

# Histogram of Gradient Orientations of Signal Plots applied to P300 Detection

Rodrigo Ramele<sup>1,\*</sup>, Ana Julia Villar<sup>1</sup> and Juan Miguel Santos<sup>1</sup>

<sup>1</sup>*Centro de Inteligencia Computacional, Computer Engineering Department, Instituto Tecnológico de Buenos Aires (ITBA), Buenos Aires, Argentina*

Correspondence\*:

Rodrigo Ramele, C1437FBH Lavarden 315, Ciudad Autónoma de Buenos Aires, Argentina  
rramele@itba.edu.ar

## 2 ABSTRACT

3 Word Count: ~~5552~~ 6144

4 The analysis of Electroencephalographic (EEG) signals is of ulterior importance to aid in the  
5 diagnosis of mental disease and to increase our understanding of the brain. Traditionally, clinical  
6 EEG has been analyzed in terms of temporal waveforms, looking at rhythms in spontaneous  
7 activity, subjectively identifying troughs and peaks in Event-Related Potentials (ERP), or by  
8 studying graphoelements in pathological sleep stages. Additionally, the discipline of Brain  
9 Computer Interfaces ([BCI](#)) requires new methods to decode patterns from non-invasive EEG  
10 signals. This field is developing alternative communication pathways to transmit volitional  
11 information from the Central Nervous System. The technology could potentially enhance the  
12 quality of life of patients affected by neurodegenerative disorders and other mental illness. This  
13 work mimics what electroencephalographers have been doing clinically, visually inspecting and  
14 categorizing phenomena within the EEG by the extraction of features from images of signal plots.  
15 These features are constructed based on the calculation of histograms of oriented gradients  
16 from pixels around the signal plot. It aims to provide a new objective framework to analyze,  
17 characterize and classify EEG signal waveforms. The feasibility of the method is outlined by  
18 detecting the P300, an ERP elicited by the oddball paradigm of rare events, and implementing  
19 an offline P300-based BCI Speller. The validity of the proposal is shown by offline processing a  
20 public dataset of Amyotrophic Lateral Sclerosis (ALS) patients and an own dataset of healthy  
21 subjects.

22 **Keywords:** [electroencephalography](#), [histogram of gradient orientations](#), [brain-computer interfaces](#), [P300](#), [SIFT](#), [amyotrophic lateral](#)  
23 [sclerosis](#), [naive-bayes near neighbours](#), [waveforms](#)

## 1 INTRODUCTION

24 Although recent advances in neuroimaging techniques, particularly radio-nuclear and radiological  
25 scanning methods (?), have diminished the prospects of the traditional Electroencephalography([EEG](#)),  
26 the advent and development of digitized devices has impelled for a revamping of this hundred years old  
27 technology. Their versatility, ease of use, temporal resolution, ease of development and production, and  
28 its proliferation as consumer devices, are pushing EEG to become the de-facto non invasive portable or  
29 ambulatory method to access and harness brain information (?).

30 A key contribution to this expansion has been the field of Brain Computer Interfaces(BCI) (?) which is  
31 the pursuit of the development of a new channel of communication particularly aimed to persons affected  
32 by neurodegenerative diseases.

33 One noteworthy aspect of this novel communication channel is the ability to transmit information from  
34 the Central Nervous System (CNS) to a computer device and from there use that information to control a  
35 wheelchair (?), as input to a speller application (?), in a Virtual Reality environment (?) or as aiding tool  
36 in a rehabilitation procedure (?). The holly grail of BCI is to implement a new complete and alternative  
37 pathway to restore lost locomotion (?).

38 EEG signals are remarkably complex and have been characterized as a multichannel non-stationary  
39 stochastic process. Additionally, they have high variability between different subjects and even between  
40 different moments for the same subject, requiring adaptive and co-adaptive calibration and learning  
41 procedures (?). Hence, this imposes an outstanding challenge that is necessary to overcome in order to  
42 extract information from raw EEG signals.

43 BCI has gained mainstream public awareness with worldwide challenge competitions like Cybathlon (??)  
44 and even been broadcasted during the inauguration ceremony of the 2014 Soccer World Cup. New  
45 developments have overcome the out-of-the-lab high-bar and they are starting to be used in real world  
46 environments (??). However, they still lack the necessary robustness, and its performance is well behind  
47 any other method of human computer interaction, including any kind of detection of residual muscular  
48 movement (?).

49 A few works have explored the idea of exploiting the signal waveform to analyze the EEG signal. In (?)  
50 an approach based on Slope Horizontal Chain Code is presented, whereas in (?) a similar procedure  
51 was implemented based on Mathematical Morphological Analysis. The seminal work of Bandt-Pompe  
52 Permutation Entropy (?) also explores succinctly this idea as a basis to establish the time series ordinal  
53 patterns. In the article (?), the authors introduce a method for classification of rhythmic EEG events like  
54 Visual Occipital Alpha Waves and Motor Imagery Rolandic Central  $\mu$  Rhythms using the Histogram  
55 of Gradient Orientations of signal plots. Inspired in that work, we propose a novel application of the  
56 developed method to classify and describe transient events, particularly the P300 Event Related Potential.  
57 The proposed approach is based on the waveform analysis of the shape of the EEG signal. The signal is  
58 drawn on a bidimensional image plot, vector gradients of pixels around the plot are obtained, and with  
59 them, the histogram of their orientations is calculated. This histogram is a direct representation of the  
60 waveform of the signal. The method is built by mimicking what regularly electroencephalographers have  
61 been performing for almost a century as it is described in (?): visually inspecting raw signal plots.

62 This paper reports a method to, (1) describe a procedure to capture the shape of a waveform of an ERP  
63 component, the P300, using histograms of gradient orientations extracted from images of signal plots,  
64 and (2) outline the way in which this procedure can be used to implement an P300-Based BCI Speller  
65 application. Its validity is verified by offline processing two datasets, one of data from ALS patients and  
66 another one from data of healthy subjects.

67 This article unfolds as follows: Section 2.1 is dedicated to explain the Feature Extraction method based  
68 on Histogram of Gradient Orientations of the Signal Plot: Section 2.1.1 shows the preprocessing pipeline,  
69 Section 2.1.2 describes the image generation of the signal plot, Section 2.1.3 presents the feature extraction  
70 procedure while Section 2.1.4 introduces the Speller Matrix Letter Identification procedure. In Section 2.2,  
71 the experimental protocol is expounded. Section 3 shows the results of applying the proposed technique. In  
72 the final Section 4 we expose our remarks, conclusions and future work.

## 2 MATERIALS AND METHODS

73 The P300 (??) is a positive deflection of the EEG signal which occurs around 300 ms after the onset of a  
74 rare and deviant stimulus that the subject is expected to attend. It is produced under the oddball paradigm (?)  
75 and it is consistent across different subjects. It has a lower amplitude ( $\pm 5\mu V$ ) compared to basal EEG  
76 activity, reaching a Signal to Noise Ratio (SNR) of around  $-15$  db estimated based on the amplitude of the  
77 P300 response signal divided by the standard deviation of the background EEG activity (?). This signal  
78 can be used to implement a speller application by means of a Speller Matrix (?). This matrix is composed  
79 of 6 rows and 6 columns of numbers and letters. The subject can focus on one character of the matrix.  
80 Figure 1 shows an example of the Speller Matrix used in the OpenVibe open source software (?), where  
81 the flashes of rows and columns provide the deviant stimulus required to elicit this physiological response.  
82 Each time a row or a column that contains the desired letter flashes, the corresponding synchronized EEG  
83 signal should also contain the P300 signature and by detecting it, the selected letter can be identified.

