

Sliding window averaging for the extraction of representative waveforms from motor unit action potential trains



Armando Malanda^{a,*}, Ignacio Rodríguez-Carreño^b, Javier Navallas^a,
Javier Rodríguez-Falces^a, Sonia Porta^a, Luis Gila^c

^a Electrical and Electronics Engineering Dept., Public University of Navarre, Pamplona, Spain

^b Economics Dept., University of Navarre, Pamplona, Spain

^c Neurophysiology Service, Navarre Hospital Complex, Pamplona, Spain

ARTICLE INFO

Article history:

Received 30 July 2015

Received in revised form 1 December 2015

Accepted 7 January 2016

Available online 14 March 2016

Keywords:

EMG

Averaging

MUAP

Waveform

Sliding window

ABSTRACT

In quantitative electromyography (EMG), the set of potentials that constitute a motor unit action potential (MUAP) train are represented by a single waveform from which various parameters are determined in order to characterize the MUAP for diagnostic analysis. Several methods that extract such a waveform are currently available, and they are, in essence, based on two operations: averaging and selection, which are performed either sample-by-sample or on the whole-potential. We present a new approach that carries out selection and averaging on a local interval basis.

We tested our algorithm with a dataset of MUAP records extracted from the *tibialis anterioris* muscle of healthy subjects and compared it with some of the most relevant state-of-the-art methods considered in a previous work (Malanda et al., J. Electromyogr. Kinesiol., 2015). The comparison covered general purpose signal processing figures of merit and clinically used MUAP waveform parameters. Significantly better results in both sets of figures of merit were obtained with the new approach. In addition, relative to the other algorithms tested, the new approach required fewer potentials from the MUAP set to obtain an accurate representative waveform.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Analysis of the motor unit action potential (MUAP) plays a central role in clinical electromyography (EMG). For quantitative MUAP analysis, intramuscular EMG signals are recorded by means of needle electrodes inserted into the muscle belly. Several trains of MUAPs are usually present in these signals, and manual, semiautomatic or fully automatic procedures [1,2] are used for separating out these trains [20]. A representative waveform is then constructed from each of these trains in order to quantitatively characterize its main features with parameters that convey clinically useful information [1,3–5]. To this end, the potentials in the set are time-aligned and averaged. Alignment is usually carried out by

superposing the different potentials in the set, so that their maximum negative peaks or their triggering points coincide in time. Alternatively, they may be aligned on the basis of maximum correlation [6].

A number of averaging methods have been proposed to extract representative waveforms from repetitive biomedical potentials in the realm of EMG (i.e., MUAP analysis) [7–10], evoked potentials [11–14,19]; and electrocardiography [15]. A descriptive review of these averaging methods, including a comparative evaluation of nine of them with a bank of intramuscular EMG signals has been recently presented [20]. In that review the authors introduced four features to characterize and classify the averaging methods: *selection*, *weighting*, *observation scope* and *operation scope*. *Selection* refers to the way that the algorithm chooses which potentials in the MUAP set to use in calculating the average. *Weighting* refers to the weights that are given to selected potentials. *Scope* refers to the locality of the search around a given inspected sample. *Observation scope* refers to the set of samples around the inspected one from which the information needed for the *selection-weighting* process is extracted. *Operation scope* is the time interval around the inspected sample over which the *selection-weighting* criteria operate: the time interval for which the selected potentials and

Abbreviations: DEP, Derivative error power; 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; NEP, Normalized error power; NPM, Number of potentials per MUAP; REP, Residual error power; SLER, Significantly large errors range; SPMF, Signal processing merit figures; SWSA, Sliding window selective averaging.

* Corresponding author. Tel.: +34 948 169312; fax: +34 948 169720.

E-mail address: malanda@unavarra.es (A. Malanda).

<http://dx.doi.org/10.1016/j.bspc.2016.01.003>

1746-8094/© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

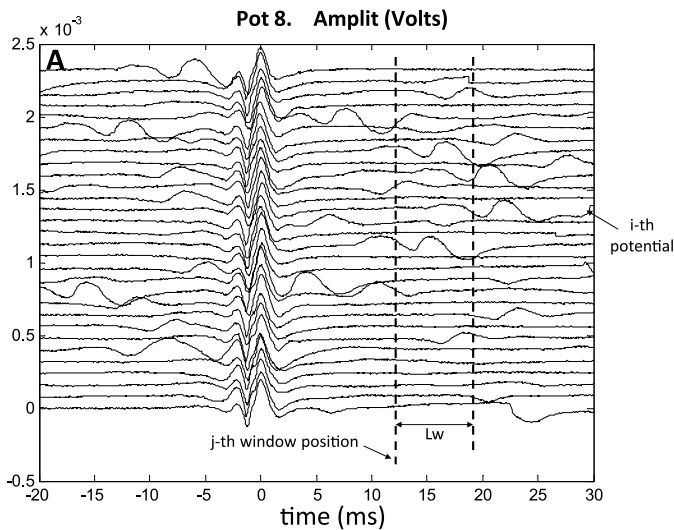


Fig. 1. MUAP potentials presented in raster mode. Sliding window for selection is shown.

the weighting coefficients remain unchanged. From the results of the analysis, *operation scope* turned out to be the most sensitive feature, and, in most of the evaluated cases, methods with one sample operation slope had better performance than those that operated on a 'whole potential' basis.

As discussed in the above-mentioned article, available averaging methods make use of one of two scopes: either one-sample scope or whole-potential scope. These two alternatives represent the extreme cases of what is normally known as local processing, in which, for processing a certain time sample, information from a limited neighbourhood of the sample is considered.

Here we present a new averaging method in which potentials are observed through a sliding window that traverses the time span of the whole potential and imposes an intermediate scope for the estimation process. The rationale behind this idea is that all the potentials in the MUAP set may contain useful information for composing the representative waveform; even if a certain potential is corrupted by one or more interfering potentials over some part of its time span, other parts may be unaffected and therefore valid for obtaining the representative waveform. Local processing provides a sensible strategy to materialize this idea, and a sliding window, a simple way to implement it (Fig. 1).

If several potentials within a MUAP set have similar shapes within a certain time interval, the common shape is more likely to be of physiological origin than due to noise or contamination. Therefore those potentials that share a common shape over a specific time interval should be the ones used to construct the shape of the representative waveform over this interval. For this reason, our algorithm selects and averages the most similar potentials within the scope of a sliding window. Once the selection and averaging process on one time interval has been completed, the algorithm slides the window along one time sample, delimiting the next interval to be analysed, from which a new set of potentials is selected and averaged. In view of these concepts, we refer to this approach as *Sliding-window selective averaging* (SWSA). In terms of the framework previously described, both the observation scope and the operational scope applied by SWSA are local, and the averaging process is based on selection of signal sections with similar shapes from several potentials in the set and uses uniform weighting of the selected potentials.

The aim of this paper is to present the SWSA approach and compare its performance with the most relevant of the

methods examined and evaluated in the above-mentioned descriptive review of averaging methods [20].

In the following section we describe the materials used in the study. Next, we give an account of SWSA as well as the methods we used to compare its performance. Then we explain the figures of merit and the gold standard used in comparisons before reporting the results of the comparative evaluation, discussing those results and offering our final conclusions.

2. Materials

The material used in this study was the same as that used in the work previously mentioned [20], with the expressed approval of the UPNA Ethical Committee. Particularly, 35 raw EMG signals were recorded from the *tibialis anterioris* muscle of seven healthy and physically active volunteer subjects who had given their informed consent before the experiments. These signals were 10 s-long and were taken while the subjects were exerting a slight to moderate muscle contraction, in the range of the current performance of signal recording with multi-MUP systems, i.e. 5–30% maximum voluntary contraction [9].

A Synergy electromyograph (Oxford Co.) and concentric needle electrodes (type DCN37; diameter 0.46 mm, recording area 0.07 mm²; Medtronic) were used for the acquisition. The EMG signals were band-pass filtered (filter setting was 3 Hz to 10 kHz), sampled (sampling rate was 20 kHz) and digitized (16-bits per sample). The digitized signals were stored on the hard disk of a PC and analyzed off-line. From these 35 EMG recordings, 175 MUAP trains were extracted using a recognized decomposition algorithm [16]; however, four MUAP train sets were lost as a result of file corruption. Each MUAP train consists of a set of potentials that have a fixed length L , sufficiently large that the waveform characteristics of the potential are fully contained within this length. In our case, $L = 1000$ samples (50 ms).

