HAL
open science

Sequential Decoding of Intramuscular EMG Signals via Estimation of a Markov Model

Jonathan Monsifrot, Eric Le Carpentier, Yannick Aoustin, Darion Farina

To cite this version:

Jonathan Monsifrot, Eric Le Carpentier, Yannick Aoustin, Darion Farina. Sequential Decoding of Intramuscular EMG Signals via Estimation of a Markov Model. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2014, 22 (5), pp.1. 10.1109/TNSRE.2014.2316547 . hal-01024878

HAL Id: hal-01024878

https://hal.science/hal-01024878

Submitted on 16 Jul 2014

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.

Sequential Decoding of Intramuscular EMG Signals via Estimation of a Markov Model

Jonathan Monsifrot*, Eric Le Carpentier†,
Yannick Aoustin*, *Member, IEEE* and Dario Farina‡, *Senior Member, IEEE*

***Abstract*—This paper addresses the sequential decoding of intramuscular single-channel EMG signals to extract the activity of individual motor neurons. A hidden Markov model is derived from the physiological generation of the EMG signal. The EMG signal is described as a sum of several action potentials (wavelet) trains, embedded in noise. For each train, the time interval between wavelets is modeled by a process that parameters are linked to the muscular activity. The parameters of this process are estimated sequentially by a Bayes filter, along with the firing instants. The method was tested on some simulated signals and an experimental one, from which the rates of detection and classification of action potentials were above 95% with respect to the reference decomposition. The method works sequentially in time, and is the first to address the problem of intramuscular EMG decomposition online. It has potential applications for man-machine interfacing based on motor neuron activities.**

***Index Terms*—Bayes methods, Biomedical signal processing, Electromyography, Hidden Markov model, Recursive estimation, Weibull distribution.**

I. Introduction

ELECTROMYOGRAPHIC (EMG) signals represent the activity of muscle fibers, as driven by the population of spinal motor neurons (MN) innervating the muscle. Thus, despite being measured peripherally, EMG signals reflect neural activity since they contain information on the activation drive sent from the spinal cord to the muscles (neural drive to the muscles).

The identification of individual MN spikes from the EMG signal is termed EMG decomposition [1]. Intramuscular EMG (iEMG) decomposition methods have been developed since decades [2] [3], with highly accurate results [4]. The main aim of these methods is to allow physiological investigations of motor unit (MU) behavior during muscle contractions, so that the neural strategies for movement control can be decoded. Being mainly a research tool, EMG decomposition has been traditionally applied off-line, often including an interaction with an expert operator for maximizing the accuracy of the result. Despite the common offline use of EMG decomposition, some available methods could be potentially implemented online. Systems that may be implemented online need to work sequentially on the data samples or on data intervals. In this paper, we propose a new sequential algorithm for EMG decomposition that allows full decomposition in a sequential way. The foreseen future application is in man-machine interaction, where sequential decoding is needed.

In [5], a statistical model of intramuscular signals has been proposed, that parameters (motor unit action potential shapes (MUAP shapes) and spike train discharge rates) are estimated off-line by means of a Monte Carlo Markov Chain technique. In [6], similar techniques are used, but the sparsity and the regularity of the input spike trains are taken into account with a minimum time interval constraint between spikes. The interspike interval is modeled by means of a log-normal probability law in [7]. These methods are very efficient even in the blind case (*i.e.* unknown waveform shapes and spike trains) compared with the performance of an expert, but are not designed to work online. Conversely, in this paper we present a method to process single-channel iEMG signals sequentially.

The main difficulty comes from the mixture of several wavelet trains in only one recording, leading to possible superimposition of several action potential shapes. This study aims for the online estimation of the discharge rate of each train, despite these interferences and despite unknown action potential shapes (although a rough initial shape is necessary). It uses some of the concepts proposed in [8] and [9], where the information carried by spike trains is encoded by action potentials waveforms and decoded offline using a Viterbi algorithm. Like [10] and [11], it tackles the online problem, but it also uses tools from reliability theory to handle the regularity of the trains.

This contribution is a first step toward the decoding of MN activities in vivo, in humans, during natural movements, and in an online fashion for the future objective of man-machine interfacing. EMG signals are indeed extensively used for man-machine interfacing, for example for the control of prosthetic devices [12] or of exoskeletons [13], by extracting global characteristics of the signal for control. For example, the EMG amplitude can be used for direct proportional control of force or speed [14]. However, as alternative to global EMG processing, the activity of populations of MNs could in principle be decoded from the EMG signal and used for controlling external devices [15]. Indeed, the natural force generation process is based on the control of the number and discharge rates of the active MNs, so that the information on MN recruitment and modulation of discharge rates would provide an ideal prediction of force [16]. Although this concept

Manuscript created March 01, 2013; revised February 24, 2014

This work was partly sponsored by the European Research Council Advanced Grant DEMOVE (contract #267888) and the European Commission via grant #280778 (MERIDIAN) (DF).

*†LUNAM Université, *Université de Nantes, †École Centrale de Nantes, *†IRCCyN (UMR CNRS 6597), 1, rue de la Noë, 44321 Nantes, France, {first-name.last-name@irccyn.ec-nantes.fr}

‡Department of Neurorehabilitation Engineering, Bernstein Center for Computational Neuroscience, University Medical Center Göttingen, Georg-August University, Von-Siebold-Str. 4, 37075 Göttingen, Germany, {dario.farina@bccn.uni-goettingen.de}

is theoretically possible, a number of challenges should be faced to make it practically feasible.

The paper is organized as follows. In Section II, we describe a framework where the signal is modeled by a hidden Markov model (HMM) and the online estimation of the state of the Markov model by means of Bayes filtering is presented. Section III is dedicated to simulation and experimental tests. Conclusion and perspectives are given in section IV.

II. Methods

A. Modeling

1) Basic hypotheses: An iEMG signal is an electrical signature of muscle activity, *i.e.* activity of muscle fibers. As remarked in [17], [18], the property of the linearity of the sum of electrical signals justifies modeling the observed signal Y as a sum of n_{M} filtered spike trains embedded in a noise W; for all discrete time indexes n:

$$Y[n] = \sum_{i=1}^{n_{\text{M}}} (h_i * U_i)[n] + W[n] \tag{1}$$

For each source, a time-invariant linear filtering effect is considered. It appears as a convolution between the input spike train U_i and an impulse response h_i:

- the spike trains U_i are sparse 0-1 processes which firing rates are linked to the muscle activity;
- the shapes of the impulse responses h_i represent the MUAP and are assumed time-invariant.

This model is a reasonable simplification of the actual iEMG signal in which wave shapes and amplitudes are in general non stationary, due to needle movement, variable neuromuscular junction transmission, and variable conduction velocities.

In a probabilistic framework, the spike train sequences and the noise sequence are stochastic processes (and denoted using upper case letters, as it is a commonly used convention for random quantities, while lower case letters are used to denote deterministic quantities; additionally, bold letters will be used to denote vectors and matrices). They are supposed mutually independent. The noise is assumed independent along time, zero-mean, gaussian, with constant variance v. The spike trains U_i are not available. The impulse responses h_i are assumed of finite length with known maximum length ℓ_{IR}. The noise variance v is unknown. The number n_{M} of firing MNs is assumed to be known and fixed, and a rough initial estimation of the wavelets h_i is necessary; thus the method proposed in the following requires an initial estimation of the number of sources and a rough estimation of the wavelets; this preliminary estimation is not discussed in this study.

We propose below a Markov representation of the system, that state will be estimated online by a Bayes filter. A side effect of the method is the reconstruction of the spike trains (a.k.a. deconvolution).

2) Markov chain for input spike trains: each input sequence $(U_i[n])_{n\in\mathbb{Z}}$ is a binary discrete process, independent of other sequences $(U_j[n])_{n\in\mathbb{Z}}$ for all $j \neq i$ (although this assumption is questionable, especially when high contraction levels imply MU synchronization). In previous articles [19] [20], these sequences were assumed to be time-independent but physiological constraints make this assumption rather unrealistic. We propose to represent these discrete-time and discrete-valued sequences by a model based on renewal theory [21].

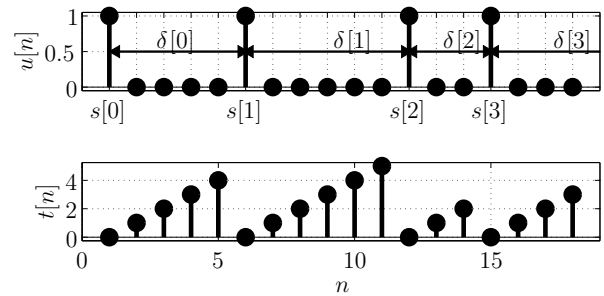


Figure 1. Spike train and corresponding sawtooth sequence

For each input source i, interspike intervals $\Delta_i[k]$ - discrete time length between spikes, $k \in \mathbb{Z}$ being the spike index - are supposed to be independent and identically distributed (i.i.d.) random variables, with a parameterized probability mass function (PMF) defined by $\mathsf{P}(\Delta_i[k] = t|\mathbf{\Theta}_i)$, for all positive natural numbers $t \geq 1$, where $\mathbf{\Theta}_i$ is an unknown parameter vector (which will be considered as random in the Bayesian framework below).

