ELSEVIER

Computer Methods and Programs in Biomedicine 51 (1996) 51–73

Computer Methods and Programs in Biomedicine

Non-linear algorithms for processing biological signals

S. Cerutti*[a], G. Carrault[b], P.J.M. Cluitmans[c], A. Kinie[b], T. Lipping[d], N. Nikolaidis[e], I. Pitas[e], M.G. Signorini[a]

[a] *Biomedical Engineering Department, Polytechnic University p.za Leonardo da Vinci 32, 20133, Milano, Italy*
[b] *Laboratoire Traitement du Signal et de l'Image — CJF-Inserm 93-04, Université de Rennes, I-35042 Rennes Cedex, France*
[c] *Department of Medical Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*
[d] *Signal Processing Laboratory, Tampere University of Technology, P.O. BOX 553, FIN-33101, Tampere, Finland*
[e] *Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki 540 06, Greece*

Abstract

This paper illustrates different approaches to the analysis of biological signals based on non-linear methods. The performance of such approaches, despite the greater methodological and computational complexity is, in many instances, more successful compared to linear approaches, in enhancing important parameters for both physiological studies and clinical protocols. The methods introduced employ median filters for pattern recognition, adaptive segmentation, data compression, prediction and data modelling as well as multivariate estimators in data clustering through median learning vector quantizers. Another approach described uses Wiener-Volterra kernel technique to obtain a satisfactory estimation and causality test among EEG recordings. Finally, methods for the assessment of non-linear dynamic behaviour are discussed and applied to the analysis of heart rate variability signal. In this way invariant parameters are studied which describe non-linear phenomena in the modelling of the physiological systems under investigation.

Keywords: Median filtering; Median learning vector quantizers; Wiener-Volterra kernel; Non-linear dynamics; Time-delay estimation; Deterministic chaos

* Corresponding author. Biomedical Engineering Department, Polytechnic University, p.za Leonardo da Vinci 32, 20133 Milano, Italy. Tel: +39 223993339; fax: +39 223993360; e-mail: cerutti@icil64.cilea.it

PII S0169-2607(96)01762-2

1. Introduction

In many studies of biological phenomena, non-linear mechanisms probably play an important role: they are certainly connected to the great interactions among sub-systems of different natures which generally constitute the system under analysis, and to the non-linear characteristics of various kinds of responses. Therefore, if one wants to correctly enhance information from these studies, one has to employ properly tuned and tested algorithms of signal processing which, by definition, ought to take into account the non-linear nature of the process. By such an approach one may avoid oversimplification, typical of classical linear modelling, which obviously is not capable of capturing information bounded to different non-linear behaviour, like entrainment, modulation phenomena, on-off mechanisms of responses, variable gain, threshold and saturation characteristics, and conclamate chaotic behaviour.

The disadvantages of non-linear methods of signal processing are: (i) that they do not generally present solutions in closed mathematical forms; and (ii) that they are generally more complex, both from a structural and from a computational point of view.

In the following we will consider different approaches of non-linear methods which will be applied to various examples taken from biological systems.

2. Non-linear algorithms of signal processing and modelling

2.1. Adaptive FIR weighted order statistic filtering

Median filtering is a non-linear filtering technique that performs particularly well in the presence of impulsive type noise. An even more important property is its ability to preserve sharp changes in the input signal. Several algorithms have been developed recently to optimise median type filters. In this paper we will apply an adaptive FIR weighted order statistic hybrid (FWH) filter to EEG burst-suppression (BS) pattern and discuss the possibilities of using adaptive median type filters in physiological signal processing.

2.1.1. Methods

The median operation, originally used in statistics, was introduced for signal processing by Tukey [1] in the 1970s. The median of a set of $N = 2k + 1$ values $\{x(n-k), \ldots, x(n), \ldots, x(n+k)\}$ can be computed by sorting the data according to their values $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(N)}$, and choosing the middle sample $x_{(k+1)}$ as the output.

The performance of median and median-type filters can be described by means of root signal analysis and statistical analysis. It has been shown by Gallagher and Wise [2] that consecutive median filtering will turn any finite length signal into a root signal, i.e., a signal which is invariant to further filtering with the particular filter. The root signals of median filter consist of regions with at least $k + 1$ consecutive identically valued samples, called constant neighbourhoods, and monotonic regions between the constant neighbourhoods, called edges. The statistical analysis of median filter has shown that median is the maximum likelihood estimate of location in the presence of uncorrelated additive double exponential or Laplacian noise [3], in other words: the behaviour of median in the case of long-tailed noise distributions is analogous to the behaviour of mean in the case of Gaussian noise. However, median filter tends to remove signal details which are small compared to the filter window. Also, it may cause edge jitter and streaking and the computation is slow in the case of long windows due to the time consuming sorting operation.

To overcome the drawbacks of the median filter several extensions have been introduced, one of which is the weighted median (WM) filter [4]. The WM filter is defined by

$$D(n) = \mathrm{Med}\{w_1 \diamond x(n-k), \ldots, w_N \diamond x(n+k)\} \tag{2.1.1}$$

where $w \diamond x$ means x repeated w times. WM filters can be optimised to minimise certain error criterion or to preserve certain patterns or structures in the signal. The latter is of special interest in image processing applications. Also the WM filter uses time order information which improves the ability of the filter to preserve details in the

signal. The WM filter can be extended to weighted order statistic (WOS) filter where the output is w_0: the largest sample in the weighted data set of Eq. 2.1.1. WM filter is a special case of WOS filter with

$$w_0 = 1/2 \sum_{i=1}^{N} = 1 \times w_i.$$

Another extension to the median filter is the FIR median hybrid (FMH) filter, where the output is the median of the outputs of k FIR subfilters, applied to the input signal:

$$D(n) = \mathrm{Med}\{Y_1(n),\ldots,Y_k(n)\}; Y_i(n) = \sum_{j=0}^{N-1} h_{ij}\, x(n-k+j), i = 1,\ldots,k. \quad (2.1.2)$$

FMH filters can be designed to preserve sharp changes in signal baseline, as median filter does while the computation time is reduced significantly. Also, better noise attenuation can be achieved due to the FIR subfilters. FMH filters have been applied to the trend detection problem in physiological signal processing [5,6].

To combine the good properties of the aforementioned filter algorithms, the FIR-WOS hybrid (FWH) filter was introduced recently. The output of the general FWH filter is the w_0: the largest value of the set $\{w_1 \diamond Y_1(n),\ldots, w_k \diamond Y_k(n)\}$, where $Y_i(n)$, $i = 1,\ldots,k$ are the outputs of FIR subfilters. Yin and Neuvo [7] present fast algorithms for the adaptation of FWH filters to minimise MAE criterion. There are two ways to adapt the FWH filters: (i) all the coefficients of the FIR subfilters and the weights of the WOS filter are adapted simultaneously; and (ii) separately optimised FIR filters are used, while only the weights of the WOS part of the filter are adapted during the adaptation of the FWH filter.

In this paper we will consider a special case of adaptive FWH filter, shown in Fig. 1. The coefficients of the FIR filter are optimised separately using the Widrow-Hoff LMS adaptation algorithm and the WOS part of the filter is optimised using the following relations [7]:

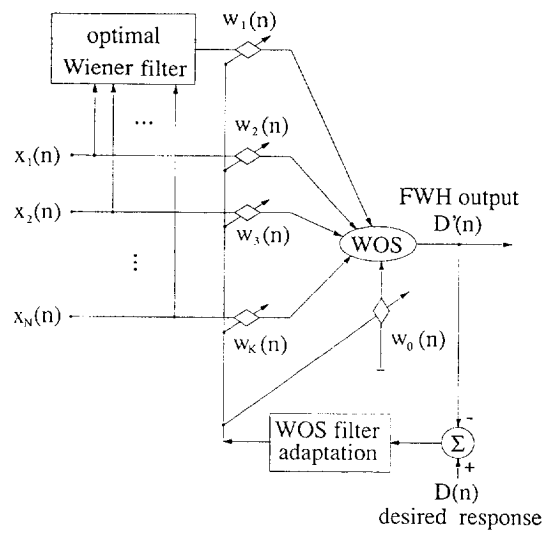

Fig. 1. The structure of the FWH filter used in the experiment.

$$w_0(n+1) = P\left[w_0(n) - 2\mu \sum_{i=2}^{K} \left(Y_{(i)}(n) - Y_{(i-1)}(n)\right) e^{Y_{(i-1)}(n)} \times (n)\Delta^{Y_{(i-1)}(n)}(n)\right], \quad (2.1.3a)$$

$$w_j(n+1) = P\left[w_j(n) + 2\mu \sum_{i=2}^{K} \left(Y_{(i)}(n) - Y_{(i-1)}(n)\right) e^{Y_{(i-1)(n)}}(n) \times \Delta^{Y_{(i-1)(n)}}(n)\, y_j^{Y_{(i-1)(n)}}(n)\right], \quad j = 1,\ldots,K, \quad (2.1.3b)$$

where

$$e^{\Gamma}(n) = d^{\Gamma}(n) - d_r s^{\Gamma}(n); \quad (2.14)$$

$$d^{\Gamma}(n) = U(D(n) - \Gamma); \quad (2.15)$$

$$\Delta^{\Gamma}(n) = d_s^{\Gamma}(n)(1 - d_s^{\Gamma}(n)); \quad (2.16)$$

$$d_s^{\Gamma}(n) = U_s\left(\{w_0,\ldots,w_K\}\{-1, y_1^{\Gamma} \times (n),\ldots, y_K^{\Gamma}(n)\}^T\right); \quad (2.17)$$

$$y_k^{\Gamma}(n) = U(Y_k(n) - \Gamma); \tag{2.18}$$

$$U(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{2.19}$$

$$U_s(X) = 1/(1 + e^{-X}); \tag{2.1.10}$$

$$P(X) = \begin{cases} X, & \text{if } X \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{2.1.11}$$

where μ is the adaptation step size. By introducing some simplifications, the algorithm can be made as fast as the Wiener filter adaptation algorithm:

$$w_0(n+1) = P[w_0(n) - 2\mu(D(n) - D'(n))] \tag{2.1.12a}$$

$$w_j(n+1) = P\big[w_j(n) + 2\mu(D(n) - D'(n))U\big(Y_j(n) - D'(n)\big)\big], \quad j = 1, \ldots, K. \tag{2.1.12b}$$

However, this algorithm is dependent on the initial weight vector and is likely to end up with a local minimum. A suitable initial weight vector can be found by random search or using genetic algorithms [7]. Another way is to start the adaptation with the algorithm of Eq. 2.1.3, and after several iterations turn to the fast algorithm of Eq. 2.1.12.

