



**HAL**  
open science

## Highly accurate real-time decomposition of single channel intramuscular EMG

Tianyi Yu, Konstantin Akhmadeev, Eric Le Carpentier, Yannick Aoustin,  
Dario Farina

► **To cite this version:**

Tianyi Yu, Konstantin Akhmadeev, Eric Le Carpentier, Yannick Aoustin, Dario Farina. Highly accurate real-time decomposition of single channel intramuscular EMG. *IEEE Transactions on Biomedical Engineering*, 2021, pp.1-1. 10.1109/TBME.2021.3104621 . hal-03353977

**HAL Id: hal-03353977**

**<https://hal.science/hal-03353977>**

Submitted on 24 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Highly accurate on-line decomposition of single channel intramuscular EMG

Tianyi Yu, Konstantin Akhmadeev, Eric Le Carpentier, Yannick Aoustin, Dario Farina, *Fellow, IEEE*

**Abstract—Objective:** Real-time intramuscular electromyography (iEMG) decomposition, as an identification procedure of individual motor neuron (MN) discharge timings from a streaming iEMG recording, has the potential to be used in human-machine interfacing. However, for these applications, the decomposition accuracy and speed of current approaches need to be improved. **Methods:** In our previous work, a real-time decomposition algorithm based on a Hidden Markov Model of EMG, using GPU-implemented Bayesian filter to estimate the spike trains of motor units (MU) and their action potentials (MUAPs), was proposed. In this paper, a substantially extended version of this algorithm that boosts the accuracy while maintaining real-time implementation, is introduced. Specifically, multiple heuristics that aim at resolving the problems leading to performance degradation, are applied to the original model. In addition, the recursive maximum likelihood (RML) estimator previously used to estimate the statistical parameters of the spike trains, is replaced by a linear regression (LR) estimator, which is computationally more efficient, in order to ensure real-time decomposition with the new heuristics. **Results:** The algorithm was validated using twenty-one experimental iEMG signals acquired from the tibialis anterior muscle of five subjects by fine wire electrodes. All signals were decomposed in real time. The decomposition accuracy depended on the level of muscle activation and was >90% when less than 10 MUs were identified, substantially exceeding previous real-time results. **Conclusion:** Single channel iEMG signals can be very accurately decomposed in real time with the proposed algorithm. **Significance:** The proposed highly accurate algorithm for single-channel iEMG decomposition has the potential of providing neural information on motor tasks for human interfacing.

**Index Terms—**Hidden Markov models, Bayes methods, Recursive estimation, Deconvolution, Electromyography decomposition, real-time decomposition, penalization.

## I. INTRODUCTION

**T**HE electromyogram (EMG) is the electrical expression of skeletal muscle fibers during a muscle contraction. This activity originates from the neural excitation of the motor neurons (MNs) in the spinal cord. The identification of individual MN discharge timings from EMG is termed EMG decomposition [1]. The information obtained by EMG decomposition has been classically used for the diagnosis of neuromuscular disorders [2], analysis of muscle architecture [3], investigation of central strategies for motor control [4], [5] as well as, more recently, for creating human-machine interfaces [6], [7], [8]. A highly accurate real-time decomposition can broaden the range of applicability of EMG processing, as well as support effectively its subsequent applications.

In the last few decades, several EMG decomposition algorithms [9], [10], [11], [12], [13], [14], [15], [16], [17], [18] have been proposed. However, most of them [9], [10], [11], [12], [13], [14], [15], either based on MUAP clustering and

TABLE I: Main notations of Hidden Markov Model (HMM) and Bayes filter

$Y$	iEMG signal
$H$	Vector of MU action potentials shapes
$\Omega$	Set of indexes of all MUs
$A$	Set of indexes of active MUs
$T$	Sawtooth sequences
$S$	Activation scenario
$W$	White noise
$\ell_{\text{IR}}$	Maximum MUAP length
$\Theta = [t_0, \beta]$	Vector containing the discrete Weibull distribution parameters: the location parameter and the concentration parameter
$t_{\text{R}}$	Shifting parameter of discrete Weibull distribution, that is the refractory period
$\text{Pr}$	Probability
w.p.	"with probability"
$Y[n]$	iEMG signal at time index $n$
$Y^n$	Vector containing the signal from time index 1 to $n$
$ ^n$	"... given $Y^n$ "
$\text{Pr}(T[n] = t[n])$	Probability of the sawtooth sequences at time index $n$ being equal to a value $t[n]$ . For all elements of the state vector, the uppercase symbols ( $T, A, S$ ) denote random variables, while the lowercase ones ( $t, a, s$ ) stand for their realizations
$n_{\text{path}}$	Number of decomposition paths (scenarios) in algorithm

template matching or blind source separation, work in an off-line manner. The algorithm proposed by Glaser et al. [16] is one of the few on-line decomposition algorithms and is applied to multichannel surface EMG signals. Another on-line decomposition method was proposed by Karimimehr et al. [18] for iEMG processing and was based on MUAP clustering and template matching. Nonetheless, this method lacked the resolution of temporal overlaps between action potentials.

In our recent work, a Bayesian filtering approach for single channel iEMG decomposition [19], following a strict time sequential approach (processing sample by sample), was firstly proposed. Moreover, a recruitment model for MUs was added [20] so that the algorithm could be adapted to a varying number of active MUs. Recently, this algorithm was implemented into a parallel computation framework [21] and achieved real time decomposition. However, the real-time constraints led to a decrease in accuracy with respect to the offline implementation. In this paper, this method is substantially extended to reach accuracy comparable to an offline implementation but maintaining online performance.

In the following, our previous method, including the Hidden Markov Model (HMM) of iEMG, a Bayesian filtering procedure estimating the parameters of MUs, and the acceleration

of the algorithm, will be briefly reviewed in Section II. The limitations of this approach that impact the accuracy when real-time constraints are posed, will also be analyzed in Section II. Then, in Section III, substantial changes are proposed in the algorithm with the aim to boost the accuracy. Further, experimental iEMG signals and an evaluation protocol will be described in Section IV. The decomposition results of experimental signals will be illustrated and analyzed in Section V.

## II. REVIEW AND ANALYSIS

In this section, a short overview of the previous proposed GPU-implemented on-line single channel iEMG decomposition algorithm, including the HMM of iEMG, the Bayes filtering, and the acceleration methods, will be provided. This description is necessary for the development of the new algorithm (Section III). Furthermore, the decomposition performance of the previous method will also be analyzed in order to point out its limitations.

### A. Hidden Markov Model

1) *Generation of the EMG signal*: The *motor unit* (MU), as the smallest functional unit of muscle contraction, comprises a MN and the muscle fibers innervated by its axon (muscle unit). When receiving input from spinal and supraspinal circuits, the MN generates trains of *action potentials* (spikes) that propagate along the MN axon and reach the muscle fibres through the neuromuscular junction, causing the fiber depolarisation. The excited muscle fibers generate a compound potential: the *motor unit action potential* (MUAP). Multiple MUs located nearby the electrodes contribute their MUAP trains to the overall interference EMG signal. Generally, the inter-spike intervals (ISI) reveal a certain regularity and have a physiological lower bound, called *refractory period*.

2) *Observation model of HMM*: Based on the EMG generation principles, a HMM of EMG was proposed in [19], [20]. The observation equation of HMM was derived from the linear model of EMG [22], [23]:

$$Y[n] = \sum_{i \in \Omega} \varphi_i(S[n]) H_i[n] + W[n] \quad (1)$$

- $n$  is the discrete time index;
- $i$  is the index of MU;
- $\Omega$  is the set of indexes of MUs, including both active and inactive ones;
- $Y$  denotes the observed iEMG signal;
- $S$  represents the activation scenario, comprising two elements:  $S[n] = (A[n], (T_j[n])_{j \in A[n]})$ .  $A[n]$  is the set of indexes of active MUs at time  $n$ ;  $(T_j[n])_{j \in A[n]}$  describes the time passed since its previous spike, for the active MU  $j$ ;
- $\varphi_i(s)$ , for each realization  $s = (a, (t_i)_{i \in a})$ , is a row vector of size  $\ell_{\text{IR}}$  with all components equal to zero; except, if  $i \in a$  and  $t_i < \ell_{\text{IR}}$ , the component in position  $t_i + 1$  is equal to 1.
- $H$  represents the MUAP waveform with finite length  $\ell_{\text{IR}}$ ;
- $W$  represents the independent identically distributed white noise sequence with unknown variance  $v$ ;

3) *State vectors of HMM*: The statistics of the ISI for each MU are modelled by the discrete Weibull distribution [19]. This distribution is defined by the vector  $\Theta_i[n]$  containing two parameters: a location parameter  $t_{0i}[n]$  and a shape parameter  $\beta_i[n]$ . Thus, the state vectors of HMM are as follows:

- $S[n] = (A[n], (T_i[n])_{i \in A[n]})$  the activation scenario,
- $H[n] = (H_i[n])_{i \in \Omega}$  the MUAP waveforms,
- $\Theta[n] = (\Theta_i[n])_{i \in \Omega}$  the inter-spike law parameters.