In reliability theory [22], it is usual to represent a discrete random variable by its survival function s or its hazard rate (or failure rate) r; for all positive natural numbers $t \geq 1$:

$$\begin{aligned} s(t, \mathbf{\Theta}_i) &= \mathsf{P}(\Delta_i[k] \geq t|\mathbf{\Theta}_i) \\ r(t, \mathbf{\Theta}_i) &= \frac{\mathsf{P}(\Delta_i[k] = t|\mathbf{\Theta}_i)}{s(t, \mathbf{\Theta}_i)} \end{aligned} \tag{2}$$

For each source, we introduce the sawtooth sequence $T_i[n]$, which corresponds to the time interval since the last spike. The value $T_i[n]$ is incremented at each time index, unless the MN fires; it is then set to zero, as seen in Fig.1.

$$T_i[n+1] = \begin{cases} 0, & \text{if the MN fires at time } n+1 \\ T_i[n]+1, & \text{otherwise} \end{cases}$$

By means of the Kronecker delta, the spike trains write $U_i[n] = \delta\left(T_i[n]\right)$. Let us use the exponent ^{n} to denote a time span till time n (e.g. $T_i^n = (T_i[j])_{1\leq j\leq n}$). Then, under the assumption that the sequence of interspike intervals is i.i.d., the following properties hold (see proof in appendix A).

• The sequence $T_i[n]$ is Markovian (for all discrete times n, given the present data $T[n]$, future data $(T_i[m])_{m>n}$ and past data $(T_i[m])_{m<n}$ are independent); for all possible time spans t (positive integer valued) since the last spike:

$$\mathsf{P}(T_i[n+1] = t|T_i^n, \mathbf{\Theta}_i) = \mathsf{P}(T_i[n+1] = t|T_i[n], \mathbf{\Theta}_i) \tag{3}$$

• The transition probability writes, by means of the hazard rate in (2):

$$\mathsf{P}(T_i[n+1] = t|T_i[n], \mathbf{\Theta}_i) = \begin{cases} r(T_i[n]+1, \mathbf{\Theta}_i), & \text{if } t = 0 \\ 1 - r(T_i[n]+1, \mathbf{\Theta}_i), & \text{if } t = T_i[n]+1 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

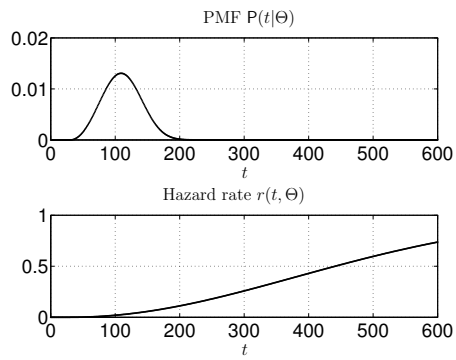


Figure 2. Weibull PMF and hazard rate ($t_0 = 120$, $\beta = 3$, $t_{\mathrm{R}} = 30$)

• The invariant distribution writes, by means of the survival function, for all $t \geq 0$:

$$\mathsf{P}(T_i[n] = t|\boldsymbol{\Theta}_i) = \frac{s(t+1, \boldsymbol{\Theta}_i)}{\mathsf{E}\{\Delta_i[k]|\boldsymbol{\Theta}_i\}} \tag{5}$$

• The firing rate in spikes per sample, that is the mean value of the binary process $U_i[n]$, is the inverse of the expectation of the interspike process, that is $1/\mathsf{E}\{\Delta_i[k]|\boldsymbol{\Theta}_i\}$.

We need an explicit form of the PMF $\mathsf{P}(\Delta_i[k] = t|\boldsymbol{\Theta}_i)$ which leads to a simple expression of the hazard rate in (2), so that the transition probabilities in (4) are easily computed. Few PMFs may lead to an explicit hazard rate. The discrete Weibull distribution, initially proposed by [23], was extended to three parameters by [24] with $\boldsymbol{\Theta}_i = [t_0, \beta, t_{\mathrm{R}}]^\top$: t_0 is a location parameter, β is a concentration parameter, and t_{R} is a shifting parameter. Although the expected value $\mathsf{E}\{\Delta_i[k]|\boldsymbol{\Theta}_i\}$ has no explicit form. It is included in an interval of length 1 (corresponding to one sample), which the middle is $(t_0 - t_{\mathrm{R}})\,\Gamma(1+1/\beta) + t_{\mathrm{R}} + 1/2$, where Γ is the Euler Gamma function [25]. Thus, it can be approximated by the middle of this interval:

$$\mathsf{E}\{\Delta_i[k]|\boldsymbol{\Theta}_i\} \approx (t_0 - t_{\mathrm{R}})\,\Gamma(1 + 1/\beta) + t_{\mathrm{R}} + 1/2 \tag{6}$$

Given $\boldsymbol{\Theta}_i = [t_0, \beta, t_{\mathrm{R}}]^\top$, the discrete Weibull PMF, the survival function s and the hazard rate r writes, for all natural numbers $t \geq t_{\mathrm{R}}$:

$$\mathsf{P}(\Delta_i[k] = t|\boldsymbol{\Theta}_i) = \mathrm{e}^{-\left(\frac{t-t_{\mathrm{R}}-1}{t_0-t_{\mathrm{R}}}\right)^\beta} - \mathrm{e}^{-\left(\frac{t-t_{\mathrm{R}}}{t_0-t_{\mathrm{R}}}\right)^\beta} \tag{7}$$

$$s(t, \boldsymbol{\Theta}_i) = \mathrm{e}^{-\left(\frac{t-t_{\mathrm{R}}-1}{t_0-t_{\mathrm{R}}}\right)^\beta} \tag{8}$$

$$r(t, \boldsymbol{\Theta}_i) = 1 - \mathrm{e}^{\left(\frac{t-1-t_{\mathrm{R}}}{t_0-t_{\mathrm{R}}}\right)^\beta - \left(\frac{t-t_{\mathrm{R}}}{t_0-t_{\mathrm{R}}}\right)^\beta} \tag{9}$$

Note that for $t < t_{\mathrm{R}}$, the PMF and the hazard rate are 0, the survival function is 1.

An example of a Weibull PMF and hazard rate is shown in Fig.2, and an associated realization of a spike train according to this PMF is shown in Fig.3. If $t_{\mathrm{R}} > 0$ or $\beta > 1$, the produced spike train is more regular than in the case of a geometric distribution, which corresponds to $t_{\mathrm{R}} = 0$ and $\beta = 1$.

In our study, t_0 and β have to be estimated, but t_{R} is assumed to be known due to the physiological constraint (a wavelet occurrence is followed by a refractory period during which no new wavelet can appear). Therefore, $\boldsymbol{\Theta}_i$ will be reduced to two components below.

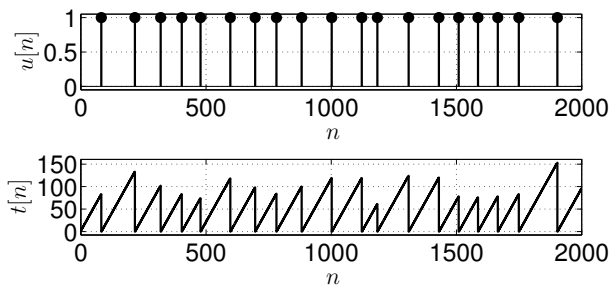


Figure 3. Spike train drawn from a Weibull PMF ($t_0 = 120$, $\beta = 3$ and $t_{\mathrm{R}} = 30$) and corresponding sawtooth sequence.

3) Hidden Markov model: Assuming that the length ℓ_{IR}, in number of samples, of the impulse responses, is lower than the refractory period t_{R}, a simple state-space representation is obtained. The random state vector is composed of:

- $\mathbf{T}[n]$, the n_{M} sojourn times $\mathbf{T}[n] = [T_i[n]]_{i \in \{1 \ldots n_{\mathrm{M}}\}}$,
- $\mathbf{H}$, the column vector made from the concatenation of the n_{M} impulse responses (each one with ℓ_{IR} coefficients),
- $\boldsymbol{\Theta}$, the column vector made from the concatenation of the n_{M} parameter vectors related to the dynamics of the sources (each one with two coefficients).

The HMM corresponds to:

- a transition law, in which the impulse responses and the interspike laws parameters are constant through time:

$$\begin{cases} \mathbf{H},\, \boldsymbol{\Theta} \;\; \text{constant along time} \\ \mathsf{P}(\mathbf{T}[n+1]|\mathbf{T}[n], \boldsymbol{\Theta}) \;\; \text{thanks to (4)} \end{cases} \tag{10}$$

- an observation equation (where $\boldsymbol{\varphi}_t$ is a row vector of size ℓ_{IR} such that all components are 0, except, if $t < \ell_{\mathrm{IR}}$, the component in position $t+1$ which is 1), which is equivalent to the representation (1).

$$Y[n] = \underbrace{\left[\boldsymbol{\varphi}_{T_1[n]} \cdots \boldsymbol{\varphi}_{T_{n_{\mathrm{M}}}[n]}\right]}_{\boldsymbol{\psi}_{\mathbf{T}[n]}} \mathbf{H} + W[n] \tag{11}$$

The Bayes filter designed to estimate $\mathbf{H}$ and $\boldsymbol{\Theta}$, with known noise variance v, is primarily presented in next part. Then, it is adapted for tracking purpose in the more realistic case of time-varying $\mathbf{H}$ and $\boldsymbol{\Theta}$ and unknown time-varying v in subsection II-B5. The setting of the filter parameters is presented in subsection II-B6, while initializations are presented in subsection II-B7.