2.1.2. Results: burst-suppression of EEG

Burst-suppression pattern of EEG signal was used to compare the performance of the Wiener filter and the adaptive FWH filter of Fig. 1. The BS pattern can be described as mixed frequency high amplitude bursts on DC-shift alternating with periods of suppressed EEG. Usually EEG signal is recorded with low time constant of analog prefilters and the DC-shift cannot be seen in the recording. In our experiment we used 5 EEG recordings of isoflurane induced BS pattern. The signals were recorded with time constant 5 s (cut-off frequency approximately 0.03 Hz), where the DC-shift is turned into a saw-tooth like pattern (Fig. 3a), and sampled at frequency 200 Hz. The ability of the filter algorithms to preserve the DC pattern while suppressing the burst activity was tested.

We chose 10 bursts from the 5 recordings (2 from each) to generate a test signal. The test signal was filtered with a high-pass filter (cut-off frequency 1.6 Hz) and a known DC-pattern (Fig. 2a) was added so that the sharp rise in the DC-pattern occurred at the beginning of bursts, as is the case with natural bursts. Seven bursts from the test signal were used to optimise the filter algorithms and then the whole test signal was filtered with the optimal filters (Fig. 2). The MAE and MSE between the outputs of the filters and the known DC-pattern are given in Table 1. Both errors are about 25% smaller for the FWH filter. Fig. 2 shows that the main advantage of the FWH filter is its ability to follow sharp changes in the signal baseline. In Fig. 3 the performance of the filters on the original BS data is shown.

2.2. Non-parametric Wiener-Volterra analysis

An evoked potential can be interpreted as the impulse response of the sensory pathway involved in the transmission of information from a sensory organ to cortical structures. Considering the nature of synaptic transmission, it is likely that non-linear properties play an important role in the input-output relationship of this process.

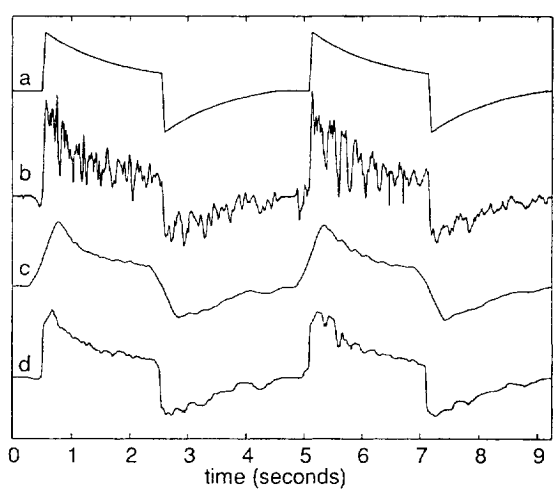

Fig. 2. (a) The DC-pattern of the test signal; (b) the test signal; (c) the test signal filtered with the Wiener filter; and (d) the test signal filtered with the optimal FWH filter.

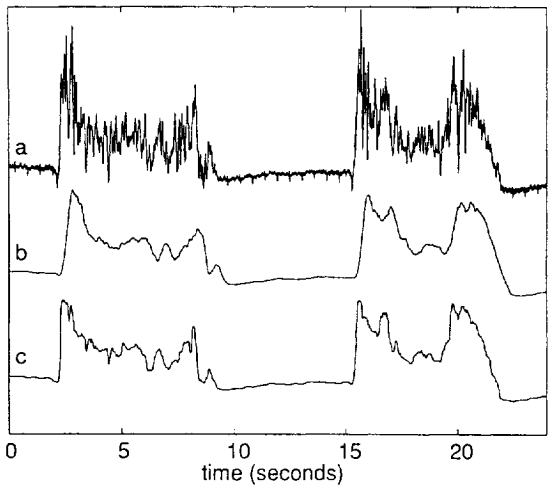

Fig. 3. (a) Original burst-suppression pattern slightly corrupted by ECG artifact; (b) BS pattern filtered with the Wiener filter; and (c) BS pattern filtered with the optimal FWH filter.

Table 1
MAE and MSE between the known DC-pattern and the filter outputs

	MAE ($\times 10^2$)	MSE ($\times 10^5$)
Wiener filter	2.96	1.81
FWH filter	2.27	1.24

Identification of the non-linear properties using a non-parametric approach based upon a Wiener-Volterra description [8] of the system properties is proposed to obtain first indication of the second-order non-linearities involved.

2.2.1. Methods

An evoked potential is the response of a neural pathway to a sensory stimulus. To improve the (often very poor) signal to noise ratio in a clinical setting, such a response most often is determined by averaging responses to multiple stimuli. If the duration of the stimuli is sufficiently short, the evoked response can be considered to be the impulse response of the sensory system under study.

Furthermore, if the system is linear this impulse response entirely characterises the system properties, and the output $y(t)$ to any input, $x(t)$ can be determined by:

$$y(t) = h_0 + \int_{-\infty}^{\infty} h_1(\tau)x(t-\tau)d\tau \qquad (2.2.1.)$$

with h_0 the average DC component of the output and h_1 the impulse response. However neural pathways, when considered as a system receiving an input at — for instance — a receptor and showing a response measured at a scalp electrode, have non-linear properties. We only have to refer to the highly non-linear nature of the synaptic transmission to illustrate this. Basically, there are two approaches to model the system properties of non-linear systems. When the nature and order of the non-linearities are modelled (e.g. based upon knowledge about the physiological mechanisms underlying the neural generators), a parametric approach using differential equations can be used.

However, when entire sensory pathways underlying scalp evoked activity are studied, there is not enough available knowledge to use this approach. A non-parametric description in terms of a Volterra series may be useful in this case.

Wiener proved that for a broad class of stationary non-linear systems this input-output relationship can be described by a Volterra series of orthogonal terms:

$$y(t) = h_0 + \int_{-\infty}^{\infty} h_1(\tau)x(t-\tau)d\tau + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_2(\tau_1, \tau_2) \times x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 + \ldots \qquad (2.2.2)$$

Here, the functions h_i are independent of the actual input $x(t)$ and are called the kernels of the system. Wiener proved that the kernels can be determined by applying Gaussian white noise (GWN) at the input and calculating the cross-correlation of the input process and the measured output $y(t)$. Krausz [9] translated this approach to a time-discrete version and proved that a set of kernels of a given system can be determined by processing the output of the system when the

input receives a Poisson process of Dirac-impulses (the discrete equivalent to GWN) at the input. Such a Poisson process is described as:

$$x(t) = \sum_{i=0}^{\infty} \delta(t - t_i),$$

with t_i the time the i-th impulse is applied. The interval between two subsequent impulses, $t_i - t_{i-1}$ is exponentially distributed:

$$P(t_i - t_{i-1} < T) = 1 - e^{-\sigma \cdot t}$$

with σ the average rate of input impulses.

The zero, first and second order kernels then can be determined by calculating the cross-correlation between the input and the measured output, y(t) and the Poisson input:

$$h_0 = \overline{y(t)}$$

$$h_1(\tau) = c_1(\tau) - h_0, \text{ with } c_1(\tau) = \frac{1}{\sigma}\overline{y(t)x(t-\tau)}$$

$$h_2(\tau_1,\tau_2) = \tfrac{1}{2}.[c_2(\tau_1,\tau_2) - h_1(\tau_1) - h_2(\tau_2) - h_0], \qquad (2.2.3.)$$

$$\text{with } c_2(\tau_1,\tau_2) = \frac{1}{\sigma^2}\overline{y(t)x(t-\tau_1)x(t-\tau_2)}$$

$$\text{and } (\tau_1 \neq \tau_2)$$

The zero-order kernel, h_0 simply is the average DC component of the measured output. It can be shown [10] that the function $c_1(\tau)$ in Eq. 2.2.3 is the AC component of the average response to each individual input impulse using a technique similar to the averaging known in the conventional evoked potential technique. The averaging process will cause an increase in the signal to noise ratio of the resulting waveform in both techniques, but two main differences exist between the two techniques: first, there is no restriction to the minimum inter-stimulus interval when using the random stimulation technique and, secondly, the final result of the new technique will contain an average effect of non-linearities caused by interference between overlapping responses.

The function c_2 in Eq. 2.2.3 can be interpreted as the AC component of the average response of the system to each pair of stimuli with an inter-stimulus interval equal to $\delta = \tau_2 - \tau_1$, and can be determined by applying the averaging technique indicated in Fig. 4. A possible way to interpret the second order non-linearities of a system is to introduce a correction in the function c_2 — which is an averaged response to two stimuli — for the response to the first stimulus which could be expected if the second stimulus were not present, i.e., the first order kernel triggered by the first stimulus of the pair.

This new function, which we call the recovery kernel is defined as: $R(\tau_1,\tau_2) = c_2(\tau_1,\tau_2) - h_1(\tau_1)$. This recovery kernel can be interpreted as the response to only the second stimulus in a pair of stimuli with inter-stimulus interval $\delta = \tau_2 - \tau_1$. For large values of δ, there will be few effects of interference (non-linearities) and therefore the recovery kernel will converge to the first order kernel for increasing δ. For very small values of δ, the effect of inhibition and refractory periods will play a significant role, and the response to the second stimulus will hardly be present in the response to a pair of stimuli. A practical but important point in the determination of these second order effects is that the number of averages available is much smaller than the number available for the $h_1(\tau)$ determination, resulting in a poorer signal to noise ratio.

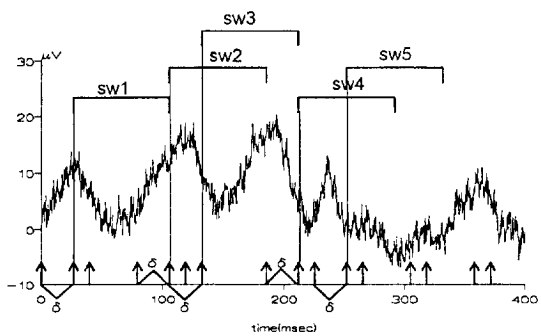

Fig. 4. Determination of c_2. Each pair of (not necessarily consecutive) stimuli with a inter-stimulus interval equal to $\tau_1 - \tau_2 = \delta$ triggers a new averaging process.

2.2.2. Results: first-order Wiener-Volterra kernel in the processing of evoked potentials

The results of a conventional evoked potential recording and the first order kernel determined using the high-frequency, random stimulation technique measured in a controlled study with cats are shown in Fig. 5. It can be seen that the shape of both waveforms up to 40 ms after the stimulus is comparable. The amplitude of the middle latency components (10–60 ms) in the first order kernel waveform however are considerably smaller than those in the conventional response. This is caused by the average effect of non-linearities mentioned above. Cluitmans however has shown [11] that the changes in amplitude and latencies relative to baseline values in both waveforms caused by changing levels of isoflurane-nitrous oxide anaesthesia were comparable. This suggests that obtaining an h_1 waveform may be an alternative to the conventional averaging technique. A potential practical advantage of the new technique is that the h_1 waveform can be obtained faster than the EP waveform, because of the relatively high average stimulus frequency allowed.

