for EA and FCA across all muscle groups. There were also significant differences between MA estimates of irregularity coefficient and those of the BPM when applied to MUAPs from myopathic muscles.

Regarding turns, neither MA nor any of the three SWSA versions presented significant differences with respect to the BPM. In contrast, FCA presented significantly large error differences in 8 out of the 16 existing scenarios (related to turns); false-turn error mean was between 8.35% and 16.15% with the 50 μV threshold and between 8.92% and 20.82% with the 25 μV threshold, and error SD was between 28.02% and 29.96% with the 50 μV threshold and between 20.54% and 40.03% with the 25 μV threshold. EA presented significantly large errors in 4 out of the 16 existing scenarios.

Finally, with regard to spike duration, FCA presented significantly large errors with respect to the BPM for the normal and both neurogenic groups; the error mean and error SD was notably high with subacute neurogenic muscles (4.86% and 22.42%, respectively). EA and MA showed significantly large errors with respect to the BPM when applied to MUAPs from myopathic and chronic neurogenic muscles. None of the three SWSA methods presented significant differences with respect to the errors of the BPM.

Table 4 shows the percentages of MWPs for which a method's error was not significantly larger than that of the BPM for each of the muscle pathology groups, averaging among the nine MWPs, using 50 potentials per MUAP train. To calculate this percentage, the four turn-related MWPs (i.e., missing turns and false turns with thresholds of 50 μV and 25 μV) were previously averaged. Similarly, Table 5 shows percentages of muscle pathology groups for which a method's MWP error was not significantly larger than that of the BPM for each MWP, averaging among the four studied groups. The last row of each of these tables shows the percentage average among the nine MWPs and the four studied groups. We will refer to this average percentage as the overall score of a method.

SWSA 3 MWP errors were not significantly larger than those of the BPM in any of the scenarios considered (score 100%). The second best method was SWSA 2, which only had significantly larger errors than the BPM with the normal group and for the area parameter (score 95.83%). The third best method was SWSA 1 (overall score 91.67%). Fourth was MA (overall score 83.33%). Finally, were EA and FCA, with significantly smaller scores: 37.5% and 25%, respectively.

Fig. 4 shows the evolution of the SPMF and the MWP overall scores as the number of potentials per MUAP train varied from 10 to 80. For the MWP, we can observe a clear ascending trend in the overall score of all the methods coming to a steady state for 50 or 60 potentials per train. This performance was above 90% for the three versions of SWSA for 40 potentials per MUAP or more, while the three other methods had lower scores (hardly above 40% in EA and FCA). For the SPMF, the ascending trend was not so evident. Here SWSA 2 and SWSA 3 had the best