4) *Transition laws of HMM*:  $H_i[n]$  and  $\Theta_i[n]$  are firstly assumed to be constant in time. Their transition laws are as follows:

$$H_i[n+1] = H_i[n] \quad (2)$$

$$\Theta_i[n+1] = \Theta_i[n] \quad (3)$$

In practice, the steady changes of  $H_i[n]$  and  $\Theta_i[n]$  are considered in the Bayes filter (see II-B2).

The transition laws for  $S[n] = (A[n], (T_i[n])_{i \in A[n]})$  follows two transition models: the renewal model and the recruitment model, respectively for the two components  $T_i[n]$  and  $A[n]$ .

In the renewal model, for each  $i \in A[n+1] \cap A[n]$ , given  $\Theta_i[n]$ , we have the transition distribution of  $T_i[n]$ :

$$T_i[n+1] = \begin{cases} 0 & \text{w.p. } r(T_i[n] + 1, \Theta_i[n]) \\ T_i[n] + 1 & \text{w.p. } 1 - r(T_i[n] + 1, \Theta_i[n]) \end{cases} \quad (4)$$

where  $r(\cdot)$  is the hazard rate function of the Discrete Weibull distribution [24]. Moreover, we should notice that  $r(t, \Theta_i[n]) = 0$ , if  $t$  is lower than the *refractory period*  $t_{\text{R}}$ , due to the physiological restriction.

In the recruitment model, the recruitment mechanism is interpreted as the variation of  $A[n]$ , the set of indexes of active MUs at time  $n$ . Given  $S[n]$ , we have:

$$A[n+1] = \begin{cases} A[n] \setminus i & \text{w.p. } 1, \text{ if } T_i[n] = t_1 \\ A[n] \cup i & \text{w.p. } \frac{\lambda}{\text{card}(\bar{A}[n])}, \text{ if } i \notin A[n] \\ A[n] & \text{w.p. } 1 - \lambda \end{cases} \quad (5)$$

where  $\text{card}(\bar{A}[n])$  is the number of inactive MUs. An  $i$ -th active MU is derecruited when  $T_i[n]$  reaches a predefined limit  $t_1$ ; a random inactive MU is recruited with predefined constant probability  $\lambda$  and initialized with  $T[n] = 0$ . Thus,  $1 - \lambda$  is the probability if the active MUs keep the same as before.

### B. Bayes filter

1) *Principles and estimations*: With increasing time, the Bayes filter recursively estimates the state vector of the HMM defined in II-A3. The posterior probability functions of the state vectors are known:

- The probability density function (PDF) of  $\Theta[n]$  given  $S^n$ ,  $H$  and  $Y^n$ . The regularity of ISI only depends on the state vector  $S^n$ . Furthermore, with the assumption of independence in MU activity in [19], this PDF can be written as the product of  $\Theta_i[n]$  given  $S^n$ . The expected value of  $\Theta_i$  given  $S^n$ , noted  $\theta_{i,S^n}$ , is estimated by a recursive maximum likelihood (RML) estimation proposed in [20].

- The PDF of  $H[n]$  given  $S^n$  and  $Y^n$ . The expected value of this PDF, noted  $\hat{H}_{S^n}^n$ , is estimated by a least-mean-square (LMS) filter. And the observation prediction denoted as  $\hat{Y}_{S^n}^{n-1}$  and its variance denoted as  $v_{S^n}$  are also provided by the LMS filter. Details and mathematical derivation of this procedure can be found in [21].
- The probability mass function (PMF) of  $S^n$  given  $Y^n$ . As presented in [20], the PMF of scenario is derived from an update-prediction scheme. For all possible realizations  $s^n$  of  $S^n$ , the update step is:

$$\Pr^n(S^n = s^n) \propto \Pr^{n-1}(S^n = s^n) g(Y[n] - \hat{Y}_{S^n}^{n-1}, v_{S^n}) \quad (6)$$

where  $g(\cdot, v)$  is a zero-mean and variance  $v$  Gaussian PDF. And the prediction step is defined as:

$$\begin{aligned} \Pr^n(S^{n+1} = s^{n+1}) &= \Pr^n(S^n = s^n) \times \\ \Pr(A[n+1] = a[n+1] | S[n] = s[n]) &\times \\ \prod_{i \in A[n+1]} \Pr(T_i[n+1] = t_i[n+1] | S^n = s^n) &\quad (7) \end{aligned}$$

where  $\Pr(A[n+1] = a[n+1] | S[n])$  and  $\Pr(T_i[n+1] = t_i[n+1] | S^n)$  depends respectively on the recruitment model and the renewal model presented in section II-A4.

2) *Adaptivity*: In order to adapt to the two non-stationary state vectors: inter-spike laws parameters  $\Theta$  and impulse responses  $H$ , a sliding window method [25] was involved in the estimation of the two state vectors. The estimation are mainly based on the last  $\ell_\infty$  observations ( $Y[n - \ell_\infty + 1 : n]$ ) and scenarios ( $S[n - \ell_\infty + 1 : n]$ ), where  $\ell_\infty$  is the maximum window length. More details about the adaptive estimation formulas can be found in [21].

3) *Initialisation*: As presented in [21], no MUs are assumed to be active at the beginning of the decomposition. Thus, the set of active MUs indexes  $A[1]$  and the sawtooth sequence  $T[1]$  are initialized as empty vector. Initial rough MUAP waveforms  $\hat{H}_{S^1}^0$ , that can be manually or automatically extracted based on [10], [26], [27], are supposed to be known as prior knowledge. And the two inter-spike law parameters of active MUs  $\hat{\theta}_{i,S^0}$ :  $t_0$  is typically initialized as  $3t_r \sim 4t_r$ ;  $\beta$  is normally set to  $2 \sim 4$ . Finally,  $n_{\text{path}}$  initial  $S^1$  are weighted with the same initial probability  $\Pr^0(S^1 = s^1)$ .

### C. Acceleration

Theoretically, the state vectors of the HMM can be estimated recursively by the Bayes filter. However, this approach has a very high computational cost in practice. As described in the renewal model of section II-A4, for each  $i \in A[n+1] \cap A[n]$ , the sawtooth sequence  $T_i[n]$  bifurcates at each time index, if  $T_i[n] > t_r$ . This implies that the number of possible scenarios for  $S^n$  grows exponentially with time. Thus, an exhaustive estimation for all scenarios is impossible. Several approaches for path pruning were proposed in [21] to reduce the number of bifurcated scenarios:

- Limiting the number of kept paths is a conventional way to deal with the large number of scenarios. The  $n_{\text{path}}$  most probable scenarios are kept at every time index, where

$n_{\text{path}}$  is chosen as a trade-off between the computational cost and the decomposition performance.

- An iEMG signal is composed of multiple prominent action potentials and many time intervals only containing background noise. A function  $Z[n]$  was introduced to distinguish the two cases, in order to avoid bifurcations of  $T_i[n]$  in the segmentation of noise. For the segments containing action potentials, the on-line decomposition algorithm bifurcates active scenarios into multiple new ones and keeps the  $n_{\text{path}}$  most probable scenarios, while in the segments of background noise, only the state vectors of the kept scenarios are updated at each time index.
- The simultaneous occurrence of two or more spikes at exactly the same time instant was considered as extremely low probability and thus excluded from the possible scenarios. This reduces the maximal number of possible bifurcation from  $n_{\text{path}} \times 2^{\text{card}(A[n+1])}$  to  $n_{\text{path}} \times (\text{card}(A[n+1]) + 1)$ .

Furthermore, in order to achieve an on-line decomposition, the algorithm was implemented into Graphics Processing Unit (GPU) parallel computing.

### D. Problems with online decomposition

The previously proposed GPU-implemented on-line single channel iEMG decomposition algorithm [21] showed high decomposition accuracy ( $> 90\%$ ) for experimental signals with less than 7 active MUs, and large decrease in performance for a greater number of active units. The performance degradation was mainly caused by limiting the number of kept paths, which was needed for online implementation.

Limiting the number of kept paths implies retaining the  $n_{\text{path}}$  most probable scenarios at each time index and assuming that one of them is the correct decomposition result. However, this assumption is not always valid mainly because of the high similarity in MUAP waveforms, high similarity in kept scenarios, and over explanation of strong interference, such as noise and variation of MUAP waveforms.

1) *High similarity in MUAP waveforms*: Due to the lack of supplementary information from other channels, single channel decomposition algorithm switches occasionally between two similar MUAP waveforms.