We discarded MUAP trains that were evaluated as unacceptable by an expert electromyographer for having an excessively noisy visual appearance or because they yielded average waveforms presented unrealistic MUAP shapes. We also discarded waveforms with a peak-to-peak amplitude lower than 0.1 mV. For SWSA sensitivity tests we included all MUAP trains with at least 40 potentials (Section 3.5). The number of MUAP trains that met the requirements for the sensitivity tests was 119. Because in our tests we wanted to measure the behaviour of the methods for different numbers of potentials in the train, we only included in the comparative tests those MUAP trains that had at least 80 potentials (Section 3.6). The number of MUAP trains that met the requirements for the comparative tests was 78.

3. Methods

3.1. Established averaging approaches used for comparison tests

Our method was compared to three different averaging methods for extracting representative MUAP waveforms from sets of potentials of MUAP trains. These methods were:

- Ensemble averaging (EA)
- Median averaging (MA) [8]
- Five-closest averaging (FCA): Average of the five potentials that are closest (as given by the Euclidean distance) to each other [20].

These methods were chosen from the nine methods used in the previously mentioned comparative study and review [20]. MA was the one that generally performed best. EA was not among the best

performers but was chosen because it is the simplest of the nine methods and is useful as a point of reference for other averaging methods. FCA generally had intermediate performance in terms of most of the evaluated parameters but in some specific scenarios had the highest scores and in others, the lowest. Besides the above considerations, these three methods were chosen because they use different observation and operation scopes and various selection and weighting modalities. Particularly, EA has one-sample observation scope, whole-potential operation scope, all-samples selection and uniform-weighting; MA, on the other hand, has one-sample observation scope, one-sample operation scope, one-sample selection and no weighting; finally, FCA has whole-potential observation and operation scopes, five samples selection and uniform weighting.

3.2. The Sliding-window selective averaging method

In our SWSA algorithm, all the potentials in the MUAP set are first aligned in time and amplitude axes. Then a sliding window is used to select sections of these potentials. For each time interval corresponding to the window location the so-called *median section* is calculated taking the median of the samples of all the sections in the train at that time interval. The potentials with shortest Euclidean distances to the *median section* are selected and averaged, obtaining a representative section for that time interval. Finally, the representative sections for each of the different time intervals obtained as the window slides along the whole potential span are assembled and averaged to form the final representative waveform.

This process is similar to what is done in Stalberg's MMA algorithm, in which the scope is one sample, whereas in the SWSA algorithm the scope is a local neighbourhood. Below is a formal and detailed description of the algorithm in four steps.

3.2.1. Alignment

Formally, the k th MUAP train is expressed as:

$$\text{MUAP}_k = \{x_{k,i}(n)\}, \quad i = 1, 2, \dots, N_k, \quad n = 0, 1, \dots, L - 1 \quad (1)$$

where i is the index that identifies the potential within the MUAP train, N_k is the number of potentials in the k th train and n is the index of the time sample. In the following, we will omit the index k because we are concerned with the generic case.

$$\text{MUAP} = \{x_i(n)\}, \quad i = 1, 2, \dots, N, \quad n = 0, 1, \dots, L - 1 \quad (2)$$

For MUAP visualization and processing, the potentials need be aligned, and here this is achieved by aligning the largest negative peak of the different potentials in the MUAP such that they coincide in the time axis. In addition, each potential may be affected by different baseline wander and secondary potentials that could increase or decrease its mean value in a random way. To accommodate the amplitude level of the potentials in the MUAP train we simply add a constant amplitude to each potential so that its Euclidean distance to the mean (i.e. the average of all the potentials in the MUAP train) is minimized.

3.2.2. Sliding window segmentation

Once the potentials are aligned, an L_w -samples-long window is slid along the time axis one sample at a time, delimiting segments of potential in the MUAP set (Fig. 1). The delimited segments are expressed as:

$$x_{ij}(n) = x_i(n+j), \quad 0 \leq n \leq L_w - 1, \quad j = 0, 1, \dots, J - 1 \quad (3)$$

where j is an index which accounts for the position of the sliding window, and J is the number of positions necessary for the sliding window to fully cover the time span of the potentials (i.e., $J = L - L_w + 1$).

3.2.3. Selection and averaging at each position

At each position j , the median segment of the set of segments is computed:

$$x_{j,\text{med}}(n) = \text{median}_{i=1,\dots,N} \{x_{ij}(n)\}, \quad 0 \leq n \leq L_w - 1 \quad (4)$$

The standard deviation (SD) of amplitudes of the potentials $x_{ij}(n)$ is also computed:

$$x_{j,\text{SD}}(n) = \text{SD}_{i=1,\dots,N} \{x_{ij}(n)\}, \quad 0 \leq n \leq L_w - 1 \quad (5)$$

And the minimum standard deviation throughout the window length is extracted:

$$X_{j,\text{min}} = \min_{n=0,\dots,L_w-1} \{x_{j,\text{SD}}(n)\} \quad (6)$$

Then, the distance between each signal segment $x_{ij}(n)$ and the median segment $x_{j,\text{med}}(n)$ is computed as the time average of the absolute value of the difference between these two segments:

$$\phi_{ij} = \text{avg}_{n=0,\dots,L_w-1} \{|x_{ij}(n) - x_{j,\text{med}}(n)|\} \quad (7)$$

Those segments that are more distant from the median segment $x_{j,\text{med}}(n)$ than $\eta \cdot X_{j,\text{min}}$ (η being a constant factor determined empirically) are excluded. The rest are considered valid and are averaged to form the representative potential at the j th position:

$$\text{ind.valid}_j = \{i = 1, \dots, N, \text{ such that } \phi_{ij} \leq \eta \cdot X_{j,\text{min}}\}, \\ j = 0, 1, \dots, J - 1 \quad (8)$$

$$y_j(n) = \text{avg}_{i \in \text{ind.valid}_j} \{x_{ij}(n)\}, \quad 0 \leq n \leq L_w - 1. \quad (9)$$

If the ind.valid_j set is empty or has only one element, the last step is iteratively repeated, increasing η by a certain amount each iteration, until this situation stops.

3.2.4. Composition of the full-length potential

At this point we have obtained a set of representative curves, $y_j(n)$, one at each position j . These now need to be assembled. First, they are placed in their corresponding time intervals, i.e., $y_j(n)$ should be delayed according to its position j :

$$y_{j,\text{del}}(n) = y_j(n - j), \quad j \leq n \leq j + L_w - 1, \quad j = 0, 1, \dots, J - 1 \quad (10)$$

where $y_{j,\text{del}}(n)$ is the delayed version of $y_j(n)$, which was obtained for the j th position.