B. Bayes filter

1) Principle: The purpose of a Bayes filter is to propagate along time the posterior probability law of the state sequence of a HMM. Let Y^n be the measured data. Since the state vector is made of discrete variables (the sawtooth sequences) and continuous ones (the impulse responses and the interspike laws parameters), we shall compute recursively, for a growing time index n:

- the posterior probability density function (PDF) of the interspike law parameters given the impulse responses

and the sawtooth sequences: $\mathsf{p}_{\boldsymbol{\Theta}|\mathbf{T}^n,Y^n,\mathbf{H}}$, which obviously simplifies to $\mathsf{p}_{\boldsymbol{\Theta}|\mathbf{T}^n}$ (since the knowledge of the data Y^n and of the filtering process $\mathbf{H}$ does not bring any information about $\boldsymbol{\Theta}$ when $\mathbf{T}^n$ is known), and finally reduces to $\prod_{i=1}^{n_{\mathrm{M}}} \mathsf{p}_{\boldsymbol{\Theta}_i|T_i^n}$ (since the n_{M} binary sequences are assumed to be independent),
- the posterior PDF of the set of the impulse responses given the sawtooth sequences: $\mathsf{p}_{\mathbf{H}|\mathbf{T}^n,Y^n}$,
- the posterior PMF of the set of random sawtooth sequences $\mathbf{T}^n$, that is, for all possible values $\mathbf{t}^n$, $\mathsf{P}(\mathbf{T}^n = \mathbf{t}^n|Y^n)$.

To simplify notation, we will use the exponent $^{|n}$ to mean "knowing data Y^n" (for example, $\mathsf{P}(\mathbf{T}^n = \mathbf{t}^n|Y^n)$ becomes $\mathsf{P}^{|n}(\mathbf{T}^n = \mathbf{t}^n)$). Derivation of the Bayes estimators follows, such as:

- the marginal maximum a posteriori (MAP) estimator for the set of sawtooth sequences (that is to say the most probable one):

$$\widehat{\mathbf{T}}^{n|n} = \arg\max_{\mathbf{t}^n} \mathsf{P}^{|n}(\mathbf{T}^n = \mathbf{t}^n) \tag{12}$$

- the minimum mean square error estimator for the continuous valued state, that is, by means of the total expectation formula, for the set of impulse responses:

$$\begin{aligned}\widehat{\mathbf{H}}^{|n} &= \mathsf{E}^{|n}\{\mathbf{H}\} \\ &= \mathsf{E}^{|n}\{\mathsf{E}^{|n}\{\mathbf{H}|\mathbf{T}^n\}\} \\ &= \sum_{\mathbf{t}^n} \underbrace{\mathsf{E}^{|n}\{\mathbf{H}|\mathbf{T}^n = \mathbf{t}^n\}}_{\widehat{\mathbf{H}}^{|n}_{\mathbf{t}^n}} \mathsf{P}^{|n}(\mathbf{T}^n = \mathbf{t}^n)\end{aligned} \tag{13}$$

and for the interspike law parameters, for all MU indices i:

$$\widehat{\boldsymbol{\Theta}}_i^{|n} = \sum_{t_i^n} \underbrace{\mathsf{E}\{\boldsymbol{\Theta}_i|T_i^n = t_i^n\}}_{\widehat{\boldsymbol{\theta}}_{t_i^n}} \mathsf{P}^{|n}(T_i^n = t_i^n) \tag{14}$$

To derive the recursion on the posterior probability laws, we will use the marginalization principle (which is well known in particle filter implementations [26]). For a fixed set of sawtooth sequences $\mathbf{T}^n$:

- unfortunately, the conditional PDF $\mathsf{p}_{\boldsymbol{\Theta}_i|T_i^n}$ of the interspike laws parameters is not simple and we will use approximations to propose an alternative estimator $\widehat{\boldsymbol{\theta}}_{t_i^n}$ (see II-B2),
- the posterior PDF $\mathsf{p}_{\mathbf{H}|\mathbf{T}^n,Y^n}$ of the set of impulse responses is gaussian, its mean $\widehat{\mathbf{H}}^{|n}_{\mathbf{T}^n}$ and variance are obtained with the standard Kalman filter [27] (see II-B3).

Finally, the probability of the sawtooth sequences is computed according to previous estimates (see II-B4). Furthermore, it is unthinkable to process all possible sawtooth sequences, since their number increases exponentially as time index n grows. At each step n, only the n_{path} most probable sequences will be kept, the number n_{path} being a parameter of the method similar to the number of particles in a particle filter [28].

2) Estimation of interspike law parameters: Up to our knowledge, it is impossible to determine an easy way to implement online minimum mean square error estimation $\mathsf{E}\{\boldsymbol{\Theta}_i|T_i^n\}$ of the interspike laws parameters $\boldsymbol{\Theta}_i$. In [29], an iterative procedure is presented to obtain a maximum likelihood (ML) estimator for a discrete Weibull distribution. We propose here a ML-based estimator, which leads to a practical online implementation. The offline ML estimation is given by:

$$\widehat{\boldsymbol{\theta}}_{t_i^n} = \arg\min_{\boldsymbol{\theta}} \underbrace{-\frac{1}{n}\ln \mathsf{P}(T_i^n = t_i^n|\boldsymbol{\Theta}_i = \boldsymbol{\theta})}_{J_{t_i^n}(\boldsymbol{\theta})}$$

Using the Markov property of the sawtooth sequences and with the initialization $J_{t_i^1}(\boldsymbol{\theta}) = -\ln \mathsf{P}(T_i[1] = t_i[1]|\boldsymbol{\Theta}_i = \boldsymbol{\theta})$, the objective function $J_{t_i^n}$ recursively writes for all $n \geq 2$:

$$J_{t_i^n}(\boldsymbol{\theta}) = \frac{1}{n} Q_{t_i^n}(\boldsymbol{\theta}) + (1 - \frac{1}{n})\, J_{t_i^{n-1}}(\boldsymbol{\theta})$$

with $Q_{t_i^n}(\boldsymbol{\theta}) = -\ln \mathsf{P}(T_i[n] = t_i[n]|\boldsymbol{\Theta}_i = \boldsymbol{\theta}, T_i[n-1] = t_i[n-1])$. $J_{t_i^1}$ and $Q_{t_i^n}$ are computed using (4) and (5); note that with a discrete Weibull law, in the special case where $\beta = 1$ is fixed, the estimator of t_0 is closed-form (see appendix B). In the general case, we will use the heuristic online implementation used in [30]; if $\widehat{\boldsymbol{\theta}}_{t_i^{n-1}}$ attains the minimum of $J_{t_i^{n-1}}(\boldsymbol{\theta})$, the gradient of the objective function gives :

$$\frac{\partial J_{t_i^n}(\widehat{\boldsymbol{\theta}}_{t_i^{n-1}})}{\partial \boldsymbol{\theta}} = \frac{1}{n} \frac{\partial Q_{t_i^n}(\widehat{\boldsymbol{\theta}}_{t_i^{n-1}})}{\partial \boldsymbol{\theta}}$$

This gradient is straightforward to compute with the transition probability given by a discrete Weibull law. The objective function is recursively minimized by a stochastic gradient update:

$$\widehat{\boldsymbol{\theta}}_{t_i^n} = \widehat{\boldsymbol{\theta}}_{t_i^{n-1}} - \frac{1}{n}\, \mathbf{G}_{t_i^n}^{-1} \frac{\partial Q_{t_i^n}(\widehat{\boldsymbol{\theta}}_{t_i^{n-1}})}{\partial \boldsymbol{\theta}} \tag{15}$$

where the Riemannian metric tensor [31] is updated for all $n \geq 2$ using:

$$\mathbf{G}_{t_i^n} = \frac{1}{n} \frac{\partial Q_{t_i^n}(\widehat{\boldsymbol{\theta}}_{t_i^{n-1}})}{\partial \boldsymbol{\theta}} \frac{\partial Q_{t_i^n}(\widehat{\boldsymbol{\theta}}_{t_i^{n-1}})}{\partial \boldsymbol{\theta}}^{\top} + (1 - \frac{1}{n})\, \mathbf{G}_{t_i^{n-1}} \tag{16}$$

Note that this can also be interpreted as a quasi-Newton online optimization where Hessian matrix approximation is obtained through the Fisher information matrix of the transition probability law [32].

