



HAL
open science

Classification of evoked potentials by Pearson's correlation in a brain-computer interface

François Cabestaing, Theresa M Vaughan, Dennis J Mcfarland, Jonathan R Wolpaw

► **To cite this version:**

François Cabestaing, Theresa M Vaughan, Dennis J Mcfarland, Jonathan R Wolpaw. Classification of evoked potentials by Pearson's correlation in a brain-computer interface. *AMSE, Modelling C Automatic Control (theory and applications)*, 2007, 67, pp.156-166. hal-00521049

HAL Id: hal-00521049

<https://hal.science/hal-00521049>

Submitted on 24 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Classification of evoked potentials by Pearson's correlation in a Brain-Computer Interface

F. Cabestaing*, T.M. Vaughan**, D.J. McFarland**, J.R. Wolpaw**

* *Laboratoire d'Automatique, Génie Informatique & Signal, CNRS UMR-8146
Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq CEDEX*

** *Laboratory of Nervous System Disorders, PO Box 509, Wadsworth Center
NYSDOH, Empire State Plaza, Albany, NY 12201-0509*

Abstract: In this paper, we describe and evaluate the performance of a linear classifier learning technique for use in a brain-computer interface. Electroencephalogram (EEG) signals acquired from individual subjects are analyzed with this technique in order to detect responses to visual stimuli. Signal processing and classification are used for implementing a palliative communication system which allows the individual to spell words. Performance with this technique is evaluated on data collected from eight individuals.

Keywords: Brain-computer interface; BCI; linear classifier; evoked potentials; P300.

1 Introduction

A brain-computer interface (BCI) is a system which allows direct communication, without using muscles or peripheral nerves, between an individual and a computer. This technique aims at giving people suffering from severe disabilities a way to communicate with their environment. For people who are locked-in, after a complete loss of muscular control, a BCI is the most promising palliative communication technique. Some patients, suffering from degenerative diseases like ALS (Amyotrophic Lateral Sclerosis), who lose all voluntary muscular control in the late stages of the disease, could benefit from this technique. Several million people could benefit from BCI systems worldwide.

The functioning principle of a BCI is sketched in figure 1. Brain activity is measured by electrical signals, acquired either on the scalp (EEG: electroencephalogram), on the cortex surface (ECoG: electrocorticogram), or directly on single neurons (SNA: Single Neuron Activity). The recorded signals are processed by the system, which translates the brain activity into commands sent to the computer.

The BCI team of the Laboratory of Nervous System Disorders at Wadsworth Center (NYS Department of Health, Albany, NY, USA) has proposed numerous paradigms that allow the use for communication of information extracted from EEG signals in order to build a BCI [1]. A first approach consists in analyzing the amplitude of rhythms normally related to muscular activity, which are mainly detected in EEG from sensorimotor areas of the cortex [2]. For example, the amplitude of the μ rhythm changes not only during actual movement, but also when the individual only imagines that he is moving. After a training period of 3-4 hours over several weeks, most people can use their sensorimotor rhythms to operate a simple BCI. This approach has been used in the Wadsworth BCI to control a cursor on a computer screen, along one and more recently two dimensions [2].

Another approach to BCI, which was initially proposed by Farwel and Donchin [3], consists