To obtain the final representative curve, the overlapping parts of the segment curves are averaged. The number of segments that actually overlap on a given time sample is given by the following expression:

$$\xi(n) = \begin{cases} n + 1, & \text{if } 0 \leq n \leq L_w - 1 \\ L_w, & \text{if } L_w \leq n \leq L - L_w - 1 \\ L - n, & \text{if } L - L_w \leq n \leq L - 1 \end{cases} \quad (11)$$

Finally, the output representative potential is obtained from $y_{j,\text{del}}(n)$ and $\xi(n)$ as

$$\hat{y}(n) = \frac{1}{\xi(n)} \times \sum_{j=1}^J y_{j,\text{del}}(n), \quad 0 \leq n \leq L - 1 \quad (12)$$

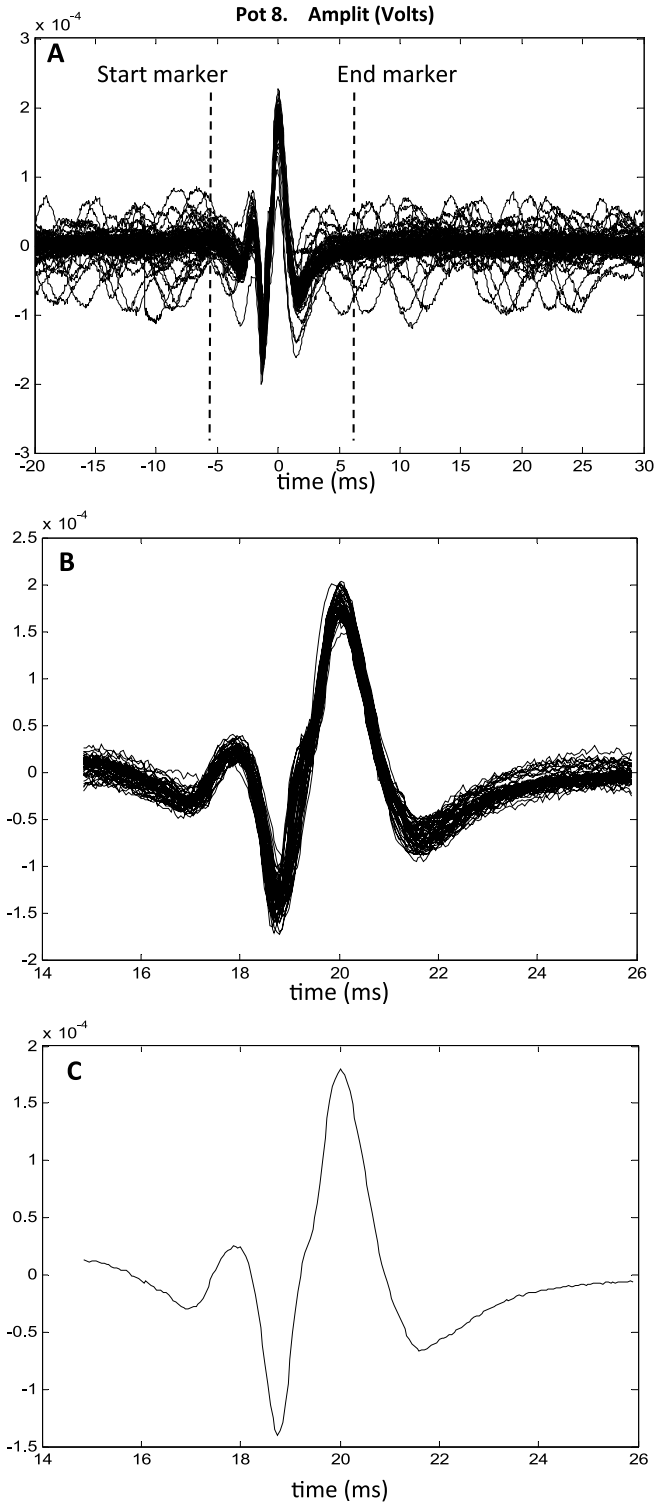


Fig. 2. (A) Ensemble of MUAP potentials and duration markers (vertical bars). (B) Selected MUAP potentials within duration limits. (C) GSMW obtained as the average of the selected potentials.

3.3. Gold standard

Gold standard MUAP waveforms (GSMWs) were used to compare quantitatively the characteristics of representative waveforms generated by different algorithms. GSMWs were obtained in a previous work [20] by time aligning the potentials in the trains, manually marking the MUAP initial and end points (Fig. 2A),

subjectively selecting a subset of potentials with similar shapes within the marked limits (Fig. 2B) and finally averaging the selected subset (Fig. 2C).

3.4. Merit figures

We used two different groups of figures of merit, also used in our previous study [20]. The first group is related to the concepts of noise and signal power in signal processing theory and are not specifically devised for the analysis of MUAP waveforms; we will refer to them as *signal processing merit figures* (SPMFs). These figures are intended to measure the similarity between the GSMW and waveforms generated by the tested algorithms. The second group of figures of merit is composed of descriptive parameters of MUAP waveforms that are typically used in quantitative EMG. The figures convey structural and physiological information about motor units and are of interest in clinical evaluation. We will refer to them as *MUAP waveform parameters* (MWPps).

3.4.1. Signal processing merit figures

We used the three figures of merit described below, that were also employed in [20].

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

$$\text{NEP} = \frac{\sum_{n_1}^{n_2} [e(n)]^2}{\sum_{n_1}^{n_2} [x(n)]^2} \quad (13)$$

where $x(n)$ is the representative MUAP potential obtained by a given method, $e(n) = x(n) - g(n)$ is what we refer to as the *error signal*, $g(n)$ being the GSMW curve, and n_1 and n_2 are the time samples corresponding to the initial and end duration cursors placed by the expert.

Residual error power (REP): signal power outside MUAP duration divided by signal power within MUAP duration:

$$\text{REP} = \frac{n_2 - n_1}{L + n_1 - n_2 - 1} \times \frac{\sum_{n=0}^{n_1-1} [x(n)]^2 + \sum_{n_2+1}^{L-1} [x(n)]^2}{\sum_{n=n_1}^{n_2-1} [x(n)]^2} \quad (14)$$

where L is the length of $x(n)$.

Derivative error power (DEP): error power of the potential derivative normalized to the power of the signal derivative within MUAP duration:

$$\text{DEP} = \frac{\sum_{n_1}^{n_2} [\delta_e(n)]^2}{\sum_{n_1}^{n_2} [\delta_x(n)]^2} \quad (15)$$

where $\delta_x(n)$ is an estimation of the time derivative of $x(n)$, which is obtained by filtering $x(n)$ first with a 4-tap averaging filter, $h_1(n)$, and then with a first order differentiator, $h_2(n)$. The impulse responses of these filters are, respectively:

$$h_1(n) = \begin{cases} 0.25 & \text{if } n = 0, \dots, 3, \\ 0 & \text{if } n < 0 \text{ or } n > 3 \end{cases} \quad (16)$$

$$h_2(n) = \begin{cases} 1 & \text{if } n = 0, \\ -1 & \text{if } n = 1, \\ 0 & \text{if } n < 0 \text{ or } n > 1 \end{cases} \quad (17)$$

Similarly $\delta_e(n)$ is an estimation of the time derivative of $e(n)$, which is also obtained by filtering $e(n)$ with $h_1(n)$ and $h_2(n)$, sequentially.

NEP gives a measure of the similarity of $x(n)$ to the GSMW curve within the time span of the potential: the more similar the two curves, the lower NEP. REP provides a measure of the power of the signal outside the potential duration limits, i.e., the baseline. The

lower REP the easier the estimation of duration and other MUAP parameters. Finally, DEP is concerned with the derivative of the MUAP representative curves within the MUAP duration. This figure of merit is sensitive to the presence of ragged (noise-like) contours in the waveforms and to the overall similarity between $x(n)$ and the GSMW curves.

3.4.2. MUAP waveform parameters

We used seven parameters that are widely accepted for MUAP waveform characterization within the realm of quantitative EMG [17,18] and that were also employed in [20]:

- Peak to peak amplitude (mV).
- Area of the rectified waveform within the MUAP duration (mV ms).
- Number of phases: the number of baseline crossings plus one within the MUAP duration. To make this measure robust against noise, a crossing was only considered to constitute a new phase if it exceeded a $20 \mu\text{V}$ threshold level above or below the baseline.
- Number of turns [3]: the number of positive and negative peaks separated from the preceding peak of opposite polarity by more than a certain amplitude threshold, $50 \mu\text{V}$ in this study.
- Irregularity coefficient [4]. This parameter measures the length of the MUAP curve normalized by the peak-to-peak amplitude and is calculated as the sum of consecutive absolute differences of the curve samples divided by the peak-to-peak amplitude.
- Spike duration [3]: the time interval between the first and the last positive peak of the MUAP (ms).
- Spike area [3]: the area of the rectified waveform within the spike duration (mV ms).

Error measurements were calculated as the difference between the MWP of the algorithm-generated waveform and the MWP of the GSMW. Relative error measurements were calculated dividing the absolute error measurements by the MWP of the GSMW.