Fig. 6 shows the recovery kernel for various values of δ measured under the same controlled laboratory conditions as Fig. 5. For reference, the first order kernel, $h_1(\tau)$ is shown in the upper graph. All waveforms are determined from a single data set. Both the poorer signal to noise ratio in the recovery kernels and the convergence towards the $h_1(\tau)$ waveform for increasing values of δ can be seen. From these data, it becomes clear that the effect of non-linearities is much smaller in the early brain-stem components (0–10 ms) when compared to the middle latency components (> 10 ms).

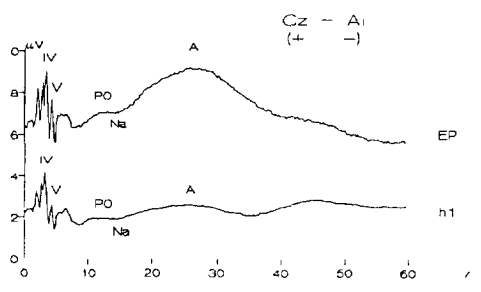

Fig. 5. Conventional (EP) and first order evoked response measured in an awake cat. Number of averages in EP: 1000 (90 s. data); Number of averages in h_1 = 10000 (100-s data).

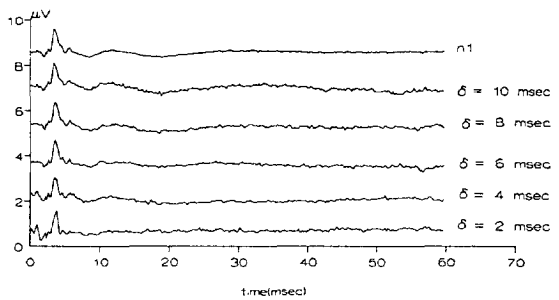

Fig. 6. Recovery kernels determined in an awake cat with auditory stimulation (number of averages = 200).

2.3. Multivariate data clustering through median learning vector quantizers

Neural networks is a rapidly expanding research field which has attracted the attention of scientists and engineers in the last decade. A large variety of artificial neural networks has been developed based on a multitude of learning techniques and having different topologies [12]. One prominent example of neural networks is the learning vector quantizer (LVQ). A novel class of learning vector quantizers (LVQs) based on multivariate order statistics is proposed in order to overcome the drawback that the estimators for obtaining the reference vectors in LVQ do not have robustness either against erroneous choices for the winner vector or against the outliers that may exist in vector-valued observations. The proposed algorithms can cope very well with multivariate data clustering problems. Therefore, they can find numerous biomedical applications, e.g. clustering of multiple physiological variables corresponding to the same patient and recorded over a time interval, or clustering of physiological variables corresponding to different patients and recorded at the same time instant.

2.3.1. Methods

The learning vector quantizer (LVQ) is an autoassociative nearest-neighbour classifier which classifies arbitrary patterns into classes using an error correction encoding procedure related to competitive learning [13] (Fig. 7). In order to make a distinction between the (standard) LVQ algorithm and the proposed variants that are based on multivariate order statistics, the LVQ algorithm will hereafter be called linear LVQ algorithm.

Let us assume a sequence of vector-valued observations $\mathbf{x}(t)$ $\mathscr{R}^p$ and a set of variable reference vectors or weights $\mathbf{w}_i(t)$; $\mathbf{w}_i \in \mathscr{R}^p$, $i = 1,2,\ldots,K\}$. Let $\mathbf{w}_i$ (0) be randomly initialised. Competitive learning tries to find the best-matching reference vector $\mathbf{w}_c(t)$ to $\mathbf{x}(t)$ (i.e., the winner) where $c = \arg\min_i \|\mathbf{x} - \mathbf{w}_i\|$ with $\|.\|$ denoting the Euclidean distance between any two vectors. This vector is updated and the process is repeated. After a large number of iterations, the different reference vectors $\mathbf{w}_i$ tend to be placed into the input space $\mathscr{R}^p$ in such a way that they approximate the probability density function (pdf) $f(\mathbf{x})$ of the multivariate input data in the sense of some minimal residual error $\epsilon = \int_\chi \|\mathbf{x} - \mathbf{w}_c\|^2 f(\mathbf{x}) d\mathbf{x}$

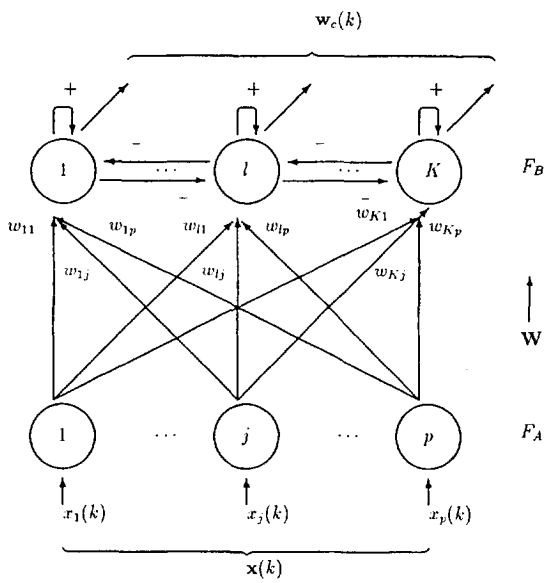

Fig. 7. The LVQ neural network.

where χ is the domain of the input vector-valued observations and $d\mathbf{x}$ is the volume differential in the space $\mathscr{R}^p$. Each reference vector is tuned to a different domain of the input data. If the stochastic-gradient-descent algorithm [14] is applied to the minimisation of ϵ in the space $\mathbf{w}_c$ and the weight vectors are updated as blocks concentrated around the winner, the following linear recursive relations result:

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + a(t)[\mathbf{x}(t) - \mathbf{w}_i(t)] \; \forall i \in \mathrm{N}_c(t)$$

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) \; \forall i \notin \mathrm{N}_c(t) \qquad (2.3.1)$$

where $a(t)$ is the adaptation step and $\mathrm{N}_c(t)$ denotes a neighbourhood around the winner. In the following, we use the notation n instead of t to denote discrete events. It can easily be seen that the reference vector for each class $i = 1,\ldots,K$ at time $n+1$ is a linear combination of the input vectors $\mathbf{x}(j) = j = 0,\ldots,n$ that have been assigned to class i. Moreover, it can be shown that in the special case of only one class following multivariate Gaussian distribution and the adaptation step sequence $a(n) = 1/(n+1)$, the winner vector is the arithmetic mean of the observations that have been assigned to the class (i.e., the maximum likelihood estimator of location). Neither in the case of multiple classes that are normally distributed nor in the case of non-Gaussian multivariate data distributions, is the linear LVQ the optimal estimator of the cluster means. In general, linear LVQ and its variations suffer from the following drawbacks:

- They do not use optimal estimators for obtaining the reference vectors $\mathbf{w}_i$ that match the pdf $f_i(\mathbf{x})$ of each class $i = 1,\ldots,K$.
- They do not have robustness against erroneous choices for the winner vector, since it is well known that linear estimators have poor robustness properties [15].
- They do not have robustness against the outliers that may exist in the vector observations.

In order to overcome these problems, we propose a novel class of learning vector quantizers that are based on order statistics [16,17]. It is well

known that operators based on order statistics have very good robustness properties. In the case of learning vector quantizers we should rely on multivariate order statistics. There is no unambiguous, universally agreeable total ordering of N p-variate samples $\mathbf{x}_i, \ldots, \mathbf{x}_N$ where $\mathbf{x}_i = (x_{1i}, \ldots, x_{pi})^T$, $i = 1, \ldots, N$. The following so-called sub-ordering principles are discussed by Barnett [17]: marginal ordering, reduced (aggregate) ordering, partial ordering, and conditional (sequential) ordering. In marginal ordering, the multivariate samples are ordered along each one of the p-dimensions:

$$x_{i(1)} \leq x_{i(2)} \leq \ldots \leq x_{i(N)} \quad i = 1, \ldots, p \tag{2.3.2}$$

i.e., the sorting is performed in each channel of the multichannel signal independently. The i-th marginal order statistic is the vector $\mathbf{x}_{(i)} = (x_{1(i)}, \ldots, x_{p(i)})^T$. The marginal median has the following definition:

$$\mathbf{x}_{\text{med}} = \begin{cases} (x_{1(\nu+1)}, \ldots, x_{p(\nu+1)})^T & N = 2\nu + \\ \left(\dfrac{x_{1(\nu)} + x_{1(\nu+1)}}{2} \right. & \\ \quad \ldots, \left. \dfrac{x_{p(\nu)} + x_{p(\nu+1)}}{2} \right)^T & N = 2\nu \end{cases} \tag{2.3.3}$$

It can be used in the following way in order to define the marginal median LVQ. Let us denote by $\mathbf{X}_i(n)$ the set of the vector observations that have been assigned to each class i, $i = 1, \ldots, K$ until time $n-1$. We find at time n the winner vector $\mathbf{w}_c(n)$ that minimises $\|\mathbf{x}(n) - \mathbf{w}_i(n)\|$, $i = 1, \ldots, K$. The marginal median LVQ (MMLVQ) updates the winner reference vector as follows:

$$\mathbf{w}_c(n+1) = \text{median}\{\mathbf{x}(n) \cup \mathbf{X}_c(n)\}. \tag{2.3.4}$$

The median operation is given by Eq. 2.3.3. Thus, all past class assignment sets $\mathbf{X}_i(n)$, $i = 1, \ldots, K$ are needed for MMLVQ. MMLVQ requires the calculation of the median of data sets of ever increasing size, as can be seen from Eq. 2.3.4. This may pose severe computational problems for relatively large n. However, for integer-valued data, a modification of the running median algorithm proposed by Huang et al. [18] can be devised to speed up median calculations by exploiting the fact that the marginal median of the already assigned samples $\mathbf{X}_i(n)$ is known.