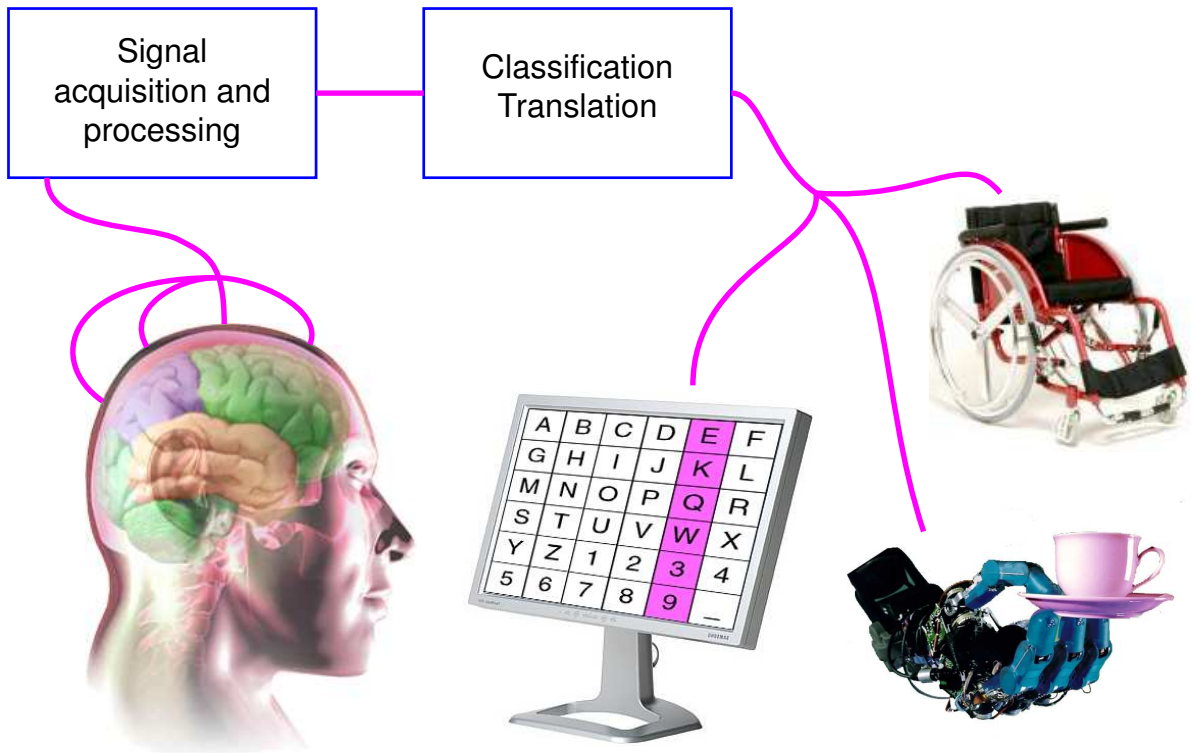


Figure 1: Structure of a brain-computer interface

in detecting in EEG signals the potentials evoked by stimuli perceived by the individual (ERP: Event Related Potentials). More specifically, P300 ERPs, that are elicited when the individual reacts in a cognitive way to visual, tactile or auditive stimuli, have been used to control a BCI. One advantage of P300-based BCIs is that they do not require user training. The P300 potentials are mainly elicited when the individual reacts to rare stimuli appearing from time to time among standard ones. This paradigm is called the 'Oddball paradigm' [4].

An ingenious visual paradigm, which satisfies the oddball condition, consists in presenting to the individual a matrix of symbols and asking him to stay focused on one of these (cf. figure 2). During the trial, columns and rows of the matrix are intensified in a random order, and the user must count the number of intensifications occurring for the selected symbol. The stimuli are the column and row intensifications. The rare stimuli are the intensifications of the column and the row including the symbol to be selected. For a majority of individuals, these rare stimuli elicit a P300 ERP which can be detected by the BCI. This paradigm has been used by several teams to build BCIs that allow simple word processing.

A	B	C	D	E	F	A	B	C	D	E	F
G	H	I	J	K	L	G	H	I	J	K	L
M	N	O	P	Q	R	M	N	O	P	Q	R
S	T	U	V	W	X	S	T	U	V	W	X
Y	Z	1	2	3	4	Y	Z	1	2	3	4
5	6	7	8	9	_	5	6	7	8	9	_

Figure 2: Matrix of symbols, here alphanumeric, whose columns and rows are individually intensified in a random order during the spelling sequence

Rare stimuli, which correspond to an intensification of the column or row including the symbol on which the user focuses his attention, elicit an ERP which must be detected in the signals. As soon as the ERP is detected, the system can determine which stimulus has elicited it, the latter being called 'target stimulus'. Other stimuli, which correspond to columns or rows not including the selected symbol, are called 'non-target stimuli'. EEG signals are recorded continuously, each signal being sampled and digitized in order to be processed by the system. The set of values corresponding to all the signals, in a time window of fixed duration starting at the stimulus onset, is called 'temporal response' to the stimulus.

