Histogram of Gradient Orientations of Signal Plots Applied to P300 Detection

The analysis of Electroencephalographic (EEG) signals is of ulterior importance to aid in the diagnosis of mental disease and to increase our understanding of the brain. Traditionally, clinical EEG has been analyzed in terms of temporal waveforms, looking at rhythms in spontaneous activity, subjectively identifying troughs and peaks in Event-Related Potentials (ERP), or by studying graphoelements in pathological sleep stages. Additionally, the discipline of Brain Computer Interfaces (BCI) requires new

methods to decode patterns from non-invasive EEG signals. This field is developing alternative communication pathways to transmit volitional information from the Central Nervous System. The technology could potentially enhance the quality of life of patients affected by neurodegenerative disorders and other mental illness. This work mimics what electroencephalographers have been doing clinically, visually inspecting, and categorizing phenomena within the EEG by the extraction of features from images of signal plots. These features are constructed based on the calculation of histograms of oriented gradients from pixels around the signal plot. It aims to provide a new objective framework to analyze, characterize and classify EEG signal waveforms. The feasibility of the method is outlined by detecting the P300, an ERP elicited by the oddball paradigm of rare events, and implementing an offline P300-based BCI Speller. The validity of the proposal is shown by offline processing a public dataset of Amyotrophic Lateral Sclerosis (ALS) patients and an own dataset of healthy subjects.

*Histograma de orientaciones de gradiente de gráficos de señal aplicados a la detección P300*

*El análisis de las señales electroencefalográficas (EEG) tiene una importancia ulterior para ayudar en el diagnóstico de enfermedades mentales y aumentar nuestra comprensión del cerebro. Tradicionalmente, el EEG clínico se ha analizado en términos de formas de onda temporales, observando los ritmos en la actividad espontánea, identificando subjetivamente valles y picos en los potenciales relacionados con eventos (ERP), o estudiando grafoelementos en etapas patológicas del sueño. Además, la disciplina de Brain Computer Interfaces (BCI) requiere nuevos métodos para decodificar patrones a partir de señales EEG no invasivas. Este campo está desarrollando vías de comunicación alternativas para transmitir información volitiva desde el Sistema Nervioso Central. La tecnología podría mejorar potencialmente la calidad de vida de los pacientes afectados por trastornos neurodegenerativos y otras enfermedades mentales. Este trabajo imita lo que los electroencefalógrafos han estado haciendo clínicamente, inspeccionando visualmente y categorizando fenómenos dentro del EEG mediante la extracción de características de imágenes de gráficos de señales. Estas características se construyen en base al cálculo de histogramas de gradientes orientados a partir de píxeles alrededor del gráfico de señal. Su objetivo es proporcionar un nuevo marco objetivo para analizar, caracterizar y clasificar las formas de onda de la señal EEG. La viabilidad del método se describe mediante la detección del P300, un ERP provocado por el paradigma extraño de eventos raros, y la implementación de un BCI Speller fuera de línea basado en P300. La validez de la propuesta se demuestra mediante el procesamiento fuera de línea de un conjunto de datos público de pacientes con esclerosis lateral amiotrófica (ELA) y un conjunto de datos propio de sujetos sanos.*

1. INTRODUCTION

Although recent advances in neuroimagining techniques, particularly radio-nuclear and radiological scanning methods (Schomer and Silva, 2010), have diminished the prospects of the traditional Electroencephalography, the advent and development of digitized devices has impelled for a revamping of this hundred years old technology. Their versatility, ease of use, temporal resolution, ease of development and production, and its proliferation as consumer devices, are pushing EEG to become the de-facto non invasive portable or ambulatory method to access and harness brain information (De Vos and Debener, 2014).

A key contribution to this expansion has been the field of Brain Computer Interfaces (Wolpaw and Wolpaw, 2012) which is the pursuit of the development of a new channel of communication particularly aimed to persons affected by neurodegenerative diseases.

One noteworthy aspect of this novel communication channel is the ability to transmit information from the Central Nervous System (CNS) to a computer device and from there use that information to control a wheelchair (Carlson and del R. Millan, 2013), as input to a speller application (Guger et al., 2009), in a Virtual Reality environment (Lotte et al., 2013) or as aiding tool in a rehabilitation procedure (Jure et al., 2016). The holly grail of BCI is to implement a new complete and alternative pathway to restore lost locomotion (Wolpaw andWolpaw, 2012). EEG signals are remarkably complex and have been characterized as a multichannel non-stationary stochastic process. Additionally, they have high variability between different subjects and even between different moments for the same subject, requiring adaptive and co-adaptive calibration and learning procedures (Clerc et al., 2016). Hence, this imposes an outstanding challenge that is necessary to overcome in order to extract information from raw EEG signals. BCI has gained mainstream public awareness with worldwide challenge competitions like Cybathlon (Riener and Seward, 2014; Novak et al., 2018) and even been broadcasted during the inauguration ceremony of the 2014 Soccer World Cup. New developments have overcome the out-of-the-lab high-bar and they are starting to be used in real world environments (Huggins et al., 2016; Guger et al., 2017). However, they still lack the necessary robustness, and its performance is well behind any other method of human computer interaction, including any kind of detection of residual muscular movement (Clerc et al., 2016).

A few works have explored the idea of exploiting the signal waveform to analyze the EEG signal. In Alvarado-González et al. (2016), an approach based on Slope Horizontal Chain Code is presented, whereas in Yamaguchi et al. (2009) a similar procedure was implemented based on Mathematical Morphological Analysis. The seminal work of Bandt-Pompe Permutation Entropy (Berger et al., 2017) also explores succinctly this idea as a basis to establish the time series ordinal patterns. In the article (Ramele et al., 2016), the authors introduce a method for classification of rhythmic EEG events like Visual Occipital Alpha Waves and Motor Imagery Rolandic Central μ Rhythms using the Histogram of Gradient Orientations of signal plots. Inspired in that work, we propose a novel application of the developed method to classify and describe transient events, particularly the P300 Event Related Potential. The proposed approach is based on the waveform analysis of the shape of the EEG signal. The signal is drawn on a bidimensional image plot, vector gradients of pixels around the plot are obtained, and with them, the histogram of their orientations is calculated.

This histogram is a direct representation of the waveform of the signal. The method is built by mimicking what regularly electroencephalographers have been performing for almost a century as it is described in Hartman (2005): visually inspecting raw signal plots. This paper reports a method to: (1) describe a procedure to capture the shape of a waveform of an ERP component, the P300, using histograms of gradient orientations extracted from images of signal plots, and (2) outline the way in which this procedure can be used to implement an P300-Based BCI Speller application. Its validity is verified by offline processing two datasets, one of data fromALS patients and another one from data of healthy subjects.

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

*1. INTRODUCCIÓN*

*Aunque los avances recientes en las técnicas de neuroimagen, en particular los métodos de exploración radionuclear y radiológica (Schomer y Silva, 2010), han disminuido las perspectivas de la electroencefalografía tradicional, la llegada y el desarrollo de dispositivos digitalizados ha impulsado una renovación de esta tecnología centenaria. . Su versatilidad, facilidad de uso, resolución temporal, facilidad de desarrollo y producción, y su proliferación como dispositivos de consumo, están empujando a EEG a convertirse en el método portátil o ambulatorio no invasivo de facto para acceder y aprovechar la información del cerebro (De Vos y Debener, 2014).*

*Una contribución clave a esta expansión ha sido el campo de las interfaces cerebro-ordenador (Wolpaw y Wolpaw, 2012), que persigue el desarrollo de un nuevo canal de comunicación especialmente dirigido a personas afectadas por enfermedades neurodegenerativas.*

*Un aspecto destacable de este novedoso canal de comunicación es la capacidad de transmitir información desde el Sistema Nervioso Central (SNC) a un dispositivo informático y desde allí usar esa información para controlar una silla de ruedas (Carlson y del R. Millan, 2013), como entrada para una aplicación de ortografía (Guger et al., 2009), en un entorno de Realidad Virtual (Lotte et al., 2013) o como herramienta de ayuda en un procedimiento de rehabilitación (Jure et al., 2016). El santo grial de BCI es implementar una nueva vía completa y alternativa para restaurar la locomoción perdida (Wolpaw and Wolpaw, 2012). Las señales EEG son notablemente complejas y se han caracterizado como un proceso estocástico no estacionario multicanal. Además, tienen una alta variabilidad entre diferentes sujetos e incluso entre diferentes momentos para un mismo sujeto, lo que requiere procedimientos de calibración y aprendizaje adaptativos y coadaptativos (Clerc et al., 2016). Por lo tanto, esto impone un desafío pendiente que es necesario superar para extraer información de las señales de EEG sin procesar. BCI se ha ganado la conciencia del público general con competencias de desafío mundial como Cybathlon (Riener y Seward, 2014; Novak et al., 2018) e incluso se transmitió durante la ceremonia de inauguración de la Copa Mundial de Fútbol de 2014. Los nuevos desarrollos han superado la barra alta fuera del laboratorio y están comenzando a usarse en entornos del mundo real (Huggins et al., 2016; Guger et al., 2017). Sin embargo, todavía carecen de la robustez necesaria y su rendimiento está muy por detrás de cualquier otro método de interacción humano-ordenador, incluido cualquier tipo de detección de movimiento muscular residual (Clerc et al., 2016).*

