Optimized Wavelets for Blind Separation of Nonstationary Surface Myoelectric Signals

Dario Farina*, Marie-Françoise Lucas, and Christian Doncarli

Abstract—Surface electromyography (EMG) signals detected over the skin surface may be mixtures of signals generated by many active muscles due to poor spatial selectivity of the recording. In this paper, we propose a new method for blind source separation (BSS) of nonstationary signals modeled as linear instantaneous mixtures. The method is based on whitening of the observations and rotation of the whitened observations. The rotation is performed by joint diagonalization of a set of spatial wavelet distributions (SWDs). The SWDs depend on the selection of the mother wavelet which can be defined by unconstrained parameters via the lattice parameterization within the multiresolution analysis framework. As the sources are classically supposed to be mutually uncorrelated, the design parameters of the mother wavelet can be blindly optimized by minimizing the average (over time lags) cross correlation between the estimated sources. The method was tested on simulated and experimental surface EMG signals and results were compared with those obtained with spatial time-frequency distributions and with second-order statistics (only spectral information). On a set of simulated signals, for 0-dB signal-to-noise ratio (SNR), the cross-correlation coefficient between original and estimated sources was 0.92 ± 0.07 with wavelet optimization, 0.74 ± 0.09 with the wavelet leading to the poorest performance, 0.85 ± 0.07 with Wigner–Ville distribution, 0.86 ± 0.07 with Choi–Williams distribution, and 0.73 ± 0.05 with second-order statistics. In experimental conditions, when the flexor carpi radialis and pronator teres were concomitantly active for 50% of the time, crosstalk was 55.2 ± 10.0% before BSS and was reduced to 15.2 ± 6.3% with wavelet optimization, 30.1 ± 15.0% with the worst wavelet, 28.3 ± 12.3% with Wigner–Ville distribution, 26.2 ± 12.0% with Choi–Williams distribution, and 35.1 ± 15.5% with second-order statistics. In conclusion, the proposed approach resulted in better performance than previous methods for the separation of nonstationary myoelectric signals.

Index Terms—Blind source separation (BSS), optimization, surface electromyography (EMG), wavelet design.

I. INTRODUCTION

Electromyographic (EMG) signals detected over the skin surface are often mixtures of signals generated by many active muscles, due to crosstalk [6]. Thus, it is sometimes not possible to identify the activity of individual muscles, in particular, when the muscles are small and close to each other. Since the EMG signals generated by closely located muscles overlap in the frequency domain, classic filtering methods fail in separating the sources. Moreover, EMG signals often overlap also in the time domain since different muscles may be active in similar time intervals. The tissues separating the muscle fibers and the recording electrodes act as a low-pass filter whose cutoff frequency depends on the distance of the source [17]. However, there is no a priori information on the specific effect of the tissues separating the sources from the recording electrodes since tissue conductivities and source location are unknown. The problem of source separation should thus be addressed blindly.

We have recently proposed the application of blind source separation (BSS) methods for the identification of individual muscle activities from mixtures of surface EMG signals [7]. The method was based on whitening and rotation of the observations (recorded signals). The rotation was performed on spatial time-frequency distributions of the whitened observations [2], [10]. A time-frequency representation of EMG signals for blind separation is convenient when the signals are nonstationary since sources may be better separated in the time-frequency plane than in the frequency domain. Theoretically, uncorrelated sources can be separated if they differ in at least one time-frequency point [2].

Another tool for time-frequency representation of signals is the wavelet transform. Wavelet coefficients are partly localized in time and frequency and form a multiscale representation of the signal, leading to localized frequency subbands with equal width on a logarithmic scale. The wavelet transform depends on the mother wavelet which has to be selected depending on the specific application. The mother wavelet defines, through translation and scaling, the basis functions over which the signal is projected.

In this paper, we develop a novel approach for BSS in which the rotation step is based on spatial wavelet transformations of the whitened observations. Moreover, we propose the parameterization of the mother wavelet and its optimization based on a blind performance criterion of source separation. The method is applied to both simulated and experimental surface EMG signals.

II. METHODS

A. Signal Model

The following linear instantaneous signal model is adopted, for time $t = 1, \ldots, T$:

$$\mathbf{x}[t] = \mathbf{A}\mathbf{s}[t] + \mathbf{n}[t]$$  \hspace{1cm} (1)
where \( \mathbf{x}[t] = [x_1[t], \ldots, x_m[t]]^T \) is the vector of size \( m \) containing the mixtures (observations), \( \mathbf{s}[t] = [s_1[t], \ldots, s_n[t]]^T \) is the vector of size \( n \) containing the sources, \( A \) is the full rank mixing matrix of size \( m \times n \) with \( m \geq n \) (the number of sources should not be larger than the number of observations), and \( \mathbf{n}[t] \) is the additive noise vector of equal power \( \sigma^2 \) on each observation. The sources are modeled as the realizations of zero mean, nonstationary, mutually uncorrelated random processes. Moreover, it is assumed that the number of sources is known a priori which is not a limiting factor in the current application. The noise is assumed to be an independent and identically distributed random process, independent of the sources and with covariance \( \sigma^2 \mathbf{I}_m \), where \( \mathbf{I}_m \) denotes the identity matrix of size \( m \). This assumption implies that the noise sequences on each observation are mutually uncorrelated and of the same power \( \sigma^2 \).

