Texto

Descripción generada automáticamente

**INSTITUTO TECNOLÓGICO DE BUENOS AIRES – ITBA**

**ESCUELA DE INGENIERÍA Y GESTIÓN**

*Análisis del rendimiento en el algoritmo de machine learning que emula ondas ERP P300 usado en un experimento con interfaces cerebro computador en pacientes con ELA (Esclerosis Lateral Amiotrófica).*

**AUTOR: Chavez Montaño, Alexander. (Leg. Nº 506218)**

**TUTOR: Ramele, Rodrigo**

**Trabajo final presentado para la obtención del título de especialista en ciencia de datos.**

**BUENOS AIRES**

**Segundo cuatrimestre, 2023**

*A mis dos emes: Macu y Marco.*



**Tabla de contenido**

1. Introducción … 5

2. Estado del arte … 6

3. Definición del problema … 8

4. Justificación del estudio … 9

5. Alcances del trabajo y limitaciones … 10

6. Hipótesis … 11

7. Objetivos … 12

8. Metodologías a usar … 13

Técnicas

Herramientas

9. Análisis exploratorio de datos … 15

9.1 La enfermedad.

9.2 El P300 Speller y el oddball paradigm (paradigma del bicho raro)

9.3 ¿Qué es un ERP (Event Related Potential)?

9.4 Los datasets y las señales

9.4.1 El ERPTemplate

9.4.2 P300-Datset

9.4.2.1 Estructura

9.5 drugsignal, getstims y getlabels.

10. Testeo del algoritmo … 21

11. Resultados

12. Referencias-Bibliografía … 5

**Abstract**

Este trabajo complementa la investigación previa realizada en los experimentos descritos en el artículo *EEG Waveform Analysis of P300 ERP with Applications to Brain Computer Interfaces* [3] en pacientes con ELA (Esclerosis Lateral Amiotrófica): parte del resultado de dicho trabajo fue la de investigar algoritmos de machine learning que emulan ondas ERP P300: tomamos ese trabajo como punto de partida para realizar un análisis exploratorio tanto del objeto de estudio como de las herramientas computacionales disponibles, para luego realizar un abanico de pruebas que arrojen distinta *performance* y nos permitan proponer mejoras en la preconfiguración de dicho algoritmo.

**1. Introducción**

La electroencefalografía es una de las herramientas clínicas que, a lo largo de las últimas décadas, se ha convertido en uno de los principales métodos para obtener imágenes en tiempo real del comportamiento cerebral de manera no invasiva, portátil y móvil más usado en el ambiente médico [1]. Dentro de la electroencefalografía tenemos un conjunto de ondas con distintas características que varían en sus propiedades físicas como amplitud o frecuencia, como también en el origen y la ubicación en las distintas zonas del cerebro. La onda P300 se obtiene de ubicar un canal en el lóbulo parietal y su comportamiento es reactivo debido a estímulos esperados pero infrecuentes relacionados con actos cognitivos.

La electroencefalografía , sin embargo, está expuesta a alteraciones no deseadas en sus resultados, ya que, por más controlado, preciso y consistente que sea el ambiente donde se realiza el experimento o la toma de muestra, estaremos sujetos a factores fuera de nuestro control. Esta problemática se suele afrontar generando ambientes de pruebas donde se pueda recrear la situación con la mayor fidelidad posible. Los experimentos pasados y éste trabajo integrador se cimentan en la base de datasets sintéticos, artificiales, con los que se simulan respuestas de ondas ERP P300 a partir de electroencefalogramas reales, con resultados previamente conocidos, a fin de trabajar en la performance de algoritmos que logren resultados con mejoras en el tiempo.

Los métodos y los procedimientos cuantitativos para automatizar la decodificación de ondas EEG como la P300 se basa en EEG no invasivo [2]. Sin embargo, los métodos de la decodificación de señales, basadas en detección de formas de onda, y además con algoritmos de machine learning son relativamente escasos.

Se pretende entonces, a través de este trabajo, darle continuidad a la investigación previa realizada en los experimentos descritos en el artículo EEG Waveform Analysis of P300 ERP with Applications to Brain Computer Interfaces [3] en pacientes con ELA (Esclerosis Lateral Amiotrófica): en éste trabajo se realiza un análisis exploratorio de electroencefalogramas llamados pasivos; pacientes que participaron del experimento pero desconociendo las reglas de interacción con el *speller*, explicado más adelante en profundidad. En una etapa posterior se “inyectan” potenciales P300 en los lugares donde sabemos de antemano suceden los eventos, y realizamos modificaciones en las propiedades de las ondas que arrojan distinta *performance* para permitirnos obtener mejoras en la preconfiguración del algoritmo.

**2. Estado del arte**

Los experimentos descritos en el en el artículo *EEG Waveform Analysis of P300 ERP with Applications to Brain Computer Interfaces* [3] sirvieron como base para este proyecto. Si bien es necesario revisar todo el artículo en mención, damos un alcance y limitaciones de este trabajo específicamente capítulo 5; el estado del arte es entonces, el experimento descrito en el artículo.

En el capítulo 3 del artículo; *Materiales y Métodos*, específicamente el punto 3.6; *Protocolo experimental*, dan la información necesaria para enmarcar los detalles del experimento cuyo objetivo central es el de evaluar el rendimiento de los algoritmos que reconocen la forma de onda P300, obtenida después de promediar segmentos de señal.

El rendimiento es evaluado mediante el procesamiento de un conjunto de datos pseudo-reales en dos modalidades, una llamada *pasiva* y otra *activa*, en alusión a la información de la que los pacientes disponen al momento del experimento. ***Las pruebas de rendimiento de este trabajo final integrador complementan las realizadas en la modalidad pasiva, al realizar modificaciones en latencia y amplitud del componente P300 de distintas magnitudes***. Los experimentos se realizan mediante la evaluación fuera de línea de la identificación del personaje de una aplicación BCI Speller basada en Visual P300.

El paradigma BCI usado en los experimentos del artículo es el Farwell and Donchin P300 Speller[12]. En él, se usa un dispositivo de traducción de pensamientos donde envía comandos a una computadora en forma de letras seleccionadas, similar a escribir en un teclado virtual. El paradigma y el P300 Speller están descritos en el capítulo 9.2 de este trabajo.

