Evaluation of intra-muscular EMG signal decomposition algorithms

Dario Farina [a, b], Roberto Colombo [c], Roberto Merletti [a,*], Henrik Baare Olsen [d]

[a] Centro di Bioingegneria, Dip. di Elettronica, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, Italy
[b] Dép. d'Automatique et Informatique appliquée, Ecole Centrale de Nantes, Nantes, France
[c] Servizio di Bioingegneria, Fondazione Salvatore Maugeri, IRCCS, Veruno, Italy
[d] Department of Physiology, National Institute of Occupational Health, Copenhagen, Denmark

Abstract

We propose and test a tool to evaluate and compare EMG signal decomposition algorithms. A model for the generation of synthetic intra-muscular EMG signals, previously described, has been used to obtain reference decomposition results. In order to evaluate the performance of decomposition algorithms it is necessary to define indexes which give a compact but complete indication about the quality of the decomposition. The indexes given by traditional detection theory are in this paper adapted to the multi-class EMG problem. Moreover, indexes related to model parameters are also introduced. It is possible in this way to compare the sensitivity of an algorithm to different signal features. An example application of the technique is presented by comparing the results obtained from a set of synthetic signals decomposed by expert operators having no information about the signal features using two different algorithms. The technique seems to be appropriate for evaluating decomposition performance and constitutes a useful tool for EMG signal researchers to identify the algorithm most appropriate for their needs.

Keywords: Electromyography; EMG signal decomposition; EMG modeling; Detection theory

1. Introduction

The study of the intra-muscular EMG signal is of great interest both for clinicians and for basic researchers because it allows the identification of single motor unit action potentials (MUAPs) with the possibility of investigating morphological abnormalities of the EMG signal and recruitment strategies of the central nervous system. The issue of decomposing an EMG signal into the constituent MUAPs is fascinating because it allows extraction of order from an apparently random signal, opens up new areas of research and triggers new questions. The ability to monitor motor unit (MU) recruitment, derecruitment and firing patterns leads in fact to the understanding of motor control strategies, myoelectric manifestations of fatigue, neuromuscular disorders and a variety of open issues in ergonomics, geriatrics, sport and rehabilitation medicine [2,6,9,13,17,19].

Since the pioneering work done in the early 80s in this field by the group led by De Luca [14,15], many techniques have been developed to implement EMG signal decomposition with various degrees of automation [9–12,16,20,22]. Diagnostic applications and the commercial success of these techniques have been lagging behind despite the enthusiasm of researchers and the number, quality and significance of scientific publications.

Decomposition algorithms are being developed and published. A very important aspect in this field is the evaluation of performance, a point not deeply investigated in the past. In order to obtain a reference decomposition result, different methods have been proposed [5]:

1. synthetic signals can be generated as a reference (model approach);
2. real signals decomposed manually can be considered as the reference;
3. recordings (by multiple electrode surfaces) from the same MU at different locations can be decomposed and the results compared; in this case the probability of incorrectly decomposing the different signals and yet having the same firing patterns is low, so when

* Corresponding author. Tel.: +39-011-5644137; fax: +39-011-5644099.

E-mail address: merletti@polito.it (R. Merletti).

the decomposition results agree for all the channels the operation is considered correct [5].

The first approach has important advantages with respect to the others [8]. It provides a priori knowledge of the exact positions and features of all the MUAPs in the signal. Moreover, a complete model is the only way to test the algorithms with signals having selected characteristics in order to test the sensitivity to different parameters. In any case it is necessary to introduce indexes of performance computed from the comparison between the result obtained with the algorithm under test and the reference.

In the following we will describe a tool to evaluate decomposition performance by having a reference decomposition result obtained by simulated signals generated by a recently developed EMG model [8]. The paper is organized as follows. In Section 2 the model proposed in [8] is briefly described and a library of 18 signals is proposed to test algorithms. In Section 3 the indexes of performance are introduced, and in Section 4, a comparison of the performance of two algorithms is reported. In Section 5, conclusions are drawn.

2. Overview of the generation model and library of test signals

A mathematical model, based on Associated Hermite (AH) series expansions of real MUAPs, to generate synthetic intra-muscular EMG signals has been developed [8]. The model receives as input about 20 parameters. They assume different importance for the various decomposition steps and permit to characterize the following EMG signal features:

- the degree of shape similarity between a MUAP detected from different detection surfaces (multi-channel detection);
- the degree of shape similarity between MUAPs of different MUs;
- the shape changes of MUAPs belonging to the same MU during time;
- the firing statistics;
- the recruitment and de-recruitment of individual MUs;
- the degree of superposition of the MUAPs in the signal;
- the amount of additive noise.

In particular, the model parameters are grouped in seven classes:

Group 1. Some algorithms have been designed to work with only one channel, others with more than one. A proper comparison must take into account the amount of information added in the case of multiple channel recordings. This is obtained by two parameters:

1. N_c is the number of generated channels.
2. v_c indicates the difference between the MUAP shapes in the different channels; the MUAP shapes of the other channels are created from the MUAP shapes of the first channel and v_c indicates how much the MUAPs in the first channel have been modified to create the others.

Group 2. The second group is related to the difference between energies and shapes of MUAPs belonging to different MUs (classes) and plays a role in the segmentation and classification phases:

1. M is the number of generated classes, that is the number of MUs.
2. ΔE_{max} is the maximum normalized energy difference between the representative MUAPs of each class; the representative MUAP of each class is defined as the one generated by the set of AH coefficients obtained by expanding a real MUAP, taken from a library (the MUAPs in the train are created by varying the AH coefficients of the representative MUAP).
3. r defines the difference in shape between the representative MUAPs of each class; it is defined as the ratio between the minimum Euclidean distance between the representative MUAPs and the square root of their mean energy.

Group 3. The third group is related to what we have defined as intra-class variability, i.e. the phenomena which make each MUAP in a motor unit action potential train (MUAPT) different from the others in the same train. The parameters in this group describe how much the waveforms belonging to the same class can vary along the MUAPT and tests the flexibility of the algorithm in the classification step and its capability of tracking progressive changes of the waveforms.

1. v_{w1} is related to random shape variability which is a global variation of the shape of the MUAP in a random way during time.
2. v_{w2} is related to shape variability with a trend in time; trend shape variability is the variability which occurs at constant steps during time (the MUAP changes shape gradually and is similar to the next and the previous in the same MUAPT but quite different from one distant in time).
3. v_{s1} and v_{s2} assume the same meaning as v_{w1} and v_{w2} for the time scale (width of the waveform) variability.

Group 4. The fourth group is related to the regularity of the firing patterns.

1. $N_{\%d}$ is the percentage of firings with respect to the total which are double discharges. Double discharge firings are firings 2–15 ms far from the previous firing belonging to the same MU [1].
2. N_{td} is the number of trains for which double discharge firings occur.
3. $N_{\%b}$ is the percentage of firings with respect to the total which are inserted in the signal in a random way (that is not according to any firing statistic).
4. $\{f_i\}_{i=1}^{M}$ are the mean firing frequencies for the M MUs.
5. J indicates the coefficient of variation (standard deviation/mean) of the interpulse interval (the distribution is gaussian).

Group 5. The fifth group describes the degree of superposition and tests the ability of the algorithm to correctly separate and classify superimposed waveforms.

1. $N_{\%s}$ is the percentage of MUAPs with respect to the grand total which are superimposed with others.
2. $N_{s,max}$ is the maximum number of MUAPs which occur in a superposition.
3. $\alpha_{\%,max}$ is the maximum degree of superposition between MUAPs (100% corresponds to the alignment of the medians of energy of all the waveforms involved in the superposition).