The overall objective of BSS is to obtain an estimate \( \hat{A} \) of \( A \), up to the standard BSS indeterminacies on ordering and scale factor (and phase in the complex case) [1]. Once \( \hat{A} \) is known, the sources can be estimated by applying the pseudoinverse matrix to the observations. As in many BSS methods, we proceed in the following two steps: 1) estimation of a “spatial whitening matrix” and 2) estimation of the “missing rotation matrix.”

### B. Whitening

The whitening step is the same as in [1] and [2] (see also [7]–[10]). Spatial whitening consists in finding a matrix \( \mathbf{W} \) of size \( n \times m \) such that

\[
\mathbf{W} \mathbf{A} \mathbf{D}^H \mathbf{W} = \mathbf{I}_n
\]

(2)

where \( H \) denotes the complex conjugate transpose of a matrix, \( \mathbf{I}_n \) denotes the identity matrix of size \( n \), and \( \mathbf{D} \) is an unknown positive–definite diagonal matrix. The matrix \( \mathbf{W} \) can be computed from the sample covariance matrix of the observations \( \mathbf{R}_{\text{swr}} \), provided that the sample covariance matrix of the sources (mutually uncorrelated) is close to diagonal. Given the \( n \) largest eigenvalues \( \lambda_1, \ldots, \lambda_n \), corresponding to the eigenvectors \( \mathbf{h}_1, \ldots, \mathbf{h}_n \) of \( \mathbf{R}_{\text{swr}} \), a spatial whitening matrix \( \mathbf{W} \) is built as [2]

\[
\mathbf{W} = [(\lambda_1 - \sigma^2)^{-1/2}\mathbf{h}_1, \ldots, (\lambda_n - \sigma^2)^{-1/2}\mathbf{h}_n]^H.
\]

(3)

An estimate \( \hat{\sigma}^2 \) of the noise variance is the average of the \( m - n \) smallest eigenvalues of \( \mathbf{R}_{\text{swr}} \) [1].

### C. Rotation

If we define \( \mathbf{U} = \mathbf{W} \mathbf{A} \mathbf{D}^{1/2} \), by definition of \( \mathbf{W} \) [see (2)], \( \mathbf{U} \) is a unitary square matrix of size \( n \). It can be shown that [22]

\[
\mathbf{A} \mathbf{D}^{1/2} = \mathbf{W}^\# \mathbf{U}
\]

(4)

where \( ^\# \) denotes the Moore–Penrose pseudoinverse of a matrix. Thus, the matrix \( \mathbf{A} \mathbf{D}^{1/2} \), equal to \( \mathbf{U} \) up to the diagonal matrix \( \mathbf{D}^{1/2} \) (which models BSS scale indeterminacy), can be recovered from \( \mathbf{W} \) and \( \mathbf{U} \).

For stationary sources, the matrix \( \mathbf{U} \) can be obtained by a procedure of diagonalization of matrices obtained by the autocovariance matrix of the whitened observations [1]. For nonstationary sources, the spatial time-frequency power distribution of the whitened observations can be used [2], [10]. However, a problem arises from the approximation of the spatial Wigner–Ville spectrum by a spatial time-frequency representation (of the Cohen’s group, for example). This approximation does not insure the diagonality of the time-frequency spatial distribution of the sources for all time-frequency locations. Thus, two issues must be blindly addressed: the choice of the most convenient approximation (i.e., the choice of the kernel of the time-frequency representation) and the selection of the time-frequency locations where the diagonalization is performed. As the choice of the kernel of the time-frequency representation was shown to have a small importance [9], we propose here a similar approach based on the diagonalization of the spatial wavelet distribution of the whitened observations, in order to have more degrees of freedom (choice of the mother wavelet).