La generación de flujo de EEG, es decir, la recopilación de la señal de los pacientes mencionada en el punto 3.6.1 del artículo, describe su estado de salud, la cantidad de pacientes, edad, de acuerdo a las normativas dictadas por la organización mundial de la salud. Incluso la disposición del paciente con respecto al dispositivo y las ubicaciones de los electrodos junto con la referencia, además de los dispositivos utilizados y la frecuencia de muestreo de la onda establecida en 250Hz.

Además, el protocolo experimental está compuesto por una cantidad de ensayos específica con el fin de deletrear una cantidad de palabras con letras predefinidas. Tienen secuencias para la matriz de 6 columnas por 6 filas con intervalos y pausas establecidos.

El preprocesamiento de la señal son las bases para la generación de las ondas que serán el objeto de estudio de este proyecto: se extraen, se filtran para eliminar ruido y también se descartan variaciones que estén por fuera de los parámetros esperados. El experimento dura alrededor de 1400s y el resultado final es una traza EEG con 4200 secciones marcadas donde 3500 de ellos están etiquetados como verdaderos y los 700 restantes como falsos. Esta información, al estar estandarizada, se almacena en formatos preestablecidos que pueden ser usados con la librería de python MNE descrita en el capítulo 8.

En cuanto a la participación de las personas en el experimento, a cuatro de los ocho se les indica que miren pasivamente la pantalla parpadeante sin centrarse en alguna letra en particular. No reciben ninguna información adicional en la pantalla. Ninguno de ellos tiene o tuvo alguna experiencia con un dispositivo BCI. Se entrega un cuestionario al final del experimento con preguntas sobre cómo se sintió el participante durante el mismo, sin dar más detalles. Los cuatro participantes restantes realizan una tarea de ortografía en la que el monitor de la computadora resalta la letra objetivo, que es la que el sujeto necesita enfocar. A lo largo de la duración de la prueba, la letra objetivo actual se informa en la parte inferior de la pantalla.

La modalidad activa no será tenida en cuenta en este trabajo.

La modalidad pasiva consiste en que, para la mitad de los pacientes, se realiza un ensamblado de la traza EEG final: se toman plantillas reales de un ERP P300 de un conjunto de datos público y se superponen con el la traza EEG generado en dichos pacientes, los cuales fueron aquellos de los que no estaban informados con detalle del experimento.

**3. Definición del problema**

La continuidad de los proyectos de investigación es vital para obtener avances y mejoras en los resultados de los experimentos. Particularmente, las investigaciones en el tratamiento de señales electroencefalográficas con modelos de machine learning son un campo de estudio relativamente nuevo y con poca disponibilidad de datos, lo que genera obstáculos que impiden realizar experimentos comparativos a gran escala. En el caso de los experimentos descritos en el artículo *EEG Waveform Analysis of P300 ERP with Applications to Brain Computer Interfaces* [3] en pacientes con ELA (Esclerosis Lateral Amiotrófica), su continuidad se encontraba pausada por razones ajenas a este documento. Es posible que, una vez reanudada estas investigaciones, se puedan dar saltos posteriores con experimentos comparativos a gran escala.

Dentro del artículo descrito, se usó un algoritmo de machine learning que ensambla electroencefalogramas de pacientes pasivos; pacientes que participaron del experimento desconociendo las reglas de interacción con los equipos, con potenciales P300 en los lugares donde sabemos de antemano que suceden los eventos evocados. Es necesario analizar y testear dicho algoritmo para que la investigación continúe.

El EDA (Análisis Exploratorio de Datos, por sus siglas en inglés) y modificaciones en las propiedades de las ondas obtenidas de ese ensamble nos permitirán ampliar el abanico de resultados que arrojan distinta *performance* en los resultados. Esto permitirá proponer mejoras en la preconfiguración del algoritmo.

**4. Justificación del estudio**

Se pretende darle continuidad al proyecto de investigación basados en los experimentos descritos en el artículo *EEG Waveform Analysis of P300 ERP with Applications to Brain Computer Interfaces* [3] en pacientes con ELA (Esclerosis Lateral Amiotrófica).

Este trabajo final integrador permitirá ampliar la gama de resultados del algoritmo mencionado en este documento. La salida del algoritmo es una onda compuesta por un electroencefalograma y potenciales ERP. Se pretende, modificando las propiedades de dicha onda, ofrecer una gama de resultados que sirvan de complemento para evaluar y proponer mejoras en la *performance* del algoritmo. A priori, y debido a las particularidades del experimento, las nuevas ondas generadas tendrán efectos distintos. No se espera verificar si el comportamiento final sigue siendo el mismo.

**5. Alcance del trabajo y limitaciones**

Los experimentos realizados en el artículo *EEG Waveform Analysis of P300 ERP with Applications to Brain Computer Interfaces* [3] en pacientes con ELA (Esclerosis Lateral Amiotrófica) permiten un sinnúmero de estudios y experimentos posteriores. El presente trabajo está enfocado únicamente al testeo del algoritmo destinado a generar una onda compuesta de manera sintética. Se espera que, con los resultados obtenidos al modificar deliberadamente las propiedades de la onda, podamos tener un espectro de respuestas que nos permitan porponer mejoras para trabajos posteriores.

**6. Hipótesis**

Los resultados en los experimentos descritos en el artículo *EEG Waveform Analysis of P300 ERP with Applications to Brain*, específicamente en los EEGs de pacientes pasivos, donde la traza es modificada de manera artificial inyectándole ERP P300, arrojan resultados de predicción cercanos al XX%. Se pretende, mediante modificaciones en la latencia y en la amplitud del potencial mencionado, verificar el rendimiento del algoritmo y comprobar un aumento en su predicción.

**7. Objetivos**

**General.**

Darle continuidad a los experimentos e investigaciones previas realizadas en el Instituto Tecnológico de Buenos Aires ITBA, analizando y obteniendo resultados del rendimiento en el algoritmo de machine learning que emula ondas ERP P300 usado en un experimento con interfaces cerebro computador en pacientes con ELA (Esclerosis Lateral Amiotrófica).

**Específicos.**

* Realizar un recorrido en la mayor cantidad de información disponible sobre los objetos de estudio: por un lado, el potencial de eventos evocado P300. Por otro, electroencefalogramas de pacientes sanos y pacientes con ELA (Esclerosis Lateral Amiotrófica), y por último todo el conjunto de señales que se obtienen producto de una interfaz cerebro computador (BCI por sus siglas en inglés) vinculado a un instrumento de comunicación llamado P300 speller.
* Completar un análisis exploratorio de datos (EDA por sus siglas en inglés) con los datasets usados en los experimentos, a través de las herramientas computacionales disponibles: librerías de python especializadas en el manejo de datos y electroencefalografía.

