EEG analysis in cerebral death condition using Wavelet Packet Decomposition

Federico Santoni[1] and Francesco Marrosu[2]

[1] CRS4 Center for Advanced Studies, Research and Development in Sardinia
[2] Cagliari University Hospital

Abstract. This paper proposes a methodology for extracting features from EEG in suspected brain death conditions. Those signals are usually caracterized by a low SNR and in such situations distortions introduced by skull/electrode/cable interface become very relevant. After a calibration procedure, the analysis of EEG signals was performed with a Wavelet Packet Decomposition filter, where the main component of noise are removed with soft thresholding strategy. We propose an estimate of error probability leading to the choice of wavelet basis and threshold parameters. As result, in absence of brain activity, the output of this filter produces really flat signals. However the presence of some activity can be examinated in WP time-frequency domain revealing distinctions between real brain activity, narrow band noise and artifacts.

1 Introduction

The electroencephalographic (EEG) recording is extensively used in Intensive Care Units (ICU) as a reliable measure of electrical brain activity. Moreover, the interpretation of EEG records aimed at diagnosis of the condition of "brain death", better expressed as electro clinical silence (ES), is a crucial step in modern medicine and represents the fundamental pre-requisite in clinical settings oriented for transplants [1]. Although EEG analysis represents a diagnostic tool among many others (e.g. cerebral arteriography, single photon emission tomography), its relatively low cost and its large availability have gained to this procedure the condition of "golden standard" in monitoring comatose patients. However, digital EEG monitoring in ICU is, nowadays, neither popular nor widespread. Among the reasons for this resistance in penetrating the common ICU practice, the most important is that EEG monitoring has yet to be proved as an easy and reliable tool. Theoretically, the analysis of a digital signal seems "prima facie" a relevant vantage over analogue recordings. However, in case of ES, since previous paper EEG yielded results that visually correlate faithfully with clinical signs (the fatal flat line), you should expect that the computerized EEG (CEEG) performs a superior task in this specific diagnosis procedure. Unfortunately, the noising background of an ICU generates insidious artifacts and the required amplification of the small voltage signals cause an amplification of the noise as well.

2 Noise removal by thresholding

The density distribution of a calibrated ES signal is almost gaussian [2]. From this premise, we can state the problem as one of estimating an unknown deterministic signal after observing a process sampled over an interval of length N. We henceforth assume that the observed samples are those of an underlying unknown signal and of white noise, where

$$x[m] = s[m] + n[m], \tag{1}$$

for $m = 1, 2, ..., N$ and $n \sim \mathcal{N}(0, \sigma^2)$.

Implementing an estimator in an orthogonal basis is intuitively appealing on account of the distribution of the noise energy in such a basis. Wavelet bases are known to concentrate the energy of piecewisesmooth signals into a few high-energy coefficients [3]. If the energy is concentrated into a few high-amplitude coefficients, such a representation can provide an accurate estimate of $s[m]$. The advantage of expressing in an orthogonal wavelet basis is two-fold:

a) if the contaminating noise samples are independent and identically distributed (i.i.d.) Gaussian, so are the coefficients, and their statistical independence is preserved.

b) intrinsic properties of the signal are preserved in a wavelet basis.

We first discuss a method for estimating the mean-square error associated with thresholding wavelet coefficients at a given level. Given a signal in some basis representation, we will threshold the coefficients and estimate the resulting error, and this error will then be used in the search for the best basis family. Succintly we report the problem as analysed in [4]. Moreover we present a new derivation used to set the error probability of our detection system.

2.1 Risk of a Wavelet-based estimator

In this section, we present a mean-square error estimator exstensively used in literature expecially for wavelet de-noising [6] . The mean-square error, or more formally the risk related to reconstruction of a vector $\boldsymbol{s}$ with a threshold T, is given by

$$R(\boldsymbol{s}, T) = E\{\|\boldsymbol{s} - \hat{\boldsymbol{s}}\|^2\} \tag{2}$$

where $\hat{\boldsymbol{s}}$ is the representation of the reconstructed vector. A signal reconstruction is obtained by applying some function of a scalar T ("thresholding") to a set of coefficients of $\boldsymbol{s}$ in a given Wavelet basis and then applying an inverse transformation. This is the general procedure we use throughout the paper.

