Local Discriminant Wavelet Packet Basis for Signal Classification in Brain Computer Interface

Victoria Peterson[1,2], Rubén Acevedo[3], Hugo Leonardo Rufiner[1,2,3] and Rubén Spies[4,5]

[1] Centro de Investigación de Señales, Sistemas e Inteligencia Computacional (Sinc(i)), Facultad de Ingeniería y Ciencias Hídricas Universidad Nacional del Litoral (FICH-UNL), Santa Fe, Argentina
[2] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Santa Fe, Argentina
[3] Facultad de Ingeniería, Universidad Nacional de Entre Ríos (FI-UNER), Entre Ríos, Argentina
[4] Instituto de Matemática Aplicada del Litoral, (IMAL), Santa Fe, Argentina
[5] Facultad de Ingeniería Química, Universidad Nacional del Litoral (FIQ-UNL), Santa Fe, Argentina

sinc(i) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
V. Peterson, R. C. Acevedo, H. L. Rufiner & R. Spies; "Local Discriminant Wavelet Packet Basis for Signal Classification in Brain Computer Interface"
Anales del VI Congreso Latinoamericano de Ingeniería Biomédica (CLAIB 2014), pp. 279, oct, 2014.

Abstract— **A Brain-Computer Interface (BCI) is a system that provides direct communication between the brain of a person and the outside world. In the present work we use a BCI based on Event Related Potentials (ERP). The aim of this paper is to efficiently solve the classification problem consisting on labeling electroencephalogram records as target (with ERP) or non-target records (without ERP). We evaluate the performance of a BCI by using the Wavelet Packet Transform with the Local Discriminant Basis (LDB) method to find an orthogonal basis that maximizes the difference between the two classes involved. The performance of the LDB patterns and the temporal data (without post-processing) are analyzed with the Fisher Linear Classifier. It is shown that the bets results are obtained with LDB patterns calculated by Daubechies 4 as filter, Sum of Squares as discriminant function and the first 18 more discriminant basis vectors.**

Keywords— **Wavelet Packet Transform, Local Discriminant Basis, Event Related Potencial, Brain Computer Interface.**

I. Introduction

If a person has a neurological desease which has disrupted or altered the neuromuscular channel through which the brain communicates and controls the person's environment, then a Brain-Computer Interface (BCI) can significantly improve the individual's quality of life. By using only brain activity and without needing any normal channel or peripheral nerves and muscles, BCI provides a person with a new way of communication with the outside world [1],[2].

It has been demonstrated [3],[4],[5], that when a subject is asked to decide to which of 2 possible categories an item belongs, and one of the 2 categories is rare and infrequent, these "rare" items will evoke in the electroencephalogram (EEG) over the parietal cortex an Event Related Potential (ERP). One of the main components of such ERP is an enhanced positive-going component with a latency of about 300 ms (called P300 wave). This is called the "oddball" paradigm [6], [7].

Although a perfect classification rate can be achieved by averaging 40 EEG records, doing so is useless for communication purposes since the transfer speed is low (about 60s per item) [7]. On the other hand, single-trial ERP detection methods would allow high communication speed, and therefore, they are strongly desirable. Traditional analysis methods are not suitable for single-trial detection since the SNR is low (around $-5dB$) and the EEG signal is non-stationary. Thus, methods able to highlight the ERP signal from the background EEG are needed. In this work, like in others (e.g. [8],[9]), the Wavelet Packet Transform (WPT) was used due to it's ability to explore the time-frequency information of ERP signals. We aim to improve the feature extraction stage for single-trial EEG signals by increasing the separability between the two classes, which in terms ought to improve the classification rate of the temporal data. We propose to evaluate BCI performance by using WPT to obtain a dictionary of orthogonal basis functions. Local Discriminant Basis (LDB) is then used to get the most discriminant orthogonal base for representation and classification of ERP signals.

