

UNIVERSIDAD NACIONAL DEL LITORAL



DOCTORADO EN INGENIERÍA

# Algoritmos para interfaces cerebro-computadora en paradigmas relacionados con el habla

Nicolás Nieto

FICH  
FACULTAD DE INGENIERÍA  
Y CIENCIAS HÍDRICAS

INTEC  
INSTITUTO DE DESARROLLO TECNOLÓGICO  
PARA LA INDUSTRIA QUÍMICA

CIMEC  
CENTRO DE INVESTIGACIÓN DE  
MÉTODOS COMPUTACIONALES

$\text{sinc}(i)$   
INSTITUTO DE INVESTIGACIÓN EN SEÑALES  
SISTEMAS E INTELIGENCIA COMPUTACIONAL

Tesis de Doctorado **2022**





UNIVERSIDAD NACIONAL DEL LITORAL  
Facultad de Ingeniería y Ciencias Hídricas  
Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional

# ALGORITMOS PARA INTERFACES CEREBRO-COMPUTADORA EN PARADIGMAS RELACIONADOS AL HABLA

Nicolás Nieto

Tesis remitida al Comité Académico del Doctorado como  
parte de los requisitos para la obtención del grado de  
DOCTOR EN INGENIERÍA  
Mención en Inteligencia Computacional, Señales y Sistemas  
de la  
UNIVERSIDAD NACIONAL DEL LITORAL

**2022**

Comisión de Posgrado, Facultad de Ingeniería y Ciencias Hídricas, Ciudad Universitaria, Paraje "El Pozo", S3000,  
Santa Fe, Argentina.







UNIVERSIDAD NACIONAL DEL LITORAL  
Facultad de Ingeniería y Ciencias Hídricas  
Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional

# ALGORITMOS PARA INTERFACES CEREBRO-COMPUTADORA EN PARADIGMAS RELACIONADOS CON EL HABLA

Nicolás Nieto

**Lugar de Trabajo:**

$\text{sinc}(i)$

Instituto de Señales, Sistemas e Inteligencia Computacional  
Facultad de Ingeniería y Ciencias Hídricas  
Universidad Nacional del Litoral

**Director:**

Dr. Hugo Leonardo Rufiner

$\text{sinc}(i)$ , CONICET-UNL

**Co-director:**

Dr. Rubén Daniel Spies

IMAL, CONICET-UNL

**Jurado Evaluador:**

Dr. Carlos Muravchik

Dr. Pablo Granitto

Dr. Rodrigo Echeveste

UNLP  
CIFASIS, CONICET-UNR  
 $\text{sinc}(i)$ , CONICET-UNL





**UNIVERSIDAD NACIONAL DEL LITORAL**  
**Facultad de Ingeniería y Ciencias Hídricas**

Santa Fe, 23 de Marzo de 2022.

Como miembros del Jurado Evaluador de la Tesis de Doctorado en Ingeniería titulada *"Algoritmos para Interfaces Cerebro-Computadora en paradigmas relacionados con el habla"*, desarrollada por el Ing. Nicolás NIETO, en el marco de la Mención "Inteligencia computacional, señales y sistemas", certificamos que hemos evaluado la Tesis y recomendamos que sea aceptada como parte de los requisitos para la obtención del título de Doctor en Ingeniería.

La aprobación final de esta disertación estará condicionada a la presentación de dos copias encuadradas de la versión final de la Tesis ante el Comité Académico del Doctorado en Ingeniería.

-----  
Dr. Pablo Granitto

-----  
Dr. Carlos Muravchik

-----  
Dr. Rodrigo Echeveste

Santa Fe, 23 de Marzo de 2022.

Certifico haber leído la Tesis, preparada bajo mi dirección en el marco de la Mención "Inteligencia computacional, señales y sistemas" y recomiendo que sea aceptada como parte de los requisitos para la obtención del título de Doctor en Ingeniería.

-----  
Dr. Rubén Spies  
Codirector de Tesis

-----  
Dr. Leonardo Rufiner  
Director de Tesis



Dr. JOSÉ LUIS MACOR  
SECRETARIO DE POSGRADO  
Facultad de Ingeniería y Cs. Hídricas

Universidad Nacional del Litoral  
Facultad de Ingeniería y  
Ciencias Hídricas  
  
Secretaría de Posgrado

Ciudad Universitaria  
C.C. 217  
Ruta Nacional Nº 168 - Km. 472,4  
(3000) Santa Fe  
Tel: (54) (0342) 4575 229  
Fax: (54) (0342) 4575 224  
E-mail: posgrado@fich.unl.edu.ar



## **DECLARACIÓN LEGAL DEL AUTOR**

Esta Tesis ha sido remitida como parte de los requisitos para la obtención del grado académico de Doctor en Ingeniería ante la Universidad Nacional del Litoral y ha sido depositada en la Biblioteca de la Facultad de Ingeniería y Ciencias Hídricas para quedar a disposición de sus lectores bajo las condiciones estipuladas por el reglamento de la mencionada Biblioteca.

Se permiten citas breves de esta Tesis sin la necesidad de un permiso especial, en la suposición de que la fuente sea correctamente citada. El portador legal del derecho de propiedad intelectual de la obra concederá por escrito solicitudes de permiso para la citación extendida o para la reproducción parcial o total de este manuscrito.



# Agradecimientos

A la familia con la que se nace y la que se elige.





# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Objetivo general . . . . .	3
1.3. Objetivos específicos . . . . .	3
<b>2. Interfaces Cerebro-Computadora</b>	<b>5</b>
2.1. Definiciones y características . . . . .	5
2.2. Sistema nervioso y lenguaje . . . . .	5
2.3. Mediciones de la actividad cerebral . . . . .	6
2.3.1. Técnicas invasivas . . . . .	7
2.3.2. Técnicas no invasivas . . . . .	8
2.4. Paradigmas . . . . .	9
2.4.1. P300 . . . . .	10
2.4.2. Potenciales evocados de estado estacionario . . . . .	10
2.4.3. Potenciales corticales lentos . . . . .	10
2.4.4. Imaginería Motora . . . . .	11
2.4.5. Paradigmas relacionados al habla . . . . .	11
2.5. Desafíos de las Interfaces Cerebro-Computadora . . . . .	12
2.5.1. Escasa cantidad de datos . . . . .	13
2.5.2. Variabilidad inter e intra sujeto . . . . .	13
2.5.3. Alta dimensionalidad y redundancia . . . . .	13
2.5.4. Baja relación señal ruido . . . . .	13
2.5.5. Aplicaciones en situaciones reales . . . . .	14
2.5.6. Aspectos éticos . . . . .	14
<b>3. Datos</b>	<b>15</b>
3.1. Datos de electroencefalografía . . . . .	15
3.1.1. Bases de datos de P300 . . . . .	15
3.1.2. Bases de datos de imaginería motora . . . . .	16
3.1.3. Bases de datos de habla imaginada . . . . .	16
3.1.4. Bases de datos de habla interna . . . . .	17
<b>4. Maquinas de aprendizaje extremo para la clasificación de habla imaginada</b>	<b>21</b>
4.1. Generalidades de la inteligencia artificial y su uso en interfaces cerebro-computadora . .	21
4.2. Máquinas de Aprendizaje Extremo . . . . .	22
4.2.1. Podado basado en relevancia de máquinas de aprendizaje extremo . . . . .	22
4.2.2. Podado estadístico pre-entrenamiento de máquinas de aprendizaje extremo . . .	27
<b>5. Análisis y clasificación del habla interna</b>	<b>31</b>
5.1. Análisis global del habla interna . . . . .	31
5.1.1. Potenciales relacionados a eventos . . . . .	31
5.1.2. Análisis espacial y espectral . . . . .	31
5.2. Clasificación de señales de habla interna . . . . .	34

5.2.1. Análisis preliminar de factibilidad . . . . .	34
5.2.2. Análisis de las características relevantes . . . . .	35
<b>6. Transporte óptimo aplicado a imaginería motora</b>	<b>39</b>
6.1. Transporte óptimo y adaptación de dominio aplicado a ICCs . . . . .	39
6.1.1. Transporte óptimo con inversión de dominios . . . . .	39
6.1.2. Resultados y comparaciones con métodos del estado del arte . . . . .	40
6.1.3. Transporte óptimo con información del clasificador . . . . .	40
<b>7. Conclusiones</b>	<b>45</b>
<b>Anexos</b>	<b>55</b>
A. Contribuciones . . . . .	57
B. Anexo I: Thinking out loud: An open-access EEG-based BCI dataset for inner speech recognition EEG-BCI . . . . .	57
C. Anexo II: Extreme Learning Machine design for dealing with unrepresentative features .	87
D. Anexo III: Podado pre-entrenamiento de maquinas de aprendizaje extremo . . . . .	113
E. Anexo IV: Preliminary feasibility analysis of inner speech as a control paradigm for brain-computer interfaces . . . . .	119
F. Anexo V: Transfer learning based on optimal transport for motor imagery brain-computer interfaces . . . . .	126

# Índice de tablas

2.1. Comparación de los principales métodos de medición de la actividad cerebral . . . . .	9
3.1. Número de ensayos dividido por participante, paradigma y clase. . . . .	20
4.1. Resultados de desempeño obtenidos para la base de datos de habla imaginada . . . . .	25
4.2. Resultados de desempeño obtenido para la base de datos de P300 . . . . .	26
6.1. Comparación de métodos de adaptación de dominio sobre datos propios . . . . .	40
6.2. Comparación de métodos de adaptación de dominio sobre datos públicos . . . . .	41
6.3. Tiempo computacional requerido por los métodos de adaptación de dominio . . . . .	41



# Índice de figuras

1.1. Cantidad de trabajos publicados en el area de interfaces-cerebro computadora por año . . . . .	2
2.1. Diagrama en bloques de una interfaz cerebro-computadora . . . . .	6
2.2. Lóbulos y áreas principales del cerebro . . . . .	7
2.3. Dispositivo “The Link” de la empresa Neuralink . . . . .	8
2.4. Homúnculo motor y somatosensorial . . . . .	11
3.1. Posiciones utilizadas de electrodos . . . . .	18
3.2. Estructura de cada estímulo . . . . .	19
4.1. Distribución de características informativas . . . . .	24
4.2. Evolución del desempeño de clasificación con respecto a los distinto número de caracte- rísticas no representativas y neuronas ocultas . . . . .	24
4.3. Valor absoluto del peso asignado por el modelo a la neurona falsa añadida . . . . .	25
4.4. Tiempo de entrenamiento de los métodos evaluados con respecto a la cantidad máximas de neuronas . . . . .	26
4.5. Histogramas típicos de activación de neuronas ocultas. . . . .	28
4.6. Desempeño de clasificación de señales de habla imaginada. . . . .	29
4.7. Desempeño de clasificación de los métodos combinados sobre datos artificialmente ge- nerados . . . . .	30
5.1. Promedio temporal de todos los trials disponibles . . . . .	32
5.2. Diferencias espaciales y espectrales en la actividad de los paradigmas habla interna y habla pronunciada . . . . .	33
5.3. Diferencias espaciales y espectrales en la actividad de los paradigmas habla interna y condición visualizada . . . . .	33
5.4. Distribución del desempeño de clasificación entre habla interna y condición visualizada. . . . .	35
5.5. Distribución del desempeño de clasificación entre habla interna y habla pronunciada . . . . .	36
5.6. Distribución del desempeño de clasificación para cada banda de frecuencia. . . . .	37
6.1. Distribución de masas en el conjunto fuente . . . . .	42



# Resumen

La inteligencia artificial, y particularmente el aprendizaje maquina, ha experimentado un inmenso crecimiento en los últimos años, adentrándose y abarcando cada vez más campos de aplicación. Por supuesto, las ciencias de la salud no han sido la excepción y se han beneficiado enormemente de estos avances. Desde el diagnóstico asistido o automático de enfermedades como el Parkinson y el Alzheimer, hasta lograr importantes mejoras en la calidad y cantidad de herramientas disponibles para personas con discapacidad, la inteligencia artificial ha conseguido mejorar enormemente la calidad de vida de las personas.