Though many thresholding functions are appliable, for the sake of simplicity, in this section we analyse strictly the risk limited to the hard thresholding rule. Soft thresholding rule will be presented in section 3.2 and for other rules see [8].

General formulation. According to Mallat notation [4] (we used it throghout the following sections) we describe the hard thresholding rule as

$$\gamma_{T}(y_x) = \begin{cases} y_x, & \text{if } |y_x| > T \\ 0, & \text{if } |y_x| \leq T. \end{cases} \tag{3}$$

for any scalar y_x where T is the threshold. Usually the measure of loss caused by recovering signal coefficients, derives from a quadratic distance which depends on T and on signal coefficients y_s:

$$\mathcal{L}\{\gamma_{T}(y_x), y_s, T\} = (y_s - \gamma_{T}(y_x))^2. \tag{4}$$

Considering the reconstructed signal $\hat{s}$ as obtained from the rule (4) applied to the wavelet coefficients in a particular orthogonal basis $\mathcal{B} = \{\boldsymbol{W}_{x_i}\}$ and with $\boldsymbol{\mathcal{W}}_x$ expressing the vector of projection (inner product) of $\boldsymbol{x}$ onto the wavelet basis $\mathcal{W}_{x_i} = \langle \boldsymbol{x}, \boldsymbol{W}_{x_i} \rangle$, the mean value of the loss is the estimation error (2)

$$E\{\mathcal{L}(\gamma_{T}(\boldsymbol{W}_x), \boldsymbol{s}, T)\} = E\{\|\boldsymbol{s} - \boldsymbol{W}_x \gamma_{T}(\boldsymbol{W}_x)\|^2\} = R(\boldsymbol{s}, T), \tag{5}$$

where for compactness $\boldsymbol{W}_x$ represents the matrix of basis functions. Representing the signal $s = \boldsymbol{W}_x \boldsymbol{\mathcal{W}}_s$ with the wavelet basis ($\boldsymbol{\mathcal{W}}_s$ is the vector with components $\langle \boldsymbol{s}, \boldsymbol{W}_{x_i} \rangle$), risk can be expressed in terms of the basis coefficients too, or

$$R(\boldsymbol{s}, T) = E\{\|\boldsymbol{W_x \mathcal{W}_s} - \boldsymbol{W_x} \gamma_{T}(\boldsymbol{\mathcal{W}_x})\|^2\} = \sum_{i=1}^{N} E\left\{|\mathcal{W}_{s_i} - \gamma_{T}(\mathcal{W}_{x_i})|^2\right\}. \tag{6}$$

To better understand the structure of R, we rewrite the function $\gamma_{T}(\mathcal{W}_{x_i}) = \mathcal{W}_{x_i} \mathcal{I}_{|\mathcal{W}_{x_i}| > T}$ where $\mathcal{I}$ is the indicator function. As briefly alluded to earlier, remember that $\mathcal{W}_{x_i} = \mathcal{W}_{s_i} + \mathcal{W}_{n_i}$ and $\mathcal{W}_{n_i} \sim N(0, \sigma^2)$.

After few calculations and considering $\mathcal{W}_{s_i}^2 \mathcal{I}_{|\mathcal{W}_{x_i}| < T} + \mathcal{W}_{s_i}^2 \mathcal{I}_{|\mathcal{W}_{x_i}| \geq T} = \mathcal{W}_{s_i}^2$ we obtain from (6)

$$R(\boldsymbol{s}, T) = \sum_{i=1}^{N} E\left\{\mathcal{W}_{s_i}^2 \mathcal{I}_{|\mathcal{W}_{x_i}| < T} + \mathcal{W}_{n_i}^2 \mathcal{I}_{|\mathcal{W}_{x_i}| \geq T}\right\}. \tag{7}$$

This equation express an intuitive concept: the contribute to the total risk is dued to the signal coefficients when the noisy coefficient $\mathcal{W}_{x_i}$ is under the threshold and to the pure noise $\mathcal{W}_{n_i}^2$ when $\mathcal{W}_{x_i}$ is over the threshold.