3) Posterior PDF and estimation of the impulse responses: For a given realization $\mathbf{t}^n$ of the sawtooth sequences $\mathbf{T}^n$, the Markov model for the impulse responses $\mathbf{H}$ reduces, for all $n \geq 1$, to:

$$\begin{cases} \mathbf{H} \text{ constant} \\ Y[n] = \boldsymbol{\psi}_{\mathbf{t}[n]}\, \mathbf{H} + W[n] \end{cases} \tag{17}$$

with $(W[n])_{n\geq 1}$ being an i.i.d. Gaussian sequence with variance v. With a Gaussian prior to represent the confidence in the initial prediction of $\mathbf{H}$, this is a standard linear gaussian model, for which the posterior law of $\mathbf{H}|\mathbf{T}^n = \mathbf{t}^n, Y^n$ is Gaussian with mean $\widehat{\mathbf{H}}^{|n}_{\mathbf{t}^n}$ and covariance matrix $\mathbf{p}_{\mathbf{t}^n}$ provided

by the standard Kalman filter. With initial priors $\widehat{\mathbf{H}}_{\mathbf{t}^0}^{|0}$ and $\mathbf{p}_{\mathbf{t}^0}$, the recursion writes for all $n \geq 1$:

$$\begin{aligned} \mathbf{k}_{\mathbf{t}^n} &= \mathbf{p}_{\mathbf{t}^{n-1}}\, \boldsymbol{\psi}_{\mathbf{t}[n]}^{\top}\, \nu_{\mathbf{t}^n}^{-1} \\ \widehat{\mathbf{H}}_{\mathbf{t}^n}^{|n} &= \widehat{\mathbf{H}}_{\mathbf{t}^{n-1}}^{|n-1} + \mathbf{k}_{\mathbf{t}^n}\, (Y[n] - \widehat{Y}_{\mathbf{t}^n}^{|n-1}) \\ \mathbf{p}_{\mathbf{t}^n} &= \mathbf{p}_{\mathbf{t}^{n-1}} - \mathbf{k}_{\mathbf{t}^n}\, \nu_{\mathbf{t}^n}\, \mathbf{k}_{\mathbf{t}^n}^{\top} \end{aligned} \tag{18}$$

In this equation, appears the prior law of the observation $Y[n]|\mathbf{T}^n = \mathbf{t}^n, Y^{n-1}$ which is itself Gaussian with mean $\widehat{Y}_{\mathbf{t}^n}^{|n-1}$ and variance $\nu_{\mathbf{t}^n}$, and is updated, for all $n \geq 0$, through:

$$\begin{aligned} \widehat{Y}_{\mathbf{t}^{n+1}}^{|n} &= \boldsymbol{\psi}_{\mathbf{t}[n+1]}\, \widehat{\mathbf{H}}_{\mathbf{t}^n}^{|n} \\ \nu_{\mathbf{t}^{n+1}} &= \boldsymbol{\psi}_{\mathbf{t}[n+1]}\, \mathbf{p}_{\mathbf{t}^n}\, \boldsymbol{\psi}_{\mathbf{t}[n+1]}^{\top} + v \end{aligned} \tag{19}$$

$\widehat{\mathbf{H}}_{\mathbf{t}^n}^{|n}$ is the conditional mean $\mathsf{E}^{|n}\{\mathbf{H}|\mathbf{T}^n = \mathbf{t}^n\}$ involved in (13), and $\mathbf{k}_{\mathbf{t}^n}$ is the well-known Kalman gain. If the additive noise is not Gaussian, the Kalman filter becomes a recursive implementation of a linear minimum mean square error estimator (LMMSE) which remains unbiased (with minimum variance among all linear estimators). An autoregressive model of the noise could be included to take into account dependency along time. However, a trade-off between model complexity and the feasibility of the Bayes filter leads us to the common assumption of white Gaussian noise.

4) Posterior probability of sawtooth sequences: In an update/prediction scheme, with an initial prior $\mathsf{P}(\mathbf{T}^1 = \mathbf{t}^1|Y^0)$, the recursion on this posterior probability writes, for all possible forks $\mathbf{t}^{n+1}$ coming from $\mathbf{t}^n$ (that are sequences in which scalar sub-sequences t_i^{n+1} are obtained from t_i^n either by $t_i^{n+1} = \{t_i^n, t_i[n]+1\}$, or by $t_i^{n+1} = \{t_i^n, 0\}$ if $t_i[n] \geq t_{\mathrm{R}}$; there can be between 1 and $2^{n_{\mathrm{M}}}$ forks):

$$\begin{aligned} &\mathsf{P}^{|n}(\mathbf{T}^n = \mathbf{t}^n) \propto \\ &\qquad g(Y[n] - \widehat{Y}_{\mathbf{t}^n}^{|n-1}, \nu_{\mathbf{t}^n})\, \mathsf{P}^{|n-1}(\mathbf{T}^n = \mathbf{t}^n) \end{aligned} \tag{20}$$

$$\begin{aligned} &\mathsf{P}^{|n}(\mathbf{T}^{n+1} = \mathbf{t}^{n+1}) = \mathsf{P}^{|n}(\mathbf{T}^n = \mathbf{t}^n) \times \\ &\prod_{i=1}^{n_{\mathrm{M}}} \begin{cases} \mathsf{E}\{r(t_i[n]+1, \boldsymbol{\Theta}_i)|T_i^n = t_i^n\}, \\ \qquad\qquad \text{if } t_i[n+1] = 0 \\ 1 - \mathsf{E}\{r(t_i[n]+1, \boldsymbol{\Theta}_i)|T_i^n = t_i^n\}, \\ \qquad\qquad \text{if } t_i[n+1] = t_i[n]+1 \\ 0, \qquad\qquad \text{otherwise} \end{cases} \end{aligned} \tag{21}$$

where $g(., \nu)$ is the PDF of a zero-mean Gaussian variable with variance ν, and $\widehat{Y}_{\mathbf{t}^n}^{|n-1}$ and $\nu_{\mathbf{t}^n}$ are provided by the Kalman filter of the previous subsection (see proof in appendix C).

The expectation $\mathsf{E}\{r(t, \boldsymbol{\Theta}_i)|T_i^n = t_i^n\}$ has no explicit expression. We use an approximation similar to the one involved in the extended Kalman filter, which consists in a linearization of $\boldsymbol{\theta} \mapsto r(t, \boldsymbol{\theta})$ around $\mathsf{E}\{\boldsymbol{\Theta}_i|T_i^n = t_i^n\}$, leading to:

$$\mathsf{E}\{r(t, \boldsymbol{\Theta}_i)|T_i^n = t_i^n\} \approx r(t, \mathsf{E}\{\boldsymbol{\Theta}_i|T_i^n = t_i^n\}) \tag{22}$$

There is no easy solution to recursively compute the expectation $\mathsf{E}\{\boldsymbol{\Theta}_i|T_i^n = t_i^n\}$. We replace it by the Recursive Maximum Likelihood estimation $\widehat{\boldsymbol{\theta}}_{t_i^n}$ obtained in II-B2. Then, the prediction step in recursion (21) becomes:

$$\begin{aligned} &\mathsf{P}^{|n}(\mathbf{T}^{n+1} = \mathbf{t}^{n+1}) \approx \mathsf{P}^{|n}(\mathbf{T}^n = \mathbf{t}^n) \times \\ &\prod_{i=1}^{n_{\mathrm{M}}} \begin{cases} r(t_i[n]+1, \widehat{\boldsymbol{\theta}}_{t_i^n}), & \text{if } t_i[n+1] = 0 \\ 1 - r(t_i[n]+1, \widehat{\boldsymbol{\theta}}_{t_i^n}), & \text{if } t_i[n+1] = t_i[n]+1 \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{23}$$

With the hazard rate of a discrete Weibull probability law, this recursion is particularly easy to implement.

5) Tracking: In algorithm 1, we report a summary of the whole process. The noise variance v is primarily estimated on a time interval of the data where only noise is present, for instance a period where the human subject does not activate any muscle. Furthermore, as the parameters of the interspike laws are known to be time varying, a slight change is introduced in (15) and (16): the factor $1/n$ is replaced by $1/\ell[n]$ where $\ell[n]$ is a window length, which grows from $\ell[1] = 1$ to ℓ_∞, according to the following recursion [30]:

$$\ell[n+1] = (1 - \frac{1}{\ell_\infty})\, \ell[n] + 1$$

ℓ_∞ is an equivalent window length which has to be set according to the desired adaptivity. This corresponds to an exponential forgetting factor $1 - \frac{1}{\ell_\infty}$, lower than 1 but next to 1. There is no tracking when $\ell_\infty = +\infty$.

Moreover, on actual iEMG signals, the impulse responses are also known to be time-varying. A usual solution is to introduce a random walk in the model, leading to a strict implementation of the Kalman filter, which was used in [33]. Nevertheless, to avoid new parameters to tune this random walk, we use the trick proposed in [34] which uses a forgetting factor to artificially increase the covariance matrix in the Kalman filter. The covariance matrix update in (18) becomes:

$$\mathbf{p}_{\mathbf{t}^n} = \frac{1}{1 - \frac{1}{\ell_\infty}}\, \left(\mathbf{p}_{\mathbf{t}^{n-1}} - \mathbf{k}_{\mathbf{t}^n}\, \nu_{\mathbf{t}^n}\, \mathbf{k}_{\mathbf{t}^n}^{\top}\right) \tag{24}$$

The noise variance v is itself unknown. We propose here an heuristic approach to estimate it by filtering the square of the estimation error $Y[n] - \boldsymbol{\psi}_{\mathbf{t}[n]}\, \widehat{\mathbf{H}}_{\mathbf{t}^n}^{|n}$ using the same adaptivity coefficients; for each possible sawtooth sequence $\mathbf{t}^n$:

$$\widehat{V}_{\mathbf{t}^n}^{|n} = \frac{1}{\ell[n]}\, \left(Y[n] - \boldsymbol{\psi}_{\mathbf{t}[n]}\, \widehat{\mathbf{H}}_{\mathbf{t}^n}^{|n}\right)^2 + (1 - \frac{1}{\ell[n]})\, \widehat{V}_{\mathbf{t}^{n-1}}^{|n-1} \tag{25}$$

The global estimation is then:

$$\widehat{V}^{|n} = \sum_{\mathbf{t}^n} \widehat{V}_{\mathbf{t}^n}^{|n}\, \mathsf{P}^{|n}(\mathbf{T}^n = \mathbf{t}^n) \tag{26}$$

The estimation $\widehat{V}^{|n}$ replaces v in (19):

$$\nu_{\mathbf{t}^{n+1}} = \boldsymbol{\psi}_{\mathbf{t}[n+1]}\, \mathbf{p}_{\mathbf{t}^n}\, \boldsymbol{\psi}_{\mathbf{t}[n+1]}^{\top} + \widehat{V}^{|n} \tag{27}$$