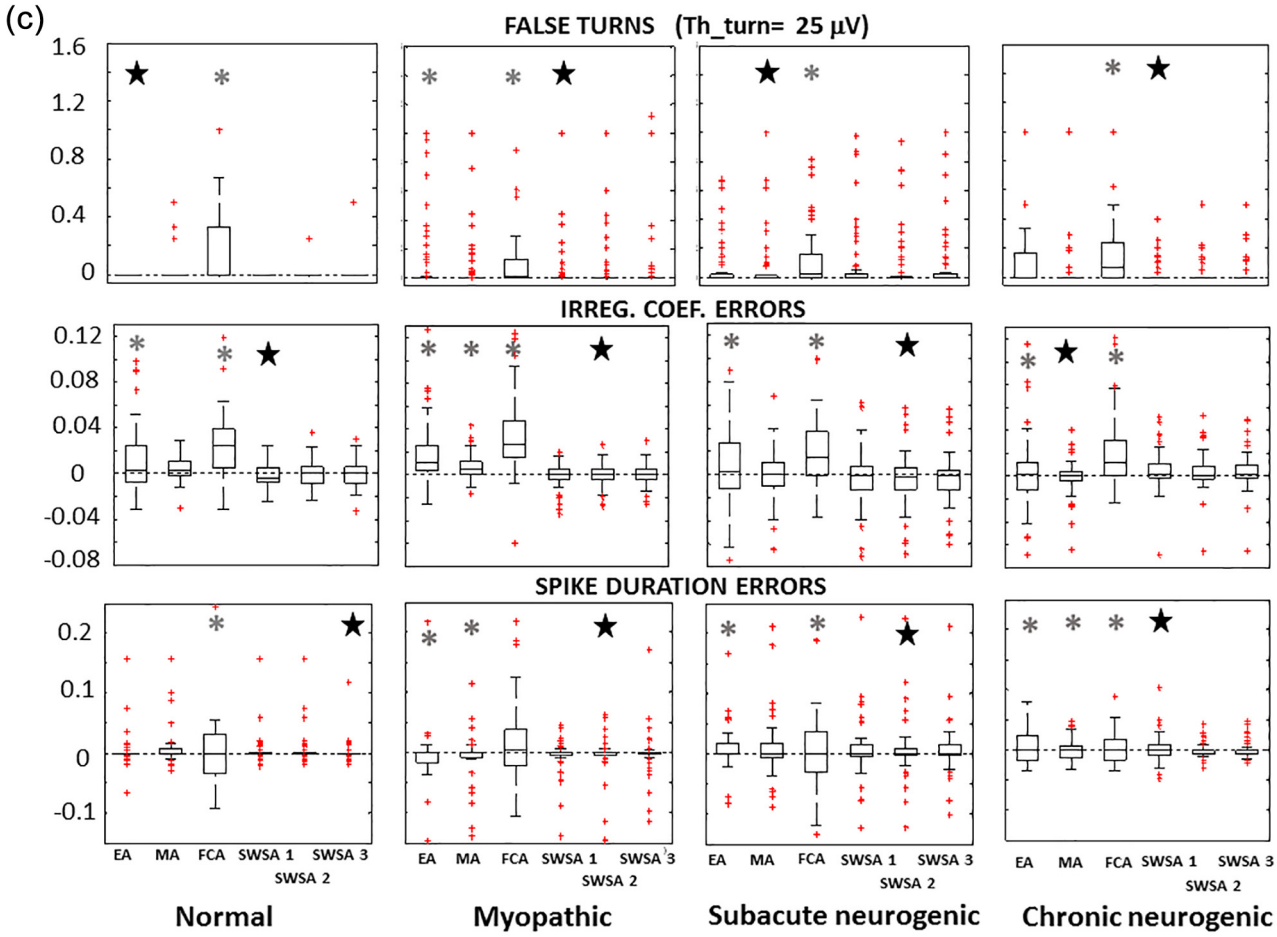

Fig. 3 (*continued*)

Table 4
Percentages of MWPs for which a method's error was not significantly larger than that of the BPM.

	EA	MA	FCA	SWSA 1	SWSA 2	SWSA 3	Mean
Normal	50	83.33	12.5	66.67	83.33	100	65.97
Myopathic	41.67	66.67	62.5	100	100	100	78.47
Subacute neurogenic	45.83	100	4.17	100	100	100	75
Chronic neurogenic	12.5	83.33	20.83	100	100	100	69.44
Mean	37.5	83.33	25	91.67	95.83	100	72.22

Table 5
Percentages of muscle pathology groups for which a method's MWP error was not significantly larger than that of the BPM.

	EA	MA	FCA	SWSA 1	SWSA 2	SWSA 3	Mean
Amplitude	75	100	25	100	100	100	83.33
Area	0	75	50	75	75	100	62.5
N. phases	50	100	0	75	100	100	70.83
Irreg. coef.	0	75	0	100	100	100	87.5
Missing turns (Th = 50 μV)	100	100	100	100	100	100	83.33
Missing turns (Th = 25 μV)	25	100	50	100	100	100	70.83
False turns (Th = 50 μV)	100	100	50	100	100	100	83.33
False turns (Th = 25 μV)	75	100	0	100	100	100	79.17
Spike dur	25	50	25	100	100	100	66.67
Mean	37.5	83.33	25	91.67	95.83	100	72.22

performance, followed by MA and SWSA 1, while EA and FCA had the worst performance. Note that FCA consistently had 0% score, for all the number of potentials per MUAP train included in the analysis.