The whitened observations \( \tilde{x}[t] \) can be written as instantaneous linear mixtures of the sources with mixing matrix \( \mathbf{U} \)

\[
\tilde{x}[t] = \mathbf{W} \mathbf{x}[t] = \mathbf{W} \mathbf{A} \mathbf{s}[t] + \mathbf{W} \mathbf{n}[t] = \mathbf{U} \mathbf{D}^{-1/2} \mathbf{s}[t] + \mathbf{W} \mathbf{n}[t].
\]

(5)

The dyadic discrete wavelet transform (DWT) decomposes a signal \( p(t) \) on a basis where all the basis functions are diluted and translated versions of a prototype function \( \psi \), called mother wavelet. The projection of the signal into these basis functions returns detail coefficients \( d_p(j,k) = \langle p(t), \psi_{j,k}(t) \rangle \), where \( \psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k) \). Approximation coefficients \( a_p(j,k) \) are obtained by projecting the signal on dilated and translated versions of the scaling function \( \varphi \) (father wavelet), \( a_p(j,k) = \langle p(t), \varphi_{j,k}(t) \rangle \), where \( \varphi_{j,k}(t) = 2^{-j/2}\varphi(2^{-j}t - k) \). The set of \( N \) coefficients \( c_p(j,k) = \{a_p(j,k), d_p(j,k)\}_{j=1,\ldots,J; k=1,\ldots,2^J; j \in \log_2 N} \) (with \( J \) the deepest level of decomposition) of the decomposition of a discrete-time signal \( p[t] \) of length \( N \) can be alternatively computed by Mallat’s algorithm with the application of a filter bank [multiresolution analysis (MRA)] [19].

The cross DWT coefficients for two signals \( p_1[t] \) and \( p_2[t] \) are defined as [23] \( c_{p_1,p_2}(j,k) = c_{p_1}(j,k) \cdot \overline{c_{p_2}(j,k)} \), where \( \overline{c} \) indicates the complex conjugate. We define the spatial wavelet distribution (SWD) of the vector \( p[t] \) of size \( n \) the matrix \( \mathbf{C}_{p}(j,k) \) with entries \( \mathbf{C}_{p}(j,k)_{i,j,h} = c_{p}(j,k) \) for \( i,h = 1, \ldots, n \). For a given time-scale location \( (j,k) \), \( \mathbf{C}_{p}(j,k) \) is a square matrix of size \( n \) whose diagonal entries contain the square magnitude of the auto wavelet coefficients, whereas nondiagonal entries contain the cross-wavelet coefficients.

From (1) and the assumptions on noise, we obtain

\[
C_{\omega\omega}(j,k) = A C_{\phi\phi}(j,k) A^H + \sigma^2 I_m
\]

(6)

where \( C_{\omega\omega}(j,k) \) and \( C_{\phi\phi}(j,k) \) are the SWDs of the observations and of the sources. Defining the following whitened and denoised SWD matrices

\[
\mathbf{C}_{\phi}(j,k) = \mathbf{W}(\mathbf{C}_{\omega}(j,k) - \sigma^2 \mathbf{I}_m) \mathbf{W}^H
\]

(7)

from (6), we obtain

\[
\mathbf{C}_{\phi}(j,k) = \mathbf{U}(\mathbf{D}^{-1} C_{\phi\phi}(j,k)) \mathbf{U}^H.
\]

(8)

If the SWDs were derived from the Affine Wigner transform, \( C_{\phi\phi}(j,k) \) would be diagonal at any time-scale location. Then,
noting that \( \mathbf{U} \) is unitary, \( \mathbf{U} \) would diagonalize \( \mathbf{C}_{XX}(j, k) \) for any \((j, k)\). Thus, \( \mathbf{U} \) could be estimated from the eigenvectors of any matrix \( \mathbf{C}_{XX}(j, k) \) with distinct eigenvalues, up to column permutations and sign changes (and phase shift in the complex case), which correspond to BSS indeterminacies. At this step, we could theoretically compute \( \mathbf{U} \) and thus recover \( \mathbf{A} \).

In practice, however, the SWDs are derived from DWTs and \( \mathbf{C}_{SS}(j, k) \) is not diagonal for any time-scale location. Thus, instead of computing \( \mathbf{U} \) by diagonalizing a single matrix \( \mathbf{C}_{XX}(j_0, k_0) \) for a single time-scale location \((j_0, k_0)\), it is rather suggested to compute \( \mathbf{U} \) from the joint diagonalization of several matrices \( \{\mathbf{C}_{XX}(j_i, k_i), \ i = 1, \ldots, K\} \), corresponding to several convenient time-scale locations \( \{j_i, k_i\}, \ i = 1, \ldots, K\) (see Section II-D). This procedure is similar to that of selecting the time-frequency points in previous approaches [1], [3], [9], [10], [13]. Joint diagonalization provides a better estimate of \( \mathbf{U} \) since an “average eigenstructure” of the whole set \( \{\mathbf{C}_{XX}(j_i, k_i), \ i = 1, \ldots, K\} \) is computed [3].

Furthermore, joint diagonalization does not require that all the matrices have distinct eigenvalues.

