POLITECNICO DI TORINO
Repository ISTITUZIONALE

Blind Source Separation. Application to Biomedical Signals

Original
Blind Source Separation. Application to Biomedical Signals / Mesin, Luca; Holobar, A.; Merletti, Roberto - In: Advanced Methods of Biomedical Signal Processing / Cerutti S., Marchesi C.. - STAMPA. - [s.l] : Wiley-IEEE Press, 2011. - ISBN 9780470422144. - pp. 379-410

Availability:
This version is available at: 11583/2440745 since:

Publisher:
Wiley-IEEE Press

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

15 May 2024

Blind Source Separation: Application to biomedical signals

Luca Mesin[*], *Aleš Holobar*[*], *Roberto Merletti*[*]

1. INTRODUCTION

Blind Source Separation (BSS) is a prominent problem in signal processing. In the past few decades, it was applied to many fields, in which separation of compound signals, simultaneously observed by different sensors, is of interest. The problem can be considered as built-up of three physical elements: sources (also called transmitters), sensors (also called receivers) and communication channels which reflect the properties of the physical medium propagating the signals form the sources to the sensors. The signals detected by the sensors are commonly referred to as observations and are assumed to be algebraic combinations of the unknown sources signals. BSS approach assumes limited a priori information on the communication channels (linearity, memory properties…) and tries to reconstruct the source signals out of the detected signals only. Analysis of the communication channels is important mainly for selection of a proper processing technique. Namely, communication channels weight and/or filter the signals coming from the sources and, together with them, determine the temporal and spectral characteristics of the detected mixtures.

An example of physical medium propagating the sound is air. Physical properties of the air determine the weights of communication channels in *Speech separation* (or *cocktail-party*) problem. This problem deals with separation of different human voices or sounds from instruments, recorded by two or more microphones during simultaneous emission of two or more sources (Koutras et al., 2000, Anemüller and Gramß, 1999). Source separation problem for sonar is discrimination of the echoes from different simultaneously present targets. Radar requires the solution of problems equivalent to those of sonar. Communication systems working under water or in an orbit also face equivalent problems of source separation, but use signals with different spectral

[*] Laboratorio di Ingegneria del Sistema Neuromuscolare (LISiN), Dipartimento di Elettronica, Politecnico di Torino, Torino

features. The problem of separating a mixture of echoes from different targets is also important in Earth Sciences, e.g. in the study of different geological layers or in the search for water or oil reservoirs.

Source separation finds important applications also in life sciences; Electroencephalographic (EEG), Electrocardiographic (ECG), Electromyographic (EMG), Mechanomyographic (MMG) signals are all compound biomedical signals, generated by several tens (EMG, MMG) or even millions (EEG) of biophysical sources. Separation of biomedical signals augments the power of human body scanning techniques and plays an important role in understanding of complex processes in biomedical phenomena (Vigàrio, 1997, Vigàrio et al., 1998, Vigàrio et al., 1999, Makeig et al., 1996). This chapter is devoted to basic descriptions of frequently used source separation methods, with focus on the biomedical applications.

2. MATHEMATICAL MODELS OF THE MIXTURES

A mathematical model of the source-sensor communication, also called a mixing process, determines an analytical relation between the source signals and the observations. *Mixing* models can be classified as follows (Lacoume, 1999).

1. *Non-linear model*: it is the most general model and very difficult to study as the source signals do not satisfy the superimposition principle (i.e., their contributions combine non-linearly to form the observed mixtures).
2. *Post non-linear model*: the process consist of linear *mixing* and an instantaneous non linear mapping of the source signals.
3. *Linear model*: it is the most widely studied model and the only one considered in this chapter. In each observation, contributions from different sources are linearly combined, i.e., superimposed to each other.

Linear mixing model can further be divided in two subgroups:

- *Convolutive mixing model*: the *mixing* process is a causal multidimensional convolution

$$\mathbf{x}(t) = \int \mathbf{A}(t-\tau)\, \mathbf{s}(\tau)\, d\tau \qquad [1]$$

  where $\mathbf{s}(t)$ are source signals from N sources, $\mathbf{x}(t)$ are observations detected by M sensors and $\mathbf{A}(t)$ is a *mixing matrix* comprising impulse responses of all the communication channels that

relate the source signals **s**(t) to the observed signals **x**(t). Convolutive mixing model is typically assumed to be causal, with memory of the source signals received in the past;

- *Instantaneous mixing model*: the signals detected in a time instant are obtained as a linear combinations of the source signals at the same instant:

$$\mathbf{x}(t) = \mathbf{A}\ \mathbf{s}(t) \qquad [2]$$

An instantaneous mixing model has no knowledge of the source samples received in the past.

In numerical implementation, **s**(*t*) is a matrix of sampled source signals (with the *T* samples of the signal from the rth source in the rth row), and **x**(*t*) is a matrix of sampled observations, with observations from different sensors in different rows. Without loss of generality, the observations **x**(*t*) are also assumed to be zero-mean.

The linear BSS problem has two types of ambiguities. Firstly, it is clear from [1] and [2] that amplitude scaling of the sources can be compensated by an inverse scaling of the corresponding elements of matrix **A**. Thus, with no a priori knowledge on the mixing matrix **A,** the power of individual source signals cannot be determined and is, by convention, set equal to 1. A second ambiguity lies in the order in which the source signals are determined.

Now, let us assume that the signals **s**(*t*) are emitted from *N* different sources, while the observations **x**(t) are detected by the *M* different sensors, where $M \geq N$. Then, in order to reconstruct the source signals, we must first estimate the mixing matrix **A**, invert it, and apply its inverse to the observed signals **x**(*t*). Thus, the unknowns of the BSS problem comprise both the elements of **A** and the source signals **s**(*t*). In the case of the discretised instantaneous model [2] with *T* samples long source signals, we must estimate $M \times N$ entries of **A** and $N \times T$ samples of the source signals, given just $M \times T$ samples of observations **x**(*t*). The number of unknowns to be determined is usually greater than the number of equations imposed by model [1] or [2] (even when $M \geq N$) and further a priori conditions on the source signals or/and the mixing matrix **A** are required to face the problem of source separation. Most of BSS methods do not use any information about the *mixing* matrix **A**. Instead, they only rely on additional information about the sources. The latter are usually considered to be uncorrelated or statistically independent. Although somehow contra-intuitive, these assumptions are often sufficient to estimate the source signals, except for the ambiguities on their amplitudes and order (as stated above).

In order to comply with practice, a random noise is usually added to the models [1] and [2]. Such a noise can be either additive or multiplicative. Typically, the noise is further assumed to be zero-

mean, temporarily and spatially white random process. Temporal whiteness implies the independence of noise samples belonging to the time series of each individual observation, whereas spatial whiteness refers to independence of samples of noise between different observations at the same time instant. Frequently, the noise is also assumed to be independent of the source signals.

3. PROCESSING TECHNIQUES

Practically all source separation techniques are based on maximisation of the distance between the estimated source signals. The definition of the distance depends on the selected a priori assumptions on the sources and generates classification of different BSS approaches. In the sequel, we will briefly describe only some of those BSS classes that found their way into the field of biomedical signal processing. Interested reader is referred to Hyvarinen et al., 2001, for more thorough and complete overview of BSS approaches.

One of the best known signal decomposition techniques is the Principal Component Analysis (PCA), also known as Karhunen-Loeve or Hotelling transform. Strictly speaking, PCA does not belong to the BSS family, as it does not truly reconstruct the original source signals. Nonetheless, it is very popular decomposition techniques and is used as a pre-processing step of numerous BSS approaches. PCA builds of on correlation of observed signals and decomposes the observations into uncorrelated signal components. If the source signals are Gaussian, uncorrelatedness also implies independence, and the signal components obtained by PCA are also statistically independent. A useful property of PCA is that it preserves the power of observations, removes any linear dependencies between the reconstructed signal components and reconstructs the signal components with maximum possible energies (under the constraint of power preservation and uncorrelatedness of the signal components). Thus, PCA is frequently used for a lossless data compression (see Section 3.1 for details).

The second large class of signal decomposition techniques is the so called Independent Component Analysis (ICA). ICA belongs to the family of BSS and imposes statistical independence of sources, meaning that all the samples of the source signals are assumed to be independent identically distributed (i.i.d.) random variables. ICA preserves the information contained in the observations and, at the same time, minimizes the mutual information of estimated source samples (mutual information is the information that the samples of the source signals have on each others). Thus, also ICA is useful in data compression, usually allowing higher compression rates than PCA.

Specific optimisation techniques used to maximise the distance between the independent sources determines further classification of the ICA methods:

1. *Algebraic* methods: matrix calculus is used to estimate the *mixing* matrix **A**;
2. *Neural networks* based methods: neural networks perform recursive estimation of weights, which define linear combinations of the mixtures; these combinations are the estimates of the sources.

In the next subsection, PCA and ICA are discussed in more details. In particular, examples of algebraic and neural network-based source separation methods are described, along with some indications about the most typical assumptions about the statistical independence of sources. References for further reading are also provided.

3.1 PCA and ICA: possible choices of distance between source signals

Assume a simple mixing model with N source signals $\mathbf{s}(t)$ and M observations $\mathbf{x}(t)$:

$$\mathbf{x}(t) = \mathbf{A}\,\mathbf{s}(t) + \mathbf{n}(t) \qquad [3]$$

where $\mathbf{n}(t)$ is a zero-mean additive Gaussian noise.

PCA is mathematical method, which determines the amount of redundancy in the observations $\mathbf{x}(t)$ and estimates a linear transformation **P**, which reduces this redundancy to a minimum. **P** is further assumed to have a unit norm, so that the total power of the observations $\mathbf{x}(t)$ is preserved. Strictly speaking, PCA does not assume any mixing model. Redundancy of information in $\mathbf{x}(t)$ is simply measured by the cross-correlation between the different observations. Therefore, although PCA can be interpreted as signal decomposition technique, the estimated principal components $\mathbf{y}(t)= \mathbf{P}\mathbf{x}(t)$ differ significantly from the original sources $\mathbf{s}(t)$ (see Subsection 3.1.1). ICA, on the other hand, employs much stronger assumption on statistical independence of sources, requires a-priori knowledge about the mixing model, and allows reconstruction of original sources $\mathbf{s}(t)$.