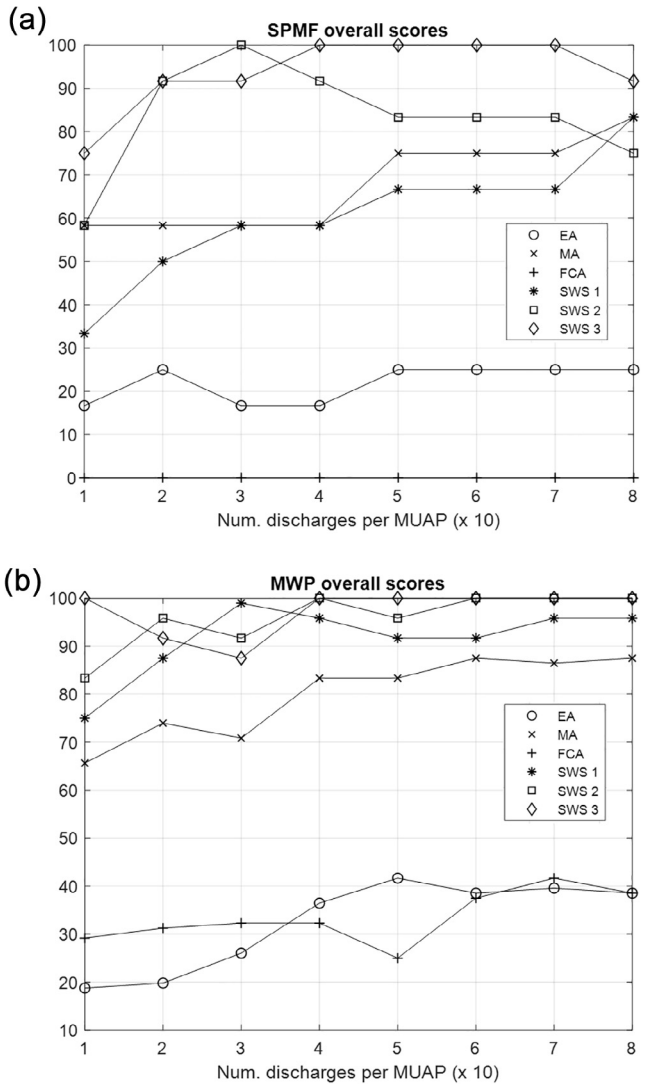

Fig. 4. Evolution curves of the SPMF (a) and the MWP (b) overall scores for the six evaluated methods with varying number of potentials per MUAP train.

4. Discussion

Several aspects of this work deserve further attention:

- On the basis of SPMF and MWP results (Tables 2–5), SWSA 3 was the method that performed best, followed by SWSA 2. SWSA 1 and MA came next, with similar levels of accuracy. EA was fifth and FCA last (see Figs. 4 and 5). These results are in good agreement with the results obtained in the previous study (Malanda et al., 2016) that only considered MUAPs from normal muscles. As pointed out in that study, the SWSA algorithm is of local scope in nature, and this permits averaging of a train of potentials with the exclusion of portions of potentials that strongly differ from the rest of the potentials in the train. Such distinct portions are expected to arise from interfering potentials fired from other motor units.
- SWSA was robust to variation in the window length. While a window length of 12.5 ms (corresponding to SWSA 3) yielded the best performance of the three explored versions (with an overall score of 100% for both SPMFs and MWPs), a window length of 7.5 ms (in SWSA 2) also performed well (with scores of 95.83% and 83.33%), and even the short, 2.5 ms, window of SWSA 1 had reasonably low error scores (91.67% for SPMFs and 66.67% for MWPs).
- The inferior performance of MA relative to the SWSA methods was evident in the spike duration parameter (Table 5) and the NEP and NDEP merit figures (Table 3), and it can also be appreciated in the irregularity coefficient of the example of Fig. 5 and the error curves of Fig. 6.
- EA tended to underestimate amplitude for MUAPs from the normal muscle group and from the two neurogenic muscle groups (Fig. 3). The same behaviour was observed with the recordings from normal muscles used in our previous study (Malanda et al., 2016), and a possible explanation was given there: misalignments among the potentials of the MUAP train lead to a reduction in the amplitude of the EA averaged curve, a well-documented phenomenon in the field of signal processing (Sörmo and Lagunas, 2005). In the present study, MA and the three SWSA versions also tended to underestimate amplitude for the normal group (Fig. 3) although this tendency was smaller than it was with EA. These findings are all in agreement with the results of the previous work. No such negative bias was observed when the methods were applied to potentials in the three pathological groups. The reasons for this inconsistent behaviour are not clear. Finally, FCA presented a positive bias in amplitude for all muscle-pathology groups (Fig. 3), a result congruent with those of the previous study and explained in the corresponding article.
- All methods presented positive bias in estimation of area across all pathology-defined groups (Fig. 3). Although amplitude and area parameters usually present high correlation, amplitude and area errors are not necessarily correlated. One of the reasons to explain this is that area errors are highly dependent on onset and end marker locations, while amplitude errors are not dependent on these locations. So, the "paradox" of negative amplitude bias and positive area bias that appeared in some scenarios (i.e., MA, EA and in the three SWSA methods for the normal group and in EA for both neurogenic groups) is only apparent.
- Related to the definition of spike duration (i.e., 'the time interval between the first and the last positive peak of the MUAP') (Stålberg et al., 1986), we considered problematic cases those in which the average MUAP obtained by one of the tested methods did not present a positive peak at either sides of the largest negative peak of the potential. Here the error in spike duration was affected in an unpredictable manner, sometimes taking on excessively large values. In these cases, the results about spike duration corresponding to this MUAP train were excluded from the study for the whole set of tested methods. This occurred in 22 cases: 4 from the normal group, 10 from the myopathic group, 4 from the subacute neurogenic group and 4 from the chronic neurogenic group.
- The performance of the methods in this study was slightly different from that in our previous study (Malanda et al., 2016). The differences referred to, of course, only concern normal muscles and a 50 μV-turn threshold, as these were the only cases tested in the previous study. The performance differences are most evident with respect to EA and MA. In the previous study, EA errors were significantly larger than BPM errors for all the MWPs and SPMFs assessed, while in the present study errors were not larger for amplitude, missing and false turns, spike duration and NBP and NDEP merit figures. Likewise, in the previous study, MA errors were significantly larger than BPM errors for area, irregularity coefficient, missing and false turns, spike duration and NEP and NDEP merit figures, while in the present study, MA errors were only larger for area. A possible explanation for these differences is that they arise from the different