2) *High similarity in scenarios*: We define highly similar scenarios, two scenarios that have the same active part in the last  $\ell_\infty$  time instants ( $S_i[n - \ell_\infty + 1 : n] = S_j[n - \ell_\infty + 1 : n]$ , where  $i$  and  $j$  are the indices of kept scenarios).

Based on the adaptivity of Bayes filter discussed in section II-B2, for any two highly similar scenarios  $i$  and  $j$ , we have:

$$\hat{H}_{S_i^n}^n \approx \hat{H}_{S_j^n}^n \quad (8)$$

$$\hat{\theta}_{i,S_i^n} \approx \hat{\theta}_{i,S_j^n} \quad (9)$$

which means that the two scenarios have almost the same estimated state vectors. Thus, they have great probability to provide the same results in the following decomposition, which will reduce the diversity of kept scenarios and the robustness of the decomposition algorithm.

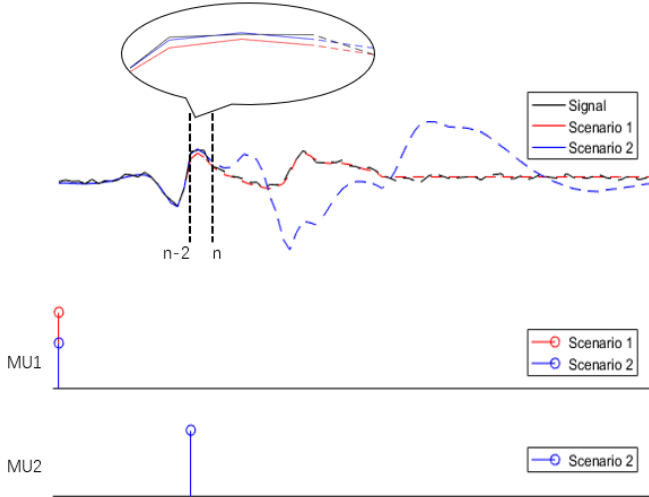


Fig. 1: Example of over explanation of noise. A signal was decomposed with two possible scenarios. Scenario 1 corresponds to the correct decomposition result, while Scenario 2 over explains the noise from time sample  $n - 2$  to  $n$  by introducing a firing for MU2.

3) *Over explanation of noise and variation of MUAP waveforms*: Since the on-line decomposition algorithm in [21] follows a strict time sequential manner, when acting on sampling points with high noise and variations in MUAP waveforms, it tends to find local optimal solutions, without considering the subsequent time samples. This phenomenon is termed over explanation.

In the example shown in Figure 1, a signal was decomposed with two possible scenarios. In the first one, only one MU fires at the beginning of the interval, while the second scenario represents the superposition of two MUs firing at close instants of time. The firing of the second MU at the time index  $n - 2$  in the second scenario is only introduced to over explain the noise from the time sample  $n - 2$  to  $n$ . However, when the signal is decomposed at time index  $n$ , without knowing the following observed signal, the second scenario could have bigger posterior probability than the first one, if they have the same prior probability.

In the decomposition of experimental signals with multiple MUs, this type of over explanation in form of superposition occurs frequently, which may result in the correct scenarios not being included in the  $n_{\text{path}}$  kept ones. The worst case occurs with multi-layers over explanation, where one MU firing over explains the noise, a second MU firing is used to correct the difference between the first MUAP waveform and the observed signal, and so on. This process can increase the number of errors substantially.

### III. HIGHLY ACCURATE ONLINE DECOMPOSITION

As shown in subsection II-D, there are substantial limitations in keeping scenarios only based on their posterior probability. In this section, we propose original approaches aimed at solving this problem and at keeping the decomposition speed within online limits.

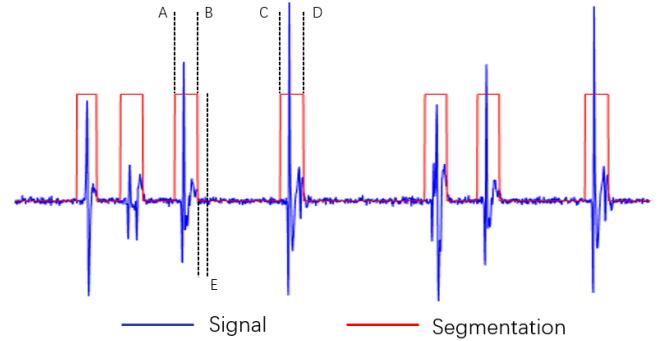


Fig. 2: Example of elimination of highly similar scenarios: in the segment AC, the high similar scenarios are eliminated at time index B, instead of checking and eliminating them at every time index. Example of the time instant for discarding scenarios with firing at the end of action potential segmentation: the scenario with a new firing at the end of segment AB will be discarded at time index E.

#### A. Limiting the inter-spike law parameters for highly similar MUs

Decomposition of the signal with highly similar MUAP waveforms is a common difficulty in single channel iEMG decomposition [28]. This can be partly counteracted by taking into account the regularity of firings in addition to MUAP waveform matching [21]. Here, we propose to limit one of the inter-spike law parameters  $\beta$  for MUs with similar MUAP shapes, in order to enhance the regularity in firing when MUAPs are very similar.  $\beta$  is typically limited to be more than 4.

#### B. Highly similar scenarios elimination

As described in subsection II-D2, highly similar kept scenarios reduce the diversity of the scenarios, leading to the decomposition being over susceptible to noise and variations in MUAP waveforms. Therefore, we propose to eliminate the highly similar kept scenarios with relative low posterior probability, and to only keep the one with highest probability.

However, an exhaustive elimination of similar scenarios at each time index is time consuming and inefficient for an on-line decomposition. According to the definition of high similarity of scenarios, we only need to make this elimination every  $\ell_{\infty}$  time sampling. More practically, we discard the similar scenarios at the last time index of each action potentials segment, whose length is much smaller than  $\ell_{\infty}$ .

Based on the measurement presented in subsection II-C, the iEMG signal can be divided into sequential action potentials and noise segments with a detection function  $Z[n]$ . The on-line decomposition algorithm in [21] only bifurcates scenarios and keeps the  $n_{\text{path}}$  most probable ones in action potentials segments. Thus, the elimination at the last time index of each action potentials segment does not only ensure the diversity of kept scenarios, but also avoids the unnecessary calculation of similar scenarios in the next noise segment.

As Figure 2 shows, an iEMG signal comprises several short prominent action potentials waveforms, such as segments AB

and CD, separated by segments of background noise, such as segment BC. In the segment AC, we only eliminate highly similar scenarios at time index B, instead of checking them at each time instant. In the noise segment BC, due to the elimination of similar scenarios, less kept scenarios need to be updated.

### C. Interdiction of spike firing in the end of action potential segmentation

Sometimes, the kept scenarios may over explain the noise at the end of action potential segmentation by introducing new firings. However, a spike firing at the end of an active segment is highly improbable since the corresponding MUAP waveform cannot match the following noise segment. Thus, this type of scenarios could be discarded at the beginning of the following noise segment.

For example, in Figure 2, in the action potential segment AB, a kept scenario fires a spike near to the instant B. From time instant B to E, due to the gradually increasing difference between the estimated signal and the observed noise, the posterior probability of this scenario will show an obvious decline. Thus, the low probable scenarios are discarded at time index E. Generally, the length of BE is set to  $\frac{\ell_{\text{IR}}}{5} \sim \frac{\ell_{\text{IR}}}{4}$ .

### D. Penalisation term of posterior probability

As presented in subsection II-D3, scenarios that over explain the observed sampling point containing strong interference by superimposing multiple MUAPs, always correspond to a locally high posterior probability, causing the possible exclusion of the correct decomposition result from the  $n_{\text{path}}$  kept scenarios. In order to resolve this issue, a penalisation term aiming at reducing local increase in posterior probability due to over explanation was added into the posterior probability formulation (6):

$$\Pr^n(S^n = s^n) \propto \Pr^{n-1}(S^n = s^n) g(Y[n] - \hat{Y}_{s^n}^{n-1}, v_{s^n}) \Pr_{\text{pl}}^n(S^n, \hat{H}_{S^{n-1}}^{n-1}, \hat{\theta}_{S^n}) \quad (10)$$

where  $\Pr_{\text{pl}}^n(S^n, \hat{H}_{S^{n-1}}^{n-1}, \hat{\theta}_{S^n})$  is the penalisation term, calculated by the observed signal  $Y[n]$ , the new bifurcated scenario  $s[n]$ , the estimated MUAP waveforms, and estimated inter-spike law parameters.