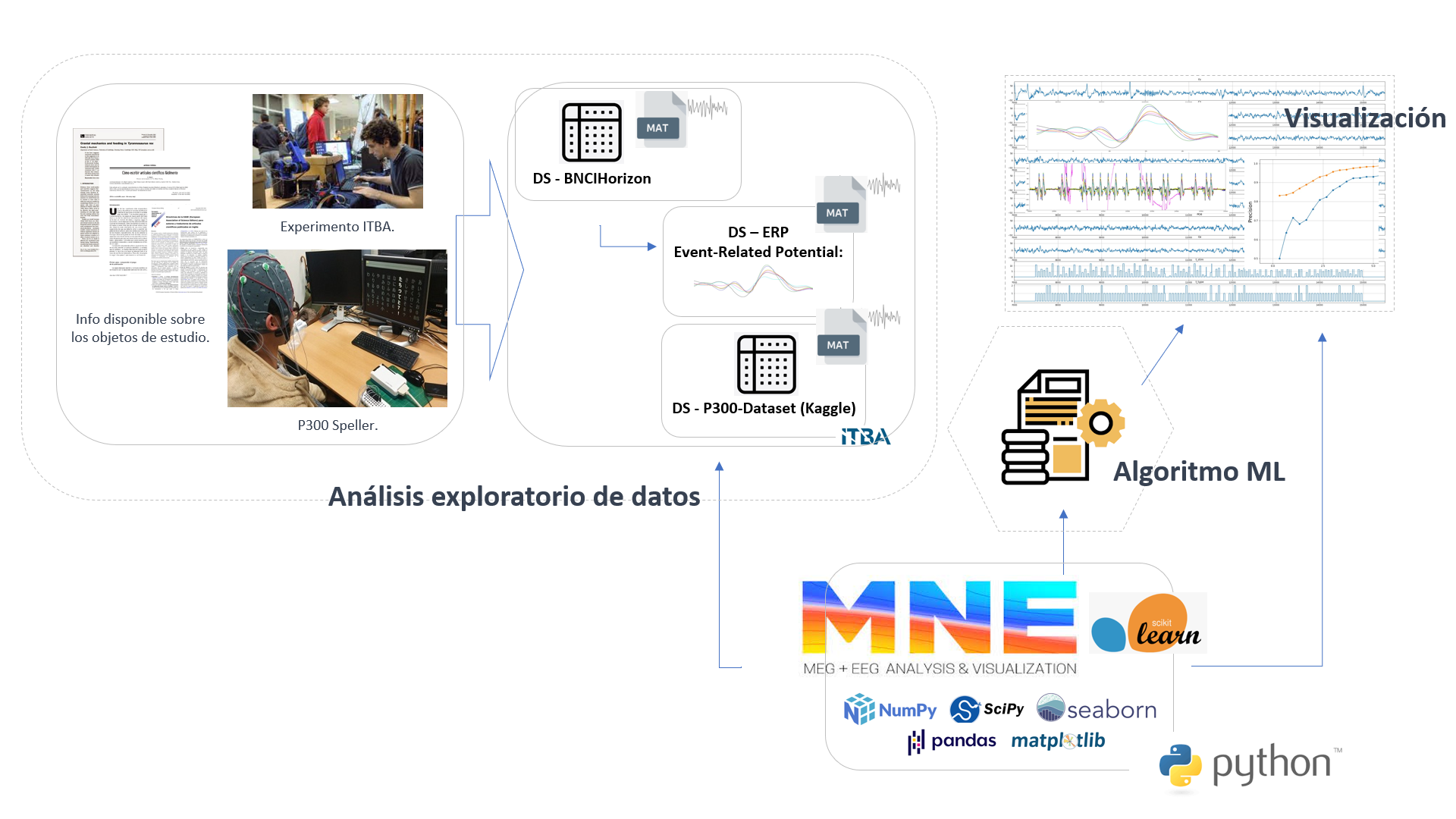
[BNCI Horizon 2020: 8. Speller P300 with ALS patients (008-2014)](http://bnci-horizon-2020.eu/database/data-sets).

[ITBA. P300 dataset of 8 healthy subjects.](https://www.kaggle.com/datasets/rramele/p300samplingdataset?resource=download)

* Modificar, en distintos rangos, las propiedades de la onda compuesta de manera sintética y verificar el rendimiento del algoritmo [*drugsignal.py*](https://github.com/faturita/python-nerv/blob/master/DrugSignal.py)

Presentar todos los resultados obtenidos de las modificaciones en las propiedades de la onda compuesta y proponer mejoras en la preconfiguración del algoritmo.

**8. Metodologías a usar**



La estructura de este trabajo de integrador está pensada en función de los objetivos en el punto anterior. Primero se ofrecerá un análisis exploratorio de datos, no solamente de los datos en sí, sino también del contexto del experimento realizado el ITBA mencionado anteriormente. Este análisis exploratorio contendrá los datasets usados en los experimentos:

* [BNCI Horizon 2020: 8. Speller P300 with ALS patients (008-2014)](http://bnci-horizon-2020.eu/database/data-sets).
* [ITBA. P300 dataset of 8 healthy subjects.](https://www.kaggle.com/datasets/rramele/p300samplingdataset?resource=download)

En segundo lugar, habrá una descripción de la preparación de los datos para ser modelados, complementado con las herramientas computacionales disponibles: librerías de python especializadas en el manejo de datos y electroencefalografía.

En tercer lugar, se aplicará el modelo con las distintas variaciones en las propiedades de la onda compuesta de manera sintética. Y en cuarto y último lugar, se mostrarán los gráficos de los resultados obtenidos de las modificaciones más representativas en las propiedades de la onda compuesta, acompañados con una propuesta de mejoras en la preconfiguración del algoritmo.

Las herramientas y/o librerías usadas en este proyecto se pueden clasificar en dos; matemáticas y de electroencefalografía, todas concentradas en el lenguaje de programación Python. Los electroencefalogramas que fueron usados son *Matlab files*: archivos de extensión .mat en versiones con funcionalidades de almacenamiento de *arrays* de n dimensiones de hasta 100.000.000 elementos por arreglo y 231 bytes por variable.

