MATHEMATICAL BIOSCIENCES
AND ENGINEERING
Volume **5**, Number **2**, April **2008**

http://www.mbejournal.org/

pp. **261–276**

AN ENVIRONMENT FOR COMPLEX BEHAVIOUR DETECTION IN BIO-POTENTIAL EXPERIMENTS

Maide Bucolo, Federica Di Grazia, Luigi Fortuna
Mattia Frasca, Francesca Sapuppo

Dipartimento di Ingegneria Elettrica, Elettronica e dei Sistemi, Universita degli Studi di Catania
Viale A. Doria 6, 95125 Catania, Italy

(Communicated by Stefano Boccaletti)

Abstract. We propose BioS (Bio-potential Study) as a new virtual data analysis and management environment.It was devised to cope with the physiological signals, in order to manage different data using advanced methods of analysis and to find a simple way to decode and interpret data. BioS has been structured as a flexible, modular, and portable environment. It includes several modules as data importing and loading, data visualization (1D, 2D, 3D), pre-processing (frequency and saturation filtering, statistical analysis), spatio-temporal processing such as power spectrum, independent component analysis (ICA) in spatial and time domain, and nonlinear analysis for the extraction of the maximum Lyapunov exponent and d_∞(d-inifnite) using optimized algorithms. The environment provides a user-friendly Graphic User Interface that allows inexperienced users to perform complex analyses and to speed up experimental data processing.

1. **Introduction.** The term "biosignal" can be used for any measurable signal produced by the human body. The measurable quantity can be electrical, magnetic (EEG, MEG respectively), or may be represented by changes in pressure or volume such as heart rate, nerve activity, renal blood flow, arterial pressure, or respiratory signals. The complex variability of physiological control systems is evident in these dynamic processes and this behaviour is motivated by the concept that physiological functions require an integration of complex networks of control systems, e.g., feedback loop, and other regulatory mechanisms enabling a body to perform a variety of activities necessary to survive. The control systems of the human body exist at molecular, subcellular, cellular, organ, and systemic levels of organization. The continuous interplay among the electrical, chemical, and mechanical components of these systems ensures that information is constantly exchanged, even if the body is at rest [18].
The biological time-series analysis is necessary to identify hidden dynamical patterns, which could be important in revealing information on the physiological mechanisms [16].

Achievements in the development of new tools for the acquisition and the use of new methodologies are of great advantage for the study of bio-potentials presenting

2000 *Mathematics Subject Classification.* 92C55.
Key words and phrases. Nonlinear Analysis, Independent Component Analysis, d-infinite (d_∞).
This work was partially supported by the EU under COST ACTION B27 ENOC.

high temporal and often high spatial resolutions. The technological and methodological efforts are made in the attempting to define and discriminate dynamical patterns in normal resting activity, in the effects of sensory stimulations, and in different pathological states [13][19][11][12]. The possibility of continuous monitoring of human activities, made possible by decreasing the memory's price, and the variety of biopotentials recorded both in space and time decreasing, have led an efficient data management strategy for such increasing amounts of data, as weel as the use of analysis methods to support the diagnosis processes decoding the data.

Numerous tools for biological signals analysis are well described in recent studies; they are used for different applications and most of them are developed under the Matlab environment; examples of commercial solutions [31] are PRANA, EMSE, while some free-ware products [32] are BrainStorm [33] and EEGLAB[7].
The platform BioS (Bio-potential Study) proposed here was devised to deal with the issues and requirements that both researchers and inexperienced users face in the analysis of bio-potentials. The aim was to create a modular and portable environment that could use well-known methodologies and examine new analysis approaches, being both a decision supporting system and a research offline virtual instrument. BioS was implemented in Matlab7 as an integrated development environment (IDE) for its flexibility, and the possibility of programming by scripts using pre-existing toolboxes. The environment provides three main modules: loading, visualization and analysis.
The loading module, described in section 2, was devised to import different types of data depending on their formats (binary, ASCII, or Matlab), their physiological origin (MEG, EEG, ECG, EOG) and medical instrumentations.
The visualization module allows for temporal, spatial, and spatio-temporal plots of raw data and analysed data, according to the different geometries adopted by the acquisition systems, as described in section 3.
The analysis module consists of two phases: the pre-processing and the processing. The pre-processing functions, reported in section 4, include statistical analysis for a first global and rapid look at the signals and frequency and saturation filtering for the extraction of artefacts and for the detection of spurious disturbs and data corruption due to instrumentation. The pre-processing methods represent a preparatory step for the processing phase included in BioS, consisting of linear and nonlinear analyses. The advantages of linear methods lay in their simple implementation and an interpretation of the results that is relatively straightforward. The two features dedicated to the linear analysis are for extraction of the power distribution and the Independent Component Analysis (ICA) approach [4][9]. The ICA is attempt to separate data into maximally independent groups, in time or space, yielding temporal-ICA (TICA)[1] and Spatial-ICA (SICA)[8][5]. An ad hoc clustering procedure has also been implemented for a quantitative comparison of Independent Components (ICs) founded in different experiments.
In recent studies, several features of nonlinear approaches have been proposed to detect important properties of the physiological phenomenon.