84 **2.1 Feature Extraction from Signal Plots**

85 In this section, the signal preprocessing, the method for generating images from signal plots, the feature  
86 extraction procedure and the Speller Matrix identification are described. Figure 2 shows a scheme of the  
87 entire process.

88 **2.1.1 Preprocessing Pipeline**

89 The data obtained by the capturing device is digitalized and a multichannel EEG signal is constructed.

90 The 6 rows and 6 columns of the Speller Matrix are intensified providing the visual stimulus. The  
91 number of a row or column is a location. A sequence of twelve randomly permuted locations  $l$  conform an  
92 intensification sequence. The whole set of twelve intensifications is repeated  $k_a$  times.

93 • **Signal Enhancement:** This stage consists of the enhancement of the SNR of the P300 pattern above  
94 the level of basal EEG. The pipeline starts by applying a notch filter to the raw digital signal, a 4th  
95 degree 10 Hz lowpass Butterworth filter and finally a decimation with a Finite Impulse Response (FIR)  
96 filter of order 30 from the original sampling frequency down to 16 Hz (?).

97 • **Artifact Removal:** For every complete sequence of 12 intensifications of 6 rows and 6 columns, a  
98 basic artifact elimination procedure is implemented by removing the entire sequence when any signal  
99 deviates above/below  $\pm 70\mu V$ .

100 • **Segmentation:** For each of the 12 intensifications of one intensification sequence, a segment  $S_i^l$   
101 of a window of  $t_{max}$  seconds of the multichannel signal is extracted, starting from the stimulus  
102 onset, corresponding to each row/column intensification  $l$  and to the intensification sequence  $i$ . As  
103 intensifications are permuted in a random order, the segments are rearranged corresponding to row  
104 flickering, labeled 1-6, whereas those corresponding to column flickering are labeled 7-12. Two of  
105 these segments should contain the P300 ERP signature time-locked to the flashing stimulus, one for  
106 the row, and one for the column.

107 • **Signal Averaging:** The P300 ERP is deeply buried under basal EEG so the standard approach to  
108 identify it is by point-to-point averaging the time-locked stacked signal segments. Hence the values  
109 which are not related to, and not time-locked to the onset of the stimulus are canceled out (?).

110 This last step determines the operation of any P300 Speller. In order to obtain an improved signal  
111 in terms of its SNR, repetitions of the sequence of row/column intensification are necessary. And,

at the same time, as long as more repetitions are needed, the ability to transfer information faster is diminished, so there is a trade-off that must be acutely determined.

The procedure to obtain the point-to-point averaged signal goes as follows:

1. Highlight randomly the rows and columns from the matrix. There is one row and one column that should match the letter selected by the subject.
2. Repeat step 1  $k_a$  times, obtaining the  $1 \leq l \leq 12$  segments  $S_1^l(n, c), \dots, S_{k_a}^l(n, c)$ , of the EEG signal where the variables  $1 \leq n \leq n_{max}$  and  $1 \leq c \leq C$  correspond to sample points and channel, respectively. The parameter  $C$  is the number of available EEG channels whereas  $n_{max} = F_s t_{max}$  is the segment length and  $F_s$  is the sampling frequency. The parameter  $k_a$  is the number of repetitions of intensifications and it is an input parameter of the algorithm.
3. Compute the Ensemble Average by

$$x^l(n, c) = \frac{1}{k_a} \sum_{i=1}^{k_a} S_i^l(n, c) \quad (1)$$

for  $1 \leq n \leq n_{max}$  and for the channels  $1 \leq c \leq C$ . This provide an averaged signal  $x^l(n, c)$  for the twelve locations  $1 \leq l \leq 12$ .

### 2.1.2 Signal Plotting

Averaged signal segments are standardized and scaled for  $1 \leq n \leq n_{max}$  and  $1 \leq c \leq C$  by

$$\tilde{x}^l(n, c) = \left\lfloor \gamma \frac{(x^l(n, c) - \bar{x}^l(c))}{\hat{\sigma}^l(c)} \right\rfloor \quad (2)$$

where  $\gamma > 0$  is an input parameter of the algorithm and it is related to the image scale. In addition,  $x^l(n, c)$  is the point-to-point averaged multichannel EEG signal for the sample point  $n$  and for channel  $c$ . Lastly,

$$\bar{x}^l(c) = \frac{1}{n_{max}} \sum_{n=1}^{n_{max}} x^l(n, c)$$

and

$$\hat{\sigma}^l(c) = \left\{ \frac{1}{n_{max} - 1} \sum_{n=1}^{n_{max}} [x^l(n, c) - \bar{x}^l(c)]^2 \right\}^{\frac{1}{2}}$$

are the mean and estimated standard deviation of  $x^l(n, c)$ ,  $1 \leq n \leq n_{max}$ , for each channel  $c$ .

Consequently, a binary image  $I^{(l,c)}$  is constructed according to

$$I^{(l,c)}(z_1, z_2) = \begin{cases} 255 & \text{if } z_1 = \gamma n \quad \text{and} \quad z_2 = \tilde{x}^l(n, c) + z^l(c) \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

with 255 being white and representing the signal's value location and 0 for black which is the background contrast, conforming a black-and-white plot of the signal. Pixel arguments  $(z_1, z_2) \in \mathbb{N} \times \mathbb{N}$  iterate over the width (based on the length of the signal segment) and height (based on the peak-to-peak amplitude) of the newly created image with  $1 \leq n \leq n_{max}$  and  $1 \leq c \leq C$ . The value  $z^l(c)$  is the image vertical position

133 where the signal's zero value has to be situated in order to fit the entire signal within the image for each  
 134 channel c:

$$z^l(c) = \left\lfloor \frac{\max_n \tilde{x}^l(n, c) - \min_n \tilde{x}^l(n, c)}{2} \right\rfloor - \left\lfloor \frac{\max_n \tilde{x}^l(n, c) + \min_n \tilde{x}^l(n, c)}{2} \right\rfloor \quad (4)$$

135 where the minimization and maximization are carried out for n varying between  $1 \leq n \leq n_{max}$ , and  $\lfloor \cdot \rfloor$   
 136 denote the rounding to the smaller nearest integer of the number.

137 In order to complete the plot  $I^{(l,c)}$  from the pixels, the Bresenham (??) algorithm is used to interpolate  
 138 straight lines between each pair of consecutive pixels.

### 139 2.1.3 Feature Extraction: Histogram of Gradient Orientations

140 The work of Edelman, Intrator and Poggio (?) on how the visual cortex sense features was the inspiration  
 141 to the development of an algorithm to identify and decode salient local information from image regions. The  
 142 Scale Invariant Feature Transform (SIFT) is a Computer Vision method proposed by ? which is composed  
 143 of two parts, the SIFT Detector and the SIFT Descriptor. The former is the procedure to identify relevant  
 144 areas of an image whereas the latter is the procedure to describe and characterize a region of an image  
 145 (i.e. patch) calculating an histogram of the ~~gradient orientations~~ angular orientations of pixel gradients. In  
 146 order to characterize EEG signal waveforms, this work proposes an alternative to the SIFT Descriptor,  
 147 the Histogram of Gradient Orientations ~~algorithm. We stripped away all the details which are not used and~~  
 148 ~~refined the procedure to enable the effective capturing of the waveform shape. (HIST) algorithm.~~

149 For each generated image  $I^{(l,c)}$ , a keypoint  $p_k$  is placed on a pixel  $(x_{p_k}, y_{p_k})$  over the image plot and a  
 150 window around the keypoint is considered. ~~A:~~ a local image patch ~~of size~~. Its size is  $X_p \times X_p$  pixels ~~and~~  
 151 is constructed by dividing the window in 16 blocks of size  $3s$  each one, where  $s$  is the scale of the local  
 152 patch and it is an input parameter of the algorithm. It is arranged in a  $4 \times 4$  grid and the pixel  $p_k$  is the  
 153 patch center, thus  $X_p = 12s$  pixels.

154 A local representation of the signal shape within the patch can be described by obtaining the gradient  
 155 orientations on each of the 16 blocks  $B_{i,j}$  with  $0 \leq i, j \leq 3$  and creating a histogram of gradients. In  
 156 order to calculate the histogram, the interval  $[0, 360]$  of possible angles is divided in 8 bins, each one of 45  
 157 degrees.

158 Hence, for each spatial bin  $0 \leq i, j \leq 3$ , corresponding to the indexes of each block  $B_{i,j}$ , the orientations  
 159 are accumulated in a 3-dimensional histogram  $h$  through the following equation:

$$h(\theta, i, j) = 3s \sum_{\mathbf{p} \in I^{(l,c)}} w_{ang}(\angle J(\mathbf{p}) - \theta) w_{ij} \left( \frac{\mathbf{p} - \mathbf{p}_k}{3s} \right) \|J(\mathbf{p})\| \quad (5)$$

160 where  $\mathbf{p}$  is a pixel from the image  $I^{(l,c)}$ ,  $\theta$  is the angle bin with  $\theta \in \{0, 45, 90, 135, 180, 225, 270, 315\}$ ,  
 161  $\|J(\mathbf{p})\|$  is the euclidean norm of the gradient vector in the pixel  $\mathbf{p}$  and it is computed using finite differences  
 162 and  $\angle J(\mathbf{p})$  is the angle of the gradient vector.

163 The contribution of each gradient vector to the histogram calculated by Equation 5 is balanced by a  
 164 trilinear interpolation. The scalar  $w_{ang}(\cdot)$  and vector  $w_{ij}(\cdot)$  functions are linear interpolations used by ?  
 165 and ? to provide a weighting contribution to the eight adjacent bins in the tridimensional histogram. They  
 166 are calculated as

$$w_{ij}(\mathbf{v}) = w(v_x - x_i)w(v_y - y_j) \quad (6)$$

167 with  $0 \leq i, j \leq 3$  and

$$w_{\text{ang}}(\alpha) = \sum_{r=-1}^1 w\left(\frac{8\alpha}{2\pi} + 8r\right) \quad (7)$$

168 where  $x_i$  and  $y_i$  are the spatial bin centers located in  $x_i, y_i \in \{-\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}\}$  and the interpolating function  
 169  $w(\cdot)$  is defined as  $w(z) = \max(0, 1 - |z|)$ . The function parameter  $\mathbf{v} = (v_x, v_y)$  is a vector variable and  $\alpha$   
 170 a scalar variable. Vector  $\mathbf{v}$  holds pixel coordinates  $(v_x, v_y)$  normalized between  $-2$  and  $2$  and combined  
 171 with the function  $w(z)$  it produces zero for every combination of  $(i, j)$  except for the  $4$  adjacent spatial  
 172 bins. On the other hand,  $r$  is an integer that can vary freely in the set  $\{-1, 0, 1\}$  and  $\alpha$  is the difference  
 173 between the gradient orientation angle and the angle bin center in radians. By following this procedure,  
 174 summands on Equation 7 are nullified except for the  $2$  adjacent angular bins.

175 These binning functions conform the trilinear interpolation that has a combined effect of sharing the  
 176 contribution of each oriented gradient between their eight adjacent bins in a tridimensional cube in the  
 177 histogram space, and zero everywhere else (?).

178 ~~Lastly, the~~ The fixed value of  $3$  is a magnification factor which corresponds to the number of pixels per  
 179 each block when  $s = 1$ . As the patch has  $16$  blocks and  $8$  bin angles are considered, for each location  $l$  and  
 180 channel  $c$  a feature called *descriptor*  $\mathbf{d}^{(l,c)}$  of  $128$  dimension is obtained. The main differences between  
 181 this implementation and the standard SIFT Descriptor are described in the Appendix on Section 5.

182 Figure 3 shows an example of a patch and a scheme of the histogram computation. In (A) a plot of the  
 183 signal and the patch centered around the keypoint is shown. In (B) the possible orientations on each patch  
 184 are illustrated. Only the upper-left four blocks are visible. The first eight orientations of the first block, are  
 185 labeled from  $1$  to  $8$  clockwise. The orientations of the second block  $B_{1,2}$  are labeled from  $9$  to  $16$ . This  
 186 labeling continues left-to-right, up-down until the eight orientations for all the sixteen blocks are assigned.  
 187 They form the corresponding descriptor  $\mathbf{d}$  of  $128$  coordinates. Finally, in (C) an enlarged image plot is  
 188 shown where the oriented gradient vector for each pixel can be seen.

#### 189 2.1.4 Speller Matrix letter Identification

##### 190 2.1.4.1 P300 ERP Extraction

191 Segments corresponding to row flickering are labeled  $1$ - $6$ , whereas those corresponding to column  
 192 flickering are labeled  $7$ - $12$ . The extraction process has the following steps:

- 193 • **Step A:** First highlight rows and columns from the matrix in a random permutation order and obtain  
 194 the Ensemble Average as detailed in steps 1, 2 and 3 in Section 2.1.1.
- 195 • **Step B:** Plot the signals  $\tilde{x}^l(n, c)$ ,  $1 \leq n \leq n_{\max}$ ,  $1 \leq c \leq C$ , according Section 2.1.2 in order to  
 196 generate the images  $I^{(l,c)}$  for rows and columns  $1 \leq l \leq 12$ .
- 197 • **Step C:** Obtain the descriptors  $\mathbf{d}^{(l,c)}$  for rows and columns from  $I^{(l,c)}$  in accordance to the method  
 198 described in Section 2.1.3.

---

**199 2.1.4.2 Calibration**

200 A trial, as defined by the BCI2000 platform (?), is every attempt to select just one letter from the speller.  
 201 A set of trials is used for calibration and once the calibration is complete it can be used to identify new  
 202 letters from new trials.

203 During the calibration phase, two descriptors  $\mathbf{d}^{(l,c)}$  are extracted for each available channel, corresponding  
 204 to the locations  $l$  of a selection of one previously instructed letter from the set of calibration trials. These  
 205 descriptors are the P300 templates, grouped together in a template set called  $T^c$ . The set is constructed  
 206 using the steps described in Section 2.1.1 and the steps A, B and C of the P300 ERP extraction process.

207 Additionally, the best performing channel,  $bpc$  is identified based on the the channel where the best  
 208 Character Recognition Rate is obtained.

**209 2.1.4.3 Letter identification**

210 In order to identify the selected letter, the template set  $T^{bpc}$  is used as a database. Thus, new unclassified  
 211 descriptors  $\mathbf{q}^{(l,bpc)}$  are computed and they are compared against the descriptors belonging to the calibration  
 212 template set  $T^{bpc}$ .

213 The Naive Bayes Nearest Neighbor (k-NBNN) (?) is a discriminative (?) semi-supervised classification  
 214 algorithm that allows the categorization of an image to one class by comparing the set of extracted  
 215 descriptors to those which are more similar from template dictionaries. This work proposes an adapted  
 216 version to obtain a unary classification scheme to identify the selected letter in the P300-Based BCI Speller,  
 217 based on the features provided by the calculated descriptors.

218 • **Step D:** Match to the calibration template  $T^{bpc}$  by computing

$$r\hat{o}w = \arg \min_{l \in \{1, \dots, 6\}} \sum_{h=1}^k \left\| \mathbf{q}^{(l,bpc)} - \mathbf{d}_h^{(bpc)} \right\|^2 \quad (8)$$

219 and

$$\hat{c}ol = \arg \min_{l \in \{7, \dots, 12\}} \sum_{h=1}^k \left\| \mathbf{q}^{(l,bpc)} - \mathbf{d}_h^{(bpc)} \right\|^2 \quad (9)$$

220 with  $\mathbf{d}_h^{(bpc)}$  belonging to the set  $N_T(\mathbf{q}^{(l,bpc)})$ , which is defined, for the best performing channel, as  
 221  $N_T(\mathbf{q}^{(l,bpc)}) = \{\mathbf{d}_h^{(bpc)} \in T^{bpc} / \mathbf{d}_h^{(bpc)} \text{ is the } k\text{-nearest neighbor of } \mathbf{q}^{(l,bpc)}\}$ . This set is obtained by  
 222 sorting all the elements in  $T^{bpc}$  based on distances between them and  $\mathbf{q}^{(l,bpc)}$ , choosing the  $k$  with  
 223 smaller values, with  $k$  a parameter of the algorithm.

224 By computing the aforementioned equations, the letter of the matrix can be determined from the intersection  
 225 of the row  $r\hat{o}w$  and column  $\hat{c}ol$ . Figure 2 shows a scheme of this process.

**226 2.2 Experimental Protocol**

227 To verify the validity of the proposed framework and method, the public dataset 008-2014 (?) published  
 228 on the BNCI-Horizon website (?) by IRCCS Fondazione Santa Lucia, is used. Additionally, an own dataset  
 229 with the same experimental conditions is generated. Both of them are utilized to perform an offline BCI  
 230 Simulation to decode the spelled words from the provided signals.

231 The algorithm is implemented on MATLAB V2017a (Mathworks Inc., Natick, MA, USA). The  
232 algorithm described in 2.1.3 is implemented on a modified version of the VLFeat (?) Computer Vision  
233 library. Furthermore, in order to enhance the impact of this paper and for a sake of reproducibility,  
234 the code of the entire algorithm, including the modified VLFeat library, has been made available at:  
235 <https://bitbucket.org/itba/hist>.

236 In the following sections the characteristics of the datasets and parameters of the identification algorithm  
237 are described.

### 238 2.2.1 P300 ALS Public Dataset

239 The experimental protocol used to generate this dataset is explained in (?) but can be summarized as  
240 follows: 8 subjects with confirmed diagnoses but on different stages of ALS disease, were recruited and  
241 accepted to perform the experiments. The Visual P300 detection task designed for this experiment consisted  
242 of spelling 7 words of 5 letters each, using the traditional P300 Speller Matrix (?). The flashing of rows  
243 and columns provide the deviant stimulus required to elicit this physiological response. The first 3 words  
244 are used for calibration and the remaining 4 words, for testing with visual feedback. A trial is every attempt  
245 to select a letter from the speller. It is composed of signal segments corresponding to  $k_a = 10$  repetitions  
246 of flashes of 6 rows and  $k_a = 10$  repetitions of flashes of 6 columns of the matrix, yielding 120 repetitions.  
247 Flashing of a row or a column is performed for 0.125s, following by a resting period (i.e. inter-stimulus  
248 interval) of the same length. After 120 repetitions an inter-trial pause is included before resuming with the  
249 following letter.

250 The recorded dataset was sampled at 256 Hz and it consisted of a scalp multichannel EEG signal for  
251 electrode channels Fz, Cz, Pz, Oz, P3, P4, PO7 and PO8, identified according to the 10-20 International  
252 System, for each one of the 8 subjects. The recording device was a research-oriented digital EEG device  
253 (g.Mobilab, g.Tec, Austria) and the data acquisition and stimuli delivery were handled by the BCI2000  
254 open source software (?).

255 In order to assess and verify the identification of the P300 response, subjects are instructed to perform a  
256 copy-spelling task. They have to fix their attention to successive letters for copying a previously determined  
257 set of words, in contrast to a free-running operation of the speller where each user decides on its own what  
258 letter to choose.

### 259 2.2.2 P300 for healthy subjects

260 We replicate the same experiment on healthy subjects using a wireless digital EEG device (g.Nautilus,  
261 g.Tec, Austria). The experimental conditions are the same as those used for the previous dataset, as detailed  
262 in section 2.2.1. The produced dataset is available in a public online repository (?).

263 Participants are recruited voluntarily and the experiment is conducted anonymously in accordance with  
264 the Declaration of Helsinki published by the World Health Organization. No monetary compensation  
265 is handed out and all participants agree and sign a written informed consent. This study is approved  
266 by the *Departamento de Investigación y Doctorado, Instituto Tecnológico de Buenos Aires (ITBA)*. All  
267 healthy subjects have normal or corrected-to-normal vision and no history of neurological disorders. The  
268 experiment is performed with 8 subjects, 6 males, 2 females, 6 right-handed, 2 left-handed, average age  
269 29.00 years, standard deviation 11.56 years, range 20-56 years.

270 EEG data is collected in a single recording session. Participants are seated in a comfortable chair, with  
 271 their vision aligned to a computer screen located one meter in front of them. The handling and processing  
 272 of the data and stimuli is conducted by the OpenVibe platform (?).

273 Gel-based active electrodes (g.LADYbird, g.Tec, Austria) are used on the same positions Fz, Cz, Pz,  
 274 Oz, P3,P4, PO7 and PO8. Reference is set to the right ear lobe and ground is preset as the AFz position.  
 275 Sampling frequency is slightly different, and is set to 250 Hz, which is the closest possible to the one used  
 276 with the other dataset.

277 2.2.3 Parameters

278 The patch size is  $X_P = 12s \times 12s$  pixels, where  $s$  is the scale of the local patch and it is an input parameter  
 279 of the algorithm. The P300 event can have a span of 400 ms and its amplitude can reach  $10\mu V$  (?). Hence  
 280 it is necessary to utilize a signal segment of size  $t_{max} = 1$  second and a size patch  $X_P$  that could capture  
 281 an entire transient event. With this purpose in consideration, the  $s$  value election is essential.

282 We propose the Equations 10 and 11 to compute the scale value in horizontal and vertical directions,  
 283 respectively.

$$s_x = \frac{\gamma \lambda F_s}{12} \quad (10)$$

$$s_y = \frac{\gamma \Delta\mu V}{12} \quad (11)$$

284 where  $\lambda$  is the length in seconds covered by the patch,  $F_s$  is the sampling frequency of the EEG signal  
 285 (downsampled to 16 Hz) and  $\Delta\mu V$  corresponds to the amplitude in microvolts that can be covered by the  
 286 height of the patch. The geometric structure of the patch ~~forces a squared configuration, then is determined~~  
 287 by the waveform to be captured, thus we discerned that by using  $s = s_x = s_y = 3$  and  $\gamma = 4$ , the local  
 288 patch and the descriptor can identify events of  $9 \mu V$  of amplitude, with a span of  $\lambda = 0.56$  seconds.  
 289 This also determines that 1 pixel represents  $\frac{1}{\gamma} = \frac{1}{4}\mu V$  on the vertical direction and  $\frac{1}{F_s \gamma} = \frac{1}{64}$  seconds  
 290 on the horizontal direction. The keypoints  $p_k$  are located at  $(x_{p_k}, y_{p_k}) = (0.55F_s \gamma, z^l(c)) = (35, z^l(c))$   
 291 for the corresponding channel  $c$  and location  $l$  (see Equation 4). In this way the whole transient event is  
 292 captured. Figure 4 shows a patch of a signal plot covering the complete amplitude (vertical direction) and  
 293 the complete span of the signal event (horizontal direction).

