SSA of biomedical signals: A linear invariant systems approach

A.M. Tomé*, A.R. Teixeira†, N. Figueiredo‡, I.M. Santos, P. Georgieva and E.W. Lang

Singular spectrum analysis (SSA) is considered from a linear invariant systems perspective. In this terminology, the extracted components are considered as outputs of a linear invariant system which corresponds to finite impulse response (FIR) filters. The number of filters is determined by the embedding dimension. We propose to explicitly define the frequency response of each filter responsible for the selection of informative components. We also introduce a subspace distance measure for clustering subspace models. We illustrate the methodology by analyzing Electroencephalograms (EEG).

KEYWORDS AND PHRASES: SSA, Linear invariant systems, Signal enhancement, Subspace distances.

1. INTRODUCTION

Projective subspace models, applied to time series data sets, can be found in literature under various names depending on the domain of application: Singular Spectrum Analysis (SSA) (for instance in climate time series analysis) [6, 5, 18] and Singular Value Decomposition (SVD) (for instance in speech enhancement) [4, 8, 10]. The aim of SSA is to achieve a decomposition of the original time series into a sum of a small number of interpretable components such as a slowly varying trend, oscillatory components and noise. While the aim of speech enhancement methods instead is simply to eliminate noise which is usually considered additive and normally distributed. In general, the processing chain comprises the following steps:

- Transformation of the time series into a sequence of lagged vectors (embedding) which are arranged into a trajectory matrix with either Toeplitz or Hankel structure.
- Estimating an orthogonal basis vector matrix, using singular value decomposition or principal component analysis.

- Projecting the multidimensional data vectors onto the new basis vectors.
- Selecting *relevant* components.
- Reconstructing the multidimensional embedded data using the possibly lower-dimensional subspace representation.
- Diagonal (or anti-diagonal) averaging to reconstitute the Toeplitz (or Hankel) structure of the reconstructed trajectory matrix.
- Reverting the embedding to yield an univariate time series.

Thus, the subspace model corresponds to an orthogonal matrix whose columns form basis vectors of the multidimensional space created by the embedding. The relevant components of the signal are obtained by projecting the data onto that basis and omitting *irrelevant* (for example noise-related) components. Finally an improved, univariate time series can be reconstructed. The selection of *relevant* components is a critical issue that can be addressed when the subspace model is generated. Recently, in [1, 2] the frequency profile of the basis vectors was studied to select those vectors (and related components) which correspond to the trends in the data.

The aim of this work is to show that, seen from a linear invariant systems perspective, the basis vectors of the time-delayed coordinate space can be interpreted as filters responding to signals in different frequency bands of the original time series. This interpretation was already discussed in [9, 8], however, here the frequency response of the filters is deduced in closed form. Using techniques from linear time-invariant systems theory to compute input - output relationships [12, 13], a closed-form expression for the frequency response of the filters is obtained such that the frequency profile of the output can be extracted easily. Another issue discussed here is how to measure a similarity between different subspace models computed for signals that have similar characteristics (see also [16]). This aspect is important because most of the selection schemes, discussed for example in signal enhancement applications, are simply based on studying eigenvalues and no attention is given to the corresponding eigenvectors. But it is equally important to know if two different data sets that lead to two closely related subspace models spread in similar directions of the underlying multidimensional space.

The paper is organized as follows: Section 2 resumes the main steps to compute the projective subspace model and discusses its application in a signal enhancement scenario. Section 2.2 presents the application of SSA using a linear invariant system approach and its interpretation as filter bank system. In Section 3 a distance measure to compare subspaces is presented. In Section 4 two examples are presented to illustrate the main characteristics of SSA using the approach and tools (namely the frequency response) of linear invariant systems. The second example (evoked potential signals) is a signal enhancement and clustering application.

2. SINGULAR SPECTRUM ANALYSIS

Time series analysis techniques often rely on embedding one-dimensional sensor signals in the space of their time-delayed coordinates. Embedding can be regarded as a mapping that transforms a one-dimensional time series into a multidimensional sequence of lagged vectors. Considering an univariate signal $(x[0], x[1], \ldots, x[N-1])$ with N samples, its multidimensional variant is obtained by $\mathbf{x}_k = (x[k-1+M-1], \ldots, x[k-1])^{\mathrm{T}}, k = 1, \ldots, K$, where $K = N - M + 1$. These lagged vectors form the columns of the data matrix $\mathbf{X}$ which is called a *trajectory matrix* [6]. The column vectors $\mathbf{x}_k$ of $\mathbf{X}$ span the space of dimension M

$$(1) \quad \mathbf{X} = \begin{bmatrix} x[M-1] & x[M] & \cdots & x[N-1] \\ x[M-2] & x[M-1] & \cdots & x[N-2] \\ x[M-3] & x[M-2] & \cdots & x[N-3] \\ \vdots & \vdots & \ddots & \vdots \\ x[0] & x[1] & \cdots & x[N-M] \end{bmatrix}.$$

Note that the trajectory matrix has identical entries along its diagonals. Such a matrix is called a Toeplitz matrix. There are other alternatives to organize a data matrix via embedding the univariate time series signal in an M-dimensional space of its time-delayed coordinates. For example, an Hankel matrix is obtained if the embedding is arranged such that identical elements occur along the anti-diagonals [6, 8]. Anyway, by computing two-point time correlations between the entries of the multi-dimensional representation of the signal, it is possible to find an orthogonal matrix $\mathbf{U}$ whose columns form an orthogonal basis of the M-dimensional space. The non-normalized correlation matrix is obtained as

$$\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$$

and its eigenvalue decomposition

$$\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}}$$

provides the related subspace model via the matrix $\mathbf{U}$ of basis vectors and corresponding eigenvalues $\mathbf{\Lambda} = $ diag$(\lambda_1, \ldots, \lambda_M)$. This subspace model is then used to manipulate the original data vectors according to

$$(2) \quad \hat{\mathbf{X}} = \mathbf{U}\mathbf{P}\mathbf{U}^{\mathrm{T}}\mathbf{X} = \mathbf{U}\mathbf{Y},$$

where

- $\mathbf{U}^{\mathrm{T}}\mathbf{X}$ corresponds to the dot product between the columns of $\mathbf{U}$ and the columns of $\mathbf{X}$.
- $\mathbf{P}$ is a diagonal selection matrix with the m-th diagonal entry equal to $p_{mm} = 1$ if the m-th row of $\mathbf{Y}$ is to be selected, or equal to $p_{mm} = 0$ if it is to be discarded.
- Each column of $\mathbf{Y} = \mathbf{P}\mathbf{U}^{\mathrm{T}}\mathbf{X}$ thus is the weighted projection of a data vector $\mathbf{x}_k$ onto a basis vector $\mathbf{u}_m, m = 1 \ldots M$.
- $\hat{\mathbf{X}}$ is the reconstructed version of the original trajectory matrix. If all $p_{mm} = 1, m = 1, \ldots, M$, the original matrix $\mathbf{X}$ is recovered because $\mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{I}$.