Dentro de las librerías matemáticas se encuentran ***NumPy*** para permitir el manejo de arreglos grandes y multidimensionales, ***SciPy*** con módulos para optimización , álgebra lineal , integración , interpolación , funciones especiales , FFT , procesamiento de señales e imágenes, entre otros, ***Matplotlib*** y ***Seaborn*** para visualización de los datos, y ***Pandas*** para la manipulación y análisis tanto de los archivos usados como fuentes de datos como para los distintos procesos intermedios en el análisis exploratorio. La librería destinada al machine learning es ***Scikit-learn***: dispone de algoritmos de clasificación, regresión y agrupamiento , que incluyen support vector machine, random forest, k -means y DBSCAN. Tiene la versatilidad para interactuar con el resto de librerías mencionadas anteriormente.

Por otra parte se usó la librería **MNE**: permite la exploración, visualización y análisis de datos neurofisiológicos humanos: MEG, EEG, sEEG, ECoG, NIRS, entre otros.

**9. Análisis Exploratorio de Datos (EDA)**

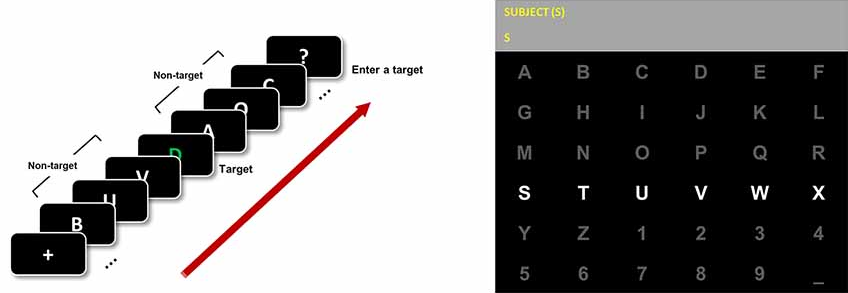
**9.1 La enfermedad.**

La esclerosis lateral amiotrófica o ELA, es una enfermedad degenerativa de las neuronas en el cerebro, el tronco cerebral y la médula espinal que controlan el movimiento de los músculos voluntarios. En la ELA, las células nerviosas (neuronas) motoras se desgastan o mueren y ya no pueden enviar mensajes a los músculos. Con el tiempo, esto lleva a debilitamiento muscular, espasmos e incapacidad para mover los brazos, las piernas y el cuerpo. La afección empeora lentamente. Cuando los músculos en la zona torácica dejan de trabajar, se vuelve difícil o imposible respirar.

En pacientes con ELA de etapas intermedias y avanzadas, es necesario el uso de dispositivos tecnológicos para la comunicación, como el P300 Speller.

**9.2 El P300 Speller y el *oddball paradigm* (paradigma del bicho raro).**

El P300 Speller es un dispositivo que conforma uno de las BCI (Brain Computer Interfaces) más usados en este tipo de aplicaciones. Su funcionamiento se acoge al paradigma del bicho raro: al usuario/paciente se le presenta una matriz de caracteres de 6 por 6 (ver figura) de manera intermitente, sucesiva y aleatoria. La tarea del usuario/paciente será enfocar su atención en los caracteres de una palabra prescrita por el investigador; es decir, un carácter a la vez. Cuando éstas contienen el carácter deseado (es decir, una fila particular y una columna determinada) se registra el potencial evocado P300 en el registro del EEC.

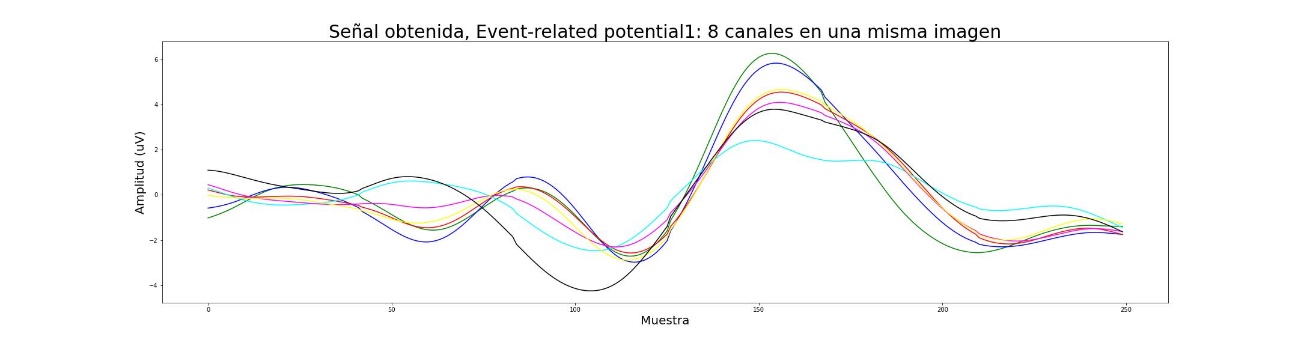
****

**9.3 ¿Qué es un ERP (Event Related Potential)?**

De forma paralela, es necesario explicar qué es una señal P300. La palabra *evocada* es clave: en medicina, se refiere a una actividad que puede ser detectada sincrónicamente después de una cantidad específica de tiempo después del inicio de un estímulo. Si estamos a la espera de que un computador nos dé una señal visual y nos la da, en nuestro cerebro ocurre un evento de este tipo. En términos médicos *es una actividad inducida* .

La onda P300 es entonces, una señal en el cerebro con amplitud positiva relacionada con eventos. Para esta investigación, los eventos serán aquellos provocados bajo el *paradigma del bicho raro*: El sujeto detecta un estímulo "objetivo" ocasional en un tren regular de estímulos estándar.

La onda P300 solo ocurre si el sujeto participa activamente en la tarea de detectar los objetivos. Su amplitud varía con la improbabilidad de los objetivos. Su latencia varía con la dificultad de discriminar el estímulo objetivo de los estímulos estándar. Una latencia pico típica cuando un sujeto adulto joven hace una discriminación simple es de 300 ms.  
En pacientes con capacidad cognitiva disminuida, el P300 es más pequeño y más tardío que en sujetos normales de la misma edad.



El ERP P300 Puede tener una duración de 400ms y su amplitud puede alcanzar los 10µV [4]. Se desconoce el origen intracerebral de la onda P300 y su papel en la cognición no se comprende con claridad. El P300 puede tener múltiples generadores intracerebrales, con el hipocampo y varias áreas de asociación de la neocorteza contribuyendo al potencial registrado en el cuero cabelludo. La onda P300 puede representar la transferencia de información a la conciencia, un proceso que involucra muchas regiones diferentes del cerebro [10].