.

ICA problem was first proposed by Jutten, 1987, and Hérault and Jutten, 199. The neural, iterative approach used by Hérault and Jutten underlines the similarities of ICA with PCA and is, for historical reasons, discussed in the next subsection. Independently from Hérault and Jutten, Bar-Ness proposed an equivalent method (Bar-Ness, 1982). Giannakis et al., 1989, addressed the issue of identificability of ICA, using cumulants of third order. Higher-order statistics were used by Laucoume and Ruiz, 1989, and by Gaeta and Laucoume, 1990, which introduced *maximum*

likelihood method for the estimation of the *mixing* matrix. The algebraic method introduced by Cardoso, 1989, and Souloumiac and Cardoso, 1991, is based on the properties of the fourth order cumulants. Inouye and Matsui, 1989, proposed an innovative solution to the problem of separation of two variables. At the same time, Comon, 1994, proposed a method for separation of N sources, while Fety, 1988, was the first to study source separation for a dynamic problem.

3.1.1 Principal components analysis (PCA)

The decomposition in principal components provides the representation of a set of signals $\mathbf{x}(t)$ as linear combinations of orthonormal components $\mathbf{y}(t)$ (called *principal components*) to be determined. For consistency reasons, signals $\mathbf{x}(t)$ are called observations here, even though PCA does not require $\mathbf{x}(t)$ to be a mixture of any sources. Principal components $\mathbf{y}(t)$ are directly related to the observations $\mathbf{x}(t)$ and are chosen to minimise the Mean Square Errors (MSE)

$$\frac{1}{T}\sum_{k=1}^{m}\int_{0}^{T} |x_k(t)-c_{ki}y_i(t)|^2 \mathrm{d}t \qquad [4]$$

where T is the observation interval, $c_{ki}y_i(t)$ is the ith approximation of the kth observation by the i-th principal component $y_i(t)$. An iterative method to obtain the principal components results directly from their definition [4] and is based on the following iterative steps.

1. Compute the first principal component minimising the sum of the m mean square errors in equation [4].
2. Compute the second principal component under the constraint of being orthogonal to the previous one(s).
3. Repeat step 2 until M principal components are reconstructed.

Exploiting the orthonormal property of the principal components, it is possible to prove that the contribution of the kth principal component to the power of the observations $\mathbf{x}(t)$ is

$$P_k = \sum_{i=1}^{n} c_{ik}^2 \qquad [5]$$

Application of abovementioned PCA procedure to a pair of surface EMG (sEMG) signals is illustrated in Figure 1. sEMG signals were recorded at the surface of the skin, above the biceps brachii muscle. Pick-up electrodes (i.e. sensors) were positioned close to each other in a linear array structure (interelectrode distance of 5 mm) and acquired electrical signals form approximately the same group of muscle fibres. As a result, both sEMG signals, x_1 and x_2 are highly correlated, as demonstrated by joint vector space representation in panel c). PCA finds the directions of maximal variance (so called, principal directions) and projects the observations, x_1 and x_2 on these directions

to reconstruct the principal components y_1 and y_2. In Figure 1.c), the first principal direction is represented by a black dashed line, the second principal direction by a black dotted line. Reconstructed principal components (i.e. projections to the principal directions) are depicted in panels 1.d) and 1.e). Note that, due to high level of redundancy in observations x_1 and x_2, more than 90 % of total power is stored in the first principal component y_1.

Figure 1 about here

According to equation [4] and Figure 1, the first principal direction is a direction of maximum variance. This suggests a second PCA computation technique. Let $\mathbf{w}_1$ be the unit norm *weight vector* representing the first principal direction of observations $\mathbf{x}(t)$. By definition, the linear combination $\mathbf{w}_1^{\mathrm{T}}\mathbf{x}$ is the first principal component with the maximum variance. The weight vector $\mathbf{w}_1$ can then be obtained as

$$\mathbf{w}_1 = \underset{\|w\|=1}{\arg\max}\, E\left[\left(\mathbf{w}^T\mathbf{x}\right)^2\right]. \qquad [6]$$

Afterwards, the projection of $\mathbf{x}$ on the subspace spanned by already reconstructed principle directions is calculated as $\mathbf{x} - \sum_{i=1}^{k-1} (\mathbf{w}_i^T\mathbf{x})\mathbf{w}_i$, and the k-th ($k \geq 2$) principal direction is calculated as

$$\mathbf{w}_k = \underset{\|w\|=1}{\arg\max}\, E\left[\left(\mathbf{w}^T\left(\mathbf{x} - \sum_{i=1}^{k-1}(\mathbf{w}_i^T\mathbf{x})\mathbf{w}_i\right)\right)^2\right]. \qquad [7]$$

This procedure is then repeated for all the remaining principal directions. Strictly speaking, principal directions reveal the directions of the maximum variance of M-dimensional random process. In the case of deterministic signals, we say that the principal directions reveal the directions of the maximum power in observations $\mathbf{x}(t)$.

Principal components, as introduced so far, reveal their usefulness in data compression, but their connection to the problem of source separation is weak. In Section 3.2, we prove that principal components of the observations $\mathbf{x}(t)$ are associated to the sources $\mathbf{s}(t)$ by an unknown rotation matrix. Method for the estimation of this unknown rotation matrix is described in Section 4.2, where a biomedical application of a PCA-based BSS method is discussed.

3.1.2 Independent Component Analysis (ICA)

Now, assume the source signals $\mathbf{s}(t)$ in [3] are random processes. In ICA, source separation is achieved by additionally supposing the source signals statistically independent, instead of being just

uncorrelated (PCA). Different measures of independence can be introduced, giving rise to different ICA methods.

When the number of observations M is greater than the number of sources N, the source signals can be estimated by applying the *separation matrix* $\mathbf{Q}$ to observations $\mathbf{x}(t)$:

$$\mathbf{y}(t) = \mathbf{Q}\,\mathbf{x}(t) \qquad [8]$$

where $\mathbf{Q}$ is generally unknown. Neglecting the influence of noise, for $\mathbf{y}(t)$ to be equal to the original sources $\mathbf{s}(t)$, we should have $\mathbf{Q} = \mathbf{A}^{\#}$, where $^{\#}$ indicates matrix pseudoinverse (see the Appendix). As $\mathbf{A}$ is unknown, additional assumption of independence of the source signals is required. One of the most intuitive ways of realizing how the assumptions on statistical independence can be used to estimate the separation matrix $\mathbf{Q}$ is based on the *central limit theorem*. Central limit theorem guarantees the linear combination of independent non-Gaussian random variables has a distribution that is "closer" to a Gaussian than the distribution of any individual variable. This implies that the samples of the vector of observations $\mathbf{x}(t)$ are "more Gaussian" than the samples of the vector of sources $\mathbf{s}(t)$. Thus, the source separation can be based on minimisation of *Gaussianity* of reconstructed sources $\mathbf{y}(t)$. All that we need is a measure of (non)Gaussianity, which is used as an objective function by a given numerical optimization technique. Many different measures of Gaussianity have been proposed. Some of them are briefly summarized in the sequel.

1. *Kurtosis*: kurtosis of a zero-mean random variable v is defined as

$$\mathrm{K}(v) = E[v^4] - 3E[v^2]^2 \qquad [9]$$

where $E[]$ stands for mathematical expectation. For a Gaussian variable v, $E[v^4] = 3E[v^2]^2$ and Kurtosis of a Gaussian variable is 0. For most non-Gaussian distributions, kurtosis is non-zero (either positive or negative). Variables with positive kurtosis are called *supergaussian* (a typical example is Laplace distribution). They have a more spiky distribution, with heavy tails and more pronounced peak with respect to a Gaussian distribution. Variables with negative kurtosis are called *subgaussian*, and have distribution that is flatter than Gaussian. A typical example of subgaussian distribution is uniform distribution. Being based on the forth-order statistic, Kurtosis is very simple to compute, but is highly sensitive to outliers. Its value might be significantly influenced by a single sample with large value. Hence, it is not appropriate for separation of noisy measurements and measurements with severe signal artefacts.

2. *Negentropy*: given the covariance matrix of a multidimensional random variable, negentropy is defined as the difference between the entropy of the considered random variable and that of a Gaussian variable with the same covariance matrix. It vanishes for Gaussian distributed variables and is positive for all other distributions. From a theoretical point of view, Negentropy is the best estimator of Gaussianity (in the sense of minimal mean square error of the estimators), but has a high computational cost as it is based on estimation of *probability density function* of unknown random variables. For this reason, it is often approximated by *k*-th order statistics, where *k* is the order of approximation (Hyvarinen, 1998, Jones e Sibson, 1987).

3. *Mutual Information*: another method for source separation by ICA is associated to *Information Theory*. Mutual information between *M* random variables is defined as

$$I(y_1,...,y_m) = \sum_{i=1}^{m} H(y_i) - H(\mathbf{y}) \qquad [10]$$

where $\mathbf{y} = [y_1,...,y_m]$ is a M-dimensional random vector. Information entropy *H* of a discrete random vector **y** is defined as $H(\mathbf{y}) = \sum_i -\mathrm{P}(\mathbf{y} = \mathbf{a}_i)\log \mathrm{P}(\mathbf{y} = \mathbf{a}_i)$, where $\mathbf{a}_i$ are the possible values of **y**. For continuous random variable with probability density $f(\mathbf{y})$, entropy *H* is defined as $H(\mathbf{y}) = -\int_{-\infty}^{\infty} f(\mathbf{y})\log(f(\mathbf{y}))d\mathbf{y}$. Mutual information is always nonnegative, and equals zero only when variables $y_1,...,y_m$ are independent. It is possible to prove (Hyvarinen, 2000) that mutual information of non correlated variables with unitary variance is equivalent to *negentropy*, except for the sign, i.e. maximization of negentropy is equivalent to minimisation of mutual information.