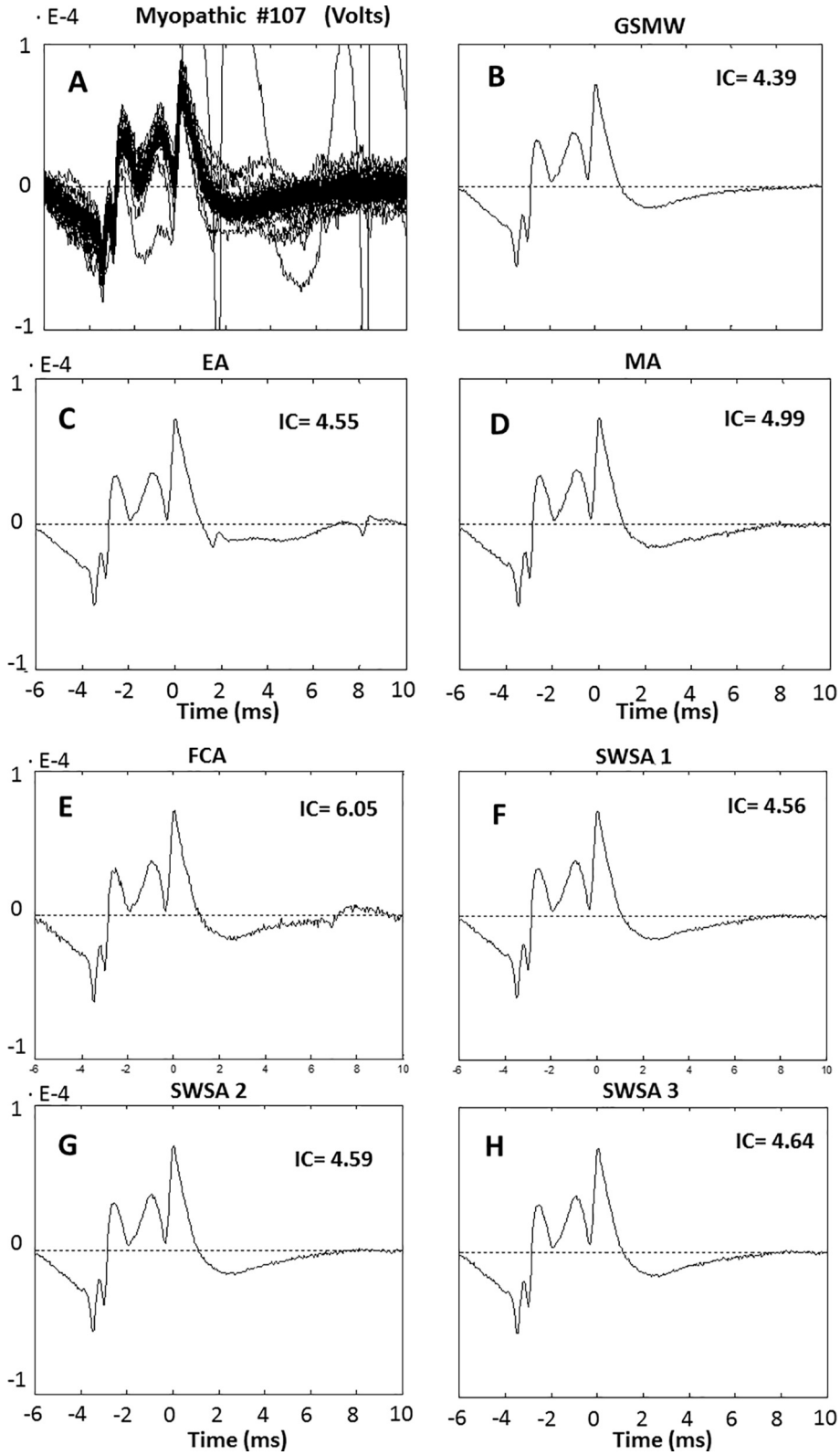

Fig. 5. Example of MUAP set (A), the GSMW (B) and the curves resulting from the application of the studied averaging methods (C–H). Note the awkward fluctuation of the EA and FCA curves around 8 ms, due to the large artefacts recorded.

protocols followed to acquire the EMG signals in the two studies. In both studies, the EMG signals were recorded in a continuous ('free run') mode, and after that a decomposition algorithm (Florestal et al., 2006) was used to extract the various MUAP trains included in the study. However, in the present study, a second check was performed manually by an expert electromyographer, and special care was taken to select only clean trains, with low degrees of contamination from noise or