The over explanation is normally in form of superposition composed of multiple layers of MUAP waveforms. For each layer of superposition, we penalize its gain of probability. The penalisation term can be written as the product of each layer penalisation:

$$\Pr_{\text{pl}}^n(S^n, \hat{H}_{S^{n-1}}^{n-1}, \hat{\theta}_{S^n}) = \prod_{i \in \Upsilon[n]} \Pr_{\text{pl},i}^n(S^n, \hat{H}_{S^{n-1}}^{n-1}, \hat{\theta}_{i,S^n}) \quad (11)$$

where  $\Pr_{\text{pl},i}^n(S^n, \hat{H}_{S^{n-1}}^{n-1}, \hat{\theta}_{i,S^n})$  is the penalisation term of each superposition layer;  $\Upsilon[n]$  is the set of MUs superposition indexes.

An example of three layers of superposition is provided in Figure 3. Three MUs fire action potentials at the time indexes A, B, and C. Since the online decomposition algorithm

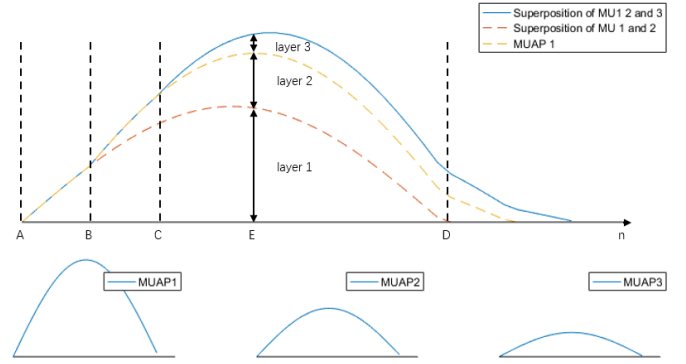


Fig. 3: Example of multiple layers superposition

respects a strict time sequential operation, the number of layers in the decomposition varies over time. if time  $n \in AB$ , Index  $\text{MU1} \in \Upsilon[n]$ ; if  $n \in BC$ , Indexes  $\text{MU1}$  and  $\text{MU2} \in \Upsilon[n]$ ; if  $n \in CD$ , Indexes  $\text{MU1}$ ,  $\text{MU2}$ , and  $\text{MU3} \in \Upsilon[n]$ . We define the order of MU indexes in  $\Upsilon[n]$  according to its spike firing time. Thus, if  $n \in CD$ ,  $\Upsilon[n] = \{1, 2, 3\}$

Theoretically, the penalisation term should restrain MUAP waveforms from over-explaining the interference, and, at the same time, does not influence the regular decomposition of superimposed action potentials. In order to reach this objective, for all  $i \in \Upsilon[n]$ , the formula of  $\Pr_{\text{pl},i}^n(S^n, \hat{H}_{S^{n-1}}^{n-1}, \hat{\theta}_{i,S^{n-1}})$  is defined as:

$$\Pr_{\text{pl},i}^n(S^n, \hat{H}_{S^{n-1}}^{n-1}, \hat{\theta}_{i,S^n}) = \frac{1}{\sqrt{2\pi}v_{s^n}} \exp(U_{\text{pl},i}^n(S^n, \hat{H}_{S^{n-1}}^{n-1})) N_{\text{pl},i} R_{\text{pl},i}(S^n, \hat{\theta}_{i,S^{n-1}}) \zeta_{\text{pl},i}(S^n) \quad (12)$$

Where  $U_{\text{pl},i}^n(S^n, \hat{H}_{S^{n-1}}^{n-1})$  is the function of penalisation unit;  $N_{\text{pl},i}$  represents the layer coefficient;  $R_{\text{pl},i}(S^n, \hat{\theta}_{i,S^{n-1}})$  denotes the regularity function of penalisation term; and  $\zeta_{\text{pl},i}(S^n)$  is the attenuation function. In the following, the four terms will be described.

1) *Function of penalisation unit:* According to the posterior probability formula (6), the local increase in posterior probability could be originated from the prior probability, or the likelihood of observation, or both of them. However, based on the prior probability function (7), the inter-spike law distribution, and the formula (4), the prior probability of the scenario firing a new spike will be smaller than scenarios without the firing of a new spike. Thus, the likelihood of the observation  $g(Y[n] - \hat{Y}_{s^n}^{n-1}, v_{s^n})$  contributes to the overall temporary gain of the posterior probability. The likelihood function is expressed as:

$$g(Y[n] - \hat{Y}_{s^n}^{n-1}, v_{s^n}) = \frac{1}{\sqrt{2\pi}v_{s^n}} \exp\left(-\frac{(Y[n] - \hat{Y}_{s^n}^{n-1})^2}{2v_{s^n}}\right) \hat{Y}_{s^n}^{n-1} = \psi(S[n]) \hat{H}_{S^{n-1}}^{n-1} \quad (13)$$

where  $\psi(s) = [\varphi_1(s), \dots, \varphi_{\text{card}(\Omega)}(s)]$ ,  $\text{card}(\Omega)$  denotes the number of MUs. The definition of  $\varphi_1(s)$  is given in subsection II-A2. The estimated observation of the signal can be also

written in form of multi-layers superposition:

$$\hat{Y}_{S^n}^{|n-1} = \sum_{i \in C[n]} \varphi_i(S_i[n]) \hat{H}_{i,S^{n-1}}^{|n-1} \quad (14)$$

With the formula (13) and (14), the contribution of each superposition layer for the posterior probability is clear. Thus, we have the penalisation unit of the  $i$ -th layer:

$$U_{pl,i}^{|n}(S^n, \hat{H}_{S^{n-1}}^{|n-1}) = - \left\| \frac{(Y[n] - \sum_{k \in C[n], k=1}^{k=i} \varphi_k(S_k[n]) \hat{H}_{k,S^{n-1}}^{|n-1})^2}{2v_{s^n}} - \frac{(Y[n] - \sum_{k \in C[n], k=1}^{k=i-1} \varphi_k(S_k[n]) \hat{H}_{k,S^{n-1}}^{|n-1})^2}{2v_{s^n}} \right\| \quad (15)$$

where  $\|\cdot\|$  is the operator of norm.

2) *Layer coefficient*: Generally, multi-layers superposition scenarios better explain the signal  $Y[n]$ . This is because the large number of layers in the superposition easily reduces the difference between the estimated and the observed signal. However, with the growing number of layers, the generated superpositions are less and less probable. Thus, we introduce a penalization term on the number of layers. The coefficient of layer is defined as:

$$N_{pl,i} = \Upsilon_{i,order} - 1 \quad (16)$$

where  $\Upsilon_{i,order}$  is the order of the  $i$ -th MU in the  $\Upsilon[n]$ . Since the first layer of superposition is not considered as the cause of over explanation, its coefficient is supposed to be zero.

3) *Regularity function*: We penalize the posterior probability of superposition to prevent them from over explaining the interference in the signal, under the condition of not influencing the normal decomposition of superimposed MUAP waveforms. Therefore, the regularity parameters of spike trains are involved into regulating the penalization terms for different cases. We have the regularity function of penalization term:

$$R_{pl,i}(S^n, \hat{\theta}_{i,S^{n-1}}) = 1 - \frac{\Pr(\Delta_i | \hat{\theta}_{i,S^{n-1}})}{\eta \Pr(\hat{t}_{0,i,S^{n-1}} | \hat{\theta}_{i,S^{n-1}})} \quad (17)$$

where  $\Pr(\Delta|\Theta)$  is the probability mass function of the ISIs distribution, which is the discrete Weibull distribution in our algorithm;  $\Delta_i$  is the length of last inter-spike of the  $i$ -th MU;  $\hat{t}_{0,i,S^{n-1}}$  is the estimated location parameter of discrete Weibull distribution, representing the location of most of ISIs in the distribution; and  $\eta$  is the coefficient of regularity function. The coefficient  $\eta$  should be more than one, typically set to two in our decomposition.

The value of  $R_{pl,i}(S^n, \hat{\theta}_{i,S^{n-1}})$  ranges from  $1 - \frac{1}{\eta}$  to 1. If  $\Delta_i$  is far from  $\hat{t}_{0,i,S^{n-1}}$ , meaning that the superposition is probably the over explanation,  $R_{pl,i}(S^n, \hat{\theta}_{i,S^{n-1}})$  approaches to 1; On the contrary, if  $\Delta_i$  is close to  $\hat{t}_{0,i,S^{n-1}}$ ,  $R_{pl,i}(S^n, \hat{\theta}_{i,S^{n-1}})$  tends to  $1 - \frac{1}{\eta}$ .