294 The number of channels  $C$  is equal to 8 for both datasets, and the number of intensification sequences  $k_a$   
 295 is fixed to 10. The parameter  $k$  used to construct the set  $N_T(\mathbf{q}^{(l,c)})$  is assigned to  $k = 7$ , which was found  
 296 empirically to achieve better results. In addition, the norm used on Equations 8 and 9 is the cosine norm,  
 297 and descriptors are normalized to  $[-1, 1]$ .

298 Lastly, in order to assess the validity of the ~~Histogram of Gradient Orientations (HIST)~~ HIST method,  
 299 the character recognition rate for both datasets is evaluated replicating the methodology proposed by  
 300 the ALS dataset's publisher, since authors ? did not report the Character Recognition Rate obtained  
 301 for this dataset. Frequency filtering, data segmentation and artifact rejection is conducted according to  
 302 Section 2.1.1 yielding 16 x 8 samples per epoch. A multichannel feature consists of time points vector (?),  
 303 formed by concatenating all the channels (?). A single-channel variant consists of using time points from a  
 304 single electrode and performing the analysis on a channel-by-channel basis. Three classification schemes  
 305 are considered as well. A multichannel version of the Stepwise Linear Discriminant Analysis (SWLDA)  
 306 classification algorithm. SWLDA is the methodology proposed by the ALS dataset's publisher. Additionally,

307 a single-channel and a multichannel variant of a linear kernel Support Vector Machine (SVM) (?) classifier  
308 are utilized. SVM has been successfully used in several BCI Competitions (?).

### 3 RESULTS

309 Table 1 shows the results of applying the HIST algorithm to the subjects of the public dataset of ALS  
310 patients. The percentage of correctly spelled letters is calculated while performing an offline BCI Simulation.  
311 From the seven words for each subject, the first three are used for calibration, and the remaining four are  
312 used for testing. The best performing channel  $bpc$  is informed as well. The target ratio is 1 : 36; hence  
313 theoretical chance level is 2.8%. It can be observed that the best performance of the letter identification  
314 method is reached in a dissimilar channel depending on the subject being studied. Table 1 and 2 show for  
315 comparison the obtained performance rates using single-channel signals with the SVM classifier. The best  
316 performing channel, where the best letter identification rate was achieved, is also depicted.

317 The Information Transfer Rate (ITR), or Bit Transfer Rate (BTR), in the case of reactive BCIs (?)  
318 depends on the amount of signal averaging required to transmit a valid and robust selection. Figure 5 shows  
319 the performance curves for varying intensification sequences for the subjects included in the dataset of  
320 ALS patients. It can be noticed that the percentage of correctly identified letters depends on the number  
321 of intensification sequences that are used to obtain the averaged signal. Moreover, when the number of  
322 intensification sequences tend to 1, which corresponds to single-intensification character recognition, the  
323 performance is reduced. As mentioned before, the SNR of the P300 obtained from only one segment of the  
324 intensification sequence is very low and the shape of its P300 component is not very well defined.

325 In Table 2 the results obtained for 8 healthy subjects are shown. It can be observed that the performance  
326 is above chance level. It is verified that HIST method has an improved performance at letter identification  
327 than SVM that process the signals on a channel by channel strategy (Wilcoxon signed-rank test,  $p = 0.004$   
328 for both datasets).

329 Tables 3 and 4 are presented in order to compare the performance of the HIST method versus multichannel  
330 SWLDA and SVM classification algorithms for both datasets. It is verified for the dataset of ALS patients  
331 that it has similar performance against other methods like SWLDA or SVM, which use a multichannel  
332 feature (Quade test with  $p = 0.55$ ) whereas for the dataset of healthy subjects significant differences are  
333 found (Quade test with  $p = 0.02$ ) where only the HIST method achieves a different performance than SVM  
334 (with multiple comparisons, significant difference of level 0.05).

335 The P300 ERP consists of two overlapping components: the P3a and P3b, the former with frontocentral  
336 distribution while the later stronger on centroparietal region (?). Hence, the standard practice is to find  
337 the stronger response on the central channel Cz (?). However, ? show that the response may also arise in  
338 occipital regions. We found that by analyzing only the waveforms, occipital channels PO8 and PO7 show  
339 higher performances for some subjects.

340 As subjects have varying *latencies* and *amplitudes* of their P300 components, they also have a varying  
341 stability of the *shape* of the generated ERP (?). Figure 6 shows 10 sample P300 templates patches for  
342 patients 8 and 3 from the dataset of ALS patients. It can be discerned that in coincidence with the  
343 performance results, the P300 signature is more clear and consistent for subject 8 (A) while for subject 3  
344 (B) the characteristic pattern is more difficult to perceive.

345 Additionally, the stability of the P300 component waveform has been extensively studied in patients  
346 with ALS (?????) where it was found that these patients have a stable P300 component, which were also

347 sustained across different sessions. In line with these results we do not find evidence of a difference in  
348 terms of the performance obtained by analyzing the waveforms (HIST) for the group of patients with ALS  
349 and the healthy group of volunteers (Mann-Whitney U Test,  $p = 0.46$ ). Particularly, the best performance  
350 is obtained for a subject from the ALS dataset for which, based on visual observation, the shape of they  
351 P300 component is consistently identified.

352 It is important to remark that when applied to binary images obtained from signal plots, the feature  
353 extraction method described in Section 2.1.3 generates sparse descriptors. Under this subspace we found  
354 that using the cosine metric yielded a significant performance improvement. On the other hand, the unary  
355 classification scheme based on the NBNN algorithm proved very beneficial for the P300 Speller Matrix.  
356 This is due to the fact that this approach solves the unbalance dataset problem which is inherent to the  
357 oddball paradigm (?).

## 4 DISCUSSION

358 Among other applications of Brain Computer Interfaces, the goal of the discipline is to provide  
359 communication assistance to people affected by neuro-degenerative diseases, who are the most likely  
360 population to benefit from BCI systems and EEG processing and analysis.

361 In this work, a method to extract an objective metric from the waveform of the plots of EEG signals is  
362 presented. Its usage to implement a valid P300-Based BCI Speller application is expounded. Additionally,  
363 its validity is evaluated using a public dataset of ALS patients and an own dataset of healthy subjects.

364 It was verified that this method has an improved performance at letter identification than other methods  
365 that process the signals on a channel by channel strategy, and it even has a comparable performance against  
366 other methods like SWLDA or SVM, which uses a multichannel feature. Furthermore, this method has the  
367 advantage that shapes of waveforms can be analyzed in an objective way. We observed that the shape of  
368 the P300 component is more stable in occipital channels, where the performance for identifying letters  
369 is higher. We additionally verified that ALS P300 signatures are stable in comparison to those of healthy  
370 subjects.

371 We believe that the use of descriptors based on histogram of gradient orientation, presented in this work,  
372 can also be utilized for deriving a shape metric in the space of the P300 signals which can complement  
373 other metrics based on time-domain as those defined by ?. It is important to notice that the analysis of  
374 waveform shapes is usually performed in a qualitative approach based on visual inspection (?), and a  
375 complementary methodology which offer a quantitative metric will be beneficial to these routinely analysis  
376 of the waveform of ERPs.