II. Materials and methods

A. Database

An Open-Access P300 speller database was used [10]. EEG records from 18 subjects were acquired by 10 electrodes in the positions shown in Fig. 1 with reference in the right ear (A2) and ground in right mastoid (M2). The records were digitalized at a rate of 256 Hz . The Farwell and Donchin oddball paradigm was used [2]. A 6-by-6 matrix containing letters and numbers are displayed on a computer screen. During the experiment, a subject is asked to spell different words. The person must focus on one character at the time. As stimulus, a row or a column of the matrix is randomly flashed. In each stimulating block, every row and column of the matrix is intensified only once. If the person is well concentrated, when the chosen character is illuminated, a relevant event occurs, i.e. an ERP signal is elicited [7]. Since there are 2 classes in-

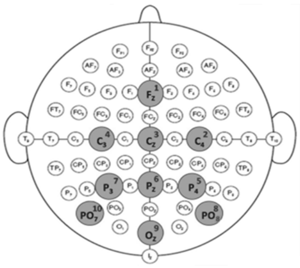

Fig. 1: Positions of the electrodes used in acquisition of EEG recording. Modified from [10].

volved, the 6-by-6 matrix results in 12 possible events (one per each row and one per each column), of which only 2 are relevant (with ERP) in every intensification block.

Each subject participated in 4 sessions. The first session was a copy-spelling run, where the individuals had to spell the words CALOR, CARINO & SUSHI. The second one was also a copy-spelling run, but only the word SUSHI had to be spelled. The other two session were free-spelling runs, i.e., the person chose the word he/she wanted to spell. In this work we used the first and second sessions as training and testing sets, respectively. Each session consisted on 15 trials, resulting in 2880 EEG epochs per channel (480 target and 2400 non-target) for the training set and 900 EEG epochs per channel (150 of them being target) for the testing set.

The pre-processing stage was done in the following steps:

1. Application of a 15^{th} order low-pass FIR filter with a cut-off frequency of $15Hz$ for each channel and for each subject [11].
2. Extraction of unique epoch of one second duration (256 samples) at the beginning of the intensification.
3. Detection and resetting of outliers beyond 6 standard deviations (i.e. any value greater than 6 was reset to 6).
4. Normalization with z-score per subject and per channel. The testing set was normalized with the mean and the standard deviation of the training set.
5. Downsampling at $64Hz$.
6. Channel concatenation. A $N \times 640$ matrix per subject was conformed, where N was the number of trials and $640 = 10\,channels \times 64\,samples$.

B. Wavelet Packet Decomposition

WPT is a generalization of multiresolution analysis. Given a vector $\Omega_{0,0} \in \mathbb{R}^n$ with $n = 2^{n_0}$, by the convolution subsampling operators constructed by a mother wavelet, the vector $\Omega_{0,0}$ is transformed into two subsequences $\Omega_{1,0}$ $\Omega_{1,1}$ of length $n/2$. In other words, the signal is partitioned into a high frequency portion called *Detail* and a low frequency portion called *Approximation*. Next, the same operators are applied to these 2 subsequences, so obtaining another two subfrequency bands of length $n/4$. This process is repeated J times. At the end, a binary tree is obtained where each node of the tree represents subspaces with different frequency localization. Each node has an associated subspace, which admits an orthogonal basis. Such basis can be obtained by going down the tree [12],[13].

There are more than $2^{2^{J-1}}$ possible orthogonal wavelet packet basis for representing our signal. In this work we used LDB method which maximize the difference between the classes of the problem. The algorithm measures the power of discrimination of each subspace in the binary tree by applying a given information cost and comparing the goodness of each node (subspace) for the classification problem. As in the best-basis search algorithm, this is done by evaluating if the union of two children nodes must be kept or not [14]. Let $\mathbf{p} = (p_i)_{i=1}^n$ and $\mathbf{q} = (q_i)_{i=1}^n$ be two non-negative sequences with $\sum p_i = \sum q_i = 1$, representing the normalized energy distributions of 2 generic signals belonging to class 1 and class 2 of the classification problem, respectively. A discriminant information function $D(\mathbf{p},\mathbf{q})$ measures "how differently" $\mathbf{p}$ and $\mathbf{q}$ are distributed. Options for D are:

- I-divergence or Antisymmetric Relative Entropy:

$$I(\mathbf{p},\mathbf{q}) \doteq \sum_{i=1}^{n} p_i \log_2 \frac{p_i}{q_i}, \tag{1}$$

where $\log_2(0) \doteq -\infty$, $\log_2(x/0) \doteq +\infty$ for $x > 0$. Note that I is not a metric since it is not symmetric.
- J-divergence or Symmetric Relative Entropy:

$$J(\mathbf{p},\mathbf{q}) \doteq I(\mathbf{p},\mathbf{q}) + I(\mathbf{q},\mathbf{p}). \tag{2}$$

- Sum of Squares:

$$W(\mathbf{p},\mathbf{q}) \doteq \|\mathbf{p}-\mathbf{q}\|^2 = \sum_{i=1}^{n} (p_i - q_i)^2. \tag{3}$$

Once the LDB is computed the projection coefficients can be used for classification purposed. However, if we want to reduce the dimensionality of the problem, then only the coefficients associated to the first few most discriminant vectors can be used as input of the classifier.

C. Classifier

We assumed that the classes involved in the problem are linearly separable, so a Linear Discriminant Analysis (LDA), more precisely the Linear Fisher method, was used. LDA assigns to a new input $\mathbf{x} \in \mathbb{R}^N$ one of two possible class labels, according to the sign of the function $\mathbf{w}^T\mathbf{x} + b$, where $\mathbf{w}$ and b are obtained after training the classifier with known examples corresponding to the training set [15]. It is well known [16] that LDA, can be a powerful tool for the classification of

ERP signals. The main advantages of LDA are: it is easy to implement, it's required $O(n^3)$ calculation and it delivers similar classification results as other more complex classification methods. It is therefore no surprise that LDA is one of the most commonly used classifiers of BCI systems, which justifies our choice of it. The performance of the classifier was based upon three indices: accuracy, sensitivity and specificity.

D. *Implementation and Experiments*

The classification results of the temporal patterns were used to compare the results achieved by LDB patterns. Before training the classifier, we balanced both the training and the testing sets of the temporal data. In other words, each patterns had the same amount of epochs with and without P300, where the latter ones where randomly selected.

WPT with decomposition level $J = 6$ was applied to the training balanced set per channel an per subject together with Daubechies 4 as mother wavelet. The choice of $J = 6$ correspond to the maximum decomposition level for a 64 length signal. On the other hand, the choice of Daubechies 4 was made because of it is widely used in EEG processing. This wavelet also has morphological similarities with the P300 component of ERP signals, reason from which it is commonly used for ERP detection [17].

We constructed LDB patterns for each discriminant measure, since LDB can be implemented with different cost functions. In order to analyze the information cost and optimize the number of discriminant vectors, LDB patterns were made to vary in increments of 2, from 2 to 64 features per channel.

The training set was used to construct the LDB basis. Both the training and the testing coefficients were computed as the associated projection of the EEG signals onto this orthogonal basis. This coefficients were then used as inputs to LDA.

The whole LDB procedure (including dataset generation) was repeated 10 times.

III. Results

For the present problem at hand it is strongly desirable to minimize the number of false positive labels. Hence, we focused our analysis on the specificity index in order to compare the goodness of classification and to select the best cost function. The third column in Table 1 shows the maximum median for the different amount of features analyzed, corresponding to the specificity index. Columns 2 and 4 show percentiles 25^{th} and 75^{th}, respectively. Finally, the last column contains the number of features needed in each case. Similarly, the Sum of Squares (3) resulted in the best performance for the other 2 indices. Fig. 2 depicts the variation of all three indices with respect to the number of features per channel using Sum of Squares.

Table 1: Best Median, P25 and P75 percentiles for the three discriminant measure corresponding to specificity index.