*Algunos trabajos han explorado la idea de explotar la forma de onda de la señal para analizar la señal de EEG. En Alvarado-González et al. (2016), se presenta un enfoque basado en Slope Horizontal Chain Code, mientras que en Yamaguchi et al. (2009) se implementó un procedimiento similar basado en el Análisis Morfológico Matemático. El trabajo seminal de Bandt-Pompe Permutation Entropy (Berger et al., 2017) también explora sucintamente esta idea como base para establecer los patrones ordinales de series temporales. En el artículo (Ramele et al., 2016), los autores presentan un método para la clasificación de eventos de EEG rítmicos como Visual Occipital Alpha Waves y Motor Imagery Rolandic Central μ Rhythms utilizando el histograma de orientaciones de gradiente de gráficos de señales. Inspirándonos en ese trabajo, proponemos una aplicación novedosa del método desarrollado para clasificar y describir eventos transitorios, particularmente el potencial relacionado con eventos P300. El enfoque propuesto se basa en el análisis de forma de onda de la forma de la señal EEG. La señal se dibuja sobre un gráfico de imagen bidimensional, se obtienen gradientes vectoriales de píxeles alrededor del gráfico y con ellos se calcula el histograma de sus orientaciones.*

*Este histograma es una representación directa de la forma de onda de la señal. El método se construye imitando lo que los electroencefalógrafos han estado realizando regularmente durante casi un siglo, como se describe en Hartman (2005): inspeccionar visualmente gráficos de señales sin procesar. Este documento informa sobre un método para: (1) describir un procedimiento para capturar la forma de onda de un componente ERP, el P300, utilizando histogramas de orientaciones de gradientes extraídos de imágenes de gráficos de señales, y (2) esbozar la forma en que esto El procedimiento se puede utilizar para implementar una aplicación BCI Speller basada en P300. Su validez se verifica procesando fuera de línea dos conjuntos de datos, uno de datos de pacientes con ELA y otro de datos de sujetos sanos.*

*Este artículo se desarrolla de la siguiente manera: la sección 2.1 está dedicada a explicar el método de extracción de características basado en histograma de orientaciones de gradiente de la gráfica de señal, la sección 2.1.1 muestra la canalización de preprocesamiento, la sección 2.1.2 describe la generación de imágenes de la gráfica de señal, la sección 2.1.3 presenta el procedimiento de extracción de características, mientras que la sección 2.1.4 presenta el procedimiento de identificación de letras de la matriz ortográfica. En el apartado 2.2 se expone el protocolo experimental. En la sección 3 se muestran los resultados de la aplicación de la técnica propuesta. En la última sección 4 exponemos nuestros comentarios, conclusiones y trabajo futuro.*

2. MATERIALS AND METHODS

The P300 (Farwell and Donchin, 1988; Knuth et al., 2006) is a positive deflection of the EEG signal which occurs around 300 ms after the onset of a rare and deviant stimulus that the subject is expected to attend. It is produced under the oddball paradigm (Wolpaw and Wolpaw, 2012) and it is consistent across different subjects. It has a lower amplitude (±5μV) compared to basal EEG activity, reaching a Signal to Noise Ratio (SNR) of around −15 db estimated based on the amplitude of the P300 response signal divided by the standard deviation of the background EEG activity (Hu et al., 2010). This signal can be used to implement a speller application by means of a Speller Matrix (Farwell and Donchin, 1988). This matrix is composed of 6 rows and 6 columns of numbers and letters. The subject can focus on one character of the matrix. Figure 1 shows an example of the Speller Matrix used in the OpenVibe open source software (Renard et al., 2010), where the flashes of rows and columns provide the deviant stimulus required to elicit this physiological response. Each time a row or a column that contains the desired letter flashes, the corresponding synchronized EEG signal should also contain the P300 signature and by detecting it, the selected letter can be identified.

2.1. Feature Extraction From Signal Plots

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

2.1.1. Preprocessing Pipeline

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

The 6 rows and 6 columns of the SpellerMatrix are intensified providing the visual stimulus. The number of a row or columna is a location. A sequence of 12 randomly permuted locations l conform an intensification sequence. The whole set of 12 intensifications is repeated ka times.

• Signal Enhancement: This stage consists of the enhancement of the SNR of the P300 pattern above the level of basal EEG. The pipeline starts by applying a notch filter to the raw digital signal, a 4th degree 10Hz lowpass Butterworth filter and finally a decimation with a Finite Impulse Response (FIR) filter of order 30 from the original sampling frequency down to 16 Hz (Krusienski et al., 2006).

• Artifact Removal: For every complete sequence of 12 intensifications of 6 rows and 6 columns, a basic artifact elimination procedure is implemented by removing the entire sequence when any signal deviates above/ bellow ±70μV.

• Segmentation: For each of the 12 intensifications of one intensification sequence, a segment Sl i of a window of tmax seconds of the multichannel signal is extracted, starting from the stimulus onset, corresponding to each row/columna intensification l and to the intensification sequence i. As intensifications are permuted in a random order, the segments are rearranged corresponding to row flickering, labeled 1–6, whereas those corresponding to column flickering are labeled 7–12. Two of these segments should contain the P300 ERP signature time-locked to the flashing stimulus, one for the row, and one for the column.

• Signal Averaging: The P300 ERP is deeply buried under basal EEG so the standard approach to identify it is by pointto-point averaging the time-locked stacked signal segments. Hence the values which are not related to, and not timelocked to the onset of the stimulus are canceled out (Liang and Bougrain, 2008).

This last step determines the operation of any P300 Speller. In order to obtain an improved signal 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 2.1.1 ka times, obtaining the 1 ≤ l ≤ 12 segments Sl 1(n, c), . . . , Sl ka (n, c), of the EEG signal where the variables 1 ≤ n ≤ nmax and 1 ≤ c ≤ C correspond to sample points and channel, respectively. The parameter C is the number of available EEG channels whereas nmax = Fs tmax is the segment length and Fs is the sampling frequency. The parameter ka is the number of repetitions of intensifications and it is an input parameter of the algorithm.

3. Compute the Ensemble Average by for 1 ≤ n ≤ nmax and for the channels 1 ≤ c ≤ C. This provide an averaged signal xl(n, c) for the twelve locations 1 ≤ l ≤ 12.

2.1.2. Signal Plotting

Averaged signal segments are standardized and scaled for 1 ≤ n ≤ nmax and 1 ≤ c ≤ C by where γ > 0 is an input parameter of the algorithm and it is related to the image scale. In addition, xl(n, c) is the point-topoint averaged multichannel EEG signal for the sample point n and for channel c. Lastly,

and

are the mean and estimated standard deviation of xl(n, c), 1 ≤ n ≤ nmax, for each channel c.

Consequently, a binary image I(l,c) is constructed according to 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 (z1, z2) ∈ N × 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 ≤ n ≤ nmax and 1 ≤ c ≤ C. The value zl(c) is the image vertical position where the signal’s zero value has to be situated in order to fit the entire signal within the image for each channel c: where the minimization and maximization are carried out for n varying between 1 ≤ n ≤ nmax, and ⌊·⌋ denote the rounding to the smaller nearest integer of the number. In order to complete the plot I(l,c) from the pixels, the Bresenham (Bresenham, 1965; Ramele et al., 2016) algorithm is used to interpolate straight lines between each pair of consecutive pixels.

*2. MATERIALES Y MÉTODOS*

*El P300 (Farwell and Donchin, 1988; Knuth et al., 2006) es una desviación positiva de la señal del EEG que ocurre alrededor de 300 ms después del inicio de un estímulo raro y desviado que se espera que el sujeto atienda. Se produce bajo el paradigma del bicho raro (Wolpaw y Wolpaw, 2012) y es consistente en diferentes temas. Tiene una amplitud más baja (±5 μV) en comparación con la actividad EEG basal, alcanzando una relación señal/ruido (SNR) de alrededor de −15 db estimada en base a la amplitud de la señal de respuesta P300 dividida por la desviación estándar de la actividad EEG de fondo ( Hu et al., 2010). Esta señal se puede utilizar para implementar una aplicación de ortografía por medio de una matriz de ortografía (Farwell y Donchin, 1988). Esta matriz es compuesto por 6 filas y 6 columnas de números y letras. El sujeto puede enfocarse en un carácter de la matriz. La Figura 1 muestra un ejemplo de Speller Matrix utilizado en el software de código abierto OpenVibe (Renard et al., 2010), donde los destellos de filas y columnas proporcionan el estímulo desviado necesario para provocar esta respuesta fisiológica. Cada vez que parpadea una fila o una columna que contiene la letra deseada, la señal EEG sincronizada correspondiente también debe contener la firma P300 y, al detectarla, se puede identificar la letra seleccionada.*

*2.1. Extracción de características de gráficos de señales*

*En esta sección, se describen el preprocesamiento de señales, el método para generar imágenes a partir de gráficos de señales, el procedimiento de extracción de características y la identificación de la matriz de ortografía. La figura 2 muestra un esquema de todo el proceso.*

*2.1.1. Canalización de preprocesamiento*

*Los datos obtenidos por el dispositivo de captura se digitalizan y se construye una señal EEG multicanal.*

*Las 6 filas y las 6 columnas de SpellerMatrix se intensifican proporcionando el estímulo visual. El número de una fila o columna es una ubicación. Una secuencia de 12 ubicaciones permutadas aleatoriamente conforman una secuencia de intensificación. Todo el conjunto de 12 intensificaciones se repite ka veces.*