In the case of the number of turns, we calculated two error measurements: missed and false turn errors. The former is a count of turns absent in an algorithm-generated waveform at instants when they are present in the GSMW. The latter is a count of turns present in an algorithm-generated waveform at time instants when they are not in the GSMW. To make these measurements robust against small waveform variations, misalignments and noise, we looked for temporal coincidence of a turn within a certain interval as opposed to within a single time sample. The interval used was 0.25 ms, which is equivalent to 5 samples in this study.

3.5. SWSA performance sensitivity

First we studied how the parameters L_w and η affect the behaviour of the SWSA method. To this end, we used the 119 available MUAP sets, taking the first 40 potentials of their trains. We varied L_w from 25 to 400 samples (1.25 to 20 ms) in steps of 25 samples (1.25 ms) and also included lengths of 15 and 20 samples (0.75 and 1.0 ms). For each value of L_w we varied η from 0.5 to 2.5 in steps of 0.5 and also included a value of 0.25. For each pair of L_w and η values, we calculated NEP, REP and DEP figures of merit.

3.6. Evaluation tests

Merit figure scores are affected by the number of potentials used to extract the representative waveform. Therefore, we performed our analysis varying the size of each MUAP set: we repeated our trials taking K potentials of each of the 78 MUAP sets, where K

was from 10 to 80 potentials in steps of 10. We called this variable the *number of potentials included in the MUAP set* (NPM). Mean and standard deviation (SD) of all merit figures, both SPMF and MWP, were calculated for each K value.

We observed that merit figure results varied depending on which K potentials were selected from the 80 potentials in the MUAP train (i.e., results might not be the same with the first ten as with the last ten). To eliminate this source of variability, for each K value and each MUAP, we repeated the previous evaluation procedure 50 times, each time selecting at random K potentials from the 80 available in the MUAP train. Then we obtained collections of 50 mean and 50 SD values of the SPMF and MWP, and finally averaged them.

In order to evaluate the significance of these results we proceeded as follows. For every merit figure and every NPM under analysis, we determined whether any of the methods yielded a mean absolute error significantly larger (using the paired t -test and P -values ≤ 0.01) than the minimum mean absolute errors obtained among the rest of tested methods. Then for each method we extracted the range of NPM values in which this situation (i.e. its mean absolute error was significantly larger than minimum mean absolute error) occurred. We will refer to this range as the *significantly large-errors range* (SLER). A large SLER value indicates that a method has a relatively low performance as compared with a method with a lower SLER value.

As representative waveforms generated by the algorithms vary with the number of potentials in the MUAP set, an interesting value for quantifying the behaviour of the algorithms is the NPM beyond which increases provide no gain in accuracy. The term *rate of convergence* will be used to refer to this concept. It was assumed that the representative waveform that an algorithm generates from the whole set of 80 potentials is the definitive output and that waveforms generated from subsets are approximations. In view of the above considerations, for each algorithm independently, we computed the P -values of merit figure score differences when $K=80$ potentials and the various trials when $K < 80$ potentials. Finally, we obtained the minimum NPM for which the yielded merit figures (SPMF or MWP error measurements) were not significantly larger (P -values ≤ 0.01) (as previously explained for the case of SLER measurements) than those obtained using 80 potentials per MUAP. We call this measure the *convergence factor*. A low *convergence factor* (10 or 20, for example) means that a method is fast, as it reaches the convergence within a few iterations. A high *convergence factor* (60 or 70, for example) means that a method takes longer to reach convergence.

4. Results

4.1. SWSA performance sensitivity

The results of our sensitivity analysis yielded specific ranges of values of parameters L_w and η for which the studied merit figures exhibited significantly better results than for other values, while no significant differences were observed within these ranges. These ranges are given in Table 1.

The intervals $150 \leq L_w \leq 175$ and $0.25 \leq \eta \leq 1.0$ define the best ranges of L_w and η values valid for most of the SPMF and MWP merit figures. On the basis of the above results, we decided to conduct our evaluation of the SWSA algorithm using three different values of the window length: $L_w = 50, 150$ and 250 samples (i.e., 2.5, 7.5 and 12.5 ms, respectively), and a value of 1.0 for the parameter η . These parameter values should enable the algorithm to minimize all of its SPMF and MWP errors, whilst maintaining some variability in the method. We will refer to these configurations of the algorithm as SWSA (1), (2) and (3), respectively.

Table 1

Ranges of values of L_w and η for which the studied merit figures exhibited significantly better results than for the rest, with respect to SPMF (rows 2–4) and MWP (rows 5–12) merit figures. L_{w1} , L_{w2} , η_1 and η_2 define the ranges of L_w and η respectively.

	L_{w1} (samples)	L_{w2} (samples)	η_1	η_2
NEP	125	225	0.25	1.0
DEP	50	175	0.25	0.25
REP	100	175	0.5	1.5
Amplitude	100	250	0.25	1.5
Area	125	350	0.25	1.0
Phases	20	375	0.25	1.0
Missing turns	150	250	0.25	1.0
False turns	50	300	0.25	1.0
Irreg. coef.	50	250	0.25	1.0
Spike duration	20	175	0.25	1.5
Spike area	15	400	0.25	1.0

Table 2

Ranges of significantly large errors (SLERs) of the SPMFs for the different averaging methods under study. (P -values ≤ 0.01). (r =range expressed in terms of NPM) (FR=full range) (None=zero range). (For the average of the last column, 'None' entries were substituted by 0s and 'FR' by '80', corresponding to the complete inspected range).

	NEP	REP	DEP	AVG
EA	FR	FR	FR	80
MA	FR	$r \leq 30$	FR	63.3
FCA	FR	FR	FR	80
SWSA (1)	$r \leq 40$	None	$r \leq 20$	20
SWSA (2)	None	FR	None	26.7
SWSA (3)	FR	FR	FR	80

Table 3

Convergence factor of the SPMFs obtained by the different averaging methods.

	NEP	REP	DEP	AVG
EA	80	80	80	80
MA	80	80	80	80
FCA	50	40	30	40
SWSA (1)	80	80	80	80
SWSA (2)	70	80	60	70
SWSA (3)	70	80	50	66.7
AVG	71.7	73.3	63.3	69.4

4.2. Signal processing figures of merit

Fig. 3 shows how NEP, REP and DEP merit figures varied with NPM, from 10 to 80 potentials, in steps of 10.

Table 2 presents the SLER values for the three SPMF and the six averaging methods studied. We include the SLER values averaged across the different merit figures (last column).

SWSA (2) significantly outperformed the rest of methods in terms of NEP and DEP in almost all NPM cases; SWSA (1) significantly outperformed the rest of methods in terms of REP in almost all NPM cases. From the numerical average of the three figures of merit, SWSA (1) and (2) are by far the best performing methods. SWSA (3), however, yielded sensibly worse results.

The rates of convergence of the studied methods, as indicated by convergence factor, were measured for NEP, REP and DEP (Table 3). FCA was the fastest method to converge (NPM between 30 and 50). The rest of methods, in most cases, only reached convergence with 70 or 80 potentials per MUAP.

4.3. MUAP waveform parameters

Table 4 presents the mean and SD of the MWP values calculated from the GSMW obtained from the collection of MUAP trains.

Correlation coefficients for GSMW parameter values obtained with 80 potentials per MUAP train are given in Table 5. There is

high correlation between amplitude, area and spike area. Note also the correlation between the number of turns and the irregularity coefficient and between the number of turns and the number of phases.

Fig. 4 shows the mean and SD of the differences in MWP values between representative curves obtained with the automatic methods and the GSMW, for NPM between 10 and 80.

Table 6 presents the SLER for the eight MWPs and the six methods studied. We include the SLER values averaged across the different merit figures (last column).

Regarding MUAP amplitude (Fig. 4A and B), error curves converged to non-zero results, i.e., they were all biased. SWSA (1), SWSA (2) and MA, curves converged to an amplitude error of around -1% ; SWSA (3), FCA, and EA converged to errors around -1.8% , 1.2% and -4% , respectively. As may be observed in Table 3, SWSA (1), SWSA (2) and MA performed best, with no significant differences among them. SWSA (3) and FCA came next, performing worse in some NPM cases. EA was significantly worse in all NPM cases. With regard to area (Fig. 4C and D), in many NPM cases, EA and FCA had significantly larger errors than the rest. All methods presented a certain amount of bias in this parameter. Error curves from SWSA (2) and SWSA (3) decreased below -1% with NPM above 50 and 70, respectively.