Measure	P25	Median	P75	#Feature
Antisymmetric	0.6970	0.7498	0.8200	22
Symmetric	0.6899	0.7576	0.8210	14
Sum of Squares	0.7355	0.7698	0.8221	18

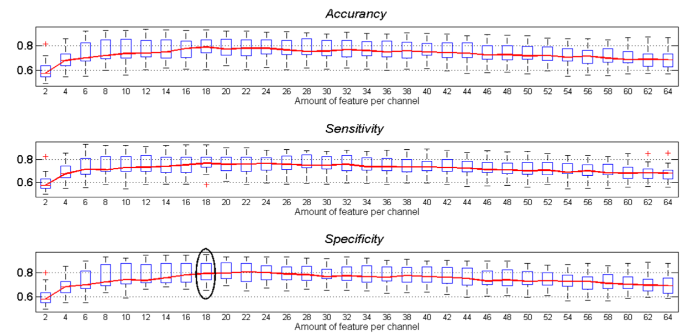

Fig. 2: Variation in the classification indices as function of the number of features per channel.

It is appropriate to mention here that for just 18 characteristics per channel, all 3 indices were between 0.7898 and 0.7925, indicating a classification success of about 78%. These results are very encouraging for an efficient online BCI system since accuracy above 70% allows communication and device control [18].

Since the main objective is to improve the performance of the whole BCI system, we compared the best LDB result with temporal data classification results. This is shown in Fig. 3.

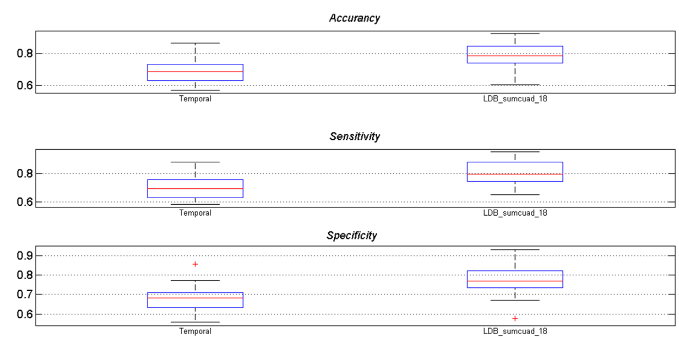

Fig. 3: Performance classification comparison between temporal data and LDB with the Sum of Squares measure with 18 characteristics per channel.

The cpu times of the classification stage were also compared. Classification with LDB with 18 features per channels was around 2000 times faster than classification with temporal data. Finally, a bit rate comparison was performed [19]. While temporal data achieved 46.51 bits/min our method achieved 115.2 bits/min, which is slightly lower than those obtained by state of the art method [18].

We have also evaluated the statistical significance of these results by computing the probability that a given experiment be better than the temporal classification patterns. In order to perform this test, we assumed statistical independence of the classification errors for each epoch and we approximated the binomial distribution for the errors by a gaussian distribution. Both assumptions are realistic: independence is obvious while the normal approximation is reasonable because of the high number of testing epochs used (51000) in classification. Our results showed that LDB Sum of Squares with 18 characteristics per channel is significantly better than temporal data with probability higher than 99.99%.

IV. Discussion

Although it is clear from Table 1 that none of the three measures presents a significant difference with respect to the others, the Sum of Squares was chosen as the best discriminant measure due to its highest specificity. In Fig. 2 one sees that the classification index increases, it reaches a maximum at about 20 characteristics per channel, and then its starts to decrease.

A comparison of the best LDB results with temporal patterns showed that the use of LDB Sum of Squares with just 18 characteristics per channel is significantly better than the classification rates obtained by temporal patterns. Moreover, the processing time of the classification stage was greatly reduced and the communication speed was significantly increased.

V. Conclusion

In this work the use of WPT together with LDB algorithm for increasing the classification rate in BCI systems was analyzed. The best result was obtained with LDB patterns constructed by Daubechies 4, decomposition level $J = 6$, Sum of Squares as discriminant measure and the first 18 basis vectors. This LDB pattern reached overall accuracy of 78.67% against a 68.67% obtained with the temporal patterns and it provided 40% increase in communication effectiveness. Also, the processing time of the classification stage was significantly trimmed down as consequence of the fact that the dimension of the original signal was reduced by more than 70%. Finally, as a general conclusion we point out that the simultaneous use of WPT and LDB constitutes a very good tool to explore and extract the most relevant features of the P300 wave embedded in the EEG signal, increasing the classification performance and improving the whole BCI system.

