Artificial Intelligence in Medicine (2006) **37**, 55—64

ELSEVIER

ARTIFICIAL INTELLIGENCE IN MEDICINE

http://www.intl.elsevierhealth.com/journals/aiim

A novel method for automated EMG decomposition and MUAP classification

C.D. Katsis [a,b], Y. Goletsis [b,c], A. Likas [b,d], D.I. Fotiadis [b,d,*], I. Sarmas [e]

[a] *Department of Medical Physics, Medical School, University of Ioannina, GR 451 10 Ioannina, Greece*
[b] *Unit of Medical Technology and Intelligent Information Systems, Department of Computer Science, University of Ioannina, P.O. Box 1186, GR 451 10 Ioannina, Greece*
[c] *Department of Economics, University of Ioannina, GR 451 10 Ioannina, Greece*
[d] *Biomedical Research Institute—FORTH, GR 451 10 Ioannina, Greece*
[e] *Department of Neurosurgery, Medical School, University of Ioannina, GR 451 10, Ioannina, Greece*

Received 9 May 2005; received in revised form 29 August 2005; accepted 13 September 2005

KEYWORDS
Quantitative electromyography;
Electromyogram decomposition;
Motor unit action potential detection and classification;
Support vector machine

Summary

Objective: This paper proposes a novel method for the extraction and classification of individual motor unit action potentials (MUAPs) from intramuscular electromyographic signals.
Methodology: The proposed method automatically detects the number of template MUAP clusters and classifies them into normal, neuropathic or myopathic. It consists of three steps: (i) preprocessing of electromyogram (EMG) recordings, (ii) MUAP detection and clustering and (iii) MUAP classification.
Results: The approach has been validated using a dataset of EMG recordings and an annotated collection of MUAPs. The correct identification rate for MUAP clustering is 93, 95 and 92% for normal, myopathic and neuropathic, respectively. Ninety-one percent of the superimposed MUAPs were correctly identified. The obtained accuracy for MUAP classification is about 86%.
Conclusion: The proposed method, apart from efficient EMG decomposition addresses automatic MUAP classification to neuropathic, myopathic or normal classes directly from raw EMG signals.

1. Introduction

Clinical electromyography analyses the electromyogram (EMG) recorded from a contracting muscle using a needle electrode to diagnose neuromuscular disorders. EMG is composed of discrete waveforms

* Corresponding author. Tel.: +30 26510 98803; fax: +30 26510 97092.
E-mail address: fotiadis@cs.uoi.gr (D.I. Fotiadis).

0933-3657/$ — see front matter
doi:10.1016/j.artmed.2005.09.002

called motor unit action potentials (MUAPs), which result from the repetitive discharges of groups of muscle fibers called motor units (MUs). The term MU refers collectively to one motoneuron and the group of muscle fibers it innervates and is the smallest unit of skeletal muscle that can be activated by volitional effort. MUAPs from different MUs tend to have distinct shapes, which remain almost the same for each discharge. The MUAPs can therefore be identified and tracked using pattern recognition techniques. The resulting information can be used to determine the origin of the weakness, i.e. neurogenic or myopathic diseases [1–3].

The changes brought about by a particular disease alter the properties of the muscle and nerve cells, causing characteristic changes in the MUAPs. Distinct MUAPs can be seen only during weak contractions when few motor units are active. When a patient maintains low level of muscle contraction, individual MUAPs can be easily recognised. As contraction intensity increases, more motor units are recruited. Different MUAPs will overlap, causing an interference pattern in which the neurophysiologist cannot detect individual MUAP shapes reliably. Usually, in clinical electromyography, neurophysiologists assess MUAPs from their shape using an oscilloscope and listening to their audio characteristics. Thus, an experienced electrophysiologist can detect abnormalities with reasonable accuracy. However, subjective MUAP assessment, although satisfactory for the detection of unequivocal abnormalities, may not be sufficient to delineate less obvious deviations or mixed patterns of abnormalities [4]. Therefore, for an effective automated MUAP assessment, a systematic handling of EMG signal must decompose the signal into MUAPs and classify each MUAP into different classes.

Although, a number of computer-based quantitative EMG analysis algorithms have been developed [5–7] practically none of them has gained wide acceptance for extensive clinical use. Most importantly, there are no uniform international criteria neither for pattern recognition of similar MUAPs nor for MUAP feature extraction [8]. Out of the two assessment tasks (i.e. MUAP detection and classification) according to our knowledge only the first one has attracted attention. Buchthal et al. [9,10] developed one of the earliest methods for quantitative EMG decomposition, where MUAPs were recorded photographically and then were selected for analysis. LeFever and DeLuca [11] used a special three channel recording electrode and a visual computer decomposition scheme based on template matching and firing statistics for MUAP identification. Stalberg et al. [8], in their original system used waveform template matching whereas more recently [12] they have used different shape parameters as input to a template matching technique. Adreassen [13] followed the manual method developed by Buchthal using template matching with four templates for the recognition of MUAP's recorded at threshold contraction. Stashuk and Qu [14] proposed a method to identify MUAPs based on power spectrum matching. Hassoun et al. [15] proposed a system called neural network extraction of repetitive vectors for electromyography (NNERVE) which uses the time domain waveform as input to a three layer artificial neural network with a ''pseudo-unsupervised'' learning algorithm for classification. McGill et al. [16] used a method based on a combination of shape recognition of the MUAPs and statistical probability of occurrence. Fang et al. [17] developed a comprehensive technique to identify single motor unit (SMU) potentials based on one-channel EMG recordings measuring waveform similarity of SMU potentials in the wavelet domain. Wu et al. [18] decomposed MUAPs of needle electrode EMG signal by means of self-organization competing neural network. Chauvet et al. [19] proposed a method that allows decomposition of EMG signals based on fuzzy logic techniques. Zennaro et al. [20] designed a decomposition software for multichannel long term EMG recordings using a wavelet based hierarchical cluster analysis algorithm, which is suitable for the study of MU discharge patterns in healthy subjects.