Alternatively, by applying a SVD decomposition to the trajectory matrix

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{V}^{\mathrm{T}}$$

the projections are obtained as

$$(3) \quad \mathbf{Y}_m = p_{mm}\mathbf{u}_m^{\mathrm{T}}\mathbf{X} = p_{mm}\lambda_m^{1/2}\mathbf{v}_m^{\mathrm{T}}.$$

The row vector $\mathbf{Y}_m$ (m-th row of $\mathbf{Y}$) is thus related with the m-th eigenvector $\mathbf{v}_m$ of the inner product matrix $(\mathbf{X}^{\mathrm{T}}\mathbf{X})$ of the data. So, the projections form a non-correlated representation of the original data.

2.1 Signal enhancement

From (2) we obtain the expansion of the reconstructed data matrix into a sum of rank-one matrices according to

$$(4) \quad \begin{aligned} \hat{\mathbf{X}} &= \mathbf{u}_1 p_{11}\mathbf{u}_1^{\mathrm{T}}\mathbf{X} + \mathbf{u}_2 p_{22}\mathbf{u}_2^{\mathrm{T}}\mathbf{X} + \cdots + \mathbf{u}_M p_{MM}\mathbf{u}_M^{\mathrm{T}}\mathbf{X} \\ &= \mathbf{u}_1\mathbf{Y}_1 + \mathbf{u}_2\mathbf{Y}_2 + \cdots + \mathbf{u}_M\mathbf{Y}_M \\ &= \sum_{m=1}^{M} \hat{\mathbf{X}}_m, \end{aligned}$$

where each element of the $1 \times K$ matrix $\mathbf{Y}_m$ is the dot product, between the m-th eigenvector and one of the columns of the data matrix, weighted by p_{mm}. As mentioned before $p_{mm} = 1$, if the m-th component is selected for the reconstruction otherwise $p_{mm} = 0$. These weights thus serve to select the components relevant for the problem at hand, which might be concerned, for example, with noise reduction or artifact removal.

In noise reduction applications, for example, the selected components correspond to the $m = 1, \ldots, L \leq M$ largest eigenvalues of the matrix $\mathbf{S}$. Then the eigenspectrum of $\hat{\mathbf{X}}$ is a truncated version of the eigenspectrum of the original data. Alternatively, it is also possible to re-scale the eigenspectrum of the reconstructed version by assigning a real value $p_{mm} \in \Re$ like, for instance, $p_{mm} = \sqrt{1 - \eta/\lambda_m}$,
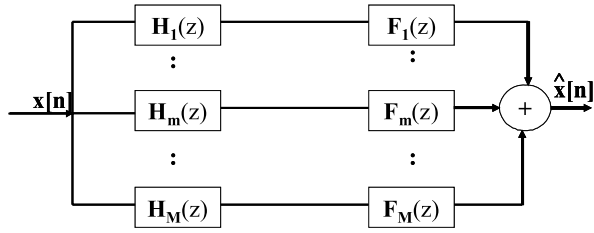
Figure 1. Filter Bank description of the processing chain: $H_m(z)$ are analysis transfer functions and $F_m(z)$ are synthesis transfer functions.

where η represents the noise variance (or energy). The eigenspectrum then scales according to the variance of the estimated noise. The amplitude of each projected component then changes according to the level of noise estimated:

$$(5) \qquad \mathbf{Y}_m = p_{mm}\lambda_m^{1/2}\mathbf{v}_m^{\mathrm{T}} = (\lambda_m - \eta)^{\frac{1}{2}}\mathbf{v}_m^{\mathrm{T}}.$$

Other possibilities to compute the weights of p_{mm} can be found, though all assume that the noise variance should be smaller than the selected eigenvalues [8, 10]. The mean of the discarded eigenvalues is often used as an estimate of the noise variance [14], but more sophisticated schemes have been considered also.

The reconstructed data matrix $\hat{\mathbf{X}}$, in general, does not exhibit identical elements along each descending diagonal like in case of the original trajectory matrix $\mathbf{X}$. In singular spectrum analysis (SSA), these distinct entries in each diagonal (or anti-diagonal) are replaced by their average in order to obtain again a Toeplitz (or an Hankel) matrix $\mathbf{X}_r$. A univariate reconstructed time series $\hat{x}[n]$ is then obtained by reverting the embedding, i.e. by forming the signal with the mean of the values along each diagonal of $\hat{\mathbf{X}}$.

2.2 SSA and filter banks

Signal enhancement as it was sketched above can also be addressed employing linear invariant systems theory. Seen from that perspective, the decomposition discussed above can be considered the result of the application of a bank of finite impulse response (FIR) filters [9], where analysis and synthesis filter pairs are connected in parallel (see Fig. 1). Hansen *et al.* [9] proposed an architecture where formally, both filters represent vectors which multiply trajectory matrices. They organize their input time series by embedding into an input trajectory matrix with Hankel structure. This matrix is common to all branches of the filter bank. At the output of each analysis filter they compute a second trajectory matrix which has a Toeplitz structure, instead. The latter type of matrix is chosen to achieve the diagonal averaging necessary to re-establish a Hankel or Toeplitz structure of the reconstructed data matrix as suggested in [3]. The coefficients of their analysis and synthesis filters correspond to the eigenvectors of the correlation matrix of the input trajectory matrix.

However, in the framework of linear invariant system theory, the filter bank structure needed to achieve the output time-series $\hat{x}[n]$ should be provided by the input time-series $x[n]$ instead of the trajectory matrices. Hence we propose an approach based on filter responses and related transfer functions rather than on matrix manipulations. The approach proposed in this work is summarized in Fig. 1. It represents the block diagram of the filter bank system where the transfer functions $H_m(z)$ of the analysis filters are related with the projection step onto the eigenvectors $\mathbf{u}_m$ of the subspace model and the transfer functions $F_m(z)$ of the synthesis filters are related with the reconstruction and diagonal averaging step. Therefore each parallel branch of the block diagram represents one summand of (4) with the diagonal averaging performed in each term. In the following we deduce closed form expressions for the transfer functions and discuss their properties.

Each row $\mathbf{Y}_m = p_{mm}\mathbf{u}_m^{\mathrm{T}}\mathbf{X}$ of the projected data matrix $\mathbf{Y}$ can be considered a filtered version of the original data sequence. But the dot product of the eigenvector with the sequence of lagged vectors $\mathbf{x}_k, k = 1, \ldots, K$ can be formulated similarly as the weighted sum of a sequence of samples of the time series. This operation, usually called convolution sum [12], is written as follows

$$(6) \qquad y_m[n] = p_{mm}\sum_{i=1}^{M} u_{im}x[n - i + 1],$$

where $(M - 1) \leq n < N$.

The row vector $\mathbf{Y}_m$ comprises $K = N - (M - 1)$ samples of the time series $y_m[n]$, starting with time index $(M - 1)$, much like in the first row of the trajectory matrix $\mathbf{X}$. The entries of the vector $\mathbf{u}_m$, the m-th column of the subspace model $\mathbf{U}$, correspond to the coefficients of a finite impulse response (FIR) filter.