En este sentido, el área de las interfaces cerebro-computadora ha sido uno de los campos con mayor desarrollo gracias a los avances de estas tecnologías. Estas interfaces representan una vital herramienta para personas con discapacidades severas, cuyo único medio posible de comunicación con su entorno es a través de estas tecnologías. Aunque el uso de estas interfaces todavía dista de alcanzar un uso masivo, ha permitido mejorar considerablemente la calidad de vida de cientos de personas que dependen de ellas, principalmente para comunicarse. El mayor obstáculo de estas tecnologías es la baja tasa de transferencia de información, requiriendo de varios segundos, e incluso minutos, para poder comunicar una sola palabra. Sumado a esto, estas interfaces presentan un desempeño de clasificación muy variable y no siempre permiten un control natural del entorno. Afortunadamente, el interés en el área ha crecido exponencialmente desde sus comienzos y son varios los grupos de investigación alrededor del mundo que intentan día a día hacer estas herramientas más accesibles y confiables.

Hace ya algunos años los paradigmas de control relacionados con el habla han cobrado especial relevancia dentro de la comunidad científica. En el contexto de una interfaz cerebro-computadora, un paradigma se define como la actividad mental ejecutada por el paciente y para ser codificada por la computadora. Mediante el uso de paradigmas relacionados con el habla, se podría, a futuro, ejecutar una acción solo con pensar en ella. De esta manera, estos paradigmas representarían una forma de control veloz, natural y de uso continuo. No obstante, el lenguaje es uno de los fenómenos más complejos y menos comprendido del ser humano, por lo que su uso como paradigma de control todavía presenta varios problemas para la comunidad.

Si bien son múltiples las causas por las que estos paradigmas no pueden, todavía, ser utilizados para controlar interfaces cerebro computadora, los principales motivos son la falta de datos públicos disponibles y la ausencia de algoritmos específicamente diseñados para tal fin. Para avanzar hacia la solución de estos problemas, en la presente tesis doctoral se han realizado diversos tendientes aportes en dirección a mejorar el uso de los paradigmas relacionados al habla como paradigma de control de las interfaces cerebro computadora.

El primer aporte es la adquisición y publicación de una base de datos de diez participantes adquirida bajo un paradigma relacionado al habla: el habla interna. En segundo lugar, y en conjunto con los datos, se desarrolló una librería que permite el simple manejo y procesamiento de estos datos, facilitando su reutilización.

Por otro lado, se desarrollaron dos métodos de podado, aplicados a las estructuras conocidas como Máquinas de Aprendizaje Extremo, para mejorar las características de estas redes y permitir que estos métodos se adapten mejor a los requerimientos específicos de una interfaz cerebro computadora.

Por último, se propusieron nuevas técnicas de transporte óptimo y adaptación de dominio, que permiten obtener un mayor provecho de los muy escasos y heterogéneos datos disponibles.





# Abstract

Artificial intelligence, and particularly machine learning, has experienced an immense growth in recent years, involving more fields of application every day. Health sciences have not been an exception and have enormously benefited from these advances. From the assisted or automatic diagnosis of diseases such as Parkinson's and Alzheimer's, to the vast improvements in the quality and quantity of tools available to people with disabilities, artificial intelligence has greatly improved people's quality of life.

In this sense, the area of brain-computer interfaces has been one of the fields with the greatest developments thanks to the advances of these technologies. These interfaces represent a vital tool for people with severe disabilities, whose only possible means of communication with the environment is through these technologies. Although the use of these interfaces is still far from reaching massive use, it has improved considerably the quality of life of hundreds of people who depend on them, mainly for communication. The strongest obstacle of these technologies is the low information transfer rate, requiring several seconds, or even minutes, to communicate a single word. In addition, these interfaces have a highly variable classification performance and do not always allow a natural control of the environment. Fortunately, interest in the area has grown exponentially since its beginning and several research groups around the world are trying to make these tools more accessible and reliable every day.

For some years now, speech-related control paradigms have gained special relevance within the scientific community. In the context of a brain-computer interface, a paradigm is defined as the mental activity executed by the patient and to be encoded by the computer. By using speech-related paradigms, one could, in the future, execute an action just by thinking about it. In this way, these paradigms would represent a form of control that is fast, natural and of continuous use. However, language is one of the most complex and least understood human phenomena and its use as a control paradigm still presents multiple problems for the community.

Although there are many reasons why these paradigms cannot, as yet, be used to control brain-computer interfaces, the main reasons are the lack of publicly available data and the absence of algorithms specifically designed for this purpose. In response to these problems, several contributions have been made in this doctoral thesis in the direction of improving the use of speech-related paradigms for controlling brain-computer interfaces.

The first major contribution is the acquisition and publication of a database of ten participants acquired under a speech-related paradigm, inner speech. Secondly, and in conjunction with the data, a library was developed that allows the simple handling and processing of this data, allowing easy reuse.

On the other hand, two pruning methods were developed, applied to the structures known as Extreme Learning Machines, to improve the characteristics of these networks and allow these methods to better adapt to the specific requirements of a brain-computer interface.

Finally, new techniques for optimal transport and domain adaptation were proposed, which allows to get more out of the very scarce and heterogeneous data available.



# Sección 1

## Introducción

### 1.1. Motivación

En Argentina, más de cien mil personas sufren cada año heridas en accidentes de tránsito, siendo el 66% de estas personas menores de 44 años<sup>1</sup>. Esta no sólo es una de las principales causas de muerte en jóvenes, sino que también representa una importante causa de discapacidad en todos los grupos etarios. Entre las complicaciones más severas causadas por estos accidentes se encuentra la pérdida total o parcial de las capacidades motoras, de comunicación y control del entorno, mayormente causadas por lesiones cerebrales o espinales. Sumado a los accidentes de tránsito, existen distintas enfermedades, como la esclerosis lateral amiotrófica (ELA), los accidentes cerebrovasculares (ACV), la parálisis cerebral y las distrofias musculares, que pueden impedir al cerebro controlar los músculos por sus vías normales. En Estados Unidos se estima que estas enfermedades afectan a alrededor de dos millones de personas (Wolpaw *et al.*, 2002). Según el Registro Nacional de Personal con Discapacidad Argentino, en 2016 se emitieron 225.865 certificados únicos de discapacidad, de los cuales el 28% correspondía a discapacidades motoras.

En los casos más severos de discapacidad los individuos pierden completamente el control voluntario de sus músculos, produciéndose lo que se denomina “síndrome de enclaustramiento” (locked-in syndrome, LIS). En algunos casos, los pacientes pueden llegar a vivir varios años en esta condición, presentando un estado cognitivo intacto (Rousseau *et al.*, 2015). Si bien no hay datos sobre la prevalencia de este síndrome en Argentina, según distintos estudios internacionales el LIS tendría una prevalencia de entre unos 0.73 a 0.8 casos por cada cien mil habitantes (Brumberg *et al.*, 2010; Pels *et al.*, 2019). Es importante destacar que, para estos pacientes, su calidad de vida se encuentra fuertemente relacionada con su capacidad de comunicación (Rousseau *et al.*, 2015).

Afortunadamente, grandes avances se realizan a diario en torno a los sistemas aumentativos alternativos de comunicación (SAAC). Estos sistemas permiten generar nuevas vías de comunicación, no solo para los pacientes con LIS, sino también para aquellos que han perdido de forma parcial su capacidad de comunicarse e interactuar con su entorno. Los SAAC cubren un gran espectro de herramientas, desde el uso de simples pictogramas y gestos hasta tecnologías más avanzadas como las interfaces humano computadora. En este tipo de interfaces los pacientes son capaces de controlar deletreadores, sillas de ruedas, cursores o incluso brazos robóticos utilizando distintas señales captadas a partir de su cuerpo.

La electromiografía (EMG) es una de las señales más utilizadas, permitiendo medir pequeños restos musculares que el paciente puede controlar de forma voluntaria y confiable. Otra tecnología ampliamente explorada en los últimos años es el seguimiento ocular mediante imágenes. Esta técnica encontró un fuerte impulso gracias a los avances en visión computacional, principalmente por el desarrollo de las redes neuronales convolucionales, y al aumento en el poder de cómputo que permitió procesar y analizar imágenes en tiempo real (Voulodimos *et al.*, 2018).

---

<sup>1</sup> <https://www.argentina.gob.ar/seguridadvial/observatoriovialnacional/estadisticas-observatorio>

Si bien estas tecnologías permiten importantes avances y mejora en la calidad de vida de cientos de pacientes, todavía existe un grupo de pacientes donde estas tecnologías no pueden ser utilizadas. Es aquí donde las interfaces cerebro computadora (ICC o BCI por sus siglas en inglés) se convierten en valiosas herramientas. Este tipo de interfaces utilizan exclusivamente señales generadas por el cerebro para controlar distintos dispositivos y le permiten a los pacientes recuperar, aunque sea de forma muy limitada, su capacidad de comunicarse e interactuar con su entorno.

Aunque el campo de las ICCs es relativamente reciente, ha experimentado un crecimiento exponencial en su desarrollo desde sus inicios en los años 90, tal como muestra la Figura 1.1. Este creciente interés se vio fuertemente impulsado por distintos factores. En primer lugar, la cantidad de usuarios finales se ha incrementado año a año, con el aumento en la expectativa de vida y la gran cantidad de personas que presentan discapacidades motrices severas. En segundo lugar, los equipos de adquisición de señales cerebrales han mejorado significativamente su capacidad y calidad de adquisición, portabilidad, precio y usabilidad, permitiendo expandir su uso a aplicaciones domésticas y siendo cada vez más accesibles a los usuarios finales. Por otro lado, el constante desarrollo y especialización de las técnicas de procesamiento, análisis y clasificación de señales cerebrales, ha generado interfaces cada vez más rápidas, robustas y confiables. Por último, un mejor entendimiento de la neurofisiología de los fenómenos cerebrales ha permitido desarrollar nuevos paradigmas de comunicación entre personas y computadoras.