4) *Attenuation function*: The local increase in posterior probability of scenarios that over explain the signal mainly occurs at the beginning of new layer of superposition, named penalized section. The attenuation function is to ensure the penalisation term keep working in the penalized section and

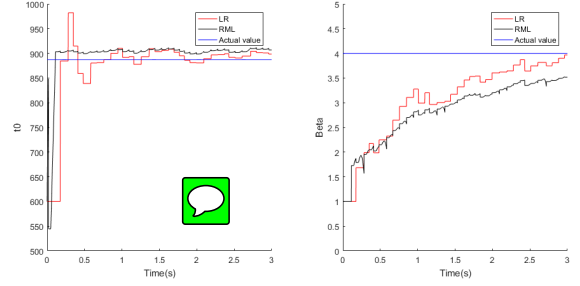


Fig. 4: Estimation of Weibull parameters with LR and RML estimator

reduce rapidly to one at the end of the section. We have the attenuation function:

$$\zeta_{pl,i}(S^n) = \begin{cases} 1 - \left(\frac{T_i[n]}{\rho_{i,pl}}\right)^3 & \text{if } T_i[n] \leq \rho_{i,pl} \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

where  $T_i[n]$  is the time passed since its previous spike of the  $i$ -th MU, defined in subsection II-A2;  $\rho_{i,pl}$  is the length of penalized section, defined as the round value of the ratio between the maximum absolute value of  $\hat{H}_{S^{n-1}}^{|n-1}$  and  $\hat{H}_{i,S^{n-1}}^{|n-1}$ , where  $\hat{H}_{S^{n-1}}^{|n-1}$  is the estimated MUAP waveforms of all MUs and  $\hat{H}_{i,S^{n-1}}^{|n-1}$  is the estimated MUAP of the  $i$ -th MU. Based on the decomposition results of previous algorithm, for each MU, the maximum absolute value of MUAP shapes is negatively correlated to the number of over explanation spikes. Therefore, the length of the penalized section is in an inverse ratio to the maximum absolute value of MUAP shapes.

### E. Estimation of inter-spike law parameters

As presented in subsection II-B1, the inter-spike law parameters were estimated by the RML estimator proposed in [20]. In the on-line estimator RML, the inter-spike law parameters were estimated at each time index with censored data, in order to provide accurate estimation for the calculation of prior probability in formula (7). However, compared to other estimation methods based on complete data, this method is time consuming due to the recursive estimation at each time index.

A number of algorithms have been proposed to estimate the discrete Weibull parameters with complete samples. Two methods proposed in [29], estimate parameters respectively based on the moments of samples and the proportion of samples. However, they present poor performance for a small number of samples. A maximum likelihood method proposed in [30] estimates recursively the parameters. When this method is implemented in the GPU parallel computation environment, it calls frequently the exponent function “double pow(double base, double exponent)” which requires high computational time. After comparing several estimation methods with inter-spikes of simulated iEMG signal (these results are not shown here), we selected a linear regression (LR) method proposed in [31]. This method models the relation of the logarithm of samples proportion and a function related to cumulative

density function, as well as weighing the regression residual based on Heteroscedastic Unreliability Weibull Estimation (HUWE). Compared to the maximum likelihood estimator, this LR method is more robust, and computationally more efficient. Details and mathematical derivation of this algorithm can be found in [31].

We compared the performance of RML and the selected LR estimator as follows. An active scenario  $S^n$  was generated with predefined inter-spike law parameters  $t_0$  and  $\beta$ , based on the renewal model and recruitment model in subsection II-A3. Given  $S^n$ , the two estimators were used to identify the parameters. Figure 4 shows the estimation results of the two estimators, which indicate that the LR estimator with complete data provides almost the same estimates as the RML estimator. The maximum sliding window length in RML was set to 0.6 s.

In the decomposition, instead of estimating inter-spike law parameters with RML estimator, the LR estimator was implemented at the end of every action potential segmentation decomposition.

#### IV. EXPERIMENTAL PROTOCOLS

##### A. Signals

Since the decomposition of simulated signals showed high performance in our previous work [21], we validated the new algorithm proposed in this paper only on experimental signals, which were more challenging.

Two sets of experimental iEMG were acquired from the tibialis anterior (TA) muscle. The first set, which is the one we used in [21], was recorded from a 26 years-old healthy man. In the recording, the subject performed five trials of an isometric force tracking task of 24 s duration. The force profile was trapezoidal with target force set to 20% or 30% of the maximal voluntary contraction (MVC). In this study, we extended the validation to a greater number of experimental signals. A second set of signals were acquired from four subjects (age range 20-25 years-old). Each subject was asked to perform four constant isometric force trials with target force respectively set to 5%, 10%, 15%, and 20% of MVC. The duration of each trail was 24 s. All signals in the two sets were recorded with wire electrodes made of Teflon coated stainless steel (50  $\mu$ m diameter; A-M Systems, Carlsborg, WA, USA) and 25G needles. Then, the signals were amplified, band-pass filtered between 100 Hz and 4 kHz, sampled at a frequency of 10kHz (OTBioelettronica MEBA amplifier), and subsequently down-sampled to 5 kHz.

The new proposed on-line iEMG decomposition algorithm was used to decode the two sets of experimental signals and the results were compared with those obtained in [21]. The activation probability  $\lambda$  and the maximum inactive time  $t_I$  of the recruitment model were respectively set to 0.03 and  $7t_R$ ; The maximum window length  $\ell_\infty$  was 1.4 s. The number of selected paths was set to 384.

##### B. Indexes of task and performance complexity

1) *Index of task complexity*: As presented in [20], the superposition percentage was used to characterise the complexity

of the decomposition task:

$$\text{Sup} = \frac{\text{Nb}_{\text{SUP}}}{\text{Nb}_{\text{SPIKES}}} \quad (19)$$

where  $\text{Nb}_{\text{SPIKES}}$  is the number of spikes in the signal and  $\text{Nb}_{\text{SUP}}$  is the number of spikes involved in the superposition of MUAP waveforms. We considered a MUAP superimposed if there was at least one other MUAP within a margin of 3 ms (less than half of the average MUAP duration) around it.

2) *Indexes of performance complexity*: Results of automatic decomposition were evaluated with the reference spike trains identified by manual decomposition. The manual decomposition was provided off-line by an expert operator using the publicly available decomposition software EMGLAB [32].

The global sensitivity and global positive predictivity were used to quantitatively evaluate the decomposition results. They were defined as follows: A MUAP was considered correctly identified (true positive) if the reference train contained a spike from the same MU within a margin of 1 ms around it. Consequently, the global sensitivity was defined as the overall number of correctly identified MUAPs from all MUs, divided by the overall number of spikes in the reference decomposition. Global positive predictivity was the number of correctly identified spikes divided by the overall number of spikes in the decomposition under evaluation.

Furthermore, an individual analysis of each MUAP train was also performed with "classification phase" indexes, including sensitivity, specificity and accuracy, as they are proposed in [33].

#### V. RESULTS

All experimental signals were decomposed by using the proposed algorithm programmed in C++ CUDA. All decompositions were performed on a Nvidia Tesla K80 GPU card with CUDA 9.0 and GCC 5.4.0 using single-precision floating-point format.

##### A. Decomposition performance

As shown in Table II, five experimental signals, including three recorded at 20% MVC and two recorded at 30% MVC, were decomposed respectively with the parallel decomposition algorithm (P.D.) proposed in [21] and the new algorithm (P.D.A.) proposed in this paper. For different experimental signals, the number of spikes ranged from 873 to 1295 and the percentage of superposition ranged from 18.10% to 28.96%. Compared to the decomposition results of the previous P.D. algorithm, the results of the new P.D.A. algorithm showed substantial increase in global predictivity (10%  $\sim$  15%) and a smaller but relevant increase in global sensitivity (3%  $\sim$  4%). Both the global sensitivity and predictivity of the five signals decomposed with the P.D.A. algorithm were above 90%. Furthermore, we noticed that the global sensitivity and predictivity decreased for signals with greater number of active MUs, due to the increase of decomposition task complexity, quantified by the superposition percentage.

It has to be noted that the results reported in Table II for the previous algorithm slightly differ from those reported in



TABLE II: Decomposition performance for the first set of experimental signals: 'Nb MUs' is the maximal number of MUs concurrently active in the signal; 'Nb sup-spikes' represents the number of spikes involved in superpositions; 'Nb spikes' denotes the overall number of spikes in the signal; 'Sup.' is the percentage of superposition; 'Algorithm: P.D.' represents using the parallel decomposition algorithm proposed in [21]; 'Algorithm: P.D.A.' denotes using the ameliorated parallel decomposition algorithm presented in this paper; 'Sens.' denotes the global sensitivity; 'Pred.' is the global predictivity.