The transfer functions $H_m(z)$ of the analysis filter can be computed by substituting in (6) every delay operation by the corresponding z transform. Therefore mapping $x[n]$ to $X(z) = \sum_{-\infty}^{\infty} x[n]z^{-n}$, $x[n \pm d]$ to $z^{\pm d}X(z)$ and $y_m[n]$ to $Y_m(z)$ [13], where z is a complex number, the filtering operation can be formulated as

$$(7) \qquad Y_m(z) = H_m(z)X(z),$$

whereby

$$(8) \quad H_m(z) = \frac{Y_m(z)}{X(z)} = p_{mm}\left(\sum_{i=1}^{M} u_{im}z^{-(i-1)}\right)$$
$$= p_{mm}(u_{1m} + u_{2m}z^{-1} + \cdots + u_{Mm}z^{-(M-1)}).$$

The transfer functions $H_m(z), m = 1, \ldots, M$ thus constitute the analysis block as it decomposes the input time series $x[n]$ into several component time series $y_m[n], m = 1, \ldots, M$.

After filtering out the subspace projections at the analysis step, these projections can be further processed by selecting only the problem-relevant components and discarding the rest. Afterwards the remaining problem-relevant projections need to be combined to the reconstructed signal at the synthesis step. To facilitate the exposition of the latter, the reconstruction step of one component $\hat{\mathbf{X}}_m$ of (4), which is given by

(9)

$$\hat{\mathbf{X}}_m = \mathbf{u}_m \mathbf{Y}_m$$

$$= \begin{bmatrix} u_{1m}y_m[M-1] & u_{1m}y_m[M] & \cdots & u_{1m}y_m[N-1] \\ u_{2m}y_m[M-1] & u_{2m}y_m[M] & \cdots & u_{2m}y_m[N-1] \\ u_{3m}y_m[M-1] & u_{3m}y_m[M] & \cdots & u_{3m}y_m[N-1] \\ \vdots & \vdots & & \vdots \\ u_{Mm}y_m[M-1] & u_{Mm}y_m[M] & \cdots & u_{Mm}y_m[N-1] \end{bmatrix}$$

is exemplified. Note that $\hat{\mathbf{X}}_m$ is a rank-one matrix as all its rows are linearly dependent. Obviously, each row is a scaled version of the same component time series $y_m[n]$. But the resulting matrix does not yet exhibit a Toeplitz structure which can be restored, however, by diagonal averaging. Interestingly, this reconstruction procedure can equally well be formulated as a convolution sum operation:

(10)
$$\hat{x}_m[n] = \frac{1}{M_d} \sum_{i=l}^{s} u_{im} y_m[n+i-1],$$

where M_d, l and s attain values according to the number of elements in the diagonals of the matrix defined in (9). More specifically, the response can be sub-divided into a transient and a steady state response according to the following distinction:

- With M elements, (10) represents a steady state response of the filter in case of $(M-1) \leq n \leq (N-M)$ with parameters $M_d = M$, $l = 1$ $s = M$.
- With less than M elements, (10) represents a transitory response of the filter in case of
  - $0 \leq n \leq (M-2)$ (lower left corner of the matrix) with parameters $M_d = n+1$, $l = M - M_d$ and $s = M$.
  - $(N-M+1) \leq n \leq (N-1)$ (upper right corner of the matrix) with parameters $M_d = N - n$, $l = 1$ and $s = M - M_d$

Note that the synthesis filter is an *anti-causal* FIR filter as the output at time index n depends on input samples with time indices $n+1, \ldots, n+M$. Both cases can be unified by formally setting $y_m[n] = 0$ for time indices $0 \leq n \leq (M-2)$ and $N \leq n \leq (N+M-2)$ and compute (10) as in the steady-state case. Therefore, the synthesis transfer function is given by:

(11)
$$F_m(z) = \frac{X_m(z)}{Y_m(z)} = \frac{1}{M} \sum_{i=1}^{M} u_{jm} z^{i-1}$$
$$= \frac{1}{M} (u_{1m} + u_{2m}z^1 + \cdots + u_{Mm}z^{(M-1)}).$$

Notice that the transfer functions of the analysis and synthesis filters differ by a scale factor and by the sign of the powers of z. Therefore the magnitudes of the frequency response of both filters are related by a scale factor and their phases are symmetric.

The transfer function of each parallel branch (see Fig. 1) resembles a cascade, formed by the projection step (analysis filter) followed by the reconstruction and diagonal averaging step (synthesis filter). Hence, the product of the two transfer functions reads

(12)
$$T_m(z) = \frac{\hat{X}_m(z)}{X(z)} = F_m(z)H_m(z) = \sum_{k=-(M-1)}^{M-1} t_{km} z^k.$$

The coefficients t_{km} are the coefficients of the product of two polynomials with the same coefficients but symmetric powers. It can easily be shown that $t_{km} = t_{-km}, k = 1, \ldots, (M-1)$. Therefore the related frequency response can be obtained by substituting $z = e^{j\omega}$, where $j = \sqrt{-1}$, in (12) which then leads to

(13)
$$T_m(e^{j\omega}) = t_{0m} + \sum_{k=1}^{M-1} 2t_{km} \cos(k\omega).$$

The frequency response is a real function, with period $\omega = 2\pi$, and it corresponds to a zero-phase filter. This means that each extracted component $\hat{x}_m[n]$ is always in-phase with its related original signal $x[n]$.

Examples of frequency responses of the eigenfilters are shown graphically in [9] but no analytic expressions of the filter responses are given. However, the present work also deduces closed-form analytical expressions for the analysis and synthesis filters. Note that frequency responses of the component filters $H_m(z)$ and $F_m(z)$ cannot be given in closed-form similar to $T_m(z)$ in (13) due to lacking symmetry properties of their coefficients [13].

In [1], the frequency profile of the eigenvectors was studied by way of the corresponding periodogram. The same goal can also be achieved by taking samples of $|H_m(e^{j\omega})|$ with $\omega = 2\pi \frac{k}{M}, k = 0, \ldots, M-1$. Note that because $H_m(e^{j\omega})$ and $F_m(e^{j\omega})$ only differ by a scaling factor and the sign of the power of the complex exponential argument, the resulting transfer function $T_m(e^{j\omega})$ is similar in shape to $|H_m(e^{j\omega})|$. Consequently, the value of $T_m(e^{j\omega})$ is given by the square of $|H_m(e^{j\omega})|$ divided by $p_{mm}M$.

The global transfer function of the block diagram represented in Fig. 1 is finally obtained by adding the transfer functions of the parallel branches. The resulting output $\hat{x}[n]$ is a sum of the selected signals $\hat{x}_m[n]$, e.g, the outputs of

the cascaded filter pairs formed by $H_m(z)$ and $F_m(z)$. Notice, that the embedding of the time-series as suggested by (1) leads to a *causal* filter for the analysis step and to an *anti-causal* filter for the synthesis step. Using alternative embedding procedures, this property of the filters can interchange.