6) Tuning: The algorithm works sample by sample sequentially and thus can provide a real time result if the computational time is sufficiently fast to be completed within one sample period. The scope of this work is to present a sequential algorithm while the numerical implementation and computational cost analysis are out of the scope of the current study. Nevertheless, it is obvious that the number of kept

Algorithm 1 Full estimation process

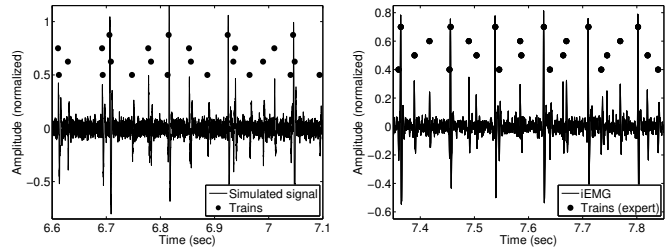
Initialize $\widehat{\mathbf{H}}^{|0}_{\mathbf{t}^0}$, $\mathbf{p}_{\mathbf{t}^0}$, $\widehat{V}^{|0}_{\mathbf{t}^0}$ ($\mathbf{t}^0$ is empty matrix)
for all initial $\mathbf{t}[1]$ **do**
  Initialize $\mathsf{P}^{|0}(\mathbf{T}[1] = \mathbf{t}[1])$, $\widehat{\boldsymbol{\theta}}_{t_i[1]}$, $\mathbf{G}_{t_i[1]}$
  Predict $\widehat{Y}^{|0}_{\mathbf{t}^1}$ and $\nu_{\mathbf{t}^1}$ with (19)
end for
for all $n \geq 1$ **do**
  new data $Y[n]$
  for all $\mathbf{t}^n$ **do**
    Compute the posterior $\mathsf{P}^{|n}(\mathbf{T}^n = \mathbf{t}^n)$ with (20)
  end for
  Select and keep the n_{path} most probable paths
  for all kept $\mathbf{t}^n$ **do**
    Update $\widehat{\mathbf{H}}^{|n}_{\mathbf{t}^n}$, $\mathbf{p}_{\mathbf{t}^n}$ with (18) (corrected by (24)), $\widehat{V}^{|n}_{\mathbf{t}^n}$ with (25)
  end for
  Compute the estimates $\widehat{\mathbf{H}}^{|n}$, $\widehat{\boldsymbol{\Theta}}^{|n}_i$, $\widehat{V}^{|n}$ with (13), (14) and (26)
  for all kept $\mathbf{t}^n$ **do**
    for all possible forks $\mathbf{t}^{n+1}$ coming from $\mathbf{t}^n$ **do**
      Compute the prior $\mathsf{P}^{|n}(\mathbf{T}^{n+1} = \mathbf{t}^{n+1})$ with (23)
      Compute $\widehat{\boldsymbol{\theta}}_{t_i^{n+1}}$ and $\mathbf{G}_{t_i^{n+1}}$ with (15) and (16)
      Predict $\widehat{Y}^{|n}_{\mathbf{t}^{n+1}}$ and $\nu_{\mathbf{t}^{n+1}}$ with (19) (corrected by (27))
    end for
  end for
end for

paths n_{path} is a crucial parameter of the method, which has to be chosen as big as possible according to the available computational power. Typically, it should be closely related to the number of parallel units in a graphical processing unit (GPU) used for general purpose processing (GPGPU). It must be underlined that there is a tradeoff between this parameter and the refractory period t_{R} (although this setting is motivated by physiological constraints): the higher the t_{R}, the lower the number of forks to investigate.

The equivalent window length ℓ_∞, used to tune the filter adaptivity, should be set in function of the human subject task type. One can imagine automatic recognition of task type ("slow and meticulous", "fast"...). Detection of abrupt changes of parameters can also be implemented [35].

The main limitation of this paper is the assumption that the number n_{M} of MUs is known, although this typically changes when force is time varying. In the next step of this research work, beyond the scope of this paper, we plan to implement a birth-and-death technique (see offline case in [36], or online case in [37]).

7) Initializations: A pre-processing was made manually by an expert to extract the number of sources, a rough initialization of MUAP shapes $\widehat{\mathbf{H}}^{|0}_{\mathbf{t}^0}$ (with a diagonal covariance matrix $\mathbf{p}_{\mathbf{t}^0}$ corresponding to a standard deviation of 10% of the initial value on each component) and a rough initialization of the noise variance $\widehat{V}^{|0}_{\mathbf{t}^0}$ on a period with no activity. A fully automated procedure will require a startup stage such as a classification/estimation procedure [8] during a learning procedure performed by the human subject.



(a) Simulated signal (b) iEMG signal

Figure 4. (a) Simulmation: sum of four convolved spike trains with additive gaussian noise. (b) iEMG signal and expert decomposition.

n_{path} initial $\mathbf{t}[1]$ are drawn from a uniform distribution between $t_{\text{R}} + 1$ and $3t_{\text{R}}$, and weighted with the same initial probability.

Simulations have shown that the initial setup of the metric tensor $\mathbf{G}_{t_i[1]}$ is crucial. We propose to operate in two steps; for each source, till the second spike appears:

- set $\beta = 1$,
- estimate only t_0 by means of the closed form solution of Appendix B which does not require $\mathbf{G}_{t_i[1]}$,
- simultaneously estimate $\mathbf{G}_{t_i[n]}$ by means of (16), initialized with $\mathbf{G}_{t_i[1]} = \mathbf{0}_{2\text{x}2}$.

Then, the constraint $\beta = 1$ is relaxed.

8) Algorithm complexity: The computational complexity is mainly due to the Kalman filter (18). For each of the proposed n_{path} paths, the computation cost of the variance-covariance matrix P_{t^n} is in $\mathcal{O}((n_{\text{M}}\ell_{\text{IR}})^3)$ [38]. The overall complexity is in $\mathcal{O}(n_{\text{path}}(n_{\text{M}}\ell_{\text{IR}})^3)$. Note that the computational cost of (19) is low due to the sparsity of the $\boldsymbol{\psi}_{\mathbf{t}}$ vector.

C. Simulated signals and experimental signal

Simulated signals were generated with the Markov model, (10) and (11). A 10 kHz sampling frequency was assumed for the simulated signals to ease comparison with the experimental signal. The filters shapes were obtained from experimental iEMG signals to make the simulation more realistic. The signal to noise ratio (SNR) was set to 11 dB. The parameters $t_{0,i}$ ranged from 70 ms to 110 ms, β_i ranged from 3 to 7, and t_{R} was set at 30 ms. Fig.4 (a) shows the sum of four convolved spike trains, and the firing instants, localized by dots. The length of the simulated and experimental runs are of eight seconds each. A jump from 90 ms to 105 ms on t_0 was simulated on source #1, at time 4 s. Although this simulation does not reflect a precise physiological mechanism, it permits to test, over a single trial, the tracking capability of the algorithm over time-varying parameters, as well as its capability of locking the values of parameters constant over time.

The experimental iEMG signals were recorded from the extensor digitorum muscle of a healthy subject (age 21 years), with a pair of wire electrodes made of Teflon coated stainless steel (A-M Systems, Carlsborg, WA, USA; diameter 50μm) inserted into the muscle belly with a 25 G needle, with insertion at approximately 45 degrees with respect to the skin plane.

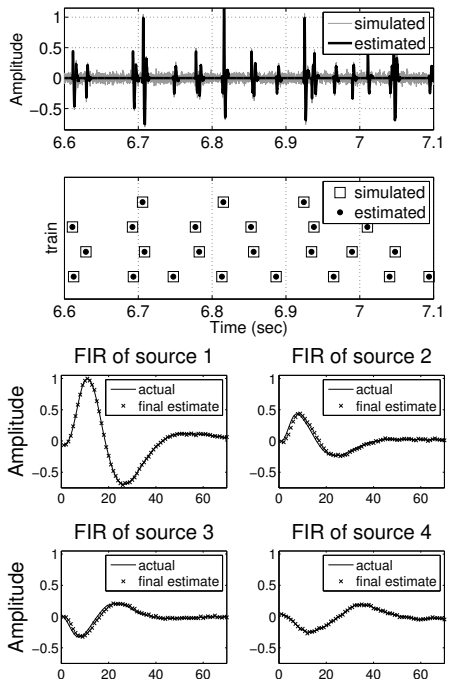


Figure 5. Reconstruction without noise of the signal for four spike trains: the reconstruction of the signal between 6.6 and 7.1 seconds with the true and reconstructed spike trains, and the estimated impulse responses (simulation)

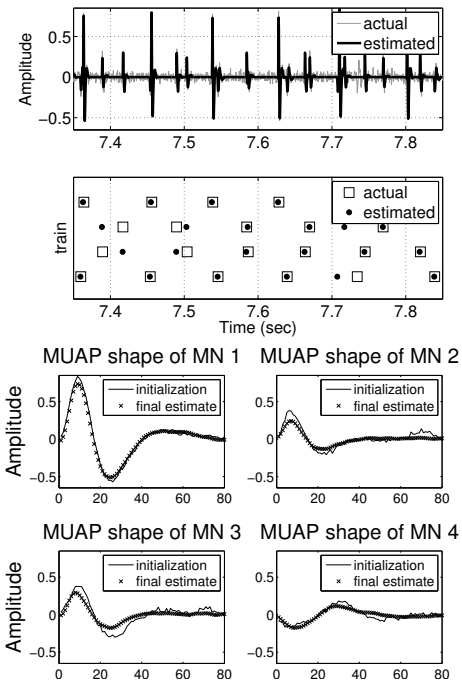


Figure 6. Reconstruction without noise of the signal for four identified spike trains of an iEMG signals: the reconstruction with the expert and estimated spike trains, and the estimated impulse responses (experimental signal)

The needle was then removed with the fine wires left into the muscle for the recording. The muscle selected corresponds to a source of EMG activity often used for the control of prostheses by transradial amputees. The iEMG signals were amplified bipolarly (Counterpoint EMG, DANTEC Medical Skovlunde, Denmark), band-pass filtered (500 Hz-5 kHz), and sampled at 10 kHz. The subject performed constant isometric contractions at 5% of the maximal voluntary contraction force (MVC) to gather iEMG signals over a period length of approximately one minute. During the experiment, the subject had access to a visual feedback of the exerted force. Fig.4 (b) presents an example of these recordings.

