Hysteresis thresholding for Wavelet denosing applied to P300 single-trial detection

Carolina Saavedra, Laurent Bougrain, Radu Ranta

7723

Hysteresis thresholding for Wavelet denoising applied to P300 single-trial detection

Carolina Saavedra, Laurent Bougrain, Radu Ranta

***Abstract*— Template-based analysis techniques are good candidates to robustly detect transient temporal graphic elements (e.g. event-related potential, k-complex, sleep spindles, vertex waves, spikes) in noisy and multi-sources electroencephalographic signals. More specifically, we present the significant impact on a large dataset of wavelet denoisings to detect evoked potentials in a single-trial P300 speller. We apply the classical thresholds selection rules algorithms and compare them with the hysteresis algorithm presented in [1] which combine the classical thresholds to detect blocks of significant wavelets coefficients based on the graph structure of the wavelet decomposition.**

I. Introduction

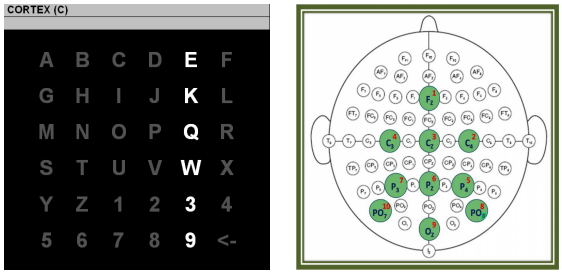
The observation of brain activity and its analysis with appropriate data analysis techniques allow to extract properties of underlying neural activity and to better understand high level functions. Several recording techniques exist providing different kinds of information at various scales. Some of them provide very local information such as multiunit activities (MUA) and local field potential (LFP) in one or several well-chosen cortical areas. Other ones provide global information about close regions such electrocorticography (ECoG) or the whole scalp such as electroencephalography (EEG).

In cognitive experiments using EEG, typically the experimental task (one trial) is repeated many times and the resulting brain activity is averaged over trials. The main reason for this averaging is the low signal-to-noise ratio (SNR) in the single-trials and average increases the SNR dramatically. The average activity allows to extract easily event-related components, which are strongly related to cognitive processes in the brain.

One major problem with transient events is how to be able to deal with the variability between trials. Thus, it is necessary to develop robust techniques based on stable features. Specific modeling techniques should be able to extract features investigating the time domain and the frequency domain. ERPs are short-time events with characteristic peaks at specific times. Nevertheless, it is useful to apply a denoising stage before the classification stage to detect the graphic-element.

II. P300 Speller: Single-trial detection

Brain-computer interface (BCI) system is a potentially powerful new communication and control option for those with severe motor disabilities [2]. A BCI system translates brain activity into commands for a computer or other devices (e.g. wheelchair, robotic arm). In other words, a BCI allows users to act on their real or virtual environment by using only brain activity. One of the well-known and powerful BCI system is the P300 speller [3] based on the non-invasive Electroencephalography (EEG) measuring from the subject's scalp [4]. This application has potentially a strong impact for patients with motor disabilities given its high rate of accuracy, reasonable speed and the fact that a long human training is not needed. Many improvements over the pioneering systems have been done and some comparisons exist [5]. This BCI system uses an oddball paradigm in which low-probability target items are inter-mixed with high-probability non-target items. The application highlights in a random order columns and rows (Fig. 1(a)). To spell one character it is necessary to flash 6 columns and 6 rows, the user pays attestation to the desired letter, when the letter is highlighted, a P300 is generated. Detecting this specific transient event, it is possible to know which letter is requested. The P300 component is a positive deflection waveform observed around 300ms after the onset of the stimulus. Almost everybody reacts on them including patients.