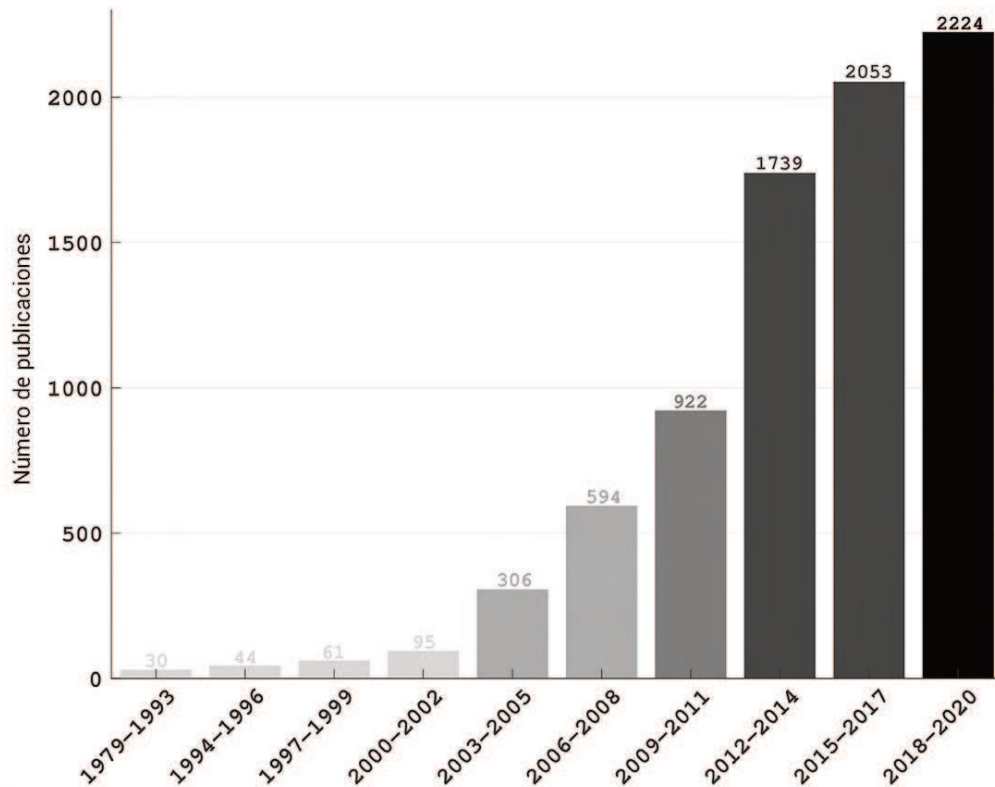


Figura 1.1: Cantidad de trabajos publicados en el area de interfaces-cerebro computadora por año. Adaptada de (Saha *et al.*, 2021).

Las ICC más difundidas son aquellas que utilizan señales de electroencefalografía (EEG), ya que estas señales se obtienen de forma no invasiva, portable, segura y con la ventaja de poseer una buena resolución temporal. Como contraparte, estas señales son no estacionarias, presentan una baja relación señal-ruido (signal to noise ratio, SNR), baja resolución espacial y son producidas por mecanismos

biológicos intrínsecamente no lineales.

Dentro de las interfaces cerebro-computadora basadas en electroencefalografía (ICC-EEG) existen distintos paradigmas de comunicación. Un paradigma se define como la actividad cerebral específica que un paciente realiza y que será aquella que se decodificará para ejecutar una acción (Abiri *et al.*, 2019). Algunos de los paradigmas de control más conocidos son el de potenciales evocados visuales P300 (McCane *et al.*, 2015), potenciales evocados visuales de estado estacionario (Allison *et al.*, 2008), imaginación motora (MI) (Ahn and Jun, 2015) y potenciales corticales lentos.

Estos paradigmas han permitido grandes avances en las ICC-EEG, no obstante, en algunas aplicaciones resultan muy lentos, requieren grandes esfuerzos por parte del paciente o no permiten interfaces naturales, limitando su uso a largo plazo. Para buscar una solución a estos problemas, hace aproximadamente una década se comenzaron a explorar nuevos paradigmas relacionados con el habla, principalmente el “habla imaginada” (Torres-García *et al.*, 2013; DaSalla *et al.*, 2009; Sereshkeh *et al.*, 2017) y el “habla interna” (Suppes *et al.*, 1997; D’Zmura *et al.*, 2009). En estos paradigmas se intenta decodificar la “voz interna” del paciente al pensar en una palabra o letra específica. Estos paradigmas presentan una gran naturalidad para controlar una ICC, ya que brindaría la posibilidad de ejecutar una orden sólo con pensar en ella. De esta forma sería posible mover un brazo robótico, una silla de rueda o un cursor en pantalla con solo pensar en la dirección deseada, tal como si se le estuviera dando una orden directa. Lamentablemente, la decodificación de estos paradigmas a través de señales de EEG es un problema abierto en la actualidad, al ser el lenguaje uno de los procesos más complejos que realiza el ser humano.

Para contribuir en la comprensión de los fenómenos neuronales del lenguaje y posibilitar su uso como paradigma de control de ICC, en la presente tesis se propone diseñar, desarrollar, formalizar y evaluar nuevos algoritmos de procesamiento y reconocimiento de señales de EEG, así como también la adquisición de nuevos datos de EEG adquiridos bajo la modalidad de habla interna.

## 1.2. Objetivo general

En la presente tesis doctoral se propone explorar, diseñar y desarrollar nuevos algoritmos de procesamiento y reconocimiento de señales de EEG, en el contexto de ICCs basadas en paradigmas relacionados con el habla.

## 1.3. Objetivos específicos

A continuación se detallan los objetivos específicos de la presente investigación:

- Diseñar algoritmos de aprendizaje maquina utilizando datos de EEG existentes. Generar discusiones teóricas y formalizaciones matemáticas de nuevas propuestas para luego compararlas con los métodos del estado del arte.
- Diseñar un protocolo de adquisición de datos de EEG bajo el paradigma de Habla Interna para su uso en una ICC.
- Generar una base de datos pública con señales de electroencefalografía, adquiridas bajo el paradigma de Habla Interna y grabadas con un equipo de alta densidad de canales.
- Proponer algoritmos de aprendizaje maquina sobre los nuevos datos adquiridos. En una primera instancia será posible utilizar los algoritmos diseñados anteriormente.
- Generar desarrollos teóricos y prácticos en el campo de la adaptación de dominio, que permitan mejorar la clasificación y adaptación de las interfaces a los usuarios finales.



## Sección 2

# Interfaces Cerebro-Computadora

### 2.1. Definiciones y características

Las interfaces cerebro-computadora (ICC o BCI por sus siglas en inglés) son sistemas que buscan transformar la actividad cerebral en comandos de control para llevar a cabo una acción deseada (Shih *et al.*, 2012). Una ICC posee cuatro etapas fundamentales presentadas en la Figura 2.1: adquisición de señales, extracción de características, clasificación y aplicaciones. La primera etapa consiste en captar, digitalizar y transmitir las señales generadas por el cerebro hacia la computadora. Existen diversas formas de medir la actividad cerebral, cada una con ventajas y desventajas que serán abordadas en forma detallada en las siguientes secciones.

Una vez digitalizadas y recibidas en la computadora, las señales son procesadas para extraer diferentes características. Técnicas de filtrado, eliminación de artefactos, reducción dimensional y transformaciones tiempo-frecuenciales han sido desarrolladas para extraer la mayor cantidad de información relacionada a la intención del paciente. Para el caso de las ICCs, estas características son principalmente espaciales, temporales y frecuenciales.

Las características obtenidas al final de esta etapa son utilizadas como entrada en la etapa de clasificación. En esta, se busca encontrar información en las características extraídas que permitan predecir la intención del usuario. A la salida de la etapa de clasificación se genera un comando que, idealmente, relaciona la señal cerebral adquirida con la intención del sujeto. Este comando puede ser utilizado luego para controlar distintas aplicaciones finales. Dentro de las más comunes se encuentran el manejo de cursores, deletreadores, neuroprótesis, sillas de rueda, exoesqueletos para rehabilitación o cualquier otro dispositivo del entorno que pueda responder a una orden digital.

Si bien no es una etapa dentro de la construcción de las interfaces, la retroalimentación que recibe el usuario es de vital importancia en las aplicaciones reales a largo plazo. Con ayuda de la retroalimentación, no solo la computadora es capaz reconocer las intenciones del paciente sino que este también aprende a modular de la mejor manera posible su actividad cerebral para que la interfaz pueda interpretarla.

### 2.2. Sistema nervioso y lenguaje

El cerebro es el órgano principal del sistema nervioso central y se encuentra protegido exteriormente por el cuero cabelludo, el pelo, el cráneo y recubierto por tres meninges: duramadre, piamadre y aracnoides. Anatómicamente, se compone de dos hemisferios subdivididos en cinco lóbulos cada uno: occipital, parietal, temporal, frontal y límbico o de la ínsula.

La unidad básica funcional del sistema nervioso es la neurona. Si bien las hay de distintos tipos y tamaños a lo largo de todo el sistema nervioso, todas se componen de un cuerpo o soma, encargado de integrar la información recibida por otras neuronas, y un axón, por donde la neurona transmite

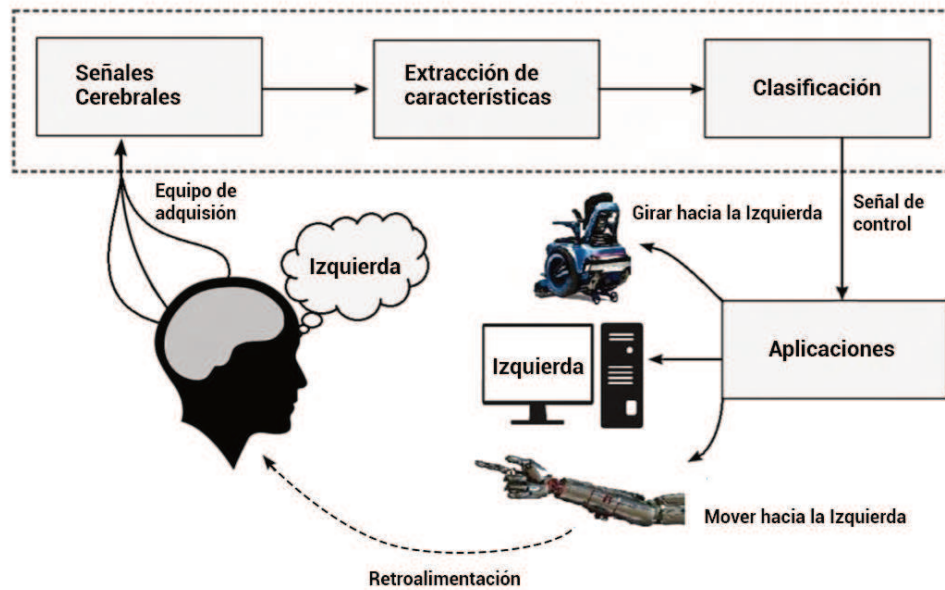


Figura 2.1: Diagrama en bloques de una interfaz cerebro-computadora y sus principales etapas.

la información hacia las demás. Las neuronas se comunican entre sí mediante señales electroquímicas, producidas en la sinapsis. Cuando el impulso nervioso llega hasta el final del axón se liberan neurotransmisores desde el botón presináptico. Estos neurotransmisores son captados por el botón postsináptico, generando un potencial de acción en esta neurona.

A nivel general, el cerebro ha evolucionado en gran parte para permitirle a los humanos la capacidad de generar el lenguaje. De hecho, gran parte del éxito evolutivo de los humanos se basó en esta capacidad. El lenguaje nos ha permitido aumentar enormemente la cantidad de información transmitida y recibida, respecto a otros tipos de comunicación. No sólo los aspectos anatómicos del tracto vocal evolucionaron para producir el lenguaje oral, sino que también, a diferencia de los chimpancés que comparten un tracto vocal similar, los humanos cuentan con la capacidad de poder reordenar distintos gestos motores para generar nuevas palabras (Lieberman, 2007). En la Figura 2.2 se detallan los principales centros relacionados con el lenguaje, junto con los principales lóbulos y referencias anatómicas del cerebro.