*• Mejora de la señal: esta etapa consiste en la mejora de la SNR del patrón P300 por encima del nivel del EEG basal. La tubería comienza aplicando un filtro de muesca a la señal digital sin procesar, un filtro Butterworth de paso bajo de 10 Hz de cuarto grado y finalmente una aniquilación con un filtro de respuesta de impulso finito (FIR) de orden 30 desde la frecuencia de muestreo original hasta 16 Hz (Krusienski et al. ., 2006).*

*• Eliminación de artefactos: para cada secuencia completa de 12 intensificaciones de 6 filas y 6 columnas, se implementa un procedimiento básico de eliminación de artefactos mediante la eliminación de toda la secuencia cuando cualquier señal se desvía por encima o por debajo de ±70 μV.*

*• Segmentación: Para cada una de las 12 intensificaciones de una secuencia de intensificación, se extrae un segmento Sl i de una ventana de tmax segundos de la señal multicanal, a partir del inicio del estímulo, correspondiente a cada fila/columna intensificación l y a la secuencia de intensificación i. A medida que las intensificaciones se permutan en un orden aleatorio, los segmentos se reorganizan correspondientes al parpadeo de las filas, etiquetados del 1 al 6, mientras que los correspondientes al parpadeo de las columnas se etiquetan del 7 al 12. Dos de estos segmentos deben contener la firma P300 ERP bloqueada en el tiempo para el estímulo intermitente, uno para la fila y otro para la columna.*

*• Promedio de señal: el P300 ERP está profundamente enterrado bajo el EEG basal, por lo que el enfoque estándar para identificarlo es promediando punto a punto los segmentos de señal apilados con bloqueo de tiempo. Por lo tanto, los valores que no están relacionados y que no están sincronizados con el inicio del estímulo se cancelan (Liang y Bougrain, 2008).*

*Este último paso determina el funcionamiento de cualquier P300 Speller. Para obtener una señal mejorada en términos de su SNR, son necesarias repeticiones de la secuencia de intensificación fila/columna. Y, al mismo tiempo, a medida que se necesitan más repeticiones, la capacidad de transferir información más rápido disminuye, por lo que existe una compensación que debe determinarse con precisión. El procedimiento para obtener la señal promediada punto a punto es el siguiente:*

*1. Resalte aleatoriamente las filas y columnas de la matriz. Hay una fila y una columna que deben coincidir con la letra seleccionada por el sujeto.*

*2. Repita el paso 2.1.1 ka veces, obteniendo los 1 ≤ l ≤ 12 segmentos Sl 1(n, c), . . . , Sl ka (n, c), de la señal EEG donde las variables 1 ≤ n ≤ nmax y 1 ≤ c ≤ C corresponden a puntos de muestra y canal, respectivamente. El parámetro C es el número de canales de EEG disponibles, mientras que nmax = Fs tmax es la longitud del segmento y Fs es la frecuencia de muestreo. El parámetro ka es el número de repeticiones de intensificaciones y es un parámetro de entrada del algoritmo.*

*3. Calcule el promedio de conjunto para 1 ≤ n ≤ nmax y para los canales 1 ≤ c ≤ C. Esto proporciona una señal promediada xl(n, c) para las doce ubicaciones 1 ≤ l ≤ 12.*

*2.1.2. Trazado de señales*

*Los segmentos de señal promediados están estandarizados y escalados para 1 ≤ n ≤ nmax y 1 ≤ c ≤ C donde γ > 0 es un parámetro de entrada del algoritmo y está relacionado con la escala de la imagen. Además, xl(n, c) es la señal de EEG multicanal promediada punto a punto para el punto de muestra n y para el canal c. Por último,*

*y*

*son la desviación estándar media y estimada de xl(n, c), 1 ≤ n ≤ nmax, para cada canal c.*

*En consecuencia, se construye una imagen binaria I(l,c) de acuerdo a que 255 es blanco y representa la ubicación del valor de la señal y 0 es negro, que es el contraste de fondo, conformando un gráfico en blanco y negro de la señal. Los argumentos de píxeles (z1, z2) ∈ N × N iteran sobre el ancho (según la longitud de la señal segmento) y la altura (basada en la amplitud de pico a pico) de la imagen recién creada con 1 ≤ n ≤ nmax y 1 ≤ c ≤ C. El valor zl(c) es la posición vertical de la imagen donde el valor cero de la señal tiene a situar para encajar toda la señal dentro de la imagen para cada canal c: donde la minimización y maximización se realizan para n variando entre 1 ≤ n ≤ nmax, y ⌊·⌋ denota el redondeo al menor entero más próximo de los número. Para completar la gráfica I(l,c) a partir de los píxeles, se utiliza el algoritmo de Bresenham (Bresenham, 1965; Ramele et al., 2016) para interpolar líneas rectas entre cada par de píxeles consecutivos.*

2.1.3. Feature Extraction: Histogram of Gradient Orientations

The work of Hubel and Wiesel (1962), on how the visual cortex sense features was the inspiration to the development of an algorithm to identify and decode salient local information from image regions. The Scale Invariant Feature Transform (SIFT) is a Computer Vision method proposed by Lowe (2004) which is composed of two parts, the SIFT Detector and the SIFT Descriptor. The former is the procedure to identify relevant areas of an image whereas the latter is the procedure to describe and characterize a region of an image (i.e. patch) calculating an histogram of the angular orientations of pixel gradients. In order to characterize EEG signal waveforms, this work proposes an alternative to the SIFT Descriptor, the Histogram of Gradient Orientations (HIST) algorithm.

For each generated image I(l,c), a keypoint pk is placed on a pixel (xpk , ypk ) over the image plot and a window around the keypoint is considered: a local image patch. Its size is Xp × Xp pixels and is constructed by dividing the window in 16 blocks of size 3s each one, where s is the scale of the local patch and it is an input parameter of the algorithm. It is arranged in a 4 × 4 grid and the pixel pk is the patch center, thus Xp = 12s pixels. A local representation of the signal shape within the patch can be described by obtaining the gradient orientations on each of the 16 blocks Bi,j with 0 ≤ i, j ≤ 3 and creating a histogram of gradients. In order to calculate the histogram, the interval [0, 360] of possible angles is divided in 8 bins, each one of 45 degrees. Hence, for each spatial bin 0 ≤ i, j ≤ 3, corresponding to the indexes of each block Bi,j, the orientations are accumulated in a 3-dimensional histogram h through the following equation:

where p is a pixel from the image I(l,c), θ is the angle bin with θ ∈ {0, 45, 90, 135, 180, 225, 270, 315}, J(p) is the euclidean normof the gradient vector in the pixel p and it is computed using finite differences and 6 J(p) is the angle of the gradient vector. The contribution of each gradient vector to the histogram calculated by Equation 5 is balanced by a trilinear interpolation. The scalar wang(·) and vector wij(·) functions are linear interpolations used by Lowe (2004) and Vedaldi and Fulkerson (2010) to provide a weighting contribution to the eight adjacent bins in the tridimensional histogram. They are calculated as where xi and yi are the spatial bin centers located in xi, yj ∈ {−3 2 ,−1 2 , 1 2 , 3 2 } and the interpolating function w(・) is defined as w(z) = max(0, 1 − |z|). The function parameter v = (vx, vy) is a vector variable and α a scalar variable. Vector v holds pixel coordinates (vx, vy) normalized between −2 and 2 and combined with the function w(z) it produces zero for every combination of (i, j) except for the 4 adjacent spatial bins. On the other hand, r is an integer that can vary freely in the set {−1, 0, 1} and α is the difference between the gradient orientation angle and the angle bin center in radians. By following this procedure, summands on Equation (7) are nullified except for the 2 adjacent angular bins. These binning functions conform the trilinear interpolation that has a combined effect of sharing the contribution of each oriented gradient between their eight adjacent bins in a tridimensional cube in the histogram space, and zero everywhere else (Mortensen and Shapiro, 2005). The fixed value of 3 is a magnification factor which corresponds to the number of pixels per each block when s = 1. As the patch has 16 blocks and 8 bin angles are considered, for each location l and channel c a feature called descriptor d(l,c) of 128 dimension is obtained. The main differences between this implementation and the standard SIFT Descriptor are described in the Appendix. Figure 3 shows an example of a patch and a scheme of the histogram computation. In Figure 3A a plot of the signal and the patch centered around the keypoint is shown. In Figure 3B the possible orientations on each patch are illustrated. Only the upper-left four blocks are visible. The first eight orientations of the first block, are labeled from 1 to 8 clockwise. The orientations of the second block B1,2 are 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 d of 128 coordinates. Finally, in (C) an enlarged image plot is shown where the oriented gradient vector for each pixel can be seen.