The bio-potential signals are frequently characterized by such standard nonlinear indicators as the maximum Lyapunov exponent (λ), the asymptotic distance d_∞(d-infinite) and the correlation dimension [30][17][20]. In the theoretical determination of the λ and d_∞, it is fundamental the knowledge of the infinite difference equations in the discrete domain function as the differential equations in the continuous domain. An example of the evaluation of d_∞ for known nonlinear systems has been

reported, particularly, in relation to the Chua's circuit [2][3]. When the laws of the systems under study are unknown and only experimental data are available, the need arises for a calculation of the asymptotic distance d∞ for generic time series. The computation of these parameters on time series involves the concepts of time-delay embedding. So the reconstruction of the state-space requires the determination of a time lag and an embedding dimension [23].Here, a new approach for the evaluation of λ and d_∞ parameter is presented and it has been implemented with a computationally optimized algorithm (DivA - Divergence Algorithm). In addition, BioS supports a graphic user interface (GUI) designed to allow inexperienced Matlab users to apply advanced signal processing techniques to their data, an example of the main GUI and its functionality is shown in Figure 1.

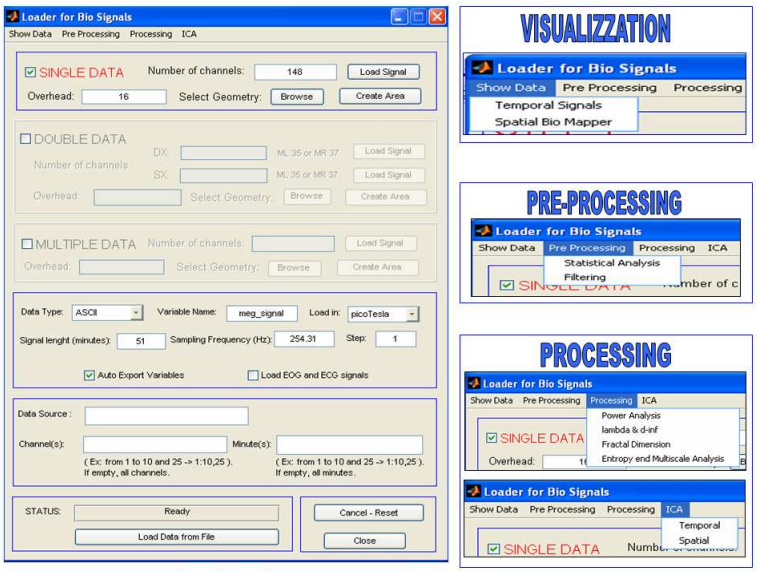


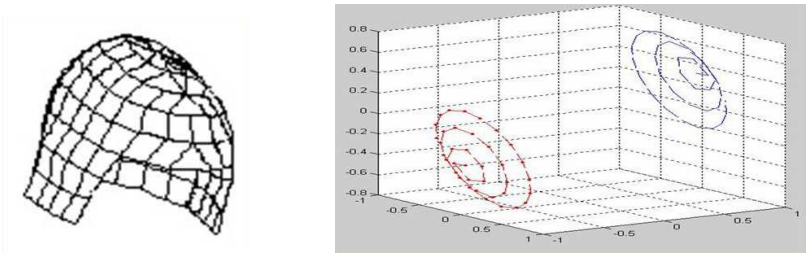
FIGURE 1. BioS main Graphic User Interface and its actual potentialities.

2. **Loading and data structures.** Biopotential data are usually recorded by different physiological systems adopting different numbers of acquisition channels organized with particular spatial distributions, therefore various geometries can be loaded using ad hoc protocols. Two opportune files containing the coordinates of the sensors (vertexes.mat) and the links existing between each channel and its three nearest neighbours (faces.mat) have to be provided with the data set to be loaded. The environment allows customers to load *single data*, as whole head for brain signals, *double data*, as two different sets of channels, one related to the left and one to the right cerebral hemisphere. Moreover the *multi data* loading, for the simultaneously recordings of nasal cycle, blood pressure, EEG, etc, has been projected for a future development.

Due to the wide range of acquisition instrumentations available in the market, the main issue dealing with the implementation of the loading system is the parameterization of several characteristics of the acquisition protocols and the bio-potential signals.

The structure "Bios-info" has been created to store variables related to data set, as the number of data storage files (single data, double data, or multiple data), the number of the acquired channels, the length of the recorded data (in minutes), the sampling frequency of the acquired data, and some variables which are specific for each analysis, such as the work-space variable name, the chosen channel/s and minute/s for the analysis. Currently it is possible to acquire signals in different data format as binary (Little-endian or Big-endian), ASCII, and Matlab. The Bios-info structure can be accessed directly from the Matlab command line to extract processing history and data info. The analysis results are given in a Matlab type and can be saved in different format as required by the user.

The environment has been tested on Magneto-encephalography (MEG) data acquired using two different systems: whole-head 148-channels (4-D Neuroimaging, San Diego, California) MEG instrument and dual 37-channel bio-magnetometers (BTI, Inc.) both located at The Scripps Research Institute (La Jolla, CA). A 3D representation of the MEG head channels distribution is shown in Figure 2.