All the aforementioned techniques deal only with MUAP detection and EMG decomposition into its constituent MUAPs. However, they do not classify MUAPs according to their pathology. To contribute to the quantification of the routine needle EMG examination, we have developed a methodology for automated MUAP detection, EMG decomposition and template MUAP classification into normal (nor), myopathic (myo) and neuropathic (neu). The proposed methodology consists of three steps. In the first step, the EMG signal is preprocessed using an algorithm that automatically detects areas of low activity and candidate MUAPs. In the second step, MUAPs are clustered and the number and shape of MUAP clusters are determined. Furthermore, superimposed MUAPs are automatically identified and decomposed into their constituents. Finally, in the third step, an unknown MUAP is classified as nor, myo or neu. In the following sections, we first describe the steps of the proposed methodology. Then, we evaluate it using a dataset of EMG recordings and an annotated collection of MUAPs. Finally, comments on the application and the results of our approach are given in Section 4.

2. Materials and methods

The proposed methodology consists of three steps (Fig. 1): in the first, preprocessing is applied to remove noise and detect areas of low activity in EMG recordings. MUAP detection, clustering and decomposition of superimposed MUAPs to their constituents is implemented in the second step. In the third step, the MUAPs are classified according to their pathology in three classes.

2.1. Dataset

Our dataset contains EMG signals from two sources: the first one has been produced by the Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus and consists of 26 EMG signals obtained from 26 subjects. The second, was obtained by the University Hospital of Ioannina, Division of Neurological Clinic, Ioannina, Greece and consists of 24 EMG recordings, from 24 patients. All EMG signals were acquired from biceps brachii following the same protocol. Ten of our subjects had no history or physical evidence of neuromuscular disease, 20 subjects suffer from myopathy and 20 subjects suffer from motor neuron disease. The annotation in each group was based on the patient history and muscle biopsy. Only subjects with no history or signs of neuromuscular disorders were considered as normal. EMG recordings were acquired from the biceps brachii, at up to 30% of the maximum voluntary contraction (MVC) level under constant isometric conditions. Each subject was asked to produce an elbow flexion at the aforementioned MVC level and to sustain it for 2 s. An UWE HS-30K digital dynamometer was used to verify the contraction level. When the desired level of contraction was accomplished, data acquisition was initiated and the subject attempted to maintain a constant level of contraction until the end of data acquisition.

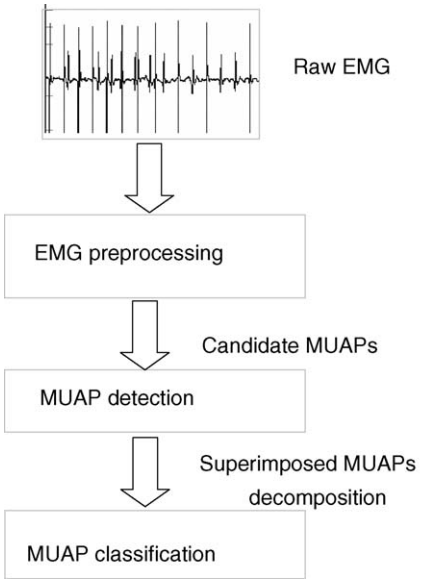

Figure 1 The proposed three step methodology.

The muscle activity was recorded by a concentric EMG needle electrode (25-mm length, 0.33-mm diameter, TECA) inserted into the muscle. A ground electrode (TECA 32-mm diameter stainless steel disc) with commercial electrode paste was taped onto the skin near the needle insertion site.

2.2. EMG preprocessing

Initially, signal preprocessing and candidate MUAP detection takes place (Fig. 2). Since EMG is contaminated by noise (due to non-targeted muscles recorded activity and electrode movement), a bandpass filter (3 Hz—8 kHz) is applied. A National Instrument's DAC is used to digitize EMG using sampling rate 20 kHz and 12-bit resolution. In order to detect the MUAPs comprising the EMG, the signal is segmented to generate possible MUAP waveforms. Areas of low activity are eliminated using a threshold T which depends on the $\max\{x_i\}$ and the mean absolute value $(1/L)\sum_{i=1}^{L}|x_i|$ of the EMG signal, where x_i are the discrete values of the EMG signal and L the number of samples. The threshold T is calculated as:

$$\text{IF } \max\{x_i\} > \frac{30}{L}\sum_{i=1}^{L}|x_i|, \quad \text{THEN } T = \frac{5}{L}\sum_{i=1}^{L}|x_i| \qquad \text{ELSE } T = \max\frac{\{x_i\}}{5}. \tag{1}$$

This threshold is used to identify peaks in the signal [21]. Peaks over the calculated threshold T are considered as candidate MUAPs. The computed threshold T for the segmentation of EMG signal is introduced to accommodate the wide range in amplitude variations in the recorded signal. A window with a constant length of 121 sampling points (i.e. ~6 ms at 20 kHz, in order to fit the main part of each MUAP) is centred at the first identified peak. If a larger peak is found in the window, the window is moved and centred at this peak; otherwise the initial 121 signal points are considered as a candidate MUAP waveform and are stored. The process continues until the end of the EMG signal is reached.