*2.1.3. Extracción de características: histograma de orientaciones de degradado*

*El trabajo de Hubel y Wiesel (1962), sobre cómo las funciones sensoriales de la corteza visual fue la inspiración para el desarrollo de un algoritmo para identificar y decodificar información local destacada de las regiones de la imagen. La Transformación de características invariantes de escala (SIFT) es un método de visión artificial propuesto por Lowe (2004) que se compone de dos partes, el detector SIFT y el descriptor SIFT. El primero es el procedimiento para identificar áreas relevantes de una imagen, mientras que el segundo es el procedimiento para describir y caracterizar una región de una imagen (es decir, un parche) calculando un histograma de las orientaciones angulares de los gradientes de píxeles. Para caracterizar las formas de onda de la señal EEG, este trabajo propone una alternativa al descriptor SIFT, el algoritmo Histogram of Gradient Orientations (HIST).*

*Para cada imagen generada I(l,c), se coloca un punto clave pk en un píxel (xpk, ypk) sobre el gráfico de la imagen y se considera una ventana alrededor del punto clave: un parche de imagen local. Su tamaño es de Xp × Xp píxeles y se construye dividiendo la ventana en 16 bloques de tamaño 3s cada uno, donde s es la escala del parche local y es un parámetro de entrada del algoritmo. Está dispuesto en una cuadrícula de 4 × 4 y el pk de píxeles es el centro del parche, por lo que Xp = 12s píxeles. Se puede describir una representación local de la forma de la señal dentro del parche obteniendo las orientaciones de los gradientes en cada uno de los 16 bloques Bi,j con 0 ≤ i, j ≤ 3 y creando un histograma de gradientes. Para calcular el histograma, el intervalo [0, 360] de ángulos posibles se divide en 8 bins, cada uno de 45 grados. Por lo tanto, para cada bin espacial 0 ≤ i, j ≤ 3, correspondiente a los índices de cada bloque Bi,j, las orientaciones se acumulan en un histograma tridimensional h a través de la siguiente ecuación:*

*donde p es un píxel de la imagen I(l,c), θ es el ángulo bin con θ ∈ {0, 45, 90, 135, 180, 225, 270, 315}, J(p) es la norma euclidiana de la vector gradiente en el píxel p y se calcula usando diferencias finitas y 6 J(p) es el ángulo del vector gradiente. La contribución de cada vector de gradiente al histograma calculado por la Ecuación 5 se equilibra mediante una interpolación trilineal. Las funciones escalar wang(·) y vectorial wij(·) son interpolaciones lineales utilizadas por Lowe (2004) y Vedaldi y Fulkerson (2010) para proporcionar una contribución de ponderación a los ocho contenedores adyacentes en el histograma tridimensional. Se calculan como donde xi y yi son los centros de bin espaciales ubicados en xi, yj ∈ {−3 2 ,−1 2 , 1 2 , 3 2 } y la función de interpolación w(・) se define como w(z) = máx(0, 1 − |z|). El parámetro de función v = (vx, vy) es una variable vectorial y α una variable escalar. El vector v contiene coordenadas de píxeles (vx, vy) normalizadas entre −2 y 2 y combinadas con la función w(z) produce cero para cada combinación de (i, j) excepto para los 4 contenedores espaciales adyacentes. Por otro lado, r es un número entero que puede variar libremente en el conjunto {−1, 0, 1} y α es la diferencia entre el ángulo de orientación del gradiente y el centro del ángulo en radianes. Siguiendo este procedimiento, los sumandos en la Ecuación (7) se anulan excepto para los 2 contenedores angulares adyacentes. Estas funciones de binning conforman la interpolación trilineal que tiene el efecto combinado de compartir la contribución de cada gradiente orientado entre sus ocho bins adyacentes en un cubo tridimensional en el espacio del histograma, y ​​cero en cualquier otro lugar (Mortensen y Shapiro, 2005). El valor fijo de 3 es un factor de aumento que corresponde al número de píxeles por cada bloque cuando s = 1. Como el parche tiene 16 bloques y se consideran 8 bin angles, para cada ubicación l y canal c se obtiene una característica llamada descriptor d(l,c) de 128 dimensiones. Las principales diferencias entre esta implementación y el SIFT Descriptor estándar se describen en el Apéndice. La Figura 3 muestra un ejemplo de un parche y un esquema del cálculo del histograma. En la Figura 3A se muestra un gráfico de la señal y el parche centrado alrededor del punto clave. En la Figura 3B se ilustran las posibles orientaciones de cada parche. Solo los cuatro bloques superiores izquierdos son visibles. Las primeras ocho orientaciones del primer bloque están etiquetadas del 1 al 8 en el sentido de las agujas del reloj. Las orientaciones del segundo bloque B1,2 están etiquetadas del 9 al 16. Este etiquetado continúa de izquierda a derecha, de arriba hacia abajo hasta que se asignan las ocho orientaciones para los dieciséis bloques. Forman el correspondiente descriptor d de 128 coordenadas. Finalmente, en (C) se muestra un diagrama de imagen ampliado donde se puede ver el vector de gradiente orientado para cada píxel.*

2.1.4. Speller Matrix Letter Identification

2.1.4.1. P300 ERP extraction

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

• Step A: First highlight rows and columns from the matrix in a random permutation order and obtain the Ensemble Average as detailed in steps 2.1.1, 2.1.1, and 2.1.1 in section 2.1.1.

• Step B: Plot the signals ˜xl(n, c), 1 ≤ n ≤ nmax, 1 ≤ c ≤ C, according section 2.1.2 in order to generate the images I(l,c) for rows and columns 1 ≤ l ≤ 12.

• Step C: Obtain the descriptors d(l,c) for rows and columns from I(l,c) in accordance to the method described in section 2.1.3.

2.1.4.2. Calibration

A trial, as defined by the BCI2000 platform (Schalk et al., 2004), is every attempt to select just one letter from the speller. A set of trials is used for calibration and once the calibration is complete it can be used to identify new letters from new trials. During the calibration phase, two descriptors d(l,c) are extracted for each available channel, corresponding to the locations l of a selection of one previously instructed letter from the set of calibration trials. These descriptors are the P300 templates, grouped together in a template set called Tc. The set is constructed using the steps described in section 2.1.1 and the steps A, B, and C of the P300 ERP extraction process. Additionally, the best performing channel, bpc is identified based on the the channel where the best Character Recognition Rate is obtained.

2.1.4.3. Letter identification

In order to identify the selected letter, the template set Tbpc is used as a database. Thus, new unclassified descriptors q(l,bpc) are computed and they are compared against the descriptors belonging to the calibration template set Tbpc. The Naive Bayes Nearest Neighbor (k-NBNN) (Boiman et al., 2008) is a discriminative (Wolpaw and Wolpaw, 2012) semisupervised classification algorithm that allows the categorization of an image to one class by comparing the set of extracted descriptors to those which are more similar from template dictionaries. This work proposes an adapted version to obtain a unary classification scheme to identify the selected letter in the P300-Based BCI Speller, based on the features provided by the calculated descriptors.

• Step D: Match to the calibration template Tbpc by computing with d (bpc) h belonging to the set NT(q(l,bpc)), which is defined, for the best performing channel, as NT(q(l,bpc)) = {d (bpc) h ∈ Tbpc/d(bpc) is the k-nearest neighbor of q(l,bpc)}. This set is obtained by sorting all the elements in Tbpc based on distances between them and q(l,bpc), choosing the k with smaller values, with k a parameter of the algorithm.

By computing the aforementioned equations, the letter of the matrix can be determined from the intersection of the row roˆw and column coˆ l. Figure 2 shows a scheme of this process.

*2.1.4. Identificación de letras de la matriz ortográfica*

*2.1.4.1. Extracción ERP P300*

*Los segmentos correspondientes al parpadeo de las filas se etiquetan del 1 al 6, mientras que los correspondientes al parpadeo de las columnas se etiquetan del 7 al 12. El proceso de extracción tiene los siguientes pasos:*

*• Paso A: Primero resalte las filas y columnas de la matriz en un orden de permutación aleatoria y obtenga el Promedio del conjunto como se detalla en los pasos 2.1.1, 2.1.1 y 2.1.1 en la sección 2.1.1.*

*• Paso B: Graficar las señales ˜xl(n, c), 1 ≤ n ≤ nmax, 1 ≤ c ≤ C, según sección 2.1.2 para generar las imágenes I(l,c) para filas y columnas 1 ≤ l ≤ 12.*

*• Paso C: Obtener los descriptores d(l,c) para filas y columnas de I(l,c) de acuerdo al método descrito en la sección 2.1.3.*

*2.1.4.2. Calibración*

*Un ensayo, tal como lo define la plataforma BCI2000 (Schalk et al., 2004), es cada intento de seleccionar solo una letra del deletreador. Se utiliza un conjunto de ensayos para la calibración y, una vez que se completa la calibración, se puede utilizar para identificar nuevas letras a partir de nuevos ensayos. Durante la fase de calibración, se extraen dos descriptores d(l,c) para cada canal disponible, correspondientes a las ubicaciones l de una selección de una letra previamente instruida del conjunto de ensayos de calibración. Estos descriptores son las plantillas P300, agrupadas en un conjunto de plantillas llamado Tc. El conjunto se construye utilizando los pasos descritos en la sección 2.1.1 y los pasos A, B y C del proceso de extracción de P300 ERP. Además, el canal con mejor rendimiento, bpc, se identifica en función del canal en el que se obtiene la mejor tasa de reconocimiento de caracteres.*

*2.1.4.3. Identificación de letras*

*Para identificar la letra seleccionada, se utiliza como base de datos el conjunto de plantillas Tbpc. Así, se calculan nuevos descriptores q(l,bpc) no clasificados y se comparan con los descriptores que pertenecen al conjunto de plantillas de calibración Tbpc. El vecino más cercano de Naive Bayes (k-NBNN) (Boiman et al., 2008) es un algoritmo de clasificación semisupervisado discriminativo (Wolpaw y Wolpaw, 2012) que permite la categorización de una imagen en una clase al comparar el conjunto de descriptores extraídos con aquellos. que son más similares a los diccionarios de plantilla. Este trabajo propone una versión adaptada para obtener un esquema de clasificación unario para identificar la letra seleccionada en el BCI Speller basado en P300, a partir de las características proporcionadas por los descriptores calculados.*

*• Paso D: Hacer coincidir con la plantilla de calibración Tbpc calculando con d (bpc) h perteneciente al conjunto NT(q(l,bpc)), que se define, para el canal de mejor rendimiento, como NT(q(l,bpc) )) = {d (bpc) h ∈ Tbpc/d(bpc) es el k vecino más cercano de q(l,bpc)}. Este conjunto se obtiene ordenando todos los elementos en Tbpc en función de las distancias entre ellos y q(l,bpc), eligiendo los k con valores más pequeños, siendo k un parámetro del algoritmo.*

*Al calcular las ecuaciones antes mencionadas, la letra de la matriz se puede determinar a partir de la intersección de la fila fila y la columna coˆ l. La Figura 2 muestra un esquema de este proceso.*

2.2. Experimental Protocol

To verify the validity of the proposed framework and method, the public dataset 008-2014 (Riccio et al., 2013) published on the BNCI-Horizon website (Brunner et al., 2014) by IRCCS Fondazione Santa Lucia, is used. Additionally, an own dataset with the same experimental conditions is generated. Both of them are utilized to perform an offline BCI Simulation to decode the spelled words from the provided signals.

The algorithm is implemented on MATLAB V2017a (Mathworks Inc., Natick, MA, USA). The algorithm described in section 2.1.3 is implemented on a modified version of the VLFeat (Vedaldi and Fulkerson, 2010) Computer Vision library. Furthermore, in order to enhance the impact of this paper and for a sake of reproducibility, the code of the entire algorithm, including the modified VLFeat library, has been made available at: <https://bitbucket.org/itba/hist>. In the following sections the characteristics of the datasets and parameters of the identification algorithm are described.

*2.2. Protocolo experimental*

*Para verificar la validez del marco y método propuesto, se utiliza el conjunto de datos público 008-2014 (Riccio et al., 2013) publicado en el sitio web de BNCI-Horizon (Brunner et al., 2014) por IRCCS Fondazione Santa Lucia. Adicionalmente, se genera un dataset propio con las mismas condiciones experimentales. Ambos se utilizan para realizar una simulación BCI fuera de línea para decodificar las palabras deletreadas de las señales proporcionadas.*

*El algoritmo se implementa en MATLAB V2017a (Mathworks Inc., Natick, MA, EE. UU.). El algoritmo descrito en la sección 2.1.3 se implementa en una versión modificada de la biblioteca VLFeat (Vedaldi y Fulkerson, 2010) Computer Vision. Además, para mejorar el impacto de este documento y en aras de la reproducibilidad, el código de todo el algoritmo, incluida la biblioteca VLFeat modificada, está disponible en: https://bitbucket.org/itba/hist. En las siguientes secciones se describen las características de los conjuntos de datos y los parámetros del algoritmo de identificación.*

2.2.1. P300 ALS Public Dataset

The experimental protocol used to generate this dataset is explained in Riccio et al. (2013) but can be summarized as follows: eight subjects with confirmed diagnoses but on different stages of ALS disease, were recruited and accepted to perform the experiments. The Visual P300 detection task designed for this experiment consisted of spelling seven words of five letters each, using the traditional P300 Speller Matrix (Farwell and Donchin, 1988). The flashing of rows and columns provide the deviant stimulus required to elicit this physiological response. The first 3 words are used for calibration and the remaining four words, for testing with visual feedback. A trial is every attempt to select a letter from the speller. It is composed of signal segments corresponding to ka = 10 repetitions of flashes of 6 rows and ka = 10 repetitions of flashes of 6 columns of the matrix, yielding 120 repetitions. Flashing of a row or a columna is performed for 0.125 s, following by a resting period (i.e., inter-stimulus interval) of the same length. After 120 repetitions an inter-trial pause is included before resuming with the following letter.

The recorded dataset was sampled at 256 Hz and it consisted of a scalp multichannel EEG signal for electrode channels Fz, Cz, Pz, Oz, P3, P4, PO7, and PO8, identified according to the 10–20 International System, for each one of the eight subjects. The recording device was a research-oriented digital EEG device (g.Mobilab, g.Tec, Austria) and the data acquisition and stimuli delivery were handled by the BCI2000 open source software (Schalk et al., 2004).

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

*2.2.1. Conjunto de datos públicos P300 ALS*

*El protocolo experimental utilizado para generar este conjunto de datos se explica en Riccio et al. (2013), pero se puede resumir de la siguiente manera: ocho sujetos con diagnósticos confirmados pero en diferentes etapas de la enfermedad de ELA fueron reclutados y aceptados para realizar los experimentos. La tarea de detección Visual P300 diseñada para este experimento consistió en deletrear siete palabras de cinco letras cada una, usando la Matriz Ortográfica P300 tradicional (Farwell y Donchin, 1988). El parpadeo de filas y columnas proporciona el estímulo desviado necesario para provocar esta respuesta fisiológica. Las primeras 3 palabras se utilizan para la calibración y las cuatro palabras restantes para la prueba con retroalimentación visual. Una prueba es cada intento de seleccionar una letra del deletreador. Está compuesto por segmentos de señal correspondientes a ka = 10 repeticiones de destellos de 6 filas y ka = 10 repeticiones de destellos de 6 columnas de la matriz, dando 120 repeticiones. El parpadeo de una fila o una columna se realiza durante 0,125 s, seguido de un período de descanso (es decir, intervalo entre estímulos) de la misma duración. Después de 120 repeticiones se incluye una pausa entre ensayos antes de continuar con la letra siguiente.*

*El conjunto de datos registrado se muestreó a 256 Hz y consistió en una señal EEG multicanal del cuero cabelludo para los canales de electrodos Fz, Cz, Pz, Oz, P3, P4, PO7 y PO8, identificados según el Sistema Internacional 10-20, para cada uno de los ocho temas. El dispositivo de registro fue un dispositivo EEG digital orientado a la investigación (g.Mobilab, g.Tec, Austria) y la adquisición de datos y la entrega de estímulos fueron manejados por el software de código abierto BCI2000 (Schalk et al., 2004).*

*Para evaluar y verificar la identificación de la respuesta P300, se instruye a los sujetos para que realicen una tarea de ortografía. Deben fijar su atención en letras sucesivas para copiar un conjunto de palabras previamente determinado, en contraste con una operación de ejecución libre del deletreador donde cada usuario decide por sí mismo qué letra elegir.*

2.2.2. P300 for Healthy Subjects

We replicate the same experiment on healthy subjects using a wireless digital EEG device (g.Nautilus, g.Tec, Austria). The experimental conditions are the same as those used for the previous dataset, as detailed in section 2.2.1. The produced dataset is available in a public online repository (Ramele et al., 2017). Participants are recruited voluntarily and the experiment is conducted anonymously in accordance with the Declaration of Helsinki published by the World Health Organization. No monetary compensation is handed out and all participants agree and sign a written informed consent. This study is approved by the Departamento de Investigación y Doctorado, Instituto Tecnológico de Buenos Aires (ITBA). All healthy subjects have normal or corrected-to-normal vision and no history of neurological disorders. The experiment is performed with 8 subjects, 6 males, 2 females, 6 right-handed, 2 left-handed, average age 29.00 years, standard deviation 11.56 years, rango 20–56 years. EEGdata is collected in a single recording session. Participants are seated in a comfortable chair, with their vision aligned to a computer screen located one meter in front of them. The handling and processing of the data and stimuli is conducted by the OpenVibe platform (Renard et al., 2010). Gel-based active electrodes (g.LADYbird, g.Tec, Austria) are used on the same positions Fz, Cz, Pz, Oz, P3,P4, PO7, and PO8. Reference is set to the right ear lobe and ground is preset as the AFz position. Sampling frequency is slightly different, and is set to 250 Hz, which is the closest possible to the one used with the other dataset.