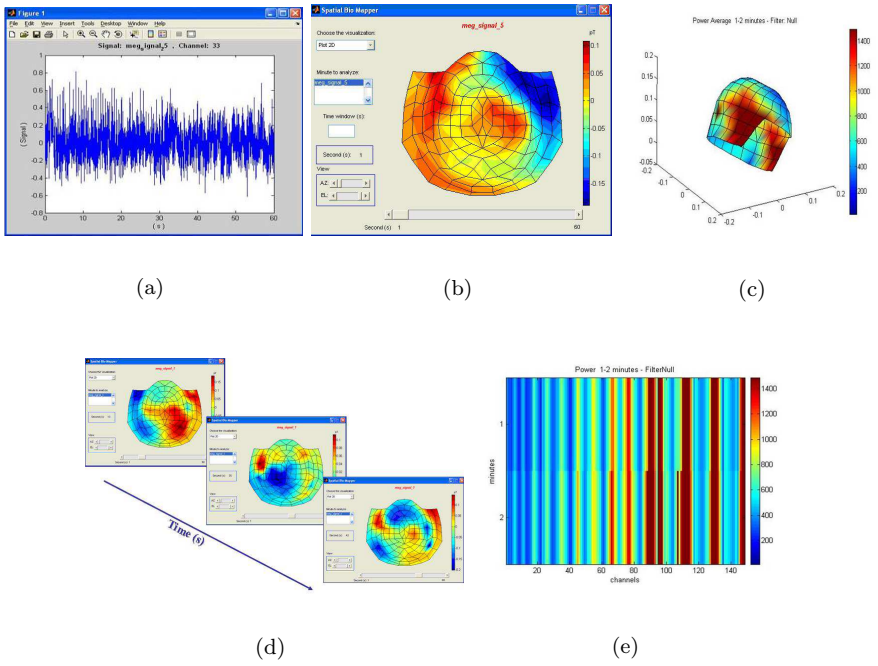


(a) (b)

FIGURE 2. Head channel distribution for:(a) whole-head 148 channels (b) dual 37channel bio-magnetometers.

3. **Visualization.** Different visualization modalities: temporal, spatial and spatio-temporal plots can be performed using the loaded files vertexes.mat and faces.mat. This represents a great potential available not only for the raw data but also for the pre-processed and processed data.The visualization in time domain is allowed by the traditional one-dimensional representation, thus, fixing a channel the signal dynamics in time is obtained (Figure 3 a). In the spatial color coded plots, 2D (Figure 3 b) and 3D (Figure 3 c), the value of the acquired channel in a unit of time is represented at the i^{th} node of the mesh corresponding to the digital reconstruction of the analyzed area, where $i = 1, 2, 3, \ldots, n$ (n is the number of channels). The intermediate areas colour representation is obtained through linear interpolation between neighbouring nodes.

As shown in Figure3 b, the sliding bar at the bottom allows to visualize 60 head maps related to the average value over each second of a single minute, and three different head maps over a second have been shown in Figure 3 d. Furthermore the parameterization of the averaging time interval allows us to push the study to the millisecond range, a time scale suitable for neuronal activity studies.
The complexity of Bio-potential data requires an environment that allows us to



(a) (b) (c)

(d) (e)

FIGURE 3. (a) One Dimensional visualization (b) Two dimensional visualization (c) Three dimensional visualization. (d) Two dimensional whole head map in time (e) Spatio-temporal map.

manage a big amount of data with high spatial and temporal resolution. Therefore it is necessary to have a procedure devoted to the visualization of the signal trends for the simultaneous comparison of multiple channels and for a first look at the data to establish the analysis strategy to be followed. For this reason, in addition to the conventional spatial maps, a spatio-temporal map (Figure 3 e), for a fixed time window, is available in BioS to visualize arising patterns in biopotentials.
Figure 3 e illustrates, as an example, the evolution in a two-minute interval of the whole-head channels value averaged over a one-minute slot. Thus, the colour of the area *(j, i)* represents the color coded value of the MEG signal averaged over the j^{th} minute for the i^{th} channel, so the image's i^{th} column represents the time evolution of the i^{th} channel by one-minute time-windows. The advantage of this representation is the possibility of having a complete view of the evolving parameters under study for all channels, without losing the spatial information. Such spatio-temporal representation can be extended to all the analysis results applied on the signals.

4. **Preprocessing.** The preprocessing module is organized in two sections: the statistical analysis, devised to apply multi-statistical methods, and *filtering*, including frequency filtering, to remove known artefacts or to extract signal frequency features, and saturation filtering in order to get rid of spurious values due to the instrumentation.
In the *statistical analysis* section three functional panels are provided: the *inside*

signal, the *between signals* and the *among signals*. The first allows the characterization of the signal related to one channel providing the extraction of the mean, the standard deviation, the histogram (Figure 4 a) and its state-space representation (Figure 4 d). The between signals examines the correlated event activity of two different channels related to the same minute or to two different ones. This panel extracts information related to the cross-correlation function (Figure 4 b) between two channels such as the maximum cross-correlation value, the respective delay, the value at zero lag, and the state space representation of one channel versus the other. In order to characterize the spatio-temporal behavior of all channels, over a selected time range, the *among signal* panel allows us to create a spatial map and the histograms related to the means and standard deviations of each channel. Moreover, through this panel the dynamics of one channel with respect to the others can be examined, obtaining the spatial maps for the three parameters of the cross-correlation (maximum value, the respective delay, the value at zero lag). In Figure 4 d an example of spatial map visualization related to cross-correlation maximum values of the channel 10^{th} with respect to the others is shown.