3. COMPARING SUBSPACES

When applying subspace analysis it is frequently of interest to compare different subspace models. Given two matrices ($\mathbf{U}_A$ and $\mathbf{U}_B$) with L_1 and L_2 eigenvectors representing two subspaces ($\mathcal{U}_A$ and $\mathcal{U}_B$) of an M-dimensional data space, we consider the following normalized distance measure of the similarity of the two subspaces

$$(14) \qquad d_{AB} = \frac{d(\mathcal{U}_A, \mathcal{U}_B)}{\sqrt{\max(L_1, L_2)}}$$
$$= \sqrt{1 - \frac{\text{trace}(\mathbf{U}_A^{\mathrm{T}} \mathbf{U}_B \mathbf{U}_B^{\mathrm{T}} \mathbf{U}_A)}{\max(L_1, L_2)}}.$$

This subspace distance measure $d(\mathcal{U}_A, \mathcal{U}_B)$ was proposed in [16, 17], where its properties concerning non-negativity, symmetry and invariance to the choice of an orthonormal basis were discussed as well. Furthermore, it was proven in [17] that this subspace distance satisfies the triangular inequality. The normalized distance $d_{AB} \in [0,1]$ provides a practical means to perform simple decisions. It reflects the difference between the eigenvectors spanning the subspaces. If two subspaces nearly coincide, they will have a small distance, and if they are almost perpendicular to each other, they will have a large distance. Moreover, the distance (14) appears to be a quite effective subspace distance measure due to its clear geometrical meaning and easy computation. Therefore it has been applied in a number of applications in face, object and video based recognition [19]. The rule $d_{AB} \leq 1/2$ provides an intuitive decision whether or not two subspaces are similar.

In the present work we extend the application of this similarity measure to cluster subspace models in the signal enhancement framework. Signal enhancement is achieved by selecting a subspace in the M-dimensional space of time-delayed coordinates, i.e., by choosing a subset of the eigenvectors which span the original data space. Suppose that a subspace $\mathcal{U}_A$ with L_1 eigenvectors is computed for a segment of a time series and another subspace $\mathcal{U}_B$ with L_2 eigenvectors is computed for another segment of the same time series. We are then interested to know if the embedded time-series lead to subspace models with eigenvectors which spread in similar directions of the multidimensional embedding space. This problem fits well to the measure considered above and is studied experimentally in the next section.

4. ILLUSTRATIVE EXAMPLES

To illustrate the concepts discussed above, we perform numerical simulations for two time series:
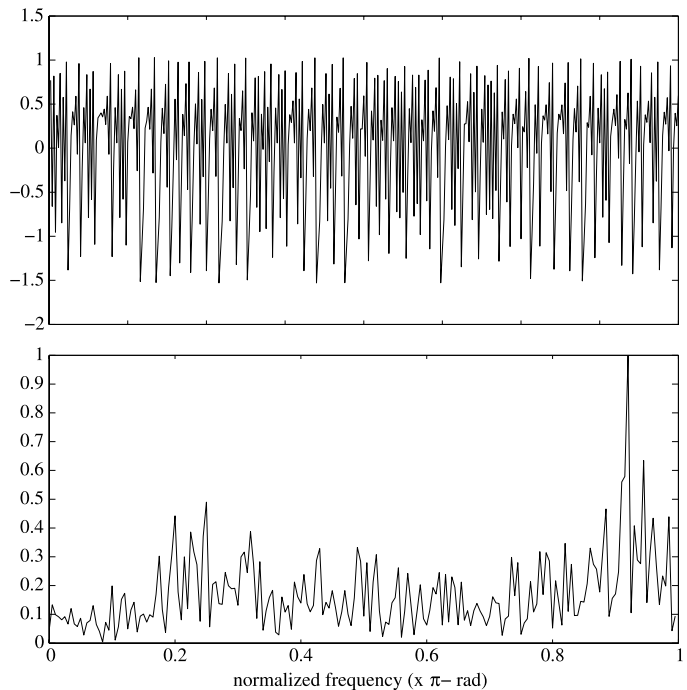

Figure 2. The Hénon time series (top) and its Discrete Fourier Transform (bottom). DFT absolute values normalized by the maximal value are shown in the lower figure.

- One nonlinear time series given by Hénon's equation.
- Electroencephalogram (EEG) signals collected during evoked potential studies.

The spectrum of the nonlinear Hénon time series spreads over all the frequency range with a slight predominance in the high-frequency band, while the EEG time series has its energy concentrated in the low-frequency band.

4.1 Hénon time series

The nonlinear Hénon time series is used to illustrate the frequency response of the filter and its relation with the embedding dimension considering that the sequence has a fixed length N. The time series results from the following dynamical equation

$$x[n+1] = 1 - ax^2[n] + bx[n-1],$$

where parameter values $a = 1.4$, $b = 0.3$ have been used in the following simulations. Fig. 2 shows $N = 400$ samples of the time series $x[n]$ and its corresponding Discrete Fourier Transform (absolute values). The spectrum of the time series is spread over the whole range of normalized frequencies and exhibits the largest amplitudes in the high frequency range close to $\omega = \pi$. The time series was analyzed applying different embedding dimensions M. The basis vectors $\mathbf{U} = (\mathbf{u}_1 \mathbf{u}_2 \ldots \mathbf{u}_M)$ (see Sec. 2) of the space of time-delayed coordinates correspond to eigenvalues ordered by decreasing value $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$. The corresponding
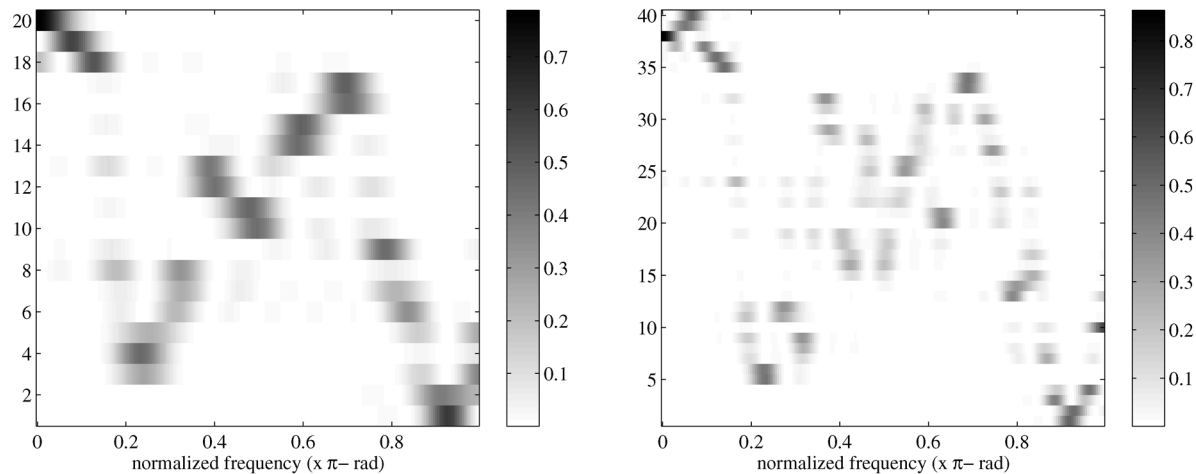
Figure 3. Frequency responses $T_m(e^{j\omega})$. Left: $M = 20$ filters, Right: $M = 40$ filters for the Hénon time series x[n].

frequency responses $T_m(e^{j\omega}), m = 1, \ldots, M$ (with $p_{mm} = 1$) of the branches of the analysis-synthesis system are shown in Fig. 3.