Las primeras teorías de lenguaje se basaron principalmente en los centros de Wernicke y Broca. El primero estaba encargado de la comprensión de las palabras mientras que el segundo se encargaba de reordenar los gestos motores para generar el habla. Las teorías más modernas sostienen que estos modelos son por lo menos incompletos, ya que con la ayuda de técnicas de neuroimágenes se pudo demostrar que distintas áreas subcorticales y basales del cerebro, incluido el cerebelo y los ganglios de la base, cumplen un rol fundamental en la generación del lenguaje. (Kandel *et al.*, 2000; Blank *et al.*, 2002; Lieberman, 2007; Brumberg *et al.*, 2010)

### 2.3. Mediciones de la actividad cerebral

A lo largo de los años se han desarrollado distintas tecnologías que permiten medir, de distintas formas, la actividad cerebral. Las primeras técnicas medían directamente el potencial eléctrico generado por las neuronas o su correspondiente campo magnético. Por otro lado, también fue posible medir indirectamente la actividad neuronal a través de la medición del flujo sanguíneo, donde las neuronas metabólicamente más activas aumentan su demanda de oxígeno y consecuentemente de sangre. En el



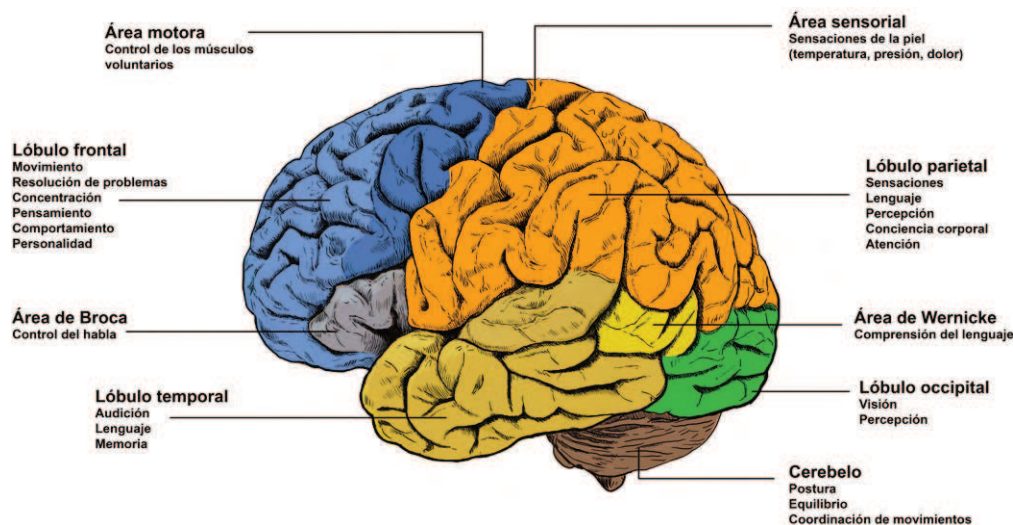


Figura 2.2: Lóbulos y áreas principales del cerebro. Fuente: Consejo Superior de Investigaciones Científicas CSIC (2021)

final de la sección la Tabla 2.1 resume de las principales técnicas de medición de actividad cerebral junto a sus características más relevantes.

### 2.3.1. Técnicas invasivas

Las técnicas de medición invasivas son aquellas que requieren de un procedimiento quirúrgico para colocar electrodos de medición. Utilizando microelectrodos implantados dentro del tejido cerebral es posible medir la actividad de una o de un pequeño grupo de neuronas (Single unit activity, SUA). Si bien esta técnica permite conocer exactamente cada uno de los disparos de un grupo de neuronas, pierde totalmente el comportamiento global del cerebro. Esta técnica es principalmente utilizada para experimentos neurocientíficos en animales. Sumando los potenciales de las neuronas registradas es posible obtener los potenciales de campo local (Local field potentials, LFP). Esta técnica también es utilizada para investigaciones neurocientíficas y permite ampliar el área de exploración a algunos cientos de neuronas.

Por otro lado, la electrocorticografía (ECoG) se realiza mediante la colocación de electrodos en la superficie del cerebro. Este procedimiento es comúnmente realizado en pacientes con epilepsia refractaria, donde las señales de ECoG permiten localizar, de la forma más precisa posible, el o los focos de actividad epiléptica. La técnica de ECoG capta la actividad conjunta de miles de neuronas al mismo tiempo. Distintos trabajos han estudiado el uso prolongado de los implantes de ECoG, concluyendo que los electrodos permanecen estables en el tiempo y pudiendo ser utilizados por pacientes de forma doméstica (Leuthardt *et al.*, 2004; Schalk *et al.*, 2007; Brumberg *et al.*, 2010).

El desarrollo de electrodos implantables es un área de activa investigación, no solo por parte de la comunidad científica, sino también por diversas empresas. Por ejemplo, la empresa Synchron<sup>1</sup> desarrolló “The Stentrode System”, un electrodo implantable similar a un stent, el cual puede ser colocado con una cirugía endovascular mínimamente invasiva y es capaz de medir señales de la corteza motora. La empresa cuenta con numerosas publicaciones científicas sobre su desarrollo (Oxley *et al.*, 2021; Raza *et al.*, 2020; Opie and Oxley, 2019). Por otro lado, la empresa Neuralink<sup>2</sup>, que en los últimos años ha tomado notable visibilidad, anunció el desarrollo de su interfaz “The Link” (Figura 2.3), aunque todavía no han comunicado precisiones. Este desarrollo busca implantar microhilos con gran cantidad

<sup>1</sup><https://synchron.com/>

<sup>2</sup><https://neuralink.com/>

de electrodos sobre distintas áreas del cerebro. La cirugía de precisión que requiere la implantación de los electrodos se realizará de forma automática mediante un robot quirúrgico desarrollado por la misma compañía.

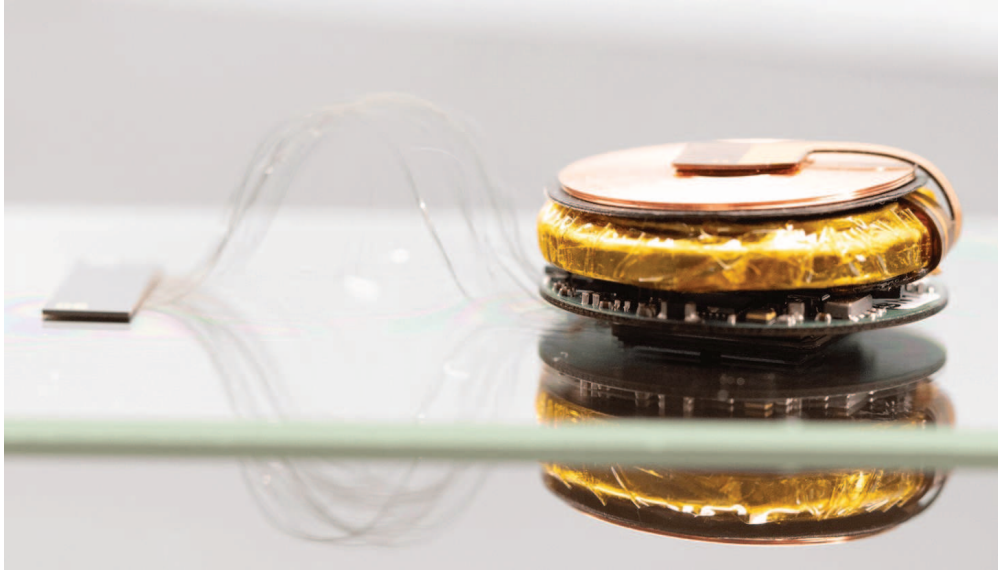


Figura 2.3: Dispositivo “The Link” de la empresa Neuralink.

### 2.3.2. Técnicas no invasivas

#### Electroencefalografía

La electroencefalografía de superficie (EEG) es la técnica estándar y más utilizada para la medición de señales cerebrales, utilizando el mismo principio de medición que el ECoG, pero haciéndolo desde la superficie de la cabeza.

Los elementos que protegen al cerebro actúan como filtros y aislantes eléctricos, mezclando y atenuando las señales de distintas poblaciones neuronales haciendo que el EEG posean una resolución espacial mucho menor comparada con las técnicas invasivas. En contraste, esta técnica permite medir distintas regiones cerebrales al mismo tiempo, permitiendo un análisis más global del cerebro.

El EEG no es capaz de medir la actividad de las estructuras subcorticales y basales del cerebro y se limita a captar la sincronización de los disparos de neuronas piramidales presentes en la corteza (Brumberg *et al.*, 2010). Por otro lado, mientras que con las señales de ECoG es posible capturar oscilaciones en una banda de frecuencias de entre 0 a 500 Hz, las señales de EEG poseen un ancho de banda mucho más acotado, pudiendo sólo captar actividad menor a 100 Hz (Brumberg *et al.*, 2010).

Es importante destacar también que el diseño de los equipos de EEG ha avanzado no solamente en mejorar la adquisición de las señales, sino también en otras características fundamentales para su uso masivo, como la portabilidad, seguridad, accesibilidad, ergonomía y estética.

#### Magnetoencefalografía

El fenómeno que mide la magnetoencefalografía (MEG) se encuentra íntimamente relacionado con el EEG, ya que esta tecnología mide la contraparte magnética de la actividad eléctrica neuronal. Cada impulso eléctrico que las neuronas producen, genera a su vez un correspondiente campo magnético. Este campo es tangencial al flujo de corriente eléctrica, por lo que, a diferencia del EEG, la MEG mide la actividad de las neuronas que se encuentran paralelas y no perpendiculares a la corteza cerebral.

Tabla 2.1: Comparación de los principales métodos de medición de la actividad cerebral. \* Valores correspondientes a (Nicolas-Alonso and Gomez-Gil, 2012).

Métodos	Actividad	Resolución Temporal [s]*	Resolución espacial [mm]*	Invasivo	Portabilidad
SUA	Eléctrica	0.003	0.05	Si	Si
LFP	Eléctrica	0.003	0.5	Si	Si
ECoG	Eléctrica	0.003	1	Si	Si
EEG	Eléctrica	0.05	10	No	Si
MEG	Magnética	0.05	5	No	No
fNIRS	Metabólica	1	5	No	Si
fMRI / PET	Metabólica	1	1	No	No

Los equipos de MEG cuentan con superconductores y aislamientos electromagnéticos especiales que permiten captar los pequeños campos magnéticos generados por las neuronas. Estas características hacen que la tecnología actual de MEG no pueda ser utilizada fuera de laboratorios ni con pacientes crónicos (Brumberg *et al.*, 2010).

### Espectroscopia funcional de infrarrojo cercano

La espectroscopia funcional de infrarrojo cercano (Functional near-infrared spectroscopy, fNIRS) es una tecnología no invasiva que se basa en la medición de los cambios de la hemoglobina oxigenada y desoxigenada en la corteza cerebral (Ayaz *et al.*, 2011). Estos cambios hemodinámicos son lentos, por lo que esta técnica tiene una mucho menor resolución temporal comparada con la de las técnicas de medición de la actividad electromagnética del cerebro.