In this paper, we focus our work on the determination of the type of stimulus, i.e. 'target' or 'non-target', which has elicited a given temporal response. In the first part of the article, we show that a linear classification technique allows one to reach this goal. We recall the principles of linear classification and explain why standard classifier training techniques are not well adapted in our case. In the second part, we show how the coefficients of the linear classifier can be determined by a statistical analysis of the temporal responses belonging to a small training set. In the last part, we present the classification results obtained using the proposed method on experimental data recorded from eight individuals.

2 Classification technique

In this part, we describe the technique used for determining the stimulus type, target or non-target, using the samples of the temporal response. The weight vector of a linear classifier is determined by a simple but efficient technique, in which the dimensions of the feature space are used independently.

2.1 Linear classifier

In the two-class case, a linear classifier is a function $h(\cdot)$ defining a mapping between the feature space \mathcal{X} , a subset of \mathbb{R}^n , and the set of labels $\mathcal{Y} = \{-1, +1\}$. It is defined by a weight vector $\mathbf{w} \in \mathbb{R}^n$ and a bias $b \in \mathbb{R}$ by:

$$h(\mathbf{x}) = \text{sign}(\mathbf{w}^T \mathbf{x} + b) = \begin{cases} +1 & \text{if } \mathbf{w}^T \mathbf{x} + b > 0 \\ -1 & \text{if } \mathbf{w}^T \mathbf{x} + b < 0 \end{cases} . \quad (1)$$

$\mathbf{w}^T \mathbf{x} + b = 0$ is the equation of a hyperplane which divides \mathbb{R}^n into two subspaces corresponding to classes. For example, with $n = 2$, a linear classifier is defined by a straight line dividing the plane into two half-planes, as indicated in figure 3.

In our case, the goal of the classification step is to determine, using the temporal responses recorded over the set of electrodes, whether the stimulus corresponds or not to a target. Let's consider the average responses to target and non-target stimuli recorded on the Cz electrode, which are represented in figure 4. As can be seen in Figure 4, only some of the samples of the temporal response are useful in discriminating the stimulus type. If two values, on the average responses to target and non-target stimuli, are equal for a given sample, the latter cannot be used to determine the stimulus type. On the other hand, if the two average values are different, this sample can be used to recognize the response type, assuming that the associated variance is small enough.

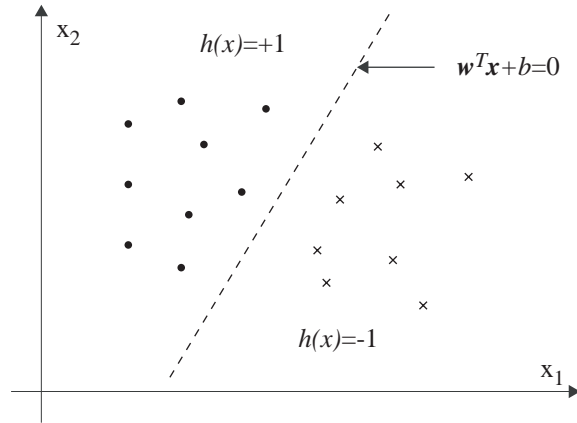


Figure 3: 2D linear classifier, separation of elements in two classes

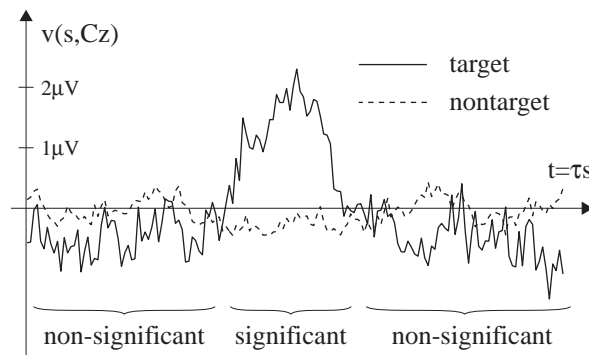


Figure 4: Average responses to target and non-target stimuli, recorded on electrode Cz of the EEG