Another definition of the multichannel median (based on the R-ordering principle) is the so-called vector median proposed by Astola et al. [19]. The vector median is the observation that has the minimum distance from all the remaining observations, i.e.:

$$\sum_{i=1}^{N} |\mathbf{x}_i - \mathbf{x}_{\text{med}}| \leq \sum_{i=1}^{N} \mathbf{x}_i - \mathbf{x}_j | \; \forall j = 1, \ldots, N. \tag{2.3.5}$$

The vector median LVQ (VMLVQ) uses the following formula to update the winner vector $\mathbf{w}_c(n)$ at step n:

$$\mathbf{w}_c(n+1) = \text{vector median } \{\mathbf{x}(n) \cup \mathbf{X}_c(n) \tag{2.3.6}$$

where $\mathbf{X}_i(n)$ is again the set of vector-valued observations that have been assigned to class i, $i = 1, \ldots, K$ so far and $\mathbf{x}(n)$ is the current observation. The vector median operator in the previous expression is the one defined in Eq. 2.3.5. The evaluation of the vector median of a data set is a rather computationally intensive operation (Eq. 2.3.5) since it requires the evaluation of n sums, each containing $n-1$ terms of the form $|\mathbf{x}_j - \mathbf{x}_i|$ and also the evaluation of the minimum of n values. In the case of the vector median LVQ we have to calculate the vector median (Eq. 2.3.6) for each time instant n. The fact that the vector median of the data in $\mathbf{X}_c(n)$ has already been evaluated can be exploited in order to speed up the computations. Both MMLVQ and VMLVQ keep track of their entire history and therefore all data samples have equal contribution to the reference vector update procedure. In the case of non-stationary data, we can evaluate the marginal or vector median using a moving window to discard the older samples as new observations become available.

The marginal weighted median LVQ (MWMLVQ) can be defined as follows. Let us denote by

$$\mathbf{w}_i(n) = \left(w_{i1}(n), w_{i2}(n), \ldots, w_{ip}(n)\right)^T \qquad (2.3.7)$$

the winner vector, i.e., $c = i$. In MWMLVQ, the elements of the winner vector are updated as follows:

$$w_{ij}(n+1) = \text{median}\left\{C_{i0} \diamond x_j(n), \ldots, C_{in} \diamond x_j(0)\right\} \qquad (2.3.8)$$

where $(C_{i0}, C_{i1}, \ldots, C_{in})^T$ is the vector of the duplication coefficients for the i-th class. The duplication coefficients can be chosen in such a way so that they heavily weigh the desired section of the data (i.e., the new observations or the old ones). If a weight C_{il} is zero, this means that the corresponding sample $\mathbf{x}(n-1)$ has not been assigned to the i-th class.

When the learning procedure of the proposed LVQ variations reaches equilibrium, it results in a partition of the domain of the input vector-valued observations called Voronoi tessellation. This means that the input space is partitioned into regions (called Voronoi neighbourhoods) bordered by hyperplanes such that each region contains a reference vector which is the nearest neighbour to any input vector within it. By doing so a clustering of the input multivariate data is achieved.

The asymptotic properties of MMLVQ and VMLVQ have also been studied. Expressions for the expected stationary state of the reference vectors when the input data are distributed according to 1-D or 2-D contaminated Gaussian or Laplacian distributions have been derived. Based on these expressions we were able to find the theoretical values for the bias in estimating the unconditional means of the underlying distributions. The bias introduced by the MMLVQ and VMLVQ was found to be smaller than the one introduced by the linear LVQ algorithm [20,21].

2.3.2. Results: comparison between linear and median learning vector quantizers (LVQ)

The performance of the proposed order statistics LVQs has been tested on finding the cluster means of a two-dimensional sample set that is described by the following probability density function:

$$\begin{aligned} f(x_1,x_2) = {} & pU([-5,20],[-5,20]) \\ & + (1-p)[eN(5,5;1,1;0) \\ & + (1-e)N(10,10;1,1;0)], \end{aligned}$$

where $U([-5,20],[-5,20])$ denotes the pdf of uniformly distributed outliers in the domain $[-5,20] \times [-5,20]$ and $N(m_{i1}, m_{i2}; \sigma_{i1}, \sigma_{i2}$ denotes a two-dimensional Gaussian distribution with mean m_{ij} and standard deviation σ_{ij} along each dimension j, $j = 1,2$ and correlation coefficient r. Furthermore, p is the probability of outlier occurrence and e the probability for one sample (that is not outlier) to be distributed according to the Gaussian pdf $N(5,5;1,1;0)$. Such a data set having $p = 0.2$ and $e = 0.5$ is shown in Fig. 8a, together with the trajectories of the weights determined by the marginal median LVQ algorithm. It must be stressed that this data set is heavily corrupted. It is clear that the MMLVQ converges close to the correct solution, i.e., the true cluster means. In contrast, the linear LVQ weights do not converge to the true cluster means (Fig. 8b), proof that the linear adaptation procedure (Eq. 2.3.1) is very susceptible to outlying input observations.

2.4. Time delay estimation using non-parametric methods

The EEG (electroencephalography) and SEEG (stereoelectroencephalography) analysis during epileptic seizures led to hypotheses concerning the behaviour and the interaction modes of the anatomical structures involved with the disease, in order to derive a possible epileptic network [22]. In signal processing terms, this latter point could be achieved by: (i) the characterisation of causal links (the interaction modes); (ii) the estimation of the delay; and (iii) the estimation of the nature of the relation (linear or non-linear), between two depth observations X_1 and X_2, reflecting the behaviour of the two structures. Only the second point is dealt with in this presentation. Several authors have attempted to evaluate the delay and the similarity between the channels

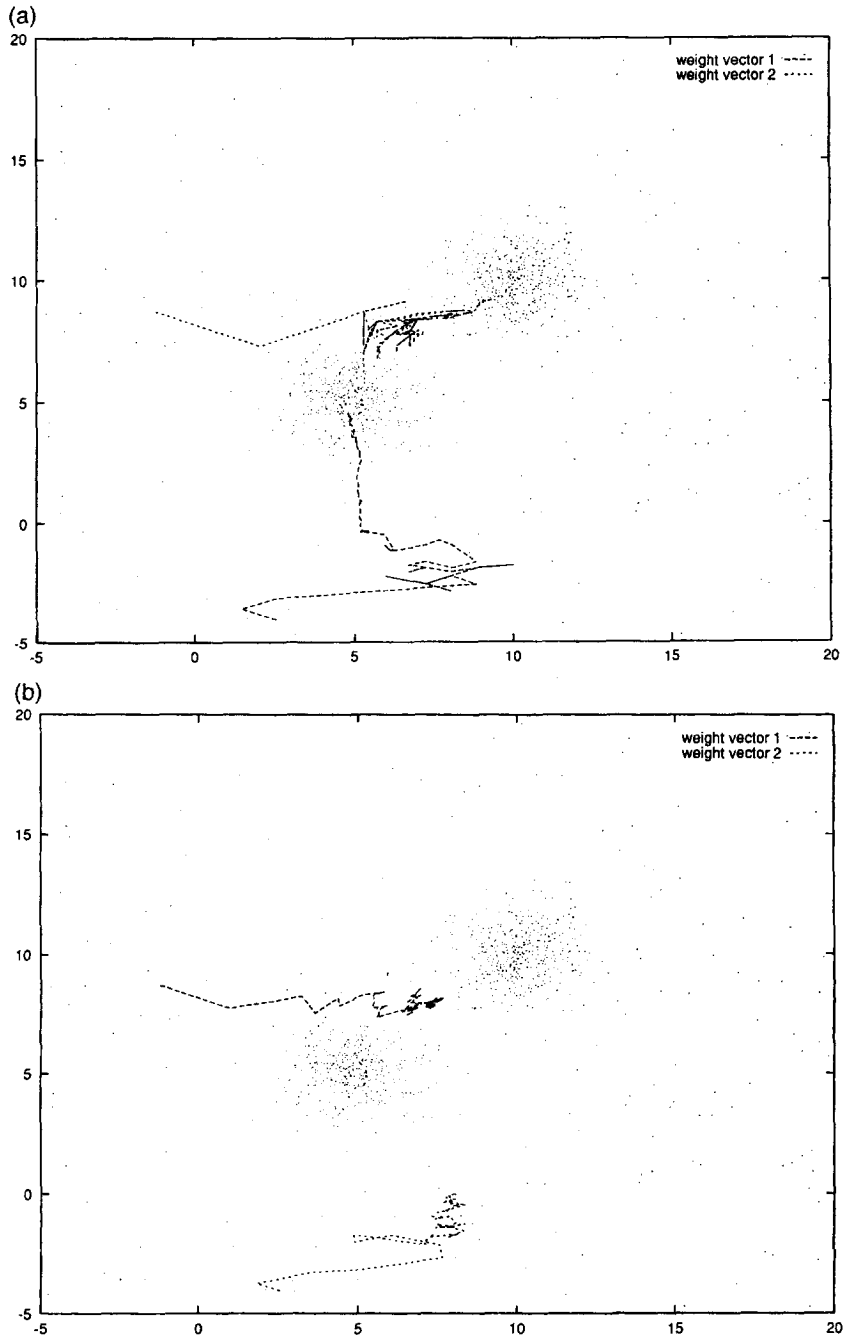

Fig. 8. (a) Trajectories of the MMLVQ reference vectors $\mathbf{w}_i(n)$ $i = 1,2$ for two classes contaminated by uniformly distributed noise; and (b) trajectories of the linear LVQ reference vectors $\mathbf{w}_i(n)$ $i = 1,2$ for two classes contaminated by uniformly distributed noise.

[23–26]. Some of them postulate that the transmission channel is linear and then, by means of (i) the intercorrelation function, or its improved extensions [27], and (ii) the coherence function coupled with the phase of the cross-spectrum, they estimate the delay and an index of causality. The works of others [23,26] take into account the possible non-linearity relationships between the two observation sequences. As is shown here, an alternative approach to EEG signal processing can benefit from the most recent theoretical insights provided by the higher order moments [28]. The present study consists of comparing four methods of time delay estimation and linear and non-linear links determination. The preliminary results obtained on simulated and in vivo signals are also reported.

2.4.1. Methods

The objective is to compare three well established methods (the linear and non-linear indexes and the transmission coefficient) versus a new one based on higher order moment: the bispectrum.

(i) The first one makes use of the normalised correlation coefficient given by:

$$r^2(\tau) = \frac{[E\{(X_1(n) - E\{X_1(n)\})(X_2(n+\tau) - E\{X_2(n)\})\}]^2}{\mathrm{var}\{X_1(n)\}\mathrm{var}\{X_2(n)\}}$$

where $E\{.\}$ and $\mathrm{var}\{.\}$ represent the expectation and the variance, respectively.

(ii) The second one, proposed by Mars and Arragon [23], is based on the computation of the marginal $H(X_1), H(X_2)$ and cross $H(X_1,X_2)$ entropies of the two processes. The transmission coefficient, evaluated for different delays τ between the channels, is then equal to:

$$T(\tau) = \frac{H(X_1(n)) + H(X_2(n+\tau)) - H(X_1(n),X_2(n+\tau))}{\min\{H(X_1(n)),H(X_2(n+\tau))\}}$$

Among the several procedures proposed for entropy evaluation, the algorithm described in [25] has been retained.