**9.4 Los datasets y las señales.**

Ya mencionado anteriormente, los electroencefalogramas que fueron usados son *Matlab files*: archivos de extensión .mat en versiones con funcionalidades de almacenamiento de *arrays* de n dimensiones de hasta 100.000.000 elementos por arreglo y 231 bytes por variable. El dataset de [BNCI Horizon](http://bnci-horizon-2020.eu/database/data-sets), el 008-2014, contiene un grupo completo de potenciales evocados P300 registrados con la interfaz cerebro computador BCI2000[11]. De éste dataset obtendremos el *ERPTemplate*. El otro, el [P300-Dataset](https://www.kaggle.com/datasets/rramele/p300samplingdataset?resource=download), está conformado por 8 EEGs de donde se extraerán algunos para realizar las pruebas del algoritmo. Todos los datasets están basados en el paradigma Farwell y Donchin [12] mencionado en el punto 9.2. Las señales usadas en este trabajo serán descritas a continuación:

**9.4.1 El ERPTemplate**

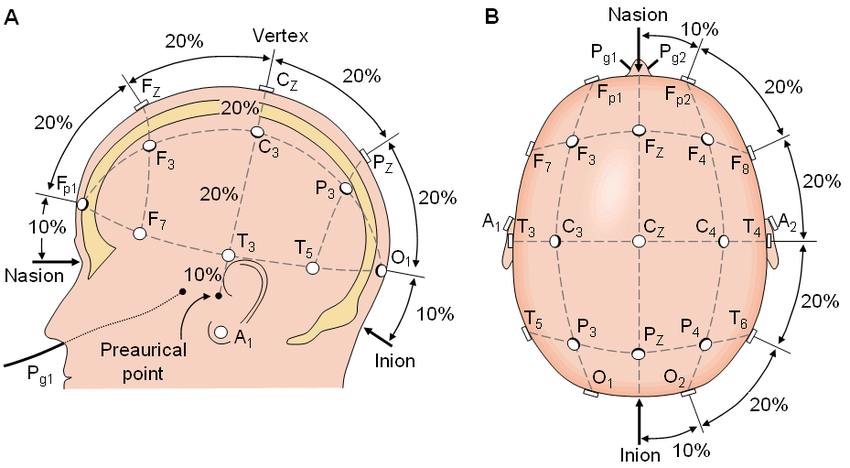
Si bien la descripción de qué es un ERP está en el punto 9.3, es importante mencionar que éste ERP se extrae artificialmente del dataset de [BNCI Horizon](http://bnci-horizon-2020.eu/database/data-sets) (008-2014) para ser “inyectado” a una señal EEG del [P300-Dataset](https://www.kaggle.com/datasets/rramele/p300samplingdataset?resource=download) con el fin de crear una señal sintética que nos permita realizar las modificaciones de latencia y amplitud en donde quede empalmado dicho ERP. El proceso está descrito en el punto 9.5. En el archivo [*a\_analisis\_ERPTemplate.ipynb*](https://github.com/alexchavez1980/repo_tesis/blob/main/a_analisis_ERPTemplate.ipynb)hay un análisis más en detalle de la estructura y propiedades de la onda.

* + 1. **P300-Dataset**

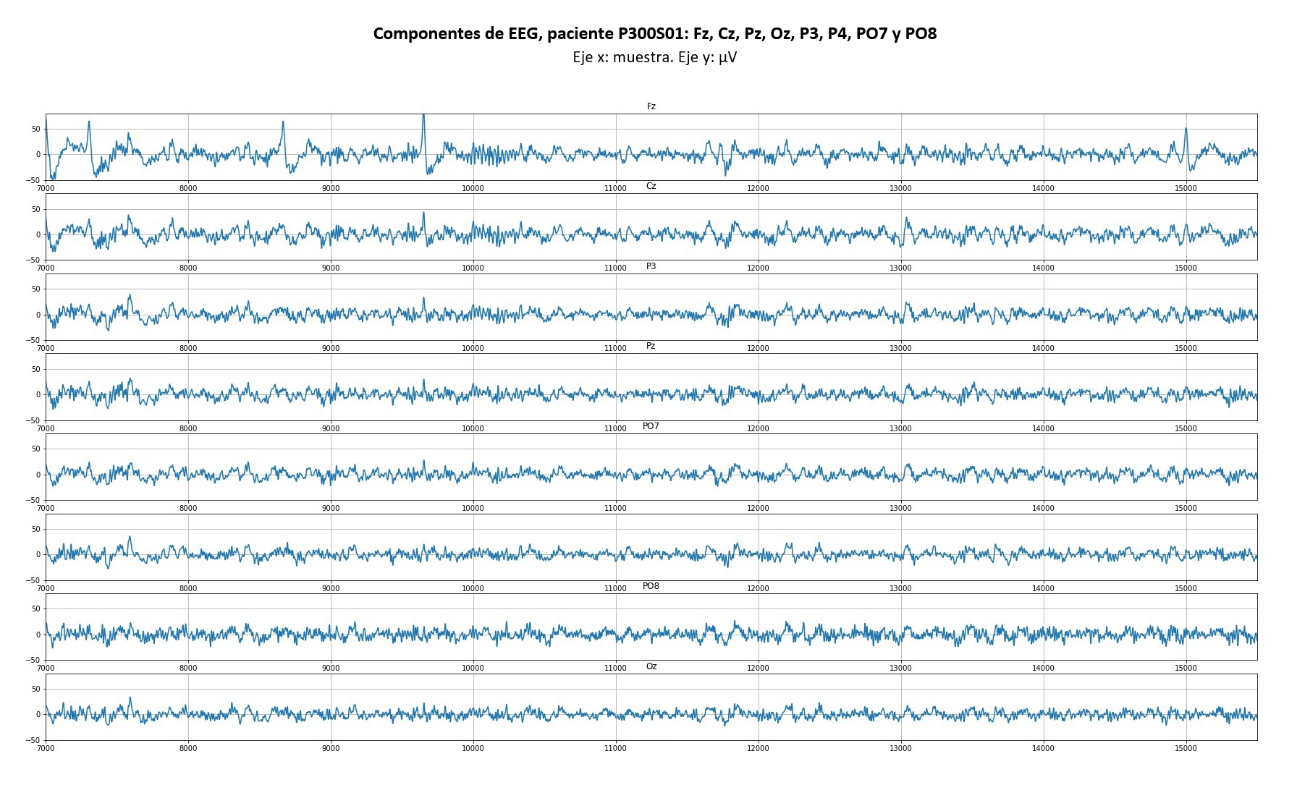
El [P300-Dataset](https://www.kaggle.com/datasets/rramele/p300samplingdataset?resource=download) está conformado por 8 EEGs distribuidos en dos grupos según la modalidad del experimento: pasivos P300S01,02,03,06 y activos P300S04, 05, 07 y 08. Este trabajo se enfoca en los pacientes pasivos: las trazas de EEG donde se superponen las pantillas ERP son de los pacientes que **no se enfocan en ninguna letra en particular**. Todo está allí, excepto el componente P300 ERP. Es por esto que se utiliza la información de marcadores para localizar los segmentos verdaderos donde se debería encontrar el P300, y esas ubicaciones de tiempo se utilizan para superponer la forma de onda de ERP extraída.

**9.4.2.1 Estructura**

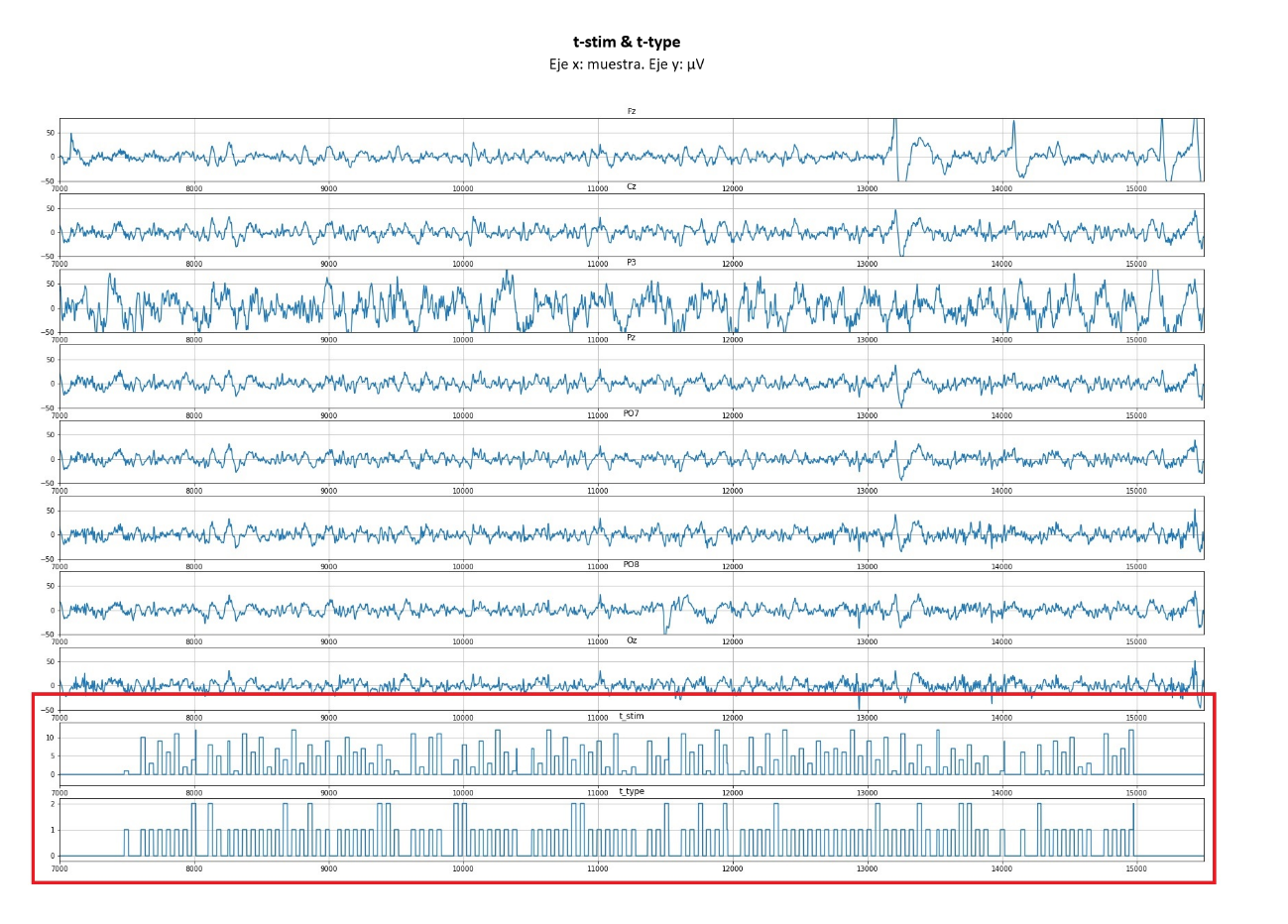
El electroencefalograma obtenido se puede observar de dos maneras distintas: Una con cada una de sus componentes por separado, y la otra como una sola señal compuesta. Se utilizan en las ubicaciones Fz, Cz, Pz, Oz, P3, P4, PO7 y PO8 según el sistema internacional 10-20 para la colocación de los electrodos extracraneales.



La referencia se establece en el lóbulo de la oreja derecha y la tierra está preestablecida como la posición AFz. La frecuencia de muestreo se establece en 250 Hz. El archivo [a\_analisis\_P300XX.ipynb](https://github.com/alexchavez1980/repo_tesis/blob/main/a_analisis_P300XX.ipynb) es el análisis de dichas señales. (DEPURAR EL ARCHIVO PARA QUE QUEDE CON LA INFO PROLIJA)



El archivo *.mat* que contiene las señales obtenidas no solamente contiene las señales del electroencefalograma, sino las marcas del experimento. Es con ellas que se identifican dónde se encuentran los tiempos en donde se han mostrado los caracteres.



**-----------------------------------------------------**

*3.6.4. Experimentos*

*Los experimentos son los siguientes:*

*Experimento 1—Rendimiento de identificación de letras: el rendimiento de identificación de letras de cada uno de estos métodos en el conjunto de datos pseudo-real generado artificialmente. La piscina de 70 P300 ERP formas de onda, obtenidas del mismo sujeto en la modalidad pasiva o de cada sujeto en la modalidad activa se utilizan para componer la onda P300 artificial en el conjunto de datos pseudo-reales. Las plantillas se seleccionan al azar.*

*Experimento 2: Ruido de latencia: en lugar de superponer los ERP P300 sobre el trazo de EEG en las ubicaciones exactas donde se sitúan los inicios de los estímulos, se agrega un retraso de latencia artificial. El valor rezagado se toma de una distribución uniforme U(0, 0.4) [s] que va de 0 a 0.4 del tamaño del segmento 1 s [74].*

*Experimento 3: Ruido de amplitud del componente: la amplitud del componente principal P3b del La plantilla ERP se modifica aleatoriamente. Este componente se define para estar ubicado desde el estímulo inicio entre 148 ms hasta 996 ms, que es alrededor de 840 ms de largo. Este elemento de forma de onda, multiplicado por un factor de ganancia, se resta de la plantilla original. Este factor de ganancia entre 0 y 1 se extrae de una distribución uniforme U(0, 1). Además, esta forma de onda sustraída es multiplicado por una ventana gaussiana con un soporte de la misma longitud [75]. Esto evita agregar cualquier discontinuidad en la señal generada artificialmente.*

*Todos estos experimentos se ejecutan utilizando el procedimiento de validación cruzada dividiendo la letra para deletrear dos juegos, conservando la estructura de los ensayos de identificación de letras. Las letras de ortografía están codificadas mientras se conserva el orden y el grupo de cada secuencia de intensificación.*

*Finalmente, el rendimiento en la identificación de letras para estos mismos métodos se evalúa ejecutando una simulación BCI fuera de línea en el conjunto de datos IIb de la competencia BCI II (2003) [76]. El protocolo de este*

*El conjunto de datos es muy similar al que se utilizó para obtener el conjunto de datos pseudo-reales. La frecuencia de muestreo de este conjunto de datos es 240, el número de letras es 73 donde los primeros 42 se usan para crear la plantilla diccionario para todos los métodos y los 31 restantes se utilizan para probar la tasa de reconocimiento de caracteres actuación. Además, en este conjunto de datos, el número de secuencias numéricas de intensificación disponibles es 15. Se agrega el método de clasificación Support Vector Machine SVM con un kernel lineal para comparación como control utilizando una característica f construida mediante la normalización de la señal en cada canal [77].*

***Este método ha demostrado su eficacia en la decodificación de P300 en varios concursos BCI [78].***

3.6.5. Classification

The same classification algorithm based on k-nearest neighbors is used for all the methods [79]. The experimental protocol used to generated the pseudo-real dataset used in the experiments 1 to 3 is composed of 35 trials to spell 7 words of 5 letters each. Each trial is composed of 10 intensification sequences of the 6 columns and 6 rows of the Speller Matrix. Fifteen trials are used to build the dictionary of templates, extracting the averaged EEG segments for the row and column that already contain the P300 ERP, hence shielding 30 different templates per channel. Figure 5 shows the set of templates while using the first 15 trials of the dataset.

Described algorithms produce a feature f for each averaged EEG segment. The aim of the classification procedure is to identify for the remaining 20 trials which of the 6 features f that were obtained for row intensification, labeled by f1, ..., 6g, and which of the 6 features for column intensification, named f7, ..., 12g are the ones that elicited the P300 response on the averaged EEG segment. The row number of the matrix can be obtained by doing with qi being the set of k-nearest neighbors of the feature fu with u varying from 1 to 6. The parameter k represents the number of neighbors chosen from the dictionary of templates. The column can be obtained in the same way, Thus, the letter identification performance can be obtained by measuring the accuracy channel-by-channel at identifying the correct letter on the matrix, coordinated by roˆw and coˆ l.

*3.6.5. Clasificación*

*El mismo algoritmo de clasificación basado en k-vecinos más cercanos se utiliza para todos los métodos [79]. El protocolo experimental utilizado para generar el conjunto de datos pseudo-reales utilizados en los experimentos 1 a 3 es compuesto por 35 ensayos para deletrear 7 palabras de 5 letras cada una. Cada prueba se compone de 10 intensificaciones secuencias de las 6 columnas y 6 filas de la matriz de ortografía. Se utilizan quince ensayos para construir el diccionario de plantillas, extrayendo los segmentos de EEG promediados para la fila y la columna que ya contienen el ERP P300, por lo tanto, protegen 30 plantillas diferentes por canal. La Figura 5 muestra el conjunto de plantillas mientras utiliza los primeros 15 ensayos del conjunto de datos.*

*Los algoritmos descritos producen una característica f para cada segmento de EEG promediado. El objetivo de la*

*El procedimiento de clasificación es identificar para los 20 ensayos restantes cuál de las 6 características f que se obtuvieron para la intensificación de fila, etiquetados por f1, ..., 6g, y cuál de las 6 características para la columna*

*intensificación, denominada f7, ..., 12g son las que provocaron la respuesta P300 en el EEG promediado segmento. El número de fila de la matriz se puede obtener haciendo siendo qi el conjunto de k-vecinos más cercanos de la característica fu con u variando de 1 a 6. El parámetro k representa el número de vecinos elegidos del diccionario de plantillas. La columna puede ser obtenidos de la misma manera,*

*Por lo tanto, el rendimiento de identificación de letras se puede obtener midiendo la precisión canal por canal para identificar la letra correcta en la matriz, coordinada por roˆw y coˆ l.*

**------------------------------------------------------**

1. Introducción 100%

2. Estado del arte 100%

3. Definición del problema 100%

4. Justificación del estudio 100%

5. Alcances del trabajo y limitaciones 100%

6. Hipótesis 100%

7. Objetivos 100%

8. Metodología 100%

Técnicas

Herramientas

9. Análisis exploratorio de datos 100%

9.1 La enfermedad. 100%

9.2 El P300 Speller y el oddball paradigm (paradigma del bicho raro). 100%

9.3 ¿Qué es un ERP (Event Related Potential)? 100%

9.4 Los datasets y las señales. 100%

9.4.1 El ERPTemplate 100%

9.4.2 P300-Dataset 100%

9.4.2.1 Estructura 100%

10. Testeo del algoritmo 0%

11. Resultados 0%

12. Referencias-Bibliografía 70%

**10. Testeo del algoritmo**

**11. Resultados**

**12. Referencias / Bibliografía**

1. Hartman, A.L. Atlas of EEG Patterns; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2005.

2. Guger, C.; Allison, B.Z.; Lebedev, M.A. Introduction. In Brain Computer Interface Research: A State of the Art Summary 6; Springer: Cham, Switzerland, 2017; pp. 1–8.