Group 6. The sixth group describes non-stationary firing patterns. It is possible to select a particular set of activation intervals for each MU and different frequency modulations.

1. *Type* is the type of activation interval pattern selected (simulating increasing–decreasing ramp force contractions, random activation intervals etc.).
2. v_f is the desired rate of change of the firing frequencies, which are increased or decreased depending on the type of activation intervals.

Group 7. The seventh group characterizes the additive noise. Although the noise associated to EMG signals may contain contributions caused by depolarizations lying further away from the electrode position [3], we have decided to use for the purposes of this paper a very simple noise model. The noise is white gaussian filtered noise with a band-pass filter with cut-off frequencies 50 and 5000 Hz. This was done for the sake of simplicity to reduce the number of parameters in the representative evaluation presented.

For technical details about the generation model the reader is invited to refer to [8].

Fig. 1 shows an example of simulated firing patterns of a signal generated by a particular set of parameters. We have used the proposed model for generating a limited subset of signals to present an example of evaluation of decomposition algorithms. With a limited number of test signals the sensitivity of the decomposition algorithms cannot be evaluated separately for all the parameters introduced in the model.

A set of 18 synthetic recordings has been generated using a sampling frequency of 51,200 Hz and real MUAPs extracted from wire EMG signals with a bandwidth of approximately 100–2500 Hz. All the signals have a duration of 10 s. The signal to noise ratio (SNR) is 20 dB for all the signals except for signals 14, 15 and 18 with a SNR of 15 dB, for signals 9 and 17 with SNR equal to 12 dB and for signals 10 and 16 with very low SNR (5 dB). All the signals are one-channel recordings with the exception of signals 14 and 18, which consist of three channels. The number of MUs is between 6 and 11.

Signals 1 and 2 (with 7 and 8 MUs, respectively) present regular statistics, no superpositions and constant mean firing frequencies. The shape difference between representative MUAPs is higher for signal 1 with respect to signal 2. Signals 3, 4 and 5, 6 (with 10, 8, 10 and 7 MUs, respectively) have similar characteristics as signals 1 and 2 but double discharge and random firings (for 3 and 4) and intra-class variability (for 5 and 6) are included. In the remaining signals (except for signal 11) superpositions have been included and for signals 11–18 re-derecruitment of MUs and variation with time in MU mean firing frequencies are introduced. Signals 7 and 8 (with 7 and 6 MUs, respectively) have the same characteristics (no irregular firings, no intra-class variability, constant firing frequencies and superpositions included) but show different maximum degree of superposition of MUAPs and different percentage of superimposed MUAPs (75% and 20% for signal 7 and 100% and 40% for signal 8). Signals 9 and 10 (with 7 and 6 MUs, respectively) have similar features as 7 and 8 but lower SNRs (12 and 5 dB, respectively). Signals 11 and 12 (both with 8 MUs) simulate an increasing followed by a decreasing force contraction. Intra-class variability is included and superpositions are present for signal 12. Signals 13 and 14 (with 8 and 11 MUs, respectively) have no intra-class variability but superpositions are included; signal 13 presents random activation intervals while signal 14 simulates an increasing force contraction. Signals 15–18 (with 8, 8, 10 and 6 MUs, respectively) simulate ramp contractions, have low SNRs, intraclass variability, superpositions and irregular firings.

The generated signals have a large variety of different features. In case specific characteristics of an algorithm need to be tested, appropriate sets of signals must be generated. The proposed library of signals should not be considered as a standard for EMG signal decomposition evaluation but just as an example of a set of signals with general characteristics. Note, for example, that all the signals have been generated with a very simple noise model and with the same sampling frequency.

The library of test signals has been distributed among

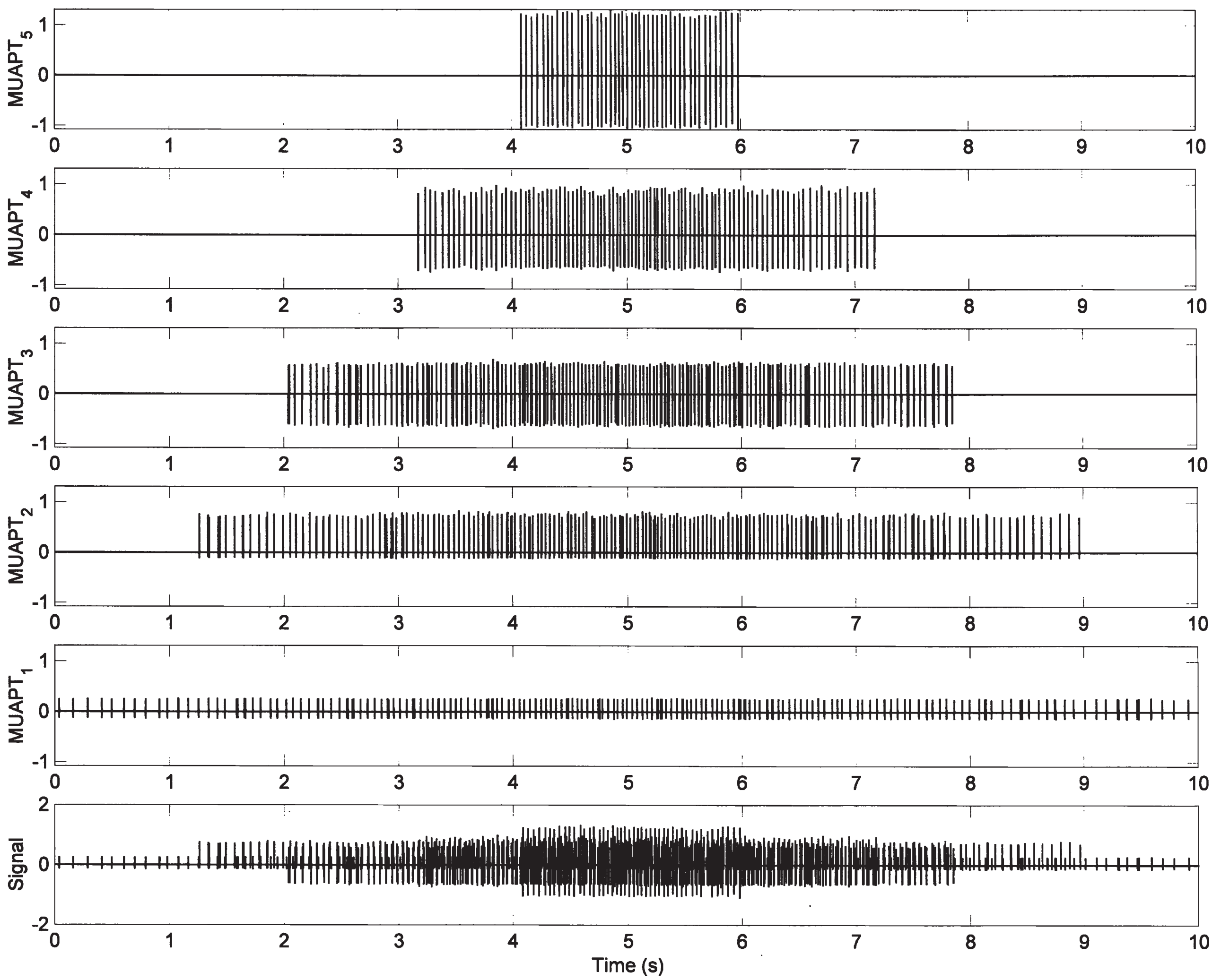

Fig. 1. Example of a synthetic signal with firing patterns generated to simulate an increasing followed by a decreasing ramp contraction. Five MUs with varying firing frequencies and double discharge firings have been simulated.

a number of laboratories which were asked to decompose the signals without any information about their characteristics except for the sampling frequency, duration and number of channels. A questionnaire, asking for information related to the time needed for the decomposition process and to the perceived quality of the results, has also been completed by the persons who decomposed the signals.