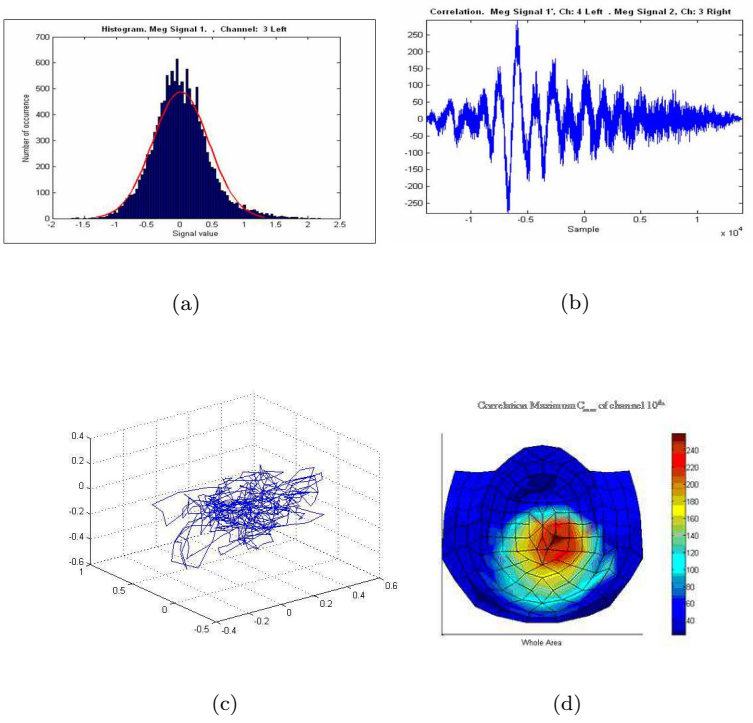
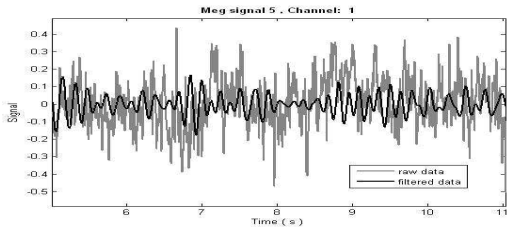


FIGURE 4. (a) Signal histogram (b) Cross-correlation analysis between two signals (c) Signal State Space representation of 10^{th} channel (d) Spatial map of the maximum of the correlation between the channel 10^{th} and the others.

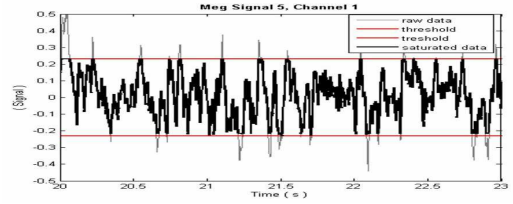
The *Frequency* section provides the possibility of projecting different custom filters or neural activity band (alpha, beta, gamma, delta, theta) filters [28]. They

can be tested on chosen channel/s and then extended to all data sets, simultaneously. An example, concerning channel 1 of the 5^{th} minute, is shown in Figure 5 a, where the grey line represents the dynamics of raw time series, while the black line is the signal filtered by the theta band filter.

Saturation filtering is suitable for channels that present spurious values. Its implementation permits us to chose the threshold over which the signal can be considered saturated and the new value to be assigned to the saturated points (Figure 5 b). The filtered signals are then available in the workspace to be fed into the preprocessing functions.



(a)



(b)

FIGURE 5. (a)Frequency filtering of a raw data (b) Saturation filtering of a raw data.

5. **Processing.** Three analysis methods were integrated in the processing module: two belonging to linear methods, the power distribution and the independent components analysis [19][15] and one concerning the nonlinear analysis methods for the extraction of the maximum Lyapunov exponent (λ) and asymptotic divergence between trajectories d_∞(d-infinite) [30]. All the methods are extended both to the spatial and temporal domain and can be applied both on raw data and pre-processed signals.

5.1. **Linear Analysis.**

5.1.1. *Power analysis.* This traditional approach is used to characterize the distribution of the power on the scalp using the maximum of the autocorrelation functions for all channels.
The autocorrelation function, represented in Equation (1) as $C_i(h)$, is calculated on

one-minute time series (N samples) for each channel ch_i and the value in zero $C_i(0)$ represents the power of the signal in that minute.

$$C_i(h) = \frac{1}{N} \sum_{k=1}^{N-1} ch_i(k) ch_i(k+h) \tag{1}$$

The power evaluated over a one-minute period is reported in Figure 6 in the spatial map visualization, meanwhile the spatio-temporal maps for a two minutes time window, minute 1 and 2, is shown in Figure 7.

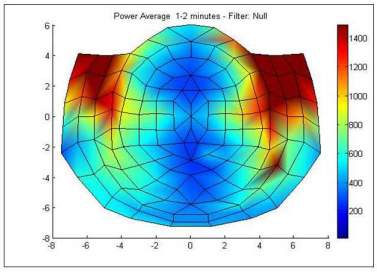


FIGURE 6. Spatial power map.

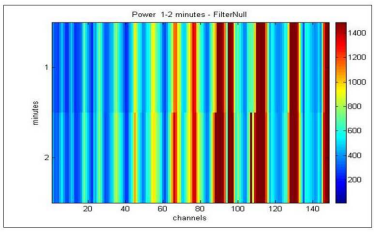


FIGURE 7. Spatio-Temporal map of a whole- head geometry.

5.1.2. *Independent Component Analysis (ICA).* ICA has become a popular mathematical method for processing data sets in biomedical signal processing. It is used to solve the blind source separation (BSS) problem, separating data into maximally independent groups in time or space, using respectively temporal-ICA (TICA) and spatial-ICA (SICA) [4]. Here, ICA is used in the temporal domain to extract common temporal features (TICs), while in the spatial domain it is used to isolate Spatial Independent Components (SICs) that can be considered as Spatial Modes (SMs).

For ICA analysis an existing Matlab tool, FastICA [9], has been used and integrated into the BioS environment. In Figure 8 the BioS output for TICA analysis over one minute is reported and it shows a clear presence of the artefact related to the heart beat signal as IC_1.