3. Ramele, R.; Villar, A.J.; Santos, J.M.; EEG Waveform Analysis of P300 ERP with Applications to Brain Computer Interfaces: Computer Engineering Department, Instituto Tecnológico de Buenos Aires (ITBA), Published: 16 November 2018.

4. Rao, R. P. N. (2013). Brain-Computer Interfacing: An Introduction. New York, NY: Cambridge University Press.

5. Skoog, D.A.; West, D.M.; Holler, F.J.; Crouch, S.R. Analytical Chemistry: An Introduction; Saunders College, Publishing: Philadelphia, PA, USA, 2000.

6. Owens, T.J.; Zandt, G.; Taylor, S.R. Seismic evidence for an ancient rift beneath the Cumberland Plateau, Tennessee: A detailed analysis of broadband teleseismic P waveforms. J. Geophys. Res. Solid Earth 1984, 89, 7783–7795.

7. Stockman, G.; Kanal, L.; Kyle, M. Structural pattern recognition of carotid pulse waves using a general waveform parsing system. Commun. ACM 1976, 19, 688–695.

8. Hartman, A.L. Atlas of EEG Patterns; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2005.

9. Picton, T.W.; The P300 Wave of the Human Event-Related Potential; Journal of clinical neurophysiology, American electroencephalographic society, 1992.

**10. The P300 wave of the human event-related potential https://pubmed.ncbi.nlm.nih.gov/1464675/**

[11] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, e J. R. Wolpaw, «BCI2000: a general-purpose brain-computer interface (BCI) system», IEEE Trans. Biomed. Eng., vol. 51, n. 6, pagg. 1034–1043, 2004

[12] L. A. Farwell e E. Donchin, «Talking off the top of your head: toward a mental prosthesis utilizing eventrelated brain potentials», Electroencephalogr. Clin. Neurophysiol., vol. 70, n. 6, pagg. 510–523, 1988.

27. Luck, S.J. An Introduction to the Event-Related Potential Technique; USA; MIT Press: Cambridge, MA, USA,

2005; Volume 78, p. 388.

**Archivos .mat** [**https://la.mathworks.com/help/matlab/import\_export/mat-file-versions.html**](https://la.mathworks.com/help/matlab/import_export/mat-file-versions.html)

|  |  |
| --- | --- |
| **Paper** | **Título** |
| brainsci-08-00199.pdf | EEG Waveform Analysis of P300 ERP with Applications to Brain Computer Interfaces |
| P300 dataset of 8 healthy subjects.pdf | P300 dataset of 8 healthy subjects.pdf |
| fncom-13-00043.pdf | Histogram of Gradient Orientations of Signal Plots Applied to P300 Detection |
| UMA-BCI Speller.pdf | UMA-BCI SPELLER: PLATAFORMA DE COMUNICACIÓN DE FÁCIL CONFIGURACIÓN BASADA EN EL BCI2000 P300 SPELLER |
| P300 Speller with patients with ALS | P300 Speller with patients with ALS |
| Picton 1992 | The P300 wave of the human Event-Related- Potential. |
| vucic2020.pdf | P300 jitter latency, brain-computer interface and amyotrophic lateral sclerosis |
| tesis\_n3966\_Gambini.pdf | Modelos de segmentación basados en regiones y contornos activos aplicados a imágenes de radar de apertura sintética |
| fnins-07-00267.pdf | MEGandEEGdataanalysiswithMNE-Python |

------------------------------------------------------------------------------------------------------------

------------------------------------------------------------------------------------------------------------

------------------------------------------------------------------------------------------------------------

*Marco conceptual*

*: define con precisión los conceptos centrales*

*del dominio del problema (Palabras Clave).*

*Marco teórico*

*: explica las teorías sobre el*

*Estado del Arte*

*: sintetiza el estado actual que se encuentre en*

*otras investigaciones o situaciones similares a la estudiada.*

*Antecedentes*

*: describe aquellos antecedentes que contemplen*

*diversas formas de resolución de ese problema , que se hayan*

*construido en forma previa a la investigación.*

------------------------------------------------------

**2.2 Marco conceptual.**

Agregar algunos concpetos y definiciones que están en los papers pero que también son útiles para el lector.Describir el experimento. BlaLa esclerosis lateral amiotrófica o ELA, es una enfermedad

------------------------------------------------------

*DIRECTAMENTE VOY A INCLUIR EL EXPERIMENTO.*

*El estado del arte es el estado de la cuestión, dónde está parada la investigación hasta ése momento.*

*La electroencefalografía es una de las herramientas clínicas que, a lo largo de las últimas décadas, se ha convertido en uno de los principales métodos para obtener imágenes en tiempo real del comportamiento cerebral de manera no invasiva, portátil y móvil más usado en el ambiente médico* ***[8]****. Sin embargo, está expuesta a alteraciones no deseadas en sus resultados, ya que, por más controlado, preciso y consistente que sea el ambiente donde se realiza el experimento o la toma de muestra, estaremos sujetos a que el objeto de estudio, en este caso es el ser humano, incurrirá en desconcentración o desenfoque al momento de hacer las pruebas y esto modificar la respuesta esperada.*

*Dentro de la electroencefalografía tenemos un conjunto de ondas con distintas características que varían en sus propiedades físicas como amplitud o frecuencia, como también en el origen y la ubicación en las distintas zonas del cerebro. La onda P300 se obtiene de ubicar un canal en el lóbulo parietal y su comportamiento es reactivo debido a estímulos esperados pero infrecuentes relacionados con actos cognitivos.*

*Los métodos y los procedimientos cuantitativos para automatizar la decodificación de ondas EEG como la P300 se basa en EEG no invasivo* ***[2]****. Sin embargo, los métodos de la decodificación de señales, basadas en detección de formas de onda, y además con algoritmos de machine learning, es relativamente escaso.*

Son ocho participantes sanos: saludables, visión normal o corregida en la normalidad, sin antecedentes de trastornos neurológicos, entre 20 y 40 años de edad, y los datos de EEG se recopilan en una sola sesión de grabación. Cada sujeto está sentado en una silla cómoda, con su vista alineada con una pantalla de computadora ubicada a un metro frente a él/ella. El manejo y procesamiento de los datos y estímulos se realiza mediante la plataforma OpenVibe.