With respect to number of phases, turns in excess, irregularity, spike duration, and spike area (Fig. 4E and F, I–P), all methods presented similar decreasing error curves with NPM. The three SWSA versions yielded the best results in almost all cases. These results were significantly better than those of the other three tested methods for all the parameters and across most of the range of NPM considered. SWSA (1) and (2) were significantly better than SWSA (3) in many of these cases, as well.

With respect to missing turns (Fig. 4G and H), error curves presented a slightly different pattern. EA presented the largest errors, which were significantly larger than those of the best methods. The best results were provided by SWSA (1) with NPM lower than 70 and by FCA with NPM of 80.

The average SLER figures of Table 6 indicate that the three SWSA versions are clearly superior to the other methods regarding MUAP waveform parameters, with SWSA (1) and SWSA (2) performing better than the SWSA (3).

Table 7 presents the convergence factors, for the eight MUAP parameters and the six methods studied. Averages across methods and across MUAP parameters are also included in the last column and the last row, respectively.

The last column and row include averages across methods and across MUAP parameters, respectively. EA and MA provided convergence factors for false turns and irregularity considerably larger than those of the other methods. For the rest of parameters the six methods presented similar convergence rates. In any case, the three SWSA versions proved superior on average to the three other tested averaging methods.

5. Discussion

Several points arising from the foregoing analysis of the results deserve further attention:

- In terms of almost all of the MUAP waveform parameters and signal processing figures of merit studied, and for any given number of potentials included in the MUAP, SWSA performed better than the other algorithms (Figs. 3 and 4). In many cases, the significantly-large-errors-range of the SWSA was smaller than those of the other algorithms (Tables 2 and 6). After SWSA, the method with best performance was MA; FCA and EA were the most error-prone. Although each algorithm has

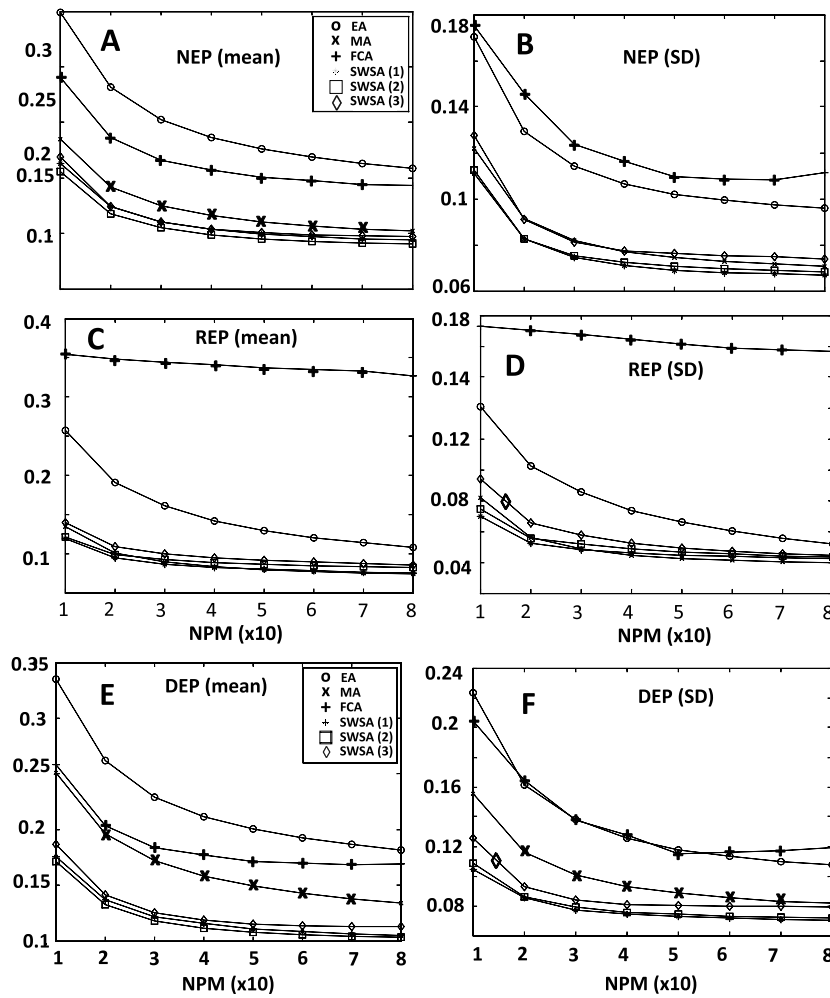


Fig. 3. Changes in SPMF results with the number of potentials per MUAP (mean in the left column and SD in the right column). (Normalized units are used in the y-axis).

Table 4 Mean and SD (in brackets) of MUAP parameter values obtained with manual measurements (GSMW).

Amplitude (mV)	Area (mV ms)	Phases	Turns	Irreg. coef.	Spike duration (ms)	Spike area (mV ms)
1.08 (0.69)	1.39 (0.78)	3.97 (1.16)	3.59 (1.32)	2.69 (0.40)	4.83 (2.61)	0.90 (0.56)

Table 5 Correlation coefficients between GSMW parameter values obtained with 80 potentials per MUAP train.

	Amplitude	Area	Phases	Turns	Irreg. coef.	Spike duration	Spike area
Amplitude	1.00	0.92	0.39	0.23	-0.15	-0.30	0.89
Area		1.00	0.35	0.21	-0.06	-0.26	0.95
Phases			1.00	0.57	0.39	-0.26	0.35
Turns				1.00	0.69	-0.19	0.29
Irreg					1.00	-0.07	-0.03
Spike duration						1.00	-0.12
Spike area							1.00

Table 6 Significantly-large-errors ranges of the MUAP waveform parameters for the different averaging methods under study (P-values ≤0.01). (r = range expressed in terms of NPM) (FR = full range) (None = zero range). (For the average of the last column, 'None' entries were substituted by 0s and 'FR' by '80', corresponding to the complete inspected range).

	Amplitude	Area	Phases	Missing turns	False turns	Irreg. coef.	Spike duration	Spike area	AVG
EA	FR	FR	r ≤ 60	FR	FR	FR	FR	FR	77.5
MA	None	r ≤ 50	r ≤ 40	r ≤ 60	FR	r ≤ 70	r = 10–50, 80	r ≤ 60	52.5
FCA	r = 80	r = 10–40, 70–80	FR	r ≤ 20	FR	FR	r ≤ 30	r ≤ 20	56.2
SWSA (1)	None	r = 20–30	None	None	None	None	r = 20	r = 20	5
SWSA (2)	None	None	None	None	None	None	None	None	0
SWSA (3)	r ≤ 20	None	r = 10	r = 10	r ≤ 40	r ≤ 50	r = 10	r = 10	18.7

Table 7
Convergence factors for the MUAP parameters obtained by the different averaging methods.