ICA assumes that the observed data X is a linear combination of underlying independent components $S \in \Re^{P \times Q}$. The proposed approach is described here by considering a matrix $X \in \Re^{N \times Q}$

$$X = WS \tag{2}$$

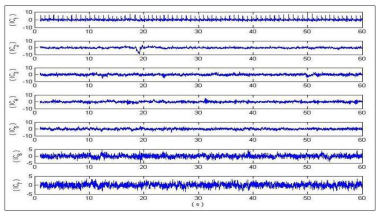


FIGURE 8. Temporal independent component analysis.

where S is an unknown source vector and the matrix $W \in \Re^{N \times P}$ is an unknown real-valued mixing matrix where P are the number of recovered ICs up to N.
Each IC is statistically independent from the others and has a non-Gaussian distribution. The matrix S can be evaluated by considering the estimated matrix $\hat{W}$ as follows:

$$S = \hat{W}^{-1} X. \tag{3}$$

The algorithm to estimate $\hat{W}$ is based on a recursive procedure [5] that includes two pre-processing stages: Centring and Whitening. The first stage makes all measurements at zero mean, and in the second one the Singular Value Decomposition (SVD) of the covariance matrix X is applied in order to have uncorrelated data with variance equal to one. Consequently, the value of P can be fixed by focusing on ICs that have the largest eigenvalues in the SVD.

In the case of TICA approach, the dimension N corresponds to the number of channels and Q is the number of time samples. Meanwhile for SICA method the number of time samples correspond to the N dimension and the Q represents the number of channels.
Concerning the SICA analysis, for a mathematical and computational issue, in order to solve N equation in Q variables with $N >> Q$, it is necessary to split the usual time window (i.e. one minute) in time sub slots and to extract from each of them the SICA components. To have a unique representation of the whole time window (one minute), it is necessary to group the SMs obtained for each sub slot, for this reason a suitable clustering strategy has been integrated in BioS in order to obtain a set of SMs that can represent the whole time window (minute). The clustering procedure is iterative and starts from a number h of classes that represent all the SICs found by analyzing each subset. Each class is characterized by a reference spatial mode (rSM) $\hat{s_i}$, for *i=1,. . . ,h*, chosen as the highest uncorrelated SM in the class. Each procedure step redistributes all SMs in the classes and updates their rSMs. In particular, each spatial mode s is correlated to all the classes by evaluating the membership indexes $m_1, \ldots, m_h$. The membership index m_i is defined by considering it as the absolute value of the cross-correlation, between the spatial mode s and the reference spatial mode $\hat{s_i}$ at lag zero:

$$m_i = \left| \sum_{n=1}^{Q} s(n) \hat{s_i}(n) \right|. \tag{4}$$

The spatial mode s is assigned to the i^{th} cluster that has the highest value of the membership index according to the following equation:

$$m_i \geq t_1 \tag{5}$$

where t_1 is the similarity threshold, increased up, by a step Δt_1, to a fixed value. The rSM of the winning class is then updated by summing its old value with the product of the learning velocity α and the classified spatial mode s:

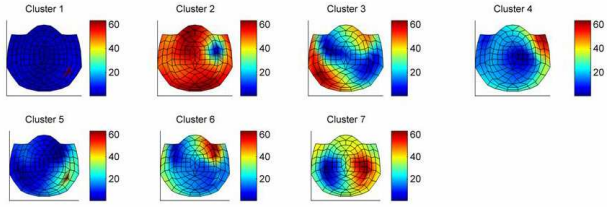
$$\hat{s}_{i,new} = \hat{s}_{i,old} + \alpha s. \tag{6}$$

Running the clustering procedure step, each SM is then attributed to a class. When all SMs are classified, the total number of classes is modified by means of the separation threshold t_2. The separation threshold t_2 is defined to reduce the number of clusters. If the rSM of the j^{th} and i^{th} classes verify the following equation:

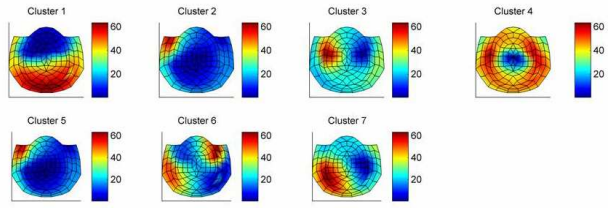
$$m_i = \left| \sum_{n=1}^{Q} s(n)\hat{s}_i(n) \right| \geq t_2, \tag{7}$$

the clusters are merged and the rSMs are averaged. All the iterative procedures restart the classification of the SMs according to the modified values of the parameters: the number of classes and the rSMs, and stops when the similarity threshold t_1 reaches the desired value.

The software also provides a quantitative comparison of ICs found in time or space in two different experiments through a second cluster algorithm based on the traditional cross-correlation analysis. Thus, the output of this clustering procedure consists of a set of common temporal/spatial ICs that represent the main features of the compared experiments dynamics. Figure 9 shows how this cross-correlation based approach permits us to extract Similarity Modes(a) and Dissimilarity Modes(b) in a comparison of two different minutes.



(a)



(b)

FIGURE 9. (a) Similarity modes (b) Dissimilarity modes in Biopotentials patterns.

5.2. **Nonlinear analysis.** The nonlinear nature and space extension of the biological systems under investigation create physiological processes analysis similar to the one associated to the extended complex systems theory.
In the present work, DivA (the acronym for Divergence Algorithm), an alternative methodology to evaluate the trajectory divergence and to calculate the maximum Lyapunov exponent λ and the asymptotic distance d_∞ in experimental data, is adopted. This implementation is computationally less onerous than the conventional ones, since it is not based on the time-delay embedding concept and no intermediate computational steps are needed to obtain the final result. Therefore, the procedure evaluates the d_∞ as the asymptotic value of the average distance between trajectories that are directly extracted from the time series. This algorithm is particularly suitable when coping with extended datasets, both in time and in space.
The proposed method has been already applied to MEG signals in order to characterize the occurrence of spatiotemporal nonlinear patterns [28].