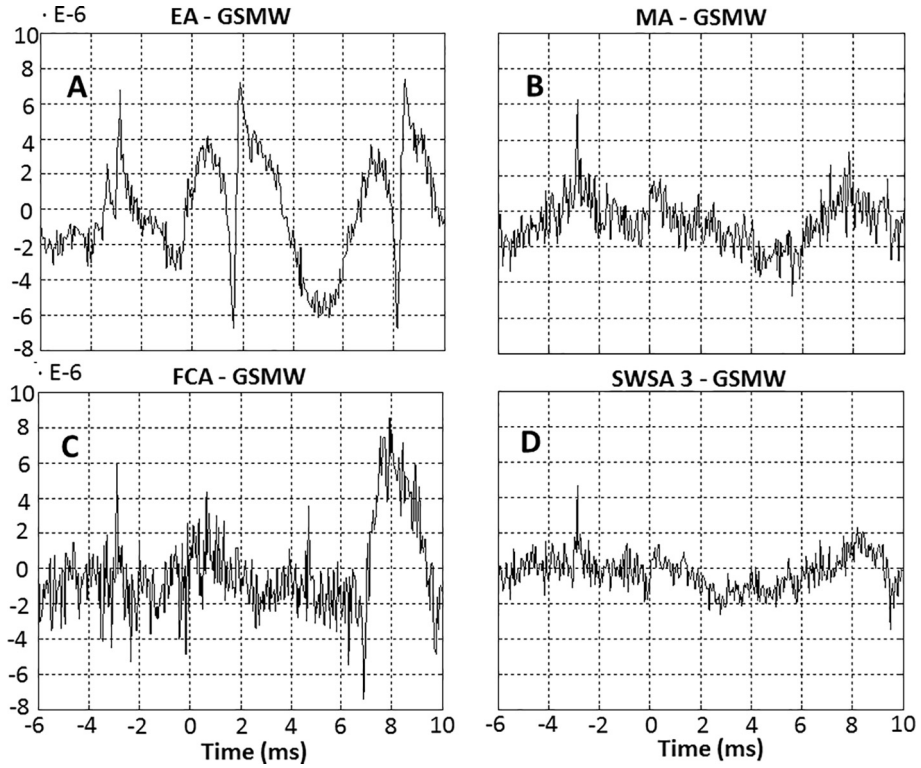

Fig. 6. Error curves of some of the averaging methods (EA, MA, FCA and SWSA 3) with respect to the GSMW in the example of Fig. 5. Note the finer level of fluctuation in the SWSA 3 error curve as compared to the other curves.

potentials from other MUAP trains. If this explanation is valid, the performance of SWSA as compared to MA or EA is expected to improve as the conditions of noise and interfering potentials gets more challenging.

- SWSA 3 exhibited broad performance differences in the two studies. While, in our previous study, significantly larger errors than those of the BPM were observed for irregularity coefficient and for the three SPMFs, in the present study there was no cases in which errors were significantly larger. In fact, SWSA 3 was the worst of the three SWSA methods in the previous study, while it was the best of the three in the current study. The reason for this discrepancy could be related to the length of the sliding window in the SWSA procedure. With longer windows (SWSA 3), the appearance of interfering potentials is more likely than with shorter windows (SWSA 1 and SWSA 2), and the capacity of the method to get rid of these potentials is accordingly diminished. If this explanation is valid, window length can be used to fit the SWSA method to different signal characteristics, setting low values (2.5–7.5 ms) in scenarios with many interfering potentials (i.e., 'free-run' recordings with a moderate level of muscle contraction) and high values (12.5 ms or more) in less noisy scenarios.
- In the present study we analysed the behavior of the method when applied to MUAP trains with a varying number of potentials per train (i.e., from 10 to 80). The overall scores increased with the number of potentials per MUAP train for both SPMF and MWP coming to a steady state for 50 or 60 potentials per train. SWSA 2 and SWSA 3 obtained the best performance of all the methods, followed by SWSA 1 and MA. EA and FCA had the worst performance (Fig. 4). These trends were congruent with the results presented in our previous works (Malanda et al., 2015, 2016).
- Satellite potentials, which are parts of the MUAP that appear outside its main spike, are sometimes present in diseased muscles, and their detection is an important aspect of the overall EMG diagnostic process. Automatic detection of satellites requires specific algorithmic strategies that are beyond the scope of the present work, which only concerns the waveform within the MUAP main spike. However, once a satellite has been detected in a given MUAP, averaging algorithms such as the SWSA may well be applied to extract the waveform.
- Several pathological conditions related to neuropathies, particularly reinnervation processes, imply an irregular function of the action potential firing mechanism in the MU. Some of the fibres of the MU may not fire in a given MUAP ('blocking'), or the firing time may be slightly desynchronized with respect to other fibres in the MU ('jitter') (Stålberg and Falck, 1997). These two phenomena give rise to MUAP trains in which the shapes of potentials can vary appreciably, even in the absence of secondary potentials or important levels of noise. This behaviour in MUAP waveforms was called 'jiggle' by Stålberg and Sonoo (1994) and Campos et al. (2000). Jiggle would be expected to present a special challenge for the SWSA method, as this method is ultimately based on the repetitiveness of the waveforms in the MUAP train, and this repetitiveness is weakened in MUAPs with high jiggle. The current study does not address this interesting case, and success of the SWSA method at dealing with jiggle cannot be guaranteed. Further studies are needed in order to evaluate the performance of the SWSA approach in this difficult scenario.
- Extracting representative waveforms from MUAP trains is a fundamental part of the process of MUAP analysis and is particularly relevant for the computation of waveform parameters and diagnosis-oriented statistical classification. In this respect,

the results presented here suggest that SWSA, given its capacity to extract more accurate and reliable MUAP curves than other existing methods, can be of service in quantitative EMG.

5. Conclusions

- The recently proposed *Sliding-window selected averaging* (SWSA) method and several other state-of-the-art averaging methods have been compared in performance at composing representative waveforms from MUAP trains, using EMG signals from normal, myopathic, subacute neurogenic and chronic neurogenic muscles and different number of potentials per MUAP train. The SWSA approach had better performance than the rest of methods in terms of signal analysis-based merit figures and MUAP waveform parameters used in the clinical practice for the different subject groups and number of potentials per MUAP train tested.
- A relatively wide range of values for the SWSA method's window length (7.5–12.5 ms) rendered good performance. With the set of MUAP trains used in the current tests, a window length of 12.5 ms gave the best performance of the algorithm. For EMG signals with many interfering potentials (i.e., 'free run' recordings at moderate levels of muscle contraction) lower window values (2.5–7.5 ms) are expected to be more appropriate, making the SWSA method outperform even more the other averaging methods.
- The behavior of the SWSA method with unstable EMG signals (i.e., MUAP trains with large values of 'jiggle') has not been tested, and new studies are required to validate the method for use in this interesting case.
- The SWSA method provides clear advances in accuracy and reliability in the field of MUAP waveform extraction for quantitative EMG.

Conflicts of interest

None of the authors has potential conflicts of interest to be disclosed.

Acknowledgement

This work has been supported by the Spanish Ministerio de Economía y Competitividad, under the TEC2014-58947-R project.