(iii) The third approach was recently introduced by Moddemeijer [26], its interest being that, unlike the previous ones, no assumption is made on the transmission channel. It consists of building a scatter plot of X_2 versus X_1 and the process $X_2(n+\tau)$ is approximated by successive linear straight lines. The non-linear regression coefficient is then:

$$h^2 = \frac{\mathrm{var}[X_2(n)] - E\{[X_2(n) - f(X_1(n-\tau)]^2\}}{\mathrm{var}[X_2(n)]}$$

The three methods are non-parametric; their maximum values, from which the time delay $\tau = D$ can be derived, range from 0 to 1 and are close to 1 if high dependency exists between the channels. The two first functions are symmetric while h^2 is not.

(iv) By definition, the bispectrum is the 2D-Fourier transform of the 3rd order moment. For a stationary real discrete centered random process $X_1(n)$, the 3rd order moment is expressed as:

$$R_{X_1X_1X_1}(m,n) = E\{X_1(k) \times X_1(k+m)X_1(k+n)\}$$

Thus, the bispectrum is:

$$B_{X_1X_1X_1}(\omega_1,\omega_2) = \sum_m \sum_n R_{X_1X_1X_1}(m,n) \times \exp(-j(\omega_1 m + \omega_2 n)) = FT\{R_{X_1X_1X_1}(m,n)$$

where $|\omega_1| \le \pi |\omega_2| \le \pi$, $\omega_1 + \omega_2 \le \pi$ and ω_i are normalised pulse frequencies and FT represents the Fourier Transform.

In the same way, the cross-bispectrum between the channels $X_1(n)$ and $X_2(n)$ can be defined as:

$$B_{X_1X_2X_1}(\omega_1,\omega_2) = FT\{R_{X_1X_2X_1}(m,n)$$

The bispectrum can be computed directly or indirectly; the direct method presented by Nikias and Pan [29], based on the Fourier Transform of each process, has been used here. If Tr is a linear transformation, then the argument of the cross-bispectrum is linear in w_1 for a given w_2. Thus, the following function can be evaluated:

$$b(\tau) = \int\int_{-\pi}^{\pi} \frac{B_{X_1X_2X_1}(\omega_1,\omega_2)}{B_{X_1X_1X_1}(\omega_1,\omega_2)} \exp(j\omega_1\tau) d\omega_1 d\omega_2$$

which is maximum for $\tau = D$. In contrast to the three previous ones, this method does not provide information about the nature of the links. This could be done by its natural extension the cross-bicoherence [30].

2.4.2. Results: comparison between linear and non-linear time delay estimation in epilepsy

2.4.2.1. Simulated experiments. Several sequences of 4 s of on-depth observations (SEEG) sampled at 200 Hz were interactively chosen, transformed, delayed and noise was added. The performances of the methods were evaluated with respect to the signal to noise ratio, to different non-linearities, described by Pijn [26], and to the observation sequences. Specific sequences, depicting different spectral characteristics, have been considered: the first one is a background activity observed between seizures; the second one, a rapid discharge measured at the beginning of the seizure; and the last one, an impulse-like sequence observed during the seizure. Only one delay was simulated and corresponded in practice to a small delay (5 samples). The bispectra and cross-bispectra were estimated over segments of 128 samples, then averaged over 20 blocks with an overlapping of 50%.

When no noise is added, on several background, discharge and impulse activities (10 different sequences of each), for all the non-linearities considered here, we have observed that the non-linear regression coefficient h^2 and the transmission coefficient T have the better behavior. In a degraded context (e.g. with additive noise), the conclusions are different. Fifty noise realisations were repeated to experimentally estimate the bias and the variance of each time delay estimator and similarity indexes. Fig. 9 represents the mean behaviour of the four functions when a quadratic transformation is applied on the input signal. It shows that only the transmission coefficient $T(\tau)$ and the non-linear index $h^2(\tau)$ are able to estimate the delay. Moreover the examination of the mean value of the maximum of the functions is close to 1 and suggests that the transmission channel could be non-linear (when one compares the linear correlation coefficient) and some causality between the two signals could exist. For another non-linear transformation and in the same measurement conditions (background activity signal, Gaussian noise, number of realisations) the results are quite different (Fig. 10). In this case, only the bispectrum is able to estimate the delay and the other functions do not inform about the causality, the possible non-linear link of the transmission channel and the delay.

A summary of the results is depicted in Tables 2 and 3 for Gaussian and non-Gaussian noises, respectively. It clearly points out that the performances are directly related to the input signal, the transformation and the noise, and underlines that great care must be taken in the interpretation of real contexts.

2.4.2.2. Processing of real signals. The four methods were tested on real signals recorded with depth electrodes on a patient suffering from temporal epilepsy. The behaviour (Fig. 11a) of three internal structures — the corpus amygdaloideum (A_{12}), the anterior cornu ammonis (B_{12}) and the posterior cornu ammonis (C_{12}) — are depicted and studied here. The similarity indexes have been computed every 5 seconds and their behaviour is depicted in Fig. 11b. They demonstrated that a high synchronism exists between each channel at time 55, which corresponds to a possible beginning of the seizure. The similarity indexes progressively decrease and again increase around the instant 75. Moreover, a closer examination of the index shows that the structures are first linearly related at the beginning of the seizure and move progressively to a non-linear relation around time 75. The delays estimated at these time instants are method dependent but are vary-

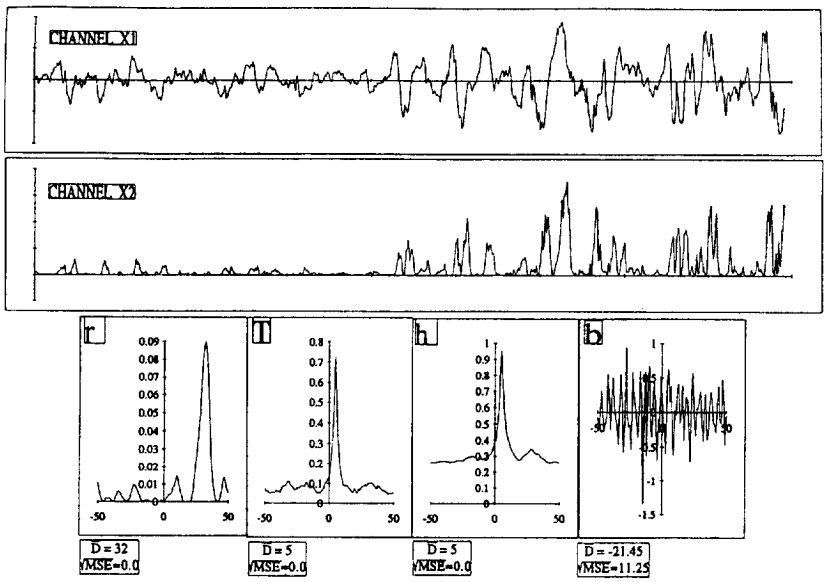

Fig. 9. Mean behavior of time delay estimators after fifty noise runs. The transformation is $X_2(n) = X_1(n-5) \times X_1(n-5) + b_2(n)$. Input sequence: background activity-S/N = 10 dB. D and MSE represent the mean value and mean standard deviation of the estimated delay.

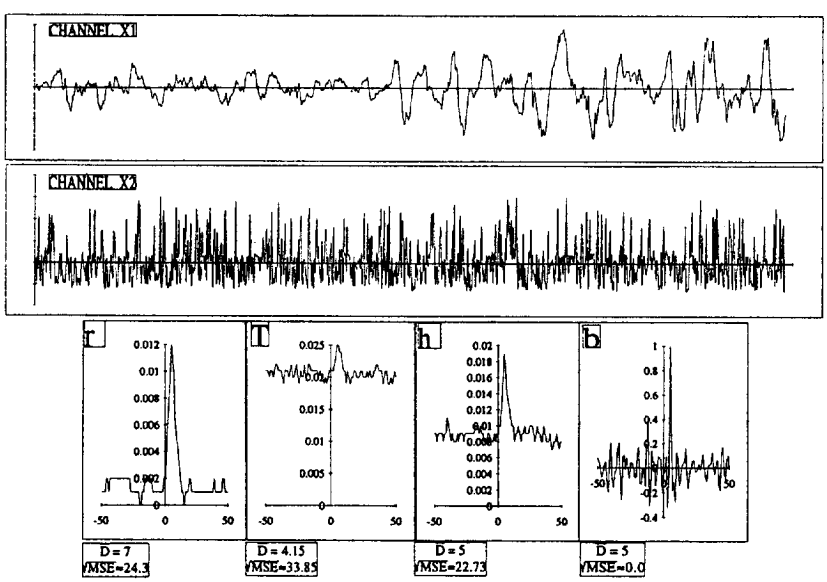

Fig. 10. Mean behavior of time delay estimators after fifty noise runs. The transformation is $X_2(n) = sgn[X_1(n-5) \times X_1(n-5) + b_2(n)$. Input sequence: background activity-S/N = 10 dB. D and MSE represent the mean value and mean standard deviation of the estimated delay.

Table 2
Comparison of the criteria after 50 Gaussian noise runs

	Non-linearity similarity index					Transmission coefficient				Linear correlation index				Interbispectrum	
Transformations	Signals	D	SD D	H_2	SD H_2	D	SD D	T	SD T	D	SD D	r^2	SD r^2	D	SD D
X	Background	−5	0	0.9	0.01	−5	0	0.53	0.04	−5	0	0.9	0.001	−5	0
	Discharge	−5	0	0.9	0.001	−5	0	0.54	0.02	−5	0	0.9	0.01	−5	0
	Impulse signal	−5	0	0.9	0.01	−5	0	0.44	0.028	−5	0	0.9	0.01	−5	0
$X \times X$	Background	−5	0	0.99	0.01	−5	0	0.75	0.001	19	0	0.18	0.001	−5.75	26.99
	Discharge	−5	0	0.98	0.01	−5	0	0.81	0.001	−8	0	0.04	0.001	−6.35	15.36
	Impulse signal	−5	0	0.92	0.001	−6.4	23.38	0.21	0.04	−8	0	0.25	0.001	−7	0
Sgn(X)sqrt	Background	−0.5	22.73	0.04	0.01	−4.75	33.85	0.04	0.03	−2.2	24.3	0.02	0.01	−5	0.3
(ABs(X))	Discharge	0.1	28.8	0.04	0.01	−9.15	25.26	0.04	0.01	3.45	24.33	0.02	0.01	−5	0
	Impulse signal	−8.5	22.33	0.04	0.01	14.9	21.5	0.04	0.01	−9.85	15.95	0.02	0.01	−3.9	25.88

D represents the estimated mean value; SD D represents the estimated standard deviation.
H_2, T or r^2, the estimated mean values of the maximum of each function, respectively; SD H_2, SD T, SD r^2, the estimated standard deviations of the maximum of each function. The grey cases indicate that the estimators have a poor behavior.