	Amplitude	Area	Phases	Missing turns	False turns	Irreg. coef.	Spike duration	Spike area	AVG
EA	10	20	70	30	80	80	30	20	42.5
MA	10	10	20	30	60	80	20	20	31.2
FCA	10	10	20	30	40	30	30	20	23.7
SWSA (1)	10	10	10	20	40	20	10	10	16.2
SWSA (2)	10	10	20	10	20	20	10	10	13.7
SWSA (3)	10	10	20	20	20	20	20	10	16.2
AVG	10	11.7	26.7	23.3	43.3	41.7	20	15	24.0

its own peculiarities, this order of performance may in part be explained by algorithm *operation scope*: SWSA, having local scope, had better performance than MA, which has one-sample scope; finally, FCA and EA, both having whole-potential scope, obtained the worst results. Contamination of MUAPs with secondary potentials (those coming from different motor units)

is expected to occur randomly along the potential time span. Implementing the selection-averaging process locally permits selecting parts of potentials in the MUAP train that are clean, while discarding other parts that are corrupted by interfering potentials. For example, if a given MUAP potential is corrupted by a secondary potential in its final portion but remains clean

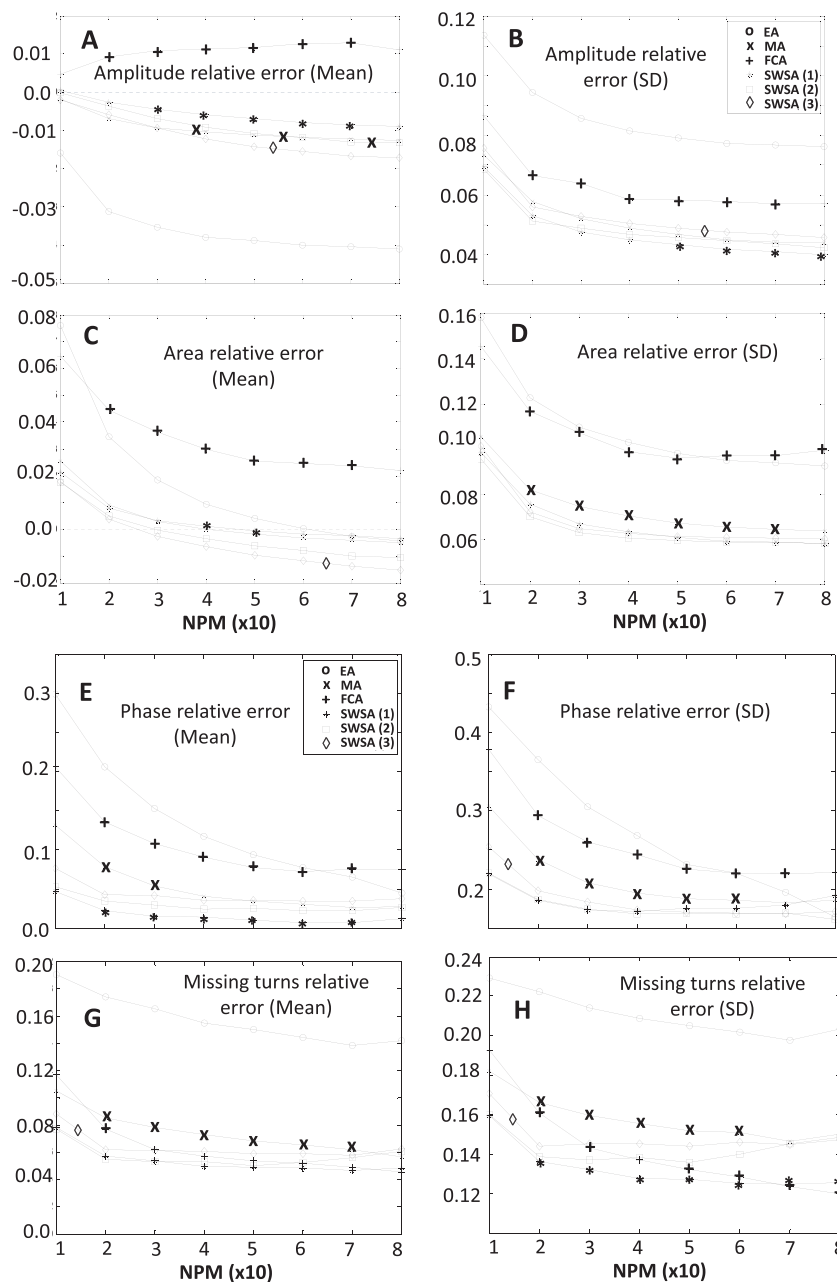


Fig. 4. Changes in MWP errors with the NPM (mean in the left column and SD in the right column). (Normalized units are used in the y-axis).

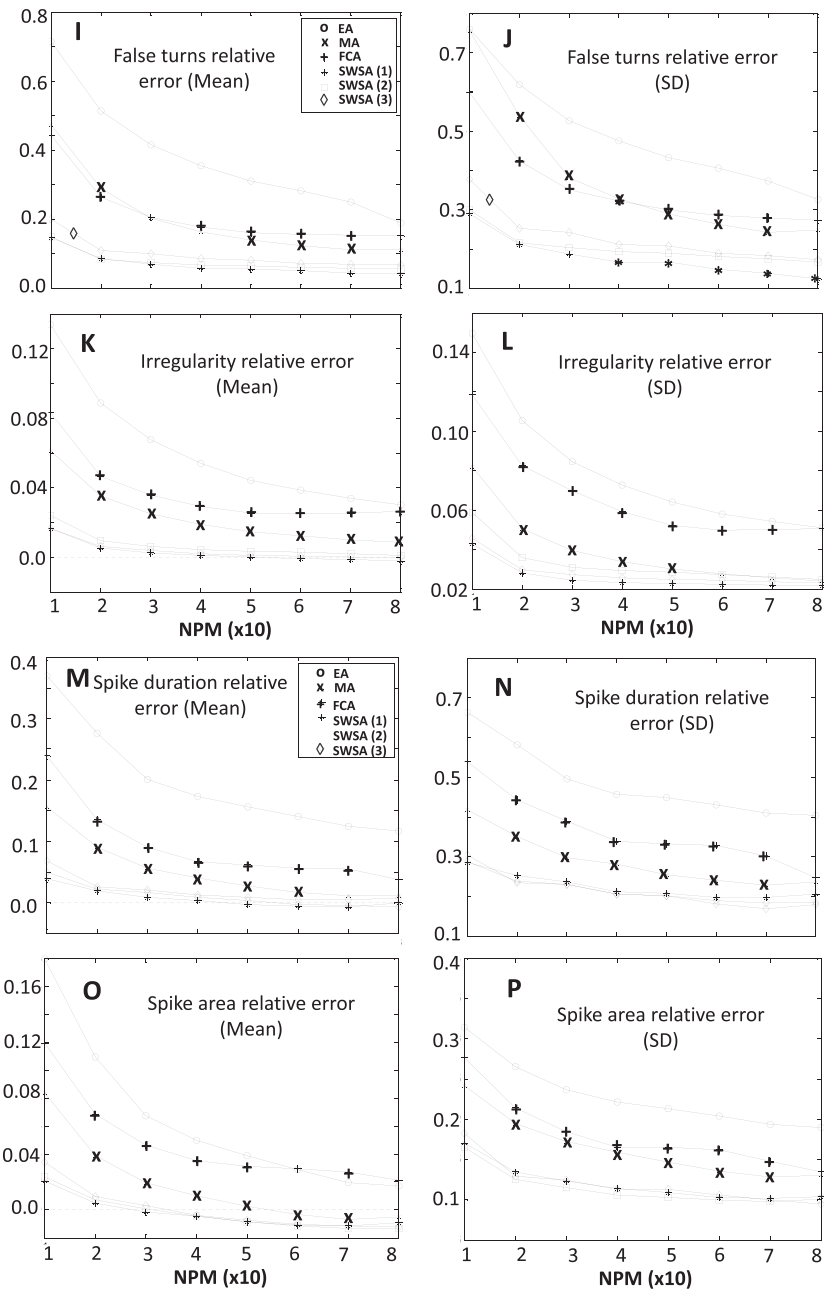


Fig. 4. (Continued).

elsewhere, local-scope algorithms, such as SWSA, may discard the corrupted portion and take into account the rest of the potential. On the other hand, whole-potential operators select, discard or weight the entire MUAP potential. In the given example they will not be able to use the clean portion of the potential, and at the same time reject its final corrupted portion. In this sense, local strategies seem better prepared than whole-potential operators for obtaining representative MUAP waveforms. At the other extreme, algorithms based on a scope of just one-sample neglect information in the waveform adjacent to this sample. Peaks (and therefore, turns) and growing and decreasing slopes are examples of local curve features that are disregarded by such algorithms. In contrast, local scope algorithms can use this local information. SWSA exploits local scope to look for similar curve segments among the set of MUAP potentials, and therefore curve features such as peaks or slopes are

implicitly considered in the process of selecting which segments to use to generate the averaged waveform at each position of the sliding window.