Algorithm 1 was applied to both the simulated signals and the experimental one, with adaptivity corresponding to a one-second equivalent window length. The algorithm realizes a sequential estimation of the dynamic parameters $\{t_0, \beta\}$ and the signature h of each spike trains. The number of selected paths n_{path} was set to 64. The number of MN for the experimental signal was estimated by an expert and assumed known (see II-B7), here 4 MNs.

The validation on experimental data was performed by comparing the results of the proposed method with those provided as reference results by manual offline decomposition of an expert operator using the EMGLAB tool [39]. The performance of the algorithm was evaluated with the indices proposed in [40], which characterize the detection and classification phase. For the proposed algorithm, both phases are realized together but the indexes can still be computed separately. The reference signals were the simulated trains in the simulation case, meaning that the firing instants are known exactly. In the experimental case, the firing instants are obtained from the expert's deconvoluted trains. Both patterns of firing instants are named *true spikes*. Thus, the spike trains proposed by Algorithm 1 and corresponding to true spikes are named *correct spikes*. Considering all MUAP trains as a unique process for the detection phase, we used as performance indexes, the sensitivity, which is the number of correct spikes relative to the number of detected spikes, and the positive predictivity, which is the number of correct spikes relative to the number of true spikes. A ± 2 samples tolerance on spike location was admitted. For the classification phase, we report the sensitivity, defined similarly to the detection phase but for each class (source), the specificity, which indicates the ability to identify MUAPs belonging to a different class, and the accuracy, indicating the quality of classification relative to the other MUAP trains.

Another (indirect) index of performance is the normalized root mean square error (NRMSE), computed as the root of the empirical mean of the square difference between the reconstructed signal $\widehat{y}$ and the actual signal realization y, normalized by the empirical standard deviation of the data.

$$\text{NRMSE} = \sqrt{\frac{\sum_{j=1}^{n} (\widehat{y}[j] - y[j])^2}{\sum_{j=1}^{n} y^2[j] - \frac{1}{n} \left(\sum y[j]\right)^2}}$$

III. Results

Fig.5 (simulated data) and Fig.6 (actual signal) present:

- processed data (simulated data or actual iEMG signal) and noiseless reconstruction,
- spike trains (actual or from expert) and estimated ones,
- actual impulse responses (for simulated data only) and estimated ones (initial and final values).

Table I reports the normalized noise standard deviation (simulation), the NRMSE for the expert's decomposition (actual signal), and the NMSRE for the online decomposition.

For the simulated data (Fig.5), the reconstruction of the spike trains did not miss any spikes over the duration of the signal. The NRMSE is very close to the actual normalized noise standard deviation meaning that the signal is almost decomposed perfectly.

For the experimental iEMG signal (Fig.6), the global reconstruction was close to the actual signal. However, several differences appear along the decomposition. Due to similarity of MUAP shapes from the second and third trains, the algorithm diverged from the expert, switching from one train to the other. Nevertheless, the NRMSE results are nearly identical, since the NRMSE is blind to errors between MU with similar shapes, and to train regularity. It has to be noted that the expert will mainly promote shape matching, whereas the algorithm will take into account the train regularity.

The performance indexes were all 100% for the simulated signals (except the detection sensitivity which was 99.9%). For the experimental signals, the indexes are calculated with respect to the expert decomposition, and are given in the table II. The algorithm detected well the MUAP shapes in quality and quantity. Double discharges may occur experimentally and cause a decrease in performance because this phenomenon is not taken into account in the model. The classification phase for the experimental signal shows that train 1 is perfectly classified, whereas the trains 2 and 3 were occasionally switched due to the similarity in the shapes of MUAPs in these two trains.

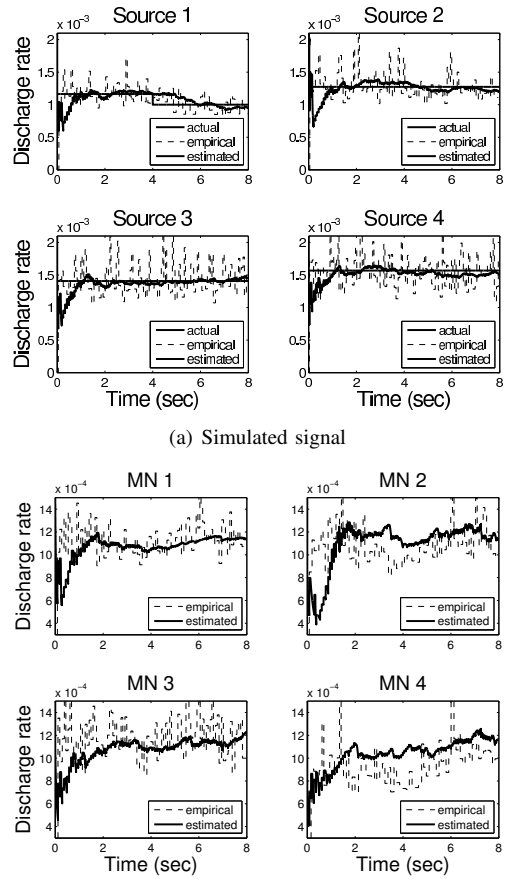
Fig.7 shows the estimated discharge rates of the sources, that is the number of spikes per unit time, computed from $\widehat{t}_0$ and $\widehat{\beta}$ by means of (6). They converge quickly to their true values, although there is variability around the true values. This variability is not due to errors in the decomposition of the spike trains by the proposed algorithm but to the need for several action potentials for an estimate of discharge rate with low variability. It has to be noted that the average discharge rate (or cumulative, which is equivalent to the average apart a multiplication factor) would be the control signal for man-machine interfaces (estimated neural drive to the muscle). On the simulated signal with the jump on discharge rate for the first source, the estimation catches up the final value after approximately 1.5 s, which is a coherent result according to the 1 s equivalent window length ℓ_∞. On the experimental signal, the estimated firing rate catches up the empirical firing rate (inverse of the current expert's interspike interval) after approximately 1 s too.

Table I
ALGORITHM EVALUATION THROUGH NRMSE.

	Simulation	Experiment
Noise standard deviation	0.301	-
NRMSE expert	-	0.318
NRMSE algorithm	0.306	0.321

Table II
DETECTION PHASE: SENSITIVITY AND POSITIVE PREDICTIVITY.
CLASSIFICATION PHASE: SENSITIVITY, SPECIFICITY AND ACCURACY.

	Sens.	Pred.	Sens.	Spec.	Acc.
Detection	95.0	96.4	-	-	-
Class. MU1	-	-	100	100	100
Class. MU2	-	-	79.5	91.9	88.6
Class. MU3	-	-	69.9	90.6	85.1
Class. MU4	-	-	90.0	98.0	96.5



(a) Simulated signal

(b) iEMG signal

Figure 7. Discharge rate and the mean value of four firing sources over 8 seconds (simulation) and four firing MNs over 8 seconds (experimental)

IV. DISCUSSION AND CONCLUSION

In this paper, we treated the case of signals with unknown binary inputs and unknown impulse responses, although a rough shape is necessary. A Markov model of sparse signals has been presented. The iEMG signals were modeled as a sum of independent convolved spike trains. The sparsity of the spike trains was exploited with the introduction of a PMF for the time between two spikes of the same train, which was modeled by a discrete Weibull distribution. Then, an online estimation method for the parameters of the Weibull distribution and impulse responses of the model was implemented. The method, which works on each sample sequentially, was tested on simulated and experimental iEMG signals, with promising results on the decomposition of superimposed MUAP trains.

The main aim of the study was the development of a sequential method for EMG decomposition, which can be implemented online and does not require the storage of the whole data sequence. The computational complexity is cubic

with respect to the number of sources. The computational time in the current implementation is high but the load can be largely reduced by implementing an ensemble Kalman filter (EnKF), an approximated Kalman filter designed for large state [41], and by using parallel computation. These technical computational aspects are out of the scope of this work.

A main issue of the work presented is the estimation (online) of the number of active MNs, that is a discrete parameter of the method. This issue will be addressed in future work. We plan to use a birth-and-death technique to process either the disappearance of MU (death) or the appearance of MU (birth) through the selection of new approximate shapes in a dictionary. This dictionary would be built from an offline learning stage using a clustering method to determine a rough shape of all possible isolated MUAP. For the time being, if the number of sources is underestimated, the algorithm will give poor estimations of MUAP shapes and dynamic parameters.