5.2.1. *λ and d_∞ from time series.* Two key aspects of chaos are the stretching of infinitesimal displacements and the existence of complex orbit-like structures, in the form of a vast variety of possible unstable orbits, confined in a region of the phase space called the attractor [30]. The stretching property is strictly related to sensitive dependence on initial conditions. A quantitative characterization of stretching properties is provided by λ. Assuming that x denotes a k-dimensional vector, and considering the dynamical system specified by the discrete map:

$$x_{n+1} = G(x_n) \tag{8}$$

Considering N couples of trajectories starting from two nearby points separated by a small distance d_o

$$\left| x_0^{(i)} - x_0^{'(i)} \right| \leq h_0 \quad x_j^{(i)} = G_n(x_0^{(i)}) \quad x_j^{'(i)} = G_n(x_0^{'(i)}) \tag{9}$$

averaging the N couples of trajectories, the mean distance between trajectories after j iteration can be defined as:

$$d_j = \frac{1}{N} \sum_{i=1}^{N} \left| x_j^{(i)} - x_j^{\prime(i)} \right| \tag{10}$$

where the $||$ operator denotes the norm. The asymptotic value of d_j is defined as:

$$d_\infty = \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} d_j \tag{11}$$

It is well known [23] that, after n iterations, the stretching phenomenon stretches the distance d_n as:

$$d_{n+1} = e^{n\lambda} \quad d_0 = e^{\lambda} \quad d_n = \Lambda d_n \tag{12}$$

where λ is the Lyapunov exponent of system. After a sufficiently large number of iterations, the folding process takes place to keep the trajectories bound in the phase space.
To take this phenomenon into account, the (12) can be considered as a first order expansion of d_∞ and, in the hypothesis that $d_n < 1$ for any n, it includes a second order correction term representing the folding action.

$$d_{n+1} = \Lambda d_n - \Gamma d_n^2 \tag{13}$$

The fixed points of (13) are

$$d^* = 0 \quad and \quad d^{**} = d_\infty = \frac{\Lambda - 1}{\Gamma}. \tag{14}$$

The characteristic values describing the evolution of nearby trajectories are Λ, Γ, and d_∞, although only two of these are actually needed because of the relationship (14). It is important to note that, while Λ is only sensitive to the stretching mechanism, d_∞ is sensitive to both the stretching and the folding mechanisms. For a continuous system, the above assumptions still apply and a similar characterization can be given [2].

The aim of DivA algorithm, given a signal s(t) (i.e. given a time series), is to compute the divergence (d_j) among trajectories x_i. The algorithm starts with the choice by the operator of an initial condition $x_0^{(0)}$ and a distance h_0, which identifies a small range $[(x_0 - h_0/2), (x_0 + h_0/2)]$ ("small" generally depends on the resolution of the signal s(t) and on the number of trajectories found).

Then, points whose y-coordinate belongs to the range $[(x_0 - h_0/2), (x_0 + h_0/2)]$, are extracted; these points represent a set of candidates to become starting points of the algorithm in relation to the equation (9). The first starting point found is assumed to be $x^* \equiv x_0^{(0)}$, this point will be used as reference point for the following steps.

The algorithm proceeds with the computing of the first derivate in x^*. Among the points in the set of the candidate starting points, only those whose derivative meets constrain (15) will represent the final set of starting points from which trajectories will be calculated.

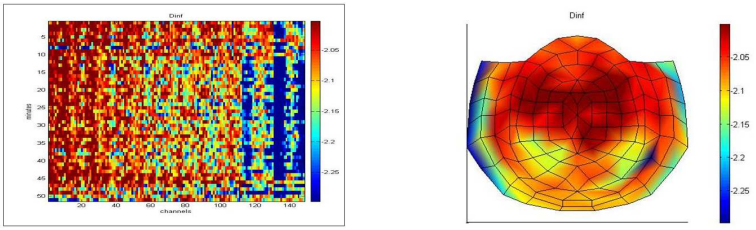
$$\left| \dot{x}_0^{(0)} - \dot{x}_0^{(i)} \right| = p \cdot var(s'(t)) \tag{15}$$

Parameter p in (15) is the slope ratio, is chosen, empirically by the user,and is small enough so that pairs of trajectories that have a different initial slope are discarded, thus decreasing the number of trajectories for the calculation of the d_j, but not so strictly so that a sufficient number of trajectories, respecting the requirement on the range and the initial slope, can be extracted. The term var(s') represents the variance of the derivative of signal s(t). The constrain on the slope has been introduced in order to collect all points having the same properties in the zero order and first order dynamics.

By means of the above described steps, a set of starting points is found: $X = (x_0^{(i)}, i = 0, \ldots, n)$. Each point $x_0^{(i)}$ identifies a trajectory made up of all the samples in the range: $[x_0^{(i)}, (x_0^{(i)} + lenght - trj - 1)]$, where *length-trj* is the length of the trajectories chosen in a way that, when the distance between them is computed, both the stretching and the folding effects are taken into account, and the asymptotic behaviour of the system can be studied. Moreover, all the combinations among points $x_0^{(i)}$ will be considered, discarding those couples whose distance (in samples) is inferior to the parameter *minimum trajectories delay*(t_{dmin}), and their differences will be computed thus obtaining d_j.