Mutual information is also related to *Kullback-Leibler divergence* defined as (Hyvarinen, 1999)

$$\delta(f,g) = -\int_{-\infty}^{\infty} f(\mathbf{y})\log(f(\mathbf{y})/g(\mathbf{y}))d\mathbf{y}. \qquad [11]$$

which can be seen as a measure of a distance between probability density functions f and g. *Kullback-Leibler divergence* is always nonnegative and vanishes if and only if the probability densities *f* and *g* are equal. In ICA, *Kullback-Leibler divergence* measures the distance between the density $f(\mathbf{y})$ and the factorised density $g(\mathbf{y})=f_1(y_1)\, f_2(y_2)\ldots f_n(y_n)$, where $f_i(y_i)$ is the marginal density of variable y_i . Mutual information and Kullback-Leibler divergence share the same practical drawbacks as Negentropy. To use them in practice, we need to somehow approximate mutual entropy of unknown random variables. As a result, although theoretically different, all

three measures of nongaussianity (i.e., Mutual information and Kullback-Leibler and Negentropy) lead to essentially same ICA algorithms.

4. *Maximum likelihood estimation:* Another well known method to estimate the independent components is *maximum likelihood (ML) estimation*. ML approach is based on *Log-likelihood* function (i.e. logarithm of likelihood) defined as (Pham et al., 1992)

$$L = \sum_{t=1}^{T}\sum_{i=1}^{n} \log(f_i(\mathbf{q}_i^T \mathbf{x}(t))) + T \log|\det(\mathbf{Q})| \qquad [12]$$

where time t is discretised into T samples and f_i is the probability density of the i-th source signal (f_i is assumed to be known). *Likelihood* can be represented as *Kullback-Leibler divergence* between the actual density of observations and factorised density of source signals. Thus, the ML approach is essentially equivalent to minimisation of mutual information.

There is an important limitation of ICA method. As already indicated by the listed measures of nongaussianity, Gaussian variables are not separable by ICA (Comon, 1994). Indeed, M-dimensional Gaussian distribution is invariant to any M-dimensional orthonormal transformation. Thus, two or more linearly combined Gaussian variables are not separable by ICA. The same applies to deterministic source signals with Gaussian distribution. ICA can separate them only if at most one source signal has a Gaussian distribution.

At the end of Section 3.1.1, we stated that PCA allows describing a set of statistical data (or a set of deterministic signals) using uncorrelated components (i.e., random variables or deterministic signals). Since PCA transformation is orthonormal, the variance (in the case of statistical data) or power (in the case of deterministic signals) of the observations is preserved by principal components (see Section 3.2). ICA is also useful to explore statistical data (deterministic signals). It provides independent random variables (independent deterministic signals) which preserve the information contained in the observations. In the sequel, two applications of PCA and ICA methods to the mixing model [3] are discussed.

3.2 Algebraic PCA method: application to an instantaneous mixing model

Algebraic method for the computation of principal components is based on correlation matrix of observations $\mathbf{x}(t)$:

$$\hat{\mathbf{R}}_{\mathbf{xx}} = \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mm} \end{bmatrix} \qquad [13]$$

where $r_{ij} = \frac{1}{T}\int_0^T x_i(t)x_j(t)dt$ (continuous signals) or $r_{ij} = \frac{1}{T}\sum_{t=1}^{T} x_i(t)x_j(t)$ (sampled signals) is the correlation between the ith and the jth observation. Note that $\hat{\mathbf{R}}_{\mathbf{xx}}$ is real, positive, and symmetric. Now, assume the observations $\mathbf{x}(t)$ are deterministic and follow the mixing model [3]. Consider the singular value decomposition (see the Appendix) of the $M \times N$ matrix $\mathbf{A}$:

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{U}^{\mathrm{T}}, \qquad [14]$$

where $\mathbf{U}_{N\times N}$ and $\mathbf{V}_{M\times M}$ (matrix) are unitary matrixes of sizes $N\times N$ and $M\times M$ (i.e. $\mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{I}$, $\mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I}$) and $\mathbf{\Lambda}$ is a diagonal $M\times N$ matrix with the N non-zero eigenvalues λ_i of $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ on the diagonal. Without loss of generality, we can assume λ_i are arranged in decreasing order. The *diagonal form* of the correlation matrix $\hat{\mathbf{R}}_{\mathbf{xx}}$ (for sampled signal) is given by:

$$\hat{\mathbf{R}}_{\mathbf{xx}} = \sum_{t=1}^{T} \mathbf{x}(t)\mathbf{x}^T(t) = \sum_{t=1}^{T} (\mathbf{A}\mathbf{s}(t) + \mathbf{n})(\mathbf{s}^T(t)\mathbf{A}^T + \mathbf{n}^T) = \mathbf{A}\mathbf{A}^T + \mathbf{I}\sigma_n^2 = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T + \mathbf{I}\sigma_n^2 \qquad [15]$$

where $\mathbf{I}\sigma_n^2$ is the covariance matrix of the noise (which, given the adopted assumptions on the white noise, is equal to the identity matrix multiplied by the noise variance). In equation [15], the normalisation of the source signals to the unit norm and the notion of uncorrelatedness of the sources signals and noise was used. It is worth noticing that the eigenvalues of $\hat{\mathbf{R}}_{\mathbf{xx}}$ sum up to the total power of observations $\mathbf{x}$(t).

Now, neglect the influence of noise and consider the relation between the eigenvectors of $\hat{\mathbf{R}}_{\mathbf{xx}}$ and the principal components $\mathbf{y}(t)$. Assume a signal $y_k(t)$ is a linear combination of the observations $\mathbf{x}(t)$. Than, $y_k(t)$ can be expressed as a linear combination of eigenvectors of $\hat{\mathbf{R}}_{\mathbf{xx}}$ (*completeness property* of the eigenvectors of the correlation matrix), multiplied by a unit norm vector **c:**

$$y_k(t) = \sum_{i=1}^{m} c_{\mathrm{i}}\mathbf{V}_{\mathrm{i}}x_{\mathrm{i}}(\mathrm{t}) = \mathbf{c}^T\mathbf{V}^T\mathbf{x} \ . \qquad [16]$$

where $\mathbf{V}_1$ is the i-th eigenvector of $\hat{\mathbf{R}}_{\mathbf{xx}}$ and $\mathbf{c} = [c_1, \ldots, c_m]$. The power of the signal $y_{\mathrm{k}}(t)$ can be expressed as

$$\sum_{t=1}^{T} y_k(t)y_k^T(t) = \mathbf{c}^T\mathbf{V}^T\sum_{t=1}^{T}\mathbf{x}(t)\mathbf{x}^T(t)\mathbf{V}\mathbf{c} = \mathbf{c}^T\mathbf{V}^T\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T\mathbf{V}\mathbf{c} = \sum_{i=1}^{m} c_i^2\lambda_i \qquad [17]$$

The right-most sum in [17] is a *convex combination* (linear combination with unitary sum of weights) of the eigenvalues, which takes a maximum at $c_i = \delta_{\mathrm{i},1}$ ($\delta_{\mathrm{i,j}}$ denoting the Kronecker delta).

Thus, the first eigenvector of $\hat{\mathbf{R}}_{\mathbf{xx}}$ indicates the direction of the maximum power (or variance) of the observations, which is, by definition, the first principal direction. The corresponding eigenvalue λ_i gives the power (variance) of the first principal component $\mathbf{V}_1^T\mathbf{x}(t)$. By repeating this procedure and limiting it to the subspace of eigenvectors $\mathbf{V}_2$ to $\mathbf{V}_M$, the second principal direction is found to be aligned with eigenvector $\mathbf{V}_2$, the third principal direction is aligned with $\mathbf{V}_3$ etc. Therefore, the eigenvectors **V** of correlation matrix $\hat{\mathbf{R}}_{\mathbf{xx}}$ reveal the principal directions of observations **x**(*t*).

It is worth noticing that a complete computation of the mixing matrix **A** requires not only the matrices **Λ** and **V**, but also the unitary matrix **U** (known as *rotation* matrix; see Section 4.2). As **U** is estimated by PCA, principal components are not sufficient to reconstruct the original source signals.

Note also that the $M \times M$ matrix $\hat{\mathbf{R}}_{\mathbf{xx}}$ has dimension larger or equal to that of **A** ($M \times N$), because $M > N$. According to equation [15], the first *N* eigenvalues and eigenvectors of $\hat{\mathbf{R}}_{\mathbf{xx}}$ provide information on the *N* source signals, whereas the additional $M - N$ eigenvalues provide information about the power (variance) of noise. This property will be used by the ICA approach in Section 4.2.

3.3 *Neural ICA method: application to instantaneous mixing model*

Indicating with $\mathbf{y}(t) = \mathbf{Q}\mathbf{x}(t)$ an estimate of the source signals in model [3], the aim of ICA methods is to compute an estimate $\mathbf{A}^{\mathbf{s}}$ of the *mixing* matrix **A**, such that $\mathbf{x}(t) = \mathbf{A}^{\mathbf{s}}\mathbf{y}(t)$. Algebraic PCA method, discussed in the previous Section, relies on a well known technique of eigenvalue decomposition (Appendix) and requires a priori knowledge of all the samples of observations **x**(*t*). This knowledge is not assumed by neural methods, which are based on an intrinsically different approach that allows real-time implementations. Neural techniques utilize iterative updates of weights $\mathbf{A}^{\mathbf{s}}$ in order to achieve convergence to the minimum of a predefined functional F[$\mathbf{A}^{\mathbf{s}}$]. F[$\mathbf{A}^{\mathbf{s}}$] measures the aforementioned distance among the different estimates of the source signals **y**(*t*) and attains the minimum when **y**(*t*) = **s**(*t*). Possible choices for F[$\mathbf{A}^{\mathbf{s}}$] are those listed in Section 3.1.

The weights $\mathbf{A}^{\mathbf{s}}$ are updated iteratively (Karhunen et al., 1997). In each iteration step, the value of the functional F[$\mathbf{A}^{\mathbf{s}}$] is decreased (i.e., the distance between the **y**(*t*) estimates of the source signals is increased). A widely used numerical minimization technique is *gradient descent algorithm* (or *stochastic gradient descent algorithm* in the case of random processes), for which the weights are updated in the direction opposite to the gradient of F[$\mathbf{A}^{\mathbf{s}}$]:

$$\mathbf{A}^{s}_{n+1} = \mathbf{A}^{s}_{n} - \mu_{n} \nabla F[\mathbf{A}^{s}]. \qquad [18]$$

where the notation ∇ is just shorthand for gradient. Performances and convergence of the minimization method depend (usually with opposite direction) on the parameter μ_n,, called *learning rate*, which determines the decrease of the weights in opposite direction of the gradient. Learning rate is usually chosen to be adaptive, with smaller values close to the minimum of the functional $F[\mathbf{A}^{s}]$.

As an example of application, we discuss the recursive neural network architecture introduced by Hérault and Jutten, 1991, (Figure 2) with the aim to separate two sources from two observations

$$\begin{cases} x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) \\ x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) \end{cases} \qquad [19]$$

where both $x_i(t)$ and $s_i(t)$ are signals with T samples. Every neuron receives the sequence of samples of observation $x_i(t)$ as an input. The outputs of all neurons are connected to all the inputs of the neurons with $j \neq i$ weighted by a scalar weight c_{ji} (Figure 2). The estimate of the output $\mathbf{y}(t) = \mathbf{Q}\ \mathbf{x}(t)$ is obtained by *separation matrix* $\mathbf{Q} = (\mathbf{I}+\mathbf{C})^{-1}$, where $\mathbf{C}$ is the matrix of weights c_{ji}. In the case of two sources considered in equation [19] we have

$$\begin{cases} y_1(t) = x_1(t) - c_{12}y_2(t) \\ y_2(t) = x_2(t) - c_{21}y_1(t) \end{cases} \qquad [20]$$

from which the following relation between source signals $\mathbf{s}(t)$ and their estimates $\mathbf{y}(t)$ arises

$$\begin{cases} y_1(t) = \dfrac{1}{1-c_{12}c_{21}}\left[\left(a_{11} - c_{12}a_{21}\right)s_1(t) + \left(a_{12} - c_{12}a_{22}\right)s_2(t)\right] \\ y_2(t) = \dfrac{1}{1-c_{12}c_{21}}\left[\left(a_{21} - c_{21}a_{11}\right)s_1(t) + \left(a_{22} - c_{21}a_{12}\right)s_2(t)\right] \end{cases} \qquad [21]$$

For the estimates to be proportional to the source signals, the following relations must hold

$$a_{12} - c_{12}a_{22} = 0\ ,\ \ a_{21} - c_{21}a_{11} = 0\ . \qquad [22]$$

In such a case y_1, y_2 are proportional to s_1, s_2, respectively. Assuming instead

$$a_{11} - c_{12}a_{21} = 0\ ,\ \ a_{22} - c_{21}a_{12} = 0 \qquad [23]$$

y_1, y_2 are proportional to s_2, s_1, respectively. The convergence of the method to either [22] or [23] is associated to the ICA ambiguity on the order of the reconstructed sources. Convergence actually occurs only for one of the solutions [22] and [23], as only one of the solutions is stable. Indeed, the condition that the loop gain $c_{21}c_{12}$ is less than 1 can be satisfied only by one of the solutions (Jutten and Hérault, 1991). For the gain $c_{21}c_{12}>1$ the method diverges.

Figure 2 about here

The theoretical study of M sources in [3] is simply an extension of the previous example with two sources. In the case of M sources, the learning rule is based on the gradient method as discussed in the sequel. Assume that the first *M-1* sources are already determined (up to the multiplicative constant)

$$\begin{cases} y_1(t) = a_{11}s_1(t) \\ \vdots \\ y_{M-1}(t) = a_{M-1,M-1}s_{M-1}(t) \end{cases} \qquad [24]$$

Substituting equation [24] into the expression for the estimation of the M-th source $y_M(t) = x_M(t) - \sum_{i=1}^{M-1} c_{Mi}y_i(t)$ and considering that $x_M(t) = \sum_{i=1}^{M} a_{Mi}s_i(t)$ we get:

$$y_M(t) = \sum_i \left(a_{Mi} - c_{Mi}a_{ii}\right)s_i(t) + a_{MM}s_M(t) \qquad [25]$$

In order to estimate the Mth source as a function $y_M(t)$ proportional to it, the weights c_{Mi} must be chosen so that the first term on the right hand side of [25] vanishes. By using the assumption of the uncorrelated source signals we have

$$E[y_M(t)^2] = \sum_i \left(a_{Mi} - c_{Mi}a_{ii}\right)^2 E[s_i(t)^2] + a_{MM}^2 E[s_M(t)^2] \qquad [26]$$

Thus, $E[y_M(t)^2]$ can be considered as the functional F[$\mathbf{A}^{\mathbf{s}}$], as its minimum is attained when the estimated source signal $y_M(t)$ is proportional to the source $s_M(t)$. Applying the gradient method to the functional $E[y_M(t)^2]$ we have

$$c_{ij}^{\ k} = c_{ij}^{\ k-1} + \mu_k E[y_i(t)y_j(t)] \qquad i \neq j\,, \qquad [27]$$

where k is the iteration step and μ_k is the positive constant determining the learning rate (i.e., the increment of the weights). In the case of the stochastic gradient method, the same equation [27] is obtained, but without the expectation operator.

There are infinite solutions corresponding to non correlated sources (Jutten and Hérault, 1991, and Jutten, 1987), but only one for which the sources are statistically independent. Thus, the rule must be modified so that the method converges to the unique solution corresponding to statistically independent sources. A further problem in the learning rule [27] is related to its symmetry: coefficients c_{12} and c_{21} vary in the same way; the solution to which the method converges is correct only if the *mixing* matrix **A** is symmetric. To avoid these problems, the learning rule [27] is substituted with

$$c_{ij}^{\ k} = c_{ij}^{\ k-1} + \mu_k E[f(y_i(t))g(y_j(t))] \qquad i \neq j \qquad [28]$$

where f and g are two different non linear, even functions (in order to break symmetry), with the same sign (in order for their product to have the same sign as $E[y_i(t)y_j(t)]$) and the direction opposite to that of the gradient is taken. It is possible to prove that, in the case of source signals with symmetrical probability densities, if iterative rule [28] converges, the obtained estimates $y_i(t)$, $y_j(t)$ are statistically independent.

4. APPLICATIONS

In this Section, examples of BSS application to the surface electromyographic (EMG) signals are discussed. Firstly, a short overview of the electrical activity of human muscle is outlined. The main focus is on generation of electrical potentials in muscle fibres and on their asynchronous merging into the detectable EMG interference patterns. The descriptions provided should serve only as a coarse introduction to the filed of electromyography. More advanced descriptions can be found in Merletti and Parker, 2004.

4.1 Physiology of human muscles

Human muscles consist of 10 to 150 mm long and 5 to 90 μm thin muscle fibres, which are attached to the bones in the tendon regions. Each muscle fibre is innervated by a single *motoneuron* which transmits the control commands from the *central nervous system* (CNS) in a form of the firing pulse trains. Several muscle fibres are innervated by the same motoneuron, forming a basic functional unit of the muscles, so called *motor unit* (MU). The number of fibres in each MU varies considerably within the same muscle and even more between different muscles. Typically, muscles comprise from several tens to several hundreds MUs.

Electrical signals, sent by CNS, propagate along a nervous fibre, terminate in *neuromuscular junctions* (NMJ) where they excite membranes of all innervated muscle fibres. Every pulse in a motoneuron induces a local depolarisation of the transmembrane potential of each muscle fibre, so called *single fibre action potential* (SFAP). The depolarised zone (i.e., SFAP) propagates without attenuation along the muscle fibre from the NMJ to the tendon endings, causing the muscle fibre to contract. The sum of single fibre action potentials corresponding to all the fibres of a single motor unit is called *motor unit action potential* (MUAP).

Several tens of MUs are simultaneously active in the muscle tissue. Their MUAPs superimpose in time and space and form highly complex interference pattern of EMG, which can be detected either within the muscle (with needle electrodes) or over the skin above the investigated muscle (with surface electrodes). The technical difficulties associated to interpretation of recorded EMG interference patterns, limit the accuracy and diagnostic value of EMG in practice and generate many source separation problems. In the sequel, only two representative examples of source separation are discussed. The first example deals with the problem of separation of EMG signals from different muscles (so called muscle crosstalk). The second example addresses identification of single MU discharge patterns, i.e. decomposition of surface EMG into constituent MUAP trains.

4.2 Separation of surface EMG signals generated by muscles close to each other (muscle crosstalk)

An important artefact of surface EMG signal is *crosstalk* from nearby muscles. Crosstalk is the signal detected over a muscle, but generated by a nearby muscle (Figure 3). This complex phenomenon depends on the properties of the propagating medium (i.e. subcutaneous tissue interposed between the muscle fibres and the detection electrodes) and on sources (i.e., the firing patterns of active MUs). The exact physical properties of interposed subcutaneous tissue are not known, hence, as little as possible a priori information on the communication channel is assumed.

Figure 3 about here

Crosstalk signals can be superimposed to the signal of interest both in time and in frequency domain and represents a serious problem for surface EMG. The distinction of signals generated by muscles close to each other is, hence, an example of a very important source separation problem. By assuming that the EMG signals generated by different muscles are statistically independent, the problem can be addressed by ICA techniques.

One of the first applications of BSS to the problem of crosstalk was proposed by Farina et al., 2004. Their work is based on separation algorithm called SOBI (*second-order blind identification*), introduced by Belouchrani et al., 1997. SOBI extends the PCA method presented in Section 3.2 and consists of two steps: 1) whitening and 2) assessment of the unknown rotation matrix **U**. The linear instantaneous mixing model [3] is assumed. Although not completely accurate, this model enables reasonable good approximation of surface EMG signals, especially when the investigated muscles are close to each other.

Step 1 - Whitening

Spatial whitening of the observations $\mathbf{x}(t)$ (decorrelation in space) follows the procedure for estimation of the principal components in Section 3.2. The $N \times M$ matrix $\mathbf{W}$ is determined such that:

$$\mathbf{W}\,\mathbf{A}\,\mathbf{A}^{\mathrm{T}}\,\mathbf{W}^{\mathrm{T}} = \mathbf{I}\,. \qquad [29]$$

By definition in [29], matrix $\mathbf{WA}=\mathbf{U}$ is unitary. Application of $\mathbf{W}$ to the observations $\mathbf{x}(t)$ yields so called *whitened observations* $\mathbf{z}(t)$:

$$\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t) = \mathbf{U}\mathbf{s}(t)+\mathbf{W}\mathbf{n}(t) \qquad [30]$$

By analogy with procedure in Section 3.2, matrix $\mathbf{W}$ can be determined from the covariance matrix of observations $\mathbf{x}(t)$:

$$\hat{\mathbf{R}}_{\mathbf{xx}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}(t)\mathbf{x}^{\mathrm{T}}(t) \qquad [31]$$

which can be factorised as

$$\hat{\mathbf{R}}_{\mathbf{xx}} \approx \mathbf{A}\,\hat{\mathbf{R}}_{\mathbf{ss}}\,\mathbf{A}^{\mathrm{T}}+\sigma^2\mathbf{I}_{\mathrm{n}} \qquad [32]$$

As the sources $\mathbf{s}(t)$ are uncorrelated, the covariance matrix $\hat{\mathbf{R}}_{\mathbf{ss}}$ is diagonal. Furthermore supposing all the sources of unitary power (ICA ambiguity on power of sources), $\hat{\mathbf{R}}_{\mathbf{ss}}$ can be made equal to identity. Under these assumptions, relations [29] and [32] indicate that matrix $\mathbf{W}$ diagonalises the matrix $\hat{\mathbf{R}}_{\mathbf{xx}}$. Thus, $\mathbf{W}$ and σ^2 can be computed from the eigenvalues and eigenvectors of $\hat{\mathbf{R}}_{\mathbf{xx}}$. Firstly, an estimate $\hat{\sigma}^2$ of the variance of the noise is obtained from the average of the M-N smallest eigenvalues of matrix $\hat{\mathbf{R}}_{\mathbf{xx}}$ (Section 3.2). Secondly, given the N greatest eigenvalues $\lambda_1,\ldots,\lambda_n$ and the correspondent eigenvectors $\mathbf{V}_1,\ldots,\mathbf{V}_M$ of $\hat{\mathbf{R}}_{\mathbf{xx}}$, $\mathbf{W}$ is given by:

$$\mathbf{W}=[(\lambda_1-\hat{\sigma}^2)^{-1/2}\,\mathbf{V}_1,\ldots,(\lambda_n-\hat{\sigma}^2)^{-1/2}\,\mathbf{V}_M]^{\mathrm{T}} \qquad [33]$$

Note that, although closely related, whitening by matrix $\mathbf{W}$ extends the PCA method described in Section 3.2, as it scales the whitened components $z_i(t)$ by factor $(\lambda_i-\hat{\sigma}^2)^{-1/2}$ to make them of unit norm (a property not required by PCA). In order to estimate the matrix $\mathbf{A}$, the unitary matrix $\mathbf{U}$ must be estimated by a rotation operation in the second step.

Step 2 - Rotation

From the matrix factorization $\mathbf{U}=\mathbf{WA}$, we have:

$$\mathbf{A} = \mathbf{W}^{\#}\mathbf{U} \qquad [34]$$

Thus, given the whitening matrix **W**, the *mixing* matrix **A** can be determined by estimating the matrix **U**. As **U** is unitary, it can be considered as an *N* dimensional rotation matrix and estimated by *joint-diagonalisation* procedure (Belouchrani et al., 2001) of the correlation matrices of whitened observations $\mathbf{z}(t)$. From the definition of the covariance matrix:

$$\hat{\mathbf{R}}_{\mathbf{zz}}(\tau) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{z}(t)\mathbf{z}(t+\tau)^{\mathrm{T}} \qquad [35]$$

we have:

$$\hat{\mathbf{R}}_{\mathbf{zz}}(\tau) \approx \mathbf{U}\,\hat{\mathbf{R}}_{\mathbf{ss}}(\tau)\,\mathbf{U}^{\mathrm{T}} \qquad \tau \neq 0 \qquad [36]$$

For non-zero lags τ the contribution of the temporarily white Gaussian noise $\mathbf{n}(t)$ vanishes. As a result, **U** can be determined from any matrix $\hat{\mathbf{R}}_{\mathbf{zz}}(\tau)$ at nonzero lag τ. A more stable procedure consists in choosing a number of matrices $\hat{\mathbf{R}}_{\mathbf{zz}}(\tau)$ for different values of $\tau \neq 0$ and determining the matrix **U** as a "best joint diagonaliser" of the set of selected matrices. By "best joint diagonaliser" we refer to a matrix which makes the matrices $\hat{\mathbf{R}}_{\mathbf{zz}}(\tau)$ as close to diagonal as possible. Ideally, **U** diagonalizes all the matrices $\hat{\mathbf{R}}_{\mathbf{zz}}(\tau)$, but this is seldom the case, mainly due the noise. Therefore, a criterion to measure a goodness of *joint-diagonalization* is required. Criterion of the choice is the sum of squares of off-diagonal elements of matrices $\mathbf{U}^{\mathrm{T}}\hat{\mathbf{R}}_{\mathbf{zz}}(\tau)\mathbf{U}$ (Belouchrani et al., 1997, Belouchrani e Amin, 1998, Belouchrani et al., 2001):

$$off(\hat{\mathbf{R}}_{\mathbf{zz}}) = \sum_{\tau=1}^{\Upsilon} \sqrt{\sum_{i}\sum_{j\neq i} \left| r_{ij}(\tau) \right|^2} \qquad [37]$$

where $r_{ij}(\tau)$ denotes the *(i,j)*-th element of selected matrix $\hat{\mathbf{R}}_{\mathbf{zz}}(\tau)$ for $\tau = 1,\ldots,\Upsilon$. Criterion [37] leads to implementation of so called Jacobi joint-diagonalization method (Cardoso, 1996), estimating the matrix **U** as

$$\mathbf{U} = \min_{\|\mathbf{U}\|=1} \arg\left(off(\mathbf{U}^{T}\hat{\mathbf{R}}_{\mathbf{zz}}\mathbf{U}) \right) \qquad [37]$$

Once the mixing matrix A is known, the sources can be estimated as $\mathbf{y}(t) = \mathbf{A}^{\#}\mathbf{x}(t)$. Exact technical description of joint-diagonalization surpasses the scope of this chapter. Interested reader is referred to Cardoso, 1996, Belouchrani et al., 1997, and Holobar et al., 2006.

An example of application of SOBI algorithm to experimental sEMG signals (Farina et al., 2004) is shown in Figure 4 (experimental setup) and Figure 5 (reduction of crosstalk). The algorithm was

applied to two forearm muscles, which allow rotation and flexion of the wrist. The two muscles are very close to each other and it is impossible to separate their EMG activity with classical methods. BSS algorithm was applied to three mixtures of signals detected over the two muscles and in an intermediate region, respectively. As demonstrated in Figure 4, it allows improving the selectivity of the detection when either a rotation or flexion of the wrist is executed.

Figures 4 and 5 about here

4.3 Separation of single motor unit action potentials from multi-channel surface EMG

The second BSS application, discussed in this section, includes the decomposition of surface EMG signals into constituent MUAP trains. As explained in Subsection 4.1, surface EMG is a compound signal comprising the contributions of different MUs. Even at moderate muscle contraction levels, many MUs contract asynchronously. Their MUAPs superimpose both in space and time and create complex interference patterns, which are very difficult to interpret. Nonetheless, surface EMG received remarkable attention over the past few decades and has become a mature measuring technique. The information extracted from the sEMG signals is currently being exploited in several different clinical studies mainly concerned with timing of muscle activation, EMG amplitude modulation and electrical manifestations of fatigue.

The development of flexible high-density (HD) arrays of surface electrodes and multi-channel amplifiers opened new possibilities of recording up to a few tens of sEMG signals over a single muscle. At the same time, source separation techniques, capable of processing and combining the information from such a multichannel recordings, emerged. De Luca et al., 2006, proposed the decomposition of four-channel sEMG, while Kleine et al., 2007, demonstrated the importance of recording many sEMG signals over the skin surface for decomposition purposes. BSS methods have also been proposed. Garcia at. al, 2005, and Nakamura et al., 2004, acquired sEMG signals with the linear array of surface electrodes, oriented transversally with respect to the muscle fibres, and demonstrated that, in this configuration and up to reasonable limitations, sEMG signals can be modelled as linear instantaneous mixtures (i.e., by the model [3]). On the other hand, Holobar and Zazula, 2004, modelled sEMG signals as linear convolutive mixtures (i.e., by the model [1]) and proposed the *Convolution Kernel Compensation* (CKC) decomposition technique. This technique proved to be highly accurate and robust; reconstructing MUAP trains of up to twenty MUs from a multichannel sEMG recordings. In the sequel, CKC decomposition technique is discussed in more details.

4.3.1 Convolution Kernel Compensation

In the case of isometric muscle contractions, sampled sEMG signals $\mathbf{x}(t)$ can be modelled as outputs of convolutive linear mixing model [1]:

$$x_i(t) = \sum_{j=1}^{N} \sum_{l=0}^{L-1} a_{ij}(l) s_j(t-l) + n_i(l), \quad i=1,...,M \qquad [38]$$

where $n_i(l)$ stands for zero-mean additive noise. Each model input $s_j(t)$ is modelled as binary pulse sequences, carrying the information about the MUAP activation times

$$s_j(t) = \sum_{r=-\infty}^{\infty} \delta[t - T_j(k)], \; j=1,...,N \qquad [39]$$

where $\delta(\tau)$ denotes the Dirac impulse and $T_j(k)$ stands for the time instant in which the *k*-th MUAP of the *j*-th MU appeared. Activation times $T_j(k)$ are supposed to be random and statistically independent (experimental observations show almost periodic discharge rates and correlated fluctuations of rate among different MUs). The channel response $a_{ij}(l)$; $l=0,1,...,L-1$, corresponds to the *L* samples long MUAP of the *j*-th MU, as detected in the *i*-th observation. The channel responses $a_{ij}(l)$ must be limited support, but can be of arbitrary shape. Hence, any physical property of the subcutaneous tissue can be taken into account.

Model [38] can be rewritten in a matrix form:

$$\mathbf{x}(t) = \mathbf{A}\bar{\mathbf{s}}(t) + \mathbf{n}(t) \qquad [40]$$

where $\mathbf{n}(t) = [n_1(t),....,n_M(t)]^T$ is a vector of white noise with a covariance matrix $\sigma_n^2\mathbf{I}$ and the mixing matrix **A** comprises all the MUAPs as detected by the different surface electrodes

$$\mathbf{A} = \begin{bmatrix} a_{11}(0) & \dots & a_{11}(L-1) & a_{12}(0) & \dots & a_{12}(L-1) & \dots \\ a_{21}(0) & \dots & a_{21}(L-1) & a_{22}(0) & \dots & a_{22}(L-1) & \dots \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots & \dots \\ a_{M1}(0) & \dots & a_{M1}(L-1) & a_{M2}(0) & \dots & a_{M2}(L-1) & \dots \end{bmatrix}. \qquad [41]$$

Vector $\bar{\mathbf{s}}(t)$ stands for an extended form of a sampled source signals $\mathbf{s}(t)$:

$$\bar{\mathbf{s}}(t) = \left[s_1(t), s_1(t-1), ..., s_1(t-L+2), ..., s_N(t), ..., s_N(t-L+2)\right]^T \qquad [42]$$

The CKC method (Holobar and Zazula, 2004) fully automates the identification of MU discharge sequences in equation [40]. In the first step, the cross-correlation vector $\mathbf{r}_{s_j\mathbf{x}} = E\left[s_j(t)\mathbf{x}^T(t)\right]$ between the *j*-th source signal and all the measurements is estimated (Holobar and Zazula, 2007). In the

second step, the j-th pulse train s_j is estimated by the linear minimum mean square error (LMMSE) estimator:

$$\hat{s}_j(t)=\mathbf{r}_{s_j\mathbf{x}}\hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\mathbf{x}(t)=\mathbf{r}_{s_j\bar{\mathbf{s}}}\mathbf{A}^T\left(\mathbf{A}\hat{\mathbf{R}}_{\bar{\mathbf{s}}\bar{\mathbf{s}}}\mathbf{A}^T+\sigma_n^2\mathbf{I}\right)^{-1}(\mathbf{A}\bar{\mathbf{s}}(t)+\mathbf{n}(t)) \qquad [43]$$

where $\hat{\mathbf{R}}_{\mathbf{xx}}=E\left[\mathbf{x}(t)\mathbf{x}^T(t)\right]$ is the correlation matrix of measurements $\mathbf{x}(t)$, and $\mathbf{r}_{s_j\bar{\mathbf{s}}}=E\left[s_j(t)\bar{\mathbf{s}}^T(t)\right]$ is the vector of cross-correlation coefficients between the j-th source and all the sources. By analogy with Subsection 4.2, $\hat{\mathbf{R}}_{\bar{\mathbf{s}}\bar{\mathbf{s}}}=\mathbf{I}$, whereas, in the case of statistically independent sources, $\mathbf{r}_{s_j\bar{\mathbf{s}}}=[\delta(1-j),\delta(2-j),...\delta(N-j)]$ equals to the unit norm vector with the j-th element equal to 1 and zeros elsewhere. When the influence of noise is neglected, the unknown mixing matrix $\mathbf{A}$ is compensated and equation [43] simplifies to

$$\hat{s}_j(t)=\mathbf{r}_{s_j\mathbf{x}}\hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\mathbf{x}(t)=\mathbf{r}_{s_j\bar{\mathbf{s}}}\mathbf{A}^T\mathbf{A}^{-T}\hat{\mathbf{R}}_{\bar{\mathbf{s}}\bar{\mathbf{s}}}\mathbf{A}^{-1}\mathbf{A}\bar{\mathbf{s}}(t)=\mathbf{r}_{s_j\bar{\mathbf{s}}}\hat{\mathbf{R}}_{\bar{\mathbf{s}}\bar{\mathbf{s}}}\bar{\mathbf{s}}(t)=s_j(t) \qquad [44]$$

Estimator [43] requires the cross-correlation vector $\mathbf{r}_{s_j\mathbf{x}}$ to be known in advance. This is never the case and Holobar and Zazula, 2007, proposed probabilistic iterative procedure for its blind estimation. In the first iteration step, the unknown cross-correlation vector is approximated by vector of measurements $\mathbf{r}_{s_j\mathbf{x}}=\mathbf{x}(t_1)$ where, without loss of generality, we assumed the j-th MU discharged at time instant t_1. Then, the first estimation of the j-th source $s_j(t)$ is computed according to [43]. In the next step, the largest peak in $\hat{s}_j(t)$ is selected as the most probable candidate for the second discharge of the j-th source, $t_2=\max_{t_1\neq t_2}\arg(\hat{s}_j(t))$, and the vector $\mathbf{r}_{s_j\mathbf{x}}$ is updated as:

$$\mathbf{r}_{s_j\mathbf{x}}=\frac{\mathbf{r}_{s_j\mathbf{x}}+\mathbf{x}(t_2)}{2} \qquad [45]$$

This procedure is then repeated, until $\mathbf{r}_{s_j\mathbf{x}}$ converges to a stable solution (Holobar and Zazula, 2007).

CKC method inherently resolves MUAP superimpositions. Moreover, it implicitly combines all the available information provided by all the observations $\mathbf{x}(t)$. By compensating for the shapes of the detected MUAPs (which are included in the mixing matrix $\mathbf{A}$), it directly estimates the impulse sources without reconstructing the detected MUAP shapes. This significantly decreases the number of unknowns to be estimated in model [40] and reduces the computational time. When required, MUAP shapes can be estimated by spike triggered averaging of sEMG signals, taking the MUAP activation times $s_j(t)$ as triggers.

The problem with the CKC method is that the convolutive model [40] increases the number of sources **s**(t) by the factor *L*. Thus, in order to decompose the sEMG signals, the number of observations must also be large (at least a few dozens). This calls for HD sEMG acquisition systems with at least several tens of pick-up electrodes arranged into closely-spaced 2D grid. An example of CKC-based sEMG decomposition is illustrated in Figure 6.

Figures 6 and 7 about here

Appendix

Eigenvalue Decomposition

A vector $\mathbf{v}_i$ which changes length but not direction when operated upon by a matrix **A** is said to be an *eigenvector* of **A**. The length scale factor is called *eigenvalue* of **A**. Eigenvalues λ_i and eigenvectors (directions) $\mathbf{v}_i$ of a matrix **A** with dimensions $M \times M$ are defined by

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i \qquad [A1]$$

where λ_i are scalars and $\mathbf{v}_i$ are *M* dimensional vectors. A matrix **A** can be represented in Jordan form as

$$\mathbf{V}^T\mathbf{A}\mathbf{V} = diag(\mathbf{J}_1,\ldots,\mathbf{J}_r) \qquad [A2]$$

where $\mathbf{V} = [\mathbf{v}_1,\ldots, \mathbf{v}_M]$, *diag* is a block diagonal matrix, *r* is the number of independent eigenvectors of **A** and $\mathbf{J}_i$ indicates the Jordan block associated to the ith eigenvalue

$$\mathbf{J}_i = \begin{bmatrix} \lambda_i & 1 & 0 & \ldots & 0 \\ 0 & \lambda_i & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \ldots & \ldots & 0 & \lambda_i \end{bmatrix} \qquad [A3]$$

The dimension of the Jordan block is the multiplicity of the correspondent eigenvalue. If matrix **A** has *M* independent eigenvectors the Jordan representation simplifies to a diagonal form:

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_m \end{bmatrix} \qquad [A4]$$

Singular Value Decomposition

A rectangular matrix **B** with dimensions $M \times N$ can be represented as

$$\mathbf{V}^T\mathbf{B}\mathbf{U} = \mathbf{\Lambda} \qquad [A5]$$

where **U** ($N \times N$ matrix) and **V** ($M \times M$ matrix) are the matrices of the right and left eigenvectors, respectively, defined as

$$\mathbf{B}\ \mathbf{u}_i = \sigma_i\ \mathbf{v}_i\ , \qquad [A6]$$

where $\sigma_1, \ldots, \sigma_p$ are the singular values (i.e., the square root of the eigenvalues of $\mathbf{B}^T\mathbf{B}$). $\mathbf{\Lambda}$ is M × N matrix with the singular values σ_i on the diagonal and zero elsewhere. The left eigenvectors $\mathbf{v}_i$ are also the eigenvectors of matrix $\mathbf{B}\mathbf{B}^T$. In tensorial notation, matrix **B** can be represented as a sum of dyadic forms

$$\mathbf{B} = \sum_{k=1}^{m} \sigma_i\ \mathbf{v}_i\ \mathbf{u}_i^T\ . \qquad [A7]$$

Pseudoinverse matrix of **B** is defined as

$$\mathbf{B}^{\#} == \sum_{k=1}^{p} \frac{1}{\sigma_i}\ \mathbf{u}_i\ \mathbf{v}_i^T \qquad [A8]$$

Consider matrix **B** as the *mixing* matrix with $M > N$. The problem of identification of the N sources $\mathbf{s}(t)$ from the M mixtures $\mathbf{x}(t)$ insists on the overdetermination of the system, so that a solution of the problem

$$\mathbf{B}\ \mathbf{s}(t) = \mathbf{x}(t) \quad \text{or} \quad \mathbf{B}\ \mathbf{s}(t) - \mathbf{x}(t) = \mathbf{0} \qquad [A9]$$

does not exist in general, as there are more independent conditions than unknowns (the independence of the conditions comes from the noise which is always superimposed to the observations $\mathbf{x}(t)$). For a solution to exists, a weaker definition of solution is introduced, i.e. the function $\mathbf{s}_d(t)$ minimising the squared error

$$\mathbf{s}_d(t) = \min_{\mathbf{s}} \arg \left\| \mathbf{B}\mathbf{s}(t) - \mathbf{x}(t) \right\|^2 \qquad [A10]$$

The *theorem of projections* for Hilbert spaces implies that

$$\mathbf{B}\ \mathbf{s}_d(t) - \mathbf{x}(t) \in Im(\mathbf{B})^{\perp} = Ker(\mathbf{B}^T) \qquad [A11]$$

where $Im(\mathbf{B})$ and $Ker(\mathbf{B})$ are the image and the kernel of the matrix **B** and the simbol $^{\perp}$ indicates the orthogonal space. Thus, we get

$$\mathbf{B}^T\ \mathbf{B}\ \mathbf{s}_d(t) = \mathbf{B}^T\mathbf{x}(t) \qquad [A12]$$

and

$$\mathbf{s}_d(t) = (\mathbf{B}^T\ \mathbf{B})^{-1}\mathbf{B}^T\mathbf{x}(t) = \mathbf{B}^{\#}\mathbf{x}(t) \qquad [A13]$$

where definition [A8] is used. Pseudoinverse multiplied by the vector of observations gives the sources $\mathbf{s}_d(t)$ which minimise the squared error [A10].

Acknowledgements.

This work was supported by the European Project CyberManS (contract n. 016712), by a Marie Curie Intra-European Fellowships within the 6th European Community Framework Programme

(contract n 023537) and by Fondazione Cassa di Risparmio di Torino and Compagnia di San Paolo di Torino.

Captions of the figures

Figure 1. Application of PCA to the pair of surface EMG signals. Signals x_1 in panel a) and x_2 in panel b) were recorded by close-by sets of electrodes, placed over the skin above the biceps brachii muscle during a low level contraction. Both electrode systems detected electrical activity of approximately the same group of muscle fibres. As a result, the signals x_1 and x_2 exhibit a high level of redundancy. As demonstrated by the joint vector space presentation in panel c), more than 90 % of variance is in the first principal direction (i.e., the direction of the first principal component). Each circle in c), depicts a pair of values ($x_1(t)$, $x_2(t)$) at a fixed time t. The first principal direction is denoted by black dashed line, the second principal direction by a black dotted line. The two principal components $y_1(t)$ in panel d) and $y_2(t)$ in panel e) were reconstructed by projecting the observations $x_1(t)$ and $x_2(t)$ on the subspaces spanned by principal directions. The first principal component $y_1(t)$ resembles the main dynamics in the observations $x_1(t)$ and $x_2(t)$, whereas $y_2(t)$ can be interpreted as a low noise uncorrelated with $y_1(t)$. Panel f) depicts the joint vector space representation of the principal components ($y_1(t)$, $y_2(t)$) after rotation of the axes depicted in c).

Figure 2. Iterative neural network architecture, introduced by Hérault and Jutten, 1991, for separation of two sources $s_1(t)$ and $s_2(t)$ out of two observations $x_1(t)$ and $x_2(t)$. Two processing blocks are depicted, with the mixing process (*left*) and separation algorithm (*right*). Separation block consist of two neurons (one per each source). Each neuron receives the samples of both observations, $x_1(t)$ and $x_2(t)$, as an inputs. Output of each neuron is multiplied by a weight c_{ij}, $j \neq i$, and fed back to the input of the other neuron. The estimate of the output $\mathbf{y}(t) = \mathbf{Qx}(t)$ is obtained with a *separation* matrix $\mathbf{Q}=(\mathbf{I}+\mathbf{C})^{-1}$, where $\mathbf{C}$ is the matrix with the weights c_{ij} (see the text for details). This method is suitable for real time implementation.

Figure 3. Sketchy representation of two muscles and two detection systems for surface EMG signals. Each detection system acquires both the EMG signal of the muscle over which the electrodes are placed and the EMG signal produced by the other, near-by muscle. This phenomenon is known as a muscle crosstalk. Its linear instantaneous model is shown in Fig. 2. A third detection system in intermediate position could record a third mixture of the two sources, as shown in Figure 4.

Figure 4. Experimental setup for the detection of surface EMG signals from two forearm muscles. The hand is fixed in an isometric brace measuring the force produced during rotation and flexion efforts. a) The subject alternates wrist rotation and flexion efforts at regular time intervals. b) EMG signal is detected with three electrode arrays placed over the pronator teres, the flexor carpi radialis and between the two muscles. The signal detected over the pronator teres (which rotates the wrist) is not zero during flexion, even if this muscle is not active during this contraction (crosstalk signal from flexor carpi radialis muscle), and vice versa (see Figure 5). Reproduced with permission from reference Farina et al., 2004.

Figure 5. Application of source separation technique to the three signals detected over the pronator teres, the flexor carpi radialis and between the two muscles with the techniques described in Figure 4. These signals are the mixtures $\mathbf{x}(t)$ of the source separation problem. SOBI algorithm provides the separation of the activity of the pronator teres, reducing the amplitude of the signal recorded over the pronator teres during flexion. The same holds for the flexor carpi radialis, for the complementary time intervals (results not shown). Reproduced with permission from reference Farina et al., 2004.

Figure 6. Experimental setup for detection of surface EMG signals from abductor pollicis brevis muscle. a) matrix of 64 surface electrodes, arranged into 13 lines and 5 columns and with the four corner electrodes missing (*upper panel*) and the isometric braces measuring the force produced during the abduction of the thumb (*lower panel*). b) Example of recorded surface EMG signals (positions of the signals reflect the spatial organization of the pick-up electrodes)

Figure 7. Surface EMG signals, recorded during a 6 s ramp-up (from 0 % to 10 % of maximum contraction level - MVC) and 6 s ramp-down (from 10 % to 0 % MVC) contraction of abductor pollicis brevis muscle and their decomposition into contributions of different motor units. a) surface EMG signals detected by the first six electrodes of the central column (Figure 6). b) the same as in a), with the portion of the signal zoomed-in. c) discharge patterns of 12 identified motor units and their dependence on the exerted muscle force. Each line corresponds to a single motor unit discharge. d) MUAP templates of eight different motor units, as reconstructed by a spike triggered averaging of the sEMG signals from the central electrode column (Figure 6), taking the identified discharge patterns as triggers.

References

ANEMÜLLER J., GRAMß T., 1999. On-line blind separation of moving sound sources. In: Cardoso, J.F., Jutten, C., Loubaton, P (Eds.), Proceedings of the First International Workshop on Independent Component Analysis and Blind Signal Separation, Aussois, France, pp. 331-334.

BAR-NESS Y., 1982. Bootstrapping adaptive interference cancelers: Some practical limitations, The Globecom. Conf., Miami, Paper F 3.7, 1251-1255.

BELOUCHRANI A., ABED-MERAIM K., CARDOSO J.F., MOULINES E., 1997. A blind source separation technique using second-order statistics, IEEE Trans. Signal Proc., vol. 45, pp. 434-443.

BELOUCHRANI A., AMIN M.G., 1998. Blind source separation based on time-frequency signal representations, IEEE Trans. Signal Proc., vol. 46, pp. 2888-2897.

BELOUCHRANI A., ABED-MERAIM K., AMIN M. G., ZOUBIR A., 2001. Joint-antidiagonalization for blind source separation, Proc. ICASSP, vol. 5, pp. 2789-2792.

BOUSBIAH-SALAH H., BELOUCHRANI A., ABED-MERAIM K., 2001. Jacobi-like algorithm for blind source separation, Electronic Letters, vol. 37, n. 16, pp. 1049-1050.

BOUSBIAH-SALAH H., BELOUCHRANI A., ABED-MERAIM K., 2001. Blind separation of non stationary sources using joint block diagonalization, Proc. IEEE Workshop on Statistical Signal Processing, pp. 448-451.

CARDOSO J.F., 1989. Sources separation using higher order moments, Proc. Internat. Conf. Acust. Speech signal Processes, Glasgow, 2109-2112.

CARDOSO J.F., SOULOUMIAC A., 1996, Jacobi angles for simultaneous diagonalization, SIAM J. Mat. Anal. Appl., vol. 17, pp. 161-164.

SOULOUMIAC A., CARDOSO J.-F., 1991. Comparaison de methodes de separation de sources, Proc. GRETSI, Juan les Pins, France.

COMON P., 1992. Independent Component Analysis, Internat. Signal Processing Workshop on High-Order Statistics, Elsevier, pp. 29-38.

COMON P., 1994. Independent Component Analysis: a new concept?, Signal Processing, vol. 36, pp. 287-314.

DE LUCA C. J, ADAM A., WOTIZ R., GILMORE L.D., NAWAB S.H., 2006. Decomposition of Surface EMG Signals, J Neurophysiol. vol. 96, pp. 1646-1657.

FARINA D., FÉVOTTE C., DONCARLI C., MERLETTI R., 2004. Blind separation of linear instantaneous mixtures of non-stationary surface myoelectric signals, IEEE Trans. Biomed. Eng., vol. 51(9), pp. 1555-67.

FETY L., 1988. Methodes de traitment d'antenne adaptèes aux radiocommunications, Doctorate Thesis, ENST.

GAETA M., LACOUNME J.-L., 1990. Source separation without a priori knowledge: The maximum likelihood solution, Proc. Eur. Signal Processing Conf., pp. 621-624.

GARCIA G.A., OKUNO R., AKAZAWA K., 2005. A decomposition algorithm for surface electrode-array electromyogram. A noninvasive, three-step approach to analyze surface EMG signals, IEEE Eng Med Biol Mag., vol. 24, pp. 63-72.

GIANNAKIS GB, INOUYE Y., MENDEL J.M., 1989. Cumulant Based Identification of Multichannel Moving Average Models, IEEE Trans. Automat. Control, vol. 34, 783-787.

HOLOBAR A., ZAZULA D., 2002. A new approach for blind source separation of convolutive mixtures of pulse trains, Proc. of BSI '02, Como, Italy, pp. 163-166.

HOLOBAR A., ZAZULA D., 2003. Surface EMG decomposition using a novel approach for blind source separation, Informatica Medica Slovenica, vol. 8, 1, pp. 2-14.

HOLOBAR A., ZAZULA D., 2007. Multichannel blind source separation using convolution kernel compensation. *IEEE trans. signal process.*, , vol. 55, pp. 4487-4496.

HOLOBAR A., OJSTERŠEK M., ZAZULA D., 2006. Distributed Jacobi joint diagonalization on clusters of personal computers. Int. J. parallel program, vol. 34, pp. 509-530.

HYVARINEN A., 1998. New approximations of differential entropy for independent component analysis and projection pursuit, Advances in Neural Information Processing Systems, vol. 10, 273-279. MIT press.

HYVARINEN A., 1999. Survey on Independent Component Analysis, Neural Computing Surveys 2, 94-128.

HYVARINEN A., E. OJA, 2000. Independent Component Analysis: Algorithm and applications, Neural Networks, vol. 13(4-5), pp. 411-430.

HYVARINEN A., KARHUNEN J., OJA E., 2001. Independent Component Analysis, John Wiley & sons, Inc. New York,.

INOUYE Y., MATSUI T., 1989. Cumulant based parameter estimation of linear systems, Proc. Workshop Higher Order Spectral analysis, Vail, Colorado, pp. 180-185.

JOLLIFFE I.T., 1986. Principal Component Analysis, Springer-Verlag.

JONES M.C., SIBSON R., 1987. What is projection pursuit? J. Of the Royal Statistical Society, ser. A, 150:1, 36.

JUTTEN C., HERAULT J., 1991. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. Signal Processing, vol. 24, pp. 1-10.

JUTTEN C., 1987. Calcul neuromimétique et traitement du signal. Analyse en composante indépendantes, Tesi di Dottorato, PhD thesis, INPG, Univ. Grenoble.

KARHUNEN J., OJA E., WANG L., VIGÁRIO R., JOUTSENSALO J., 1997. A class of neural networks for independent component analysis. IEEE Trans. on Neural Networks, vol. 8(3), pp. 486-504.

KLEINE B.U., VAN DIJK J.P., LAPATKI B.G., ZWARTS M.J., STEGMAN D.F., 2007. Using two-dimensional spatial information in decomposition of surface EMG signals, J of Electromyogr Kinesiol. vol. 5, pp. 535-548.

KOUTRAS A., DERMATAS E., KOKKINAKIS G., 2000. Blind separation of speakers in noisy environments: A neural network approach, Neural Network World, vol. 10(4), pp. 619-630.

LACOUME J.L., 1999. A survey of source separation, ICA'99, 1-6, Aussois, January 11-15.

MAKEIG S., BELL A.J., JUNG T.-P., SEJNOWSKI T.-J., 1996. Independent component analysis of electroencephalographic data. Advances in Neural Information Processing Systems 8, pp. 145-151, MIT press.

MERLETTI R., PARKER P.A., 2004. Electromyography: physiology, engineering, and non-invasive applications, IEEE Press and John Wiley & Sons.

NAKAMURA H., YOSHIDA M., KOTANI M., AKAZAWA K., MORITANI T., 2004. The application of independent component analysis to the multi-channel surface electromyographic signals for separation of motor unit action potential trains: part I-measuring techniques, J. Electromyog. Kinesiol., vol. 14, pp. 423-432.

PHAM D.-T., GARRAT P., JUTTEN C., 1992. Separation of a mixture of independent sources through a maximum likelihood approach. Proc. EUSIPCO, pp. 771-774.

RUIZ P., LACOUME J.L., 1989. Extraction of independent sources from correlated inputs, in Proc. Workshop on Higher Order Spectral Analysis, Vail, Colorado, pp. 146-151.

SOROUCHYARI E., 1991. Blind Source Separation, Part III: Stability analysis, Signal Processing, vol. 24, pp. 21-29.

VIGÁRIO R.,1997. Extraction of ocular artifacts from EEG using independent component analysis. Electroenceph. clin. Neurophysiol., vol. 103(3), pp. 395-404.

VIGÁRIO R., JOUSMÄKI V., HÄMÄLÄINEN M., HARI R., OJA E., 1998. Independent component analysis for identification of artifacts in magnetoencephalographic recordings. Advances in Neural Information Processing 10 (Proc. NIPS'97), pp. 229-235, Cambridge, MA. MIT press.

VIGÁRIO R., SÄRELÄ J., JOUSMÄKI V., OJA E., 1999. Independent component analysis in decomposition of auditory and somatosensory evoked fields. Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99), pp. 167-172, Aussois, France.

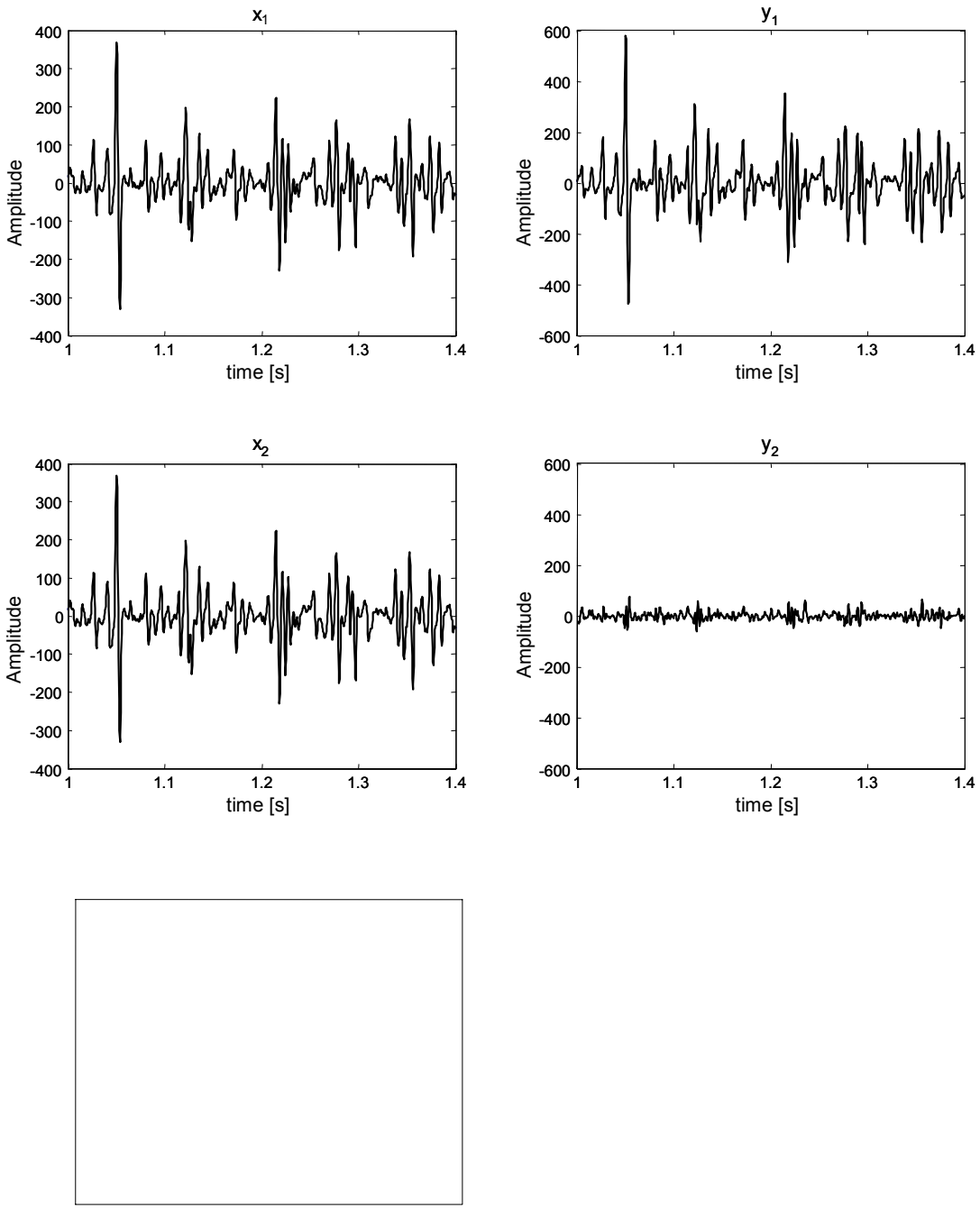
x1
y1
x2
y2
Amplitude
time [s]
400
300
200
100
0
-100
-200
-300
-400
600
400
200
0
-200
-400
-600
1
1.1
1.2
1.3
1.4

Figure 1