Wavelet Packet Decomposition and Rectangular approximation. The Wavelet Packed Decomposition (WPD) differs from Wavelet Decomposition (WD) since it decomposes approximation and detail coefficients, building a complete tree [10].

This difference allows to better locate the signal activities into time-frequency(scale) plan in a more complete way than WD and to monitorize with the same detail level all frequency bands. This feature is really important for both detection and analysis purposes, as discussed in next section. Consider a level j of WPD decomposition of the signal s. This level contains N coefficients distributed into 2^j basis and, as previously affirmed for WD, only few coefficients are significantly greater than zero. Obviously the signal energy is conserved in the WP domain so, $\mathcal{E}_s = \sum_{i=0}^{N} |s_i|^2 = \sum_{i=0}^{N} |\mathcal{W}_{s_i}|^2$.

Now we introduce an approximation assuming that the distribution of coefficients $\mathcal{W}_{s_i}$ is constant for a subset of indexes $\mathcal{N}_c \equiv \{i : \mathcal{W}_{s_i} \neq 0\}$ and $n_c = \mathrm{card}(\mathcal{N}_c) \leq N$

$$\mathcal{W}_{s_i} = \begin{cases} \pm\sqrt{\frac{\mathcal{E}_s}{n_c}}, & \text{if } i \in \mathcal{N}_c \\ 0, & \text{if } i \notin \mathcal{N}_c. \end{cases} \tag{8}$$

In other words, we approximate with a rectangular function the distribution of WP coefficients in a way to maintain the signal energy. Obviously this function works well when there are few coefficients to approximate in the WP domain. In following sections we assume almost one WP coefficient of signal $> T$ and we include the energy loss after thresholding with an efficiency factor $k = \frac{\mathcal{E}_s}{\mathcal{E}_s true} < 1$. This factor is strictly related with n_c and those parameters define the compression efficency of a WPD with a given basis.

3 Detection and analysis

The aim of this paper is to demonstrate the effective reliability of a detection and analysis system of EEG signals of patients in critical condition based on WPD. In this optics we describe the qualitative behavior of the risk. In the next paragraph we show the relation between error probability and compression and therefore suggest which kind of wavelet basis use for decomposition. If detection shows some evidence of life, it's important to define some analysis procedure to recover and reconstruct this embedded and unknown activity. We used the algorithm implemented by Donoho and Johnstone based on the needing of some visual feature (smoothness) of the reconstructed signal. Their works demonstrated the optimality, under certain condition, of the use of soft-thresholding rule when minimizing the risk $R(\boldsymbol{s}, T)$ in a minimax sense [6]. In this paper, since the noise considered is zero-mean, we define

$$\mathrm{SNR} = 10 \log \left(\frac{\mathrm{var}(\boldsymbol{s})}{\sigma^2} \right) \tag{9}$$

as measure of Signal to Noise power Ratio.

3.1 Detection

To evaluate the error probability of a system based on hard or soft thresholding, we consider two situations. In the first case, the patient is alive so his EEG signal can be modeled with the additive model (1). The probability of a null detection (i.e. to obtain a vector of zeroes $\underline{\mathbf{0}}$) is

$$P(e)^{+} = P(\hat{\boldsymbol{s}} = \underline{\mathbf{0}}) = P(\boldsymbol{W_x}\gamma_{\boldsymbol{T}}\ \boldsymbol{(\mathcal{W}_x) = \underline{0}}) = \prod_{i=0}^{N} P(|\mathcal{W}_{x_i}| < T), \qquad (10)$$

since we are considering orthogonal bases and uncorrelated noise. Using the same approximation described in (8), the equation (10) can be rewritten as

$$P(e)^{+} = \prod_{i \in \mathcal{N}_c} P\left(|\mathcal{W}_{n_i} + \sqrt{\frac{\mathcal{E}_s}{n_c}}| < T\right) \prod_{i \notin \mathcal{N}_c} P(|\mathcal{W}_{n_i}| < T)$$

$$= P\left(|\mathcal{W}_n + \sqrt{\frac{\mathcal{E}_s}{n_c}}| < T\right)^{n_c} \cdot P(|\mathcal{W}_n| < T)^{N-n_c} = p_{\mathcal{E}_s}^{n_c} . p_n^{N-n_c} \qquad (11)$$

In the second case, we consider an ES signal ($\boldsymbol{s} = \mathbf{0}$) so it can be considered as pure noise ($\mathcal{W}_{x_i} = \mathcal{W}_{n_i}$). The probability of detect something different from noise is

$$P(e)^{-} = P(\hat{\boldsymbol{s}} \neq \mathbf{0}) = 1 - \prod_{i=0}^{N} P(|\mathcal{W}_{n_i}| < T) \qquad (12)$$

$$= 1 - P(|\mathcal{W}_n| < T)^N = 1 - p_n^N$$

The whole error probability of our system can be written as

$$P(e) = P(e|\boldsymbol{s} \neq \mathbf{0}) + P(e|\boldsymbol{s} = \mathbf{0}) = P(e)^{+} + P(e)^{-} \qquad (13)$$

$$= 1 - p_n^N \left[1 - \left(\frac{p_{\mathcal{E}_s}}{p_n}\right)^{n_c}\right]$$

where $P(e)^{+}$represents the incidence of false negatives and $P(e)^{-}$ the incidence of false positives.

Straightforward considerations on (13) demonstrate the need of a high compression ($n_c \ll N$) to keep low the error probability. But this is the case our approximation works better. Our purpose is to obtain some reasonable values of loss (k) and compression (n_c) for a clinical acceptable error. We plot the equation (13) in function of $\frac{n_c}{N}\%$, considering $k = \frac{1}{2}$ and for some SNR values (fig.1). For low values of compression the incidence of $P(e)^{-}$ is predominant while, for very high value of compression ($\sim 97\%$), the contribution of false negatives is the most significant and limits the lower bound of the error.

For example, to obtain a $P(e) < 10^{-4}$ with SNR$= 0$, we need of a basis with a compression of $\frac{n_c}{N} \sim 97\%$ with a loss of 50%. We made a lot of tests with EEG signals and we found that Coiflet4 family [10] satisfies these constraints.

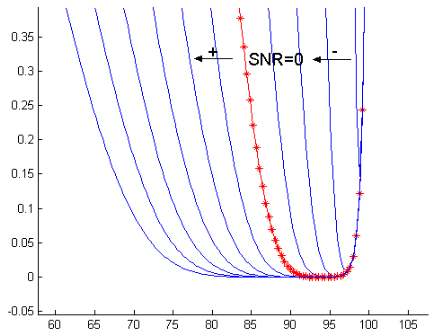

Fig. 1. Error probability for different SNR.

3.2 Analysis

The SURE algorithm is a method presented by Donoho and Johnstone [5] for the soft-thresholding procedure defined by

$$\chi(y_x) = \text{sign}(y_x) \begin{cases} |y_x| - T, \text{ if } |y_x| - T > 0 \\ \\ 0, \quad \text{ if } |y_x| - T \leq 0 \end{cases} \tag{14}$$

Besides the benefits of this strategy, largely described in [6], the weak differentiability of this thresholding function makes straightforward applicable the Stein method [9] to estimate the risk $R(s,T)$ in an unbiased fashion. The optimal threshold T will be chosen to minimize this approximation for each decomposition level. This "adaptative" method is really quasi-optimal in the "dense" situation, i.e. when the energy is distributed over many coefficients, but it has a serious drawback in situations of extreme sparsity or, in other words, when the signal compression is high. Donoho proposed a pre-testing procedure to In this case, exploiting that, in probability, if $\mathcal{W}n \sim N(0,\sigma^2), \|\mathcal{W}n\|_{l_n^\infty} \leq T_f = \sigma\sqrt{2\log(n)}$ [7], the threshold is set to T_f. As usual, noise variance is approximated with $\hat{\sigma}^2 = \left(\frac{\text{MAD}(\boldsymbol{\mathcal{W}}_x)}{0.6745}\right)^2$.