Therefore, a basic linear classifier can be defined on the basis of a limited number of samples, found to be informative on some specific electrodes [3]. The weight vector of this classifier is determined empirically using the average responses to target and non-target stimuli. For example, one can use the samples of the signal recorded on the Cz electrode between 200 and 600 milliseconds after stimulus onset, since it is known to show typical P300 type evoked potentials. Statistically, for a P300 ERP, an informative sample takes a higher value for a target stimulus than for a non-target one. In this case, one assigns a positive value to the weight of this sample, in order to get a positive term for this sample in the scalar product $w^T x$ of equation (1). The informative dimensions of feature space and the associated coefficients must be determined for each individual, since the ERPs are very variable in space and time between different people.

Rather than selecting a few particular feature dimensions, we recommend using all the dimensions of the feature space in the linear classifier. The temporal response to a stimulus is therefore represented by a vector, denoted x , with $n_e \times n_s$ coordinates, including all the dimensions of the feature space. n_e is the number of electrodes of the EEG cap, each one delivering a signal digitized in n_s samples. With a sampling rate τ , the response duration is then $n_s \tau$. For example, with 64 electrodes sampled at 300Hz, a one-second response is described by a vector with 19200 coordinates. With such a representation, the spatial and temporal orders of features are lost, since the coordinates of a vector are not ordered.

2.2 Drawbacks of learning techniques

In the supervised case, classifier parameters are determined using a learning set, composed of pairs $(\mathbf{x}_i, y_i) \in \mathcal{X} \times \mathcal{Y}$ of elements \mathbf{x}_i of the feature space for which the labels are known. This set is used as a whole by batch learning techniques, or element by element by iterative learning techniques. The reader can refer to [5] for a detailed presentation of learning techniques.

Fisher's discriminant analysis is the reference batch learning method [6]. All the elements of the learning set are projected onto the straight line perpendicular to the hyperplane defined by \mathbf{w} , in order to define two subsets $m^+ = \{\mathbf{w}^T \mathbf{x}_i \mid y_i = +1\}$ and $m^- = \{\mathbf{w}^T \mathbf{x}_i \mid y_i = -1\}$ corresponding to the two classes. On this straight line, the two subsets are well separated when their two centers of gravity are distant, whereas in each subset the variance of positions remains small. In order to determine the best separating hyperplane, Fisher proposes to maximize the criterion:

$$J(\mathbf{w}) = \frac{[\text{avg}(m^+) - \text{avg}(m^-)]^2}{\text{var}(m^+) + \text{var}(m^-)}, \quad (2)$$

in which $\text{avg}(\cdot)$ and $\text{var}(\cdot)$ denote respectively the average value and the variance of positions of all the points of a set.

Among iterative learning techniques, the simplest and most widely used is the perceptron learning rule, originally described by Rosenblatt [7]. The elements of the learning set, which in this case is not always defined a priori, are used one after the other in order to update the weight vector of the linear classifier. Let \mathbf{w}_k denote the weight vector estimated using the first k elements of the learning set and \hat{y}_k the estimation of the class obtained with \mathbf{x}_k . If the classifier makes a mistake for one of the elements of the learning set (i.e. if the predicted value \hat{y}_k is different from the actual one y_k for this element), the weight vector is updated following the rule:

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \eta y_k \mathbf{x}_k, \quad (3)$$

in which η is a constant called the learning rate.