(a) (b)

Fig. 1. (a) A 6x6 P300 speller; (b) The 10 recorded EEG channels

The task of the P300 speller system is to recognize the ERP components from the noisy EEG background signal. It is found difficult to accomplish this target on the base of a single-trial because the magnitude of the EEG background activities is usually one-order larger than the one of the ERP components, that means the ERP components in single-trial recordings are almost covered by the background neural activities. Moreover, non-invasive electrodes produce a noisy signal because the skull dampens signals. Thus, ERP detection usually needs to average responses of repeated stimulations. Due to the averaging operation, the background EEG activities are reduced and the ERP components are enhanced and evident. From a practical point of view, an important issue is to reduce the number of repetitions, in order to obtain high communication bit-rates. The methodology has been improved but a gap still exists to enable single-trial recognition. To validate our study, we used a database

obtained from first-time users of the P300 speller application implemented within the BCI2000 platform [4].

III. Database

10 healthy subjects with similar characteristics (sleep duration, drugs, age, etc.) recorded by the Neuroimaging Laboratory of Universidad Autnoma Metropolitana (Mexico) were used. 10 channels (Fz, C3, Cz, C4, P3, Pz, P4, PO7, PO8, Oz) have been recorded at 256 sps using the g.tec gUSBamp EEG amplifier, a right ear reference and a right mastoid ground (Fig. 1(b)). An 8th order bandpass filter, 0.1-60 Hz and a 60 Hz Notch have been used. The stimulus is highlighted for 62.5 ms with an inter-stimuli interval of 125 ms.

A complete description of the parameters used for the speller and the data are available in BCI2000 and Matlab formats on the database website: http://akimpech.izt.uam.mx/p300db.

To spell one character it is necessary to flash 6 columns and 6 rows. To write properly a character it is necessary to detect the only column/row over 6 that contains a P300. Thus we can see the task as a 6 classes problem rather than a 2 classes problem. The main reason is that if we want to detect if there is a P300 response or not, we have to do it for each column/row. It is possible that the classifier wrongly detect as a P300 more than one response. At the opposite, it is also possible that the classifier wrongly detect none P300. Even using additional information such as the higher posterior probability or the higher distance from the border to select the only P300 over the 6 responses, the classifier has more difficulty. So we use a multiclassifier (see section V). Doing it, it is necessary to check if all classes are balanced for training, so the numbers of samples per class are computed. Because classes are unbalanced, it is necessary to create new data for the classes with less samples. To do this, we permute the P300 wave in the row/column. After this process a matrix is of size 1080×15360 is obtained to train the classifier.

IV. Methods

Recent literature presents solutions for denoising ERP using specific wavelets [6],[7], [8]. We focused our attention on order 3 coiflets [9]. We want to test it on our data set and compare its performance with support vector machines, without denoising. We are especially interested by the single-trial detection which is the worst case because the response is very noisy if no averaging is performed.

A. Denoising by wavelet decomposition

The Wavelets Transform (WT) was developed as an alternative to the Short Fourier Transform to overcome the resolution problem [10]. WT is a windowing technique with variable regions size. The main idea is to represent a signal $x(t)$ in terms of displaced and shifted version of a mother wavelet $\Psi(t)$.

$$\Psi_{a,b}(t) = |a|^{-\frac{1}{2}} \Psi\left(\frac{t-b}{a}\right) \tag{1}$$

where a and b are the scale and translation parameters respectively.

The signal coefficients are obtained by the convolution of the original signal and the different version of the mother wavelet.

$$W_\Psi X(a,b) = \langle x(t) | \Psi_{a,b}(t) \rangle \tag{2}$$

The coefficients refer to the similarity between the signal and the wavelet at the current scale and time position.

B. Discrete Wavelet Decomposition

In wavelet analysis, we often use the discrete form of equations (1) and (2). More precisely, the discrete orthogonal wavelet decomposition is obtained using a discretized scale $a_j = 2^j$ (dyadic step), while the time shift $b = p_j$ is obtained such as, on a given scale $j+1$, there are twice less coefficients than on the previous scale j. The wavelet coefficients, obtained by convolution, as in (2), will be noted as w^{j,p_j}.