377 The goal of this work is to answer the question if a P300 component could be solely determined by  
378 inspecting automatically their waveforms. We conclude affirmatively, though two very important issues  
379 still remain:

380 First, the stability of the P300 in terms of its shape is crucial: the averaging procedure, montages, the  
381 signal to noise ratio and spatial filters all of them are non-physiological factors that affect the stability of  
382 the shape of the P300 ERP. We tested a preliminary approach to assess if the morphological shape of the  
383 P300 of the averaged signal can be stabilized by applying different alignments of the stacked segments (see  
384 Figure 2) and we verified that there is a better performance when a correct segment alignment is applied.  
385 We applied Dynamic Time Warping (DTW) (?) to automate the alignment procedure but we were unable

386 to find a substantial improvement. Further work to study the stability of the shape of the P300 signature  
387 component needs to be addressed.

388 The second problem is the amplitude variation of the P300. We propose a solution by standardizing the  
389 signal, shown in Equation 2. It has the effect of normalizing the peak-to-peak amplitude, moderating its  
390 variation. It has also the advantage of reducing noise that was not reduced by the averaging procedure. It is  
391 important to remark that the averaged signal variance depends on the number of segments used to compute  
392 it (?). The standardizing process converts the signal to unit signal variance which makes it independent of  
393 the number  $k_a$  of signals averaged. Although this is initially an advantageous approach, the standardizing  
394 process reduces the amplitude of any significant P300 complex diminishing its automatic interpretation  
395 capability.

396 To further extend the capabilities of this method, it would be desirable to implement a multichannel  
397 version. The straightforward extension of concatenating the obtained descriptors results in high dimensional  
398 feature vector, while other variants that merge descriptors per channel may diminish the mutual information  
399 between different channels. Hitherto variants using color versions of SIFT (?), where different color bands  
400 are mapped to electrode channels, have been explored without substantial success.

401 In our opinion, the best benefit of the presented method is that a closer collaboration of the field of  
402 BCI with physicians can be fostered (?), since this procedure intent to imitate human visual observation.  
403 Automatic classification of patterns in EEG that are specifically identified by their shapes like K-Complex,  
404 Vertex Waves, Positive Occipital Sharp Transient (?) are a prospect future work to be considered. We are  
405 currently working in unpublished material analyzing K-Complex components that could eventually provide  
406 assistance to physicians to locate these EEG patterns, specially in long recording periods, frequent in sleep  
407 research (?). Additionally, it can be used for artifact removal which is performed on many occasions by  
408 visually inspecting signals. This is due to the fact that the descriptors are a direct representation of the shape  
409 of signal waveforms. In line with these applications, it can be used to build a database (?) of quantitative  
410 representations of waveforms and improve atlases (?), which are currently based on qualitative descriptions  
411 of signal shapes.

## CONFLICT OF INTEREST STATEMENT

412 The authors declare that the research was conducted in the absence of any commercial or financial  
413 relationships that could be construed as a potential conflict of interest.

## AUTHOR CONTRIBUTIONS

414 This work is part of the PhD thesis of RR which is directed by JS and codirected by AV.

## FUNDING

415 This project was supported by the ITBACyT-15 funding program issued by ITBA University from Buenos  
416 Aires, Argentina.

## 5 APPENDIX

417 [This section describes the differences between the HIST algorithm proposed in this work and the SIFT](#)  
418 [Descriptor \(?\).](#)

419 The two most important modifications are:

- 420 • SIFT Detector and custom frame: The SIFT Detector provides the keypoint localization information  
421 in the standard SIFT method. The keypoint localization information is stored in a *frame* data structure  
422 which is composed of the keypoint center location ( $x_{kp}, y_{kp}$ ), patch scale  $s$  and patch orientation  $\phi$ :  
423 ( $x_{kp}, y_{kp}, s, \phi$ ). In the HIST proposal the keypoint location and patch parameters are directly specified  
424 over the plot image in order to detect the signal waveform (see Section 2.2.3) and the SIFT Detector  
425 is not used.
- 426 • Patch scale: Whereas in the standard SIFT implementation the patch is a squared region and there is  
427 only one SIFT scale parameter, in HIST a different scale parameter can be assigned to the horizontal  
428 and vertical axis. This is a very important modification because otherwise signal plots which extend  
429 only on the horizontal direction of the plot image could not be entirely covered. By using a rectangular  
430 patch, there isn't any constraint on its size and it can be adjusted by neurophysiological priors to map  
431 any expected waveform based on Equations 10 and 11.

432 Additionally, other changes are implemented where specific steps of the SIFT Descriptor were not found  
433 to be useful to characterize signal waveforms:

- 434 • Patch orientation: We verified experimentally that the patch orientation  $\phi$  does not provide any extra  
435 utility for the extraction of characteristic waveforms from plots. Hence, this patch orientation is fixed  
436 to zero (vertical, pointing upwards in Figure 4).
- 437 • Rotations: SIFT was designed to allow affine invariance, i.e. to be robust to rotations and scale  
438 modifications of patterns in images. It was not found, so far, of any utility to rotate the patch to  
439 capture the signal waveform.
- 440 • Octave selection: A gradient image is used to obtain the oriented gradients and calculate the histogram  
441 of gradient orientations. In SIFT, these gradient images are downsampled and smoothed by a  
442 Gaussian filter. The SIFT Descriptor calls *octave* to each downsampling level (??). The standard  
443 SIFT Descriptor estimates the octave to use on the gradient image based on the image size and patch  
444 parameters. The HIST method uses only the zero octave which means that the gradient image has the  
445 same size as the original image, without any downsampling operation.
- 446 • Gradient image smoothing: Additionally, the SIFT Descriptor performs an initial smoothing operation  
447 by applying a Gaussian filter on the gradient image regardless of the octave. In the HIST method, this  
448 operation is not implemented.
- 449 • Descriptor Gaussian weighting: On the standard SIFT Descriptor, a Gaussian weighting operation  
450 is performed on the calculated SIFT descriptor to increase the importance of gradients from pixels  
451 closer to the center of the patch. For the HIST method, this is found to be in detriment of the waveform  
452 characterization and is not used.
- 453 • SIFT descriptor codification: The SIFT descriptor  $d$  is a 128-dimension feature vector, as described  
454 in Section 2.1.3. Histogram values are floating point numbers, all positive, and they are accumulated  
455 on each coordinate of this vector. Once all the gradients are summarized, the following operations are  
456 performed:
  - 457 • The descriptor is  $\ell$ -2 normalized (i.e all the values are divided by the euclidean norm of the  
458 descriptor).
  - 459 • Each value is clamped to 0.2. This means that any value above 0.2 is set to 0.2.
  - 460 • The descriptor is  $\ell$ -2 re-normalized again (?).

461     This generates a 128-vector of floating point numbers, between  $[0, 1]$ . In the HIST implementation,  
462     these values are rescaled to  $[-1, 1]$  in order to use the cosine distance (?) on Equations 9 and 8.  
463     Finally, output values are cast to floating point numbers (i.e. floats), yielding an effective 128-vector  
464     of floats between  $[-1, 1]$ .

**Table 1.** Character recognition rates for the public dataset of ALS patients using the Histogram of Gradient (HIST) calculated from single-channel plots. Performance rates using single-channel signals with the SVM classifier are shown for comparison. The best performing channel  $bpc$  for each method is visualized

Participant	$bpc$	HIST	$bpc$	Single Channel SVM
1	Cz	35%	Cz	15%
2	Fz	85%	PO8	25%
3	Cz	25%	Fz	5%
4	PO8	55%	Oz	5%
5	PO7	40%	P3	25%
6	PO7	60%	PO8	20%
7	PO8	80%	Fz	30%
8	PO7	95%	PO7	85%

**Table 2.** Character recognition rates for the own dataset of healthy subjects using the Histogram of Gradient (HIST) calculated from single-channel plots. Performance rates using single-channel signals with the SVM classifier are shown for comparison. The best performing channel  $bpc$  for each method is visualized.