2.3. MUAP clustering and detection of superimposed MUAPs

The second step consists of two stages (Fig. 3):

(a) MUAP clustering;

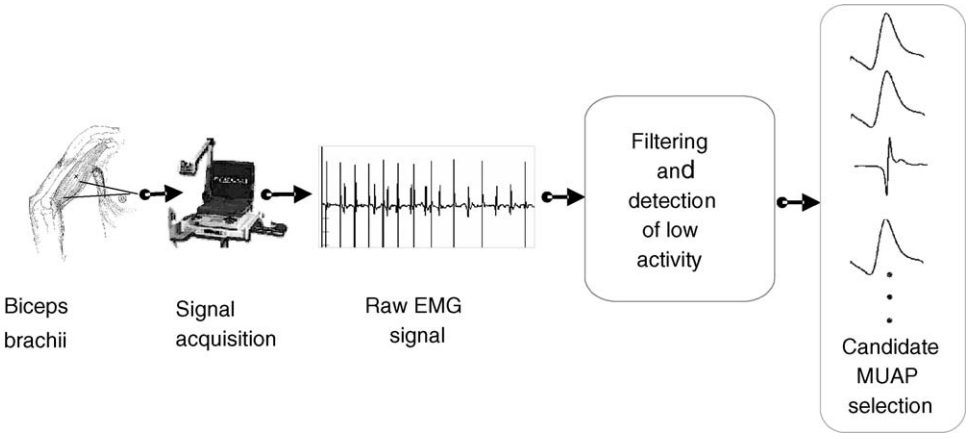

Figure 2 EMG preprocessing.

(b) detection and decomposition of superimposed MUAPs.

2.3.1. MUAP clustering

In this stage, MUAP clusters are automatically detected and for each cluster the average or *template* shape is determined. The procedure for the detection of the number of clusters in EMG data is based on the minimization of a regularized cost function J [22] with respect to the distance of the candidate MUAPs from the cluster centres (first term in Eq. (2)) and with respect to the distance of the cluster centres from each other (second term in Eq. (2)):

$$J = \sum_{\mu=1}^{p}\sum_{\nu=1}^{k} I(\mathbf{y}^{(\nu)}|\mathbf{x}^{(\mu)})||\mathbf{x}^{(\mu)} - \mathbf{y}^{(\nu)}||^2 + \sum_{\mu=1}^{p}\sum_{\nu=1}^{k} \tilde{\lambda}_{\nu}\tilde{I}(\mathbf{y}^{(\nu)}|\mathbf{x}^{(\mu)})||\mathbf{y}^{(\nu)} - \mathbf{y}^{(w)}||^2, \quad (2)$$

where $I(\mathbf{y}^{(\nu)}|\mathbf{x}^{(\mu)})$ is an indicator function which equals to 1, if $\nu = \text{argmin}_l||\mathbf{x}^{(\mu)} - \mathbf{y}^{(l)}||^2$ and to 0 otherwise; $\tilde{I}(\mathbf{y}^{(\nu)}|\mathbf{x}^{(\mu)})$ is an indicator function equal to 1 if $\mathbf{y}^{(\nu)} \in N_{\mathbf{y}^{(w)}}$ and 0 otherwise; $w = \text{argmin}_l ||\mathbf{x}^{(\mu)} - \mathbf{y}^{(l)}||^2$; $N_{\mathbf{y}^{(w)}}$ is the neighbourhood of the cluster centre $\mathbf{y}^{(w)}$; $\mathbf{x}^{(\mu)}$ is the μth feature vector, $\mathbf{x}^{(\mu)} \in {}^{n}$; p is the number of patterns, $\{\mathbf{x}^{(\mu)}: \mu = 1, 2, \ldots, p\}$ and k is the number of cluster centres. As already mentioned, the cost function J consists of two parts. The first part is related to the distribution of the cluster centres in order to minimize the sum of squared distance from each input pattern to the nearest cluster centre. The regularization (second) term of the cost function further requires that the sum of squared distances from a cluster to its nearby clusters is minimum. Minimization of the cost function results in the minimization of the sum of squared distances of the data points from the respective nearest cluster centre as well as the sum of the squared distances of the individual cluster centres from the neighbourhood cluster centres. Small values of the neighbourhood encourage the formation of more distinct cluster centres, while large values of the neighbourhood encourage the formation of fewer distinct cluster centres. The

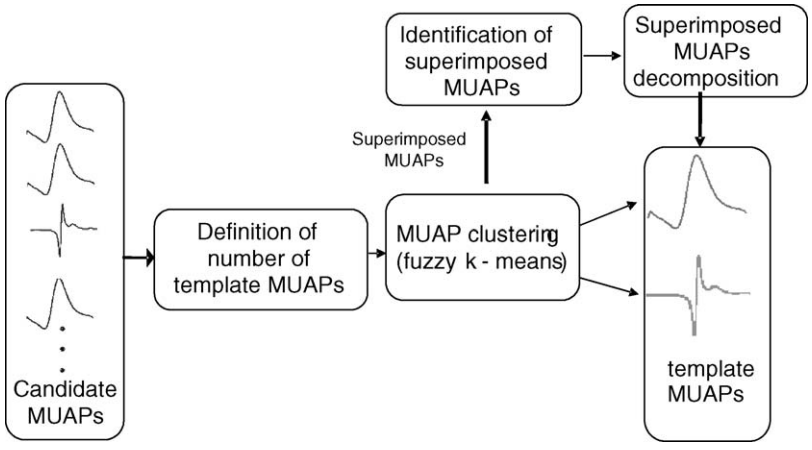

Figure 3 MUAP clustering, detection and decomposition of superimposed MUAPs.

neighbourhood is identified as a scale parameter and the number of clusters is obtained at varying values of the scale parameter. The number of cluster centres is then obtained based on persistence over the largest range of the scale parameter.