- (b) The study shows that all the averaging methods suffer from amplitude bias. (See Figs. 4A and Fig. 5). After convergence, the amplitude bias was around -4% for EA, 1.2% for FCA and from -0.9% to -1.8% for the rest of methods. One possible explanation for this would involve hypothesizing a certain selection bias in the manual procedure for extracting the GSMW. A natural tendency of the expert electromyographer to select potentials that, apart from being similar to each other, were also somewhat larger than the rest could explain the slight and negative bias in MA and SWSA (Fig. 5C–E). Apart from this, the larger negative bias in EA (Fig. 5C) could be explained as follows. Given that small misalignments of the potentials in the MUAP set are practically inevitable, as EA performs no selection or weighting

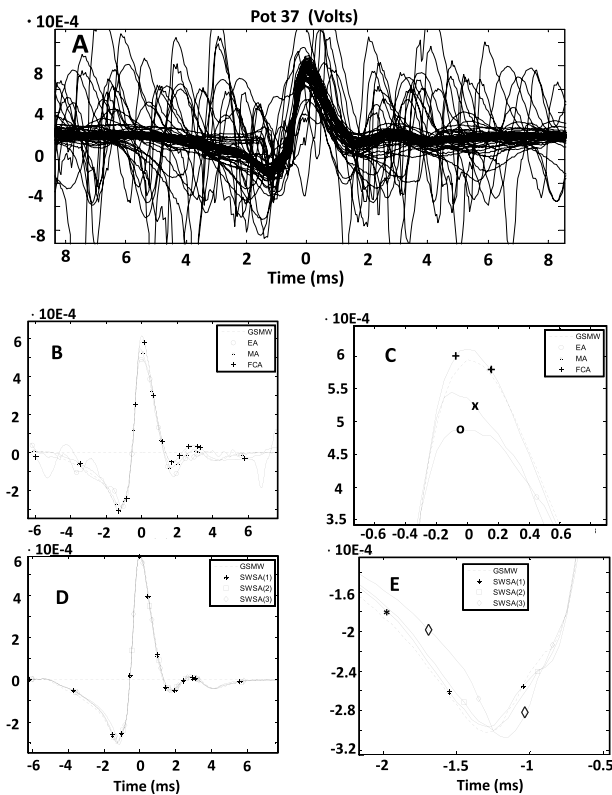


Fig. 5. Example of a MUAP set illustrating the different behaviour of the studied averaging methods with respect to MUAP amplitude. (See Section 5 point b for explanation).

of samples or potentials at all, misaligned curves cause a decrease in the averaged peak-to-peak amplitude that increases with the number of potentials used for the average. This is a well-documented phenomenon in the realm of curve averaging [13]. MA, FCA and SWSA, in contrast, implement some form of selection or weighting, either on a whole-curve or on a sample-by-sample scale, and consequently only a small fraction of the MUAP potentials has important weight in calculating the average. This also occurs with the manual extraction procedure performed to obtain the GSMW. This results in a smaller decrease in the amplitude of the average waveform yielded by MA, SWSA, FCA (and also by the GSMW), which explains the lower bias of MA, FCA and SWSA with respect to EA. Besides, the fraction of potentials involved in the averaging by MA, FCA and SWSA are probably more similar among themselves and to the GSMW than the rest of potentials in the MUAP train, which favours even more a lower bias in these methods, as compared to EA. Finally, the same curve misalignment phenomenon could account for the positive amplitude bias of FCA (Fig. 5C), as the number of potentials selected in this method is five, which in the majority of evaluated MUAPs is lower than the number of those manually selected for the GSMW. (In our study this number ranged from 3 to 88, with a mean of 21.0 and SD values of 16.6).

- (c) As MUAP area is strongly correlated to MUAP amplitude, the amplitude bias exhibited by the different averaging methods led to corresponding bias in the MUAP area (Fig. 4C).
- (d) Errors in the detection of turns reflect differences in the shape of the averaged waveform relative to the GSMW. False turns errors were most frequent with EA, followed by FCA and MA (Fig. 4I–J). False turns in EA are evident in the example of Fig. 6. The difference in the shape of an analysed curve with respect to a GSMW is also indicated by NEP, and, not surprisingly, EA, MA

and FCA were the methods with highest NEP values (Fig. 3A and B). False turn errors can also be caused by jagged curve profiles due to noise or to the averaging process itself, as occurs in the case of MA in the example of Fig. 6.

- (e) The irregularity coefficient measures the normalized length of the potential [4]. So, irregularity errors reflect discrepancy in the time course between averaged waveform and the GSMW. These errors are larger with EA, especially with noisy MUAPs (see Figs. 4K and L and Fig. 6). Jagged curves yielded by MA may also be associated with irregularity values higher than normal (Fig. 6).
- (f) Spike area is directly dependent on the potential amplitude and on the spike duration. MA and SWSA yielded negative amplitude errors upon convergence (Fig. 4A). Similarly, they produced negative spike area errors (Fig. 4O). In contrast, FCA had positive error values in spike area, which was in agreement with the positive errors in amplitude and spike duration (Fig. 4A and M). EA presented negative bias for the amplitude, positive bias for spike duration and positive bias in spike area. It seems in this case that spike duration has relatively more influence on spike area than the peak-to-peak amplitude.
- (g) The high correlation between MUAP waveform parameters (Table 5) can explain the relative similarity of some of the error curves, in particular, the error curves for area (Fig. 4C and D) and spike area (Fig. 4O and P); and the error curves for number of phases (Fig. 4E and F), false turns (Fig. 4I and J) and irregularity coefficient (Fig. 4K and L).
- (h) With respect to the MWP, convergence, as indicated by average convergence factor, was generally quickest for the three SWSA methods and slowest for EA (Table 7). However, with respect to SPFM, FCA was the fastest method to converge and all the rest converged very similarly (Table 3). No explanation can be given for this discrepancy in results.
- (i) Convergence rate differed according to the parameter evaluated, being generally quicker for MUAP amplitude, area and spike area (Table 7). Convergence was slower for the estimation of the number of turns and irregularity; presumably these two parameters demand a more detailed and precise representative waveform than that required for the rest.
- (j) When long windows are used in SWSA it is more likely that any segment of the MUAP train be affected by secondary potentials than when short windows are used. Therefore, the resulting averaged potential will be in general more corrupted by these interferences when long windows are used. However, very short windows will not benefit from the strategy of using information from a neighborhood of samples (a local interval), not just from a single sample, to decide how to calculate the average. These two issues seem to focus in opposite directions and a tradeoff should be found. From our results (Tables 2 and 6 and Figs. 3–4) it seems that windows with lengths between 50 and 150 samples (2.5 and 7.5 ms) are the best compromise.
- (k) Selection of factor η is also important. If η is very small, very frequently only one segment (i.e., the median segment) will be selected for the local neighbourhood, disregarding the benefit of averaging segments from several potentials; the process will tend to work as the MA method. On the other hand, if η is large, for every analysed sample the segments from almost all potentials will be included for the average, blocking the selection process; in this case SWSA resembles the EA method. Again a compromise should be found. Our performance sensitivity study indicates that a value of η in the interval (0.25, 1.0) is suitable for most of the figures of merit that we evaluated (Table 1).
- (l) The EMG data used in this comparative study was obtained only from normal subjects. Further studies are required to confirm that use of EMG data from patients with myopathic or neurogenic conditions would yield similar results and trends as those elucidated by our evaluations.