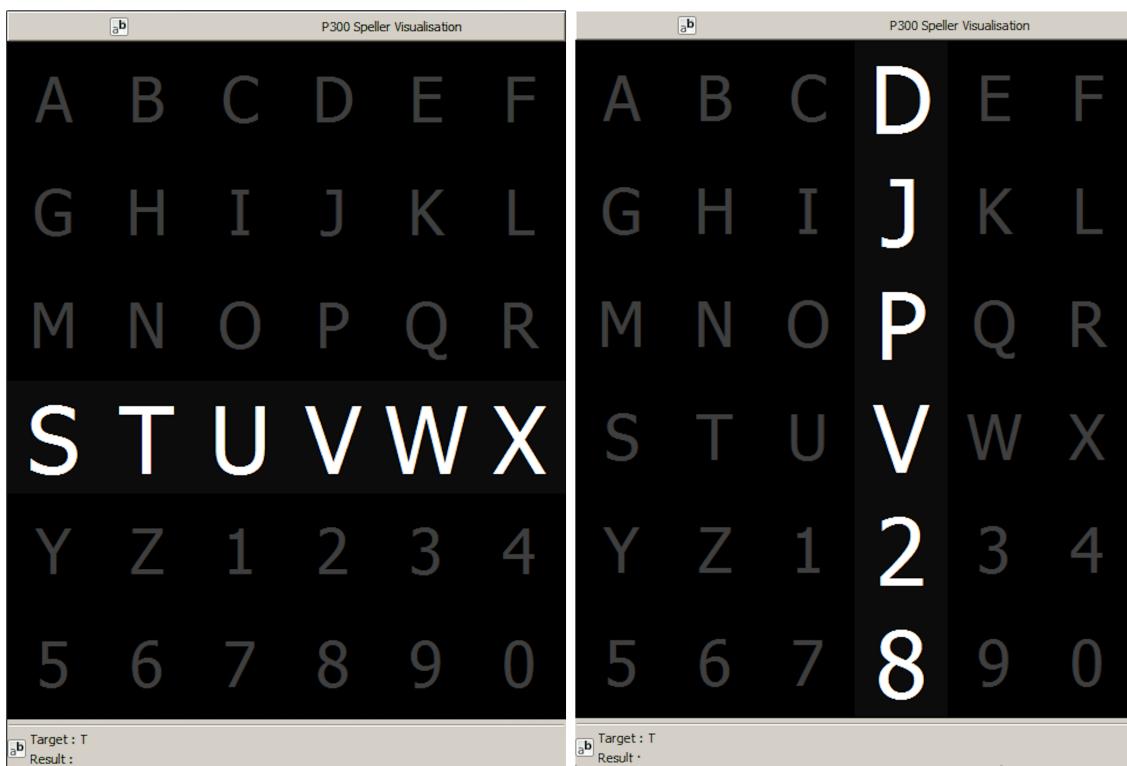
Participant	$bpc$	HIST	$bpc$	Single Channel SVM
1	Oz	40%	Cz	10%
2	PO7	30%	Cz	5%
3	P4	40%	P3	10%
4	P4	45%	P4	35%
5	P4	60%	P3	10%
6	Pz	50%	P4	25%
7	PO7	70%	P3	30%
8	P4	50%	PO7	10%

**Table 3.** Character recognition rates and the best performing channel  $bpc$  for the public dataset of ALS patients using the Histogram of Gradient (HIST) (repeated here for comparison purposes). Performance rates obtained by SWLDA and SVM classification algorithms with a multichannel concatenated feature.

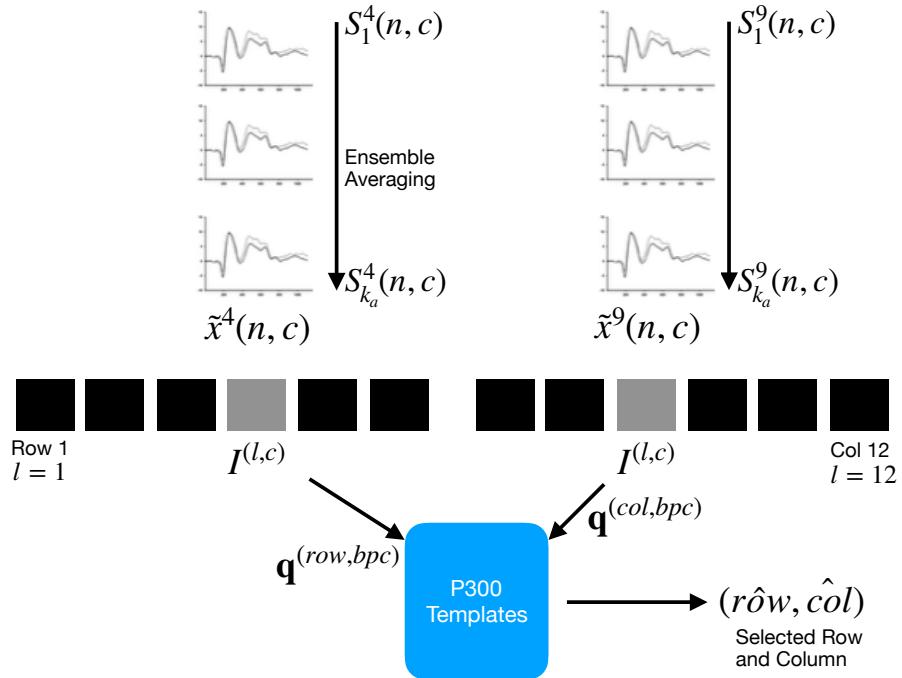
Participant	$bpc$ for HIST	HIST	Multichannel SWLDA	Multichannel SVM
1	Cz	35%	45%	40%
2	Fz	85%	30%	50%
3	Cz	25%	65%	55%
4	PO8	55%	40%	50%
5	PO7	40%	35%	45%
6	PO7	60%	35%	70%
7	PO8	80%	60%	35%
8	PO7	95%	90%	95%

**Table 4.** Character recognition rates and the best performing channel *bpc* for the own dataset of healthy subjects using the Histogram of Gradient (HIST) (repeated here for comparison purposes). Performance rates obtained by SWLDA and SVM classification algorithms with a multichannel concatenated feature.

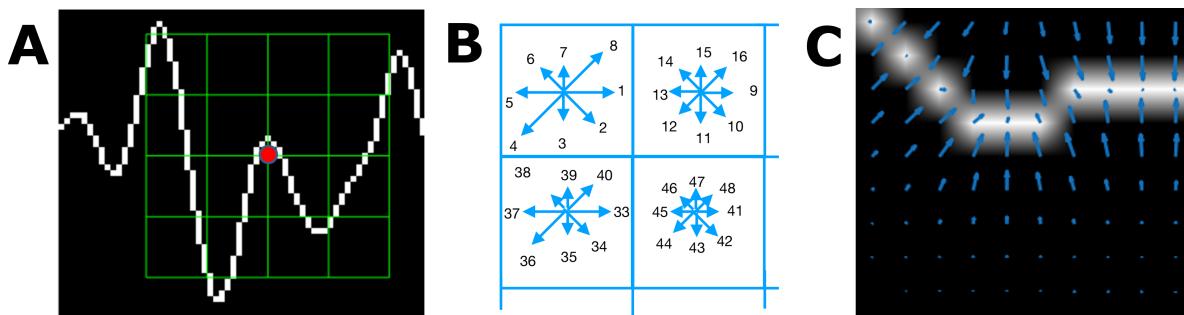
Participant	<i>bpc</i> for HIST	HIST	Multichannel SWLDA	Multichannel SVM
1	Oz	40%	65%	40%
2	PO7	30%	15%	10%
3	P4	40%	50%	25%
4	P4	45%	40%	20%
5	P4	60%	30%	20%
6	Pz	50%	35%	30%
7	PO7	70%	25%	30%
8	P4	50%	35%	20%