When using a finite depth decomposition, the dilatations (thus the scales) are limited, *i.e.*, $j = 1 \dots M$. In this case, wavelets are not sufficient to represent an arbitrary signal containing low frequencies. Therefore, father wavelets (also known as scaling functions) are introduced in the decomposition: as the mother wavelets, they have unit norm, but they do not have zero mean. For a given family of scaled and translated mother wavelets, we have a family of (translated) scaling functions $\phi^{M,p}$ and the corresponding coefficients w^{M,p_M} (see [11] for full details and fast algorithm implementation).

Using this notations, a noisy discretized signal $\mathbf{z} = \mathbf{x} + \mathbf{n}$, where $\mathbf{z}$ is the given discrete–time recorded signal, $\mathbf{x}$ is the noise–free unknown version of $\mathbf{z}$ and $\mathbf{n}$ the noise. The orthogonal wavelet decomposition of $\mathbf{z}$ can be written as:

$$\mathbf{z} = \sum_{p_j,j} w_z^{j,p_j}\ \psi^{j,p_j} + \sum_{p}^{M} w_z^{M,p_M}\ \phi^{M,p_M}, \tag{3}$$

where j is the scale, p the position, ψ the wavelet, ϕ the scaling function and M the analysis depth [11].

In wavelet analysis, we often talk about approximations and detail coefficients. The approximations coefficients correspond to the high-scale, low-frequency components of the signal. These coefficients contain the most important part of the signal, the part with the "essence" of the signal. The detail coefficients are the low-scale, high-frequency components, i.e. all the small features who made the signal "unique".

C. Wavelet denoising

The fundamental hypothesis of wavelet denoising is that wavelets are correlated with the informative signal and not correlated with the noise, which globally means that large coefficients correspond to signal and small coefficients correspond to noise.

Therefore, noise cancelling can be performed by thresholding: only large coefficients will then be used to reconstruct

the informative signal (see Antoniadis *et al.* [12], [13] for a review of denoising methods).

1) Classical thresholds: With the notation from (3), an estimate of the denoised signal $\hat{\mathbf{x}}$ can be obtained by inverse wavelet transform of the coefficients vector $\hat{\mathbf{w}}_x$. These coefficients are obtained by modifying $\mathbf{w}_z = [w_z^{M,p} w_z^{j,p}]$ (generally, the approximation coefficients $w_z^{M,p}$ remain unchanged):

$$\hat{w}_x = g(w_z) w_z,$$

with $g(w_z)$ a shrinkage function applied on the measured signal coefficients. As the noise-free coefficient vector is presumably sparse, the simplest solution is provided by the well-known hard shrinkage (thresholding):

$$3MBg(w_z) = \max\left(0, \text{sign}(|w_z| - T)\right), \tag{4}$$

where T is a threshold value. The most widely used is the universal threshold T_U proposed by Donoho and Johnstone in their algorithm $VisuShrink$ (designated $Visu$ in the sequel) [14]. $Visu$ is used to achieve *complete asymptotic elimination of the normal Gaussian noise* and it can be shown (using extreme values statistics) that this is achieved by setting $T_U = \sqrt{2 \log N}$. This method is very appealing because of its simplicity and its visually attractive results, however, its focus on eliminating all noise often leads to less precise reconstruction of the signal of interest and that represents a major drawback.

A different approach is proposed by the $Sure$ algorithm [15], that aims to *estimate as precisely as possible the "clean" signal* by minimizing an estimate of the mean squared error (MSE) between the denoised signal and the original one, known as the *Stein Unbiased Risk Estimator*. This risk is minimized by exhaustively searching the optimal threshold T_S among the coefficients w_z. The algorithm was developed under specific conditions on the function g (weak differentiability). In particular, the most widely used is the soft thresholding function:

$$g(w_z) = \max\left(0, \frac{|w_z| - T}{|w_z|}\right) \tag{5}$$

The obtained threshold T_S (or thresholds, as the method is usually implemented by scale) are lower than the T_U and the obtained signal has a noisier appearance.