Table 3
Comparison of the criteria after fifty non-Gaussian noise runs

	Non-linearity similarity index					Transmission coefficient				Linear correlation index				Interbispectrum	
Transformations	Signals	D	SD D	H_2	SD H_2	D	SD D	T	SD T	D	SD D	r^2	SD r^2	D	SD D
X	Background	−5	0	0.91	0.01	−5	0	0.59	0.022	−5	0	0.91	0.01	−5	0
	Discharge	−5	0	0.92	0.01	−5	0	0.77	0.024	−5	0	0.92	0.01	−5	0
	Impulse signal	−5	0	0.9	0.01	−5	0	0.7	0.014	−5	0	0.91	0.01	−5	0
$X \times X$	Background	−5	0	1	0	−5	0	0.8	0.001	18	0	0.19	0.001	−19	0
	Discharge	−5	0	1	0.01	−5	0	0.76	0.001	−5	0	0.18	0.001	−4	0
	Impulse signal	−5	0	0.92	0.001	−60.4	34.59	0.21	0.024	−8	0	0.25	0.001	−7	0
Sgn(X)sqrt	Background	−11.25	31.56	0.28	0.01	−5	0	0.23	0.05	−6.6	18.09	0.02	0.01	1.15	31.98
(Abs(X))	Discharge	−6.85	4.49	0.35	0.014	−5	0	0.51	0.026	−4.05	10.36	0.1	0.022	1.8	23.19
	Impulse signal	−8.55	33.97	0.28	0.01	−5	0	0.43	0.02	−4.15	12.34	0.02	0.01	−4.55	29.45

The non-Gaussian noise is generated by passing a Gaussian noise through a non-linear filter and by subtracting the mean value.
D represents the estimated mean value; SD D represents the estimated standard deviation.
H2, T or r^2, the estimated mean values of the maximum of each function, respectively; SD H2, SD T, SD r^2, the estimated standard deviations of the maximum of each function. The grey cases indicate that the estimators have a poor behavior.

ing in the same way. In other words, they clearly show that the signal, observed on A_{12}, drives the other two. The relations between C and B are more difficult to interpret (B is first in advance and then delayed).

2.5. *Non-linear dynamics: a deterministic chaotic approach*

The non-linear behaviour of biological systems may be also studied through the dynamic characteristics of the generated signals. Under certain hypotheses, the correspondent time series may provide information about the complexity of the system under investigation. The sequence of samples is studied in the space-state approach which is able to depict periodical or non-periodical solutions, stable and equilibrium points, and evaluation of geometrical dimension, up to erratic behaviour which, in many cases, can derive from a chaotic attractor. To discriminate between random and chaotic behavior is a fundamental task

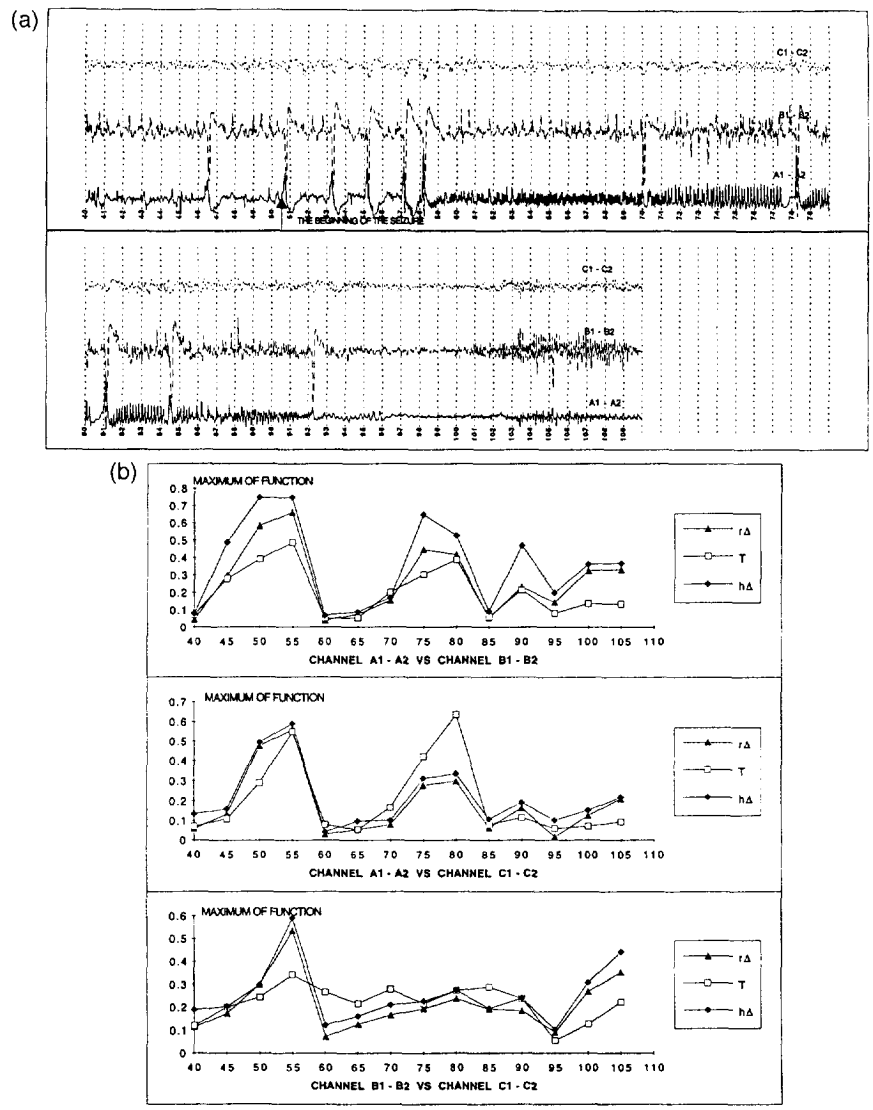

Fig. 11. (a) Example of temporal epileptic seizure; and (b) behaviour of the correlation indexes computed every 5 s between time instant 40–105 s on the signal depicted in Fig. 3a.

for correct comprehension of the system dynamics. In this paper some basic methods for obtaining the invariant characteristics of the system attractor through the calculation of fractal (non-integer) dimension, entropy and chaotic indexes like Lyapunov exponents, are proposed. The described application is aimed at studying complex neural controlling mechanisms of heart rate variability (HRV) regulation.

2.5.1. Methods

As the model of the system is unknown, we have to reconstruct the system state, starting from a measured time series.

2.5.1.1. State-space reconstruction. The dynamic of a system with n state variables is completely described by a vector $\mathbf{v}_m(k)$ obtained from a single measured variable through an embedding procedure. That represents the scalar data with

the sequence of discrete time series of the original sampled signal $x(k)$, where

$$\mathbf{v}_m(k) = [x(k), x(k+\tau), \ldots, x(k+(m-1)\tau)]$$

$$k = 1, N-(m-1)\tau$$

is a vector embedded in a space of dimension m and delayed by an interval τ.

This is established by Takens' theorem. He demonstrates that the attractor reconstructed through this procedure is definitely different from the original but it the same qualitative features, thus preserving geometrical invariant as fractal dimension and Lyapunov exponents, if $m \geq 2n+1$ [31].

2.5.1.2. Choice of m and τ values. This is a fundamental step in order to fulfill a correct analysis of attractor properties starting from a discrete series. The components of the reconstructed state vector will be too uncorrelated if the time delay is too high. On the other hand, too small time delays could bring the system attractor dynamics on the bisector of the m-dimensional space. Then it would be impossible to fully investigate the complex structure of the system attractor. The existence of underlying structures with different degrees of complexity can be properly revealed depending on the choice of the reconstruction step. Two different criteria have been used to choose the optimum τ step value:

(i) assuming the first zero of the autocorrelation function; and
(ii) measuring the first local minimum of the mutual information criterion [32]. It evaluates the joint probability density P between p and q where $[p,q] = [x(\tau), x(t+\tau)]$.

These criteria provide very different estimations of τ interval: in this way more than one value exists for the time delay choices which have to be evaluated, for example, by using them in the estimation of attractor fractal dimension, as more fully explained below.

Embedding dimension m can be determined by an a posteriori method which estimates the system complexity and identifies the optimum value as $m \cong 2D+1$. The method we propose in this paper is the test on false nearest neighbors (FNN), based on the evaluation of the distance between couples of points in the state-space, for growing values of m, until self crossings of trajectory, which are generated by a projection in a low-dimensional space, can be excluded [33].

After the reconstruction procedure, the quantitative analysis of the attractor properties allows determination of the fractal dimension of system attractor and the Lyapunov exponents. The flow charts depicted in Figs. 12 and 13 summarize the conceptual procedure for assessing the two invariant parameters.

2.5.1.3. Fractal dimension D. The geometric dimension of the system attractor is described by the fractal dimension parameter whose non-integer values are imposed by the fractal structure of the attractor in the state-space.

The estimation of the fractal dimension for experimental data sets can be performed using

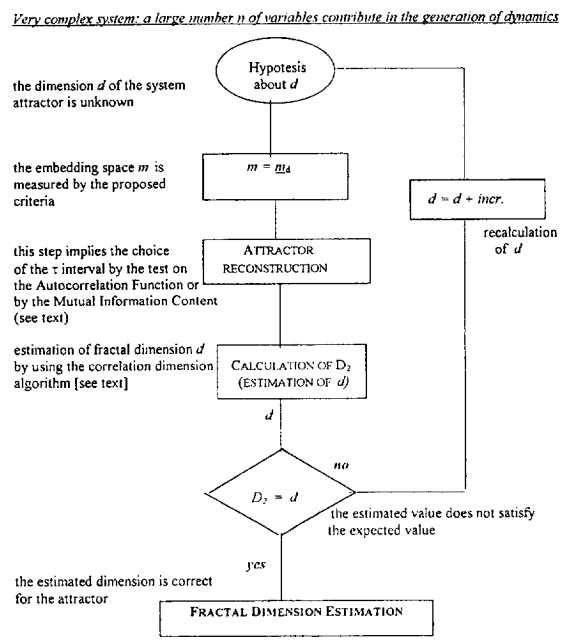

Fig. 12. Flow chart of the procedure for the estimation of fractal dimension of the system attractor when the model of the system is unknown. The illustrated procedure summarizes the conceptual steps as presented in the method section.