3. Indexes of performance

Indexes of performance are introduced to evaluate and compare the algorithms. The definition of indexes to compactly describe the quality of the decomposition process can be partly derived from traditional detection theory [7] adapted to the multi-class problem of EMG signal decomposition. Moreover, the use of a reference decomposition result (in this case given by synthetic signals) introduces problems related to the definition of event occurrences and correspondence between reference and detected classes. Since we will use as reference signals generated by a model, in the following we will refer to the model result rather than to the reference result. It is nonetheless clear that the proposed indexes of performance can be used with any suitable reference assumed correct (for example, real signals decomposed manually).

Any decomposition algorithm operates on the signal in two basic steps: the segmentation and the classification phase. The segmentation phase consists of the detection of the MUAPs in the signal and the classification phase consists of their assignment to clusters (MUs). It is useful to characterize an algorithm's performance separately for these two steps. Note that the decomposition process can be divided into more than two steps (resolving superpositions is usually performed separately, for example) but from the result point of view the whole process can be separated into segmentation, the capability of detecting an event, and classification, the ability to differentiate between different MU activities.

The first issue to address is the assignment of model classes to detected classes. In particular, problems arise when the number M_A of detected MUs differs from the generated number M_I of MUs [4]; in this case it is not trivial to determine the correspondence between detected and model classes (MUs). The establishment of correspondence between classes implies the definition of a criterion to assign a detected occurrence to an effective one, taking into account that the firing instant is defined in a different way by different algorithms and by the generation model. For example, the firing instant can be defined as the time corresponding to the median of energy, to the peak value, to the "beginning" or "end" of the MUAP, to a zero crossing etc. So the difference in the firing instant definition can be as high as 3–4 ms (for example, in the case of the difference between the beginning and the end of the waveform) and not negligible. Therefore, it is expected to find a systematic difference (offset in the detection) together with a random difference between the model and the detected times of occurrence. The systematic error should not be taken into account (it must be estimated and compensated for in some way) while the random error should be limited because it determines an error in instantaneous firing rate estimation. If an error δt_o is present in the detection of firing time, the error δf_r in instantaneous firing rate estimation is given by:

$$\delta f_r = 2 f_r^2 \delta t_o$$

with f_r being the instantaneous firing rate.

If we consider 40 Hz as the upper limit of the firing rate, in the case of δt_o=0.5 ms we obtain a relative error in f_r of 4% which has been considered acceptable.

So, after the estimation and compensation for the systematic error, a detection will be considered correct if the estimated time of occurrence is within ±0.5 ms from a real one. The problem is how to assign a detected class to a model class. Note that the systematic error depends on the class; in fact, for example, the time difference between the end of a MUAP and the beginning depends on the duration of the waveform. For this reason it is not possible to estimate the systematic error before the establishment of class correspondence. It is also worth noting that the averaged template of each detected class cannot be used to define a correspondence between classes because it is influenced by classification errors and its definition can change with the algorithm. Moreover, in the case of variation of MUAP shape along the train, the definition of an averaged template over the whole train is meaningless.

3.1. Association between model and detected classes

A class is characterized by the times of occurrence of its MUAPs. The times of occurrence is the only information used to establish the correspondence between classes since the matching between averaged templates would depend on the classification errors, as explained previously. To determine the correspondence between a model and a detected class, each model class is compared with all the detected classes (given by the algorithm under test) by computing the distances between the model times of occurrence and the estimated ones. The distances are computed in order to assign a detected MUAP to the closest model MUAP, that is at each step the distance between the nearest model and detected occurrence is computed and the correspondence selected:

$$^0\Delta_k^{IA} = \min_{i,a}\{|t_i - t_a|\} \quad k=1, \ldots, \min\{N_I, N_A\} \tag{1}$$

with t_i the time of occurrence of the model MUAP i of model class I, t_a the time of occurrence of the detected MUAP a of detected class A, N_I the total number of generated MUAPs of model class I and N_A the total number of detected MUAPs of detected class A. After association, the associated MUAPs are excluded from the computation of the next distances.

Note that the offset is not known a priori so at this step it is assumed equal to zero and not taken into account in Eq. (1). The estimation of the offset in the detection is described below. It is possible to give a graphical representation of the computed distances by plotting their histograms, as shown in Fig. 2.

M_A histograms (class bins of 1 ms) are in this way associated to each model generated class. The offset is estimated at this point from the vectors $^0\Delta^{IA}$. It is computed as the mean of the distances in the bin corresponding to the peak value of the histogram of $^0\Delta^{IA}$. The distances are then computed again subtracting the estimated offset θ^{IA}, giving the definitive Δ^{IA}:

$$\Delta_k^{IA} = \min_{i,a}\{|t_i - t_a - \theta^{IA}|\} \quad k=1, \ldots, \min\{N_I, N_A\} \tag{2}$$

Note that for each model class I, M_A offset values θ^{IA} are estimated, corresponding to the offset values to be applied in case of correspondence between model class I and each of the M_A detected classes.

Then for each vector of distances the number N_{AI} of occurrences within an interval of 1 ms centered at zero (now that the offset has been compensated for) is computed. These occurrences are assumed as true classified positives (MUAPs correctly detected and classified) and the number N_{AI} corresponds to the number of true classified positives in the case of assuming a correspondence between the model class I and the detected class A.

The model class I is associated to the detected class A which corresponds to the maximum N_{AI} (that is, each model class is associated to the detected class resulting in the maximum number of true classified positives).

It is possible that more than one model class is in this

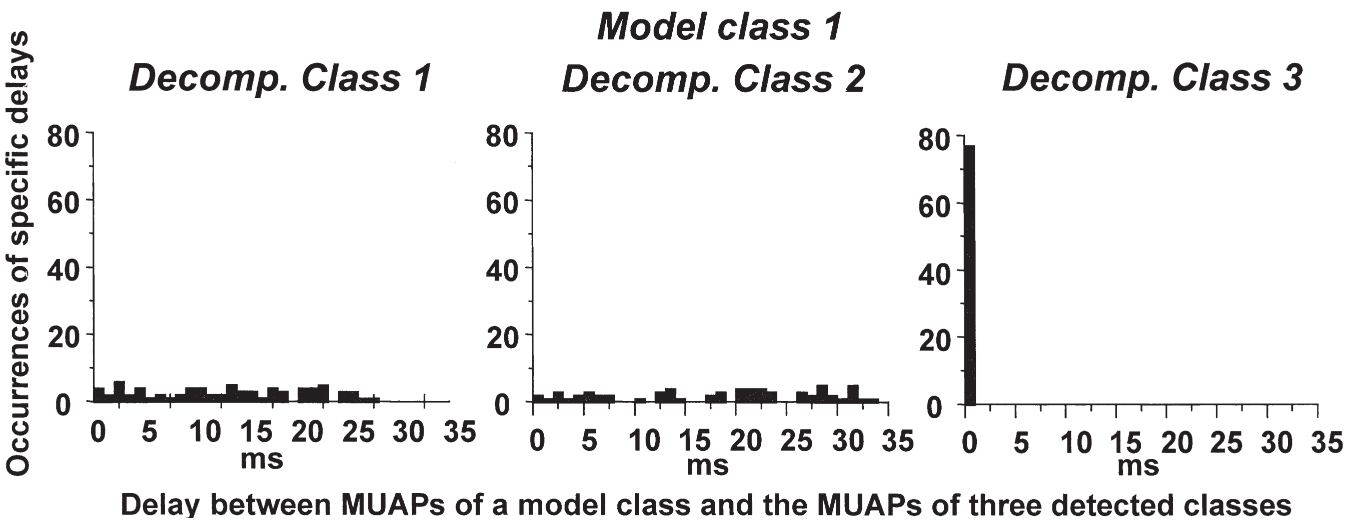

Fig. 2. Graphical representation of the computed time delays between model and detected occurrences. A model class is compared with three detected classes. The offset is not yet compensated for and is estimated as the average of the time delays in the peak bin of the histograms.

way associated to the same detected class (merge of two classes); in this case the correspondence resulting in the greatest number of true classified positives is chosen. The not selected model classes are associated to other not associated detected classes with the same criterion of maximizing the number of true classified positives. If there are no other detected classes not yet associated, the remaining model classes are not assigned and are considered missed. A class is also considered missed if it corresponds to a number of true classified positives smaller than 20% of the total number of actual MUAPs belonging to that model class.