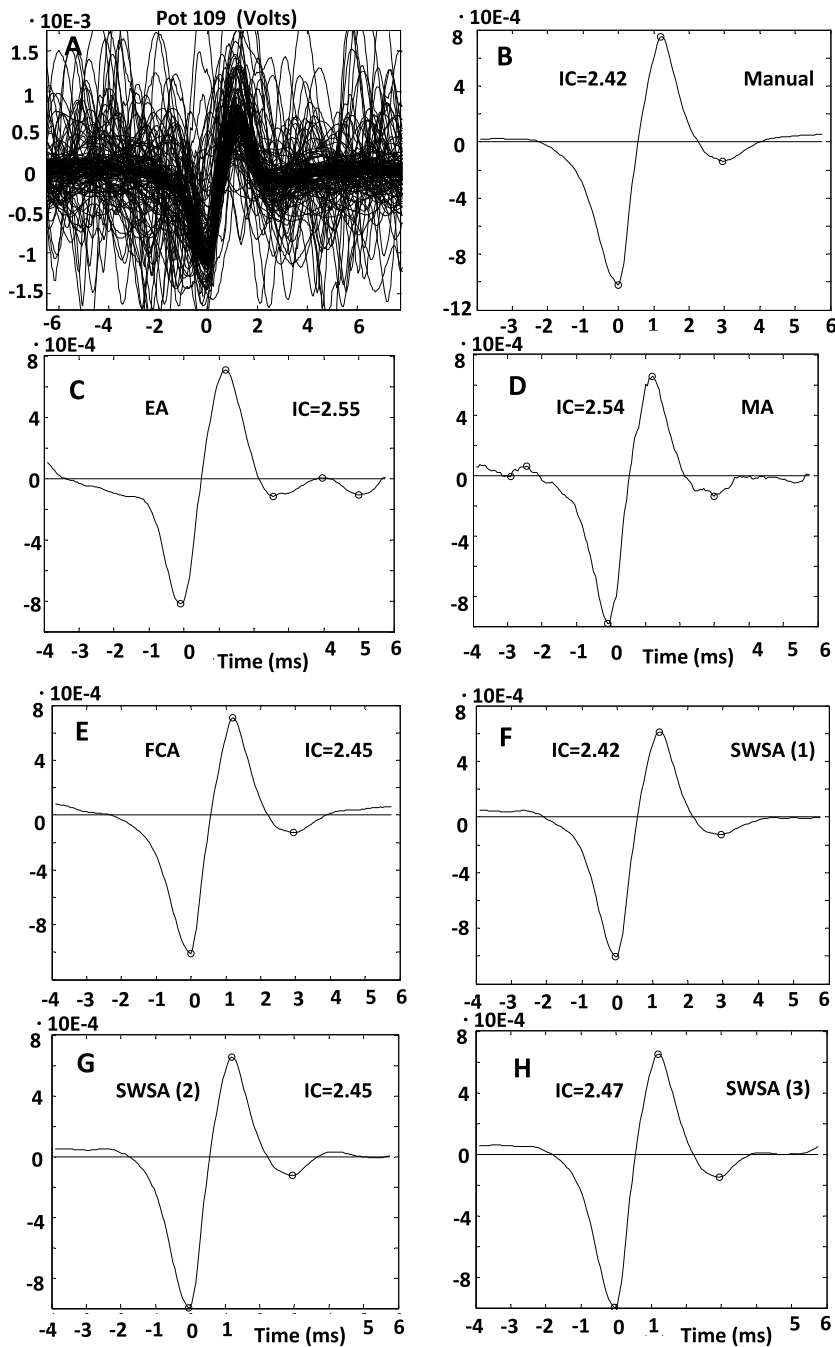


Fig. 6. Example of MUAP set illustrating the different behaviour of the studied averaging methods with respect to MUAP turns and the irregularity coefficient. (See Section 5 points d and e).

6. Conclusions

- 1- A new averaging algorithm for the extraction of representative waveforms from MUAP sets has been presented. The *Sliding-window selective averaging* is based on selection and averaging segments of potentials within the local scope of a sliding window.
- 2- In terms of several signal processing figures of merit and of various commonly-used quantitative MUAP waveform parameters, the new algorithm performed markedly better than three other relevant averaging algorithms at processing trains of real MUAP recordings.
- 3- With regard to MUAP waveform parameters, the new algorithm converged faster than the other three methods tested, that is,

it generally required a smaller number of MUAP potentials to attain maximal accuracy.

- 4- The best performance of the algorithm was found for a length of the sliding window in the range of 5 to 12.5 ms and for the internal parameter η in the range from 0.25 to 1.0.
- 5- Further studies are necessary to evaluate how the *Sliding-window selective averaging* algorithm performs with EMG signals from muscles with neuromuscular pathological dysfunctions.

Acknowledgement

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

References

- [1] S.D. Nandedkar, Quantitative analysis: objective and interactive assessment in electromyography, in: T.E. Bertonini (Ed.), *Clinical Evaluation and Diagnostic Tests for Neuromuscular Disorders*, Butterworth Heinemann, Amsterdam, 2002.
- [2] R. Merletti, P.A. Parker (Eds.), *Electromyography, Physiology, Engineering and Non Invasive Applications*, IEEE Press Series in Biomedical Engineering, Piscataway, NJ, 2004.
- [3] E.V. Stålberg, S. Andreassen, B. Falck, H. Lang, A. Rosenfalck, W. Trojaborg, Quantitative analysis of individual motor unit potentials: a proposition for standardized terminology and criteria of measurement, *J. Clin. Neurophysiol.* 3 (1986) 313–348.
- [4] E. Zalewska, I. Hausmanowa-Petrusewicz, *IEEE Trans. Biomed. Eng.* 42 (6) (1995) 616–620.
- [5] J. Kimura, Routine clinical electromyography, in: T.E. Bertonini (Ed.), *Clinical Evaluation and Diagnostic Tests for Neuromuscular Disorders*, Butterworth Heinemann, Amsterdam, 2002.
- [6] C. Campos, A. Malanda, L. Gila, V. Segura, M.I. Lasanta, J. Artieda, Quantification of jiggle in real electromyographic signals, *Muscle Nerve* 23 (2000) 1022–1034.
- [7] E. Stålberg, L. Antoni, Computer aided EMG analysis, in: J.E. Desmedt (Ed.), *Computer Aided Electromyography*. Progress in Clinical Neurophysiology, vol. 10, Karger, Basel, 1983, pp. 186–234.
- [8] S. Nandedkar, D.B. Sanders, Median averaging of electromyographic motor unit action potentials, *Med. Biol. Eng. Comput.* 27 (1989) 556–571.
- [9] E. Stålberg, B. Falck, M. Sonoo, S. Stålberg, M. Aström, Multi-MUP EMG analysis—a two year experience in daily clinical work, *Electroencephalogr. Clin. Neurophysiol.* 97 (Jun (3)) (1994) 145–154.
- [10] S.D. Nandedkar, P.E. Barkhaus, A. Charles, Multi-motor unit action potential analysis (MMA), *Muscle Nerve* 18 (10) (1995) 1155–1166.
- [11] M. Hoke, B. Ross, R. Wickesberg, B. Lütkenhöner, Weighted averaging—theory and application to electric response audiometry, *Electroencephalogr. Clin. Neurophysiol.* 57 (5) (1984) 484–489.
- [12] R. Mühler, H. von Specht, Sorted averaging—principle and application to auditory brainstem responses, *Scand. Audiol.* 28 (3) (1999) 145–149.
- [13] L. Sörmo, P. Lagunas, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*, Elsevier Academic Press, Amsterdam, 2005.
- [14] T. Rahne, H. von Specht, R. Mühler, Sorted averaging—application to auditory event-related responses, *J. Neurosci. Methods* 172 (Jul (1)) (2008) 74–78.
- [15] J.M. Leski, Robust weighted averaging, *IEEE Trans. Biomed. Eng.* 49 (8) (2002) 796–804.
- [16] J.R. Florestal, P.A. Mathieu, A. Malanda, Automatic decomposition of intramuscular electromyographic signals, *IEEE Trans. Biomed. Eng.* 53 (5) (2006) 832–839.
- [17] E. Stålberg, S.D. Nandedkar, D.B. Sanders, B. Falck, Quantitative motor unit potential analysis, *J. Clin. Neurophysiol.* 13 (1996) 401–422.
- [18] E. Zalewska, I. Hausmanowa-Petrusewicz, Effectiveness of motor unit potentials classification using various parameters and indexes, *Clin. Neurophysiol.* 111 (8) (2000) 1380–1387.
- [19] Z. Leonowicz, J. Karvanen, S.L. Shishkin, Trimmed estimators for robust averaging of 5 event-related potentials, *J. Neurosci. Methods* 142 (1) (2005) 17–26.
- [20] A. Malanda, J. Navallas, J. Rodriguez-Falces, I. Rodriguez Carreño, L. Gila, Averaging methods for extracting representative waveforms from motor unit action potential trains, *J. Electromyogr. Kinesiol.* 25 (4) (2015) 581–595.