D. Selection of the Time-Scale Locations for Joint Diagonalization

The matrix \( \mathbf{U} \) is not a diagonalization matrix for all time-scale locations \((j, k)\). We can restate the problem by saying that \( \mathbf{U} \) can be found as the matrix that diagonalizes the SWD of the whitened observations for the time-scale points for which the SWD of the sources is diagonal. It is thus necessary to find blindly a set of time-scale locations \( \{j_i, k_i\}, \ i = 1, \ldots, K\) for which \( \mathbf{C}_{SS}(j, k) \) is actually diagonal. To do so, we observe that the matrix \( \mathbf{C}_{SS}(j, k) \) is diagonal for all time-scale locations in which only one source is present (it follows from the definition of the cross-wavelet transform). In this case, there would be only one nonzero term on the diagonal [13]. A diagonal matrix \( \mathbf{C}_{SS}(j, k) \) with only one nonzero entry on the diagonal is referred to as a single autoterm matrix [10], [13]. Such matrices have one nonzero eigenvalue. Since the eigenvalues of \( \mathbf{C}_{XX}(j, k) \) are the same as those of \( \mathbf{C}_{SS}(j, k) \) (it follows from the fact that \( \mathbf{U} \) is unitary and from (8)) [1], the source single autoterm locations can be detected as those where \( \mathbf{C}_{XX}(j, k) \) has one dominant eigenvalue. This leads to the following selection criterion, proposed in [10] and [13] for the time-frequency approach:

Select \((j, k)\) if 
\[
CR(j, k) = \max_{\epsilon_{\text{SAT}}} \left| \text{eig}(\mathbf{C}_{XX}(j, k)) \right| \geq 1 - \epsilon_{\text{SAT}}
\]  

(9)

where \( \epsilon_{\text{SAT}} \) denotes the set of eigenvalues of \( \mathbf{C}_{XX}(j, k) \) and \( \epsilon_{\text{SAT}} \) is close to zero. To overcome the problem of choosing the threshold \( \epsilon_{\text{SAT}} \), a fixed number of time-scale locations corresponding to the highest values of \( CR(j, k) \) will be chosen in the following results. This ranking assures in practice the selection of single autoterm corresponding to all the sources if the number of selected terms is much larger than the number of sources. The single autoterm criterion allows the blind selection of the time-scale points in which the SWD of the sources is diagonal.

E. Parametrization of the Mother Wavelet

The SWD of the whitened observations [see (8)] depends on the selected mother wavelet. In the MRA framework, the scaling function \( \varphi \) and its associated wavelet \( \psi \) are related to two filters \( h \) and \( g \) by the two-scale recursive relations
\[
\varphi(t/2) = \sqrt{2} \sum_n h[n] \varphi(t-n) \quad \text{and} \quad \psi(t/2) = \sqrt{2} \sum_n g[n] \varphi(t-n).
\]

In the case of orthogonal wavelets, \( g \) can be deduced from \( h \) from the relation \( g[k] = (-1)^{j-k} h[1-k] \). Consequently, by restricting to the orthogonal case, \( h \) defines \( \psi \). The problem of choosing a specific mother wavelet becomes the problem of selecting a finite set of filter coefficients (parametrization of the wavelet).

To generate an orthogonal MRA wavelet, \( h \) must satisfy some constraints. For a finite impulse response (FIR) filter of length \( L \), there are \( L/2+1 \) sufficient conditions to ensure the existence and orthogonality of the scaling function and wavelets [15], [16]. Thus, \( L/2+1 \) degrees of freedom remain to design the filter \( h \). The lattice parameterization described by Vaidyanathan [25] offers the opportunity to design \( h \) via unconstrained optimization: the \( L \) coefficients of \( h \) can be expressed in term of \( L/2 \) new free parameters. We denote \( \theta \) the design parameter vector. For instance, if \( L = 6 \), we need a two-component design vector \( \theta = [\alpha, \beta] \), and \( h \) is given by [5], [21]

\[
i = 0, 1: h[i] = [(1 + (-1)^i \cos \alpha + \sin \alpha) \\
\times (1 - (-1)^i \cos \beta - \sin \beta) \\
+ (-1)^i \sin \beta \cos \alpha)] / (4\sqrt{2})
\]

\[
i = 2,3: h[i] = [1 + \cos(\alpha - \beta) + (-1)^i \sin(\alpha - \beta)] / (2\sqrt{2})
\]

\[
i = 4,5: h[i] = \sqrt{2} - h(i-4) - h(i-2).
\]

(10)

For other values of \( L \), expressions of \( h \) are given in [5] and [21].

With this wavelet parameterization, there are infinite available SWDs \( \mathbf{C}_{XX}(j, k) \), which depend on the design parameter vector \( \theta \), to represent the whitened observations.

F. Optimization Criterion

With wavelet parameterization, (8) is rewritten as

\[
\mathbf{C}_{XX}(j, k) = \mathbf{U}^{\theta} (\mathbf{D}^{-1} \mathbf{C}_{SS}(j, k)) \mathbf{U}^{\theta H}
\]

(11)

which leads to an estimated source matrix \( \mathbf{s}^{\theta} \) that depends on the parameterization vector \( \theta \). Different values of \( \theta \) may lead to different quality in source separation. In order to choose the optimal \( \theta \) value (and thus the optimal wavelet), a blind criterion of performance is needed. We propose to compute the absolute value of the cross-correlation function between estimated sources over time lags, centered around zero (number of time lags equal to signal length), and to minimize its average value. This criterion is based on the observation that the separation is optimal if the reconstructed sources are uncorrelated for any time lag.

G. Summary of the Method

The method is based on the following steps.

1) Compute the covariance matrix of the observations for time lag zero \( \mathbf{R}_{xx} = (1/T) \sum_{t=1}^{T} x[t] x[t]^H \) and note the \( n \)
largest eigenvalues \( \lambda_1, \ldots, \lambda_n \), corresponding to the eigenvectors \( \mathbf{h}_1, \ldots, \mathbf{h}_n \) of \( \mathbf{R}_{xx} \).

2) Estimate the noise variance as the average of the \( m - n \) smallest eigenvalues of \( \mathbf{R}_{xx} \).

3) Build a spatial whitening matrix \( \mathbf{W} \) as \( \mathbf{W} = \left[(\lambda_1 - \sigma^2)^{-1/2}\mathbf{h}_1, \ldots, (\lambda_n - \sigma^2)^{-1/2}\mathbf{h}_n \right]^H \).

4) For each value of \( \theta \) in a grid of predefined values, repeat the following steps.
   a) Compute the SVD \( \mathbf{C}^\theta_{xx}(j, k) \) of the whitened observations as a matrix with entries
   \[
   [\mathbf{C}^\theta_{xx}(j, k)]_{i,h} = c_{\mathbf{w} \mathbf{h}}(j, k), \quad \text{for} \quad i, h = 1, \ldots, n, n.
   \]
   b) A unitary matrix \( \hat{\mathbf{U}}^\theta \) is estimated by joint diagonalization of a set of matrices \( \mathbf{C}^\theta_{xx}(j_1, k_1), \ldots, (j_i, k_i), \ i = 1, \ldots, K \). The time-scale points are chosen through the criterion in (9).
   c) The sources \( \hat{\mathbf{s}}^\theta[t] \) are recovered by applying the pseudoinverse of the estimated mixing matrix \( \hat{\mathbf{A}} \mathbf{D}_1/2 = \mathbf{W}^H \hat{\mathbf{U}}^\theta \) to the observations.
   d) Compute the average \( \omega^\theta \) of the absolute value of the cross-correlation function between the pairs of estimated sources (in case \( n = 2 \), the cross-correlation function is \( R_{s_1 s_2}[\tau] = (1/T) \sum_{t=1}^{T} s_1[t] s_2[t + \tau]^H \); in case \( n > 2 \), the average of the absolute value of the cross-correlation function over all pairs of estimated sources is computed).
   e) Select the \( \theta \) value (\( \theta_{\text{opt}} \)) that leads to the minimum \( \omega^\theta \) (criterion).
   f) The estimated sources are \( \hat{\mathbf{s}}^\theta_{\text{opt}}[t] \), obtained from \( \theta_{\text{opt}} \).

H. Simulation of Surface EMG

The BSS method was tested on both simulated and experimental signals. In both cases, two sources and three observations have been considered. A cylindrical, multilayered volume conductor model has been used for simulating surface EMG signals generated by two muscles. The volume conductor includes bone (radius 20 mm), muscle tissue (26 mm thick), subcutaneous layer (3 mm), and skin (1 mm) (Fig. 1) with conductivities as in [8]. The simulated muscle fibers had a length of 100 mm with end plate located in the middle of the fiber. The motor unit conduction velocity had a Gaussian distribution with mean 4 m/s and standard deviation (SD) 0.3 m/s [24]. Motor units (number of muscle fibers per motor unit in the range 50–1000) were recruited according to the size principle [12], with recruitment function described by Fuglevand et al. [11]. Fiber density in the motor units was 20 fibers/mm² [11] and motor units had intermingled territories. Two small muscles with elliptical shapes (semiaxes 3.5 mm and 10 mm; cross-sectional area 110 mm²) were simulated (Fig. 1). Single differential recording systems (10-mm interelectrode distance) were located in three detection points, two directly over the two muscles and one in the middle. Each muscle was activated with a Gaussian excitation profile with SD 0.2 ms and maximum value 40% or 80% of the maximum excitation. The two profiles overlapped by two degrees, leading to four simulated conditions (Fig. 1). For each simulation condition, the BSS methods were tested on 20 signal realizations obtained by randomly locating the motor units within the muscles and adding a different noise realization.

I. Experimental Protocol

The experimental protocol used for recording EMG signals was similar to that described in [7]. Six male subjects (age, mean ±SD, 28.1 ± 2.5 years) were asked to alternatively activate the flexor carpi radialis and pronator teres muscles. The exercise consisted of a series of cycles each made of 1) 3 s of flexion (flexor carpi radialis) at 50% of the maximal voluntary contraction (MVC), 2) 1-s rest, 3) 3 s of rotation (pronator teres) at 50% MVC, and 4) 1-s rest. The cycle, lasting 8 s, was repeated for 100 s. The subjects were asked to perform flexion and rotation as selectively as possible by using force feedback on both rotation and flexion.

Surface EMG signals were detected with three linear adhesive arrays (LISIn–SPES Medica, Italy) of four electrodes (1 × 5 mm) with 10-mm interelectrode distance, in bipolar configuration. Only the central bipolar signal provided by each array was used for further analysis, leading to three bipolar recordings. EMG signals were amplified (amplifier LISIn–Ottonio Bioelettronica, Torino, Italy), bandpass filtered (±3 dB bandwidth, 10–500 Hz), sampled at 2048 Hz, and converted to digital data by a 12 bit analog-to-digital (AD) converter. Before placement of the arrays, the skin was lightly abraded with abrasive paste and the muscles of interest were identified by palpation during a few test contractions of flexion and rotation of the wrist. The arrays were placed parallel to each other in the direction of the muscle fibers. The first array (leading to the signal called trace #1) was located over the pronator teres, the second (trace #2) over the flexor carpi radialis and the third (trace #3) in between the two muscles. The transverse distance between adjacent arrays was approximately 10 mm.

From the exercise definition, the two investigated muscles were not active together and a signal recorded over an active muscle without activity of the other muscle was assumed as an original source, i.e., the first (respectively, the second) source corresponded to the signal recorded by the first (second) array during the activity period of only pronator teres (flexor carpi radialis). Muscle activity periods were selected from the recorded forces. Thus, source 1 (respectively, source 2) consisted of the trace #1 (respectively, #2) inside the activity period of muscle #1 (respectively, #2) and zero outside. It was thus possible to obtain the original sources (for method performance evaluation) and the activity of each muscle from three detection points without (in case of ideally selective activation) activity of the other muscle.

In order to generate observations with different degrees of temporal overlap between the sources, synthetic observations were built as follows. Traces #1, #2, and #3 were segmented into epochs corresponding to the activity of muscle #1 and muscle #2. The synthetic observation #1 (respectively, #2 and #3) was obtained by mixing two time-shifted versions of trace #1 (respectively, #2 and #3) recorded during activity of, respectively, muscle #1 and muscle #2. This allowed the generation of experimental observations in which the two sources had
variable time overlap. The characteristics of these observations were the same as if the signals would have been recorded by the three arrays with the two muscles active in partially overlapping time intervals. The activity of each muscle independently was recorded at the three locations, thus mixtures of the three traces represented concomitant muscle activity. The original recordings corresponded to time overlap 0% which is of limited interest since the two sources could be separated in time domain by simple segmentation. Time overlaps of 50% and 75% were tested. Moreover, since the task was cyclic, several realizations of sources and observations were available. Results from the first ten cycles were used (averaged) in this study.

J. Signal Analysis

The proposed BSS algorithm was applied to both the simulated and experimental signals with optimization of the mother wavelet, as described previously, and, for comparison, with the time-frequency approach applied in [7], with Wigner–Ville and Choi–Williams ($\pi = 1$) kernel, and with the second-order blind identification (SOBI) approach proposed in [1], which was designed for stationary signal analysis. The number of time-frequency and time-scale points for joint diagonalization was ten in all cases, as previous work has shown only a moderate improvement in performance when increasing the number of time-frequency points above ten [7]. For wavelet optimization, a filter length $L = 4$ was chosen in all cases. This led to one free parameter which was varied among 39 values uniformly spaced between $-\pi$ and $\pi$. The increase in filter length and/or number of values assigned to the free parameters would lead to selection among a larger number of mother wavelets. However, it was observed (results not shown) that, while the computational complexity increases, there was not a substantial improvement of performance for experimental signals when increasing $L$ or the number of discretization steps with respect to the chosen values. This was probably due to the fact that simplified model structure (uncorrelated sources, instantaneous mixtures; see Section IV) was the main limiting factor in performance for experimental recordings.

The reference sources (not used by the BSS algorithm) were used for testing performance. The measure of performance was the cross correlation between the original and estimated (after BSS) sources (average value over the sources) for both simulated and experimental tests. For this purpose, the cross correlation was computed between all pairs of estimated and original sources and the maximum values identified the correct association between estimated and original sources.
(thus, solving a posteriori the BSS indeterminacy, association between original and estimated sources could be solved in practice also without knowledge of the original sources). In experimental conditions, the power of the signal outside the interval of activation of the associated muscle was computed relative to the total signal power as a measure of crosstalk before and after BSS. This index of crosstalk is peculiar in the sense that the larger the overlapping is, the smaller the crosstalk index becomes. This choice was, however, preferred over others since it quantifies the signal power outside the activation interval of a muscle which may be a way to quantify crosstalk in practical applications (with selective muscle activation and recording over several muscles). It is noted that the choice of the crosstalk index is not critical since in all cases the same index was compared for the same degree of overlapping across the different methods. Comparison of this index for different degrees of overlapping is not of interest.

III. RESULTS

A. Simulated Signals

Fig. 2 shows an example of source separation with optimized wavelet. Table I reports the performance index with joint diagonalization of SWDs computed with the best wavelet (i.e., the wavelet leading to the best performance, as selected a posteriori when the true sources are known), the optimized wavelet (selected with the blind criterion), and the wavelet leading to the worst performance. Results are also reported for the joint diagonalization of the spatial Wigner–Ville and Choi–Williams distributions [2] and of the autocorrelation matrix of the whitened observations (SOBI algorithm [1]). The optimized wavelet provided results close to the best wavelet in all conditions, indicating that the blind optimization criterion provides an estimate of the wavelet leading to the best separation. However, the relation between the blind criterion and the a posteriori performance index was not linear, not perfectly monotonous, (Fig. 3), and the optimized wavelet did not always corresponded to the best wavelet, as also evidenced in Table I. Performance depended on the wavelet (Fig. 3) and there was no wavelet which was optimal in all cases, thus the necessity for signal-based optimization. The wavelet method with optimization of the mother wavelet always led to better results than previous time-frequency methods. In the third simulated condition for 10 dB SNR, for example, the performance index was $0.92 \pm 0.07$ with wavelet optimization, $0.74 \pm 0.09$ with the wavelet leading to the poorest performance, $0.85 \pm 0.07$ with Wigner–Ville distribution, $0.86 \pm 0.07$ with Choi–Williams distribution, and $0.73 \pm 0.05$ with SOBI. The SOBI algorithm provided poor results in comparison to the nonstationary methods, indicating that separation of the sources was better achieved in the time-frequency plane than using only the frequency information. Fig. 4 shows the comparison between original and estimated sources for some of the methods compared.

B. Experimental Signals

Fig. 5 shows an example (75% overlap) of sources separated by the proposed approach in experimental recordings. Table II

![Figure 2. Example of separation performed by whitening and rotation of the spatial wavelet distribution with optimized mother wavelet. (a) One of the three observations. (b) The first original source and its estimate. (c) The second original source with its estimate. The simulation corresponds to the third condition in Fig. 1 and no noise was added (nu: normalized units).](image-url)
Fig. 3. (a) Performance index as a function of the free parameter that defines the mother wavelet. Results are reported for one simulated signal from the third simulation set (Fig. 1) and 10-dB SNR. (b) Optimal and worst wavelet for the case shown in (a). (c) Scatter plot between the blind criterion for optimization (average cross correlation between estimated sources) and the a posteriori performance criterion (cross correlation between estimated and original sources) for signals simulated within the third simulation set (Fig. 1) and no noise. Note that the minimum value of the blind criterion corresponds to almost optimal performance but that the relation is not linear and not perfectly monotonous.

Fig. 4. Example of source separation by different BSS methods. (a) Original (solid lines) and estimated (with optimized wavelet) (dashed lines) source. Sources estimated with (b) optimized wavelet, (c) worst wavelet, (d) Choi–Williams, and (e) SOBI. Cross-correlation coefficients between original and estimated sources (average over the two sources) are also reported. The simulation corresponds to the third condition in Fig. 1, with 10-dB SNR (nu: normalized units).

reports the crosstalk index and cross correlation between estimated and original sources for the different methods tested. The BSS approach with optimized wavelet led to the highest crosstalk reduction and best source reconstruction as compared to the other approaches.

IV. DISCUSSION

We have proposed an innovative approach for BSS based on spatial wavelet distributions. The approach follows previous work [1], [2], [10] in which Cohen’s class time-frequency representations were used for the rotation of the whitened observations. We have also proposed the parameterization of the mother wavelet and its optimization in order to minimize the reconstruction error in the separation (measured by the cross correlation between the estimated sources). Optimization of the mother wavelet proved to be superior to random selection of the wavelet and to previous approaches.

A. Blind Separation of Surface EMG Signals

Surface EMG signals detected over the skin surface may be mixtures of signals generated by many muscles due to poor spatial selectivity of the recording [6]. Since the sources may overlap in both temporal and frequency domain and there is no a priori information, BSS approaches should be applied for identification of individual muscle activities [7].

As discussed in [7], the model adopted in this study, i.e., linear instantaneous mixtures, does not properly describe the generation of the surface EMG. Indeed the action potentials generated
by each muscle fiber are filtered by the tissues interposed between the sources and the detection electrodes. Thus, the mixture is convolutive and not linear instantaneous. In [7], we considered simulated mixtures which were generated in a linear instantaneous manner, so that the simulations exactly reproduced the model assumed for the development of the algorithm [see (1)]. In this paper, the mixtures were generated taking into account the convolutive effect of the tissues. The current validation in simulation is superior than the one we previously performed [7] since it considers realistic mixtures of EMG signals. The results shown indicate that in the conditions tested (small muscles may fail in more general conditions (for example, larger and more spaced muscles) when the convolutive nature of the mixtures may not be totally uncorrelated between sources is active. In each estimated source, the activity of the other source is reduced with respect to the observations, although not totally suppressed (nu: normalized units).

The main contribution of this study is the proposal of a novel way to represent the whitened observations for the purpose of source separation. The method is based on the DWT which provides information on both time and scale. However, performance strongly depends on the mother wavelet (Fig. 3 and Tables I and II). Convenient selection of the mother wavelet, which provides an additional degree of freedom, determined the best results among the methods tested, both in simulation and experimental recordings.

The optimization criterion for the mother wavelet was the minimization of the average absolute value of the cross-correlation between pairs of estimated sources. The criterion was proven to partly predict the performance of the BSS approach (compare best and optimized wavelet in Table I), although the relation between a posteriori performance index and blind criterion was not linear (Fig. 3) and not perfectly monotonous (Table I). The method for blind wavelet selection may be improved in further work, however, in the current version it allowed to blindly select a mother wavelet leading to the best separation among a number of other

| Table II | Crosstalk Index and Performance Index (Cross-Correlation Coefficient Between Original and Estimated Sources) (Mean ± SD, Over Six Subjects) for Different BSS Methods. Results Are Compared for Rotation Performed With the Wavelet Blindly Optimized (Optimized Wavelet), the Wavelet Leading to the Worst Performance (Worst Wavelet), Wigner–Ville, Choi–Williams Distributions, and Second-Order Statistics (SOBI). The Crosstalk Index Is Reported Also for the Observations |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Overlapping    | Crosstalk index | Performance index |
| %              | %               | %               | %               | %               |
| Observations   | 55.2 ± 10.0 %   | 19.8 ± 4.5 %    | --              | --              |
| Optimized      | 15.2 ± 6.3 %    | 5.1 ± 3.2 %     | 0.73 ± 0.15     | 0.73 ± 0.14     |
| Worst wavelet  | 30.1 ± 15.0 %   | 12.1 ± 5.2 %    | 0.59 ± 0.17     | 0.59 ± 0.19     |
| Wigner-Ville   | 28.3 ± 12.3 %   | 12.2 ± 5.6 %    | 0.60 ± 0.19     | 0.60 ± 0.21     |
| Choi-Williams  | 26.2 ± 12.0 %   | 11.2 ± 4.9 %    | 0.65 ± 0.15     | 0.65 ± 0.15     |
| SOBI           | 35.1 ± 15.5 %   | 16.3 ± 5.0 %    | 0.61 ± 0.11     | 0.58 ± 0.11     |

Fig. 5. Example of separation of experimental signals for 75% overlap in time. (a) Observations #1 and #2. (b) The two original sources estimated as the EMG over each muscle when the other muscle was not active. (c) Sources estimated with the proposed BSS method with optimized wavelet. (d) Comparison of a portion of original (solid lines) and estimated (dashed lines) sources in a time interval centered in the middle of the duration of the observations. The dashed line on the left indicates the time instant after which the first source is not active. The dashed line on the right indicates the time instant after which the second source is active. In each estimated source, the activity of the other source is reduced with respect to the observations, although not totally suppressed (nu: normalized units).
methods. The main contribution of this work is the proposal of a new representation of the whitened observations in which there are more degrees of freedom for optimizing performance. Similarly, we have previously proven that wavelet optimization substantially improves performance of algorithms for signal compression [20] and classification [18].

V. CONCLUSION

A new approach, based on SWD, for blind separation of nonstationary surface EMG signals has been proposed. The approach showed better performance than previous methods. Results on experimental signals pointed out that further improvement in performance may be reached with the use of a convolutive, rather than linear instantaneous, mixture model.

REFERENCES


Marie-Françoise Lucas graduated in electrical engineering from the Ecole Nationale Supérieure de Mécanique, Nantes, France, in 1984 and received the Ph.D. degree in signal processing from the Université de Nantes, Nantes, France, in 1986. Currently, she is an Associate Professor at the Ecole Centrale de Nantes, Nantes, France. Her main research interests are in the area of signal processing and decision. Her main application domain is biomedical signals.

Christian Doncarli graduated in electrical engineering from the Ecole Nationale Supérieure de Mécanique, Nantes, France, in 1971 and received the Ph.D. degree in signal processing from the Université de Nantes, Nantes, France, in 1976. Currently, he is the Professor of Signal Processing at the Ecole Centrale de Nantes, Nantes, France. His research theoretical interests concern decision problems (classification, detection, blind source separation), especially in transformed domains like time-frequency or time scale. His application domains are biomedical and audio signals.