Another classical threshold T_M is the minimax one which is used as a fixed threshold chosen to yield minimax performance for mean square error against an ideal procedure. The minimax principle is used in statistics in order to design estimators. Since the de-noised signal can be assimilated to the estimator of the unknown regression function, the minimax estimator is the one that realizes the minimum of the maximum mean square error obtained for the worst function in a given set. As for the T_S the obtained value will be lower than the universal threshold. Consequently, we preserve more information but the denoised signal has a noisier appearance.

2) Hysteresis threshold: We have proposed in [1] the following heuristic combination: use a $Visu$ computed high threshold $T_h = T_U$ to detect *blocks* of significant wavelet coefficients, and a $Sure$ or $Minimax$ low threshold $T_l = T_S(T_M)$ to fix the limits of the selected blocks. The method, based on the graph structure of the wavelet decomposition, is detailed in [1]. The main idea is the following: using in the graph structure, the selected neighborhood of a "very large" coefficient $|w_z| > T_U$ will be formed by the connected "large enough" coefficients selected by $Sure$ ($|w_z| > T_S$). The obtained two thresholds algorithm was called hysteresis denoising. The resulting blocks, without *a priori* predefined shapes, will naturally integrate two important properties of real signals, as defined by [16]: **persistence** and **clustering**. Persistence implies that "large/small values of wavelet coefficients tend to propagate across scales", which means that the binary trees of wavelet coefficients (Fig. 2(a)) tend to contain similar amplitude coefficients. Clustering, defined as the fact that "if a particular wavelet coefficients is large/small, then adjacent coefficients are very likely to also be large/small", considers that connected neighboring coefficients belonging to the same scale should be selected together if one of them is superior to the threshold (Fig. 2(b)). If both persistence and clustering are considered, all the coefficients are linked together in a complete graph (Fig. 2(c)).

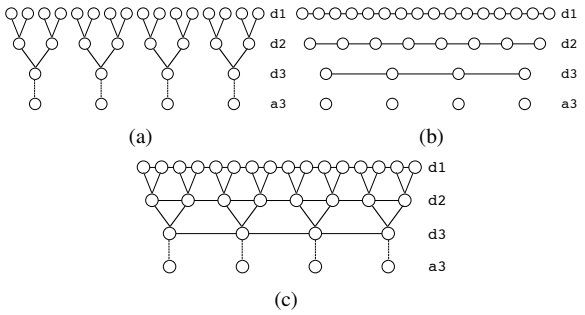


Fig. 2. Graph types: (a) tree; (b) scale; (c) complete.

As in real applications the noise is neither white nor scaled to unitary variance, the thresholding must be adapted by multiplication by the noise standard deviation, estimated separately for each level of the wavelet decomposition.

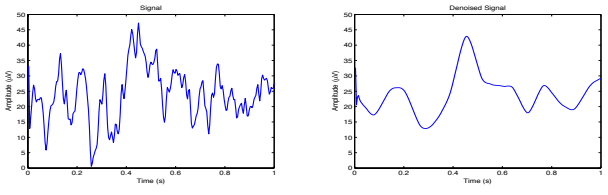


Fig. 3. (a) P300 response on CZ ; (b) Denoised signal using order 3 coiflets at level 4 with the complete hysteresis thresholding.

D. Classifier

A copy-spelling run with 16 letters and 15 repetitions were used to train a multi-class support vector machine (SVM) per

trial and repetition, to recognize rows and columns containig P300 events. A linear kernel function is used becuase it is more robust than polynomial one for very noisy data with high variability. The penalty parameter C that controls the trade off between errors on training data and margin maximization is equal to 1.

Multi-class SVM is based on several binary classifiers, the method "1-against-1" was used to solve the multi-class SVM problem. We used libsvm, a library for Support Vector Machines by Chih-Chung Chang and Chih-Jen Lin and its simple interface for matlab [17]. We test the SVM model with a free-spelling run (user-selected words) of 3 words that correspond approximately to 15 characters per subject.

V. Results

In this section, we compare the letter accuracy on the P300-speller using SVMs as classifiers with : i) no denoising ii) classical thresholds and iii) the hysteresis approaches developed in [1].

We remain here that order 3 coiflets were chosen due to their good performances compare to B-Splines on this application [18]. We present in details the results we obtained for a fourth-level decomposition (corresponding to the fourth-level approximation and the first four levels of detail). The results for the other levels of decomposition have been computed but the ranking of the compared methods is the same for all levels and the best performances are obtained for the fourth level of decomposition.

Fig. 4 shows that all wavelet denoisings improve significantly the letter accuracy according to a confidence interval of 95% (the probability to find by chance the correct letter is 1 over 36). The choice of the hysteresis option seems to not have a strong impact on the results.

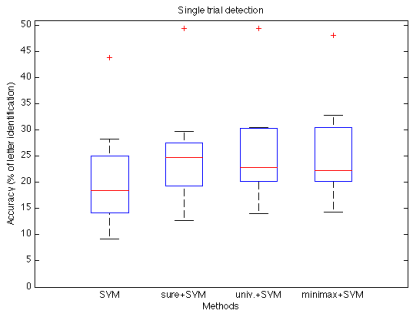
Table I shows the average letter accuracy for all subjects using classical (T_u, T_s, T_m) and hysteresis (T_{ht}, T_{hs}, T_{hc}) thresholdings. We can see that the minimax threshold T_m has a better performance compare to the other classical thresholds. The hysteresis thresholds improve the classical ones. More specifically, the complete T_{hc} approach has the best performance. If we compare the best performances of hysteresis and classical families, we can see that it is possible to improve the relative performance of the 1.49% that represents 43 letters more.

VI. Conclusion

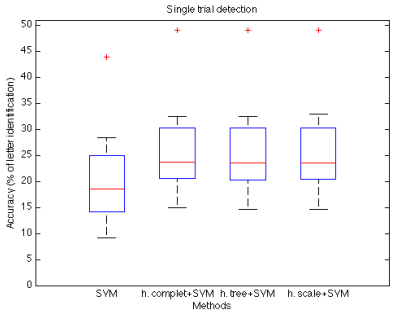
We show the improvement of a wavelet denoising on P300 single trial detection. More specifically, we present the improvement of the hysteresis thresholdings compare to classical ones.

In this paper, we present the impact to apply wavelet denoising as a pre-processing stage to recognize evoked potential, from an EEG background, in a single-trial P300 speller. We compare the classical algorithms to detect thresholds based on noise estimation from a signal and the hysteresis aproach presented in section IV-C, wich combines classical thresholds to get better results. The order 3 Coiflets wavelets was used in the experiments according to there performance for ERP denoising in literature.

The improvement of wavelet denoising is visible in terms of letter accuracy. Algorithms based on hysteresis improves performance of classical algorithms due to combination of them. The best result was obtain by hysteresis complete, who found 10 letters more than the second best one.



(a)



(b)

Fig. 4. Letter accuracy Box-plots using a) classical thresholds b) hysteresis minimax thresholds for Wavelet denoising for P300 single-trial detection. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