These well known learning methods, and their equivalents, are not well adapted to our case, either because they require too many computations, or because the learning set does not include enough observations. Indeed, our learning set includes a limited number (several hundreds) of time responses associated with a given type of stimulus, while the feature space has a much higher dimension (several thousands). Maximizing Fisher's discriminant criterion requires computing a cross-correlation matrix, a square matrix with size $n_e n_s \times n_e n_s$, and its inverse. This computation would require hours of processing time for a square matrix with size 19200×19200 even on a powerful computer. Moreover, the result would probably be very imprecise because of the accumulation of thousands of unavoidable computation errors due to arithmetic rounding. On the other hand, for iterative learning methods, the problem of the limited size of the learning set arises. Learning several thousands of coefficients using an iterative method requires at least a comparable number of elements in the learning set. One usually considers that the ratio between the number of learning elements and the number of coefficients should be at least ten, which is far from our case.

3 Proposed tuning technique

Pearson's correlation coefficient, usually denoted by r , is a statistical analysis tool that we use to quantify the degree to which a given feature of a time series predicts the criterion. This coefficient, which measures the correlation between two series $X = x_i, 1 \leq i \leq n_i$ and $Y = y_i, 1 \leq i \leq n_i$, is computed by the equation:

$$r = \frac{n_i \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{[n_i \sum x_i^2 - (\sum x_i)^2] [n_i \sum y_i^2 - (\sum y_i)^2]}}. \quad (4)$$

r reflects the degree of proportionality between the two series by a value ranging between -1 and $+1$. If all the values of the first series are strictly proportional to those of the second series, r is either $+1$ or -1 . If the two series show no correlation, r is equal to zero.

Let n_i denote the number of elements in the learning set and Y the series of values $y_i, 1 \leq i \leq n_i$ defining the type of stimulus for each element, with $+1$ for a target stimulus and -1 for a non-target one. $X(s, e)$ is the series of values $x_i(s, e), 1 \leq i \leq n_i$, of the particular sample s of electrode e in the n_i time responses. We compute the Pearson's correlation coefficient $r(s, e)$ between the series $X(s, e)$ and Y according to equation 4. The higher the absolute value of $r(s, e)$, the most discriminant the sample s of electrode e to determine the type of stimulus. For example, one can verify this property on the curve in figure 5, which represents the values of the $r(s, e)$ coefficients for all the samples s of the $e=Cz$ electrode.

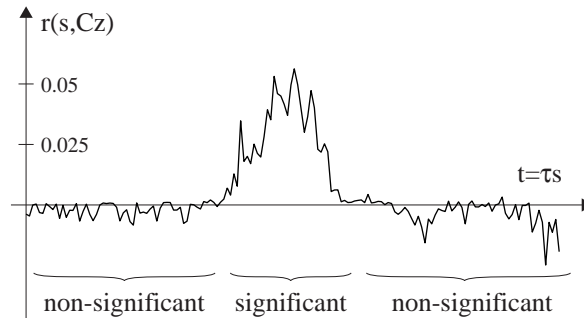


Figure 5: Pearson's correlation coefficient vs. time, Cz electrode

Rather than selecting the most informative samples of the time response, we use all the coefficients $r(s, e)$ as the weight vector w of a linear classifier defined on a feature space with dimensions $n_e \times n_s$. If a sample is informative, its value is multiplied by a non-null coefficient and accumulated to the sum. If it is not informative, its value is multiplied by a very small coefficient and therefore has no influence on the result. On the other hand, the value of a sample in the average response to a target stimulus can be either greater or lower than the value of the same sample in the average response to a non-target stimulus. The sign of $r(s, e)$ takes this property into account, since the product $r(s, e)x(s, e)$ is statistically greater than zero for a sample of a target response.

This technique for determining the coefficients of the linear classifier, which is a batch type approach, uses each axis of the feature space independently. It is very efficient in terms of computation time, since its complexity is only $o(n)$, where n is the dimension of the feature space. For comparison, standard batch techniques, such as maximization of Fisher's discriminant, have a $o(n^3)$ complexity. However, our technique does not yield a solution that maximizes any optimality criterion, and therefore it can hardly be described as a 'learning' technique.