Aunque esta técnica es mucho más reciente que el EEG, distintos trabajos han demostrado su potencial para ser utilizadas en interfaces cerebro-computadora. (Sitaram *et al.*, 2007; Ayaz *et al.*, 2011).

### Neuroimágenes

Las técnicas de neuroimágenes más utilizadas en el campo de las neurociencias son la resonancia magnética funcional (Functional magnetic resonance imaging, MRI) y la tomografía por emisión de positrones (Positron emission tomography, PET). La primera tiene un principio de medición similar al del fNIRS, detectando cambios en el flujo sanguíneo y relacionando los mismo con la actividad metabólica de las neuronas.. En contraste al fNIRS, el fMRI puede medir regiones internas del cerebro con una gran resolución espacial. Sin embargo, los resonadores son equipos sumamente costosos y no pueden ser transportados. Por su parte, el PET también es una técnica de registro metabólico, pero utiliza elementos radiactivos o radiomarcadores para detectar las zonas que presentan una mayor actividad metabólica y mayor demanda de oxígeno. Si bien estas tecnologías no pueden ser utilizadas para la construcción de ICCs, han aportado valiosa información sobre los mecanismos y procesos cerebrales.

## 2.4. Paradigmas

Como definimos anteriormente, en el campo de las ICCs, un paradigma se define como la actividad cerebral voluntaria que un usuario ejecuta para generar una respuesta deseada. Los paradigmas son fundamentales para las aplicaciones reales de las interfaces, ya que cada uno cuenta con características particulares que se relacionan con la intención del usuario.

Los paradigmas pueden clasificarse en dos grandes grupos: exógenos y endógenos. En los primeros, la actividad cerebral es generada cómo una respuesta involuntaria a una señal externa y no es el usuario

quien controla la actividad cerebral. Por el contrario, en los paradigmas endógenos el usuario produce los cambios en la actividad cerebral de forma voluntaria y no en respuesta a un estímulo externo.

En esta sección se detallarán los paradigmas más utilizados para el control de ICCs, sus mecanismos neuronales, sus características principales y sus ventajas y desventajas de uso.

### 2.4.1. P300

Este paradigma se basa en un potencial positivo que aparece como respuesta a un estímulo inesperado 300 ms después del mismo. Aunque los estímulos más utilizados son visuales, es posible construir una ICC mediante estímulos auditivos o somatosensoriales. El uso de estímulos visuales requiere que el paciente mantenga un control ocular confiable, limitando el uso de este paradigma en pacientes con discapacidades severas. A su vez, las técnicas de seguimiento ocular mediante imágenes han demostrado una gran precisión y una respuesta rápida, convirtiéndolas en candidatas más apropiadas para este tipo de pacientes.

Cómo principal ventaja, este paradigma no requiere de entrenamiento alguno por parte del usuario y permite la elección de múltiples opciones.

Sin embargo, la actividad evocada va perdiendo intensidad con el tiempo, ya que el usuario se acostumbra a los estímulos y la intensidad de la respuesta depende de la probabilidad de ocurrencia del mismo (Ravden and Polich, 1998; Farwell and Donchin, 1988). Por otro lado, este paradigma genera una gran fatiga visual, limitando su uso continuo. Por último, este paradigma tiene una baja tasa de transferencia de información (20-25 bits/min) (Nicolas-Alonso and Gomez-Gil, 2012). Esta unidad es ampliamente utilizada para comparar distintas ICCs (Wolpaw *et al.*, 2002). Estas condiciones descartan este paradigma para aplicaciones a largo plazo, aunque sigue representando una solución rápida en algunos casos particulares.

### 2.4.2. Potenciales evocados de estado estacionario

Este paradigma exógeno, también conocidos como SSVEP (Steady state visual evoked potential), es producido por la sincronización neuronal ante un estímulo externo que oscila a una frecuencia fija. Las neuronas, principalmente aquellas ubicadas en la corteza visual, se sincronizan a la misma frecuencia del estímulo externo y generan un pico de energía en la misma frecuencia (Norcia *et al.*, 2015). En distintos estudios se demostró un aumento en la intensidad en la frecuencia de oscilación relacionado no solo con la atención directa, sino también con la indirecta (cover and over attention) (Morgan *et al.*, 1996). Este paradigma permite también la selección entre un gran número de opciones y tiene una de las mejores tasas de transferencia en la actualidad (60-100 bit/min) (Nicolas-Alonso and Gomez-Gil, 2012). Sin embargo, la necesidad de un mínimo control ocular y la fatiga visual limita fuertemente su uso a largo plazo.

### 2.4.3. Potenciales corticales lentos

Los potenciales corticales lentos, (Slow cortical potentials, SCP) son producidos por modulaciones voluntarias que duran de uno a varios segundos y producen variaciones de potencia en frecuencias inferiores a 1 Hz (Birbaumer *et al.*, 1990). Una disminución en la potencia se relaciona con un aumento en la actividad neuronal, propio de tareas cognitivas más complejas. Por el contrario, un aumento en este potencial es generado por el cerebro en estado de reposo (Birbaumer *et al.*, 1990). Es importante mencionar que las modulaciones necesarias para generar los SCP de forma voluntaria pueden ser entrenadas (Hinterberger *et al.*, 2004).

Como principal ventaja, este paradigma no requiere de estímulos externos para su generación y puede ser utilizado en casos extremos de síndromes de enclaustramiento (Kübler *et al.*, 1999; Birbaumer *et al.*, 2003; Iversen *et al.*, 2008). Sin embargo, la tasa de transferencia de este paradigma es muy baja (5-12 bits/min) (Nicolas-Alonso and Gomez-Gil, 2012), ya que, como su nombre indica, los cambios modulados son muy lentos.

### 2.4.4. Imaginería Motora

La imaginería motora es uno de los paradigmas endógenos más estudiados en la actualidad y ha superado ampliamente al paradigma SCP en cuanto a tasa de transferencia (3-35 bits/min) (Nicolas-Alonso and Gomez-Gil, 2012). Este paradigma tiene sólidas bases fisiológicas, ya que tanto el planeamiento como la ejecución de los movimientos son campos profundamente estudiados en las neurociencias. Además, la corteza somatosensorial y motora se encuentran bien caracterizadas, tal como se puede observar en la Figura 3. La resolución espacial del EEG permite diferenciar, con bastante confianza, las regiones motoras que se encuentran en regiones de la corteza más lejanas, como las que controlan los movimientos de pies, manos y lengua. Sin embargo, la distinción entre movimientos más finos, cómo la diferenciación entre el movimiento de cada dedo individualmente o los gestos del habla, representan todavía un problema abierto y complejo. Si bien este paradigma resulta de gran utilidad y naturalidad para algunas aplicaciones, como la neurorehabilitación y el control de exoesqueletos, no representa un paradigma totalmente natural para otras aplicaciones, especialmente aquellas que involucren comunicación.

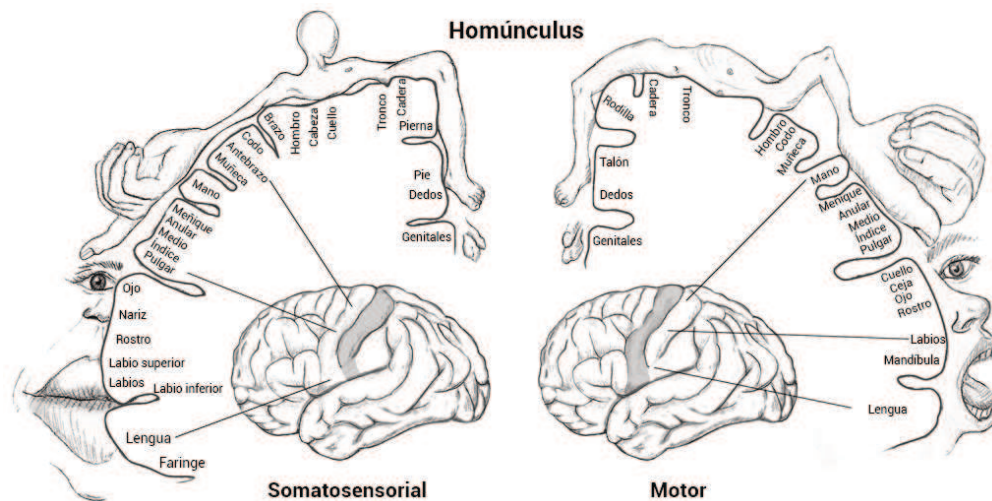


Figura 2.4: Homúnculo motor y somatosensorial. Adaptada de (EBMC, 2015)

### 2.4.5. Paradigmas relacionados al habla

El habla es por excelencia la forma de comunicación más eficiente del ser humano, en cuanto a cantidad de información transmitida por minuto. Este es uno de los principales motivos para el creciente interés en el estudio de la decodificación del habla como paradigma de control en las ICCs. Esto permitiría generar un salto cuantitativo en la, actualmente, limitada tasa de transferencia de información de las ICCs. Adicionalmente, el control mediante la decodificación del habla representaría una forma de control mucho más natural comparada con otros paradigmas.

Sin embargo, el habla es uno de los procesos más complejos y menos conocidos del ser humano. Gran parte del proceso del habla, principalmente aquellos relacionado con la comprensión de fonemas y palabras, se encuentra lateralizado en el hemisferio izquierdo, incluso para personas de dominancia zurda (Kandel *et al.*, 2000). Como se mostró anteriormente en la Figura 2.2, las regiones neuronales relacionadas al habla se encuentran pobremente definidas y ampliamente distribuidas por todo el cerebro.

Si bien los nombres de los paradigmas relacionados al habla fueron usados de forma casi indistinta, existen diferencias teóricas y fisiológicas entre ellos. Gran parte de este análisis y discusiones se encuentran en el trabajo presentado en el Anexo I.

### Habla Silenciosa

En este paradigma (Silent speech), el usuario realiza los movimientos de fonación para generar una palabra, pero sin emitir sonido. Por lo general, para la decodificación de este paradigma se utilizan señales de electromiografía (EMG) del tracto vocal, lengua, labios, músculos faríngeos y laríngeos, y no señales cerebrales. También relacionado se encuentra el paradigma de “habla intentada” (Cooney *et al.*, 2018), donde un paciente, sin poseer la capacidad de emitir sonido, intenta hablar. Este paradigma es utilizado principalmente para aplicaciones de neurorehabilitación o en ambientes con importante ruido acústico, pero no como paradigma de control de una ICC.

### Habla Imaginada

En el habla imaginada (Imagined speech), los usuarios deben concentrarse en los gestos articulatorios de la boca, labios y lengua, pero sin realizar ningún movimiento (Schultz *et al.*, 2017). En este sentido, el habla imaginada es un caso particular de imaginación motora, representando un problema complejo por ser el control motor del habla ejecutado por zonas próximas de la corteza motora. Este problema es similar al de detección de movimientos de dedos individuales, o el control muscular en la escritura (Brumberg *et al.*, 2010). Este paradigma ha sido explorado utilizando señales de EEG (DaSalla *et al.*, 2009; Zhao and Rudzicz, 2015; Brigham and Kumar, 2010; Sereshkeh *et al.*, 2017; Cooney *et al.*, 2019; Pressel-Coreto *et al.*, 2016) y electrocorticografía (ECoG) (Leuthardt *et al.*, 2004; Pei *et al.*, 2011; Guenther *et al.*, 2009).