The split of one model class into two or more detected classes is treated by choosing the association with the same criterion used in the case of merge of classes (maximize the number of true classified positives). In this way the model classes are associated to the detected ones and in the case of a different number of model and detected classes the assignment is made in order to generate the maximum total number of true classified positives.

At the end of the assignment procedure for each model class the information about the associated detected class and the offset in the MUAP detections is available. Note that for each model class a different offset value is possible.

3.2. *Detection phase*

The detection phase can be viewed as a traditional two-class detection problem. True positives are in this case defined as the detected occurrences corresponding to actual occurrences (independent of the classification step) and consist of the detected occurrences correctly classified plus the occurrences detected but incorrectly classified. The first group, which coincides with the true positives of the classification phase, can be derived directly from the class assignment procedure. The second group is identified by comparing the not associated model times of occurrence with the not associated detected times of occurrence. The comparison is made by compensating for the estimated offset in the detection, that is from each model time of occurrence the offset of the model class to which it belongs is subtracted.

False negatives are in this phase defined as actual occurrences not detected by the algorithm. The number of false negatives ${}^{(s)}$FN is the difference between the total number of model MUAPs and the number of true positives ${}^{(s)}$TP.

Finally, false positives are detected occurrences not corresponding to effective ones; their number ${}^{(s)}$FP is the difference between the total number of detected occurrences and ${}^{(s)}$TP.

True negatives cannot be computed for the detection phase.

${}^{(s)}$TP and ${}^{(s)}$FN are also computed separately for the superimposed MUAPs (${}^{(s)}\mathrm{TP}_s$, ${}^{(s)}\mathrm{FN}_s$) and for the MUAPs with irregular firing statistics (double discharge MUAPs or MUAPs placed randomly in the signal) (${}^{(s)}\mathrm{TP}_r$, ${}^{(s)}\mathrm{FN}_r$).

The significance of these definitions is reported in Fig. 3(a).

The traditional sensitivity ${}^{(s)}$Se and positive predictivity ${}^{(s)}P$ [7] for the detection phase can therefore be defined as:

$$^{(s)}\mathrm{Se}=\frac{^{(s)}\mathrm{TP}}{^{(s)}\mathrm{TP}+{}^{(s)}\mathrm{FN}} \quad \text{for each not missed model class}$$

$$^{(s)}P=\frac{^{(s)}\mathrm{TP}}{^{(s)}\mathrm{TP}+{}^{(s)}\mathrm{FP}}$$

${}^{(s)}$Se is computed (a) for each not missed model class,

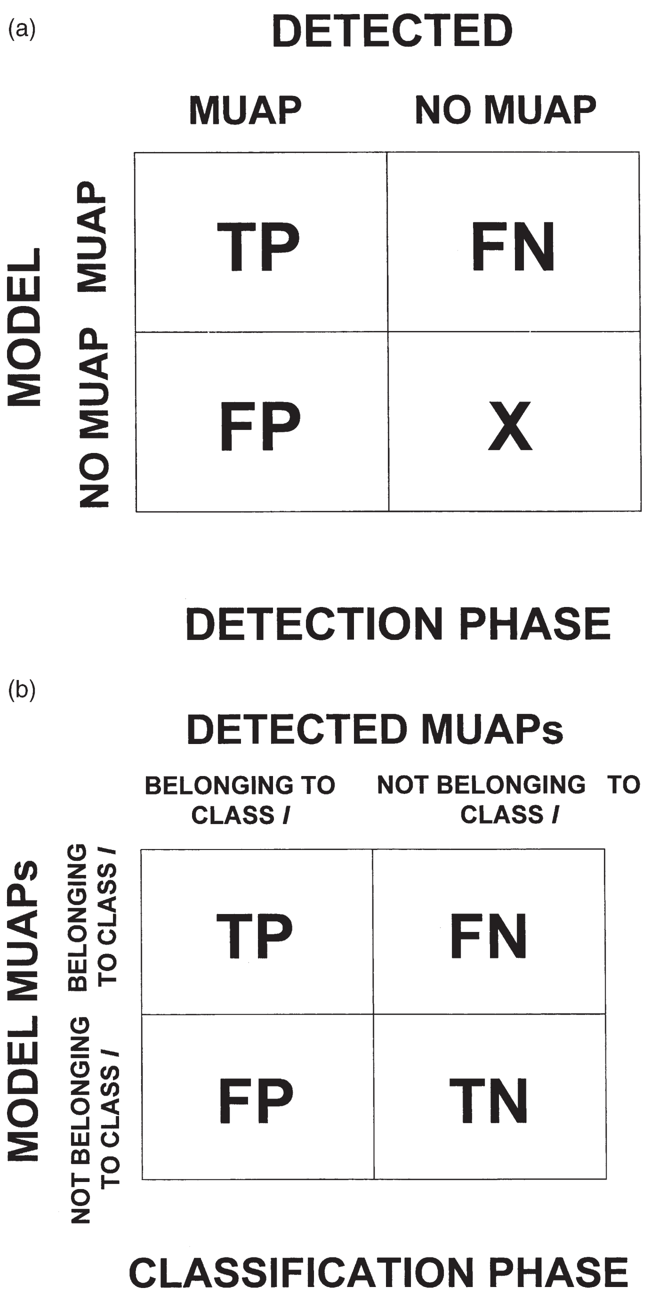

Fig. 3. Definition of true positives, false positives, true negatives and false negatives for the detection (a) and for the classification (b) phase (for class I).

(b) as the mean on all the not missed model classes, (c) for the superimposed firings (${}^{(s)}Se_s$) and (d) for the irregular firings (${}^{(s)}Se_r$).

3.3. Classification phase

True positives, false positives, true negatives and false negatives have a different definition for the classification phase with respect to that relative to the detection phase.

True positives for class I are the detected occurrences correctly classified as belonging to class I and their number ${}^{(c)}TP$ is computed directly during the assignment procedure.

False positives for class I are detected occurrences corresponding to actual occurrences but incorrectly classified as belonging to class I. Their number ${}^{(c)}FP$ is the difference between the number of correctly detected occurrences classified in class I and ${}^{(c)}TP$ of class I.

True negatives are, for each not missed model class I, the number of detected occurrences not belonging to class I and not assigned by the algorithm to class I. Note that the classification of an occurrence can be incorrect and the occurrence be a true negative. For example, in the case of four classes, if a MUAP belonging to class 3 is classified in class 4 it is a true negative for class 1 and 2.

False negatives for class I are correctly detected occurrences of class I incorrectly classified to a class different from I.

Fig. 3(b) summarizes the meaning of the definitions introduced.