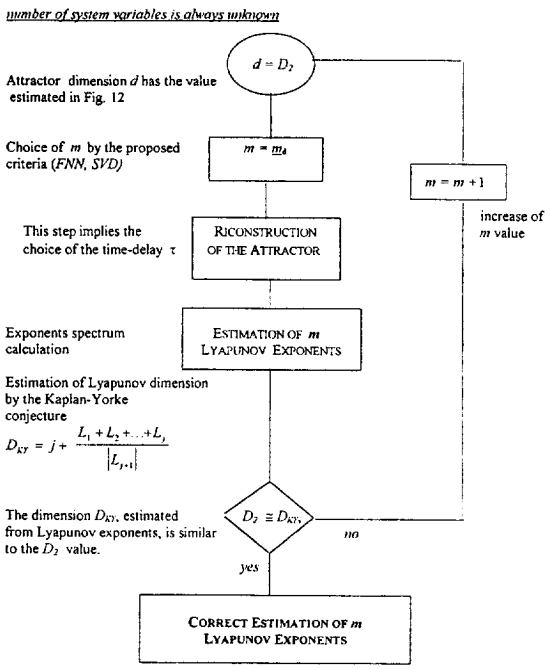

Fig. 13. Flow chart of the estimation of Lyapunov exponents spectrum starting from a trajectory reconstructed from a time series and taking as reference value the estimation D_2 of the fractal dimension. The operative rule, for the choice of the correct exponent number, based on the comparison between D_2 and D_{KY} value, is illustrated.

the correlation dimension parameter (D_2). D_2 provides an inferior bound for the fractal dimension D of the system attractor as it was demonstrated that $D_2 \leq D$ [34]. Its value will reach the true dimension value if the whole attractor is investigated: this means that a long enough series of data have to be considered. D_2 could also represent an index of system complexity in terms of number of state variables which contribute to generate dynamics in the state-space. The algorithm evaluates the spatial correlation between attractor points calculating a correlation integral for each different m dimension:

$$C_m(\epsilon) = \lim_{N\to\infty} \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} \vartheta$$

$$-\left(\epsilon - \left|v_m(k_i) - v_m(k_j)\right|\right)$$

when N is the number points, $v_m(k)$ the state vector reconstructed in the m-dimensional space, ϵ the correlation length and θ the Heaviside distance.

For small ϵ values, it can be demonstrated that $C(\epsilon) \cong \epsilon^{D_2}$. Consequently, D_2 value is obtained as the slope of the linear zone in the plot of $C(\epsilon)$ versus ϵ in log-log scale. The slope of the curve is an increasing function of the embedding value m until a saturation is noticed, providing the D_2 value. This corresponds, in the flow chart of Fig. 12, to change the hypothesis about the dimension D which is unknown and, in the same time, to increase the m value. The procedure stops when the calculated and the supposed value of D do not change for any increase of parameters. Saturation should not happen for non-deterministic series but spurious saturation could be found if the signal is a stochastic colored noise. For this reason, it is necessary to perform a determinism test on time series by destroying the Fourier phases and substituting them with a sequence of random numbers [35]. If the saturation of D_2 parameter is due to a stochastic colored noise it is likely that it does not change after randomization: on the contrary, deterministic signals do not show saturation anymore after randomization.

2.5.1.4. Lyapunov exponents. Chaotic systems, in particular, show marked dependence from initial conditions. That means that small differences at the starting time will be relevant after few time intervals of system evolution, as trajectory of chaotic system attractor exponentially diverges. Lyapunov exponents (L_i) represent a quantitative measurement of this phenomenon: a positive L_i means chaos. The larger the exponent, the more chaotic the system.

In the trajectory evolution, after a time t, an infinitesimal hypersphere of radius r centered in a point that belongs to the attractor in the F dimension hyperspace, will become an ellipsoid. Between its principal axis $r(t)$ and the initial radius r_0, a correspondence is then established

$$r(t) = r_0 e^{L_i t}$$

where L_i is the i-th Lyapunov exponent (measured in bit/s).

Each principal direction of the ellipsoid undergoes a deformation which is univocally described by one Lyapunov exponent. Stretching and folding process originates a divergence and then always in chaotic systems, almost one L_i is positive. Negative values of exponents indicate a convergent trajectory instead. For a periodic signal, due to the superimposition of all trajectories, L_i is equal to 0 and no convergence or divergence is noticed. Completely casual signals are characterized by equal value exponents as they have the same number of convergent and divergent trajectories. The number of L_i equals the number of system dimension and the entire set of L_i is called spectrum of exponents. Each L_i identifies a direction in the state-space. The method we present here is from Eckmann et al. [36]. If the system evolution law is known, as in the classical attractors (i.e., of Lorenz, Rossler, Henon and so on), the estimation of L_i from trajectory is correct. In biological time series a large amount of data is required in order to enhance the dynamics of trajectory evolution in the whole attractor, thus assuring a more consistent statistical analysis with regard to the attractor properties. Furthermore, as the state-space dimension of the system attractor is unknown, the algorithm calculates a spectrum of exponents whose number equals the m dimension (m is a parameter of the algorithm). Thus if m is too large the so-called 'spurious exponents' are introduced by the algorithm, changing the spectrum in an erroneous way. Fig. 13 illustrates the estimation procedure of LE, by providing an operative criterion for the choice of the correct dimension of the state-space reconstruction and then for the correct LE spectrum estimation.

Kaplan and Yorke conjectured that from the LE spectrum one could also obtain information about the fractal dimension of the attractor. Lyapunov or Kaplan-Yorke (D_L or D_{KY}) dimension is defined as follows:

$$D_{KY} = j + \frac{L_1 + L_2 + \ldots + L_j}{|L_{j+1}|}$$

j is the progressive number of exponents obtained when adding the exponents themselves one at a time (in decreasing value) until the sum remains positive. The $j+1$ exponent changes the sign of the sum (from positive to negative) [37]. The idea is that only the positive dynamical contributions, coming from the stretching of trajectories, are involved in the generation of the fractal structure of the attractor.

The conjecture provides a link between dynamic and geometric properties of the system attractor. In the block diagram the value of D_L is compared with the D_2 value previously estimated and the correct dimension m of the spectrum is obtained when the two estimations of fractal dimension provide similar values.

This approach also allows estimation of the entropy value K. It is defined as the sum of positive exponent values, supposing that the stretching behavior generates fractal structure of the attractor.

2.5.2. Results: non-linear dynamics in heart rate variability (HRV) signal analysis

The methods previously introduced have been applied to HRV signal which contains information about the regulation, of the cardiovascular system, by the autonomic nervous system (ANS) through the action of its sympathetic and parasympathetic branches. The coupling between these two physiological oscillators generates the variety of behaviour one can identify in the HRV signal.

The series of R-R intervals in the 24 h was collected from ECG Holter recordings through the automatic recognition of QRS maxima [38]. Nine normal subjects (N) and patients with cardiovascular diseases and nine with severe heart failure (HF) and 7 with orthotopic transplants (T) were studied to assess if non-linear mechanisms were identifiable in cardiovascular regulating dynamics. Series of approximately 20 000 or more points were selected in the 24 h, separating day and night epochs. The entire amount of the series is approximately 100 000 or more samples.

Correlation dimension and entropy show a decrease passing from normal to pathological subjects. $D_2 = 6.8 \pm 1.41$ for N, 5.2 ± 1.8 for HF and

4.5 ± 1.5 for T patients. HRV signal in pathological subjects seems more predictable and then depleted of information in respect of the normal ones. The number of state variables involved in the normal HRV regulation is high and the state-space representations of trajectories also confirms it. It is difficult to identify orbits or structures in N subjects by using the 2-dimension maps; for pathological and less complex patients instead interesting pseudo-regular orbits are recognisable.

Finally, the positivity of the Lyapunov exponents, when the entire spectrum is calculated, is a good confirmation of the chaotic dynamics in the HRV signal. For really noisy data the overall spectrum of λ_i also provides the estimation of entropy (sum of positive exponents) and D_{KY}. This confirms an information (or complexity) loss in pathological subjects. First λ_i value decreases passing from healthy to diseased subjects and from the day to the night epoch.

3. Conclusion

The methods introduced in the present paper, as well as the reported results, confirm the hypothesis that non-linear algorithms of signal processing are capable of enhancing important information from both the physiological and clinical points of view. In many instances they present a better performance in respect to the linear ones. On the other hand, it is important to stress the fact that many of them are still at the development phase and need further evaluation and validation. From a review of the presented methods of signal processing it is possible to establish the following conclusive remarks.

Burst Suppression (BS) pattern appears in EEG signal after ischaemic brain damage or in deep anaesthesia. In clinical work the pattern can occasionally be seen during clinical operations. The more important, in the case of treatment of intractable epilepsy by means of deep anaesthesia, BS EEG is often monitored for days and computerised analysis of the data is needed. In this paper we showed that due to the non-linearities in the BS pattern, median type algorithms can have an advantage over linear methods in processing and detection of the pattern. Successful detection of the DC-shift during bursts enables the distinguishing of BS pattern from similar patterns during normal sleep. It has been pointed out recently that epileptic discharges can occur during BS and they are often difficult to discriminate from bursts [39]. Using the Tsay's non-linearity test, Loula et al. [40] found the respiration signal was essentially of non-linear characteristics. We suggest that algorithms based on non-linear median type operation can be useful in processing this kind of signal.

The disadvantage of using median type filters has been the lack of theoretical tools describing the performance of the filters. However, the optimisation algorithms significantly reduce the risk of getting unwanted results. Much work is currently being done to further extend the theoretical basis of median type filtering. In the case of long filter windows, which usually occurs in processing physiological signals, the sorting operation is time consuming. However, this disadvantage can be overcome by choosing a filter structure which reduces the number of inputs to the WOS operation by a suitable combination of FIR subfilters. For a certain application, a sufficiently good initial weight vector, common to all cases, can usually be found and fast adaptation algorithm of Eq. 2.1.12 can be used. Median type filters have been successfully applied to trend detection problem in physiological signal processing. The recent development in the theoretical background of median type filtering broadens the area of applications of these filters significantly. These applications include, e.g. pattern recognition, adaptive segmentation, data compression, prediction, and data modelling. The performance of median type methods should be compared with that of conventional methods and in those applications where the enhancement in performance outweighs the increased computational complexity, the new methods should be used.

One of the major drawbacks of the conventional evoked potential technique in clinical applications is the relatively long data acquisition period necessary to obtain sufficient averages for reaching an acceptable signal to noise ratio. Using the technique presented here, this disadvantage is

significantly decreased. A paper currently under preparation (van de Velde et al.) shows that in auditory evoked potential monitoring a decrease in necessary data acquisition time is required to obtain a satisfactory signal to noise ratio of up to 60%.