VI. Acknowledgements

The authors would like to thank Prof. Naoki Saito for Matlab's mldb7 toolbox used to run the LDB algorithm.

References

1. Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G, Vaugham T M. Brain computer interfaces for communication and control *Clinical Neurophysiology.* 2002;113:767–791.
2. Farwell L A, Donchin E. Talking off the top of your head: toward a metal prosthesis utilizing event-related brain potentials *Electroencephalography and clinical neurophysiology.* 1988;70:510–523.
3. Donchin E, Karis D, Bashore TR, Coles M G H, Gratton G. Cognitive psychophysiology and human information processing *M.G.H. Coles, E. Donchin and S.W. Porges (Eds.), Psychophysiology: Systems, Processes, and Applications. Guilford Press.* 1986a:244–267.
4. Hillyard S A, Kutas M. Electrophysiology of cognitive processing *Annu. Rev. Psychol.* 1983;34:33–61.
5. Pritchard W S. Psychophysiology of P300 *Psychol. Bull..* 1981;89:506–540.
6. Hiyard S A, Kutas M. Electrophysiology of congnitive processing *Annual Reviews Phychol..* 1983;34:33–61.
7. Donchin E, Spencer K M, Wijesinghe R. The mental prosthesis: Assessing the Speed of a P300-Based Brain-Computer Interface *Transactions on Rehabilitation Engineering.* 2000;8:174–179.
8. Peterson V, Atum Y, Jauregui F, Gareis I, Acevedo R, Rufiner L. Detección de potenciales evocados relacionados a eventos en interfaces cerebro-computadora mediante transformada wavelet *Revista Ingeniería Biomédica.* 2013;7:51–57.
9. Raz J, Dickerson L, Turetsky B. A Wavelet Packet Model of Evoked Potentials *Brain and Language.* 1999;66:61–88.
10. Ledesma-Ramirez C, Bojorges-Valdez E, Yaez-Suarez O, Saavedra C, Boygrain L, Gentiletti G. An Open-Access P300 Speller Database in *Fourth international BCI meeting, Monterrey, USA, California* 2010.
11. Bougrain L, Saavedra C, Ranta R. Finally, what is the best filter for P300 detection *Tools for Brain-Computer Interaction (TOBI) Workshop III.* 2012;1.
12. Rufiner H L. *Aná lisis de la señ al de voz mediante diccionarios discretos.* Ediciones UNL, Coleccin Ciencia y Tcnica 2009.
13. Mallat S. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way.* Elsevier Inc. 2009.
14. Saito N, Coifman R R. Local Discriminant Bases and Their Applications *Journal of Mathematical Imaging and Vision.* 1995;5:337–358.
15. Fisher R A. The use of multiple measurements in taxonomic problems *Annals of Eugenics.* 2002;7:179–188.
16. Blankertz Benjamin, Lemm Steven, Treder Matthias, Hauf Stefan, Müler Klaus Robert. Single-trial analysis and classification of ERP component- A tutorial *Neuroimage.* 2011;56:814–825.
17. Samar V J. Wavelet Analysis of Neuroelectric Waveforms: A Conceptual Tutorial *Brain and Language.* 1999;66:7–60.
18. Li K, Raju V N, Sankar R, Arbel Y, Donchin E. Advances and Challenges in Signal Analysis for Single Trial P300-BCI in *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems*;2(ll):87–94 2011.
19. Wolpaw J R, Ramoser H, McFarland D J, Pfurtscheller G. EEG-based communication: Improved accuracy by response verification *IEEE Trans. Neural Syst. Rehab. Eng.* 1998;6:326–333.

Author: Victoria Peterson
Institute: Sinc(i) - FICH - UNL
Street: Ruta Nacional N 168 - Km 472.4
City: Santa Fe
Country: Argentina
Email: vpeterson@santafe-conicet.gov.ar