4 Performance evaluation and discussion

4.1 Experimental data

Eight people (six men and two women between 24 and 50 years of age) used the BCI system to provide the data used in the following experiment. Some had already used a BCI system, but no one had experience with an ERP-based BCI. The data, acquired during five independent sessions for each individual, were recorded in similar experimental conditions over several months.

The user sat in front of a video monitor and viewed the matrix of alphanumeric symbols presented as in figure 2. During each trial, he or she had to focus on a given symbol and to count the number of flashes occurring for this target character. All the data were collected in the so-called copy mode (the target symbol was designated a priori to the user by the computer). During each session, the user had to spell nine words or series of symbols, namely the eight words of the sentence "The quick brown fox jumps over the lazy dog" and a number with three digits. These series were chosen to span a large part of the matrix of symbols during each trial. Columns and rows were intensified for 100 milliseconds with a period of 175 milliseconds. For spelling one symbol, each column and each row was intensified 15 times, in a random order, which corresponds to 180 stimuli, among which 30 were targets.

The EEG signals were acquired using a 64 electrodes cap (Electro-Cap International Inc.) organized following the standard international 10 – 20 system. All the electrodes were referenced to the right ear lobe and grounded to the right mastoid. Signals were amplified by a factor of 20 000 (SA Electronics amplifier), band-pass filtered between 0.1 and 60Hz, sampled at 240Hz and recorded. Data collection and stimulus generation were handled entirely by the BCI2000 software [8].

4.2 Validation and comparison protocol

A potential evoked by a stimulus appearing within 800 milliseconds following the stimulus onset was considered informative. This duration corresponds to 192 samples for each electrode when the signal is sampled at 240Hz. Taking into account the 64 electrodes, this leads to processing 12288 features for each time response. In most other classification techniques, the first step consists in removing many dimensions from the feature space to reduce complexity, which allows for a more efficient learning step. Dimension reduction is obtained by subsampling the signals in either the temporal or spatial domain, or in both. In order to check the validity of these subsamplings and to validate simultaneously our classification technique, we analyzed the data in the following four configurations:

1. with the initial 12288 dimensions of the feature space, i.e. 64 electrodes sampled at 240Hz.
2. using only 8 electrodes, selected by an expert as the most interesting for detecting ERPs, reducing the dimensions to 1536.
3. by subsampling the signal down to 20Hz, with low-pass filtering and decimation, reducing the dimensions to 1024.

4. using simultaneously both reductions, i.e. 8 electrodes sampled at 20Hz, which yields a feature space with 128 dimensions.

With five sessions for each individual, several learning/test combinations could be used. We selected two options: 1) learning classifier parameters on the data of the first session and testing them on the remaining four sessions; 2) learning the parameters on each session and testing them on the following one. Under these conditions, seven pairs of (learning, testing) sessions were available for each individual. For learning the coefficients, all the time responses available in the session were used, i.e. 35 symbols, each one including 180 time responses, yielding 6300 elements in the learning set.

The performance of the proposed technique was evaluated using the correct classification rate for each pair (learning, testing) previously described. We also studied the evolution of the correct classification rate versus the number of stimuli used for testing, considering a number of flashes ranging between 1 and 15.

4.3 Learning results

An example of a learning result, for a single session of a given individual in the four configurations of the feature space, is presented in figure 6. Each image represents the weight vector of the classifier, the horizontal direction corresponding to the sample number (time), the vertical direction to the electrode number, and the gray level to the absolute value of the weight determined for this axis of the feature space. White corresponds to a large absolute value of the weight, which means that this axis is informative in terms of stimulus discrimination, whereas black corresponds to a null weight.