From ${}^{(c)}TP$, ${}^{(c)}FP$, ${}^{(c)}TN$ and ${}^{(c)}FN$ it is possible to compute sensitivity, specificity and accuracy of the algorithm for each class. They are defined as follows [7]:

$${}^{(c)}Se=\frac{{}^{(c)}TP}{{}^{(c)}TP+{}^{(c)}FN} \quad \text{for each not missed model class}$$

$${}^{(c)}Sp=\frac{{}^{(c)}TN}{{}^{(c)}TN+{}^{(c)}FP} \quad \text{for each not missed model class}$$

$${}^{(c)}Ac=\frac{{}^{(c)}TP+{}^{(c)}TN}{{}^{(c)}TP+{}^{(c)}TN+{}^{(c)}FP+{}^{(c)}FN} \quad \text{for each not missed model class}$$

These indexes are computed (a) for each not missed model class, (b) as the mean on all the not missed model classes, (c) for superimposed firings (${}^{(c)}Se_s$, ${}^{(c)}Sp_s$, ${}^{(c)}Ac_s$) and (d) for irregular firings (${}^{(c)}Se_r$, ${}^{(c)}Sp_r$, ${}^{(c)}Ac_r$).

The indexes introduced until now give an indication about detection and classification phases. Nevertheless it is important to underline that it is possible that some detected MUAPs not assigned to any class are not included in the final result of the decomposition. The indexes related to the detection phase can thus indicate different performance depending on the way in which unclassified MUAPs are treated by the algorithm (disregarded or assigned to special classes). In any case the indexes proposed will give an indication of the decomposition result obtained by the user, that is about the information the user receives after the decomposition process.

3.4. Global performance indexes

A few indexes are introduced to describe the performance of the estimation of global parameters in order to give an indication of the algorithm's performance when general features of the signal rather than exact firing patterns are of interest.

Two indexes to characterize the ability of the algorithm to estimate the number of active MUs are introduced. The first is the difference between the number of model classes and the number of classes estimated by the algorithm before the previously described assignment procedure (Δc_1) and may be positive or negative. The second is the difference between the number of model classes and the number of detected classes associated to model classes after the assignment procedure (Δc_2), so it represents the number of missed classes and can assume only positive values. For example, we may have 10 model classes and 12 detected classes resulting in $\Delta c_1=-2$. Nine of the detected classes may be associated to nine model classes corresponding to $\Delta c_2=+1$. Thus, in this example, we have two extra detected classes before assignment and one missed class and this may indicate that a class has not been detected by the algorithm while three classes have been split into two classes each.

Other global indexes are related to mean firing rate estimation and activation interval detection. The normalized difference between the estimated mean firing rate and the actual mean firing rate of each not missed model class is Δf_i (defined only for constant firing rates). Note that the mean firing rate is in this case simply computed as the inverse of the mean interpulse interval taking into account all the detected firings; better estimations can be obtained by more sophisticated techniques, such as the one proposed in [21].

Finally, τ_p is defined as the percentage of time for which the algorithm indicates correctly that the MU is active with respect to the total activation time. τ_n is the percentage of time for which the algorithm indicates correctly that the MU is not active with respect to the total non-activation time (Fig. 4). MU i is considered not active after the time of occurrence of a MUAP belonging to it and not followed by another MUAP of MU i within an interval of 200 ms (corresponding to an instantaneous frequency of 5 Hz).

Δc_1, Δc_2, Δf_i, τ_p and τ_n give an indication of the quality of the global information obtained by the decomposition process.

4. Example application for algorithms A and B

Two algorithms have been tested with the library of 18 test signals described above. The signals have been decomposed with the two algorithms by expert operators who had no information about the signal characteristics. Each operator answered a questionnaire about the time required for the decomposition process and about his/her perceived complexity of the signals. The operators agreed that most of the 18 signals looked like real ones. Algorithm B has no manual parts while algorithm A permits a manual decision in the case of ambiguous classifications. Due to the manual editing, the time required to decompose a signal with algorithm A is considerably higher than that required by algorithm B. Moreover, algorithm A operates with three-channel signals while algorithm B can operate only with one-channel signals. In the case of one-channel signals the same signal has been used by algorithm A for all three channels, while in the case of three-channel signals the first channel has been used by algorithm B for the decomposition. Finally, algorithm A operates only with a sampling frequency of 51,200 Hz (this is the reason why all the signals were generated with that sampling frequency) while algorithm B permits one to select the sampling frequency.

Fig. 5 shows an example of the result of the decomposition process compared with the reference model.

The difference in the number of detected and model classes was zero for algorithm A in 17 out of 18 cases; one resulted in $\Delta c_1=\Delta c_2=+1$ (one missed class). Algor-

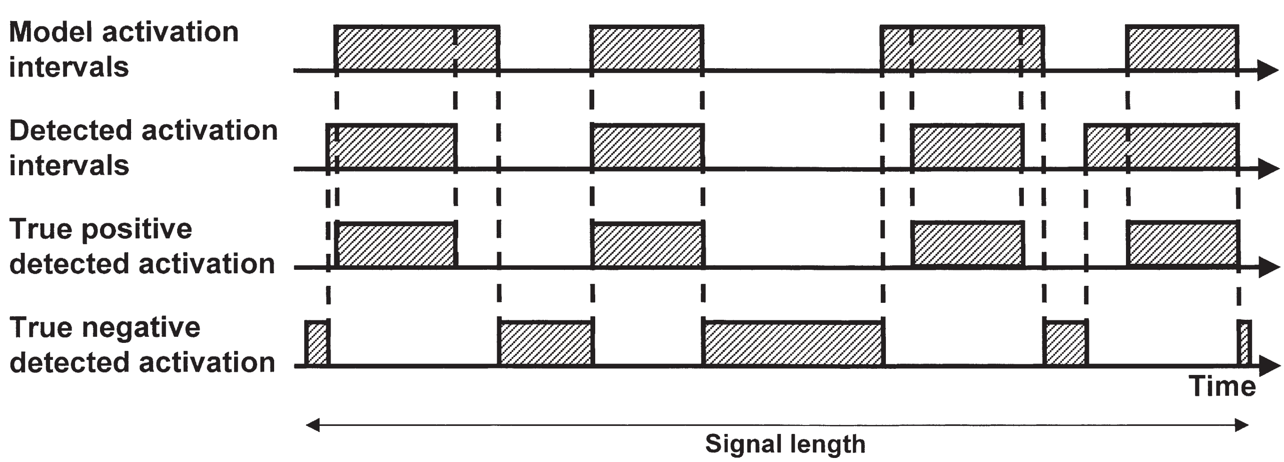

Fig. 4. Definition of τ_p (percentage of true positive detected activation) and τ_n (percentage of true negative detected activation) from the comparison between the model activation intervals and the detected ones.