*2.2.2. P300 para sujetos sanos*

*Reproducimos el mismo experimento en sujetos sanos utilizando un dispositivo EEG digital inalámbrico (g.Nautilus, g.Tec, Austria). Las condiciones experimentales son las mismas que las utilizadas para el conjunto de datos anterior, como se detalla en la sección 2.2.1. El conjunto de datos producido está disponible en un repositorio público en línea (Ramele et al., 2017). Los participantes son reclutados voluntariamente y el experimento se realiza de forma anónima de acuerdo con la Declaración de Helsinki publicada por la Organización Mundial de la Salud. No se entrega compensación monetaria y todos los participantes están de acuerdo y firman un consentimiento informado por escrito. Este estudio está aprobado por el Departamento de Investigación y Doctorado, Instituto Tecnológico de Buenos Aires (ITBA). Todos los sujetos sanos tienen una visión normal o corregida a normal y no tienen antecedentes de trastornos neurológicos. El experimento se realiza con 8 sujetos, 6 hombres, 2 mujeres, 6 diestros, 2 zurdos, edad promedio 29,00 años, desviación estándar 11,56 años, rango 20–56 años. EEGdata se recopila en una sola sesión de registro. Los participantes están sentados en una silla cómoda, con su visión alineada con la pantalla de una computadora ubicada a un metro frente a ellos. El manejo y procesamiento de los datos y estímulos es realizado por la plataforma OpenVibe (Renard et al., 2010). Los electrodos activos a base de gel (g.LADYbird, g.Tec, Austria) se utilizan en las mismas posiciones Fz, Cz, Pz, Oz, P3,P4, PO7 y PO8. La referencia se establece en el lóbulo de la oreja derecha y el suelo se preestablece como la posición AFz. La frecuencia de muestreo es ligeramente diferente y se establece en 250 Hz, que es lo más cercano posible a la utilizada con el otro conjunto de datos.*

2.2.3. Parameters

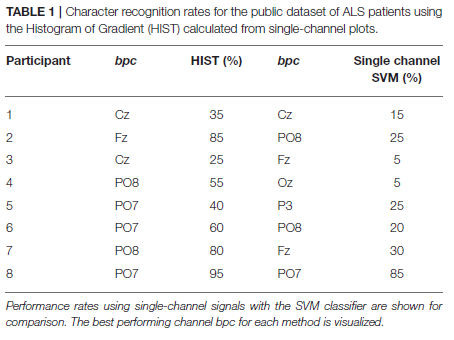
The patch size is XP = 12s × 12s pixels, where s is the scale of the local patch and it is an input parameter of the algorithm. The P300 event can have a span of 400 ms and its amplitude can reach 10μV (Rao, 2013). Hence it is necessary to utilize a signal segment of size tmax = 1 second and a size patch XP that could capture an entire transient event. With this purpose in consideration, the s value election is essential. We propose the Equations (10) and (11) to compute the scale value in horizontal and vertical directions, respectively.

where λ is the length in seconds covered by the patch, Fs is the sampling frequency of the EEG signal (downsampled to 16 Hz) and 1μV corresponds to the amplitude in microvolts that can be covered by the height of the patch. The geometric structure of the patch is determined by the waveform to be captured, thus we discerned that by using s = sx = sy = 3 and g = 4, the local patch and the descriptor can identify events of 9 μV of amplitude, with a span of λ = 0.56 s. This also determines that 1 pixel represents 1 γ = 1 4μV on the vertical direction and 1 Fs γ = 1 64 s on the horizontal direction. The keypoints pk are located at

(xpk , ypk ) = (0.55Fs γ , zl(c)) = (35, zl(c)) for the corresponding channel c and location l (see Equation 4). In this way the whole transient event is captured. Figure 4 shows a patch of a signal plot covering the complete amplitude (vertical direction) and the complete span of the signal event (horizontal direction). The number of channels C is equal to 8 for both datasets, and the number of intensification sequences ka is fixed to 10. The parameter k used to construct the set NT(q(l,c)) is assigned to k = 7, which was found empirically to achieve better results. In addition, the norm used on Equations (8) and (9) is the cosine norm, and descriptors are normalized to [−1, 1].

Lastly, in order to assess the validity of the HIST method, the character recognition rate for both datasets is evaluated replicating the methodology proposed by the ALS dataset’s publisher, since authors Riccio et al. (2013) did not report the Character Recognition Rate obtained for this dataset.

Frequency filtering, data segmentation and artifact rejection is conducted according to section 2.1.1 yielding 16 x 8 samples per epoch. A multichannel feature consists of time points vector (Lotte et al., 2018), formed by concatenating all the channels (Krusienski et al., 2006). A single-channel variant consists of using time points from a single electrode and performing the analysis on a channel-by-channel basis. Three classification schemes are considered as well. A multichannel version of the Stepwise Linear Discriminant Analysis (SWLDA) classification algorithm. SWLDA is the methodology proposed by the ALS dataset’s publisher. Additionally, a single-channel and a multichannel variant of a linear kernel Support Vector Machine (SVM) (Scholkopf and Smola, 2001) classifier are



utilized. SVM has been successfully used in several BCI Competitions (Rakotomamonjy and Guigue, 2008).

*2.2.3. Parámetros*

*El tamaño del parche es XP = 12s × 12s píxeles, donde s es la escala del parche local y es un parámetro de entrada del algoritmo. El evento P300 puede tener un lapso de 400 ms y su amplitud puede alcanzar los 10 μV (Rao, 2013). Por lo tanto, es necesario utilizar un segmento de señal de tamaño tmax = 1 segundo y un tamaño de parche XP que pueda capturar un evento transitorio completo. Con este propósito en consideración, la elección del valor de s es esencial. Proponemos las Ecuaciones (10) y (11) para calcular el valor de la escala en las direcciones horizontal y vertical, respectivamente.*

*donde λ es la longitud en segundos que cubre el parche, Fs es la frecuencia de muestreo de la señal EEG (remuestreada a 16 Hz) y 1μV corresponde a la amplitud en microvoltios que puede cubrir la altura del parche. La estructura geométrica del parche está determinada por la forma de onda a capturar, así discernimos que usando s = sx = sy = 3 y g = 4, el parche local y el descriptor pueden identificar eventos de 9 μV de amplitud, con una intervalo de λ = 0,56 s. Esto también determina que 1 píxel representa 1 γ = 1 4 μV en la dirección vertical y 1 Fs γ = 1 64 s en la dirección horizontal. Los puntos clave pk se encuentran en*

*(xpk, ypk) = (0.55Fs γ, zl(c)) = (35, zl(c)) para el canal c correspondiente y la ubicación l (ver Ecuación 4). De esta manera se captura todo el evento transitorio. La figura 4 muestra un parche de un gráfico de señal que cubre la amplitud completa (dirección vertical) y el intervalo completo del evento de señal (dirección horizontal). El número de canales C es igual a 8 para*

*ambos conjuntos de datos, y el número de secuencias de intensificación ka se fija en 10. El parámetro k utilizado para construir el conjunto NT(q(l,c)) se asigna a k = 7, que se encontró empíricamente para lograr mejores resultados. Además, la norma utilizada en las Ecuaciones (8) y (9) es la norma del coseno y los descriptores se normalizan a [−1, 1].*

*Por último, para evaluar la validez del método HIST, se evalúa la tasa de reconocimiento de caracteres para ambos conjuntos de datos replicando la metodología propuesta por el editor del conjunto de datos ALS, ya que los autores Riccio et al. (2013) no informaron la tasa de reconocimiento de caracteres obtenida para este conjunto de datos.*

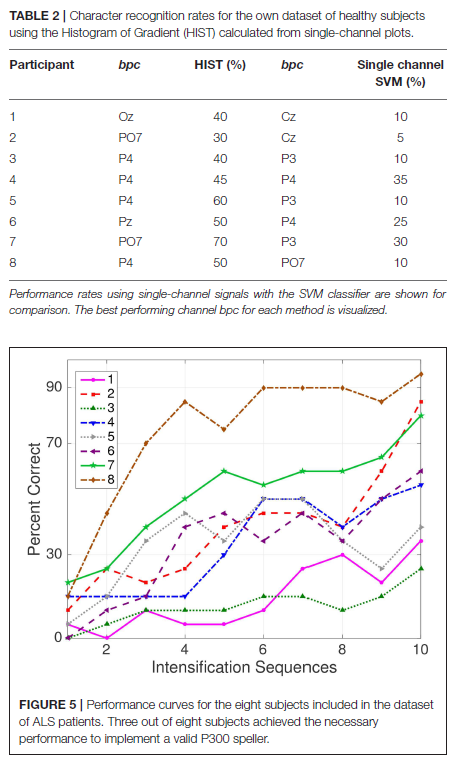
*El filtrado de frecuencia, la segmentación de datos y el rechazo de artefactos se realizan de acuerdo con la sección 2.1.1, lo que genera 16 x 8 muestras por época. Una característica multicanal consiste en un vector de puntos de tiempo (Lotte et al., 2018), formado por la concatenación de todos los canales (Krusienski et al., 2006). Una variante de un solo canal consiste en usar puntos de tiempo de un solo electrodo y realizar el análisis canal por canal. También se consideran tres esquemas de clasificación. Una versión multicanal del algoritmo de clasificación Stepwise Linear Discriminant Analysis (SWLDA). SWLDA es la metodología propuesta por el editor del conjunto de datos ALS. Además, una variante de un solo canal y una variante multicanal de un clasificador de máquina de vectores de soporte (SVM) de kernel lineal (Scholkopf y Smola, 2001) son utilizado. SVM se ha utilizado con éxito en varias competencias BCI (Rakotomamonjy y Guigue, 2008).*

3. RESULTS

Table 1 shows the results of applying the HIST algorithm to the subjects of the public dataset of ALS patients. The percentage of correctly spelled letters is calculated while performing an offline BCI Simulation. From the seven words for each subject, the first three are used for calibration, and the remaining four are used for testing. The best performing channel bpc is informed as well. The target ratio is 1:36; hence theoretical chance level is 2.8%. It can be observed that the best performance of the letter identification method is reached in a dissimilar channel depending on the subject being studied. Tables 1, 2 show for comparison the obtained performance rates using single-channel signals with the SVMclassifier. The best performing channel, where the best letter identification rate was achieved, is also depicted. The Information Transfer Rate (ITR), or Bit Transfer Rate (BTR), in the case of reactive BCIs (Wolpaw andWolpaw, 2012) depends on the amount of signal averaging required to transmit a valid and robust selection. Figure 5 shows the performance curves for varying intensification sequences for the subjects included in the dataset of ALS patients. It can be noticed that the percentage of correctly identified letters depends on the number of intensification sequences that are used to obtain the averaged signal. Moreover, when the number of intensification sequences tend to 1, which corresponds to single-intensification carácter recognition, the performance is reduced. As mentioned before, the SNR of the P300 obtained from only one segment of the intensification sequence is very low and the shape of its P300 component is not very well defined. In Table 2 the results obtained for 8 healthy subjects are shown. It can be observed that the performance is above chance level. It is verified that HIST method has an improved performance at letter identification than SVM that process the signals on a channel by channel strategy (Wilcoxon signed-rank test, p = 0.004 for both datasets).