In order to compute the number of MUAP clusters, we assume initially a large number of clusters (e.g. 16 clusters). This is considered to be satisfactory since the maximum number of template MUAP clusters that can be identified with needle EMG at low to moderate force levels is at most 12–14 [23–25]. At the end of each clustering epoch, cluster centres in the same neighbourhood are combined, while clusters with small number of MUAPs (<3) are removed. The number of clusters is obtained at a given "neighbourhood scale" and the corresponding plot is generated as it is shown in Fig. 4.

Once the number of clusters is detected and in order to obtain the template MUAP's shape, the fuzzy *k*-means algorithm is used. The algorithm starts with a selection of an initial set of prototypes, which implies the partition of the feature vectors into *k* clusters. Each cluster is represented by a prototype (the template MUAP), which is computed as the centre of the feature vectors belonging to that cluster. Each of the feature vectors is assigned to the cluster whose prototype is its closest neighbour. The new prototypes are computed from the results of a new partition and this process is repeated until the prototypes' displacements from one iteration to the next become negligible ($<5 \times 10^{-4}$). More specifically, the fuzzy *k*-means

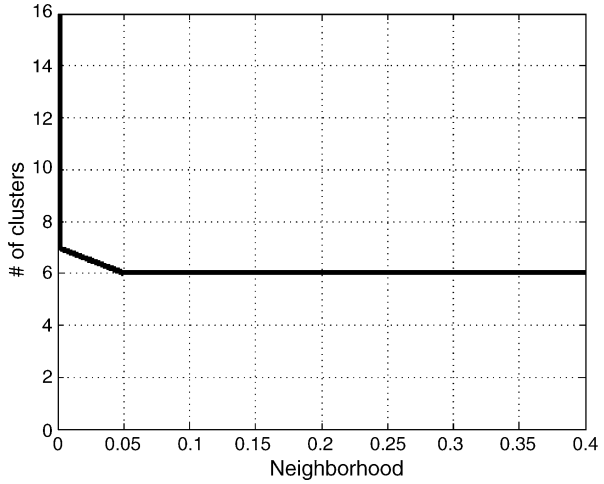

Figure 4 Starting with a large number of classes, at the end of each clustering epoch, cluster centers in the same neighbourhood are combined, while clusters with small number of MUAPs are removed. The number of clusters is obtained at a given "neighbourhood scale" and the corresponding plot is generated. The number of clusters *k*, is selected as the number of clusters, which persists over the largest range of neighbourhood (here $k = 6$).

algorithm is based on the minimization of the following objective function (with respect to a fuzzy *k*-partition (U) and a set of *k* prototypes ($y^{(\nu)}$)):

$$J_q = (U, y^{(\nu)}) = \sum_{\mu=1}^{r}\sum_{\nu=1}^{k}(u_{\nu\mu})^q ||x^{(\mu)} - y^{(\nu)}||^2, \quad k \le r, \tag{3}$$

where, $x^{(\mu)}$ is the μth feature vector, $y^{(\nu)}$ the centre of the νth cluster ($x^{(\mu)}$, $y^{(\nu)} \in {}^{121}$), $u_{\nu\mu}$ the degree of membership of $x^{(\mu)}$ in the νth cluster, $||x^{(\mu)} - y^{(\nu)}||^2$ the distance between $x^{(\mu)}$ and $y^{(\nu)}$, r the number of data points and k is the number of clusters. The parameter q is the weighting exponent for $u_{\nu\mu}$ which controls the "fuzziness" of the resulting clusters[1] (in our method $q = 1.5$).

2.3.2. Detection and decomposition of superimposed MUAPs

The EMG signal recorded even at low to moderate contractions contains superimposed potentials produced from different MUAPs overlap. Candidate MUAPs with degree of membership $u_{\nu\mu} < 0.8$ are considered as superimposed. This threshold value was chosen heuristically after extensive testing in collaboration with a medical expert. The following decomposition approach is used for the superimposed MUAPs: first, the crosscorrelation between the superimposed waveform and the template MUAP that has the largest degree of membership is computed. This MUAP is time shifted as many sampling points are impaired by the crosscorrelation and subtracted from the superimposed waveform. In the same way, a crosscorrelation is carried out again between the residual waveform, and the template MUAP that has the second degree of membership. If the maximum waveform value of the residual signal is larger than the detected threshold T (computed in the EMG preprocessing step), it is assumed that the superimposed waveform contains another template MUAP and a new crosscorrelation is computed between the residual waveform and the next template MUAP in terms of degree of membership. Otherwise the process ends. The decomposition procedure is depicted in Fig. 5.

2.4. MUAP classification

For all MUAP waveforms, the 6 ms long MUAP segments are expanded to 25 ms on the original EMG

[1] The parameter q determines the degree of fuzziness of the final solution, which is the degree of overlap between classes; q ranges from 1 to $+\infty$. When $q = 1$ the solution is a hard partition. As q approaches infinity the solution approaches the highest degree of fuzziness.

signal where the position of the identified MUAP peak was marked during segmentation. This is due to the fact that MUAP duration is in most cases longer than 6 ms (in most of the cases the MUAP duration does not exceed 18 ms) [26]. Due to superpositions in the expanded window, in order to eliminate discrepancies from the class average, the standard deviation (S.D.) for each sampling point of all MUAPs in a cluster is calculated. Values of points beyond ± 1.5 S.D. from the average are excluded from the computation of the cluster average [21].

In order to classify the template MUAPs into normal, myopathic and neuropathic, a support vector machine (SVM) classifier is employed [27–29]. A classification task based on SVM usually involves training and testing data, which consist of a number of data instances. Each instance in the training set contains one "target value" (class labels) and several "attributes". Although initially developed for binary classification problems, SVMs can be adapted to deal with multi-class problems using the one-against-one method [30]. This method constructs $k(k-1)/2$ classifiers (where k is the number of classes) where each one is trained using data from two classes. Although other methods for multi-class SVMs exist, the above-mentioned approach has been chosen due to the low training time required and its comparable performance [31]. The goal of the SVM is to produce a model, which predicts a target value of data instances in the testing set in which only the attributes are given. Let a training set of instance-label pairs be (x_i, y_i), $i = 1, \ldots, p$ where $x_i \in {}^n$ is the training vector of original data, belonging to one of three classes (nor, myo or neu), p is the number of the template MUAPs in the training set and $y_i \in \{-1, 1\}$ indicates the (one of the two) class of x_i. The support vector machine requires the solution of the following optimization problem:

$$\min_{w,b,\xi}\left(\frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + C\sum_{i=1}^{p}\xi_{\mathrm{i}}\right), \tag{4}$$

$$\text{subject to } y_i(\mathbf{w}^{\mathrm{T}}\phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0,$$

where b is the bias term, $\mathbf{w}$ is a vector perpendicular to the hyperplane $\langle \mathbf{w}, b\rangle$, ξ the factor of classification error and $C > 0$ is the penalty on parameter of the error term. The training vectors x_i are mapped into a higher dimensional space F by the function ϕ: $R^n \rightarrow F$, where F is a feature space where the data are separable. SVM finds a separating hyperplane with the maximal geometric margin and minimal empirical risk R_{emp} in this higher dimensional space. R_{emp} is defined as:

$$R_{\mathrm{emp}}(a) = \frac{1}{2p}\sum_{i=1}^{p}|y_i - f(x_i, a)|, \tag{5}$$

where f is the decision function defined as:

$$f(x) = \sum_{i=1}^{p} y_i a_i K(x_i, x) + b, \tag{6}$$

with $K(x_i, x_j) \equiv \phi(x_i)^{\mathrm{T}}\phi(x_j)$ being the kernel function, a_i the weighting factors and b is the bias term. In our case the kernel is a radial basis function (RBF), which is defined as:

$$K(x_i, x_j) = \exp(-\gamma||x_i - x_j||^2), \quad \gamma > 0, \tag{7}$$

where $\gamma = \frac{1}{2\sigma^2}$ (and σ is the standard deviation) is a parameter of the kernel. The RBF kernel non-line-

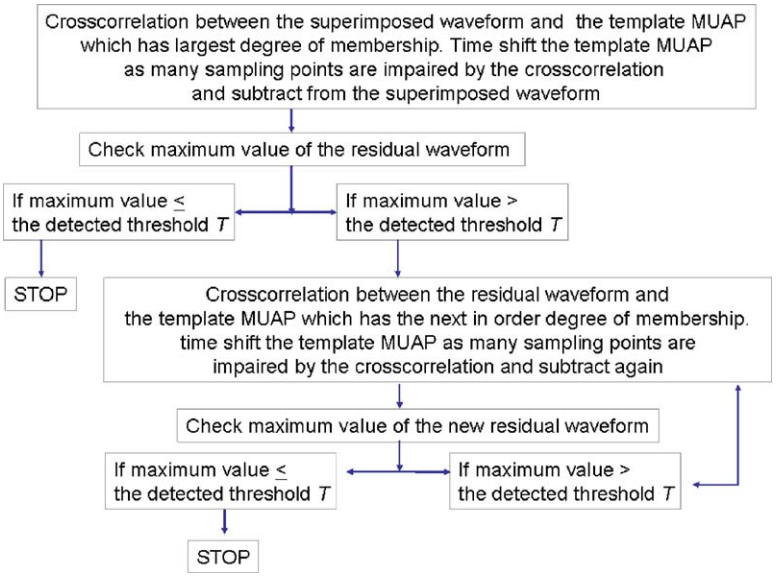

Figure 5 Schematic representation of the superimposed MUAP decomposition procedure to their constituents.

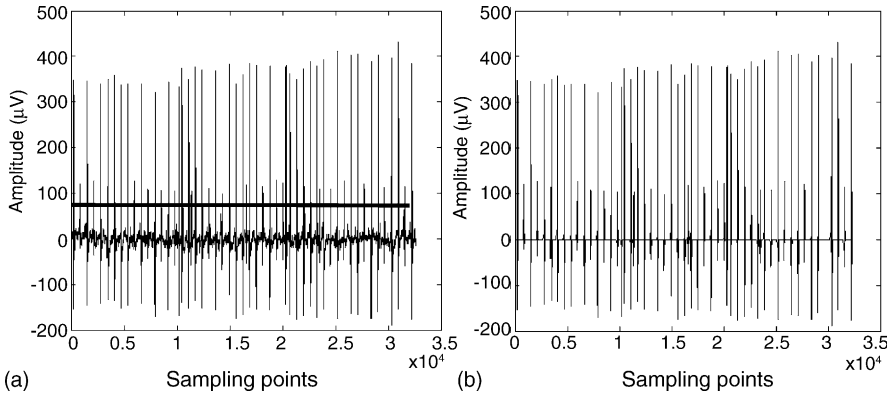

Figure 6 (a) Raw signal (the horizontal line corresponds to the computed detection threshold *T*, in this case $T = 86.4$) and (b) the signal after preprocessing.

arly maps samples into a higher dimensional space, so it can handle cases when the relation between class labels and attributes is non-linear. The parameters γ and C were defined heuristically (more specifically γ was set to $2^{-2.25}$ and C to $2^{6.25}$).

In our application, we have used the SVM training algorithm provided by the LIBSVM library [32,33], which has been proven computationally effective [34].

3. Results

The preprocessing of the 50 EMG signals resulted in 2969 candidate MUAPs. The computed threshold T ranges from 30 to 100 μV, which values are within the limits reported in other works [12,30]. The implementation of the preprocessing step for a specific EMG signal is shown in Fig. 6.

Following the preprocessing step, the MUAPs are decomposed (if superimposed) and clustered. To measure the performance of the proposed methodology the following measures were used[2]:

$$\text{correct identification rate (CIR)} = \left(1 - \frac{\sum |\text{mismatches}|}{\sum \text{clusters detected by neurophysiologist}}\right) \times 100\% \quad (8)$$

$$\text{success rate (SR)} = \left(\frac{\sum \text{clusters detected by method}}{\sum \text{clusters detected by neurophysiologist}}\right) \times 100\% \quad (9)$$

[2] Although, we consider CIR to provide more accurate information on method's performance we also calculate the SR, for comparison reasons.

The clustering step results in high CIRs: 93, 95 and 92% for nor, myo and neu classes, respectively. Table 1 presents the SRs for each of the three MUAP classes, while Table 2 provides a comparison of our approach with other methods reported. Our methodology identified correctly 91% of the superimposed MUAPs. In Fig. 7, a superimposed MUAP is decomposed into its constituents (template MUAPs).

For the evaluation of the last step of our methodology, an annotated database was created. An expert neurophysiologist was asked to characterize every template MUAP produced from the second step of the proposed methodology. The resulting MUAP dataset consists of 231 template MUAPs. Ninety-five MUAPs were characterized as normal, 55 as myopathic and 81 as neuropathic. The database was randomly divided in two datasets, a training and a testing one. More specifically out of the 231 MUAPs, 115 used for training and 116 for testing. In order to obtain an accurate estimation of the classifying performance, the experiment was repeated ten times, with different (random) data splits. The average classifying accuracy (number of correctly classified divided by the total number of classified MUAPs) was 86.14%. The classifying performance is measured in terms of sensitivity and specificity. Table 3 presents the obtained (average) results for the MUAP diagnosis (classification step) while Table 4 presents the respective confusion matrix.

Table 1 MUAP detection success rate (in parenthesis: the number of identified classes to the total number of classes)

MUAP classes	Success rate (%)
nor	94.74 (90/95)
myo	96.36 (53/55)
neu	95.06 (77/81)
Total	95.24 (220/231)

Table 2 Comparison of the results of our work with other existing works

References	MVC level (%)	Muscle	No. of subjects	Success rate (%) (EMG decomposition)
Christodoulou and Pattichis [21]	Up to 30	Biceps brachii	40	95–97
Mc Gill et al. [16]	20	Biceps brachii	–	30–70
Stashuk [23]	–	–	–	88.7
Stashuk and Paoli [24]	–	–	–	80.8
Loudon et al. [6]	Up to 20	–	–	95
This work	Up to 30	Biceps brachii	50	95.24

4. Discussion

A novel methodology for automated EMG decomposition and MUAP classification based on raw EMG signals has been developed. It performs decomposition and automated diagnosis of EMG signals recorded from biceps brachii muscle during up to 30% of MVC level under constant isometric conditions. The methodology was designed for conventional needle EMG examination by analysing signals acquired using a standard concentric needle electrode and a conventional electromyograph. The automated detection of the number of clusters in EMG data is based on the minimization of the regularized cost function, while the shape and detection of superimposed MUAPs are produced using the fuzzy *k*-means method. Finally, SVM was used for MUAP classification.

As far as the EMG decomposition stage is concerned (step 2 of stage 2) the results of the EMG clustering are comparable and in some cases higher than similar works reported in the literature as it is shown in Table 2. However, attention should be paid when comparing the results of the various EMG decomposition methods since they may involve different MVC levels [6,16].

Table 3 Classification performance

Pathology	Sensitivity (%)	Specificity (%)
nor	76.84	95.58
myo	92.73	88.64
neu	92.59	96.00

Classification accuracy: 86.14%.

Table 4 Confusion matrix

MUAPs	Classified as nor	Classified as myo	Classified as neu
nor[a]	73	18	4
myo[a]	2	51	2
neu[a]	4	2	75

[a] According to the medical expert.

The proposed methodology applies novel techniques for the MUAP clustering and the identification and decomposition of superimposed MUAPs. Moreover, according to our knowledge, the proposed methodology is the only one in the literature performing fully automated MUAP classification from raw EMG signals. In our approach, the number of template MUAP clusters is automatically calculated using no a priori knowledge. The main advantage of

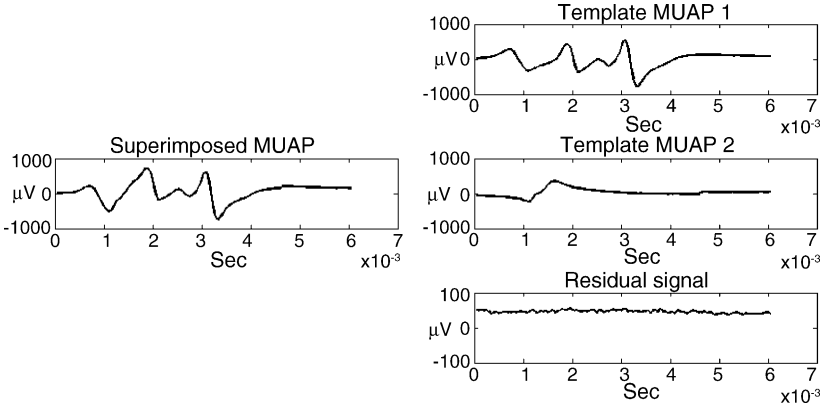

Figure 7 A superimposed MUAP decomposed into its constituents. Two template MUAPs are identified (both of them are correctly classified during step 3; template MUAP 1 being neu and template MUAP 2 nor). Template MUAP 2 was time shifted by −1.5 ms before subtraction. The decomposition process stopped after the identification of the second MUAP since the residual signal was below the detection threshold (82.3 μV).

our approach is the minimum use of tuned parameters. Data driven calculation of thresholds is another advantage since it enhances method's adaptability to different EMG signals. The proposed approach effectively deals with signals of MVC level up to 30%.

The use of SVM is advantageous since it improves the efficiency of the methodology. This can be confirmed comparing with the nearest neighbourhood classifier (*N*—*N*). SVM demonstrated a 9% increase in performance (the classification accuracy for the *N*—*N* was around 77%). Looking at a more technical level inside the SVM, it should be noted that the selection of the kernel *K* is of major importance for the performance of the classifier. In our case, an RBF kernel has been applied. Alternative approaches, such as linear or polynomial were not used due to the nature of our problem: the linear kernel cannot handle non-linear separable problems; the polynomial kernel has more hyperparameters than the RBF kernel, fact that influences the complexity of model selection.

EMG recordings of 2 s were used for validation. Longer EMG recordings cause a geometrical increase of computational time. It was also observed that often, due to waveform variability (especially in cases of motor neuron diseases) MUAP classes coming from the same motor unit, although looking familiar, were not grouped together. Another issue is the selection of the length of the segmentation window. In our work, a 6 ms window was chosen as covering the main MUAP spike duration in most of the disease cases. A shorter window will fail to contain the main MUAP spike in the case of motor neuron diseases where MUAPs usually have larger duration. This could break a long MUAP into two artificial potentials. A shorter segmentation window will result in the identification of more potential occurrences during the classification process in the case of normal or myopathic signals, since only the main spike will be included.

The proposed approach can provide a valuable tool to neurophysiologists for both MUAP detection and classification. Applying our approach when an EMG signal is processed, its constituent MUAPs are identified and classified according to their pathology. In this way, valuable information is provided to the medical expert in order to facilitate EMG diagnosis. The proposed method can easily be integrated in existing software packages, while its fully automated nature ensures that it can be easily used by a non-IT expert. Future work will integrate EMG analysis to a hybrid diagnostic system for neuromuscular diseases, which will also exploit recorded clinical data, apart from the acquired EMG signals, to provide diagnosis. Moreover, future work must address MUAP classification based on extracted MUAP features.

5. Conclusions

A novel methodology for automated EMG decomposition and MUAP diagnosis is presented. The proposed approach addresses the automatic MUAP classification to neuropathic, myopathic or normal classes directly from a raw EMG signal. This information is provided to the doctor who will make the final diagnosis to the acquired EMG signal. The superimposed MUAPs are automatically detected, decomposed into their constituents and classified to nor, myo or neu groups. The methodology has been validated using a dataset of EMG recordings and an annotated collection of MUAPs. The obtained accuracy of MUAP classification and the success rate of MUAP clustering for normal, myopathic and neuropathic are high. The proposed methodology can be integrated in a decision support system and become a valuable tool in everyday clinical practice for MUAP detection and classification.

Acknowledgments

The authors would like to thank Prof. C.S. Pattichis, University of Cyprus, for the provision of EMG data and his helpful comments. The work is partially funded by the EU (AUBADE project: IST-2002-507605).

References

[1] DeLuca CJ. Towards understanding the EMG signal, 4th ed., Baltimore: Williams & Wilkinson; 1978.
[2] Krarup C. Pitfalls in electrodiagnosis. J Neurophysiol 1999; 81:1115—26.
[3] McGill KC. Optimal resolution of superimposed action potentials. IEEE Trans Biomed Eng 2002;49:640—50.
[4] Richfield EK, Cohen BA, Albers JW. Review of quantitative and automated needle electromyographic analyses. IEEE Trans Biomed Eng 1981;506—14.
[5] Gerber A. A new framework and computer program for quantitative EMG signal analysis. IEEE Trans Biomed Eng 1984;31:857—63.
[6] Loudon GH, Jones NB, Sehmi AS. New signal processing techniques for the decomposition of EMG signals. Med Biol Eng Comput 1992;30:591—9.
[7] Nandedkar SD, Barkhaus PE, Charles A. Multimotor action potential analysis (MMA). Muscle Nerve 1995;18:1155—66.
[8] Stalberg E, Andreassen S, Falck B, Lang H, Rosenfalck A, Trojaborg W. Quantitative analysis of individual motor unit potentials: a proposition for standardized technology and criteria for measurement. J Clin Neurophysiol 1986;3:313—48.

[9] Buchthal F, Guld C, Rosenfalck P. Action potential parameters in normal human muscle and their dependence on physical variables. Acta Physiol Scand 1954;32:200—15.
[10] Buchthal F. An introduction to electromyography. Copenhagen: Gyldendal; 1957.
[11] LeFever RS, DeLuca CJ. A procedure for decomposing the myoelectric signal into its constituent action potentials: part I, execution and test for accuracy. Technique, theory and implementation. IEEE Trans Biomed Eng 1982;29:149—57.
[12] Stalberg E, Falck B, Sonoo M, Stalberg S, Astrom M. Multi-MUP EMG analysis—a two year experience in daily clinical work. In: Electroencephalography and clinical neurophysiol. Amsterdam, The Netherlands: Elsevier Science; 1995. pp. 145—154.
[13] Andreassen S. Methods for computer aided measurement of motor unit parameters. In: Ellington RJ, et al., editors. Proceedings of the London Symp., EEG suppl. 13—20. 1987.
[14] Stashuk DS, Qu H. Automatic decomposition of selective needle detected myoelectric signals. IEEE Trans Biomed Eng 1988;35:1—10.
[15] Hassoun MH, Wang C, Spitzer AR. NNERVE: neural network extraction of repetitive vectors for electromyography—part II: algorithm. IEEE Trans Biomed Eng 1994;41:1053—61.
[16] McGill K, Cummins K, Dorfman L. Automatic decomposition of the clinical electromyogram. IEEE Trans Biomed Eng 1985;32:470—6.
[17] Fang J, Agarwal GC, Shahani BT. Decomposition of multiunit electromyographic signals. IEEE Trans Biomed Eng 1999; 46:685—97.
[18] Wu S, Gong YX, Xue C, Zhi Z. Extraction of MUAP from NEMG signal using self-organization competing NN. Hefei 230026: Dept. of Electronic Science & Tech., Univ. of Science & Tech. of China; 2001.
[19] Chauvet E, Fokapu O, Hogrel JY, Gamet D, Duchene J. Automatic identification of motor unit action potential trains from electromyographic signals using fuzzy techniques. Med Biol Eng Comput 2003;41:646—53.
[20] Zennaro D, Wellig P, Koch VM, Moschytz GS, Laubli T. A software package for the decomposition of long-term multichannel EMG signals using wavelet coefficients. IEEE Trans Biomed Eng 2003;50:58—69.
[21] Christodoulou CI, Pattichis CS. Unsupervised Pattern Recognition for the Classification of EMG Signals. IEEE Trans Biomed Eng 1999;46:169—78.
[22] Kothari R, Pitts D. On finding the number of clusters. Patt Rec Lett 1999;20:405—16.
[23] Stashuk D. Decomposition quantitative analysis of clinical electromyographic signals. Med Eng Phys 1999;21:389—404.
[24] Stashuk D, Paoli GM. Robust supervised classification of motor unit action potentials. Med Biol Eng Comput 1998; 36:75—82.
[25] Nandedkar SD, Barkhaus PE, Charles A. Multi-motor unit action potential analysis (MMA). Muscle Nerve 1995; 18:1155—66.
[26] Engel AG, Armstrong C. 2nd ed., Myology, vol. 1, 2nd ed. United States: McGraw-Hill Inc.; 1994.
[27] Boser B, Guyon I, Vapnik V. A training algorithm for optimal margin classifiers. In: Proceedings of Fifth Annual Workshop on Computational Learning Theory. ACM Press; 1992.
[28] Cortes C, Vapnik V. Support-vector network. Machine Learn 1995;32:273—97.
[29] Chang C, Lin CJ. LIBSVM, a library for support vector machines. Available: http://www.csie.ntu.edu.tw/∼cjlin/libsvm [Online]; 2004 [accessed: 8 September].
[30] Knerr S, Personnaz L, Dreyfus G. Single layer learning revisited: a stepwise procedure for building and training a neural network. In: Fogelman F, Hérault J, editors. Proceedings of NATO workshop Les Arcs 1989. Springer; 1989.
[31] Hsu C-W, Lin C-J. A comparison of methods for multi-class support vector machines. IEEE Trans Neural Networks 2002;13:415—25.
[32] Chang CC, Lin CJ. LIBSVM: introduction and benchmarks. Available: http://www.csie.ntu.edu.tw/∼cjlin/libsvm [Online]; 2004 [Accessed: 8 September].
[33] Platt J. Sequential minimal optimization: a fast algorithm for training support vector machines. Microsoft Research. Technical Report MSR-TR-98-14. Washington, United States: Redmond; 1998.
[34] Keerthi SS, Shevade SK, Bhattacharyya C, Murthy KR. Improvements to Platt's SMO algorithm for SVM classifier design. Indian Institute of Science, Dept. of Computer Science and Automation. Technical Report. Bangalore, India; 1999.