References

Antoni L. AiD: computer programs for automatic EMG analysis. Sweden: Balinge; 1983.
Campos C, Malanda A, Gila L, Segura V, Lasanta MI, Artieda J. Quantification of jiggle in real electromyographic signals. Muscle Nerve 2000;23:1022–34.
Rodríguez-Carreño I, Gila-Useros L, Malanda-Trigueros A, García-Gurtubay I, Navallas-Irujo J, Rodríguez-Falces J. Application of a novel automatic duration method measurement based on the wavelet transform on pathological motor unit action potentials. Clin Neurophysiol 2010;121:1574–83.
Florestal JR, Mathieu PA, Malanda A. Automatic decomposition of intramuscular electromyographic signals. IEEE Trans Biomed Eng 2006;53:832–9.
Hoke M, Ross B, Wickesberg R, Lütkenhöner B. Weighted averaging–theory and application to electric response audiometry. Electroencephalogr Clin Neurophysiol 1984;57(5):484–9.
Kimura J. Routine clinical electromyography. In: Bertorini TE, editor. Clinical evaluation and diagnostic tests for neuromuscular disorders. Amsterdam: Butterworth Heinemann; 2002.
Leonowicz Z, Karvanen J, Shishkin SL. Trimmed estimators for robust averaging ofevent-related potentials. J Neurosci Methods 2005;142(1):17–26.
Leski JM. Robust weighted averaging. IEEE Trans Biomed Eng 2002L;49(8):796–804.
Malanda A, Navallas J, Rodriguez-Falces J, Rodriguez Carreño I, Gila L. Averaging methods for extracting representative waveforms from motor unit action potential trains. J Electromyogr Kinesiol 2015;25(4):581–95.
Malanda A, Rodriguez-Carreño I, Navallas J, Rodriguez-Falces J, Porta S, Gila L. Sliding window averaging for the extraction of representative waveforms form motor unit action potential trains. Biomed Signal Process Control 2016;27:32–43.
Merletti R, Parker PA, editors. Electromyography, Physiology, Engineering and non invasive applications. Piscataway, NJ: IEEE Press Series in Biomedical Engineering; 2004.
Mühler R, von Specht H. Sorted averaging–principle and application to auditory brainstem responses. Scand Audiol 1999;28(3):145–9.
Nandedkar S, Sanders DB. Median averaging of electromyographic motor unit action potentials. Med Biol Eng Comput 1989;27:556–71.
Nandedkar SD, Barkhaus PE, Charles A. Multi-motor unit action potential analysis (MMA). Muscle Nerve 1995;18:1155–66.
Nandedkar SD, Barkhaus PE. Quantitative analysis: objective and interactive assessment in electromyography. In: Bertorini TE, editor. Clinical evaluation and diagnostic tests for neuromuscular disorders. Amsterdam: Butterworth Heinemann; 2002.
Navallas J, Malanda A, Gila L, Rodríguez J, Rodríguez I, Florestal JR, Mathieu PA. An algorithm for optimal discharge selection for MUAP waveform extraction. Torino: Congress of the ISEK; 2006.
Pfeiffer G, Kunz K. Turn and phase counts of individual motor unit potentials: correlation and reliability. Electroencephalogr Clin Neurophysiol 1992;85:161–5.
Rahne T, von Specht H, Mühler R. Sorted averaging–application to auditory event-related responses. J Neurosci Methods 2008;15:74–8.
Sörmo L, Lagunas P. Bioelectrical signal processing in cardiac and neurological applications. Elsevier Academic Press; 2005.
Stålberg E, Falck B, Sonoo M, Stålberg S, Aström M. Multi-MUP EMG analysis – a two year experience in daily clinical work. Electroencephalogr Clin Neurophysiol 1995;97(3):145–54.
Stålberg E, Antoni L. Computer aided EMG analysis. In: Desmedt JE, editor. Computer aided electromyography. Progress in clinical neurophysiology, vol. 10; 1983. p. 186–234.
Stålberg E, Nandedkar SD, Sanders DB, Falck B. Quantitative motor unit potential analysis. J Clin Neurophysiol 1996;13:401–22.
Stålberg EV, Andreassen S, Falck B, Lang H, Rosenfalck A, Trojaborg W. Quantitative analysis of individual motor unit potentials: a proposition for standarized terminology and criteria of measurement. J Clin Neurophysiol 1986;3:313–48.
Stålberg E, Sonoo M. Assessment of variability in the shape of the motor unit action potential, the "jiggle", at consecutive discharges. Muscle Nerve 1994;17:1135–44.
Stålberg E, Falck B. The role of electromyography in neurology. Electroencephalogr Clin Neurophysiol 1997;103:579–98.
Stewart CR, Nandedkar SD, Massey JM, Gilchrist JM, Barkhaus PE, Sanders DB. Evaluation of an automatic method of measuring features of motor unit action potentials. Muscle Nerve 1989;12(2):141–8.
Zalewska E, Hausmanowa-Petrusewicz I. Effectiveness of motor unit potentials classification using various parameters and indexes. Clin Neurophysiol 2000;111(8):1380–7.
Zalewska E, Hausmanowa-Petrusewicz I. Evaluation of MUAP shape irregularity – a new concept of quantification. IEEE Trans Biomed Eng 1995;42(6):616–20.