*3. RESULTADOS*

*La Tabla 1 muestra los resultados de aplicar el algoritmo HIST a los sujetos del conjunto de datos público de pacientes con ELA. El porcentaje de letras escritas correctamente se calcula mientras se realiza una simulación BCI fuera de línea. De las siete palabras de cada tema, las tres primeras se utilizan para la calibración y las cuatro restantes para la prueba. También se informa el canal bpc con mejor rendimiento. La relación objetivo es 1:36; por lo tanto, el nivel de probabilidad teórico es del 2,8%. Se puede observar que el mejor desempeño del método de identificación de letras se alcanza en un canal disímil dependiendo del tema que se estudie. Las tablas 1, 2 muestran a modo de comparación las tasas de rendimiento obtenidas utilizando señales de un solo canal con el clasificador SVM. También se representa el canal con mejor desempeño, donde se logró la mejor tasa de identificación de letras. La Tasa de Transferencia de Información (ITR), o Tasa de Transferencia de Bits (BTR), en el caso de BCI reactivos (Wolpaw y Wolpaw, 2012) depende de la cantidad de promedio de señal requerida para transmitir una selección válida y robusta. La figura 5 muestra las curvas de rendimiento para secuencias de intensificación variables para los sujetos incluidos en el conjunto de datos de pacientes con ELA. Se puede notar que el porcentaje de letras correctamente identificadas depende del número de secuencias de intensificación que se utilicen para obtener la señal promediada. Además, cuando el número de secuencias de intensificación tiende a 1, lo que corresponde al reconocimiento de carácter de intensificación única, el rendimiento se reduce. Como se mencionó anteriormente, la SNR del P300 obtenida de un solo segmento de la secuencia de intensificación es muy baja y la forma de su componente P300 no está muy bien definida. En la Tabla 2 se muestran los resultados obtenidos para 8 sujetos sanos. Se puede observar que el rendimiento está por encima del nivel de probabilidad. Se verifica que el método HIST tiene un rendimiento mejorado en la identificación de letras que SVM que procesa las señales en una estrategia de canal por canal (prueba de rango con signo de Wilcoxon, p = 0.004 para ambos conjuntos de datos).*

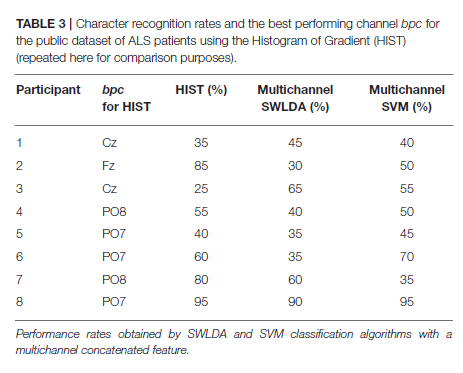


Tables 3, 4 are presented in order to compare the performance of the HIST method versus multichannel SWLDA and SVM classification algorithms for both datasets. It is verified for the dataset of ALS patients that it has similar performance against other methods like SWLDA or SVM, which use a multichannel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

The P300 ERP consists of two overlapping components: the P3a and P3b, the former with frontocentral distribution while the later stronger on centroparietal region (Polich, 2007). Hence, the standard practice is to find the stronger response on the central channel Cz (Riccio et al., 2013). However, Krusienski et al. (2006) show that the response may also arise in occipital regions. We found that by analyzing only the waveforms, occipital channels PO8 and PO7 show higher performances for some subjects.

*Las tablas 3 y 4 se presentan para comparar el rendimiento del método HIST frente a los algoritmos de clasificación SWLDA y SVM multicanal para ambos conjuntos de datos. Se verifica para el conjunto de datos de pacientes con ELA que tiene un rendimiento similar frente a otros métodos como SWLDA o SVM, que utilizan una función multicanal (prueba Quade con p = 0,55), mientras que para el conjunto de datos de sujetos sanos se encuentran diferencias significativas (prueba Quade con p = 0.02) donde solo el método HIST logra un desempeño diferente al SVM (con múltiples comparaciones, diferencia significativa de nivel 0.05).*

*El ERP P300 consta de dos componentes superpuestos: el P3a y el P3b, el primero con distribución frontocentral mientras que el segundo más fuerte en la región centroparietal (Polich, 2007). Por lo tanto, la práctica estándar es encontrar la respuesta más fuerte en el canal central Cz (Riccio et al., 2013). Sin embargo, Krusienski et al. (2006) muestran que la respuesta también puede surgir en las regiones occipitales. Descubrimos que al analizar solo las formas de onda, los canales occipitales PO8 y PO7 muestran rendimientos más altos para algunos sujetos.*



As subjects have varying latencies and amplitudes of their P300 components, they also have a varying stability of the shape of the generated ERP (Nam et al., 2010). Figure 6 shows 10 sample P300 templates patches for patients 8 and 3 from the dataset of ALS patients. It can be discerned that in coincidence with the performance results, the P300 signature is more clear and consistent for subject 8 (A) while for subject 3 (B) the characteristic pattern is more difficult to perceive.

Additionally, the stability of the P300 component waveform has been extensively studied in patients with ALS (Sellers et al., 2006; Madarame et al., 2008; Nijboer and Broermann, 2009; Mak et al., 2012; McCane et al., 2015) where it was found that these patients have a stable P300 component, which were also sustained across different sessions. In line with these results we do not find evidence of a difference in terms of the performance obtained by analyzing the waveforms (HIST) for the group of patients with ALS and the healthy group of volunteers (Mann– Whitney U-Test, p = 0.46). Particularly, the best performance is obtained for a subject from the ALS dataset for which, based on visual observation, the shape of they P300 component is consistently identified.

It is important to remark that when applied to binary images obtained from signal plots, the feature extraction method described in section 2.1.3 generates sparse descriptors. Under this subspace we found that using the cosine metric yielded a significant performance improvement. On the other hand, the unary classification scheme based on the NBNN algorithm proved very beneficial for the P300 Speller Matrix. This is due to the fact that this approach solves the unbalance dataset problem which is inherent to the oddball paradigm (Tibon and Levy, 2015).

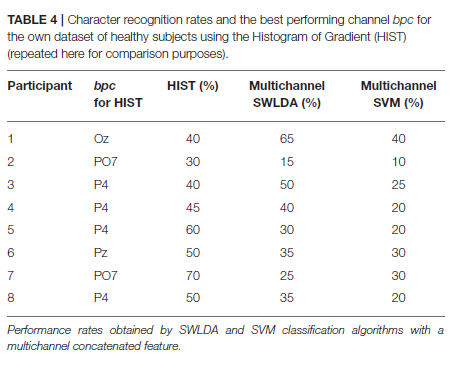
*Como los sujetos tienen latencias y amplitudes variables de sus componentes P300, también tienen una estabilidad variable de la forma del ERP generado (Nam et al., 2010). La figura 6 muestra 10 muestras de parches de plantillas P300 para los pacientes 8 y 3 del conjunto de datos de pacientes con ELA. Se puede discernir que, en coincidencia con los resultados de rendimiento, la firma P300 es más clara y consistente para el sujeto 8 (A) mientras que para el sujeto 3 (B) el patrón característico es más difícil de percibir.*

*Además, la estabilidad de la forma de onda del componente P300 se ha estudiado ampliamente en pacientes con ELA (Sellers et al., 2006; Madarame et al., 2008; Nijboer and Broermann, 2009; Mak et al., 2012; McCane et al., 2015) donde se encontró que estos pacientes tienen un componente P300 estable, que también se mantuvo en diferentes sesiones. En línea con estos resultados no encontramos evidencia de una diferencia en cuanto al rendimiento obtenido al analizar las formas de onda (HIST) para el grupo de pacientes con ELA y el grupo de voluntarios sanos (Mann-Whitney U-Test, p = 0,46 ). En particular, el mejor rendimiento se obtiene para un sujeto del conjunto de datos ALS para el cual, según la observación visual, la forma del componente P300 se identifica consistentemente.*

*Es importante señalar que cuando se aplica a imágenes binarias obtenidas a partir de gráficos de señales, el método de extracción de características descrito en la sección 2.1.3 genera descriptores dispersos. Bajo este subespacio encontramos que el uso de la métrica del coseno produjo una mejora significativa en el rendimiento. Por otro lado, el esquema de clasificación unario basado en el algoritmo NBNN demostró ser muy beneficioso para P300 Speller Matrix. Esto se debe al hecho de que este enfoque resuelve el problema del conjunto de datos desequilibrado que es inherente al paradigma del bicho raro (Tibon y Levy, 2015).*

4. DISCUSSION

Among other applications of Brain Computer Interfaces, the goal of the discipline is to provide communication assistance to people affected by neuro-degenerative diseases, who are the most likely



population to benefit from BCI systems and EEG processing and analysis.