### Habla Interna

Por su parte, el habla interna (Inner speech), se define como el proceso interiorizado donde una persona piensa en significados puros, generalmente asociado a una imagen auditiva de la propia voz interior (Schultz *et al.*, 2017). En forma práctica, el uso del paradigma de habla interna permitiría ejecutar una acción cuando el sujeto imagina su voz dando una instrucción directa a la interfaz. A diferencia del habla imaginada y silenciosa, en el habla interna no se conservan las propiedades fonológicas ni las cualidades de cambio de turno de un diálogo externo (Schultz *et al.*, 2017; Alderson-Day and Fernyhough, 2015; Gonzalez-Lopez *et al.*, 2020). En comparación con los procesos del sistema motor, el procesamiento del lenguaje es un fenómeno más complejo e implica e involucra distintas áreas corticales (Pei *et al.*, 2011; Indefrey and Levelt, 2004). Hasta el día de hoy no hay certezas de que los paradigmas relacionados al habla puedan ser distinguidos a partir de señales de EEG. En este sentido, el habla interna representaría un paradigma aún más natural al no requerir de una concentración especial sobre los gestos articulatorios del habla. El fenómeno del habla interna fue ampliamente estudiado desde las neurociencias y la medicina, para dar una explicación a las alucinaciones auditivas presente especialmente en pacientes con esquizofrenia. Este fenómeno se atribuye a una falla en la auto identificación de la propia voz interna (Frith and Done, 1988), principalmente producida por la falta de una copia eferente del estímulo producido. Este fenómeno genera activaciones en la corteza frontal inferior izquierda, la corteza temporal, parahipocampal y cerebelosa (Shergill *et al.*, 2003). Este paradigma ha sido también estudiado para aplicaciones relacionadas con las ICCs utilizando tanto EEG (D’Zmura *et al.*, 2009; Deng *et al.*, 2010; Suppes *et al.*, 1997), ECoG (Pei *et al.*, 2011). fMRI como PET (Fiez and Petersen, 1998; Price, 2000; Hickok and Poeppel, 2007; McGuire *et al.*, 1996). Hasta el día de hoy no hay certezas sobre si es posible distinguir estos paradigmas a partir de señales de EEG.

## 2.5. Desafíos de las Interfaces Cerebro-Computadora

Las ICCs-EEG presentan diversos y complejos desafíos en todas sus etapas. En la siguiente sección, describiremos algunos de los más importantes, enfocándonos en aquellas que afectan principalmente a las etapas de extracción de características y clasificación, por ser puntos centrales de esta tesis.

### 2.5.1. Escasa cantidad de datos

Los datos de electroencefalografía públicamente disponibles son muy limitados ya que su adquisición requiere generalmente de mucho tiempo de grabación, instalaciones especiales y equipos costosos. Sumado a esto, solo un muy pequeño número de publicaciones en el área permiten el acceso público a los datos utilizados, limitando la reutilización de estos datos y restringiendo fuertemente la reproducibilidad de los resultados obtenidos. Por último, un gran porcentaje de los datos son adquiridos en participantes sanos, siendo muy difícil conseguir datos de pacientes reales.

### 2.5.2. Variabilidad inter e intra sujeto

Sumado a la escasez de datos públicamente disponibles, otra de las principales complicaciones de las ICCs es la gran heterogeneidad de los datos, principalmente causada por la variabilidad inter como intra sujeto.

Las variaciones intrasujeto tienen dos orígenes fundamentales. El primero es debido a alteraciones ajenas al sujeto, como la diferencia en posición o impedancia de los electrodos o la cantidad de ruido electromagnético presente en el medio. Por otro lado, existen variaciones propias del sujeto, que pueden darse a corto plazo, como la variación en la concentración, la fatiga y el entusiasmo, o a largo plazo, principalmente debido a fenómenos de plasticidad neuronal, acostumbamiento a estímulos externos o entrenamiento del mismo sujeto.

Por su parte, las variaciones intersujeto se producen por las diferencias anatómicas en el cerebro de los individuos y estructuras que lo recubren, los distintos ajustes del equipo sobre el paciente y la diferencia intrínseca en la ejecución mental de las acciones requeridas.

Esta gran variabilidad hace que las ICCs deban ser altamente sujeto-específicas y personalizadas, debiendo ajustarse con el tiempo a los propios cambios intersujeto (Samek *et al.*, 2013). Afortunadamente, en los últimos años distintos métodos de adaptación de dominio y alineamiento fueron propuestos para poder utilizar al máximo la escasa cantidad de datos disponibles, reduciendo los tiempos de entrenamiento y mejorando el desempeño de clasificación.

### 2.5.3. Alta dimensionalidad y redundancia

Las señales de EEG presentan una alta dimensionalidad, principalmente debido a que los equipos de adquisición pueden tener una gran cantidad de canales, hasta 256, y una tasa de muestreo de hasta 16 kHz por canal<sup>3</sup>, de lo cual se obtendrían un total de 4096000 características por segundo.

Espacialmente, los canales cercanos capturan la actividad de poblaciones neuronales similares, produciendo que la información de estos canales sea altamente redundante. Por otro lado, las muestras temporalmente continuas contienen información muy similar, ya que desde las señales de EEG solo pueden medir oscilaciones relativamente lentas, menores a 100 o 50 Hz. Estas dos características hacen que las etapas de extracción, reducción y selección de características tengan una vital importancia en las interfaces cerebro-computadora.

### 2.5.4. Baja relación señal ruido

La actividad producida por las poblaciones neuronales se encuentra en el orden de los micro voltios, medida desde el EEG. Esto hace que las señales sean muy sensibles a estar contaminadas con otros tipos de ruido, como el electromagnético o la actividad muscular, la cual es varios órdenes de magnitud mayor que la actividad neuronal.

Si bien los algoritmos de eliminación de ruido y artefactos han mejorado en los últimos años, la baja relación señal ruido sigue siendo un problema inevitable en las EEG-ICCs, principalmente en equipos de bajo costo.

---

<sup>3</sup>[https://www.biosemi.com/activetwo\\_full\\_specs.htm](https://www.biosemi.com/activetwo_full_specs.htm)

### 2.5.5. Aplicaciones en situaciones reales

Las aplicaciones en situaciones reales implican para las ICCs una respuesta en tiempo real, en entornos domésticos, con equipos comerciales y utilizada por pacientes con severas discapacidades. Según el cuestionario presentado en (Vansteensel *et al.*, 2017), los investigadores del área destacaron como su mayor preocupación el traspaso de estas tecnologías desde los laboratorios hacia los entornos reales.

### 2.5.6. Aspectos éticos

Las interfaces cerebro-computadora, especialmente cuando involucran personas con discapacidad, presentan complejos dilemas éticos. Algunos de ellos han sido debatidos profundamente desde las ciencias médicas, cómo la dificultad de pedir consentimiento informado a un paciente con LIS o situaciones donde un familiar directo del paciente desea que este sea sometido a una cirugía para utilizar ICCs invasivas.

Sumado a estos dilemas generales, existen dilemas íntimamente relacionados con las ICCs, como el manejo de las expectativas y datos sensibles, la capacidad, a futuro, de leer la mente del paciente, poniendo en peligro la privacidad del mismo (Vlek *et al.*, 2012). Aunque relativamente poca atención se le presta a estos problemas, es de vital importancia dar estos debates a tiempo, ya que tendrán a futuro una gran influencia en la opinión pública y en la aceptación masiva de las ICCs.



## Sección 3

# Datos

### 3.1. Datos de electroencefalografía

Los datos constituyen la base para el avance de la ciencia en su conjunto, y son un pilar fundamental para el desarrollo de nuevas tecnologías, especialmente en el área del aprendizaje maquina. En el caso particular de las neurociencias se presenta la complejidad adicional de que la generación de datos es muy costosa, tanto desde el tiempo como de los equipamientos e instalaciones necesarias para adquirir datos. Además, los datos capturados en un paradigma no pueden ser directamente utilizados para el desarrollo de otros paradigmas.

Si bien existen algunos repositorios que contienen datos adquiridos en diversas condiciones y modalidades, son pocos los datos disponibles de EEG generados en el contexto de ICCs. Además, a pesar de que encontrarse disponibles bases de datos adquiridas en paradigmas como P300, imaginaria motora y habla imaginada, no existe ninguna base de datos pública adquirida bajo el paradigma de habla interna.

En la presente sección, se realizará una breve introducción y descripción de todas las bases de datos utilizadas a lo largo de esta tesis, tanto para el diseño como para la evaluación de los distintos algoritmos propuestos. Por último, se presentarán los puntos más relevantes del procedimiento realizado para adquirir una base de datos nueva bajo el paradigma de habla interna.

#### 3.1.1. Bases de datos de P300

Esta base de datos fue introducida en (Ledesma-Ramirez *et al.*, 2010) y contiene los registros de EEG de 25 participantes sanos utilizando una ICC basada en el paradigma de P300 para controlar un deletreador. Los datos fueron registrados utilizando un equipo gUSBamp, a una frecuencia de muestreo de 256 Hz y utilizando 10 canales (Fz, C3, Cz, C4, P3, Pz, P4, PO7, PO8, Oz, acorde al sistema el sistema 10/20). Cada segmento (trial) cuenta con 1000 ms de grabación comenzando en el momento en el que el estímulo fue presentado. Cada segmento tiene entonces 2560 características (features), correspondientes a 256 muestras  $\times$  10 canales.

El protocolo de estimulación contaba con una pantalla donde se presentaba una matriz de 6 filas por 6 columnas. Cada fila o columna destelló durante un periodo de 62,5 ms con intervalos entre estímulos de 125 ms. Para cada carácter, el bloque de estimulación (destello de las 6 filas y 6 columnas) se repitió 15 veces el bloque de estímulos. De esta forma, la decodificación de cada carácter de la palabra a deletrear demora aproximadamente 34 segundos.

Cada sujeto participó en cuatro sesiones, pero solamente las dos primeras fueron utilizadas por tener la etiqueta correspondiente a la presencia del potencial P300. Cada sujeto deletreó 21 caracteres, por lo que cada registro contiene 3780 trials en total (21 caracteres  $\times$  12 estímulos  $\times$  15 repeticiones), donde 630 de estos presentan el potencial P300. En la etapa de preprocesamiento, los registros de EEG fueron filtrados entre 0,1 Hz y 12 Hz mediante un filtro pasabanda Butterworth de 4° orden.

Esta base de datos fue utilizada en el trabajo presentado en el Anexo II.

### 3.1.2. Bases de datos de imagería motora

Dos bases de datos de imagería motora fueron utilizados para el trabajo presentado en el Anexo IV. Estos datos tienen la particularidad de estar grabados en dos sesiones distintas, por lo que permiten atacar el problema de la variabilidad inter sesión en un mismo sujeto.

La primera base de datos fue adquirida en (Peterson *et al.*, 2019) y no se encuentra públicamente disponible. Los datos se registraron en 10 participantes sin experiencia previa en el uso de ICCs (3 mujeres, 4 personas de dominancia zurda, edad media  $\pm$  desvío estándar =  $25,45 \pm 2,50$  años). Las sesiones de adquisición tuvieron una separación máxima de 5 días.