Subj	NO	T_S	T_U	T_M	T_{hs}	T_{ht}	T_{hc}	10Hz
APM	62.35	59.61	64.31	**65.88**	61.57	62.75	61.57	**65.88**
ASG	52.50	57.92	62.50	61.67	63.33	**64.58**	**64.58**	63.75
DCM	57.22	63.61	**70.83**	67.78	69.17	69.17	68.06	70.00
ELC	56.30	62.96	71.85	67.78	70.74	70.74	71.85	**72.22**
LAC	75.24	80.95	**82.38**	80.00	**82.38**	80.48	80.48	81.90
LAG	48.41	50.43	**55.94**	55.36	55.36	53.62	53.04	55.65
GCE	33.81	38.10	37.62	38.57	**39.52**	37.62	36.67	37.62
WFG	43.89	44.44	44.44	43.61	43.89	44.17	44.17	**45.56**
LGP	61.33	63.33	**63.67**	**63.67**	**63.67**	62.67	63.00	62.67
ACS	40.00	39.52	**46.19**	45.71	45.24	44.29	44.76	45.24
DMA	30.00	36.11	35.56	36.67	35.56	**37.78**	36.37	35.56
JLP	33.70	34.44	**40.37**	34.81	39.26	35.19	34.81	36.67
JSC	25.24	34.76	40.95	38.57	41.43	40.48	40.48	**41.90**
PGA	25.61	31.23	37.54	37.19	37.89	37.19	37.54	**38.25**
JST	71.79	71.79	72.82	72.82	**73.33**	72.82	72.82	71.79
ASR	6.03	22.84	**35.87**	29.84	33.65	33.97	33.97	34.92
mean	45.21	49.50	**53.93**	52.50	52.52	52.97	51.78	53.72

TABLE I

Accuracy of all subjects using classical and hysteresis thresholdings.

As a future work we want to investigate, differents thresholds selection rules to improve the hysteresis results and how to estimate noise without relying on a the assumption of noise is white $N(0,1)$.

REFERENCES

[1] R. Ranta and V. Louis-Dorr, "Hysteresis thresholding: A graph-based wavelet block denoising algorithm," *The Open Signal Processing Journal*, vol. 3, pp. 6–12, 2010, http://benthamscience.com/open/tosigpj/.

[2] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, and T. Vaughan, "Brain-computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[3] L. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.

[4] G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, 2004.

[5] D. Krusienski, E. Sellers, F. Cabestaing, S. Bayoudh, D. McFarland, T. Vaughan, and J. Wolpaw, "A comparison of classification techniques for the P300 Speller," *Journal of neural engineering*, vol. 3, p. 299, 2006.

[6] R. Quiroga and H. Garcia, "Single-trial event-related potentials with wavelet denoising," *Clinical Neurophysiology*, vol. 114, no. 2, pp. 376–390, 2003.

[7] Y. Yong, N. Hurley, and G. Silvestre, "Single-trial EEG classification for brain-computer interface using wavelet decomposition," in *European Signal Processing Conference, EUSIPCO 2005*. Citeseer, 2005.

[8] C. Gupta, Y. Khan, R. Palaniappan, and F. Sepulveda, "Wavelet Framework for Improved Target Detection in Oddball Paradigms Using P300 and Gamma Band Analysis."

[9] M. Unser, "On the Asymptotic Convergence of E-Spline Wavelets to Gabor Functions," *IEEE Transactions on Information Theory*, vol. 38, no. 2, 1992.

[10] S. Mallat, *A wavelet tour of signal processing*. Academic Pr, 1999.

[11] ——, *A wavelet tour of signal processing*. Academic Press, 1999.

[12] A. Antoniadis, J. Bigot, and T. Sapatinas, "Wavelet estimators in nonparametric regression: a comparative simulation study," *Journal of Statistical Software*, vol. 6, no. 6, pp. 1–83, 2001, http://www-lmc.imag.fr/lmc-sms/Anestis.Antoniadis/HTTP/publis-anto.html.

[13] A. Antoniadis, "Wavelet methods in statistics: Some recent developements and their applications," *Statistics Surveys*, vol. 1, pp. 16–55, 2007.

[14] D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[15] ——, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, pp. 1200–1224, 1995.

[16] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," vol. 46, no. 4, pp. 886–902, 1998.

[17] C. Chang and C. Lin, "LIBSVM: a library for support vector machines," 2001.

[18] C. Saavedra and L. Bougrain, "Wavelet denoising for P300 single-trial detection," in *Proccedings of the 5th french conference on computational neuroscience - Neurocomp'10*, Neurocomp, Ed., Lyon France, 10 2010, pp. 227–231. [Online]. Available: http://hal.inria.fr/inria-00549218/en/