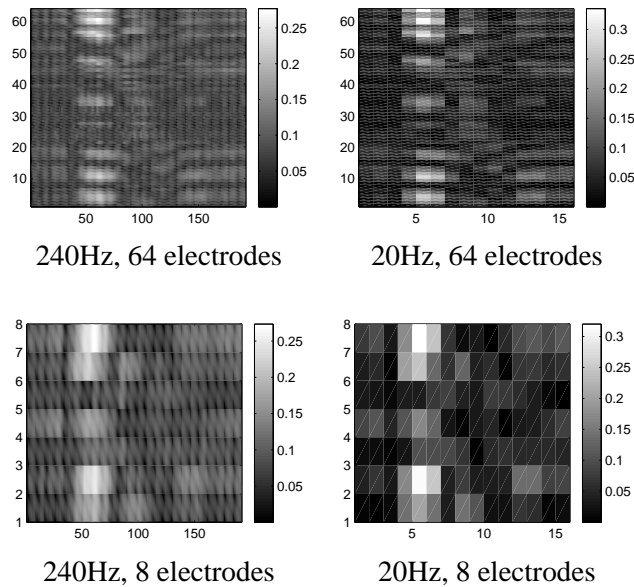


Figure 6: Learning results in four configurations. The weight vector is represented as an image: the abscissa is the sample number, the ordinate the electrode number, the gray level is proportional to the absolute value of the weight

As can be seen in Figure 6, some samples of the time response are very informative (e.g., near the sixty-fifth sample for 240Hz and near the sixth sample for 20Hz). These samples correspond

quite precisely to a time offset of 300 milliseconds, which confirms their relation to P300 ERPs. In addition, a reduction of the resolution, either spatial or temporal, yields consistent weight vectors, since their aspects are very similar when they are represented by an image.

4.4 Correct classification rate

The correct classification rate allows one to verify the correct behavior of the proposed learning technique. In table 1, we indicate the correct classification rates obtained for the eight individuals (A-H) with the seven available pairs (learning session, testing session). The method yields high correct classification rates, although closely related to the individual's ability to use an ERP-based BCI. There is no noticeable evolution of this rate during successive sessions, presumably because ERPs are unlearned responses on which learning has very little influence.

	1-2	1-3	1-4	1-5	2-3	3-4	4-5
A	66,67	55,56	44,44	36,11	69,44	41,67	41,67
B	78,38	75,68	71,05	72,97	59,46	78,95	64,86
C	86,11	80,56	83,33	88,89	66,67	88,89	88,89
D	66,67	86,11	75	30,56	75	72,22	36,11
E	97,22	96,55	97,22	97,22	100	100	100
F	62,16	77,42	57,89	78,38	83,87	68,42	83,78
G	86,11	80,56	100	91,67	88,89	97,22	94,44
H	66,67	63,89	16,67*	83,33	72,22	11,11*	83,33

Table 1: Correct classification rate (in %) for each individual versus the pair (learning session, testing session). * during session 4, the individual H did not stay well focused on the task

In figure 7, we show the evolution of the correct classification rate versus the number of intensifications series used for testing. This rate increases with the number of intensifications, which is explained by the noise lowering effect of averaging. There is also a disparity among individuals in terms of ERPs, since some users can reach high rates (70% to 100%), with only six intensifications, whereas for some others the rate evolution with the number of intensifications is progressive.

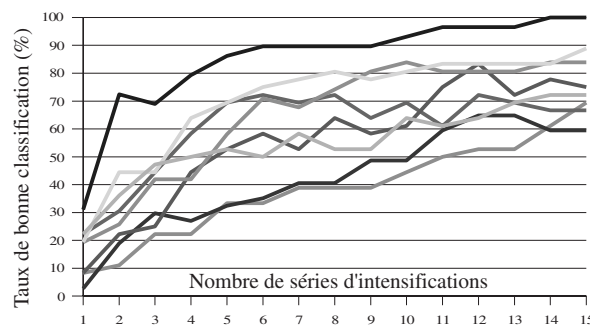


Figure 7: Evolution of the correct classification rate versus the number of intensifications

As previously indicated, the method has been applied to data in four different configurations of spatial and temporal resolutions. This way, it is possible to compare the correct classification rates obtained versus the dimension of the feature space. The result of this comparison is

summarized in table 2, which presents the evolution of the correct classification rate for each individual among all sessions when the resolution is changed.