In Fig.2 we apply WP soft thresholding on a EEG signal measured from a patient in Brain Silence condition (Fig.2c, 2d) compared to the analogue application on a signal deriving from a normal patient (Fig.2a, 2b).

4 Future work

In this paper a preprocessing system for de-noising and computerized analysis of EEG signals in critical conditions has been presented. The basic algorithm is based on Wavelet Packed Decomposition and our inspection has

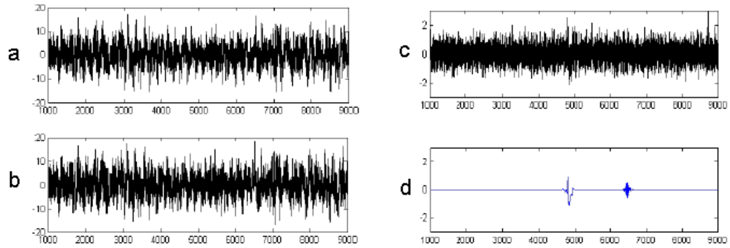

Fig. 2. Normal EEG thresholded and compared with ES EEG.

been dedicated to the orthogonal basis. This kind of approach was first developed by Donoho and Johnstone with their SURE algorithm based on a soft-thresholding strategy [5]. Starting from this miliary stone, we derive a really simple methods to relate the risk, and consequenlty the error probability of a detection system based upon these concepts, to the compression properties of the underlying set of bases chosen for denoising. We project to apply this algorithm to long EEG acquistion (setting a sample frequency of 128 Hz and an acquisition of 12 h, we have a set of ~ 5.5 million of samples for each EEG derivation). The aim is to build an automatic pre-analyzer to help the expert to only examine the really interesting parts.

References

1. J. Doyle (1995) The diagnosis of brain death: a checklist approach. The Online Journal of Anesthesiology, vol.2,(http://www.gasnet.org/esia/1995/march/).
2. F. Santoni (2001) Calibration algorithm for EEG Acquisition Systems. CRS4 Internal Report 02/26
3. I. Daubechies (1988) Orthonormal bases of compactly supported wavelets. Com. Pure and Appl. Math., vol. XLI, pp. 909-996.
4. H.Krim, D.Tucker, S. Mallat and D. Donoho (1999) On denoising and best signal representation. IEEE Trans. Inform. Theory, vol. 45, no.7, pp. 2225-2238
5. D.Donoho, I.Johnstone (1995) Adapting to unknown smoothness via wavelet shrinkage. Journal of the American Statistical Association, vol. 90, pp.1200-1224.
6. D.Donoho (1995) De-noising by soft-thresholding. IEEE Trans. Inform. Theory, vol. 41, pp.613-627.
7. S. I. Resnick (1987) Extreme values, regular variation and point processes. Appl. Probab. Series, New York:Springer-Verlag.
8. H.Y. Gao (1998) Wavelet Shrinkage denoising Using the Non-Negative Garrote. Journal of Computational and Graphical Statistics, vol.7, no.4, pp.469-488.
9. C. M. Stein (1981) Estimation of the mean of a multivariate normal distribution. Ann. Statist.,vol. 9, no. 6, pp. 1135-1151.
10. Coifman, R.R., Y. Meyer, M.V. Wickerhauser (1992) Wavelet analysis and signal processing. Wavelets and their applications, M.B. Ruskai et al. (Eds.), pp. 153-178, Jones and Bartlett.