Neural Network Approaches To The Processing of Experimental Electro-Myographic Data from Non-Invasive Sensors

Francesco Carlo Morabito, Maurizio Campolo
University "Mediterranea" of Reggio Calabria
Faculty of Engineering, DIMET
Reggio Calabria (Italy)
E-mail: morabito@unirc.it, campolo@ing.unirc.it

Abstract

Learning theories and algorithms for both supervised and unsupervised Neural Networks (NNs) have already been accepted as relevant tools to cope with difficult problems based on the processing of experimental electromagnetic data. These kinds of problems are typically formulated as inverse problems. In this paper, in particular, the electrical signals under investigations derive from experimental electromyogram interference patterns measured on human subjects by means of non-invasive sensors (surface ElectroMyoGraphic, sEMG, data). The monitoring and the analysis of dynamic sEMG data reveals important information on muscles activity and can be used to clinicians for both preventing dramatic illness evolution and improving athletes performance. The paper proposes the use of Independent Component Analysis (ICA), an unsupervised learning technique, in order to process raw sEMG data by reducing the typical "cross-talk" effect on the electric interference pattern measured by the surface sensors. The ICA is implemented by means of a multi-layer NN scheme. Since the IC extraction is based on the assumption of stationarity of the involved sEMG recording, which is often inappropriate in the case of biomedical data, we also propose a technique for dealing with non-stationary recordings. The basic tool is the wavelet (time-frequency) decomposition, that allows us to detect and analyse time-varying signals. An auto-associative NN that exploits wavelet coefficients as an input vector is also used as simple detector of non-stationarity based on a measure of reconstruction error. The proposed approach not only yields encouraging results to the problem at hand, but suggests a general approach to solve similar relevant problems in some other experimental applications of electromagnetics.

1. Introduction

Most relevant medical problems are today faced by processing (by visual inspection or some automatic means) electrical signals detected on the human body. Evaluation of patient populations often includes the use of ancillary tests for diagnosis and/or prognosis. Data sets collected from these diagnostic tests, such as the Electroencephalogram (EEG), the Electromyogram (EMG), the Electrocardiogram (ECG) and, more recently, functional Magnetic Resonance Imaging (fMRI), tend to be complex, large and high-dimensional. The trend towards digitization of the traditionally analog EEG, EMG and ECG signals has coincided with the development of computing power and multivariate signal processing techniques capable of manipulating and analyzing such large data sets [Akay M.,1997].

The use of Independent Component Analysis (ICA), an unsupervised learning technique which generalizes Principal Component Analysis (PCA), commonly implemented through Neural Network (NN) schemes, is proposed in this study to process experimental biomedical data. Applied to sEMG (surface ElectroMyoGraphy) data, ICA results in numerous spatially-independent patterns, each associated with a unique time-course, providing a way to separate different electrical signals coming from different muscle activities [Jung T.P., 2000]. In contrast to the variable nature of the surface EMG recorded from a single muscle in isolation, ICA of the sEMG from several muscles simultaneously allows the detection of highly reproducible components for example in the sEMG of the face and the throat during swallowing and in the sEMG of arm muscles during reaching movements [McKeown M.J., 2002].

The researches reported in the present study show important applications in the study of some neurological diseases, and in the monitoring of athletic activities for improving significantly the potential of athletes as well as the capabilities of normal subjects in daily actions, since it makes it possible, in principle, to enhance motor coordination. Also, musculo-skeletal disorders are the first cause of patient-physician encounters in the industrialized countries [IEEE Engineering in Medicine and Biology, 2001].

This paper is organized as follows. In Section 2 the type of data coming from electrical activity of muscles will be discussed. In Section 3 we shall propose the McKeown idea of motion through integration of sub-movements [McKeown M.J., 200b]. The computational model incorporating sub-movements will be presented in Section 4. Section 5 is devoted to the proposal of NN schemes to implement ICA. Section 6 will report the results achieved by processing the experimental data. The assumption of stationarity of the electrical signals will be relaxed in Section 7, where the wavelet approach will be proposed. Finally, some conclusions are drawn.

2. ElectroMyographic Data

When skeletal muscle fibers contract, they conduct electrical activity (action potentials, APs) that can be measured by electrodes affixed to the surface of the skin above muscles [Akay M., 1997]. As the APs pass by the electrodes, spikes of electrical activity are observed and pulses of muscle fiber contractions are produced. Small functional groups of muscle fibers, termed motor units (MUs), contract synchronously, resulting in a motor unit action potential (MUAP). To sustain force, an MU is repeatedly activated by the central nervous system several times per second. The repetition, or average, firing rate is often between 5 and 30 times per second (or faster). The electromyographic (EMG) signal is widely used as a suitable means to have access to physiological processes involved in producing joint movements. The information extracted from the EMG signals can be exploited in several different applications. The typical sensors used for EMG are needle (unipolar or bipolar) sensors. The experimental data here analysed come from non-invasive surface EMG sensors, that present the cross-talk effect, i.e., they detect electrical activities from several muscles simultaneously in action.

3. Sensorimotor integration of sub-movements

A growing body of evidence suggests movements which appear smooth to the naked eye are actually composed of the temporal and spatial superposition of discrete sub-movements precisely recruited and coordinated by the central nervous system [Harris C.M., 1998]. However, the spatial and temporal overlap of sub-movements has generally made it impossible, with the common computational tools available to the neuroscientist, to isolate the effects of individual sub-movements [Sejnowski T.J., 1998].

Extensive computational expertise is required to adequately interpret the data gleaned from the experiments. Detection of non-stationarity in the sEMG and kinematic variables is necessary to detect the onset of temporally overlapping sub-movements. We investigate the information-theoretic considerations of channel capacity and bandwidth as important determinants in the selection and sensorimotor integration of individual sub-movements.

4. Computational Models incorporating Sub-movements

Some computational approaches have attempted to model reaching movements as incorporating sub-movements; however, they have not addressed many of the unanswered questions regarding the characteristics of sub-movements. Others have attempted to model reaching movements without considering sub-movements at all. Smoothness, an empirical observation of motor movements, has often used as a cost function to optimise the models. Rather than define sub-movements on the basis of the velocity profiles, in this project the sub-movements are defined on the basis of muscular activity. Empirically, experienced physical therapists describe “efficiency” of motor movements as subjects progressively recover. At some stage of rehabilitation, people are able to mimic normal kinematics but still complain of muscle aching and fatigue due to excessive muscle co-contraction.

Intuitively, sub-movements are groups of muscles that have the tendency to activate together following a common neural input. We assert that a sub-movements is “hard-wired” by adulthood, in the sense that it may be encoded in the spinal cord as part of a Central Pattern Generator (CPG), and also partly reflect the anatomical distribution across several muscles of a single nerve root exiting the spinal cord. To suggest a computational model of sub-movements, we initially make the stationarity assumption. Since the EMG is an indirect measure of the neural command to the muscle, the Mutual Information (MI) can be used as a metric to infer the recordings from two EMG electrodes contain common neural input. M. McKeown has proposed using ICA for the analysis of sEMGs, demonstrating that the Independent Components (ICs) are more strongly coupled with ongoing brain rhythms (EEG) than the sEMGs recordings of individual muscles [McKeown M.J., 2000a]. The ICA model can be used to provide a useful starting point for the rigorous definition of a sub-movement upon which more elaborate models can be created. Consider numerous simultaneous sEMG recordings deriving from several electrodes distributed over many muscles during a coordinated cortically-controlled movement. If we model the sEMGs recorded from each electrode to be the linear superposition of activity from different group of muscles (possibly encoded with CPGs) that tend to co-activate, the, the goal is to estimate the cortical modulation of the commonly influenced muscles. A single sub-movement is defined as m(t) = U C(t), t=t0→tn, where m is a column vector, with mj representing the muscle electrical activity contributing to the jth electrode as a function of time, U is a stationary column vector representing the relative weighting that a given cortical command gives to the different muscle areas, and C(t) is the unknown scalar neural command over time. If several, e.g. p, sub-movements during a complex movement are temporally (and spatially) overlapping, the linear combination of $m_k(t_k)$ outputs M(t), the total muscle electrical activity over the duration of the whole movement and M_j is the electrical activity recorded from the jth electrode, Ck represents the relative activation of the kth sub-movement by an independent cortical command, and the matrix $U_{j,k}$ has as its columns, U_k, the vectors defining the different sub-movements. If we assume that for a given time-period, say T, a constant number of sub-movements, c, are simultaneously active, thus, we have M = UC, where M is the matrix of the electrical activity, C is the matrix of presumed independent cortical commands, and U is a matrix defining the sub-movements. The goal is then, given the recordings from the electrodes, and not knowing U, to estimate the different cortical influences, **C**. If the C_k are assumed to be independent, and c can be estimated, this is possible through the ICA.

5. Neural models of ICA

Independent Component Analysis (ICA) can easily be introduced as a straightforward evolution of the well-known statistical technique referred to as Principal Component Analysis (PCA). Nevertheless, it is also possible to investigate the main ideas behind ICA from the perspectives of both learning/neural systems and signal processing (blind source separation). A good definition of ICA can be found in [Lee T.W., 1998]: ICA is a method for finding a linear non-orthogonal co-ordinate system in any multivariate data. The directions of the axes of this co-ordinate system are determined by both the second and higher order statistics of the original data. The goal is to perform a linear transformation which makes the resulting variables as statistically independent from each other as possible. In contrast to correlation-based transformations such as PCA, ICA not only decorrelates the signals, through second-order statistics, but also reduces higher-order statistical dependencies. Blind source separation by ICA has received attention because of its potential applications in signal processing. Here, the goal is to recover independent sources given only sensor observation that are unknown linear mixtures of the latent (unobserved), possibly independent, source signals. In parallel to blind source separation researches, the ICA emerged within the framework of unsupervised learning. In particular, Linsker [Linsker R.] firstly proposed an algorithm based on information theory that was then used to maximize the mutual information between the inputs and the outputs of a NN. Each neuron of an "output" layer should be able to encode features that are as statistically independent as possible from other neurons over another ensemble of "inputs". The statistical independence of the outputs implies that the multivariate probability density function (pdf) of the outputs can be factorised as a product of marginal pdf's. Bell and Sejnowski [Bell A.J., 1995], derived stochastic gradient learning rules for achieving the prescribed maximization. The same Authors put the problem in terms of an information-theoretic framework and demonstrated the separation and deconvolution of linearly mixed sources [Bell A.J., 1996].

Among the various approaches proposed in the literature to implement the ICA, the approach used by McKeown [Lee T.W., 1999] is the algorithm developed by Bell and Sejnowski [Bell A.J., 1995] which is based on an Infomax NN, where a self-organizing algorithm is used to maximize the information transferred in a network of non-linear units. The general framework of ICA is now simply described as the blind separation problem, typically introduced by the "cocktail party problem": we have n different sources $\underline{s}_j$ (that is, the speakers i=1,…,n) and m different linear mixtures $\underline{x}_j$ (that is, the microphones j=1,…,m). By referring to $\underline{\underline{x}}$ as the matrix of the observed signals, and as $\underline{\underline{s}}$ the matrix of the independent components, the matrix $\underline{\underline{W}}$, called unmixing matrix, satisfies the following property:

$$\underline{\underline{s}} = \underline{\underline{W}} \cdot \underline{\underline{x}} \tag{1}$$

or, by defining the mixing matrix $\underline{\underline{A}}$ as:

$$\underline{\underline{x}} = \underline{\underline{A}} \cdot \underline{\underline{s}} \tag{2}$$

then the mixing and unmixing matrixes are related by the following equation:

$$\underline{\underline{W}}^{-1} = \underline{\underline{A}} \,. \tag{3}$$

5.1 The ICA based on the information maximization by using a neural network approach

Bell and Sejnowski derived a self-organizing learning algorithm to maximize the information transferred to a NN of non-linear units. The non-linear transfer functions pick up the higher-order moments of the statistical distribution of the input data, and, moreover, they are able to reduce the redundancy in the output data. Higher-order methods use information on the distribution of $\underline{\underline{x}}$ that is not contained in the covariance matrix. This fact becomes meaningful when the distribution of $\underline{\underline{x}}$ is non Gaussian, since it is possible to assume that the covariance matrix of a zero mean Gaussian variable, contains the whole information carried by this variable. By defining the differential entropy for a continuous random variable x as:

$$H(x) = -\int_{-\infty}^{+\infty} f_x(x) \cdot \ln[f_x(x)] \cdot dx \tag{4}$$

when $f_x(x)$ is the probability density function of the considered variable. The conditional differential entropy is defined as follows:

$$H(y\,|\,x) = -\int_{-\infty}^{+\infty} f_x(x) \int_{-\infty}^{+\infty} f_y(y\,|\,x) \cdot \ln[f_y(y\,|\,x)] \cdot dy \cdot dx \tag{5}$$

It represents to the variations that occur in the information carried by y when x is observed. Finally the mutual information between two variables x and y is given by:

$$MI(x,y) = H(x) - H(x\,|\,y) = H(y) - H(y\,|\,x) \,. \tag{6}$$

This quantity measures the information that is added to x when y is observed, or to y when x is observed. The mutual information of (x, y) zeroes, when and only when the variables are independent. The Bell-Sejnowski approach is based on the use of a NN able to minimize the mutual information between the input $\underline{x}$ and the output $\underline{y}$ of the neural network where $\underline{y}$ are the independent components. If we suppose to have noise-free input data, $\underline{y}$ can be obtained from $\underline{x}$ by a deterministic manner: in this case, $H(\underline{y}|\underline{x})$ assumes its lowest value (-∞). The problem in this case is that the density functions of the unknown components cannot be computed, and therefore the $H(\underline{y}|\underline{x})$ is difficult to be estimated. This drawback can be overcame by taking into account that, if $\underline{y}$ can be computed from $\underline{x}$ by an invertible continuous deterministic mapping, the maximization of $MI(\underline{x}|\underline{y})$ corresponds to maximize the entropy of the outputs. In the NN case, we have to maximize the H($\underline{y}$) with respect to the network parameters $\underline{\underline{w}}$. If we have just one input x and one output y, if the mapping from x to y is defined as y=g(x), and if *g(•)* has a unique inverse, then the probability density function of y can be computed as:

$$f_y(y) = \left|\frac{\partial y}{\partial x}\right|^{-1} \cdot f_x(x) \quad (7)$$

The differential entropy of y is given by:

$$H(y) = -E[\ln(f_y))] = -\int_{-\infty}^{+\infty} f_y(y) \cdot \ln[f_y(y)]\, dy = \quad (8)$$

$$= E\left[\ln\left|\frac{\partial y}{dx}\right|\right] - E[\ln(f_x(x))]$$

To maximize the differential entropy, we need to maximize just the first term. This maximization is carried out by a stochastic gradient ascent learning rule, where the update step can be computed as:

$$\Delta w \propto \frac{\partial H}{\partial w} = \frac{\partial}{\partial w}\left(\ln\left|\frac{\partial y}{\partial x}\right|\right) = \left(\frac{\partial y}{\partial x}\right)^{-1} \cdot \frac{\partial}{\partial w}\left(\frac{\partial y}{\partial x}\right) \quad (9)$$

If *g(•)* becomes the logistic transfer function, of the scaled and translated input:

$$y = \frac{1}{1 + e^{-(w \cdot x + w_0)}} \quad (10)$$

the update term can be rewritten as the update step for the weight w:

$$\Delta w \propto \frac{1}{w} + x \cdot (1 - 2y) \quad (11)$$

and the update step for the bias weight can be computed as:

$$\Delta w_0 \propto 1 - 2y \quad (12)$$

In the most general multivariate case, we have:

$$f_{y_1, y_2, \ldots, y_N}(y_1, y_2, \ldots, y_N) = |\underline{\underline{J}}|^{-1} \cdot f_{x_1, x_2, \ldots, x_N}(x_1, x_2, \ldots, x_N) \quad (13)$$

where $\underline{\underline{J}}$, is the transformation Jacobian. The update step for the matrix weight becomes:

$$\underline{\underline{\Delta W}} \propto \underline{\underline{W}}^{-T} + (\underline{1} - 2\underline{y}) \cdot \underline{x}^T \quad (14)$$

where $\underline{1}$ is a unit column vector and the update step for the bias weight vector can be computed as:

$$\underline{\Delta w}_0 \propto \underline{1} - 2\underline{y} \quad (15)$$

The input data are measurements of N different input sources, and, therefore, they can be referred to as a matrix $\underline{\underline{x}}$, where the i-th column represents the i-th sample of the each source. The inputs of the neural network are $\underline{\underline{h}} = \underline{\underline{W}} \cdot \underline{\underline{x}}_s$ and $\underline{\underline{x}}_s$ are called sphered data. The sphered data are computed by zero-meaning the input data $\underline{\underline{x}}$ and by sphering these data with the following matrix operation:

$$\underline{\underline{x}}_s = \underline{\underline{S}} \cdot \underline{\underline{x}}_0 \quad (16)$$

$$\underline{\underline{x}}_0 = \underline{\underline{x}} - E[\underline{\underline{x}}] \quad (17)$$

$$\underline{\underline{S}} = 2\left(\sqrt{E[\underline{\underline{x}}_0 \cdot \underline{\underline{x}}_0^T]}\right)^{-1} \quad (18)$$

where $\underline{\underline{S}}$ is called sphering matrix, and it is used to speed the convergence. The infomax NN estimate the matrix $\underline{\underline{y}}$, where the i-th column represents the i-th sample of the each independent component. The architecture of the neural network is depicted in Figure 1.

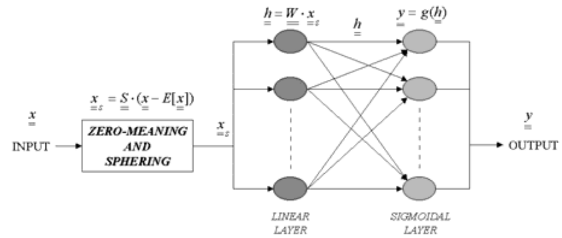


Figure 1- Architecture of the Infomax Neural Network

5.2 ICA-NN scheme based on contrast functions

The Infomax NN described in the previous Section has some limitations, both on the kind of source signals pdf and in the computational load. In this Section we will describe a different NN scheme to extract ICs that is most suitable to solve our problem. The proposed NN is also useful to cope with time-varying mixtures [Koivunen V., 2001].

The goal of ICA is to make a transform into a signal space in which the signals are statistically independent. Sometimes independence can be attained, especially in blind source separation in which the original signals are linear mixtures of independent source components and the goal of ICA is to invert the unknown mixing operation. Even when independence is not possible, the ICA transformation produces useful component signals that are non-Gaussian. The ICA allows us to approximately take into account all higher-order correlations and make the signals truly independent. Higher order statistics are needed to determine ICA expansion. In the framework of NNs, the ICA structure is that of a linear network that after learning is of the purely feed-forward type. However, during learning non-linearity must be used for separating sources. We assume here that we have a set of noisy linear mixtures representing the observed signal. By denoting with $x_k = [x_k(1), \ldots, x_k(M)]^T$ the M-dimensional kth data vector corresponding to the measurements carried out at discrete point, we can write the ICA signal model in the vector form:

$$\underline{x}_k = \underline{\underline{A}}\underline{s}_k + \underline{n}_k. \quad (19)$$

Here $\underline{s}_k$ is the source vector consisting of the independent signal components (sources), $\underline{s}_k$(i), i=1, N, $\underline{\underline{A}}$ = [$\underline{a}$(1), …, $\underline{a}$(N)] is a constant MxN “mixing matrix” whose columns $\underline{a}$(i) are the basis vectors of ICA, and $\underline{n}_k$ denotes possible corrupting noise, often omitted, because it is not possible to distinguish noise from source signals. The source separation aim is to determine $\underline{s}_k$, knowing only $\underline{x}_k$. Several assumptions must be made in ICA, in particular, only one of the source signals is allowed to have a Gaussian marginal distribution. Typically, the basis vectors $\underline{a}$(i) are normalized to unit length and arranged according to the powers E [$\underline{s}_k$ (i) ^{2}] in a similar way as in standard PCA. In PCA, the data model has the same form, but the coefficient $\underline{s}_k(i)$ are required to have sequentially maximal powers (variances), and the basis vectors $\underline{a}(i)$ are constrained to be mutually orthonormal. Usually, the basis vectors of ICA are not mutually orthogonal, in order to better characterize the data. The ICA allows to determine a sparse encoding of the input vector, where histograms show a high probability of a large response as well as of no response at all. The code increases first-order redundancy (histograms) by decreasing higher-order redundancy. This redundancy transformation can be described in terms of kurtosis, that is defined by (E[.] denotes expectation):

$$k[s(i)^4] = E[s(i)^4]-3[E[s(i)^2]]^2. \qquad (20)$$

The separation capability of various algorithms depends on the kurtosis [Ref, Kar]. It is possible to realize the estimation procedure by using a feed-forward scheme. The inputs of the NN are the M components of the vector $\underline{x}$. In the hidden layer, we have N nodes. The first layer of weights carry out a MxN whitening (and compression) of the input vector, After this, the sources are separated by means of an orthonormal matrix ($\underline{\underline{W}}^T\underline{\underline{W}} = \underline{\underline{I}}_N$) that the NN should learn. The ICA network, firstly proposed in [Karhunen J., 1997] is shown in Figure 2. Non-linearity (i.e., hyperbolic tangent function) must be used in learning the separating matrix. The learning algorithm here used is described in [Karhunen J., 1997] and can be summarized as follows: whitening of the original data $\underline{x}$ by $\underline{v} = \underline{\underline{D}}^{-1/2}\,\underline{\underline{E}}^T\,\underline{x}$, where $\underline{\underline{E}}$ is the matrix of the eigenvectors of $\underline{x}$ and $\underline{\underline{D}}$ is the diagonal matrix of eigenvalues that produces a starting point for an iterative process that finds vector $\underline{\underline{W}}$. The learning rule is:

$$\underline{\underline{W}}(k+1) = E[\underline{v}\, g(\underline{\underline{W}}(k)^T\underline{v}) - g(\underline{\underline{W}}(k)^T\underline{v})\,\underline{\underline{W}}(k)], \qquad (21)$$

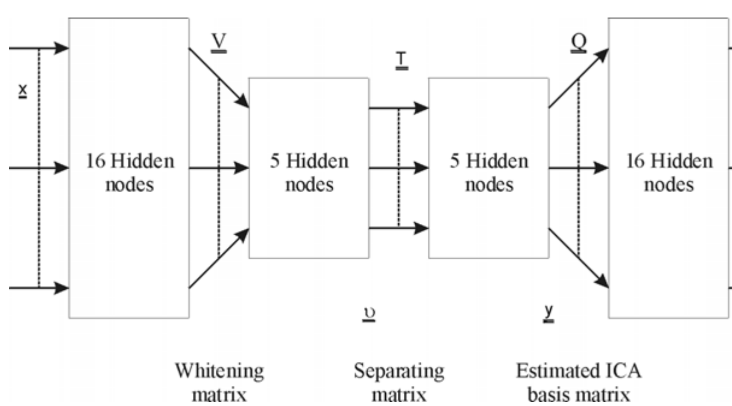
where g(.) is the hyperbolic tangent. After finding W, the IC's can be found by linear combination $\underline{y} = \underline{\underline{W}}^T\underline{v}$ and the mixing matrix $\underline{\underline{A}}$ by $\underline{\underline{A}} = \underline{\underline{E}}\,\underline{\underline{D}}^{1/2}\,\underline{\underline{W}}$.

The use of ICA network allows us to determine the ICA separating matrix.

6. Experimental EMG data processing results

The ICA-NN scheme proposed in the previous Section has been used to extract ICs from sEMG recordings. In what follows, we will report some results that have been achieved in this study. The following Table reports the correspondence between the placements of sEMG electrodes and the related muscles.

Figure 3 reports an example of the signal acquired during about 2 s of exercise (corresponding to pointing the monitor of a computer with alternatively the right and the left hand). Figure 4 reports the time-course of the 6th ICs, that appears to be mostly correlated with the 4th sEMG sensor.



$\underline{y} = \underline{\underline{W}}^T\,\underline{\underline{V}}\,\underline{x}$
Components of $\underline{y}$ as independent as possible

Weight matrix $\underline{\underline{Q}}$ minimizes the MSE error
$E\{\|\underline{x} - \underline{\underline{Q}}\,\underline{y}\|^2\}$

$\underline{\underline{W}}^T$ Orthonormal ($\underline{\underline{W}}^T\underline{\underline{W}} = \underline{\underline{I}}$) separation matrix that the network should learn

Fig. 2- The Neural Network feed-forward scheme for computing ICA.

1	SPec Superior Pectoralis	9	DBic Distal Bicep
2	IPec Inferior Pectoralis	10	PTri Proximal Tricep
3	LPec Lateral Pectoralis	11	DTri Distal Tricep
4	LDel Lateral Deltoid	12	PWEx Proximal Wrist Extensors
5	ADel Anterior Deltoid	13	DWEx Distal Wrist Extensors
6	MTrp Medial Trapezius	14	PWFl Proximal Wrist Flexors
7	LTrp Lateral Trapezius	15	DWFl Distal Wrist Flexors
8	PBic Proximal Bicep	16	APB Abductor Pollicus Brevis

Table 1: Correspondence between the electrode locations and the investigated muscles

Each ICs consists of a temporally independent waveform and a spatial distribution over the electrodes. The spatial distributions of the electrodes is shown on a cartoon body. The diagram has been obtained by making use of the MATLAB Toolbox for Electrophysiological Data Analysis, Version 3.2 (S. Makeig, et al, available online, http://www.cnl.salk.edu/scott/ica.html).

The electrodes are positioned according to Table 1. The colouring of each electrodes is proportional to the particular IC contributes to the electrode's raw recording. In the example, it is shown that the 6^{th} ICs mostly contributes to the 4^{th} electrode reading. Note the unmixing of the related components, basically activating just one electrode. Figures 6 to 8 reports the same signals for the 16^{th} electrode and the 16^{th} ICs. In this case, the 16^{th} component mainly activates the same electrode.

Measuring the ICs of sEMG will provide a more reliable and robust measure of motor performance than interpreting the activity of each individual muscles in isolation [Jung T.P., 2001].

There are practical advantages of separating the sEMG signals into temporally ICs, namely, the ICs are less susceptible to changes in position of the electrodes, and therefore more suitable for serially monitoring performance; the ICs are, in addition, more likely to correspond to brain activations [Jung T.P., 2001], by looking for common cortical influences in the muscle activity.

As previously mentioned, the experiment described in the present Section have been carried out by using a Neural Network scheme to implement ICA. It is, of course, possible, to use different techniques to implement ICA, however, it could be demonstrated that the use of a NN approach is equivalent to other approaches, like maximum likelihood estimation. The NN scheme is most suitable to achieve hardware implementation.

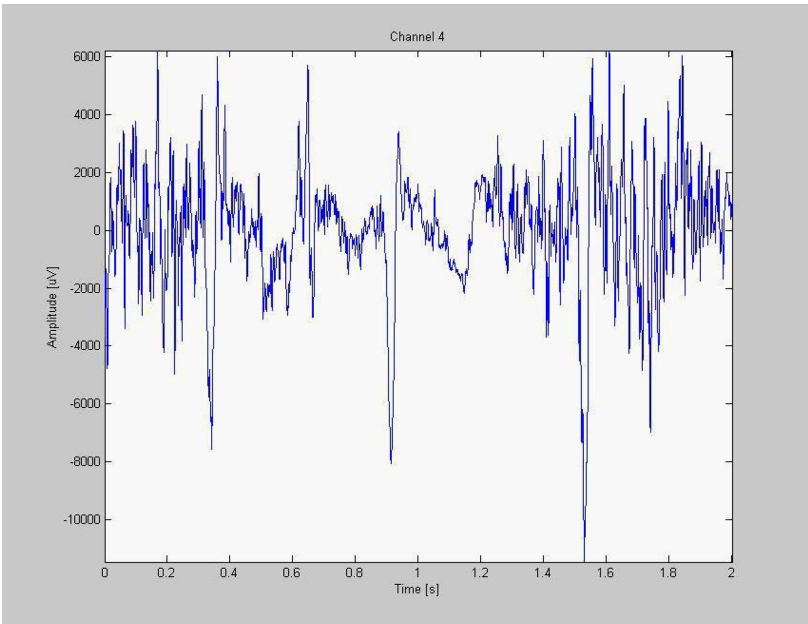


Figure 3: Raw EMG recording from the 4th electrode

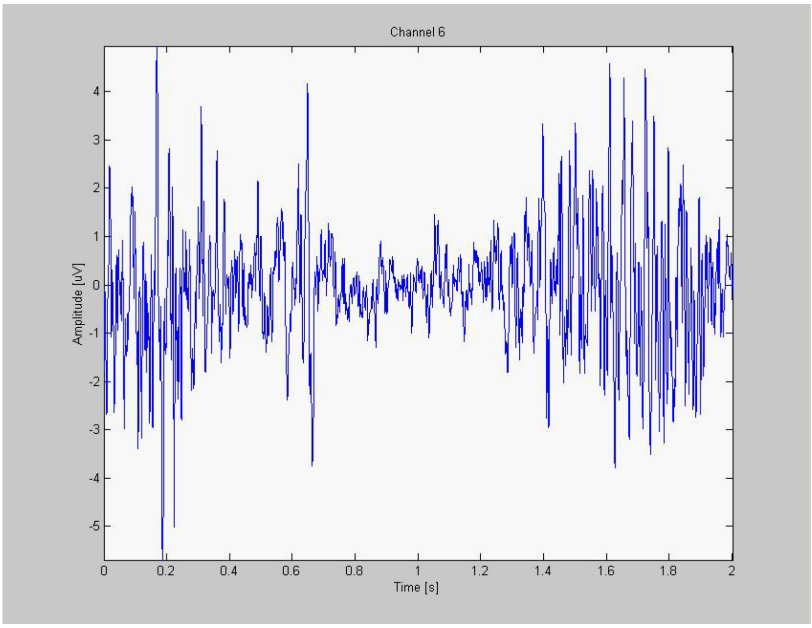


Figure 4: Time-course of the 6th extracted ICs

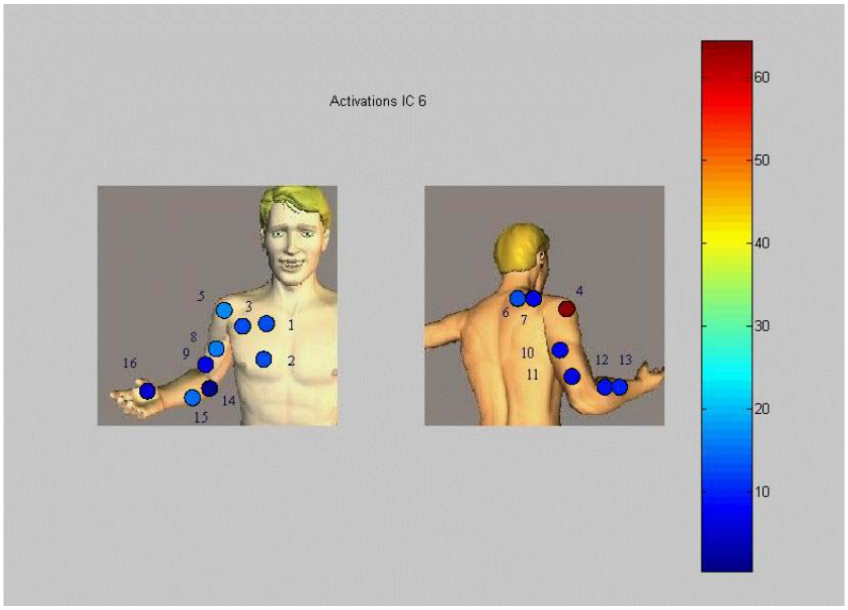


Figure 5: Spatial distribution of the activations corresponding to the 6^{th} ICs

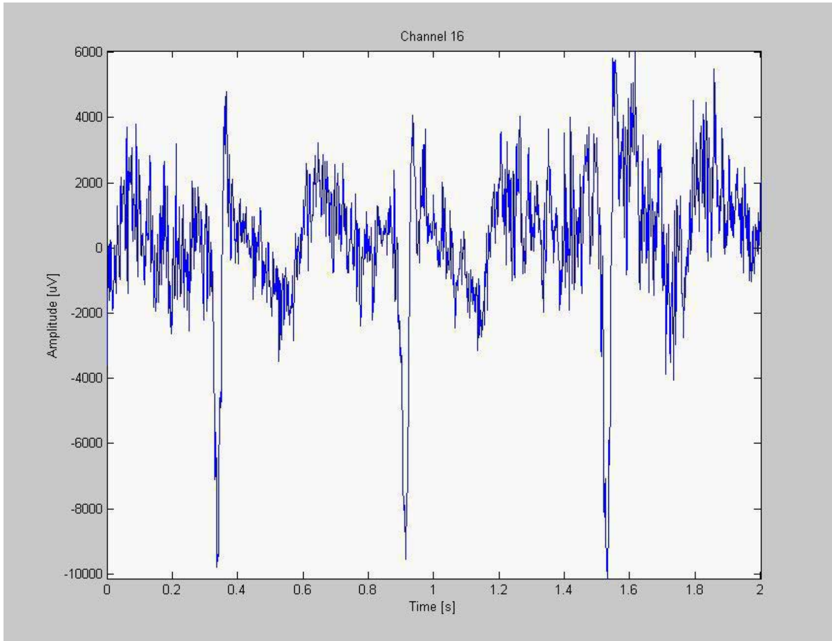


Figure 6: Raw EMG recording from the 16th electrode

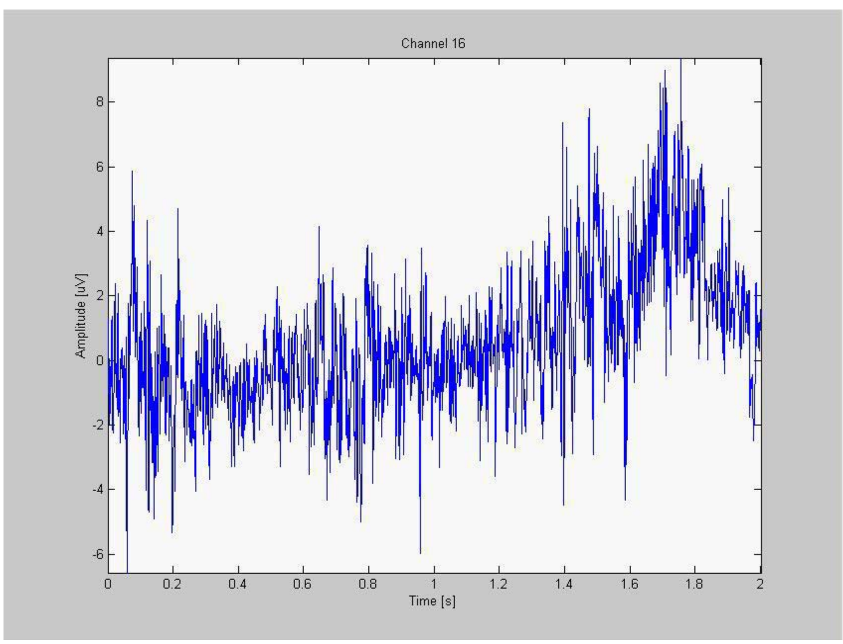


Figure 7: Time-course of the 16th extracted ICs

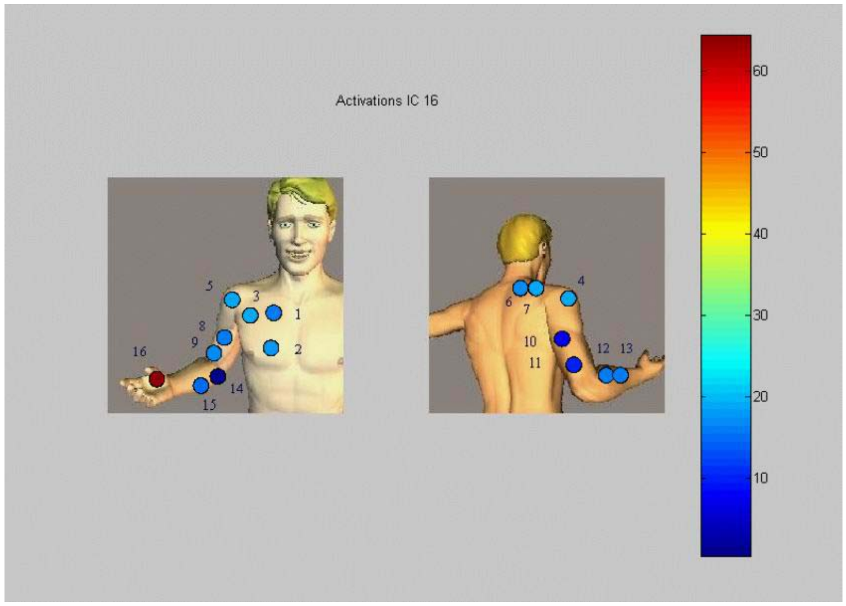


Figure 8: Spatial distribution of the activations corresponding to the 16th ICs

7. Treatment of non-stationarity

The extraction of ICs is based on the assumption of stationarity among different trials of the same experiment. In the practice, for such sEMG data, this is a hardly acceptable assumption. We would like now to propose a time-frequency approach to the analysis of sEMG data (or their ICs counterparts) that allows to cope with signal non-stationarity. The sEMG is indeed non-stationary as its statistical properties change over time. The MUAPs (Motor Unit Action Potentials) are transients that exist for a short period of time: for that reason, time-frequency methods are useful to characterize the localized frequency content of each MUAP. The use of a time-frequency representation also allows, in principle, to detect the onset of sub-movements, according to what we explained in the previous Sections. We have carried out the wavelet analysis in both the time domain of sEMG and of the ICs, in order to show that this kind of analysis should be carried out on the original space (the IC space is generated by already making a stationarity assumption).

The wavelet transform also guarantees to possibility of not specifying in advance the key signal features and the optimal basis functions needed to project the signal in order to highlight the features. An orthogonal wavelet transform is characterized by two functions:

1) the scaling function,

$$\phi(x) = \sqrt{2}\sum_{k\in Z} h(k)\phi(2x-k) \tag{22}$$

and 2) its associated wavelet:

$$\psi(x) = \sqrt{2}\sum_{k\in Z} g(k)\phi(2x-k) \tag{23}$$

where g(k) is a suitable weighting sequence (function).

The sequence h (k) is the so-called refinement filter. The wavelet basis functions are constructed by dyadic dilation (index j) and translation (index k) of the mother wavelet:

$$\psi_{jk} = 2^{\cdot j/2}\psi(x/2^{-j}-k) \tag{24}$$

The sequences h and g can be selected such that $\{\psi_{jk}\}_{(jk)\in Z^2}$ constitutes an orthonormal basis of L2, the space of finite energy functions. This orthogonality permits the wavelet coefficients $d_j(k) = \langle f, \psi_{jk}\rangle$ and the approximation coefficients $c_j(k) = \langle f, \phi_{jk}\rangle$ of any function f(x) to be obtained by inner product with the corresponding basis functions. In practice, the decomposition is only carried out over a finite number of scales J. The wavelet transform with a depth J is then given by:

$$f(x) = \sum_{j=1}^{J}\sum_{k\in Z} d_j(k)\psi_{jk} + \sum_{k\in Z} c_J(k)\phi_{Jk}\ . \tag{25}$$

In the present study, we shall use the WT in order to derive a set of features that can reveal singularity of the signal (corresponding to the onset of activity of single muscles) and to detect the precursors of the non-stationarity. A set of features derived from the inspection of the scale-dilation plane have been used as input vector of an auto-associative NN that is able to alarm the user about modification of the energy content of the spectrum. The features are extracted by considering the correspondence between singularities of a function and local maxima of its wavelet transform. A singularity corresponds to pairs of modulus maxima across several scales. Feature extraction is accomplished by the computation of the singularity degree (peakiness), i.e., the local Lipschitz regularity, which is estimated from the wavelet coefficients decay [Mallat S., 1992, Arkidis N.S., 2002].

Figures 9 and 10 reports the amplitude sEMG signal for channel 4th, and the wavelet transform obtained by using Daubechies 1 and 4 mother wavelet. The modulus maxima plots have been drawn and a thresholding operator is used in order to reduce the number of effective wavelet coefficients needed to represent the original functions. Once the features have been extracted by inspecting the modulus maxima plot, we can use the corresponding nonzero coefficients in order to predict the raising of non stationarity. A MLP NN with an input layer of corresponding size acts as a bottleneck network (the output size is the same of the input one, while the hidden layer size is considerably reduced). The NN fed by the wavelet coefficients computes the estimation of the corresponding wavelet coefficients at the output: a reconstruction error is computed. If the error overcomes a prescribed threshold level, the non-stationarity signal is activated and the following trials are used to compute a novel matrix (ICs) weights. The use of a MLP-NN is not obliged to ensure accuracy or success in the reconstruction; for example, a different compression scheme could be used, like the Singular Value Decomposition. The bottleneck layer is, in principle, able to work as principal component extractor, but the idea here is to build a compressed representation which is deliberately redundant. The reconstruction error could be sub-optimal with respect to different schemes, but optimality comes at the expenses of quite low fault tolerance. Finally, the MLP NN can be implemented easily in a FPGA hardware chip. A typical case of non-stationarity is the onset of fatigue. The Figure 11 describes how the activation intervals [Micera S., 2001] of the muscles during the exercise cycle are determined starting from the ICs.

The standard approach to determine on-off activation patterns is to process each epoch by means of a double threshold statistical detector [Bonato P., 1998, Balestra G., 2001] to obtain the muscle detection intervals. We have compared the results achieved by our method with the one described and we have found an improvement of about 20% in the performance.

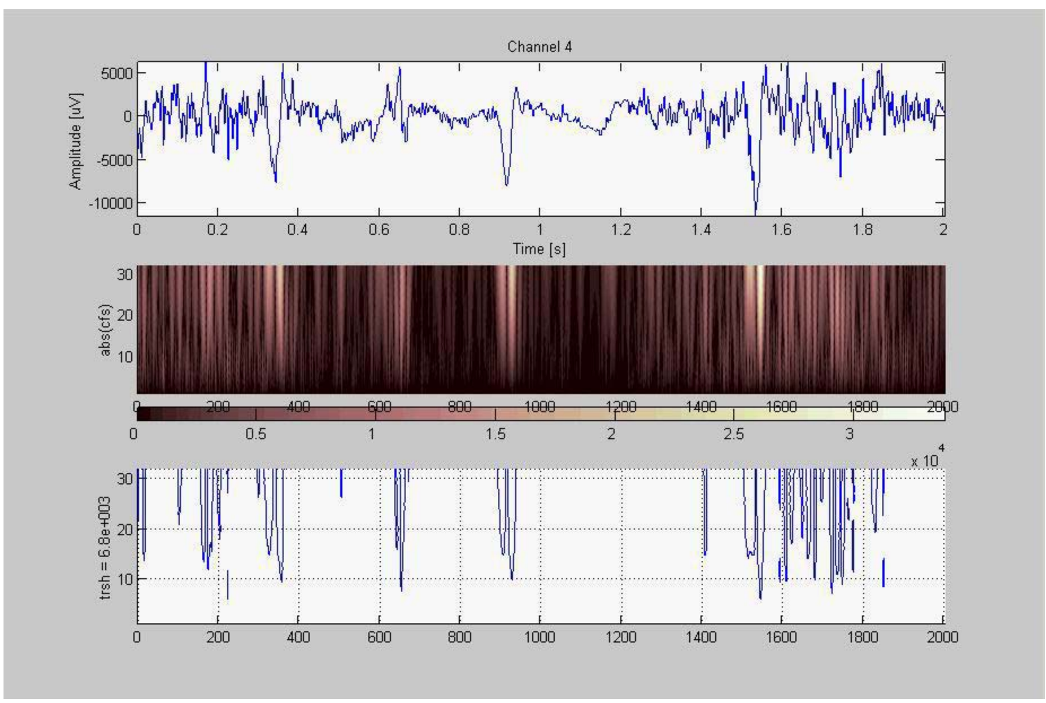


Figure 9: The wavelet transform of the 4th sEMG channel (mother wavelet, Daubechies 1): the raw data recording (top), the plot of the absolute values of the WT coefficients (middle) and the modulus maxima extracted (bottom). A thresholding is applied to suppress WM that are not of interest. White colour corresponds to high value of the coefficients. If one uses a wavelet with one vanishing moment, then the bottom plot corresponds to the maxima of the smoothed first-order derivative of the function.

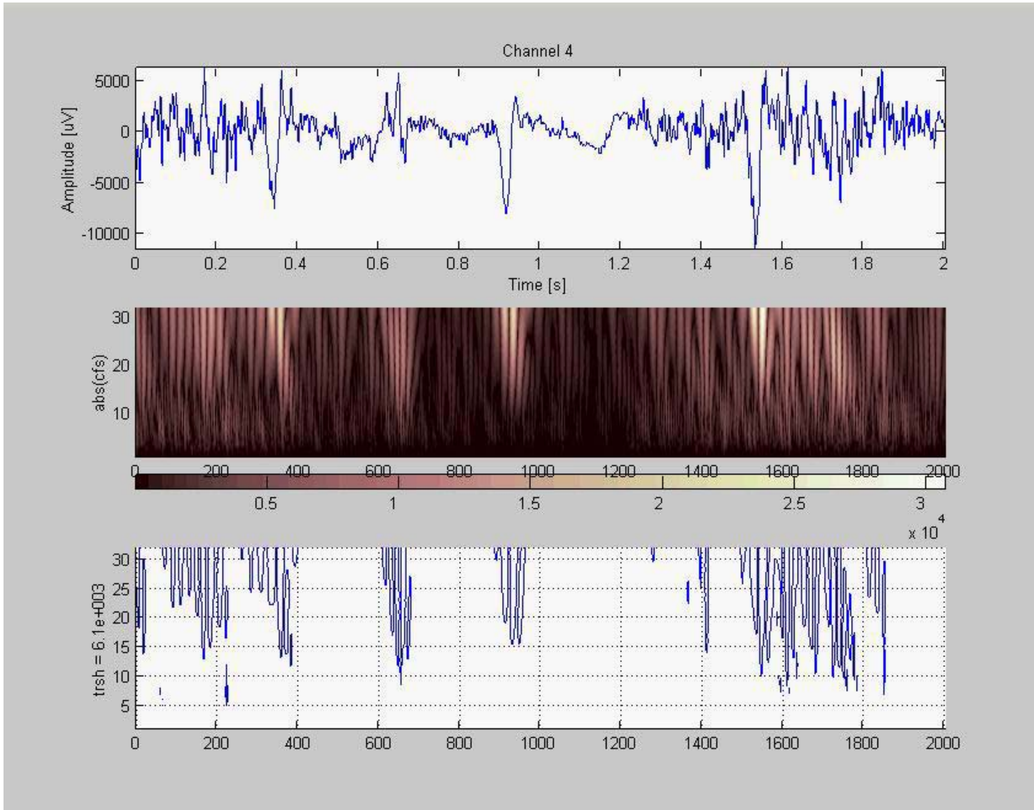


Figure 10: The wavelet transform of the 4th sEMG channel (mother wavelet, Daubechies 4): the raw data recording (top), the plot of the absolute values of the WT coefficients (middle) and the modulus maxima extracted (bottom). White colour corresponds to high value of the coefficients. A wavelet function with 4 vanishing moments is used.

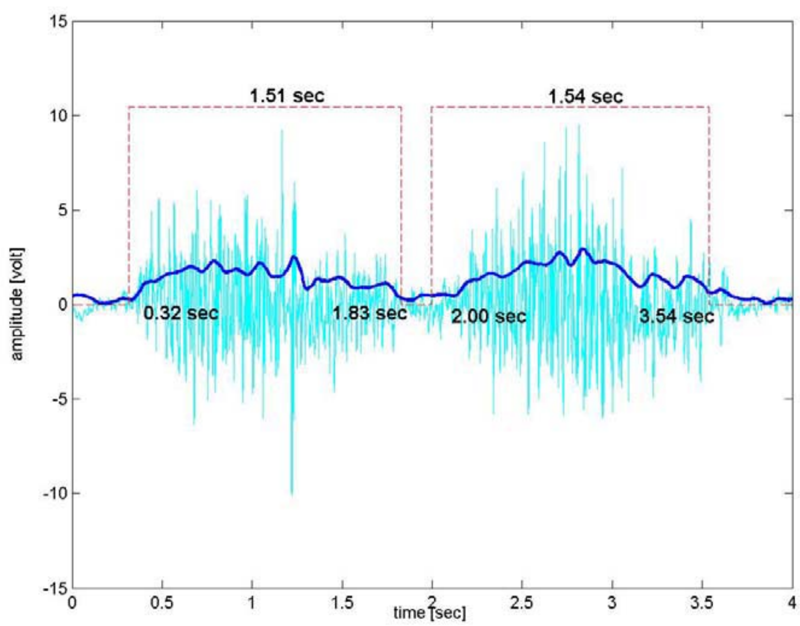


Figure 11: The determination of the activation intervals (the wavelet envelope is used).

8. Conclusion

The paper proposed the use of some NNs to process experimental electrical data derived from non-invasive sEMG experiments. The original (raw) data have been analysed by a neural IC processor aiming to obtain signals that can be easily correlated to cortical activity. The assumption of stationarity is then relaxed in order to cope with time-varying mixing systems, more adherent to the biophysical problem at hand. An auto-associative NN exploits the features obtained by wavelet transforming the raw data for making a quick and efficient prediction of non-stationarity. The results we have shown can be considered just as preliminary to solve the difficult problem.

Acknowledgment

The work described in this paper has been carried out within a research cooperation co-ordinated by Prof. Martin J. McKeown, Duke University, Durham NC, USA. We would like to explicitly thank him and his group for proposing us to work on this very cross-disciplinary subject. Most of the computation described in this work have been carried out by Dr. Greco and Dr. Costantino, two post-graduate students that we gratefully acknowledge.

REFERENCES

Akay M., *Time-frequency and Wavelets in Biomedical Signal Processing*, Piscataway, NJ, IEEE Press, 1997.

Balestra G., Frassinelli S., Knaflitz M. Molinari F., *Time-Frequency Analysis of Surface Myoelectric Signals Durino Athletic Movement*, IEEE Engineering in Medicine and Biology, Special Issue on Monitoring Muscles in Motion, Vol. 20, No.6, Nov/Dec 2001.

Bonato P., D'Alessio T., and Knaflitz M., *A statistical method for the measurement of muscle activation intervals from surface myoelectric signal during gait*, IEEE Trans. Biomed. Eng., Vol.45, pp.287-299,1998.

Arkidis N.S., Abel E.W., Faster A., *Interscale Wavelet Maximum: A Fine to Coarse Algorithm for Wavelet Analysis of the EMG Interference Pattern*, IEEE Trans. On Biom. Eng., Vol. 49, No. 4, pp.337-344, April, 2002.

Bell A.J. and Sejnowski T.J., *An Information-Maximization Approach to Blind Separation and Blind Deconvolution,* Neural Computation*, Vol.*7, pp.1129-1159, 1995.

Bell A.J. and Sejnowski T.J., *Learning the higher-order structure of a natural sound*, Network Computation in Neural Systems, Vol.7, 1996.

Harris C.M., *On the optimal control of behaviour: A stochastic perspective*, J. Neurosci. Meth, Vol. 83, pp. 73-88, 1998.

IEEE Engineering in Medicine and Biology, Special Issue on Monitoring Muscles in Motion, Vol. 20, No.6, Nov/Dec 2001.

Jung T.P., Makeig S., Lee T.W., McKeown M.J., Brown G., Bell A.J. and Sejnowski T.J, *Independent Component Analysis of Biomedical Signals* – The 2nd Int'I Workshop on Independent Component Analysis and Signal Separation, 2000.

Jung T.P., Makeig S., McKeown M.J., Bell A.J., Lee T.W., Sejnowski T.J., *Imaging brain dynamics using independent component analysis*. P IEEE. 89(7): 1107-22, 2001

Karhunen J., Oja E., et al, *A class of neural networks for independent component analysis*, IEEE Trans. on Neural Networks, Vol. 8, N. 3, pp.486-504, 1997.

Koivunen V., Enescu M., Oja E., *Adaptive algorithm for blind separation from noisy time-varying mixtures*, Neural Computation, Vol.13, pp.2339-2357, 2001.

Lee T.W., *Independent Component Analysis: Theory and Applications*, Kluwer Academic Publishers, Boston, 1998.

Lee T.W., Girolami M., and Sejnowski T.J., *Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources*, Neural Computation, Vol. 11, N.2, pp. 417-441, 1999.

Linsker R., *An Application of the Principle of Maximum Information Preserving to Linear Systems*, In Advances in Neural Information Processing Systems 1.

Mallat S. and Hwang W.L., *Singularity detection and processing with wavelets*, IEEE Trans. Inform. Theory, Vol.38, pp. 617-643, 1992.

McKeown M.J., *Phasic and Tonic Coupling between EEG & EMG Revealed with ICA*, J. Clin. Neurophysiology, 2000a, in press

McKeown M.J., *Cortical activation related to arm movement combinations*, Muscle Nerve, Vol 9, Suppl. 9.4, 2000b.

McKeown M.J., Torpey D.C., Gehm W.C.. *Non-Invasive Monitoring of Functionally Distinct Muscle Activations during Swallowing Clinical Neurophysiology* 2002 (in press).

Micera S., Vannozzi G., Sabatini A.M., Dario P., *Improving Detection of Muscle Activation intervals*, IEEE Engineering in Medicine and Biology, Vol. 20, No. 6 Nov-Dec 2001 pp. 38-46.

Sejnowski T.J., *Making smooth moves*, Nature, Vol.394, Aug 1998.

Francesco Carlo Morabito was born and educated in Italy, where he worked in the University since 1989 after a three-year experience as a researcher in the electronic industry. Since December 2001, he serves as Dean of the Faculty of Engineering of the "Mediterranea" University of Reggio Calabria, Italy, where he teaches (full professor) Electrical Engineering and Neural Networks. He was a member of INNS since 1992 and served as Board of Governors member for the last three years (2000-2002). He served as Secretary of INNS in 2002 and he is now serving as Secretary. He chairs the Regional Italy SIG since 1997, taught and co-ordinated Tutorials at the IJCNN1999, 2000, 2001. Since 2002, he served as an action editor for Neural Networks. He also serves as both steering and scientific committee of Italian Society for Neural Networks (SIREN). He chaired the Eduardo Caianiello Scientific Prize. He is a senior Member of IEEE. He taught at the EPFL, Lausanne, Switzerland at the Ph.D. courses in 1999 and worked as a referee for both international research projects and Ph.D. thesis in the field of Neural Networks. He was general chair of AMSE-ISIS'97 and '98, and AEDEM 2001, member of the scientific committees of many international conferences in Europe, Asia, USA and Africa.

He was a visiting researcher at the Max-Planck Institut fuer Plasmaphysik, NET Team, Asdex-Upgrade, ITER Joint Team, Garching bei Muenchen, Germany for some years. He was scientific responsible for various national and international scientific research project.

Listed in Who's Who in the World, Who's Who in Science and Engineering; International Biographical Centre. He is the author of 1 book, 3 book chapters, and more than 150 scientific papers. He is the recipient of the 2002 "Henry Coanda" Prize from Rumanian Academy for his scientific activity in the field of Neural Networks.

Maurizio Campolo was born in Reggio Calabria, Italy. He has been working since 1993 as computer science researcher at the University Mediterranea of Reggio Calabria. His research interests are in the field of Neural Networks and Information Technology.

He is the author of more than 20 scientific papers and worked as a technician for some national research projects. He was member of the scientific committe of international conference SSCC98 in Africa and member of the organizing committe of AMSE-SIS97, ENDE98 and HBD2000 in Italy. Since 1993 he is the network administrator and webmaster of the Faculty of Engineering. He also worked with NeuroLab.