	A	B	C	D	E	F	G	H
Modification of electrode number (64 downto 8), constant sampling rate								
	Average variation of the rate compared to table I							
64 downto 8 (240)	+10,19	-4,46	+8,89	+10,79	+2,73	+12,21	+4,92	+8,47
64 downto 8 (20)	+9,02	-5	+8,99	+8,86	+3,77	+12,17	+5,79	+7,46
Modification of sampling rate (240 downto 20), constant electrode number								
	Average variation of the rate compared to table I							
240 downto 20 (64)	+0,53	-1,29	-2,75	-1,75	-3,09	-2,44	-3,25	-0,53
240 downto 20 (8)	-0,63	-1,83	-2,65	-3,68	-2,05	-2,48	-2,38	-1,53

Table 2: Correct classification rate (in %) versus modifications of spatial and temporal resolutions

Let us consider first the effect of changing the spatial resolution. Electrode selection allows an increase in the correct classification rate for seven of the eight individuals (e.g., up to 12% increase for individual F). We believe that this improvement comes from the reduction of the noise introduced in the response by non informative electrodes. The noise is not perfectly removed by the learning technique, which is non-optimal. On the other hand, for individual B, reducing the resolution down to eight electrodes leads to a 4.5% decrease in the correct classification rate, which means that some useful information was available in the cancelled signals. In sum, a priori reduction of the number of electrodes is a useful technique, but the informative electrodes should be selected on an individual basis.

The effect of reducing the temporal resolution of signals is that it globally leads to a decrease in the correct classification rate, although limited to about 3% with our experimental data. A priori subsampling removes, in a way not adapted to each individual, some information that was initially available in the signals. This result suggests that it is best to keep all the samples of the original signal and let the learning technique select the informative ones.

5 Conclusion

In this article, we have described a technique for the classification of time responses to visual stimuli that can be easily implemented in a brain-computer interface. Its algorithmic complexity is low compared to those of the standard learning techniques described in the literature. Moreover, this technique can process raw signals coming from the EEG recording system, without requiring a reduction of data resolution. Using experimental data, we have shown that a reduction of the number of electrodes can noticeably increase the correct classification rate. In contrast, reducing the sampling rate decreases the correct classification rate.

Acknowledgements

The authors acknowledge the support of the french Nord-Pas de Calais regional council, of FEDER, of the IRRH (Institut Régional de Recherche sur le Handicap), of NIH (under grants

NICHHD HD30146 et NIBIB/NINDS EB00856) and of the Foundation James S. McDonnell. They thank Dr. Elizabeth Winter Wolpaw for her help in improving this manuscript.

References

- [1] J. R. Wolpaw, D. J. McFarland, T. M. Vaughan, and G. Schalk, "The Wadsworth Center brain-computer interface (BCI) research and development program," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 1–4, 2003.
- [2] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans," *Proceedings of the National Academy of Science of the USA*, vol. 101, no. 51, pp. 17 849–17 854, 2004.
- [3] L. A. Farwell and E. Donchin, "Talking off the top of your head: A mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, pp. 510–513, 1988.
- [4] D. Friedman, Y. M. Cycowicz, and H. Gatea, "The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty," *Neuroscience & Biobehavioral Reviews*, Elsevier, vol. 25, pp. 355–373, 2001.
- [5] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification, second edition*. Wiley-Interscience Publication, 2000.
- [6] R. A. Fisher, "The statistical utilization of multiple measurements," *Annals of Eugenics*, vol. 8, pp. 376–386, 1938.
- [7] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65, pp. 386–408, 1958.
- [8] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 1034–1043, 2004.