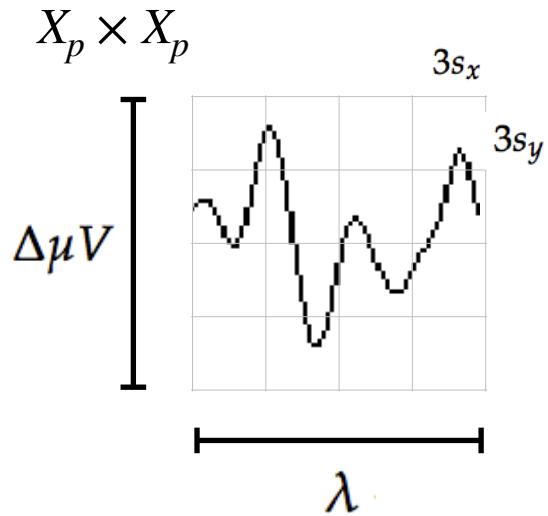
**Figure 1.** Example of the  $6 \times 6$  Speller Matrix used in the study obtained from the OpenVibe software. Rows and columns flash in random permutations.



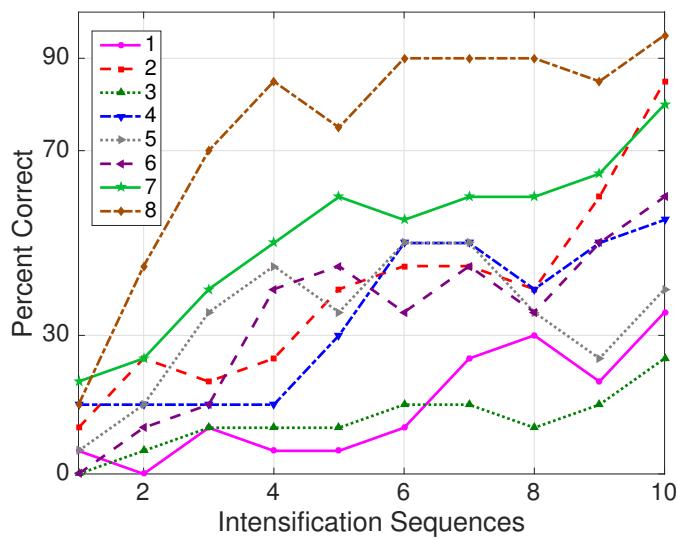
**Figure 2.** For each column and row, an averaged, standardized and scaled signal  $\tilde{x}^l(n, c)$  is obtained from the segments  $S_i^l$  corresponding to the  $k_a$  intensification sequences with  $1 \leq i \leq k_a$  and location  $l$  varying between 1 and 12. From the averaged signal, the image  $I^{(l,c)}$  of the signal plot is generated and each descriptor is computed. By comparing each descriptor against the set of templates, the P300 ERP can be detected, and finally the desired letter from the matrix can be inferred.



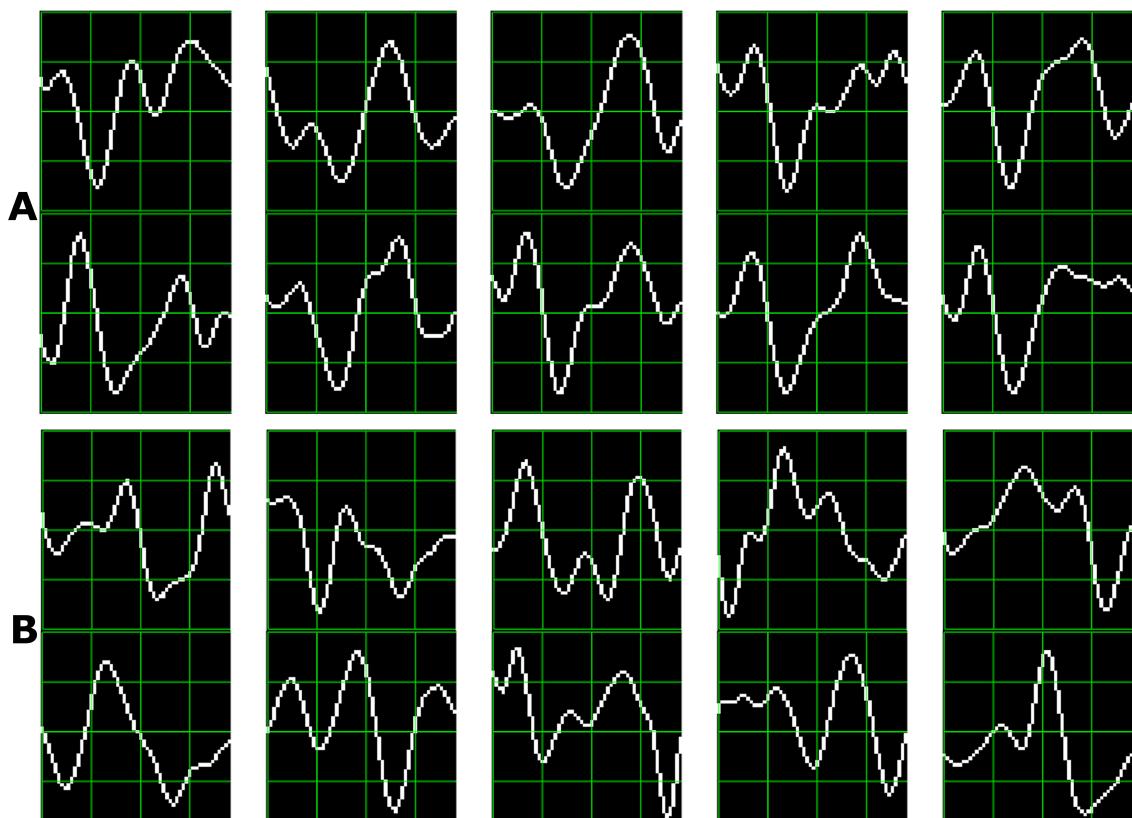
**Figure 3.** (A) Example of a plot of the signal, a keypoint and the corresponding patch. (B) A scheme of the orientation's histogram computation. Only the upper-left four blocks are visible. The first eight orientations of the first block, are labeled from 1 to 8 clockwise. The orientation of the second block  $B_{1,2}$  is labeled from 9 to 16. This labeling continues left-to-right, up-down until the eight orientations for all the sixteen blocks are assigned. They form the corresponding descriptor of 128 coordinates. The length of each arrow represent represents the value of the histogram on each direction for each block. (C) Vector field of oriented gradients. Each pixel is assigned an orientation and magnitude calculated using finite differences.



**Figure 4.** The scale of local patch is selected in order to capture the whole transient event. The size of the patch is  $X_p \times X_p$  pixels. The vertical size consists of 4 blocks of size  $3s_y$  pixels which is high enough as to contain the signal  $\Delta\mu V$ , the peak-to-peak amplitude of the transient event. The horizontal size includes 4 blocks of  $3s_x$  and covers the entire duration in seconds of the transient signal event,  $\lambda$ .



**Figure 5.** Performance curves for the eight subjects included in the dataset of ALS patients. Three out of eight subjects achieved the necessary performance to implement a valid P300 speller.



**Figure 6.** Ten sample P300 template patches for subjects 8 (A) and 3 (B) of the ALS Dataset. Downward deflection is positive polarity.