Solving this estimation problem will be the next step toward the use of the number of active MNs and their discharge frequencies to control external devices, by mimicking the natural way of movement control based on recruitment and rate coding of MNs. The signal processing stage presented in this paper is indeed a first stage which needs to be followed by a post-processing stage which will aim to derive the link between the obtained decomposition and the developed force or kinematics.

Appendix

A. *Proof of formulas* (3) *and* (4)

Let us drop the index i in the proof since all sources are independent, and the parameter $\mathbf{\Theta}_i$ for sake of simplicity. Let us consider a discrete time n between k-th spike at time $S[k]$ and strictly before the $(k+1)$-th spike at time $S[k+1]$ of the source. The knowledge of T^n is obviously equivalent to the one of $\{\Delta^{k-1}, T[n]\}$; then:

$$\begin{aligned}\mathsf{P}(T[n+1]=0|T[n]=t,T^{n-1}) &= \mathsf{P}(T[n+1]=0|T[n]=t,\Delta^{k-1}) \\ &= \frac{\mathsf{P}(T[n+1]=0,T[n]=t|\Delta^{k-1})}{\mathsf{P}(T[n]=t|\Delta^{k-1})}\end{aligned}$$

The event $\{T[n+1]=0,\, T[n]=t\}$ corresponds to a spike with a sojourn time $\Delta[k]$ of $t+1$, and the position of the last spike $S[k]$ at $n-t$. The event $T[n]=t$ means that the sojourn time is greater than $t+1$, and the position of the last spike is still at $n-t$; then:

$$\begin{aligned}\mathsf{P}(T[n+1]=0|T[n]=t,T^{n-1}) &= \frac{\mathsf{P}(\Delta[k]=t+1,S[k]=n-t|\Delta^{k-1})}{\mathsf{P}(\Delta[k]\geq t+1,S[k]=n-t|\Delta^{k-1})} \\ &= \frac{\mathsf{P}(\Delta[k]=t+1|S[k]=n-t,\Delta^{k-1})}{\mathsf{P}(\Delta[k]\geq t+1|S[k]=n-t,\Delta^{k-1})}\end{aligned}$$

Supposing that the last firing time $S[k]$ is independent of the sojourn time sequence Δ^k, and that this sequence is i.i.d., then:

$$\mathsf{P}(T[n+1]=0|T[n]=t,T^{n-1}) = \frac{\mathsf{P}(\Delta[k]=t+1)}{\mathsf{P}(\Delta[k]\geq t+1)}$$

which is the hazard rate $r(t)$ of the interspike process defined in (2). It is easy to show that the invariant distribution of $T[n]$, that is the solution PMF $\rho(t)$ such that

$$\rho(t') = \sum_{t\geq 0} \mathsf{P}(T[n+1]=t'|T[n]=t)\,\rho(t)$$

is given (using the fact that the sum of the survival function is nothing but the mean value of the interspike process) by:

$$\rho(t) = \frac{1-\sum_{\tau=1}^{t}\mathsf{P}(\Delta=\tau)}{\mathsf{E}\{\Delta[k]\}} = \frac{s(t+1)}{\mathsf{E}\{\Delta[k]\}}$$

The mean value of the sequence $U[n]=\delta(T[n])$ is:

$$\mathsf{E}\{\delta(T[n])\} = \mathsf{P}(T[n]=0) = \rho(0) = 1/\mathsf{E}\{\Delta[k]\}$$

B. *Closed-form ML estimation of* t_0 *when* $\beta=1$

Let us drop the index i and note $\lambda = 1-\exp\left[\frac{-1}{t_0-t_{\mathrm{R}}}\right]$. If $\beta=1$, $\mathsf{E}\{\Delta[k]|\mathbf{\Lambda}\} = t_{\mathrm{R}}+\frac{1}{\Lambda}$ and, for all $t>t_{\mathrm{R}}$, $r(t,\lambda)=\lambda$ and $s(t+1,\lambda)=(1-\lambda)^{t-t_{\mathrm{R}}}$. Denoting $M=t[1]-t_{\mathrm{R}}$, N_0 the number of firing instants and N_1 the number of instants where the sojour time is superior to the refractory period, the objective function writes:

$$J_{t^n}(\lambda) = \frac{1}{n}\ln(\lambda t_{\mathrm{R}}+1) - \frac{N_0+1}{n}\ln\lambda - \frac{N_1+M}{n}\ln(1-\lambda)$$

Optimizing $J_{t^n}(\lambda)$ with respect to λ leads to the solution:

$$\hat{\lambda} = -\frac{1}{2t_{\mathrm{R}}}\left(1-\frac{N_0t_{\mathrm{R}}-1}{N_0+N_1+M}\right) + \frac{1}{2t_{\mathrm{R}}}\sqrt{\left(1-\frac{N_0t_{\mathrm{R}}-1}{N_0+N_1+M}\right)^2 + 4t_{\mathrm{R}}\frac{N_0+1}{N_0+N_1+M}}$$

Then the solution in t_0 is given by $\hat{t}_0 = t_{\mathrm{R}} + \frac{1}{\ln\frac{1}{1-\hat{\lambda}}}$

C. *Proof or recursion* (20) *and* (21)

Using Bayes rule:

$$\begin{aligned}\mathsf{P}(\mathbf{T}^n=\mathbf{t}^n|Y^n) &= \mathsf{P}(\mathbf{T}^n=\mathbf{t}^n|Y[n],Y^{n-1}) \\ &\propto \mathrm{p}_{Y[n]|\mathbf{T}^n=\mathbf{t}^n,Y^{n-1}}(Y[n])\,\mathsf{P}(\mathbf{T}^n=\mathbf{t}^n|Y^{n-1})\end{aligned}$$

The proportionality factor $\mathrm{p}_{Y[n]|Y^{n-1}}(Y[n])$ is dropped since it does not depend on $\mathbf{t}^n$. $\mathrm{p}_{Y[n]|\mathbf{T}^n=\mathbf{t}^n,Y^{n-1}}$ is the Gaussian PDF with mean $\widehat{y}_{\mathbf{t}^n}(Y^{n-1})$ and variance $\nu_{\mathbf{t}^n}$ provided by the Kalman filter. The prediction step writes:

$$\begin{aligned}&\mathsf{P}(\mathbf{T}^{n+1}=\mathbf{t}^{n+1}|Y^n) = \\ &\quad \mathsf{P}(\mathbf{T}[n+1]=\mathbf{t}[n+1]|\mathbf{T}^n=\mathbf{t}^n,Y^n)\mathsf{P}(\mathbf{T}^n=\mathbf{t}^n|Y^n)\end{aligned}$$

Using successively the fact that given $\mathbf{T}^n$, Y^n does not bring information about $\mathbf{T}[n+1]$, the independance of the sources and the total probability law:

$$\begin{aligned}&\mathsf{P}(\mathbf{T}[n+1]=\mathbf{t}[n+1]|\mathbf{T}^n=\mathbf{t}^n,Y^n) \\ &\quad = \mathsf{P}(\mathbf{T}[n+1]=\mathbf{t}[n+1]|\mathbf{T}^n=\mathbf{t}^n) \\ &\quad = \prod_{i=1}^{n_{\mathrm{M}}}\mathsf{P}(T_i[n+1]=t_i[n+1]|T_i^n) \\ &\quad = \prod_{i=1}^{n_{\mathrm{M}}}\mathsf{E}\{\mathsf{P}(T_i[n+1]=t_i[n+1]|T_i^n,\mathbf{\Theta}_i)|T_i^n\}\end{aligned}$$

Using Markov property, and respecting (4), one obtains (21).

Acknowledgments

The authors thank Dr. C. Laine and Dr. T. Lorrain for the recording and the analysis of the experimental iEMG signals.

References

[1] B. Mambrito and C. De Luca, "A Technique for the Detection, Decomposition and Analysis of the EMG Signal," *Electroencephalography and Clinical Neurophysiology*, pp. 58:175–188, 1984.

[2] R. Lefever and C. De Luca, "A Procedure for Decomposing the Myoelectric Signal Into Its Constituent Action Potentials - Part I: Technique, Theory, and Implementation," *IEEE Trans. Biomed. Eng.*, vol. 29, no. 3, pp. 149–157, Mar. 1982.

[3] ——, "A Procedure for Decomposing the Myoelectric Signal Into Its Constituent Action Potentials - Part II: Execution and Test for Accuracy," *IEEE Trans. Biomed. Eng.*, vol. 29, no. 3, pp. 158–164, Mar. 1982.

[4] H. Marateb, S. Muceli, K. Mcgill, R. Merletti, and D. Farina, "Robust decomposition of single-channel intramuscular EMG signals at low force levels," *IEEE Trans. Autom. Control*, vol. 8, no. 6, pp. 1–13, 2011.

[5] D. Ge, E. Le Carpentier, J. Idier, and D. Farina, "Spike sorting by stochastic simulation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 3, pp. 249–259, 2011.

[6] G. Kail, J.-Y. Tourneret, F. Hlawatsch, and N. Dobigeon, "Blind deconvolution of sparse pulse sequences under a minimum distance constraint: A partially collapsed Gibbs sampler method," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2727–2743, Jun. 2012.

[7] C. Pouzat, M. Delescluse, P. Viot, and J. Diebolt, "Improved spike-sorting by modeling firing statistics and burst-dependent spike amplitude attenuation: A Markov Chain Monte Carlo approach." *J. of Neurophysiology*, vol. 91, pp. 2910–2928, 2004.