Index	Duration (s)	Force (MVC%)	Nb MUs	Nb spikes	Sup.(%)	Algorithm: P.D.		Algorithm: P.D.A.	
						Sens. (%)	Pred.(%)	Sens. (%)	Pred.(%)
1.1	24	20	5	873	18.10	90.84	82.60	94.27	92.78
1.2	24	20	5	936	18.38	94.87	88.01	97.65	97.65
1.3	24	20	6	933	17.15	92.82	82.35	95.15	92.50
1.4	24	30	8	1282	28.39	86.04	76.70	90.48	91.41
1.5	24	30	8	1295	28.96	87.03	77.04	91.97	90.05

TABLE III: Decomposition performance for the second set of experimental signals. The meaning of indexes are the same as table II

Index	Duration (s)	Force (MVC%)	Nb MUs	Nb spikes	Sup.(%)	Algorithm: P.D.		Algorithm: P.D.A.	
						Sens. (%)	Pred.(%)	Sens. (%)	Pred.(%)
2.1	24	20	8	1269	17.89	85.50	79.14	93.30	91.29
2.2	24	15	7	903	24.58	83.50	86.77	98.12	98.23
2.3	24	10	7	746	10.86	95.31	87.24	98.93	98.40
2.4	24	5	3	354	4.24	96.33	90.69	98.87	97.77
3.1	24	20	7	890	13.03	89.21	81.44	94.27	93.85
3.2	24	15	5	664	10.39	84.04	73.91	91.11	88.71
3.3	24	10	5	503	9.47	91.65	83.06	91.65	91.11
3.4	24	5	3	229	4.35	97.32	88.45	97.66	94.50
4.1	24	20	11	1194	27.55	53.18	43.61	83.95	84.60
4.2	24	15	11	1040	24.04	51.44	37.18	88.19	88.12
4.3	24	10	9	1010	24.95	84.75	73.67	90.10	91.83
4.4	24	5	5	481	14.35	94.39	85.34	96.88	95.69
5.1	24	20	10	1510	23.31	78.35	66.69	82.05	84.75
5.2	24	15	8	993	13.49	89.53	75.53	94.66	91.89
5.3	24	10	5	617	8.75	95.53	82.35	98.87	99.35
5.4	24	5	3	353	3.12	99.44	76.70	100	100

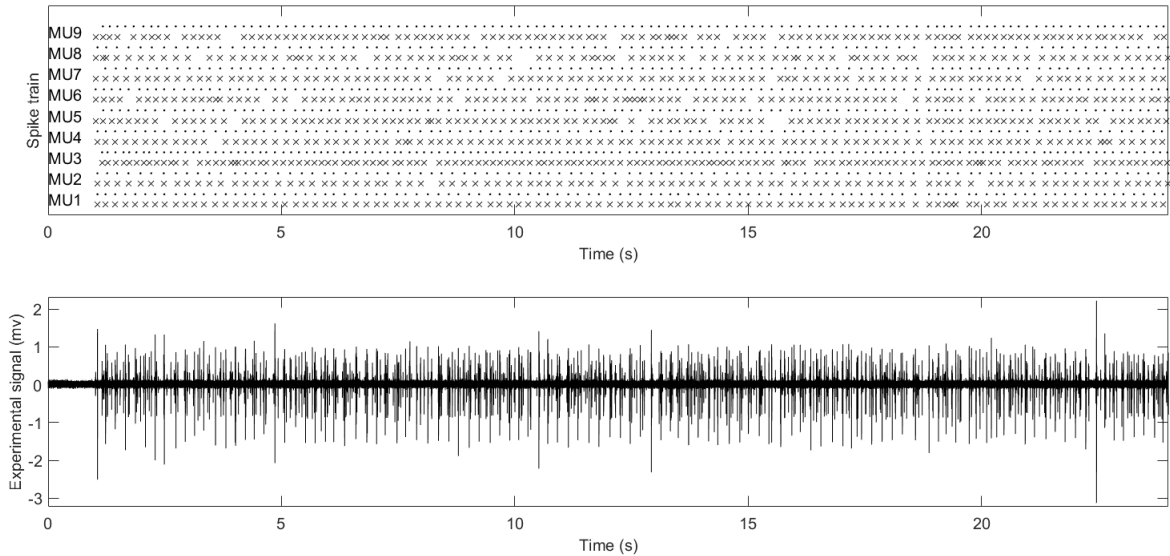


Fig. 5: Comparison of P.D.A. automatic (crosses, 'x') and reference (points, '.') decompositions (upper panel) and the experimental signal corresponding to index 4.3 in Table II (lower panel).

[21], since in the current paper we did not fine tune the lower limit for the inter-spike law parameters  $\beta$  (the same limits are set for the previous and the new algorithm).

Besides five experimental signals in Table II, 16 experimental signals were also automatically decomposed in order to validate the new P.D.A. algorithm. As shown in Table

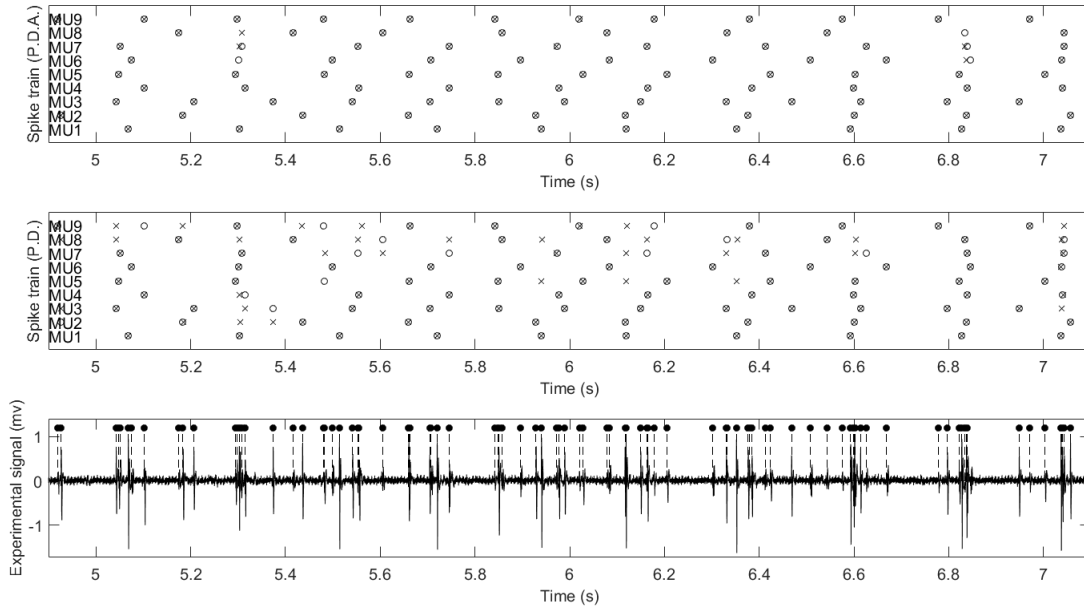


Fig. 6: An extract of the experimental signal decomposition shown in figure 5; circles ‘o’ and crosses ‘x’ represent respectively the spikes from the reference and automatic decompositions; ‘P.D.’ represents “using the parallel decomposition algorithm” proposed in [21]; ‘P.D.A.’ denotes “using the advanced parallel decomposition algorithm” presented in this paper.

III, the 16 signals were divided in four groups. Each group contained four signals recorded from the same subject at 5%, 10%, 15%, and 20% MVC. For signals in the same group, the number of active MUs, number of spikes, and the percentage of superposition increased with increasing force level. An exception is the percentage of superposition of signal 2.2. Based on the manual decomposition results of this signal, we found that the high correlation between its 1<sup>st</sup> MU and 2<sup>nd</sup> MU led to the abnormal high superposition percentage. For the 16 experimental signals, the number of MUs ranged from 3 to 11 and the percentage of superposition ranged from 3.12% to 27.55%. Since the force profile was constant, some of the signals showed a higher complexity compared to the ones in Table II.

For the decomposition performance of the 16 experimental signals in Table III, both the global sensitivity and predictivity of 13 of them with number of active MUs less than 10 were above 90%, while the global sensitivity and predictivity of the other three with ten or eleven active MUs were above 80%. Consistent with the decomposition performance shown in Table II, the performances of the new P.D.A. algorithm in Table III is much superior than for the previous P.D. algorithm. Moreover, we note that the relative increase in global sensitivity and predictivity becomes larger for signals at higher forces.

In the following, we analyse in detail the results of the decomposition for the signal 4.3 with 9 MUs, as a representative case.

A global view of the decomposition results is given in Figure 5. In the upper panel, nine decomposed spike trains

TABLE IV: Decomposition performance for the experimental signal decomposition shown in figure 5: for each MU, ‘Sens.’ denotes the sensitivity; ‘Pred.’ is the predictivity; ‘Acc.’ represents the accuracy.

MU	Sens.	Spec.	Acc.
MU1	99.07	99.75	99.67
MU2	92.93	99.39	98.70
MU3	92.96	98.98	98.06
MU4	94.34	99.14	98.59
MU5	91.38	99.38	98.38
MU6	89.83	99.14	97.95
MU7	82.35	97.98	96.30
MU8	80.21	97.54	95.79
MU9	86.29	98.89	97.22

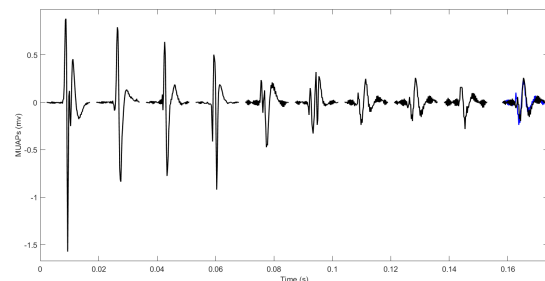


Fig. 7: Nine MUAP shapes (manually-extracted dictionary) for the signal presented in Figure 5, and a comparison between the 7th one and the 8th one.

(crosses) of each MU of P.D.A. are correlated with the reference results (points); in the lower panel, the decomposed

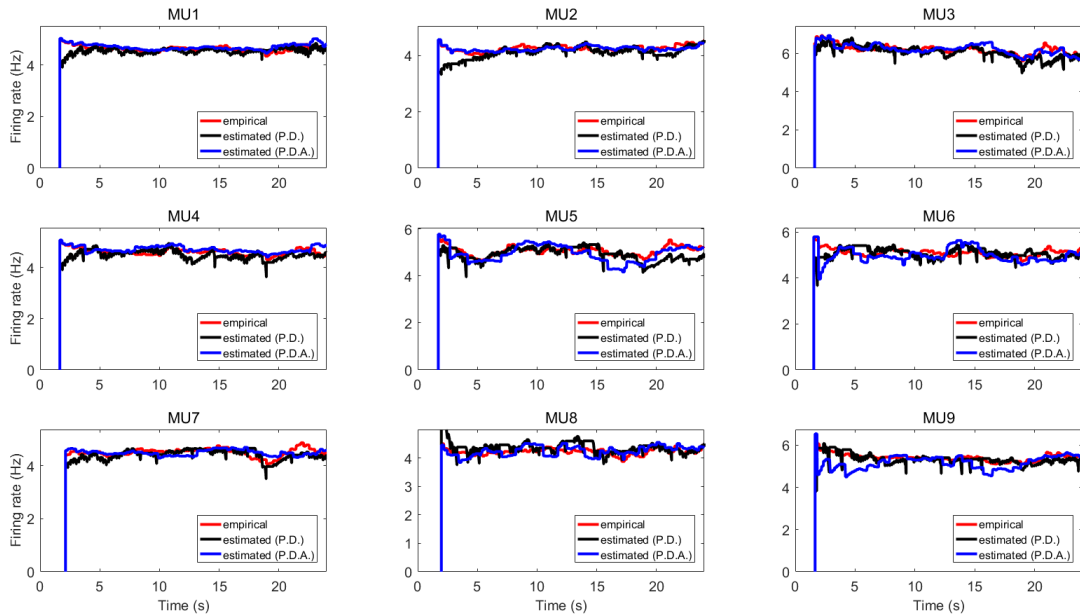


Fig. 8: Firing rates for the iEMG presented in figure 5: the red line (empirical) represents the firing rates estimated using reference decomposition; black line (estimated P.D.) represents the firing rates calculated via parameters of discrete Weibull distribution estimated with RML estimator [19], [20]; blue line (estimated P.D.A.) denotes the firing rates calculated via parameters of discrete Weibull distribution estimated with LR estimator (see section III-E).

iEMG signal is shown.

Figure 6 provides a detailed view of the decomposition results containing for approximately two seconds of recorded signal. The first panel shows the comparison between the identified spike timings (crosses) by P.D.A. and the reference results (circles); in the second panel, the identified spike timings by the P.D. algorithm correlate to the reference; the last panel shows the corresponding two seconds interval of recorded signal and the location of spikes in the reference result. Although two mistakes occurred at 5.3 s and 6.83 s, the P.D.A. algorithm performed generally well, successfully processing several complex superpositions (see first panel of Figure 6). However, the P.D. algorithm failed to process several superpositions containing small amplitude MUAP waveforms, where MUAP waveforms are illustrated in Figure 7. In addition, we note multiple problems in the decomposition with the P.D. algorithm, such as: over explanations at 5.05 s, 5.55 s, 5.95 s, 6.13 s, 6.36 s, and 6.60 s, and switches between the 7th MU and the 8th MU at 5.57 s, 5.60 s, and 5.74 s, were all corrected in the decomposition by the P.D.A. algorithm.

For the classification phase of P.D.A., Table IV shows the individual (per MU) performance indexes. Figure 7 shows the MUAP waveforms of nine MUs with a detailed comparison between MU 7 and 8. Based on Table IV and the Figure 7, we notice that the first four MUs were well classified, due to their larger MUAP waveforms, while the relative lower sensitivity for the 7<sup>th</sup> and 8<sup>th</sup> MU was caused by their smaller and similar MUAP waveforms, which may lead to the over explanation problem and switches with each other. Compared to the P.D. algorithm, the P.D.A. algorithm substantially improves

the decomposition performance, but still cannot eliminate all classification mistakes.

The algorithm estimates the parameters of the inter-spike intervals distribution, used to calculate the firing rates. The corresponding firing rates are illustrated in Figure 8. Empirical ones were estimated as the inverse of the moving average of subsequent inter-spike intervals in the reference decomposition. The estimated ones of P.D.A. were calculated with the parameters  $t_0$  and  $\beta$  estimated by the LR estimator. Based on the decomposed spike trains of P.D.A., we also estimated the parameters  $t_0$  and  $\beta$  with RML estimator used in P.D. and calculated its firing rates. The three highly coincident firing rates lines demonstrate that the replacement of inter-spike law parameters estimator has little influence on the decomposition performance.

### B. Decomposition time

Table V provides the execution time for all experimental signals for P.D.A. and P.D.. We notice that all signals, except the signal 5.1 with 10 MUs, could be decomposed in real time using 384 paths. The long decomposition time of signal 5.1 was due to the large maximum length of MUAP waveforms (around 10 ms), compared to others (6 ms ~ 8 ms). However, this signal could still be decomposed in real time with the P.D.A. algorithm using 256 paths. The execution time was 23.82 s. Moreover, we notice that the execution time of P.D.A. is almost the same as the one of P.D. for all signals. This is because in the P.D.A., time saved from the inter-spike law parameters estimation was used to regulate the diversity of scenarios and penalize the posterior probabilities.

TABLE V: Execution time for decomposition of experimental signals. The meaning of indexes are the same as table II

Index	Duration (s)	Force (MVC%)	Nb MUs	Nb spikes	Sup.(%)	Time(s)	
						Algorithm: P.D.	Algorithm: P.D.A.
1.1	24	20	5	873	18.10	18.04	17.80
1.2	24	20	5	936	18.38	18.31	17.85
1.3	24	20	6	933	17.15	17.36	16.16
1.4	24	30	8	1282	28.39	18.46	18.16
1.5	24	30	8	1295	28.96	19.53	18.81
2.1	24	20	8	1269	17.89	18.44	18.52
2.2	24	15	7	903	24.58	16.84	16.60
2.3	24	10	7	746	10.86	18.40	16.98
2.4	24	5	3	354	4.24	14.32	15.43
3.1	24	20	7	890	13.03	20.33	18.50
3.2	24	15	5	664	10.39	18.47	18.38
3.3	24	10	5	503	9.47	17.18	17.24
3.4	24	5	3	229	4.35	13.49	14.86
4.1	24	20	11	1194	27.55	22.50	22.84
4.2	24	15	11	1040	24.04	20.58	20.71
4.3	24	10	9	1010	24.95	19.27	17.97
4.4	24	5	5	481	14.35	18.26	17.33
5.1	24	20	10	1510	23.31	26.50	26.96
5.2	24	15	8	993	13.49	20.82	19.93
5.3	24	10	5	617	97.08	18.16	19.32
5.4	24	5	3	353	99.72	13.46	16.13

## VI. CONCLUSION AND PERSPECTIVES

In our previous works [19], [20], [21], a real-time decomposition algorithm based on a HMM of the EMG, using Bayesian filtering to estimate the unknown parameters of discharge series of MUs, and accelerated in the parallel computation environment, was proposed. Several iEMG signals with number of MUs up to ten were successfully decomposed; however, the decomposition should be under delicate parameter tuning and it showed relative low performance for experimental signals with number of active MUs greater than six.

In this paper we presented several methods to solve the main issues of the previous algorithm, including the replacement of time-consuming RML estimator by a more computationally efficient LR estimator for the inter-spike law parameters, and four heuristics. The proposed approach was validated with 21 experimental signals acquired from five subjects, thus substantially enlarging the validation dataset with respect to our previous work. The results showed that the new algorithm can provide much higher accuracy in decomposition in real time, compared to our previous work [21]. Moreover, the proposed algorithm could decompose experimental signals with a larger numbers of MUs than the real-time decomposition method proposed in [16]. For comparison, the real-time decomposition method proposed in [18] does not process superposed waveforms and thus provides much lower performance than the proposed one.

As indicated in the analysis of decomposition results, although this approach is able to improve greatly the decomposition performance, some mistakes still remain in the classification. The main reason is due to the limited information of single channel signals. Thus, a multi-channel version of the presented algorithm, which may further boost the decomposition accuracy, is our planned future objective. Another limitation is the relatively low decomposition speed for signals with a large number of active MUs, which restricts

the approach to relatively low contraction forces. Tracking large numbers of MUs requires a greater number of scenarios to ensure high accuracy, leading to a large computational complexity, which currently cannot be achieved in real time. This difficulty may be overcome by further simplification of our model or implementation in multi-GPUs.

## ACKNOWLEDGMENT

This work was supported by the European Research Council synergy project "Natural BionicS" (#810346).

## REFERENCES

- [1] B. Mambrito and C. D. Luca, "A Technique for the Detection, Decomposition and Analysis of the EMG Signal," *Electroencephalography and Clinical Neurophysiology*, vol. 58, pp. 175–188, 1984.
- [2] T. Kamali, R. Boostani, and H. Parsaei, "A multi-classifier approach to MUAP classification for diagnosis of neuromuscular disorders," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 22, pp. 191–200, 2014.
- [3] Z. C. Lateva, K. C. McGill, and M. E. Johanson, "The innervation and organization of motor units in a series-fibered human muscle: the brachioradialis," *European J. of Applied Physiology*, vol. 108, pp. 1530–1541, 2010.
- [4] C. J. D. Luca and Z. Erim, "Common drive of motor units in regulation of muscle force," *Trends in Neurosciences*, vol. 17, pp. 299–305, 1994.
- [5] A. Adam and C. J. D. Luca, "Recruitment order of motor units in human vastus lateralis muscle is maintained during fatiguing contractions," *Trends in Neurosciences*, vol. 90, pp. 2919–2927, 2003.
- [6] D. Farina and A. Holobar, "Human machine interfacing by decoding the surface electromyogram," *IEEE Signal Processing Magazine*, vol. 32, pp. 115–120, 2015.
- [7] F. Negro, S. Muceli, M. Castronovo, and D. Farina, "Multi-channel intramuscular and surface EMG decomposition by convolutive blind source separation," *J. of Neural Engineering*, vol. 13, no. 2, 2016.
- [8] D. Farina, I. Vujaklija, and M. Sartori, "Man/machine interface based on the discharge timings of spinal motor neurons after targeted muscle reinnervation," *Nature biomedical engineering*, vol. 1, 2017.
- [9] D. Ge, E. Le Carpentier, J. Idier, and D. Farina, "Spike sorting by stochastic simulation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 19, no. 3, pp. 249–259, 2011.
- [10] J. Florestal, P. A. Mathieu, and A. Malanda, "Automated Decomposition of Intramuscular Electromyographic Signals," *IEEE Trans. on Biomedical Engineering*, vol. 53, no. 5, pp. 832–839, 2006.

- [11] J. Florestal, P. Mathieu, and K. McGill, "Automatic decomposition of multichannel intramuscular EMG signals," *J. of Electromyography and Kinesiology*, vol. 19, pp. 1–9, 2009.
- [12] H. R. Marateb, S. Muceli, K. C. McGill, R. Merletti, and D. Farina, "Robust decomposition of single-channel intramuscular emg signals at low force levels," *J. of Neural Engineering*, vol. 8, no. 6, p. 066015, 2011.
- [13] S. H. Nawab, S.-S. Chang, and C. J. De Luca, "High-yield decomposition of surface EMG signals," *Clinical Neurophysiology*, vol. 121, no. 10, pp. 1602–1615, Oct. 2010.
- [14] F. Negro, S. Muceli, A. M. Castronovo, A. Holobar, and D. Farina, "Multi-channel intramuscular and surface emg decomposition by convolutive blind source separation," *J. of Neural Engineering*, vol. 13, no. 2, p. 026027, 2016.
- [15] J. Roussel, P. Ravier, and M. Haritopoulos, "Decomposition of Multi-Channel Intramuscular EMG Signals by Cyclostationary-Based Blind Source Separation," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 2035–2045, 2017.
- [16] V. Glaser, A. Holobar, and D. Zazula, "Real-Time Motor Unit Identification From High-Density Surface EMG," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 6, pp. 949–958, 2013.
- [17] A. Holobar and D. Zazula, "Multichannel blind source separation using convolution kernel compensation," *IEEE Trans. Signal Process*, pp. 55:4487–96, 2007.
- [18] S. Karimimehr, H. R. Marateb, S. Muceli, M. Mansourian, M. A. Mananas, and D. Farina, "A Real-Time Method for Decoding the Neural Drive to Muscles Using Single-Channel Intra-Muscular EMG Recordings," *Int. J. of Neural Systems*, vol. 27, no. 6, p. 1750025, 2017.
- [19] J. Monsifrot, E. Le Carpentier, Y. Aoustin, and D. Farina, "Sequential Decoding of Intramuscular EMG Signals via Estimation of a Markov Model," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 5, pp. 1030–40, 2014.
- [20] T. Yu, K. Akhmadeev, E. Le Carpentier, Y. Aoustin, R. Gross, Y. Péréon, and D. Farina, "Recursive decomposition of electromyographic signals with a varying number of active sources: Bayesian modeling and filtering," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 428–440, 2019.
- [21] T. Yu, K. Akhmadeev, É. Le Carpentier, Y. Aoustin, and D. Farina, "On-line recursive decomposition of intramuscular emg signals using gpu-implemented bayesian filtering," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 6, pp. 1806–1818, 2019.
- [22] D. Farina, A. Crosetti, and R. Merletti, "A model for the generation of synthetic intramuscular EMG signals to test decomposition algorithms," *IEEE Trans. on Biomedical Engineering*, vol. 48, no. 1, pp. 66–77, Jan. 2001.
- [23] D. Stashuk, "EMG signal decomposition: how can it be accomplished and used?" *J. of Electromyography and Kinesiology*, vol. 11, no. 3, pp. 151–173, 2001.
- [24] V. Barbu and N. Limnios, "Reliability theory for discrete-time semi-Markov systems," in *Semi-Markov Chains and Hidden Semi-Markov Models toward Applications*, ser. Lecture Notes in Statistics. Springer New York, 2008, vol. 191, pp. 1–30.
- [25] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Massachusetts and London: The MIT Press, 1983.
- [26] K. C. McGill, K. L. Cummins, and L. J. Dorfman, "Automatic decomposition of the clinical electromyogram," *IEEE Trans. on biomedical engineering*, vol. 32, no. 7, 1985.
- [27] C. Katsis, Y. Goletsis, A. Likas, D. Fotiadis, and I. Sarmas, "A novel method for automated EMG decomposition and MUAP classification," *Artificial Intelligence in Medicine*, vol. 37, pp. 55–64, 2006.
- [28] S. H. Nawab, R. P. Wotiz, and C. J. De Luca, "Decomposition of indwelling emg signals," *Journal of Applied Physiology*, vol. 105, no. 2, pp. 700–710, 2008.
- [29] M. Khan, A. Khaliq, and A. Abouammoh, "On estimating parameters in a discrete weibull distribution," *IEEE Trans. on Reliability*, vol. 38, no. 3, pp. 348–350, Aug. 1989.
- [30] K. Kulasekera, "Approximate MLE's of the parameters of a discrete Weibull distribution with Type-I censored data," *Microelectronics Reliability*, vol. 34, no. 7, pp. 1185–1188, 1994.
- [31] A. Fernandez and M. Vazquez, "Improved estimation of weibull parameters considering unreliability uncertainties," *IEEE Transactions on Reliability*, vol. 61, no. 1, pp. 32–40, 2011.
- [32] K. McGill, Z. Lateva, and H. Marateb, "EMGLAB: An interactive EMG decomposition program," *J. of Neuroscience Methods*, vol. 149, no. 2, pp. 121–133, 2005.
- [33] D. Farina, R. Colombo, R. Merletti, and H. B. Olsen, "Evaluation of intra-muscular EMG signal decomposition algorithms," *J. of Electromyography and Kinesiology*, vol. 11, pp. 175–187, 2001.