The frequency response of the first two filters of the filter bank with $M = 20$ parallel FIR filters are centered around $\omega = 0.9\pi$. Hence, they cover the high frequency band of the signal. The next region covered by both filter bank models is around $\omega = 0.2\pi$ which corresponds to the second largest peaks in the spectrum of the time series. Generally the pattern of frequency responses is very similar for different embedding dimensions as can be concluded from the examples given above. However, the bandwidth of the filters decreases as the embedding dimension M increases. Hence, the number of filters covering any specific region of the spectrum, i.e. any given frequency band, also changes.

By deliberately increasing the mean of the time series $x[n]$ by adding a constant amplitude $c = 3$ to $x[n]$, the Discrete Fourier Transform (DFT) exhibits an additional peak at $\omega = 0$ which is the strongest peak of the spectrum. Consequently, the filter corresponding to $m = 1$ turns from a bandpass filter centered at high frequency to a low-pass filter. Except from this, the general profile of frequency responses of the other filters is very similar in case of both input sequences $x[n]$ and $x[n] + c$, respectively.

We now compare two subspace models $\mathcal{U}_A$ and $\mathcal{U}_B$ corresponding to series $x[n]$ and $x[n] + c$, respectively. Fig. 4 shows the measures of similarity between these subspaces with dimensions $L_1 < M$ and $L_2 = L_1$. The largest normalized distance $d_{AB} = 1$ is achieved when the subspaces have dimension $L_1 = L_2 = 1$. When $L_1 = L_2 = 40$, both subspaces correspond to the complete space, hence their normalized distance is $d_{AB} = 0$. The two subspaces can be considered to differ substantially if $d_{AB} > 1/2$ holds, resulting in $L_1, L_2 < 5$. However, notice the abrupt change occurring when $L_1, L_2 = 38$ which obviously corresponds to the inclusion of the low-pass filter in the subspace model $\mathcal{U}_A$ (see Fig. 3).
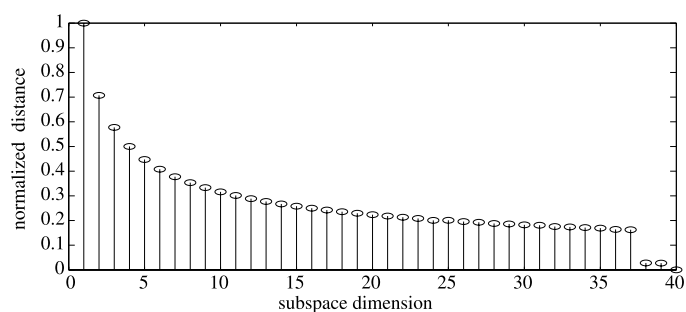

Figure 4. The dependence of the distance (14) on $L_1 = L_2$, the dimension of the subspaces, for the Hénon time series $x[n]$ and $x[n] + c$.

4.2 Evoked response potential signals

Evoked response potentials (ERPs) represent transient components in the electroencephalogram (EEG) generated in response to a stimulus, e.g. a visual or auditory stimulus. To render the ERP signal visible, a large number of single-trial responses are required to perform an ensemble average over all trials. However, this simple methodology has some drawbacks. The ERP averaging might not cancel some artifacts induced by eye movements or blinks if they are time-locked to the experimental events. Thus it is common practice in ERP studies to reject all EEG epochs contaminated with artifacts. The averaging process has the further drawback of masking the single-trial variability of the task-related responses with respect to amplitude or latency, for example. The fact that the same stimulus can elicit somewhat different signals has been discussed for decades. Despite these drawbacks, cognitive brain studies even consider a grand average to characterize dominant peaks. Grand average signals are taken over the ensemble averages obtained in several sessions with different participants. There-
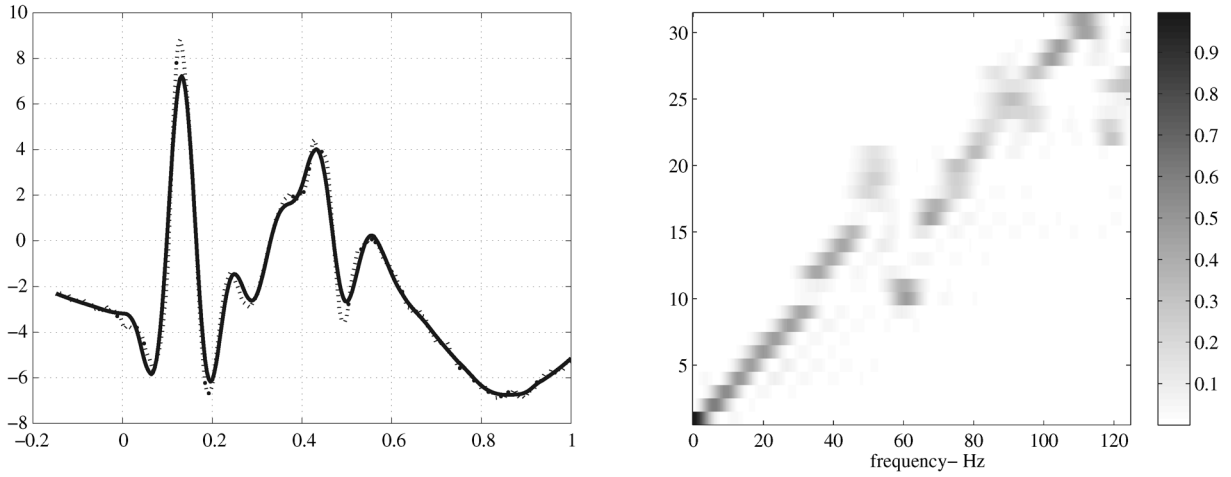
Figure 5. The grand-average (occipital channel) and frequency response of the filters $T_m(z)$. Left: Grand-average (dot) and the enhanced version (line) with the subspace model of the signal. Right: Frequency responses of $M = 31$ filters, the first $L_1 = 3$ filters form the subspace model.

fore, evoked response potential studies deal with signals that represent different levels of analysis: the single-trial segment, the ensemble average and the grand-average. The application of SSA to single-trial signals is mostly oriented towards signal enhancement. But it would be interesting to have some information about the signal under analysis. Subsequently, a projective subspace analysis of ERPs is pursued. The resulting subspace models of the single-trial signals are clustered using the subspace distance defined in (14). Ensemble averages are computed for every cluster and compared with the corresponding ensemble average of visually selected trials. Note that in spite of not being optimal, the ensemble average waveform is often the goal in practice or it constitutes a reference signal [11, 7] of a single-trial enhancement algorithm.

The data set analyzed comprises 32 sessions (two sessions per participant) with roughly 250 trials per session. Experimental protocol and data acquisition are fully described in [15]. The stimulus consisted of overlapping pictures of faces and houses. The participant's task was to determine, during each trial, if the relevant stimulus (house or face, depending on the condition) had the same identity as the relevant stimulus presented on the previous trial, i.e., if it was the same house or the same person. Disregarding any eventual differences between conditions, early potentials like the famous P100 response amplitude are clearly visible in ensemble averages, mostly in occipital derivations [15].

The ensemble average of a particular channel $x_s[n]$ of a session is estimated according to

$$(15) \qquad x_s[n] = \frac{1}{T} \sum_{t=1}^{T} x_t[n]$$
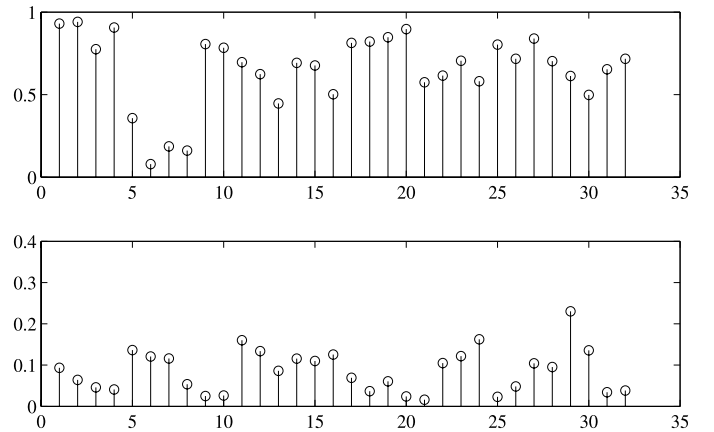
$$n = -B, -B+1, \ldots, 0, \ldots, R-1,$$

Figure 6. Grand-average versus ensemble average: correlation coefficients (top); distance between respective subspace models (bottom).

where t denotes the trial number and B and R stand for the number of samples before and after stimulus onset at $n = 0$. As referred to before, traditionally, the EEG is visually inspected and segments with excessive artifacts (for instance eye movement artifacts, drifts) are annotated. Therefore the ensemble average (15) includes only T single trials ERPs free of artifacts. For visualization purposes, the ensemble averages $x_s[n]$ are further averaged over the population of subjects to form the grand-average $x[n]$. Fig. 5 exemplifies such a grand-average signal of one occipital channel. The ERP oscillation is characterized by an initial negative peak (denoted as either N1 or N70, where 70 indicates $70ms$ time lag after stimulus onset), followed by a large positive peak (denoted as either P1 or P100) followed by another negative peak (denoted as either N2 or N170). In the $200ms$ window after
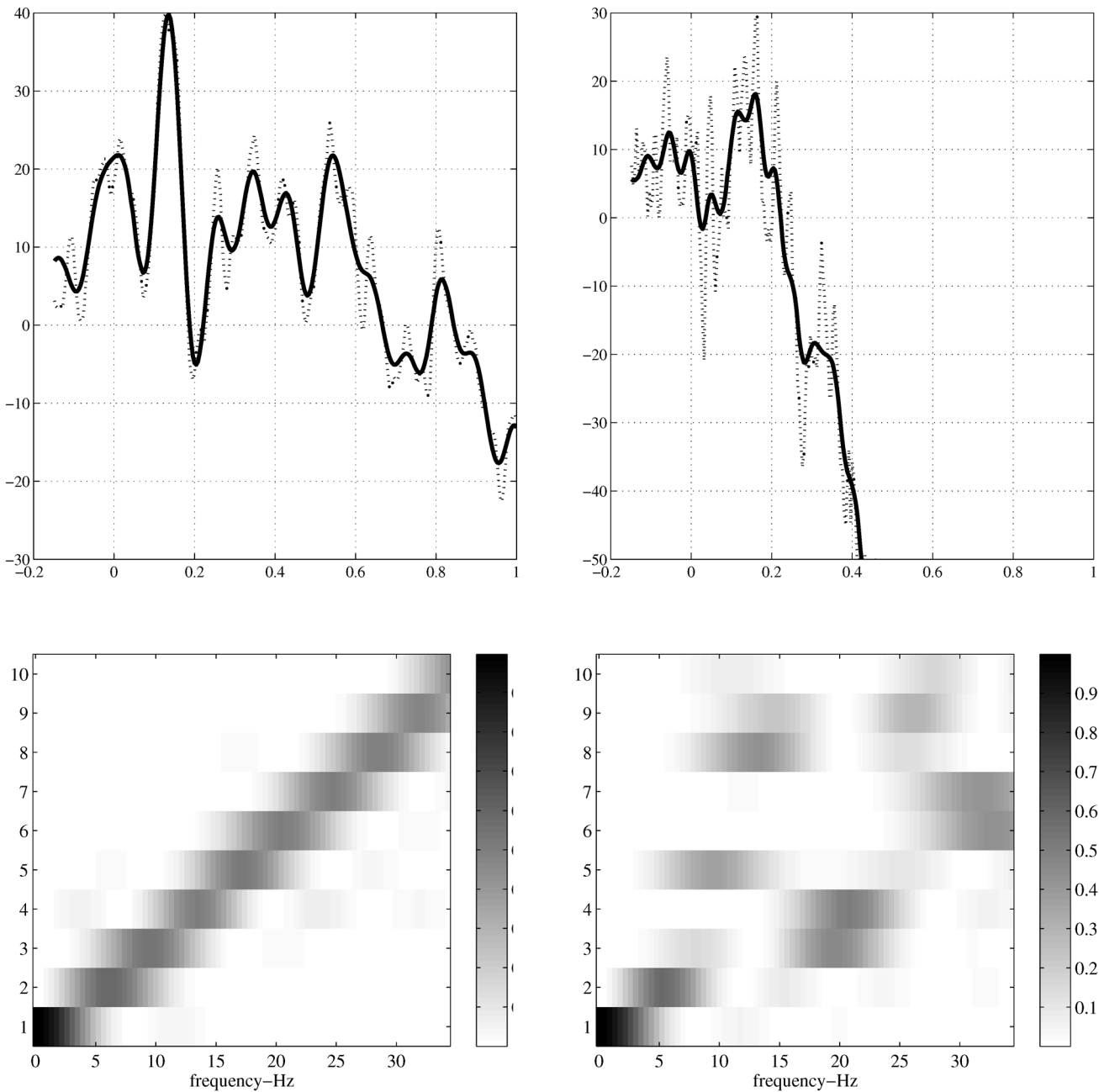
Figure 7. Single-trial analysis and frequency response of the filters. Top: input signal (dashed), output (line); Bottom-frequency responses of $L = 10$ filters.

the stimulus onset all these peaks are visible. However, as Fig. 5 illustrates, the $P100$ wave, appearing roughly $100ms$ after the stimulus, is the most prominent peak of ERPs. The ensemble averages $x_s[n]$ exhibit peaks with similar latencies. However, their amplitudes and the waveforms differ from session to session. Fig. 6 (top) presents correlation coefficients between $x[n]$ and each $x_s[n]$. Remarkably, these correlation coefficients are less than 0.5 in 4 of the cases.

SSA was applied to the single trial evoked potential and the subspace models were computed using a similar strat-

egy: the embedding dimension is chosen to be $M = 31$ and $N = 300$ samples ($B = 37$ before stimulus) are used to estimate **S**. From these signals ($x[n]$ and $x_s[n]$) subspace models were estimated with $L = 3$ dimensions, corresponding to $L_1 = L_2 = 3$ filters of the related filter bank model. The subspace models are then compared employing the normalized distance measure discussed before. From all the results shown in Fig. 6, globally, a similar conclusion can be drawn: when the correlation coefficient is high, the subspace distance is low. Note, however, the exception of the signal
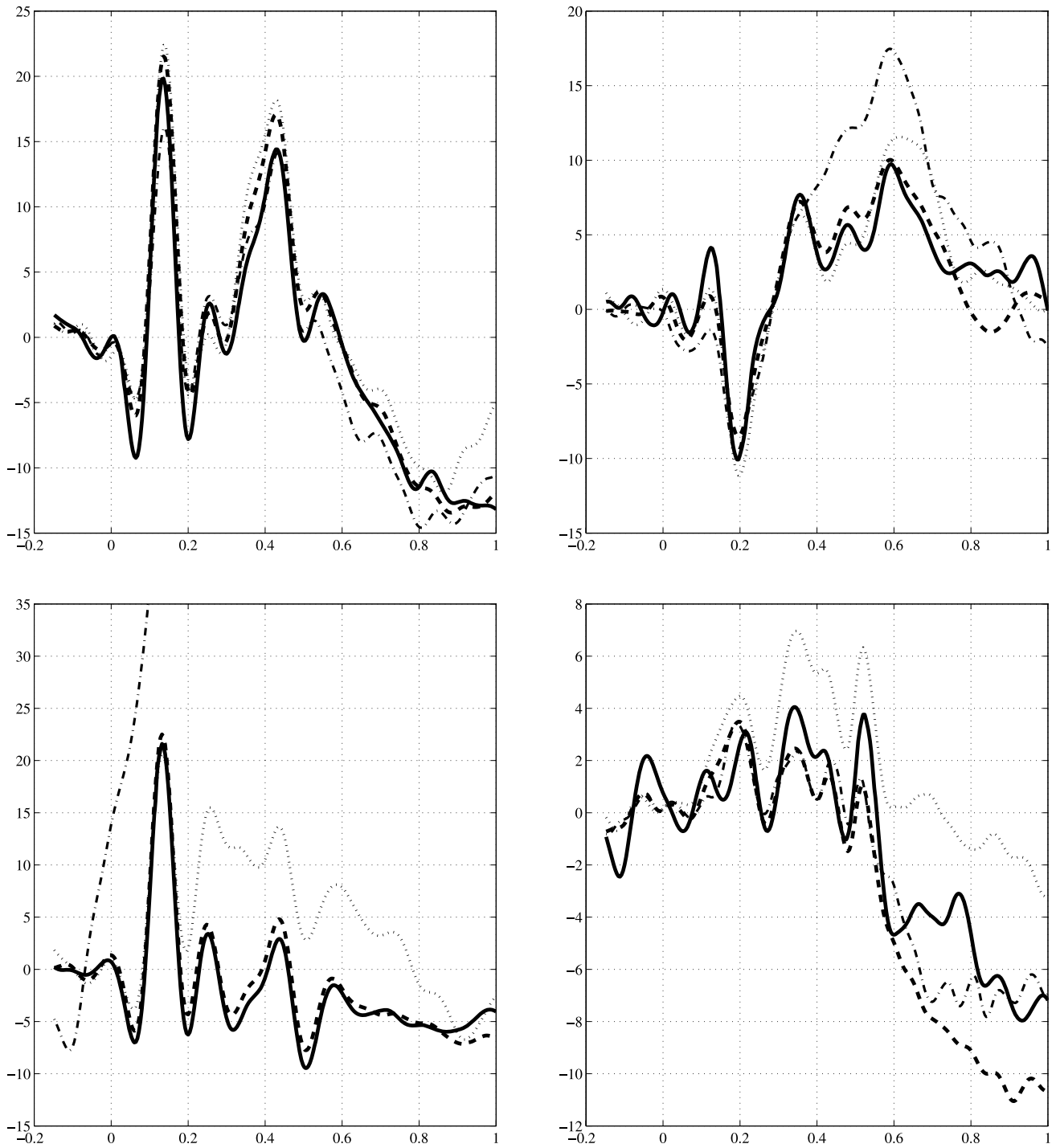
Figure 8. Ensemble averages of 4 *sessions: Dashed: Ensemble of visually selected trials; Ensemble averages in: Line: cluster A with $d_{AB} < 0.1$ and dot: cluster B with $0.1 \leq d_{AB} < 0.2$ and dash-dot: cluster C with $d_{AB} \geq 0.2$.*

corresponding to session 29 that exhibits the largest subspace distance and a large correlation coefficient. Computing a grand-average over the subject population, a smoother signal results from the subspace model with $L_1 = 3$. Fig. 5 shows the frequency response of the corresponding filters, where the first 10 filters have their bandpass increasing with

the frequency f. The input signals have a low-pass spectral profile decreasing in amplitude with f, and with highest percentage of energy up to $35Hz$.

The goal of the analysis is to achieve single-trial signal enhancement and, simultaneously, to perform an ensemble averaging of signals whose subspace models are considered

similar according to the chosen metric. For each single trial signal, irrespective of the presence of possible artifacts, a corresponding subspace model is estimated. The subspace model is applied to the signal and its enhanced version (output of the filter bank with $L_2 = 3$) is taken to perform ensemble averaging over the corresponding cluster.

To form the clusters, the grand-average signal was chosen as the corresponding reference model ($\mathcal{U}_A, L_1 = 3$) for comparison. Three clusters were considered to group the single-trial subspace models ($\mathcal{U}_B, L_2 = 3$). Clustering was effected using the normalized distance measure according to the following rules: a single-trial subspace model belongs to cluster A if its distance to the grand-average subspace model is $d_{AB} < 0.1$, and it belongs to cluster B if $0.1 \leq d_{AB} < 0.2$ or it belongs to cluster C if $d_{AB} \geq 0.2$.

Fig. 7 shows two single-trial signals and its enhanced versions and the corresponding frequency response of the filters. The first single trial signal, shown in the top left row, exhibits a pronounced peak at a latency of roughly $100ms$. The corresponding frequency responses of the filters, shown in the bottom left, have a profile similar to the filters of the reference model ($d_{AB} = 0.02$). The second single trial signal, shown in the top right row, has a drift starting at $t = 200ms$ after stimulus onset and the related filter output signal has no clearly defined peak close to $t = 100ms$. There is also a visible difference in the frequency response of the filters when compared to the reference model ($d_{AB} = 0.39$), the third filter of the subspace model moves to the $20Hz$ region.

Fig. 8 shows ensemble averages of the single-trial signals of each of the three clusters and of visually selected trials corresponding to 4 sessions (the number 2, 5, 10 and 29 of Fig. 6). The sessions 2 and 10 exhibit prominent $P100$ peaks with large amplitudes ($\simeq 20\mu V$) in two of the clusters and in the visually selected group (Fig. 8 on left). The cluster C of session 10 is formed with trials whose signals have artifacts and the ensemble average has no meaning. The session 5 has $P100$ peak with smaller amplitude but in the cluster A the peak is larger than in the other ensemble averages (Fig. 8 on right-top). Also notice that the ensemble average of the cluster C presents a slow wave with large amplitude after $t > 0.4s$ and the peak around $t = 0.1s$ is not positive. The session 29 has not a peak with the expected latency. The ensemble average of the visually selected trials has a peak close to $0.2s$ and cluster B has the peak with similar latency. However cluster A has a very low peak closer to $t = 0.1s$ but it is not the largest peak of the wave.

Whatever is the session there is always one ensemble average that is very similar to the ensemble average of visually selected trials. And in most of the cases the number of signals in the cluster is smaller than the number of trials visually selected. This justifies the application of SSA to enhance single-trial signals. The application of the distance measure provides ensemble averages of signals in different clusters. In this way it is possible to separate the contribution of possible artifacts and isolate them in a group like cluster C in most of the sessions. This aspect can be important to avoid the visual inspection of the recordings, a very time consuming task, usually performed before the ERP analysis.

5. CONCLUSIONS

The interpretation of SSA as a bank of filters can be useful to attain a more clear-cut insight into the outcomes of the method. By the frequency responses of the filter bank, corresponding to the basis vectors of the subspace, the frequency content of the different components can be easily attained. SSA filters are data adaptive and the relevance of one component to the energy of input signal is deduced from the corresponding eigenvalue. Moreover, the frequency profile of each component is determined only at the projection step. However, in order to get a component in phase with the input signal, the diagonal averaging is required. Therefore, by applying a linear invariant system theory approach, analytical expressions of the frequency response are deduced in this work. These results thus corroborate the properties of the SVD/SSA steps referred to in previous works [9, 3]. The possibility of having outputs that are in phase with the input is an important aspect in applications where measures have to be taken using a small number of samples (like evoked potential signals). This work also discusses the application of a similarity measure to compare subspace models of the single-trial signal enhancement task. The preliminary results in evoked potential studies are very promising and encouraging to proceed with a more quantitative evaluation of the results.

REFERENCES

[1] ALEXANDROV, T., GOLYANDINA, N. and SPIROV, A. Singular spectrum analysis of gene expression profiles of early drosophila embryo: exponential-in-distance patterns. *Res. Let. Signal Proc.*, 2008:1–5, 2008.

[2] ALEXANDROV, T. A method of trend extraction using singular spectrum analysis. *REVSTAT*, 7(1):1–22, 2009.

[3] DOLOGLOU, I. and CARAYANNIS, G. Physical interpretation of signal reconstruction from reduced rank matrices. *IEEE Transactions on Signal Processing*, 39(7):1681–1682, 1991.

[4] EPHRAIM, Y. and VAN TREES, H. L. A signal subspace approach for speech enhancement. *IEEE Transactions on Acoustic, Speech and Signal Processing*, 3(4):251–266, 1995.

[5] GHIL, M., ALLEN, M. R., DETTINGER, M. D., IDE, K. et al. Advanced spectral methods for climatic time series. *Reviews of Geophysics*, 40(1):1–41, 2002.

[6] GOLYANDINA, N., NEKRUTKIN, V. and ZHIGLJAVSKY, A. *Analysis of Time Series Structure: SSA and Related Techniques.* Chapman & HALL/CRC, 2001. MR1823012

[7] HAJIPOUR, S., SHAMSOLLAHIL, M. B., MAMAGHANIAN, H. and ABOOTALEBI, V. Extracting single trial visual evoked potentials using iterative generalized eigen value decomposition. In *IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2008*, pages 233–237, 2008.

[8] HANSEN, P. C. and JENSEN, S. H. Subspace-based noise reduction for speech signals via diagonal and triangular matrix decompositions: Survey and analysis. *Eurasip Journal on Advances in Signal Processing*, Vol 2007, 2007. MR2366752

[9] HANSEN, P. C. and JENSEN, S. H. FIR filter representations of reduced-rank noise reduction. *IEEE Transactions on Signal Processing*, 46(6):1737–1741, 1998.

[10] HERMUS, K., WAMBACQ, P. and VAN HAMME, H. A review of signal subspace speech enhancement and its application to noise robust speech recognition. *Eurasip Journal on Advances in Signal Processing*, 2007.

[11] IYER, D. and ZOURIDAKIS, G. Single-trial evoked potential estimation: Comparison between independent component analysis and wavelet denoising. *Clinical Neurophysiology*, 118:495–504, 2007.

[12] JACKSON, L. B. *Signals, Systems and Transforms.* Addison-Wesley, 1991.

[13] JACKSON, L. B. *Digital Filters and Signal Processing.* Kluwer Academic Publishers, 1996.

[14] LIAVAS, A. P. and REGALIA, P. A. On the behavior of information theoretic criteria for model order selection. *IEEE Transactions on Signal Processing*, 49(8):1689–1695, 2001.

[15] SANTOS, I. M., IGLESIAS, J., OLIVARES, E. I. and YOUNG, A. W. Differential effects of object-based attention on evoked potentials to fearful and disgusted faces. *Neuropsychologia*, 46(5):1468–1479, 2008.

[16] SUN, X. and CHENG, Q. On subspace distance. In A. Campilho and M-Kamel, editors, *ICIAR 2006*, volume LNCS 4142, pages 81–89. Springer-Verlag, 2006.

[17] SUN, X., WANG, LIWEI and FENG, J. Further results on the subspace distance. *Pattern Recognition*, 40:328–329, 2007.

[18] TEIXEIRA, A. R., TOMÉ, A. M., LANG, E. W., GRUBER, P. and MARTINS DA SILVA, A. Automatic removal of high-amplitude artifacts from single-channnel electroencephalograms. *Computer Methods and Programs in Biomedicine*, 83(2):125–138, 2006.

[19] WANG, L., WANG, X. and FENG, J. Subspace distance analysis with application to adaptive bayesian algorithm for face recognition. *Pattern Recognition*, 39(3):456–464, 2006. doi: DOI: 10.1016/j.patcog.2005.08.015.

A.M. Tomé
DETI, IEETA (Signal Processing Laboratory)
Universidade de Aveiro, 3810-193 Aveiro, Portugal
E-mail address: ana@ua.pt

A.R. Teixeira
DETI, IEETA (Signal Processing Laboratory)
Universidade de Aveiro
3810-193 Aveiro, Portugal
E-mail address: ateixeira@ua.pt

N. Figueiredo
DETI, IEETA (Signal Processing Laboratory)
Universidade de Aveiro, 3810-193 Aveiro, Portugal
E-mail address: nuno.figueiredo@ua.pt

I.M. Santos
Dep. Ciências de Educação
Universidade de Aveiro
3810-193 Aveiro, Portugal
E-mail address: isabel.santos@ua.pt

P. Georgieva
DETI, IEETA (Signal Processing Laboratory)
Universidade de Aveiro, 3810-193 Aveiro, Portugal
E-mail address: petia@ua.pt

E.W. Lang
CIML Group, Biophysics
University of Regensburg
93040 Regensburg, Germany
E-mail address: elmar.lang@biologie.uni-regensburg.de