[8] R. Gut and G. S. Moschytz, "High-precision EMG signal decomposition using communication techniques," *IEEE Trans. Signal Process.*, vol. 48, no. 9, pp. 2487–2494, 2000.

[9] J. A. Herbst, S. Gammeter, D. Ferrero, and R. H. Hahnloser, "Spike sorting with hidden Markov models," *J. of Neuroscience Methods*, vol. 174, no. 1, pp. 126–134, 2008.

[10] F. Franke, M. Natora, C. Boucsein, M. Munk, and K. Obermayer, "An online spike detection and spike classification algorithm capable of instantaneous resolution of overlapping spikes," *J. of Computational Neuroscience*, vol. 29, no. 1-2, pp. 127–148, 2010.

[11] T. Haga, O. Fukayama, Y. Takayama, T. Hoshino, and K. Mabuchi, "Efficient sequential bayesian inference method for real-time detection and sorting of overlapped neural spikes," *J. of Neuroscience Methods*, vol. 219, no. 1, pp. 92–103, 2013.

[12] C. Castellini, P. van der Smagt, G. Sandini, and G. Hirzinger, "Surface EMG for force control of mechanical hands," in *Proc. 2008 IEEE Robotics and Automation Conf.*, May 2008, pp. 725–730.

[13] E. Guizzo and H. Goldstein, "The rise of the body bots [robotic exoskeletons]," *IEEE Spectrum*, vol. 42, no. 10, pp. 50–56, Oct. 2005.

[14] M. Reaz, M. Hussain, and F. Mohd-Yasin, "Techniques of EMG signal analysis: detection, processing, classification and applications," *Biological Procedures Online*, vol. 8, pp. 11–35, 2006.

[15] D. Farina and F. Negro, "Accessing the neural drive to muscle and translation to neurorehabilitation technologies," *IEEE Reviews in Biomedical Engineering*, vol. 5, pp. 3–14, 2012.

[16] F. Negro, A. Holobar, and D. Farina, "Fluctuations in isometric muscle force can be described by one linear projection of low-frequency components of motor unit discharge rates." *J. of Physiology*, vol. 587, no. Pt 24, pp. 5925–38, 2009.

[17] 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.

[18] D. Farina, A. Crosetti, and R. Merletti, "A model for the generation of synthetic intramuscular EMG signals to test decomposition algorithms," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 1, pp. 66–77, Jan. 2001.

[19] J. Monsifrot, E. Le Carpentier, D. Farina, and Y. Aoustin, "Sequential estimation of intramuscular EMG model parameters for prosthesis control," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Syst., Workshop on Robotics for Neurology and Rehabilitation*, San Francisco, Ca, 2011.

[20] Y. Li, L. Smith, L. Hargrove, D. Weber, and G. Loeb, "Sparse Optimal Motor Estimation (SOME) for extracting commands for prosthetic limbs," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 1, pp. 104–111, Jan. 2013.

[21] R. Pyke, "Markov renewal processes: Definitions and preliminary properties," *The Ann. of Math. Stat.*, vol. 32, no. 4, pp. 1231–1242, 1961.

[22] 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 Stat. Springer New York, 2008, vol. 191, pp. 1–30.

[23] T. Nakagawa and S. Osaki, "The discrete Weibull distribution," *IEEE Trans. Reliab.*, vol. R-24, no. 5, pp. 300–301, Dec. 1975.

[24] R. Lockhart and M. Stephens, "Estimation and tests of fit for the three-parameter Weibull distribution," *J. of the Roy. Stat. Soc.. Series B (Methodological)*, vol. 56, no. 3, pp. 491–500, 1994.

[25] C. de Luca and W. Forrest, "Some properties of motor unit action potential trains recorded during constant force isometric contractions in man," *Kybernetik*, vol. 12, pp. 160–168, 1973.

[26] T. Schon, F. Gustafsson, and P.-J. Nordlund, "Marginalized particle filters for mixed linear nonlinear state-space models," *IEEE Trans. Signal Process.*, vol. 53, pp. 2279–2289, 2005.

[27] R. Kalman, "A new approach to linear filtering and prediction problems," *Trans. of the ASME - J. of Basic Eng.*, pp. 35–45, Mar. 1960.

[28] A. Doucet, N. De Freitas, and N. Gordon, *Sequential Monte Carlo methods in practice*. Springer, 2001.

[29] 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.

[30] L. Ljung and T. Söderström, *Theory and practice of recursive identification*. Massachusetts and London: The MIT Press, 1983.

[31] N. Schraudolph, J. Yu, and S. Guenter, "A stochastic quasi-Newton method for online convex optimization," in *Proc. of 11th Int. Conf. on Artificial Intell. and Stat.*, 2007.

[32] S. Yi, D. Wierstra, T. Schaul, and J. Schmidhuber, "Stochastic search using the natural gradient," in *Proc. of the 26th Annu. Int. Conf. on Mach. Learning*, Montreal, Quebec, Canada, 2009, pp. 1161–1168.

[33] R. M. Studer, R. J. de Figueiredo, and G. S. Moschytz, "An algorithm for sequential signal estimation and system identification for EMG signals," *IEEE Trans. Biomed. Eng.*, no. 3, pp. 285–295, 1984.

[34] Q. Xia, M. Rao, Y. Ying, and X. Shen, "Adaptive fading Kalman filter with an application," *Automatica*, vol. 30, no. 8, pp. 1333–1338, 1994.

[35] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes - Theory and Application*. Prentice-Hall, Inc., 1993.

[36] M. Stephens, "Bayesian analysis of mixture models with an unknown number of components - An alternative to reversible jump methods," *The Ann. of Stat.*, vol. 28, no. 1, pp. 40–74, 2000.

[37] S. Särkkä, A. Vehtari, and J. Lampinen, "Rao-Blackwellized particle filter for multiple target tracking," *Inform. Fusion*, vol. 8, no. 1, pp. 2–15, 2007.

[38] M. Verhaegen and P. Van Dooren, "Numerical aspects of different Kalman filter implementations," *IEEE Trans. Autom. Control*, vol. 31, no. 10, pp. 907–917, 1986.

[39] K. McGill, Z. Lateva, and H. Marateb, "EMGLAB: An interactive EMG decomposition program," *J. of Neuroscience Methods*, vol. 149, no. 2, pp. 121–133, Dec. 2005.

[40] D. Farina, R. Colombo, R. Merletti, and H. Baare Olsen, "Evaluation of intra-muscular EMG signal decomposition algorithms," *J. of Electromyography and Kinesiology*, vol. 11, pp. 175–187, 2001.

[41] G. Evensen, "The Ensemble Kalman Filter: theoretical formulation and practical implementation," *Ocean Dynamics*, vol. 53, no. 4, pp. 343–367, Nov. 2003.

Jonathan Monsifrot was born in Ris-Orangis, France, on 1986. He received the M.Sc. degree in Signal processing from Ecole Centrale Nantes, France, in 2009 and the Ph.D. degree from the University of Nantes, France, in 2013. His research interests are modelling system in probabilistic frameworks, bayesian inference and sequential estimation.

Eric Le Carpentier received the Ph.D. degree in automatic control from the Nantes University, Nantes, France, in 1990. Since 1992, he has been an Associate Professor at Ecole Centrale de Nantes, France, and at the Institut de Recherche en Communications et Cybernétique de Nantes. His current research interests include signal processing, stochastic simulation for estimation, detection, tracking and control, with applications in biomedical signals, robotics and precise geo-location.

Yannick Aoustin is currently a Lecturer Excluding class at the Department of Physics, University of Nantes, France. He earned his Ph.D. degree in Automatic in 1987 and his research degree for leading research of PhD students in 2006. His research interests include mechanical systems under actuated legged robots, bipedal, nonlinear observers, and biomechanics, He has many publications on journals circulated internationally and premier conferences. He is also an associate editor of International Journal of Advanced Robotic Systems.

Dario Farina (M'01-SM'09) received the M.Sc. degree in electronics engineering from Politecnico di Torino, Italy, in 1998, and the Ph.D. degrees in automatic control and computer science and in electronics and communications engineering from the Ecole Centrale de Nantes, Nantes, France, and Politecnico di Torino, respectively, in 2002. In 2002 through 2004, he was a Research Assistant Professor at Politecnico di Torino and in 2004 to 2008 an Associate Professor in biomedical engineering at Aalborg University, Aalborg, Denmark. From 2008 to 2010, he was a Full Professor in motor control and biomedical signal processing and Head of the Research Group on neural engineering and neurophysiology of movement at Aalborg University. In 2010, he was appointed Full Professor and Founding Chair of the Department of Neurorehabilitation Engineering, University Medical Center Göttingen, Georg-August University, Germany, within the Bernstein Center for Computational Neuroscience. He is also the Chair for NeuroInformatics of the Bernstein Focus Neurotechnology Göttingen. His research focuses on biomedical signal processing, modeling, neurorehabilitation technology, and neural control of movement. Within these areas, he has (co)-authored approximately 300 papers in peer-reviewed journals and over 300 among conference papers/abstracts, book chapters and encyclopedia contributions. Since 2012, Dr. Farina has been the President of the International Society of Electrophysiology and Kinesiology (ISEK). He is the recipient of the 2010 IEEE Engineering in Medicine and Biology Society Early Career Achievement Award for his contributions to biomedical signal processing and to electrophysiology and a Fellow of the American Institute for Medical and Biological Engineering (AIMBE). He is an Associate Editor of Medical and Biological Engineering and Computing and of the Journal of Electromyography and Kinesiology. He is an Associate Editor of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING.