Gradient Convolution Kernel Compensation Applied to Surface Electromyograms

Aleš Holobar[1] and Damjan Zazula[2]

[1] LISiN, Politecnico di Torino, Torino, Italy
ales.holobar@delen.polito.it
http://storm.uni-mb.si
[2] Faculty of Electrical Engineering and Computer Science,University of Maribor, Maribor, Slovenia

Abstract. This paper introduces gradient based method for robust assessment of the sparse pulse sources, such as motor unit innervation pulse trains in the filed of electromyography. The method employs multichannel recordings and is based on Convolution Kernel Compensation (CKC). In the first step, the unknown mixing channels (convolution kernels) are compensated, while in the second step the natural gradient algorithm is used to blindly optimize the estimated source pulse trains. The method was tested on the simulated mixtures with random mixing matrices, on synthetic surface electromyograms and on real surface electromyograms, recorded from the external anal sphincter muscle. The results prove the method is highly robust to noise and enables complete reconstruction of up to 10 concurrently active motor units.

1 Introduction

Biomedical signals are important, but very complex source of information. They typically comprise contributions of many concurrently active sources, such as neurons and muscle fibers. The sources are commonly considered statistically independent (or at most weakly correlated), but their mixing process is virtually unknown. Therefore, the acquired signals must be decomposed blindly.

When it comes to neurophysiology, electromyograms (EMG) are one of the most active research areas. They measure electrical activity of skeletal muscles and provide an insight into peripheral properties of skeletal muscles and control strategies of human motor cortex [1]. Their field of application ranges from clinical assessments of neuromuscular disorders and objective evaluations of medical treatments to basic investigations of different physiological phenomena (e.g. cramps, muscle reinnervation, etc.). Comprising action potentials (AP) from several tens of concurrently active motor units (MU), EMG signal is commonly considered a random interference pattern which is very difficult to interpret [2]. This is especially true in the case of surface electromyography [1], which deals with measuring the electrical activity of human muscles on the surface of the skin. Bad electrode-to-skin contact and presence of noise hinder the decomposition of surface EMG and make the extraction of clinically relevant information difficult.

M.E. Davies et al. (Eds.): ICA 2007, LNCS 4666, pp. 617–624, 2007.

Recent development of high-density surface electrode arrays enabled acquisition of several tens or even hundreds of EMG channels. Different pattern recognition techniques, capable of dealing with such amount of data were also proposed. Kleine et al. [3] studied the importance of two-dimensional spatial filters, Gazzoni et al. [4] introduced the template matching segmentation and classification technique, while Wood et al. [5] employed the finite element analysis. Blind source separation decomposition techniques have also been proposed. Garcia et al. [6] modelled the EMG signal as an instantaneous mixture of motor unit action potential (MUAP) trains, while Holobar and Zazula [7] proposed Convolution Kernel Compensation (CKC) to deal with the convolutive mixtures of motor unit innervation pulse trains (IPT). The latter proved to be highly efficient as it enables the complete reconstruction of up to 30 concurrently active MUs from a good quality multichannel surface EMG.

In this paper the gradient-based extension of the CKC method [7] applied to the low-quality noisy signals is addressed. This extension is of paramount importance for clinical practice, where recoding environment cannot be strictly controlled. This paper is organized in five sections. In Section 2, the assumed data model is presented and the classic CKC approach is briefly summarized. Section 3 introduces its gradient-based extension, while in Section 4 the results of tests on simulated and real EMG signals are presented. Section 5 discusses the results and concludes the paper.

2 Data Model and Convolution Kernel Compensation

Suppose M convolutive measurements are simultaneously observed and denote their sampled vector by $\mathbf{x}(n) = [x_1(n), ..., x_M(n)]^T$, where $x_i(n)$ stands for the n-th sample of the i-th measurement. In the case of linear time-invariant (LTI) multiple-input multiple-output (MIMO) system, $\mathbf{x}(n)$ can be written as:

$$\mathbf{x}\left(n\right) = \mathbf{H}\bar{\mathbf{t}}\left(n\right) + \omega\left(n\right) \tag{1}$$

where $\omega(n) = [\omega_1(n), ..., \omega_M(n)]^T$ is a zero-mean spatially and temporally white additive noise vector, $\bar{\mathbf{t}}\left(n\right) = [t_1\left(n\right), t_1\left(n-1\right), ..., t_1\left(n-L+1\right), ..., t_N\left(n\right), ... \; t_N\left(n-L+1\right)]^T$ is the extended version of vector of input signals from N sources $\mathbf{t}\left(n\right) = [t_1\left(n\right), ..., t_N\left(n\right)]^T$, and the mixing matrix $\mathbf{H}$ comprises all the channel responses (convolution kernels) $\mathbf{h}_{ij} = [h_{ij}\left(0\right), ..., h_{ij}\left(L-1\right)]$ of length of L samples:

$$\mathbf{H} = \begin{bmatrix} h_{11}(0) & \dots & h_{11}(L-1) & h_{12}(0) & \dots & h_{12}(L-1) & \dots \\ h_{21}(0) & \dots & h_{21}(L-1) & h_{22}(0) & \dots & h_{22}(L-1) & \dots \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots & \dots \\ h_{M1}(0) & \dots & h_{M1}(L-1) & h_{M2}(0) & \dots & h_{M2}(L-1) & \dots \end{bmatrix} \tag{2}$$

In surface electromyography, the channel response $\mathbf{h}_{ij}$ corresponds to the j-th MUAP, as detected by the i-th measurement, while each model input $t_j\left(n\right)$ is a

sample of an IPT, modelled as a binary pulse sequence carrying the information about the MUAP triggering times only:

$$t_j(n) = \sum_{k=-\infty}^{\infty} \delta\left[n - T_j(k)\right] \tag{3}$$

where $\delta(.)$ denotes the Dirac impulse and $T_j(k)$ stands for the time instant in which the k-th MUAP of the j-th MU appears.

2.1 Convolution Kernel Compensation

CKC method compensates the unknown mixing matrix $\mathbf{H}$ in model (1) and directly estimates the innervation pulse trains $\hat{t}_j(n)$ [7]:

$$\hat{t}_j(n) = \mathbf{c}_{t_j\mathbf{x}}^T \mathbf{C}_{\mathbf{xx}}^{-\mathbf{1}} \mathbf{x}(n) \tag{4}$$

where $\mathbf{C}_{\mathbf{xx}} = E\left(\mathbf{x}(n)\mathbf{x}^T(n)\right)$ is the correlation matrix of measurements, $\mathbf{c}_{t_j\mathbf{x}} = E\left(\mathbf{x}(n)t_j^T(n)\right)$ is cross-correlation vector, and $E(.)$ denotes mathematical expectation. Estimator (4) is linear minimum mean square error (LMMSE) estimator of the j-th IPT and requires the cross-correlation vector $\mathbf{c}_{t_j\mathbf{x}}$ to be known in advance. This is rarely the case and Holobar and Zazula [7] proposed probabilistic iterative procedure for its blind estimation. In the first iteration step, the unknown cross-correlation vector $\mathbf{c}_{t_j\mathbf{x}}$ is approximated by vector of measurements $\hat{\mathbf{c}}_{t_j\mathbf{x}} = \mathbf{x}(n_1)$ where, without loss of generality, we assumed the j-th MU discharged at time instant n_1. Then, the first estimation of the j-th IPT yields

$$\hat{t}_j(n) = \hat{\mathbf{c}}_{t_j\mathbf{x}}^T \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{x}(n) \tag{5}$$

In the second step, the largest peak in $\hat{t}_j(n)$ is selected as the most probable candidate for the second discharge of the j-th source, $n_2 = \underset{\arg}{\max}(t_j(n))\,|_{n_2 \neq n_1}$, and the vector $\hat{\mathbf{c}}_{t_j\mathbf{x}}$ is updated as:

$$\hat{\mathbf{c}}_{t_j\mathbf{x}} = \frac{\hat{\mathbf{c}}_{t_j\mathbf{x}} + \mathbf{x}(n_2)}{2} \tag{6}$$

This procedure is then repeated, with a special attention paid to the separation of superimposed pulse sources (note that more than one source may be active at instant n_1). Interested reader is referred to [7] for further description of classic CKC approach.

3 Gradient Descent Optimization of Estimated Pulse Trains

Let us use shorthand notation $\mathbf{w}_{j,k} = \mathbf{C}_{\mathbf{xx}}^{-1}\hat{\mathbf{c}}_{t_j\mathbf{x}}$ to denote the estimation of the j-th separation vector in the k-th iteration step and let $F(\hat{t}_j) = \sum_m f(\hat{t}_j(m))$ denote

sample cost function in the manifold of pulse trains, with arbitrary differentiable scalar function $f(t)$ applied to each sample of pulse train $\hat{t}_j(n)$. General iteration step for natural gradient descent algorithm is defined as [8]:

$$\mathbf{w}_{j,k+1} = \mathbf{w}_{j,k} - \eta(k)\mathbf{G}\Delta\mathbf{w}_{j,k} \tag{7}$$

where $\eta(k)$ is the learning rate and $\mathbf{G}$ is Riemannian metric tensor embedding the manifold of pulse trains $\mathbf{t}\,(n)$ into the manifold of measurements $\mathbf{x}\,(n)$. For smooth manifolds, the induced metric tensor $\mathbf{G}$ can be computed as:

$$\mathbf{G} = \mathbf{H}^{-1}\mathbf{H}^{-T} \tag{8}$$

where, in our case, $\mathbf{H}^{-1}$ stands for the Jacobian of the system $\mathbf{t}\,(n) = \mathbf{H}^{-1}\mathbf{x}\,(n)$. Taking the common convention on the amplitude ambiguity of trains $\mathbf{t}\,(n)$ into account, we assume $\mathbf{C}_{\overline{\mathbf{tt}}} = \mathbf{I}$. Hence, $\mathbf{G}$ can be written as $\mathbf{G} = \mathbf{H}^{-1}\mathbf{C}_{\overline{\mathbf{tt}}}^{-1}\mathbf{H}^{-T} = \mathbf{C}_{\mathbf{xx}}^{-1}$. Finally, by expressing the gradient $\Delta\mathbf{w}_{j,k}$ in terms of cost function $F(\hat{t}_j)$ we derive to the following gradient update rule:

$$\mathbf{w}_{j,k+1} = \mathbf{w}_{j,k} - \eta(k)\sum_m \frac{\partial f(\hat{t}_j(m))}{\partial \hat{t}_j(m)}\mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{x}(m) \tag{9}$$

There is yet another possible interpretation of update rule (9). By rewriting (9) in terms of (5) we get the update rule for the cross-correlation vector $\hat{\mathbf{c}}_{t_j\mathbf{x}}$:

$$\hat{\mathbf{c}}_{t_j\mathbf{x}} = \hat{\mathbf{c}}_{t_j\mathbf{x}} - \eta(k)\sum_m \frac{\partial f(\hat{t}_j(m))}{\partial \hat{t}_j(m)}\mathbf{x}(m) \tag{10}$$

which provides insight into the convergence properties of the algorithm (9). Namely, by selecting the $\frac{\partial f(t)}{\partial t}$ to be concave even function, e.g. $\frac{\partial f(t)}{\partial t} = t^2$, the peaks in $\hat{t}_j(n)$ get reinforced (i.e. the corresponding measurement vectors $\mathbf{x}(m)$ in (10) get multiplied by large weights), while the base-line noise (i.e. values close to zero) is suppressed. The steeper the function $\frac{\partial f(t)}{\partial t}$, the larger the weights multiplying the peaks in $\hat{t}_j(n)$, and the faster the convergence of (10). However, by setting the peak weights too high, we jeopardize the stability of convergence as it may happen that the highest peak in $\hat{t}_j(n)$ outweights all the others. In such a case, $\hat{\mathbf{c}}_{t_j\mathbf{x}}$ converges to $\mathbf{x}(m_p)$, where m_p denotes the time instant of the largest peak in $\hat{t}_j(n)$. Typically, $\frac{\partial f(t)}{\partial t} = |t|$ or $\frac{\partial f(t)}{\partial t} = t^2$ prove to be a good compromise between the speed and stability of convergence, yielding the cost functions $F(\hat{t}_j) = \frac{1}{2}\sum_m \hat{t}_j(m)\sqrt{\hat{t}_j^2(m)}$ and $F(\hat{t}_j) = \frac{1}{3}\sum_m \hat{t}_j^3(m)$, respectively.

Gradient descent algorithm (9) still requires a good initial approximation of $t_j(n)$ in order to converge to the genuine solution (i.e. the peaks in $\hat{t}_j(n)$ must represent the true train pulses). As demonstrated in the next section, the CKC approximation (5) with $\hat{\mathbf{c}}_{t_j\mathbf{x}} = \mathbf{x}(n_1)$ proves to be a good initialization point. Required initial time instants n_1 can be selected from the activity index $I_A(n) = \mathbf{x}^T(n)\mathbf{C}_{\mathbf{xx}}^{-\mathbf{1}}\mathbf{x}(n)$, as suggested in [7].

4 Simulation and Experimental Results

Gradient CKC method was applied to three different sets of test signals. The first two experiments evaluated the influence of noise in the case of multichannel synthetic measurements with well-conditioned random mixing matrix $\mathbf{H}$ and in the case of badly-conditioned synthetic surface EMG measurements, while in the third experiment, the method was applied to recordings of external anal sphincter muscle. In all three experiments the scalar function $f(t) = \frac{1}{3}t^3$ was used in (9), while in each iteration step the adaptive learning rate $\eta(k)$ was adjusted according to bisection. The sensitivity of gradient CKC algorithm, false alarm rate and the number of reconstructed pulse trains were observed and compared to the results of classic CKC approach.

Experiment 1: Ten simulation runs were performed, with the number of sources N set equal to 10. In each run, random input pulse trains $t_j(n) = \sum_{k=1}^{200} \delta\left(n - k \cdot 100 + T_j(k)\right)$ were generated with the mean inter-pulse interval (IPI) set equal to 100 samples and the values $T_j(k); k = 1, 2, \ldots, 200$ uniformly distributed on the interval $[-10, 10]$. The length of simulated pulse trains $\mathbf{t}$ was 20,000 samples. Random zero-mean mixing matrix $\mathbf{H}$ was generated, with $L = 10$ samples long convolution kernels. The number of observations M was set equal to 25. Seven realizations of Gaussian zero-mean noise per each generated signal (1) were simulated, with SNR ranging from 20 dB to -10 dB. In order to increase the number of measurements, 9 delayed repetitions of each original measurement were used as additional measurements [7]. As a result, the number of extended pulse trains increased to 190, while the number of observations was fixed at 250. Each mixture was decomposed two times - by classic and gradient CKC. The results are summarized in Fig. 1. The CKC gradient method converged after 15 iterations, on average.

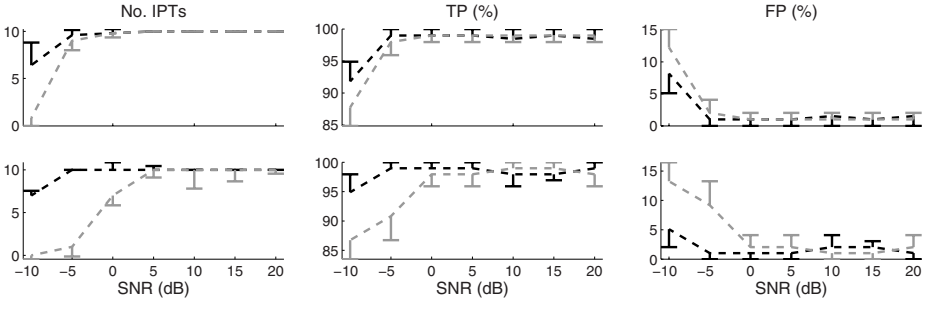

Fig. 1. Number of reconstructed IPTs (*left column*), True Positive rate (TP) (*central column*) and False Positive rate (FP) (*right column*) for gradient CKC (*black line*) and classic CKC (*gray line*). The results are averaged over 10 simulation runs (*error bars* indicate std. deviations). In each run, random mixing matrix $\mathbf{H}$ was generated, with condition number set equal to 240 ± 10 (*upper row*) and 1200 ± 50 (*bottom row*).

Experiment 2: Synthetic surface EMG signals were generated by cylindrical volume conductor model consisting of bone, muscle, subcutaneous, and skin tissues [2]. Biceps Brachii muscle with 200 MUs and 200 mm^2 cross-section was simulated. The distribution of the MU locations was random and the fibers of a MU were randomly scattered in circular MU territory, with a density of 20 fibers/mm^2. Exponentially distributed innervation numbers ranged from 25 to 2500. The surface-recorded MUAP comprised the sum of the action potentials of the muscle fibers belonging to the MU. The MUs had muscle fiber conduction velocities of 4 ± 0.3 m/s. The recording system was a grid of 13×5 electrodes of circular shape (1-mm radius) with 5-mm interelectrode distance in both directions. The EMG signals were additionally corrupted by additive zero-mean Gaussian noise between 20 and 0 dB SNR. Ten seconds long contraction at 10% excitation level (constant over time) was simulated what resulted in 105 active MUs. MU IPTs were based on a motor unit population recruitment model [3] with the recruitment and the peak discharge rate set to 8 and 35 pulses per second. In likefashion to the *Experiment 1*,9 delayed repetitions of each original measurement were added to the original set of measurements. The results of CKC decomposition, averaged over 25 Monte Carlo runs are depicted in Fig. 2.

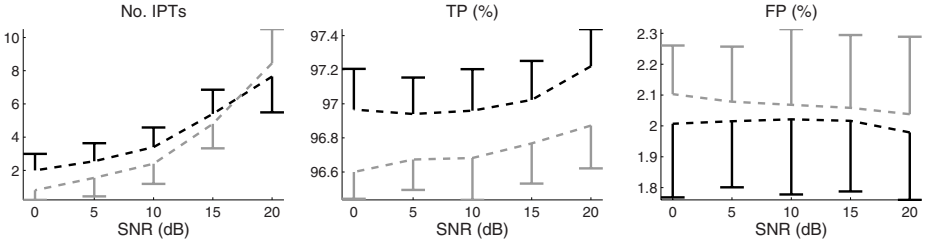

Fig. 2. Number of reconstructed IPTs (*left panel*), True Positive rate (TP) (*central panel*) and False Positive rate (FP) (*right panel*) for gradient CKC (*black line*) and classic CKC (*gray line*) when decomposing synthetic surface EMG (*Experiment 2*). The results are averaged over 25 simulation runs (*error bars* indicate std. deviations).

Experiment 3: The last experiment was conducted on real surface EMG signals, recorded by a 48-channel cylindrical anal probe from the external anal sphincter muscle. The electrodes (1×10 mm) were arranged in 3 circumferential arrays of 16 electrodes each. Interelectrode and inter-array distance was 2.7 mm and 5 mm, respectively. The experiment was conducted in Gynecological Clinic at University of Tübingen, Germany, and was approved by the local ethics committee. Six subjects participated to the experiment. The signals were acquired during three 10 s long maximum voluntary contractions. The EMG signals were amplified, band-pass filtered (3 dB bandwidth, 10 Hz-500 Hz), sampled at 2 kHz, and converted to digital form by a 12-bit A/D converter. The acquired set of measurements was additionally extended by 9 delayed repetitions of each

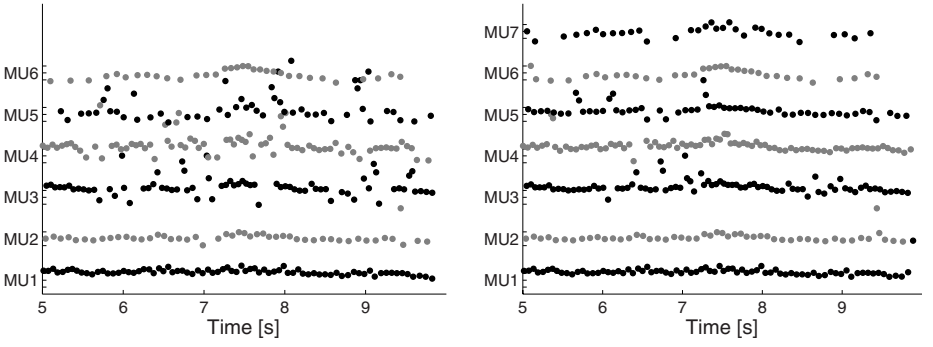

Fig. 3. IPTs reconstructed from real surface EMG by classic CKC (*left panel*) and by gradient CKC (*right panel*). Each plotted dot corresponds to single MU discharge. EMG signals were recorded from the external anal sphincter muscle.

Table 1. Number of MUs (mean ± std. dev.) reconstructed from real EMG signals

Subject	A	B	C	D	E	F
classic CKC	2.7 ± 0.6	3.7 ± 1.2	3.3 ± 0.6	3.3 ± 0.6	3.3 ± 0.6	2.3 ± 1.2
gradient CKC	6.3 ± 1.2	8.3 ± 1.2	4.7 ± 0.6	5.7 ± 1.2	8.0 ± 2.0	3.3 ± 0.6

measurements. The results of decomposition are summarized in Table 1 and exemplified by Fig. 3.

5 Discussion and Conclusions

The gradient CKC method proved to be highly efficient. In low noise environments, it is equivalent to the classic CKC approach. In the presence of severe noise, however, it provides superior accuracy. The results on synthetic measurements with random mixing matrix $\mathbf{H}$ proved that almost complete reconstruction of pulse trains at the SNR of -5 dB is possible. In the case of synthetic surface EMG, the robustness to noise was reduced. This was mainly due to the badly conditioned mixing process (in the case of surface EMG, typical condition number of the mixing matrix $\mathbf{H}$ is about 10^7). Nevertheless, up to 5 pulse trains were reconstructed down to the SNR of 10 dB, with the pulse accuracy exceeding the 97%. When compared to the classic CKC approach at low SNR, the gradient CKC yielded 30% increase in the number of reconstructed IPTs and 0.5% increase in the accuracy of the reconstructed pulse trains. Similar results were observed in the case of real surface EMG signals, acquired from the external anal sphincter. The general quality of acquired signals was low, mainly due to the bad electrode-mucosa contact and movement of the anal probe with respect to the muscle fibers. The gradient CKC technique reconstructed 6.2 ± 2.3 MUs per contraction (compared to 3.1 ± 0.8 MUs reconstructed by the classic CKC).

MU discharge patterns reconstructed by gradient CKC also exhibited higher regularity than those of the classic CKC method, as verified by careful visual inspection.

It is concluded that gradient CKC is highly robust to noise. In low SNR environment, it yields the performance superior to the classic CKC approach and has the potential to be used in regular clinical practice, where the quality of acquired signals cannot be strictly controlled.

Acknowledgment

The authors are sincerely grateful to Prof. R. Merletti from Politecnico di Torino, Italy, and to Prof. P. Enck, Dr. H Hinninghofen and Dr. T. Kreutz from University of Tübingen, Germany, for their measurements of the external anal sphincter. This research was supported by a Marie Curie Intra-European Fellowships within the 6th European Community Framework Programme (Contract No. MEIF-CT-2006-023537), and by German-Italian research project TASI.

References

1. Merletti, R., Parker, P.A.: Electromyography: physiology, engineering, and noninvasive applications, IEEE Press and John Wiley & Sons (2004)
2. Farina, D., Merletti, R.: A novel approach for precise simulation of the EMG signals detected by surface electrodes. IEEE trans. Biomed. Eng. 48, 637–646 (2001)
3. Kleine, B.U., van Dijk, J.P., Lapatki, B.G., Zwarts, M.J., Stegman, D.F.: Using two-dimensional spatial information in decomposition of surface EMG signals, J. of Electromy and Kinesiol (2006)
4. Gazzoni, M., Farina, D., Merletti, R.: A new method for the extraction and classification of single motor unit action potentials from surface EMG signals. J. of Neurosc. Methods 136, 165–177 (2004)
5. Wood, S.M., Jarratt, J.A., Barker, A.T., Brown, B.H.: Surface electromyography using electrode arrays: a study of motor neuron diseases. Muscle Nerve 24, 223–230 (2001)
6. Gracia, G.A., Okuno, R., Akazawa, K.: A Decomposition Algorithm for Surface Electrode-Array Electromyogram: A Noninvasive, Three-Steep Approach to Analyze Surface EMG Signals. IEEE Eng. in Med. Biol. Magazine 5, 63–71 (2005)
7. Holobar, A., Zazula, D.: Multichannel Blind Source Separation Using Convolution Kernel Compensation, IEEE Trans.Sig. Process., 56 (2007)
8. Amari, S.I.: Natural Gradient Works Efficiently in Learning. Neural Computation 10, 251–276 (1998)
9. Fuglevand, A.J., Winter, D.A., Patla, A.E.: Models of recruitment and rate coding organization in motor unit pools. J. Neurophysiol 70, 2470–2488 (1993)