Para la adquisición se utilizó un sistema de EEG portátil de 64 canales, con una frecuencia de muestreo de 512 Hz. Los electrodos se colocaron de acuerdo con el sistema internacional 10-20, utilizando CPz como referencia y AFz como electrodos de tierra. Las señales de EEG se filtraron entre 0,5 Hz y 40 Hz. Los sujetos realizaron dos tareas mentales, reposo e imaginación del movimiento de agarre de su mano dominante. Cada sesión se componía de cuatro bloques de estimulación separados por breves descansos. Cada bloque constaba de 40 ensayos (20 para cada condición), lo que suponía un total de 160 ensayos al final de cada sesión. No se proporcionó ninguna retroalimentación al participante durante las sesiones.

El segundo conjunto de datos se encuentra disponible en forma pública en BNCI Horizons 2020<sup>1</sup> bajo el código de BNCI2015001. Es el conjunto de datos presentado en (Faller *et al.*, 2012). Los datos G se obtuvieron de 12 voluntarios sanos sin experiencia previa en el manejo de ICCs (5 mujeres, edad  $24,8 \pm 3,0$  años). Las sesiones de adquisición tuvieron una separación máxima de 5 días. Los registros de EEG se adquirieron con el sistema de electrodos activos g.GAMMAsys (FC3, FCz, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CPz y CP4) junto con un amplificador g.USBamp a una frecuencia de muestreo de 512 Hz. Durante la adquisición, las señales se filtraron entre 0,5 y 100 Hz, aplicando adicionalmente un filtro Notch en 50 Hz. En cada sesión, los participantes realizaron 5 series de 40 ensayos de imagería motora de manos y pies. Sólo en la primera ejecución de cada sesión no se proporcionó retroalimentación visual.

### 3.1.3. Bases de datos de habla imaginada

La base de datos de habla imaginada fue introducida en (DaSalla *et al.*, 2009) y fue utilizada para los desarrollos del Anexo II, III y IV. Esta base de datos se encontraba disponible en el siguiente enlace<sup>2</sup>, aunque actualmente el sitio no pertenece a una dirección válida.

La base de datos contiene señales de EEG registradas en tres sujetos (1 mujer, edades entre 26 y 29 años, todos los participantes diestros). El equipo de adquisición utilizado fue un BioSemi ActiveTwo System con 64+8 electrodos activos a una frecuencia de muestreo de 2048 Hz, aunque luego fue submuestreada a 256 Hz. Cada trial tiene una duración de 3 s, con 1 s pre estímulo. Las señales fueron luego filtradas entre 1 y 45 Hz para luego aplicar el filtro de patrones espaciales comunes (Common spatial patterns, CSP).

En el protocolo de estimulación, los participantes debían imaginar el movimiento de la boca al pronunciar las vocales /a/ y /u/. Estas vocales fueron específicamente seleccionadas por poseer los gestos motores más distintos. La imaginación de la vocal /a/ implica la imaginación de la apertura de la boca, mientras que la vocal /u/ implica el redondeo de los labios. Por último, se adquirió una condición de reposo. Cada condición fue grabada 50 veces por participante.

En dicho trabajo, se realizaron comparaciones por pares de condiciones (C1 = /a/ frente a reposo; C2 = /u/ frente a reposo; C3 = /a/ frente a /u/). Para cada condición, un filtro espacial fue aprendido y solamente la proyección de las señales sobre los primeros cuatro filtros espaciales fueron utilizados. Los datos disponibles contaban con el procesamiento descripto.

<sup>1</sup><http://bnci-horizon-2020.eu/>

<sup>2</sup>[http://www.brainliner.jp/data/brainliner-admin/Speech\\_Imagery\\_Dataset](http://www.brainliner.jp/data/brainliner-admin/Speech_Imagery_Dataset)

### 3.1.4. Bases de datos de habla interna

Las bases de datos descriptas anteriormente han permitido el desarrollo de algoritmos que enfrentan los distintos desafíos de las ICCs. Sin embargo, ante la ausencia de datos públicos de habla interna, la adquisición de una base de datos con estas características fue un punto central en el desarrollo de esta tesis. La descripción detallada de la adquisición, procesamiento y análisis de los datos obtenidos se encuentra en el trabajo presentado en el Anexo I. Dicho artículo se encuentra publicado en la revista *Nature Scientific Data*.

#### Aspectos generales

El estudio se llevó a cabo en la Universidad Torcuato Di Tella de la Ciudad Autónoma de Buenos Aires, Argentina, bajo la supervisión de Juan Kamienkowski y con la colaboración de Victoria Peterson. La universidad cuenta con un laboratorio dedicado a la adquisición de datos, con una sala electromagnéticamente blindada y dos computadoras de escritorio para uso exclusivo. El equipo de electroencefalografía utilizado fue el BioSemi ActiveTwo<sup>3</sup>, de 128 canales de EEG y 8 canales externos. El protocolo experimental fue aprobado por el Comité Asesor de Ética y Seguridad en el Trabajo Experimental (CEySTE, CCT-CONICET, Santa Fe, Argentina)<sup>4</sup>.

#### Participantes

Diez voluntarios, cuatro mujeres y seis hombres, con una edad promedio de  $34 \pm 10$  años, participaron del estudio. Todos los participantes eran diestros, Argentinos hablantes nativos de Español y ninguno poseía experiencia previa en el uso de ICCs ni presentaban dificultades auditivas, motoras o comunicacionales.

Cada participante asistió a una grabación de aproximadamente dos horas. En una primera instancia, se explicó el experimento para luego realizar un breve cuestionario junto con la lectura y firma del acuerdo de confidencialidad. Una vez dentro de la sala de adquisición, entre 30 y 45 minutos fueron utilizados para la puesta a punto del equipo y la preparación y montaje de los 128+8 electrodos, ubicados tal como muestra la Figura 3.1. En ese mismo intervalo, a cada participante se le mostró detalladamente el protocolo de estimulación para familiarizarse tanto con el entorno como con las instrucciones que aparecerían luego en la pantalla. Las luces en la habitación se apagaban al momento de la adquisición, por lo que la pantalla presentaba un leve tono gris para evitar encandilamiento y fatiga ocular.

El registro se llevó a cabo en tres bloques de estimulación consecutivos, permitiendo descansos de algunos minutos entre cada bloque. La duración del último bloque fue acordada con cada participante, principalmente dependiendo de su estado de cansancio y atención.

#### Protocolo de estimulación

El protocolo de estimulación fue diseñado utilizando la librería Psychtoolbox-3 (Brainard, 1997) en MatLab (MATLAB, 2010). Al comienzo de cada ensayo (trial), un punto blanco aparecía en el centro de la pantalla, donde los participantes fueron instruidos de mantener su mirada. Luego, un triángulo, cuya punta apuntaba en la dirección de la orden a ejecutar, se presentaba por 0.5 s. Las palabras (clases) fueron específicamente seleccionadas para analizar el potencial de uso del habla interna como paradigma de control de una ICC: “arriba”, “abajo”, “derecha”, “izquierda”. Una vez que el triángulo desaparecía, los participantes fueron instruidos de ejecutar la acción correspondiente a la clase y paradigma solicitado. El intervalo de acción tuvo una duración de 2.5 s, terminando cuando el punto blanco de la pantalla se transformaba en azul, indicando que el participante debía detener la acción que se encontraba ejecutando. Luego, un intervalo de descanso con una duración variable entre 1.5 y 2 s fue dado a los participantes. El resumen de cada estímulo puede observarse en la Figura 3.2.

---

<sup>3</sup><https://www.biosemi.com/products.htm>

<sup>4</sup><https://santafe.conicet.gov.ar/ceyste/>

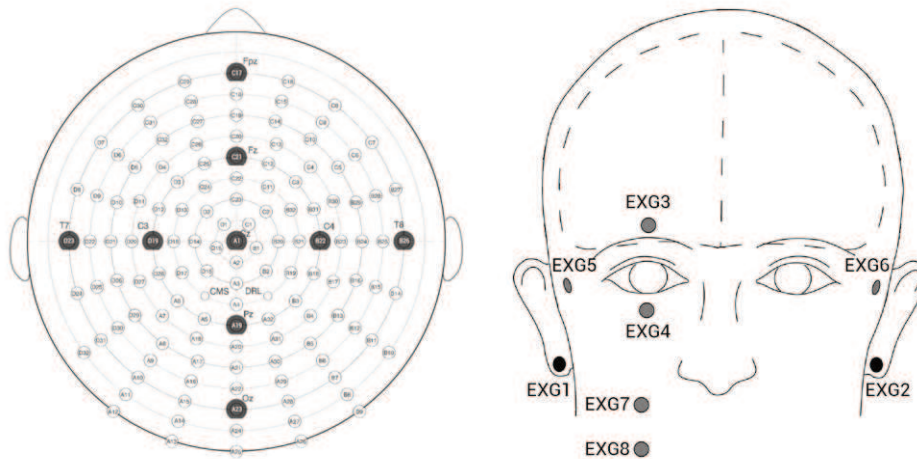


Figura 3.1: Posiciones utilizadas de electrodos. Izquierda: Electrodo de electroencefalografía, nomenclatura acorde a fabricante. Derecha: Electrodo externo.

### Paradigmas

La base de datos no fue diseñada solamente para conocer la factibilidad práctica del habla interna como paradigma de control, sino también para intentar desglosar las diferentes actividades que se suceden en el cerebro para producir la voz interna. Es por esto que, además del ya mencionado paradigma de habla interna, se adquirieron señales en otros dos paradigmas relacionados: habla pronunciada y condición visualizada. En el primero se le pidió a los participantes repetir la palabra presentada en voz alta. Por otro lado, la condición visualizada fue propuesta con la intención de capturar información sobre el sentido de las clases propuestas, ya que las mismas presentan una alta carga espacial. En este paradigma se le indicó a los participantes que se concentren en intentar mover el punto central de la pantalla en la dirección indicada. En ninguno de los paradigmas se proveyó de retroalimentación a los participantes.

Si bien utilizar tiempo del experimento para grabar otros paradigmas disminuye la cantidad de ensayos totales en el paradigma de habla interna, permite obtener una base de datos más completa desde el punto de vista neurocientífico, con la posibilidad de recabar más información sobre el comportamiento del habla interna.

En un principio, sería posible utilizar las señales de habla pronunciada para, luego de la eliminación del ruido muscular, encontrar los centros motores que se activan con la pronunciación de cada palabra. De esta forma podría ser posible identificar si centros similares se activan también en el habla interna. Un análisis similar podría llevarse a cabo con la condición visualizada, intentando encontrar centros neuronales que se activen en ambos paradigmas.

### Procesamiento de datos

Junto con la publicación de los datos, y para facilitar su uso posterior, una librería para la carga, el filtrar, el procesamiento y la visualización de los datos fue propuesta. La librería fue enteramente desarrollada en Python (Van Rossum and Drake, 2009), utilizando funciones de los paquetes MNE (Gramfort *et al.*, 2014), Numpy (Oliphant, 2006), Scipy (Virtanen *et al.*, 2020), y Pickle (Van Rossum, 2020).