The possibilities of obtaining information about the non-linear properties of sensory pathways using this technique seem to be limited to the characterisation of second order effects. The small number of averages available for the determination of higher order terms is insufficient to obtain a reasonable signal to noise ratio. Until now, we did not investigate the second order effects in a quantitative manner and therefore few statements can be made or conclusions drawn about the practical meaning of these data.

The Wiener-Volterra technique described here has been applied in various clinical studies, mainly during anaesthesia. The main advantage is that the first order kernel, $h_1(\tau)$ can be used as a substitute for the conventional evoked response providing a better signal to noise ratio in a smaller data acquisition period. The usefulness of the second-order effects in clinical practice remains unclear. It must be emphasized that the data presented here are acquired under optimal laboratory conditions and it will be difficult to further improve the signal to noise ratio. Nevertheless this technique allows the investigation of non-linear properties without using parametric models and as such it provides a powerful tool for basic neurophysiological brain research.

The use of multivariate estimators of location that are based on data ordering gives the proposed LVQ variations robustness against erroneous choices of the winner vector, a feature that can be very important when overlapping clusters exist. Furthermore, outlying observations, i.e. observations that do not belong to any of the clusters, have little effect on the weight vectors. These properties result in an enhanced performance of the proposed algorithms. Moreover, the performance of both MMLVQ and VMLVQ does not depend on the choice of any parameters, unlike the linear LVQ whose performance depends on the choice of an appropriate adaptation step sequence $a(t)$. Finally, when the input data are integer-valued the learning procedure of MMLVQ does not involve any floating point arithmetic. If this feature is used in conjunction with the running median calculation algorithm, a very computationally efficient realisation of MMLVQ can result.

The clustering ability of the proposed algorithms can be of great importance in many biomedical applications where multivariate data are involved, e.g. in cases where simultaneous recording and interpretation of a number of physiological variables is required. In such cases the new algorithms can be used to segment in an unsupervised and robust way the input space into a number of meaningful entities, thus revealing the inherent structure of the input data as well as the relations between the recorded variables.

Results on simulated and experimental signals may lead to estimate causality among EEG signals. This study, based on a quantitative comparison, has shown that the interpretation of real data is not an easy task due to the sensitivity of the methods with respect to the nature of the signal and the non-linear transformation. This study raises others fundamental questions about (i) the delay, (ii) the measurement of the causality, (iii) the estimation of the non-linearity. A major problem is to estimate objectively the performance of methods on real data. For example, what does the delay means when an observed sequence is built from a mixture of signals with different delays? How could one take into account the dynamic behaviour of the transmission channel as suggested by the example here described? At a methodological level, we are now developing dependency tests based upon the statistics of the coherence function [27] and the cross-bicoherence [30] that could help in solving the measurement of the causality.

The failure of non-parametric methods for some non-linear transformations calls for the evaluation of linear and non-linear models (such as Volterra series).

The calculation of various and different parameters in the long-term (24-h) ECG recordings suggests that non-linear dynamics are involved in the generation mechanism of HRV signal. In particular, geometric invariants such as D_2

(correlation dimension) and K (entropy) show that the signals involved are characterised, in the normal cases, by a high value of complexity, which decreases when considering the pathological subjects. Further, the fact that a positive Lyapunov exponent is always found in the normals, brings us to the conclusion that a possible chaotic mechanism might be considered in the complex physiological model underneath. Such a positivity significantly decreases when passing to the pathological subjects. The plotting of the samples in the state-space diagram presents various orbits and trajectories which may be considered as a realisation of a strange attractor phenomenon: in particular, systems with lower complexity (generally less than 3.0, such as in transplanted heart patients) show such a behaviour directly in a two-dimensional plot. The interpretation of results seems consistent with a complex cardiovascular physiology which presents characteristics of multidimensionality, redundancy and shows, in the long-term regulation, chaotic properties as well as fractal geometrical structures.

References

[1] J.W. Tukey, Non-linear (non-superposable) methods for smoothing data, Cong. Rec. EASCOM (1974) 673.

[2] N.C. Gallager Jr. and G.W. Wise, A theoretical analysis of the properties of median filters, IEEE Trans. ASSP ASSP-29 (1981) 1136–1141.

[3] I. Pitas and A.N. Venetsanopoulos, Non-linear digital filters: principles and applications (Kluwer, Boston, 1990).

[4] L. Yin, R. Yang, M. Gabbouj and Y. Neuvo, Weighted median filters: a tutorial, IEEE Trans. Circuits Syst. (in press).

[5] P. Heinonen and Y. Neuvo, FIR-median hybrid filters, IEEE Trans. ASSP ASSP-35 (1987) 832–838.

[6] T. Lipping, P. Loula, V. Jäntti and A. Yli-Hankala, DC-level detection of burst-suppression EEG, Meth. Inform. Med. 33 (1994) 35–38.

[7] L. Yin and Y. Neuvo, Fast adaptation and performance characteristics of FIR-WOS hybrid filters, IEEE Trans. Signal Processing 42 (1994) 1610–1628.

[8] N. Wiener, Non-linear problems in random theory (Wiley, New York, 1958).

[9] H.I. Krausz, Identification of non-linear systems using random impulse train inputs, Biol. Cybern. 19 (1975) 217–230.

[10] P.J.M. Cluitmans and J.E.W. Beneken, Non-linear analysis of sensory evoked potentials, pp. 419–420. Proc. Ann. Conf. IEEE EMBS, Orlando, FL, USA, 1991.

[11] P.J.M. Cluitmans, Neurophysiological monitoring of anesthetic depth, Ph.D. Thesis, Eindhoven, 1990.

[12] Special issue on neural networks, I: Theory and modelling; II: Analysis, techniques, and applications, Proc. IEEE 78, 9–10, pp. 1409–1680, September–October 1990.

[13] Kohonen T., Self-organisation and associative memory, 3rd edn. (Springer-Verlag, Berlin, Heidelberg, 1989).

[14] J.G. Proakis and D.G. Manolakis, Introduction to digital signal processing (Macmillan, New York, 1988).

[15] P.J. Huber, Robust statistics (Wiley, New York, 1981).

[16] I. Pitas and A.N. Venetsanopoulos, Non-linear digital filters: principles and applications (Kluwer Academic Publishers, Hingham, MA, 1990).

[17] V. Barnett, The ordering of multivariate data, J. R. Statist. Soc. A 139 (3) (1976) 318–354.

[18] T.S. Huang, G.J. Yang and G.Y. Tang, A fast two-dimensional median filtering algorithm, IEEE Trans. Acoustics, Speech Signal Processing 27 (1) (1979) 13–18.

[19] J. Astola, P. Haavisto and Y. Neuvo, Vector median filters, Proc. IEEE 78 (4) (1990) 678–689.

[20] C. Kotropoulos, I. Pitas and M. Gabbouj, Marginal median learning vector quantizer, in European Signal Processing Conference '94, Edinburgh, UK, 1994, pp. 1496–1499.

[21] C. Kotropoulos, N. Nikolaidis, I. Pitas and P. Kiniklis, Marginal median learning vector quantizer, IEEE Trans. Signal Processing (submitted).

[22] P. Chauvel, P. Buser, J.-M. Badier, C. Liegiois-Chauvel and P. Marquis, Bancaud, La 'zone épileptogène' chez l'homme: représentation des évènements intercritiques par cartes spatio-temporelles, Revue Neurologique 143 (1987) 443–450.

[23] N.J.I. Mars and G.W. Arragon, Time delay estimation in non-linear systems, IEEE Trans. ASSP ASSP-29 (3) (1981) 619–621.

[24] J. Gotman, Measurement of small time differences between EEG channels: method and application to epileptic seizure propagation, Electroencephalogr. Clin. Neurophysiol. 56 (1983) 501–504.

[25] R. Moddemeijer, Delay estimation with application to electroencephalograms in epilepsy, Ph.D. Thesis, University of Twente, Enschede, 1989.

[26] J.P. Pijn, Quantitative evaluation of EEG signal in epilepsy, Ph.D. Thesis, Amsterdam, 1990.

[27] G.C. Carter, Coherence and time delay estimation, Proc. IEEE 75, (2) (1987) 236–250.

[28] C.-L. Nikias and M.-R. Raghuveer, Bispectrum estimation: a digital signal processing framework, IEEE Trans. ASSP ASSP-36 11 (1988) 1706–1714.

[29] C.-L. Nikias and R. Pan, Time delay estimation in unknown Gaussian spatially correlated noise, Proc. IEEE 75 (7) (1988) 869–891.

[30] B. Saltzberg and D. William, Electrophysiological measures of neuronal interactive coupling, linear and nonlinear relationships among multiple electroencephalographic recordings, Int. J. Biomed. Computing 18 (1986) 77–87.

[31] F. Takens, Detecting strange attractors in turbulence, in Dynamical systems and turbulence, Lecture notes in Mathematics, eds. D.A. Rand and L.S. Young, 898, pp. 366–381 (Springer, Berlin, 1981).

[32] H.L. Fraser and Swinney, Independent coordinates for strange attractors from mutual information, Phys. Rev. A 33 (1986) 1134–1140.

[33] H.D.I. Abarbanel, R. Brown, J.J. Sidorowich and L. Tsimring, The analysis of observed chaotic data in physical systems, Rev. Mod. Phys. 65 (1993) 4.

[34] P. Grassberger and I. Procaccia, Measuring the strangeness of strange attractors, Physica 9D (1983) 189–208.

[35] A. Provenzale, L.A. Smith, R. Vio and G. Murante, Distinguishing between low dimensional dynamics and randomness in measured time series, Physica D 58 (1992) 31–49.

[36] J.P. Eckmann, O.S. Kamphorst, D. Ruelle and S. Ciliberto, Lyapunov exponents from time series, Phys. Rev. A 34 (1986) 4971–4979.

[37] P. Brown, Bryant and H.D.I. Abarbanel, Computing the Lyapunov spectrum of a dynamical system from an observed time series, Phys. Rev. A 43 (6) (1991) 2787–2806.

[38] S. Cerutti, G. Baselli, A.M. Bianchi and M.G. Signorini, Spectral techniques of analysis for blood pressure and heart rate signals, in Blood pressure and heart rate variability, eds. M. Di Rienzo et al. (IOS Press, 1993).

[39] V. Jäntti, K. Eriksson, K. Hartikainen and G.A. Baer, Epileptic EEG discharges during burst suppression, Neuropediatrics 25 (1994) 271–273.

[40] P. Loula, T. Lipping, V. Jäntti and A. Yli-Hankala, Non-linear interpretation of respiratory sinus arrhythmia in anaesthesia, Meth. Inform. Med. 33 (1994) 52–57.