The d_∞, representing the asymptotic value of d_j is then extracted and used as a parameter for characterizing the nonlinear dynamics of the system. Moreover, from the computed curve d_j, the maximum Lyapunov exponent can be extracted as the initial slope of the curve. This extraction can be computed in different ways, polynomial fit, custom equation fit or in an empiric way. This last method is the one used here.

This method can be used for the characterization of time series coming from measurements performed on real systems when the laws and structures are unknown and chaotic dynamics are suspected. Since it is computationally efficient, it can be easily applied to large data sets [20].
The output representation can be obtained in spatio-temporal and spatial maps for any geometry as show in Figure 10 a and b respectively, were the d_∞ has been calculated over the minute 5.



(a) (b)

FIGURE 10. (a) Spatio-temporal d_∞ maps. (b) Average d_∞ head maps.

6. **Conclusions.** The BioS platform developed in Matlab7 is a user-friendly interactive environment allowing bio-potential data processing and results representation through images.
The analyses provided by BioS are useful for both inexperienced users, to study bio-potential data without programming, and experienced ones, to speed up experimental data analysis. Thus, both the experienced and the inexperienced users will be able to utilize BioS features such as data visualization, pre-processing, and processing. This makes the Bios and all its functionality a valid and user-friendly support to medical interpretations of data for clinical purposes.
The software architecture has been designed to be modular and flexible in order to meet any future requirements. This software allows users to process biological data from different acquisition systems, characterized by different sensor distribution geometries (single area, multiple areas) and by different data formats (binary, ASCII and Matlab).
Concerning visualization, BioS allows a temporal, spatial, and spatio-temporal representation not only for raw data, but also for the analyzed signals.
The BioS pre-processing feature permits users to perform statistical analyses and frequency and saturation filtering, which are useful in preparing the data for the analyses supported by the processing phase.

Currently, Bios includes three analysis methods, two belonging to linear methods, the power distribution and the independent components analysis, and one related to the nonlinear analysis methods for the extraction of the maximum Lyapunov exponent (λ) and asymptotic divergence between trajectories d_∞(d-infinite). All the methods are extended both to the spatial and temporal domain and can be applied on both raw data and pre-processed signals. Moreover, to perform the nonlinear analysis, a new optimized algorithm, DivA, has been designed and integrated.

All BioS features were tested on MEG data acquired through two different instrumentations and geometries, but it is also possible to work with other types of data (i.e. EEG data in txt format). Furthermore a section dedicated to multiple data analysis (such as simultaneous recordings of nasal cycle, blood pressure, EEG) has been designed for a future development.

Thanks to its modularity, BioS could support other pre-processing or processing features. For example, in order to overcome the problem of limitations in the study of spatially-extended biological systems in the presence of short times, it has been considered the opportunity to study the interactions between dynamical systems, commonly quantified using linear techniques, i.e. coherency [21][22], or other methods [10][27] such as "synchronization likelihood" [29], which is based on the statistical interdependence of brain activity signals, as an index of "functional connectivity". With the same aim, other nonlinear methods, such as entropies (Sample [26], Approximate [25] and Multiscale Entropy [6]) and fractal dimensions characterization (Higuchi Dimension [14] and Detrended Fluctuation Analysis [24]), are going to be implemented.
Moreover, a future development is represented by porting the Bios code on grid computing architectures, parallelizing the data processing with respect to the space (areas or channels) and time (minutes, seconds or milliseconds) domain. This would offer the possibility of a massive characterization of bio-potential data and statistical analyses, which are at the bases of Data Bases creation supporting diagnosis, treatment, and prognosis in clinical studies.

Acknowledgements. Thanks to D. Shannahoff-Khalsa for his fruitful collaboration working on Analysis of MEG data for treatment of Obsessive Compulsive Disorder with yoga breathing technique. MEG data provided by the Institute for Nonlinear Science, University of California, San Diego, La Jolla, California, USA.

REFERENCES

[1] B.B. Biswal and J.L. Ulmer, *Blind source separation of multiple signal sources of fMRI data sets using independent component analysis*, J. Comput. Assist. Tomogr, **23** (1999), 265–271.
[2] A. Bonasera, M. Bucolo, L. Fortuna and A. Rizzo, *The d-infinite parameter to characterise chaotic dynamics*, Proceedings of the IEEE-INNS- ENNS International Joint Conference on Neural Networks, **5** (2000), 24–27.
[3] A. Bonasera, M. Bucolo, L. Fortuna and A. Rizzo, *A New Characterization of Chaotic Dynamics: the d-infinite Parameter*, Nonlinear Phenomena in Complex Systems, **6(3)** (2003), 779–786.
[4] M. Bucolo, L. Fortuna, N. Frasca, M. La Rosa, D. S. Shannhoff-Khalsa, L. Schult and J.A. Wright, *Independent component analysis of magnetoencephalography data*, "Proceedings 23rd EMBC", (2001).
[5] G. Bucolo, M. Bucolo, M. Frasca, M. La Rosa, D. Shannahoff-Khalsa and M. Sorbello, *Spatial Modes in Magnetoencephalography Spatio-Temporal Patterns*, Proceedings of the 25th Annual International Conference of the IEEE EMBS, Cancun, Mexico, (2003).
[6] M. Costa , A.L. Goldberger and C.-K.Peng, *Multiscale entropy analysis of complex physiologic time series*, Phys. Rev. Lett., **89(6)** (2002), 68-102 .
[7] A. Delorme and S. Makeig, *EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics*, Journal of Neuroscience Methods, **134** (2002), 9–21.
[8] M.J. McKeown ,S. Makeig ,G.G. Brown ,T.-P. Jung , S.S. Kindermann and T.J. Sejnowski, *Analysis of fMRI data by blind separation into independent spatial components*, Human Brain Mapping, **6** (1998), 160–188.
[9] The FastICA package is Copyright (C) 1996-2005 by Hugo Gvert, Jarmo Hurri, Jaakko Srel, and Aapo Hyvrinen.