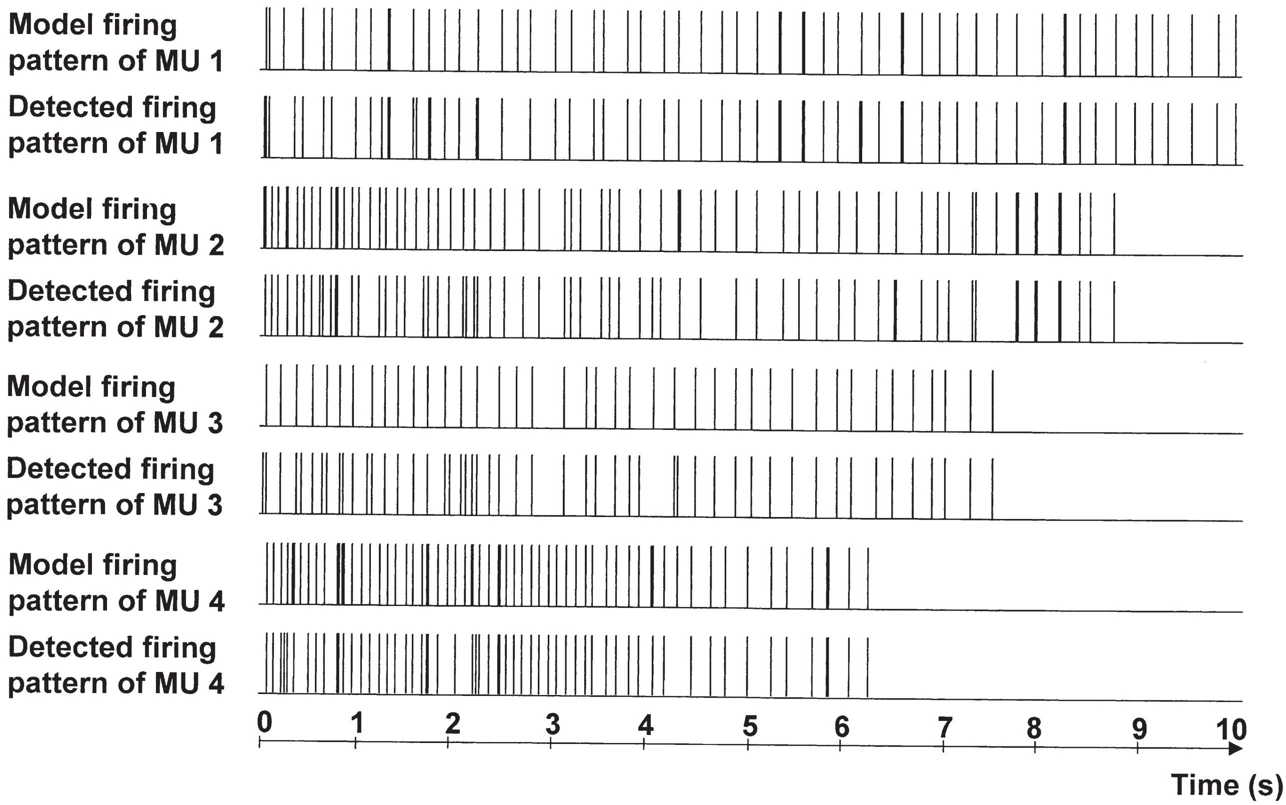

Fig. 5. Comparison of the firing patterns generated by the model and those detected by algorithm B for 4 MUs of one of the signals of the test set.

ithm B gave a worse performance in the estimation of the number of MUs, resulting in a non-zero difference between the number of model and detected classes for five out of 18 signals, with a maximum difference of three.

Table 1 shows the comparison between sensitivity (mean of the values obtained for each not missed model class) and positive predictivity of the detection phase for the two algorithms for all the test signals. Table 2 reports the results for the classification phase (mean of the values obtained for each not missed model class).

Algorithm A shows sensitivity and positive predictivity for the detection phase higher than 97% in all cases and sensitivity, specificity and accuracy of the classification phase higher than 99% (except for signal 17 for which the algorithm shows a sensitivity of about 96%). This algorithm shows good robustness to irregular statistics, superpositions and shape variability along the train. Resolution of superpositions is performed by algorithm A with no significant decrease in the performance (signals 7–10 compared to 1–6); moreover, the algorithm shows a low sensitivity to additive noise at very low SNR (signals 10 and 16). Algorithm B offers sensitivity higher than about 84% and predictivity higher than about 98% for the detection phase. Sensitivity, specificity and accuracy for the classification phase are higher than 94% for all signals except for signals 15 and

Table 1
Mean sensitivity and predictivity for the segmentation phase obtained with the two algorithms for the 18 signals

Signal no.	$^{(s)}$Se		$^{(s)}P$	
	Alg. A	Alg. B	Alg. A	Alg. B
1	100.0	99.4	100.0	99.7
2	100.0	98.9	100.0	97.8
3	100.0	99.9	100.0	100.0
4	100.0	99.5	100.0	100.0
5	100.0	91.0	99.9	96.3
6	99.9	89.7	99.9	93.0
7	100.0	98.9	100.0	100.0
8	99.9	90.3	99.9	99.6
9	100.0	98.9	100.0	100.0
10	99.8	91.3	99.1	97.7
11	100.0	95.9	100.0	94.5
12	99.5	96.1	100.0	95.4
13	99.9	97.5	100.0	99.2
14	99.3	89.7	99.9	84.6
15	99.6	90.7	98.5	92.6
16	98.0	94.5	98.9	96.5
17	97.6	90.0	97.1	79.4
18	97.2	95.0	99.9	95.3

Table 2
Mean sensitivity, specificity and accuracy for the classification phase obtained with the two algorithms for the 18 signals

Signal no.	${}^{(c)}$Se		${}^{(c)}$Sp		${}^{(c)}$Ac	
	Alg. A	Alg. B	Alg. A	Alg. B	Alg. A	Alg. B
1	100.0	99.7	100.0	99.9	100.0	99.9
2	100.0	100.0	100.0	100.0	100.0	100.0
3	100.0	99.9	100.0	99.9	100.0	99.9
4	99.9	99.8	99.9	99.9	99.9	99.9
5	99.6	99.4	99.9	99.9	99.9	99.6
6	100.0	99.7	100.0	99.9	100.0	99.8
7	100.0	99.8	100.0	99.9	100.0	99.9
8	100.0	97.5	100.0	99.4	100.0	99.1
9	100.0	99.6	100.0	99.9	100.0	99.9
10	99.9	98.1	99.9	99.6	99.9	99.4
11	99.6	97.9	99.9	99.6	99.9	99.2
12	99.8	96.3	99.9	99.6	99.9	99.2
13	99.9	98.4	99.9	99.7	99.9	99.5
14	99.5	94.2	99.9	99.3	99.9	98.7
15	98.0	88.0	99.7	98.3	99.5	97.0
16	99.2	94.7	99.9	99.2	99.8	98.6
17	96.6	84.9	99.5	98.1	99.2	96.7
18	100.0	98.2	100.0	99.6	100.0	99.4

17 (for which the sensitivity is 88% and 84.9%, respectively).

Table 3 reports the same performance indexes specifically for superimposed MUAPs. The results are of course reported only for the signals which contain superimposed MUAPs. Algorithm A shows sensitivity for segmentation and sensitivity, specificity and accuracy for classification higher than 93% in all the cases, showing good performance in resolving superpositions in the case of low SNR, intra-class variability and irregular firing patterns. Algorithm B shows worse performance, especially for signals 12–18, indicating that the resolution of superpositions is affected by the presence of re-derecruitment of MUs. Table 4 shows similar results computed specifically for irregular firings. Again algorithm A performs better than algorithm B. Anyway both algorithms seem to be almost unaffected by the presence of irregular firings.

Estimation of the mean firing rate for the signals with constant mean firing frequencies (signals 1–10) is performed by algorithm A with a maximum averaged (over all of the not missed model classes) error $\overline{\Delta f}$=0.2% and by algorithm B with $\overline{\Delta f}$=8.8%. Both algorithms show a very low relative error for all the 10 test signals, meaning that

Table 3
Mean sensitivity for the segmentation phase and mean sensitivity, specificity and accuracy for the classification phase, computed only for the superimposed MUAPs, obtained with the two algorithms for the signals with superimposed MUAPs

Signal no.	${}^{(s)}$Se		${}^{(c)}$Se		${}^{(c)}$Sp		${}^{(c)}$Ac	
	Alg. A	Alg. B	Alg. A	Alg. B	Alg. A	Alg. B	Alg. A	Alg. B
7	100.0	94.4	100.0	98.7	100.0	99.9	100.0	99.8
8	100.0	94.4	100.0	93.3	100.0	98.5	100.0	97.6
9	100.0	95.2	100.0	98.4	100.0	99.8	100.0	99.7
10	96.7	92.9	100.0	94.9	100.0	98.9	100.0	98.3
12	97.5	86.6	98.9	91.7	99.8	98.6	99.7	97.6
13	100.0	85.5	100.0	87.7	100.0	98.4	100.0	97.1
14	98.4	79.8	98.8	86.3	99.9	98.4	99.8	97.2
15	96.8	82.3	95.5	83.3	99.4	97.6	98.9	95.7
16	93.9	86.7	98.7	86.3	99.8	98.0	99.6	96.5
17	94.0	82.9	93.1	75.6	99.0	97.1	98.2	94.8
18	93.9	87.8	100.0	95.0	100.0	98.8	100.0	98.1

Table 4
Mean sensitivity for the segmentation phase and mean sensitivity, specificity and accuracy for the classification phase, computed only for double discharge and random firings, obtained with the two algorithms for the signals with irregular firings

Signal no.	$^{(s)}$Se		$^{(c)}$Se		$^{(c)}$Sp		$^{(c)}$Ac	
	Alg. A	Alg. B	Alg. A	Alg. B	Alg. A	Alg. B	Alg. A	Alg. B
3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
4	100.0	99.4	100.0	100.0	100.0	100.0	100.0	100.0
14	98.6	92.6	100.0	94.9	100.0	99.0	100.0	98.3
15	97.9	77.1	97.9	89.3	99.4	97.8	99.0	96.1
16	100.0	96.6	97.3	95.4	99.6	99.3	99.3	98.8
17	98.1	91.1	98.6	81.3	99.7	97.8	99.5	96.2
18	100.0	97.6	100.0	99.1	100.0	99.9	100.0	99.7

if only mean firing rate is of interest the two algorithms are equivalent and the faster one (algorithm B) should be preferred.

Finally we present the results obtained in the estimation of re-decruitment times. Fig. 6 shows the comparison of the two algorithms for activation interval estimation. The indexes τ_p and τ_n averaged over all of the not missed model classes are shown in the case of signals with re-derecruitment of MUs (signals 11–18).

τ_p is on average smaller than τ_n for both algorithms. Algorithm A provides a good estimation of the recruitment times with τ_p and τ_n higher than 97% in all the cases, and demonstrates a very accurate estimation of the time intervals of activity. Algorithm B gives worse

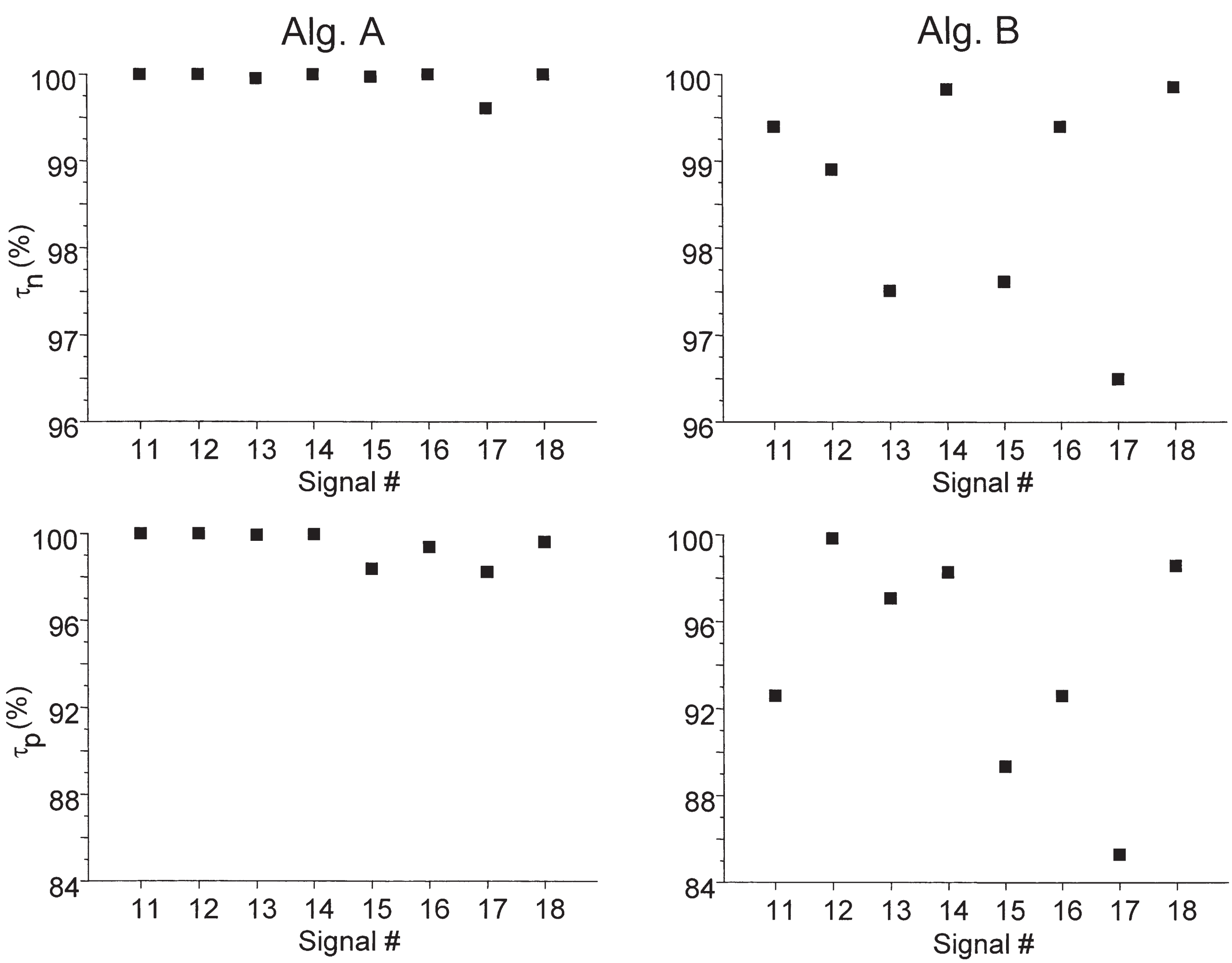

Fig. 6. τ_p and τ_n (mean for all the not missed model classes) for the eight signals with re-derecruitment of MUs, for algorithms A and B.

performance resulting in τ_n higher than 96% and τ_p higher than 84% in all cases, which is, however, still a good estimation of the re-decruitment times.

Both algorithms have shown good performances either for firing pattern estimation or for global indication about MU activity. Algorithm A works better almost in any situation but it has to be noted that the difference can be negligible for particular applications. The time required for the decomposition and the degree of interaction between the algorithm and the operator (manual parts of the decomposition process) should also be taken into account. On the basis of the field of application and on an evaluation technique like the one proposed, the user can decide which of the available algorithms is the most suitable for his/her needs. Fast and relatively accurate algorithms can be used for example in clinical applications, while very accurate algorithms are needed in particular research fields (in this case the speed can be of minor importance).

5. Conclusions

At present, the most common method for evaluation of EMG signal decomposition programs is the comparison with a reference manual decomposition of real signals. However, it has been shown that manual decomposition is subjective to some extent, even if performed by expert operators [18]. Furthermore, with real signals it is not possible to deal with signals of preset complexity. Moreover, the lack of a standard way to report the results and to define the quality of the decomposition made difficult in the past to compare performance and to show different capabilities of different algorithms in treating specific signal features.

In this paper an objective evaluation method for EMG signal decomposition algorithms has been presented. A generation model, already described [8], has been used to generate synthetic signals assumed as the reference and a number of indexes of performance have been defined. Some of these indexes are derived from traditional detection theory adapted to the EMG signal decomposition problem and to the comparison with a reference result; others are defined on the basis of the characteristics of the proposed model. We have also proposed a library of 18 test signals with different general features created to evaluate the sensitivity of the algorithms to different signal characteristics. Two decomposition algorithms have been tested and compared with the aim of showing a representative application of the method.

The tool appeared appropriate for the evaluation of the two algorithms as indicated by the questionnaire completed by the operators and by the results obtained. In particular, both operators indicated that the complexity of the test signals was such that most of the signals could be considered as real ones. On the other hand, the results obtained with the test signals show that the complexity of the signals was appropriate to be able to characterize the algorithms with respect to sensitivity to different signal features.

The model and the programs to compute the indexes of performance constitute a useful tool for developers of decomposition algorithms to properly report their results. Furthermore, it can be used by clinical EMG signal researchers who want to increase their knowledge of and experience with the EMG signal decomposition problem.

Acknowledgements

The authors are grateful to Andrea Crosetti for his precious help. This work was undertaken as an activity within the project “PRevention of muscular disorders in Operation of Computer Input Devices (PROCID)”, a concerted action financed under the European Union research programme BIOMED-2 (BMH-98-3903). The authors are grateful to the research groups among the PROCID partners who participated in the evaluation process by decomposing the synthetic signals. This work was also supported in part by Camera di Commercio di Torino, Compagnia di San Paolo, Fondazione CRT di Torino, Fondazione S. Maugeri, Regione Piemonte and Italian Ministry of Health.

References

[1] AAEE glossary of terms in clinical electromyography. American Association of Electromyography and Electrodiagnosis, Muscle and Nerve, 1997.
[2] Barkhaus P, Nandedkar S, Sanders DB. Quantitative EMG in inflammatory myopathy. Muscle Nerve 1990;13:247–53.
[3] Basmajian JV, De Luca CJ. Muscle alive. Williams & Wilkins, 1979.
[4] Colombo R, Crosetti A, Farina D, Merletti R. Evaluation of needle EMG decomposition algorithms with synthetic test signals. In: Proc. Symposium on Muscular Disorders in Computer Users; Copenhagen, 1999:73–7.
[5] De Luca CJ. Reflections on EMG signal decomposition. In: Desmedt JE, editor. Computer-aided electromyography and expert system. Elsevier Pub, 1989:33–7.
[6] Doherty TJ. Age related changes in the numbers and physiological properties of human motor units. PhD thesis, 1993.
[7] Egan JP. Signal detection theory and ROC analysis. Academic Press, 1975.
[8] Farina D, Crosetti A, Merletti R. A model for the generation of synthetic intra-muscular EMG signals to test decomposition algorithms. IEEE Trans Biomed Eng, in press, Jan. 2001.
[9] Forsman M, Kadefors R, Zhang Q, Birch L, Palmerud G. Motor unit recruitment in the trapezius muscle during arm movements and in VDU precision work. Int J Ind Ergonomics 1999;24(6):619–30.
[10] Gerber A, Studer RM, Figueiredo JP, Moschytz GS. A new framework and computer program for quantitative EMG signal analysis. IEEE Trans Biomed Eng 1984;31:857–63.

[11] Gut R, Moschytz GS. High-precision EMG signal decomposition using communication techniques. IEEE Trans Signal Processing 2000;48:2487–99.

[12] Haas WF, Meyer M. An automatic EMG decomposition system for routine clinical examinations and clinical research. ARTMUP — automatic recognition and tracking of motor unit potentials. In: Desmedt JE, editor. Computer-aided electromyography and expert system., 1989:67–81.

[13] Kadefors R, Forsman M, Zoega B, Herberts P. Recruitment of low threshold motor units in the trapezius muscle in different static arm positions. Ergonomics 1999;42(2):359–75.

[14] LeFever RS, DeLuca CJ. A procedure for decomposing the myoelectric signal into its constituent action potentials — part 1: technique, theory, and implementation. IEEE Trans Biomed Eng 1982;29:149–57.

[15] LeFever RS, Xenakis AP, DeLuca CJ. A procedure for decomposing the myoelectric signal into its constituent action potentials — part 2: execution and test for accuracy. IEEE Trans Biomed Eng 1982;29:158–65.

[16] McGill K, Cummins KL, Dorfman LJ. Automatic decomposition of the clinical electromyogram. IEEE Trans Biomed Eng 1985;32:470–7.

[17] Sonoo M, Stalberg E. The ability of MUP parameters to discriminate between normal and neurogenic MUPs in concentric needle EMG: analysis of the MUP "thickness" and the proposal of "size index". EEG Clin Neurophysiol 1993;89:291–303.

[18] Søgaard K, Jensen BR, Pilegaard M, Sjøgaard G. Analysis of consistency for motor unit identification during dynamic force. In: Proc. Symposium on Muscular Disorders in Computer Users, 1999:206–7.

[19] Stalberg E, Falck B, Sonoo M, Stalberg S, Astrom M. Multi MUP EMG analysis — a two years experience in daily clinical work. EEG Clin Neurophysiol 1995;97:145–54.

[20] Stashuk D, De Luca CJ. Update on the decomposition and analysis of EMG signals. In: Desmedt JE, editor. Computer-aided electromyography and expert system, 1989:38–53.

[21] Stashuk DW, Qu Y. Robust method for estimating motor unit firing pattern statistics. Med Biol Eng Comput 1998;35:1–8.

[22] Wellig P, Moschytz GS, Laubli T. Decomposition of EMG signals using time–frequency features. Proc EMBS'98 1996;20:1497–500.

Dario Farina graduated summa cum laude in Electronics Engineering from Politecnico di Torino, Torino, Italy, and is a registered professional engineer. During 1998 he was a fellow of the Laboratory for Neuromuscular System Engineering in Torino, Italy. Since 1999 he has been a Ph.D. candidate at the Department of Electronics of Politecnico di Torino within a cotutelle program with the École Centrale de Nantes, Nantes, France, and is involved in teaching activities in electronics and mathematics at Politecnico di Torino. He is author/co-author of one book, more than ten papers on international journals and about thirty conference abstracts and papers on national journals. His main interests are in the areas of signal processing applied to biomedical signals and modeling of biological systems.

Roberto Colombo received a degree in Electronics Engineering from Politecnico di Milano, Milano, Ital, in 1980. Since 1981 he has worked at Fondazione S. Maugeri, Clinica del Lavore e della Riabilitazione IRCCS-Veruno as a biomedical engineer. Dr. Colombo's research concerns mobility impairments and rehabilitation, spasticity and muscle force measurement, cardio-respiratory rhythms and mechanics and speech production mechanisms.

Roberto Merletti graduated in Electronics Engineering from Politecnico di Torino, Torino, Italy, and received MS and Ph.D. degrees from the Ohio State University, Columbus. He taught biomedical instrumentation and biomedical signal processing at the BME department of Boston University, Boston, MA, from 1989 to 1994. He is currently Associate Professor at the Department of Electronics of Politecnico di Torino and Director of the Laboratory for Neuromuscular System Engineering in Torino, Italy. His research focuses on surface electromyography, myoelectric signal processing, electrical stimulation, and neuromuscular control.

Henrik Baare Olsen received his MS degree in Electrical Engineering from the Technical University of Denmark in 1993. Since 1995 he has worked at the National Institute of Occupational Health, Denmark. He has developed hardware and software for intra-muscular data acquisition and decomposition techniques. Presently he is developing a decomposition program to handle long time motor unit identification. His research interests are focused on understanding mechanisms of the upper extremity motor control and the risk of developing musculoskeletal disorders during computer work.