In this work, a method to extract an objective metric from the waveform of the plots of EEG signals is presented. Its usage to implement a valid P300-Based BCI Speller application is expounded. Additionally, its validity is evaluated using a public dataset of ALS patients and an own dataset of healthy subjects. It was verified that this method has an improved performance at letter identification than other methods that process the signals on a channel by channel strategy, and it even has a comparable performance against other methods like SWLDA or SVM, which uses a multichannel feature. Furthermore, this method has the advantage that shapes of waveforms can be analyzed in an objective way. We observed that the shape of the P300 component is more stable in occipital channels, where the performance for identifying letters is higher.We additionally verified that ALS P300 signatures are stable in comparison to those of healthy subjects.

We believe that the use of descriptors based on histogram of gradient orientation, presented in this work, can also be utilized for deriving a shape metric in the space of the P300 signals which can complement other metrics based on time-domain as those defined by Mak et al. (2012). It is important to notice that the analysis of waveform shapes is usually performed in a qualitative approach based on visual inspection (Sellers et al., 2006), and a complementary methodology which offer a quantitative metric will be beneficial to these routinely analysis of the waveform of ERPs.

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

First, the stability of the P300 in terms of its shape is crucial: the averaging procedure, montages, the signal to noise ratio and spatial filters all of them are non-physiological factors that affect the stability of the shape of the P300 ERP.We tested a preliminary approach to assess if the morphological shape of the P300 of the averaged signal can be stabilized by applying different alignments of the stacked segments (see Figure 2) and we verified that there is a better performance when a correct segment alignment is applied. We applied Dynamic Time Warping (DTW) (Casarotto et al., 2005) to automate the alignment procedure but we were unable to find a substantial improvement. Further work to study the stability of the shape of the P300 signature component needs to be addressed.

The second problem is the amplitude variation of the P300. We propose a solution by standardizing the signal, shown in Equation (2). It has the effect of normalizing the peak-to-peak amplitude, moderating its variation. It has also the advantage of reducing noise that was not reduced by the averaging procedure. It is important to remark that the averaged signal variance depends on the number of segments used to compute it (Van Drongelen, 2006). The standardizing process converts the signal to unit signal variance which makes it independent of the number ka of signals averaged. Although this is initially an advantageous approach, the standardizing process reduces the amplitude of any significant P300 complex diminishing its automatic interpretation capability.

To further extend the capabilities of this method, it would be desirable to implement a multichannel version. The straightforward extension of concatenating the obtained descriptors results in high dimensional feature vector, while other variants that merge descriptors per channel may diminish the mutual information between different channels. Hitherto variants using color versions of SIFT (Van De Sande et al., 2010), where different color bands are mapped to electrode channels, have been explored without substantial success.

In our opinion, the best benefit of the presentedmethod is that a closer collaboration of the field of BCI with physicians can be fostered (Chavarriaga et al., 2017), since this procedure intent to imitate human visual observation. Automatic classification of patterns in EEG that are specifically identified by their shapes like K-Complex, Vertex Waves, Positive Occipital Sharp Transient (Hartman, 2005) are a prospect future work to be considered. We are currently working in unpublished material analyzing K-Complex components that could eventually provide assistance to physicians to locate these EEG patterns, specially in long recording periods, frequent in sleep research (Michel and Murray, 2012). Additionally, it can be used for artifact removal which is performed on many occasions by visually inspecting signals. This is due to the fact that the descriptors are a direct representation of the shape of signal waveforms. In line with these applications, it can be used to build a database (Chavarriaga et al., 2017) of quantitative representations of waveforms and improve atlases (Hartman, 2005), which are currently based on qualitative descriptions of signal shapes.

*4. DISCUSIÓN*

*Entre otras aplicaciones de las interfaces cerebro-computadora, el objetivo de la disciplina es brindar asistencia en la comunicación a las personas afectadas por enfermedades neurodegenerativas, que son las más propensas a población se beneficie de los sistemas BCI y el procesamiento y análisis de EEG.*

*En este trabajo, se presenta un método para extraer una métrica objetiva de la forma de onda de los gráficos de señales de EEG. Se expone su uso para implementar una aplicación BCI Speller válida basada en P300. Además, su validez se evalúa utilizando un conjunto de datos público de pacientes con ELA y un conjunto de datos propio de sujetos sanos. Se verificó que este método tiene un desempeño mejorado en la identificación de letras que otros métodos que procesan las señales en una estrategia de canal por canal, e incluso tiene un desempeño comparable con otros métodos como SWLDA o SVM, que utilizan una característica multicanal. Además, este método tiene la ventaja de que las formas de las formas de onda se pueden analizar de forma objetiva. Observamos que la forma del componente P300 es más estable en los canales occipitales, donde el rendimiento para identificar letras es mayor. Además, verificamos que las firmas ALS P300 son estables en comparación con las de sujetos sanos.*

*Creemos que el uso de descriptores basados ​​en histogramas de orientación de gradiente, presentados en este trabajo, también puede utilizarse para derivar una métrica de forma en el espacio de las señales P300 que puede complementar otras métricas basadas en el dominio del tiempo como las definidas por Mak. et al. (2012). Es importante notar que el análisis de las formas de las formas de onda generalmente se realiza con un enfoque cualitativo basado en la inspección visual (Sellers et al., 2006), y una metodología complementaria que ofrezca una métrica cuantitativa será beneficiosa para estos análisis rutinarios de la forma de onda. de ERP.*

*El objetivo de este trabajo es responder a la pregunta de si un componente P300 podría determinarse únicamente inspeccionando automáticamente sus formas de onda. Concluimos afirmativamente, aunque aún quedan dos cuestiones muy importantes:*

*Primero, la estabilidad del P300 en términos de su forma es crucial: el procedimiento de promediación, los montajes, la relación señal/ruido y los filtros espaciales son factores no fisiológicos que afectan la estabilidad de la forma del P300 ERP. probamos un enfoque preliminar para evaluar si la forma morfológica del P300 de la señal promediada se puede estabilizar aplicando diferentes alineaciones de los segmentos apilados (ver Figura 2) y verificamos que hay un mejor rendimiento cuando se aplica una alineación de segmento correcta. Aplicamos Dynamic Time Warping (DTW) (Casarotto et al., 2005) para automatizar el procedimiento de alineación, pero no pudimos encontrar una mejora sustancial. Se debe abordar el trabajo adicional para estudiar la estabilidad de la forma del componente de firma P300.*

*El segundo problema es la variación de amplitud del P300. Proponemos una solución estandarizando la señal, que se muestra en la Ecuación (2). Tiene el efecto de normalizar la amplitud pico a pico, moderando su variación. También tiene la ventaja de reducir el ruido que no fue reducido por el procedimiento de promedio. Es importante señalar que la varianza de la señal promediada depende del número de segmentos utilizados para calcularla (Van Drongelen, 2006). El proceso de estandarización convierte la señal en una varianza de señal unitaria, lo que la hace independiente del número ka de señales promediadas. Aunque este es inicialmente un enfoque ventajoso, el proceso de estandarización reduce la amplitud de cualquier complejo P300 significativo, lo que disminuye su capacidad de interpretación automática.*

*Para ampliar aún más las capacidades de este método, sería deseable implementar una versión multicanal. La extensión directa de concatenar los descriptores obtenidos da como resultado un vector de características de alta dimensión, mientras que otras variantes que combinan descriptores por canal pueden disminuir la información mutua entre diferentes canales. Hasta ahora, se han explorado sin éxito variantes que utilizan versiones en color de SIFT (Van De Sande et al., 2010), donde se asignan diferentes bandas de color a los canales de los electrodos.*

*En nuestra opinión, el mayor beneficio del método presentado es que se puede fomentar una colaboración más estrecha del campo de BCI con los médicos (Chavarriaga et al., 2017), ya que este procedimiento intenta imitar la observación visual humana. La clasificación automática de patrones en EEG que se identifican específicamente por sus formas como K-Complex, Vertex Waves, Positive Occipital Sharp Transient (Hartman, 2005) son un trabajo futuro prospectivo a considerar. Actualmente estamos trabajando en material inédito analizando los componentes del K-Complex que eventualmente podrían ayudar a los médicos a localizar estos patrones de EEG, especialmente en largos períodos de registro, frecuentes en la investigación del sueño (Michel y Murray, 2012). Además, se puede utilizar para la eliminación de artefactos, lo que se realiza en muchas ocasiones mediante la inspección visual de las señales. Esto se debe al hecho de que los descriptores son una representación directa de la forma de las formas de onda de la señal. De acuerdo con estas aplicaciones, se puede utilizar para construir una base de datos (Chavarriaga et al., 2017) de representaciones cuantitativas de formas de onda y mejorar atlas (Hartman, 2005), que actualmente se basan en descripciones cualitativas de formas de señales.*

ETHICS STATEMENT

Participants are recruited voluntarily and the experiment is conducted anonymously in accordance with the declaration of Helsinki published by the World Health Organization. No monetary compensation is handed out and all participants agree and sign a written informed consent. This study is approved by the Departamento de Investigación y Doctorado, Instituto Tecnológico de Buenos Aires (ITBA).