[10] P.J. Franaszczuk and G.K. Bergey, *An autoregressive method for the measurement of synchronization of interictal and ictal EEG signals*, Biol Cybern, **81** (1999).
[11] A. Fuchs, V. K. Jirsa, and J. A. S. Kelso, *Theory of the Relation between Human Brain Activity (MEG) and Hand Movements*, NeuroImage, **11** (2000), 359–369.
[12] A. Fuchs, V. K. Jirsa, and J. A. S. Kelso, *Spatiotemporal forward solution of the EEG and MEG using network modelling*, IEEE Transaction on Medical Imaging, **21(5)** (2002).
[13] H. Haken, *Nonlinearities in Biology: The Brain as an Example*, Lecture Notes in Physics,Springer Berlin/Heidelberg GmbH, **542** (2000), 427–445.
[14] T. Higuchi, *Relationship between the fractal dimension and the power law index for a time series: A numerical investigation*, Physica D, **46** (1990), 254–264 .
[15] A. Hyvarinen and E. Oja, *Independent component analysis: algorithms and applications*, Neural Networks, **13** (2000), 411–430.
[16] PCh. Ivanov ,MG Rosenblum ,C-K Peng ,J. Mietus ,S. Havlin ,HE. Stanley and AL. Goldberger , *Scaling behaviour of heartbeat intervals obtained by wavelet-based time series analysis*, Nature, **383** (1996), 323–327.
[17] Z. J. Kowalik, A. Schnitzler, H. J. Freund and O. W. Witte, *Local Lyapunov exponents detect epileptic zones in spike-less interictal MEG recordings*, Clinical Neurophysiology, **112** (2001), 60–67.
[18] L. A. Lipsitz, *Physiological Complexity, Aging, and the Path to Frailty*, Sci. Aging Knowl. Environ., **16** (2004), pe16.
[19] S.Makeig, A.J.Bell, T.P. Jung and T.J.Sejnowski, *Independent component analysis of electroencephalic data*, Advances in Neural Information Processing Systems, **8** (1996), 145–151.
[20] G. Nolte, T. Sander, A. Lueschow, and B. A. Pearlmutter, *Nonlinear time series analysis of human alpha rhythm*, Proceedings of the BIOMAG, Jena (2002).
[21] P.L. Nunez, R. Srinivasan, A.F. Westdorp, R.S. Wijesinghe, D.M. Tucker, R.B. Silberstein and P.J. Cadusch, *EEG coherency. I. Statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales*, Electroenceph. Clin. Neurophysiol., **103** (1997).
[22] P.L. Nunez, R.B. Silberstein, Z. Shi, M.R. Carpenter, R. Srinivasan, D.M. Tucker, S.M. Doran, P.J. Cadusch and R.S. Wijesinghe, *EEG coherency. II. Experimental comparisons of multiple measures*, Clin. Neurophysiol., **110** (1999).
[23] U. Parlitz, *Nonlinear Time-Series Analysis in Nonlinear Modeling Advanced Black-Box Techniques*, Eds. J.A.K. Suykens and J. Vandewalle Kluwer Academic Publishers (1998).
[24] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley and A.L. Goldberger, *Mosaic organizations of DNA nucleotides*, Phys. Rev. E , **49** (1994), 1685-1689.
[25] S.M. Pincus, *Approximate entropy (ApEn) as complexity measure*, Chaos, **5(1)** (1995), 110–117.
[26] J.S. Richman and J.R. Moorman, *Physiological time-series analysis using approximate entropy and sample entropy*, Heart Circ. Physiol., **278** (2000), 2039–2049.
[27] M.G. Rosenblum and J. Kurths, *Analysing synchronization phenomena from bivariate data by means of the Hilbert transform*, Non-linear Analysis of Physiological Data, Springer Series for Synergetics, Springer, (1998) , 91–100 .
[28] F. Sapuppo, E. Umana, M. Frasca, M. La Rosa, D. Shannahoff-Khalsa, L. Fortuna and M. Bucolo, *Complex Spatio-Temporal Feature in MEG Data*, Mathematical Biosciences and Engineering, **3(4)** (2006), 697–716.
[29] C.J.Stam and B.W. Van Dijk, *Synchronization likelihood: an unbiased measure of generalized synchronization in multivariate data sets*, Physica D, **163** (2002), 236–241 .
[30] S. H. Strogatz, Nonlinear Dynamics and Chaos. Perseus Book, Cambridge, Massachusetts. Webster J.G. (1998), Medical Instrumentation, Wiley.
[31] [http://www.phitools.com/index.html], [http://www.cortechsolutions.com/].
[32] [http://www.sprweb.org/repository/index.html].
[33] [http://neuroimage.usc.edu/brainstorm/].

Received on October 30, 2007. Accepted on February 3, 2008.

E-mail address: maide.bucolo@diees.unict.it
E-mail address: fdigra@diees.unict.it
E-mail address: lfortuna@diees.unict.it
E-mail address: mfrasca@diees.unict.it
E-mail address: fsapuppo@diees.unict.it