El procesamiento propuesto consta de tres puntos principales y fue aplicado sobre los datos originales para generar una base de datos estructurada que facilite su posterior uso para técnicas de aprendizaje maquina. En una primera instancia, los datos fueron re-referenciados a los canales de actividad no neuronal adquiridos “EXG1” y “EXG2”. Este paso permite eliminar todo voltaje común

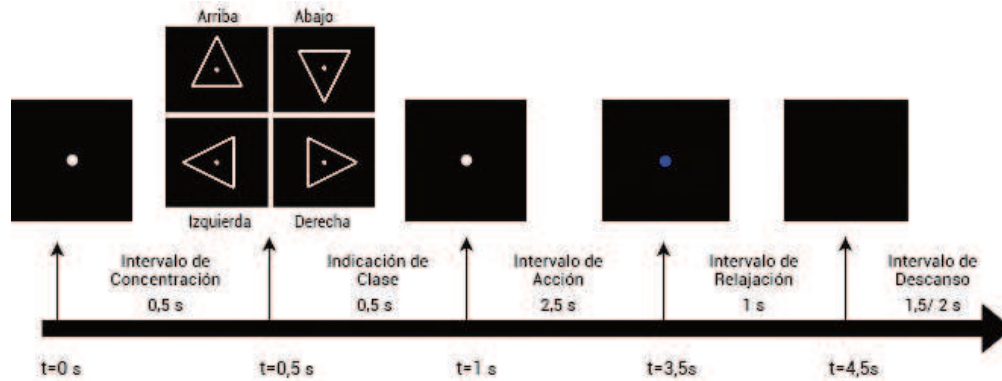


Figura 3.2: Estructura de cada estímulo. La pantalla presentada al participante en cada intervalo de tiempo se trazó en la flecha superior de la figura. El tiempo relativo y el tiempo global se representaron por encima y por debajo de la flecha, respectivamente

del cuerpo y ayuda a eliminar el ruido electromagnético. Luego, las señales fueron filtradas entre 0.5 y 100 Hz, aplicando adicionalmente un filtro Notch en 100 Hz. La decisión de utilizar un filtro amplio en frecuencia fue tomada para permitir a futuros usuarios la aplicación de filtros más angostos, sin necesidad de tener que reprocesar todos los datos. Por último, para la eliminación de artefactos oculares y musculares, un análisis de componentes independientes (Independent components analysis, ICA) fue realizado. En el mismo, se generaron el máximo de componentes posibles, 128, para reducir al máximo la cantidad de información eliminada por el método. Este procesamiento fue realizado por la función de MNE correspondiente, donde se implementó el método infomax ICA, presentado en (Bell and Sejnowski, 1995).

Al dejar públicamente disponible la librería para ejecutar el procesamiento propuesto, es posible que futuros usuarios generen procesamientos distintos con solo cambiar unas pocas líneas de código.

### Principales aportes y trabajos futuros

En total, más de 9 horas de grabaciones fueron adquiridas, contabilizando 5640 ensayos finales, distribuidos por participante, paradigma y clase como se muestra en la Tabla 3.1. Tanto las grabaciones procesadas como las originales fueron cargadas, junto con la librería completa, permitiendo una fácil y rápida reutilización por parte de futuros usuarios.

A la fecha, más de doscientas personas han descargado la base de datos desde el repositorio de OpenNeuro<sup>5</sup>. Por otra parte, distintos usuarios han comenzado a interactuar con el repositorio de Github<sup>6</sup>.

Si bien en el trabajo se presenta un análisis global de las características de las señales, un análisis detallado, participante por participantes, deberá ser realizado para encontrar información valiosa que permita conocer las diferencias y similitudes de los paradigmas entre sí y la potencialidad del habla interna como paradigma de control.

<sup>5</sup><https://openneuro.org/datasets/ds003626/versions/2.0.0>

<sup>6</sup>[https://github.com/N-Nieto/Inner\\_Speech\\_Dataset](https://github.com/N-Nieto/Inner_Speech_Dataset)

Tabla 3.1: Número de ensayos dividido por participante, paradigma y clase.

Participantes	Habla Pronunciada				Habla Interna				Condición Visualizada			
	Arriba	Abajo	Derecha	Izquierda	Arriba	Abajo	Derecha	Izquierda	Arriba	Abajo	Derecha	Izquierda
sub-01	25	25	25	25	50	50	50	50	50	50	50	50
sub-02	30	30	30	30	60	60	60	60	60	60	60	60
sub-03	25	25	25	25	45	45	45	45	55	55	55	55
sub-04	30	30	30	30	60	60	60	60	60	60	60	60
sub-05	30	30	30	30	60	60	60	60	60	60	60	60
sub-06	27	27	27	27	54	54	54	54	54	54	54	54
sub-07	30	30	30	30	60	60	60	60	60	60	60	60
sub-08	25	25	25	25	50	50	50	50	50	50	50	50
sub-09	30	30	30	30	60	60	60	60	60	60	60	60
sub-10	30	30	30	30	60	60	60	60	60	60	60	60
Sub Total	282	282	282	282	559	559	559	559	569	569	569	569
<b>Total</b>	<b>1128</b>				<b>2236</b>				<b>2276</b>			

## Sección 4

# Maquinas de aprendizaje extremo para la clasificación de habla imaginada

### 4.1. Generalidades de la inteligencia artificial y su uso en interfaces cerebro-computadora

La inteligencia artificial influencia casi todos los aspectos de nuestra vida cotidiana, incluso cuando no somos completamente consciente de ello. Las películas y noticias que vemos, los productos y servicios que se nos ofrecen, las rutas que utilizamos para desplazarnos y la posibilidad de acceder a un crédito bancario, son sólo algunos ejemplos de las crecientes aplicaciones de la inteligencia artificial en el mundo.

Pese al gran avance de esta área, recién hace unos pocos años que se discuten temas relacionados con las cuestiones éticas y legales que conllevan estas tecnologías. Por ejemplo, la creación de contenido falso mediante “deep fakes” ha demostrado tener una gran influencia en los procesos electorales democráticos de varios países, incluida la Argentina. Por otro lado, los algoritmos toman decisiones de alto impacto de forma cada vez más autónomas, ya sea conduciendo un vehículo, sugiriendo candidatos para un puesto laboral o incluso generando diagnósticos médicos (Esteva *et al.*, 2017; De Fauw *et al.*, 2018). Las decisiones que toman estos algoritmos pueden presentar sesgos raciales (Buolamwini and Gebru, 2018), de género (Larrazabal *et al.*, 2020), etnia, entre otros (Wiens *et al.*, 2019; Char *et al.*, 2018). Los algoritmos de inteligencia artificial son capaces de captar y reproducir sesgos presentes en los datos donde son entrenados, principalmente producidos por las personas que los generan y etiquetan. Es importante destacar también que no son sólo los datos los responsables de generar estos sesgos, sino que también las decisiones de diseño de los modelos tienen un impacto significativo en la reproducción y amplificación de sesgos (Hooker, 2021). Si bien el campo de la “justicia algorítmica”, (machine learning fairness) ha tenido un gran crecimiento en los últimos años, mucho trabajo queda por delante para evitar que estos sesgos tengan un impacto negativo concreto en la vida de las personas. Si bien no es una solución definitiva, generar grupos de trabajos diversos es un buen primer paso para evitar, o por lo menos visibilizar, potenciales sesgos algorítmicos.

Sin escapar a estas cuestiones éticas, las interfaces cerebro computadora se han visto enormemente favorecidas por los avances de la inteligencia artificial. Esta juega un papel fundamental en dos etapas cruciales de las ICCs: la etapa de extracción de características y la etapa de clasificación. Dentro de la primer etapa, el desarrollo de algoritmos como el análisis de componentes independientes (Zhukov *et al.*, 2000; Iriarte *et al.*, 2003), el análisis de componentes principales (Bernat *et al.*, 2008), los filtros espaciales (Blankertz *et al.*, 2007; McFarland *et al.*, 1997) y las transformaciones tiempo-frecuenciales (Mallat, 1999) permitieron ampliar y mejorar el uso de las ICCs. En cuanto a la etapa de clasificación, los primeros algoritmos utilizados fueron clasificadores adaptativos, de entrenamiento supervisado, como el análisis de discriminante lineal (LDA) (Zhang *et al.*, 2013) o las máquinas de soporte vectorial

(SCV) (Jrad *et al.*, 2011). En los últimos años, el crecimiento del aprendizaje profundo ha incidido también en el campo de las ICCs. Con este enfoque, el algoritmo une la etapa de extracción de características y la etapa de clasificación. De esta forma, es el mismo algoritmo quien encuentra las características que le son más útiles para la tarea de clasificación. El aprendizaje profundo ha generado una gran revolución en diversas áreas donde se disponen grandes cantidades de datos. Por su parte, en el área de las ICCs, que aún presenta pocos datos de entrenamiento, todavía no es clara la ventaja de este tipo de enfoques sobre los métodos tradicionales (Lotte *et al.*, 2018).

En la presente sección, describiremos los algoritmos propuestos para atacar el problema de clasificación de señales de EEG en el contexto de interfaces cerebro computadora en los paradigmas relacionados al habla.

## 4.2. Máquinas de Aprendizaje Extremo

Las máquinas de aprendizaje extremo (Extreme Learning Machines, ELM), son redes neuronales artificiales con una única capa oculta que fueron propuestas en (Huang *et al.*, 2004, 2006, 2011) y poseen dos características destacadas que las diferencian de los tradicionales perceptrones multicapas (Multi Layer Perceptron, MLP). La primera es que los pesos que conectan la capa de entrada con la capa oculta no sólo son generados sino que se fijan de forma aleatoria, sin modificarse durante el entrenamiento. Por otro lado, los pesos que conectan la capa oculta con la capa de salida son calculados mediante la inversa generalizada de Moore-Penrose. Estas características permiten obtener una red con un entrenamiento “extremadamente” rápido y con una buena capacidad de generalización.

La originalidad de estas redes ha sido puesta en discusión, ya que el concepto de redes neuronales artificiales con pesos aleatorios fue inicialmente propuesto en (Schmidt *et al.*, 1992), con la diferencia que en dicho trabajo se propone calcular los pesos de salida mediante el método de Fisher. Sin entrar en la discusión sobre la originalidad de la técnica, este tipo de redes neuronales ha sido ampliamente utilizadas en diversos campos de aplicación, incluidas las neurociencias y las ICCs (Murugavel and Ramakrishnan, 2016; Song and Zhang, 2013; Yuan *et al.*, 2011; Zhao *et al.*, 2018; Shi and Lu, 2013; Ding *et al.*, 2015, 2017; Liang *et al.*, 2006).

Una de las grandes ventajas de este tipo de redes es que posee un solo hiperparámetro para ajustar: la cantidad de neuronas en la capa oculta  $M$ . Una capa oculta subdimensionada puede resultar en un mal desempeño de clasificación, mientras que una capa oculta sobredimensionada puede generar un entrenamiento más costoso sin una mejora de clasificación. En (Huang and Babri, 1998) se ha demostrado que  $N$  ejemplos de entrenamiento pueden ser aprendidos con exactamente  $N$  neuronas ocultas, proponiéndose este número como cota superior en la cantidad de neuronas en la construcción de estas redes neuronales.

Las señales de EEG presentan una complejidad particular para las ELMs: su gran redundancia de información. Este problema genera que no todas las neuronas de la capa oculta resulten útiles para la red. Para intentar solucionar, o atenuar, este problema, la eliminación o “podado” de estas neuronas ha sido un tema de investigación activo dentro de las ELMs (Rong *et al.*, 2008; Miche *et al.*, 2010; Tavares *et al.*, 2014; Alencar *et al.*, 2016). Las técnicas de podado generan redes más compactas, con mejor poder de generalización y una menor variabilidad. Otra de las técnicas utilizadas sobre las ELMs para mejorar su capacidad de generalización son los métodos de regularización (Deng *et al.*, 2009). Mediante estos métodos se evitan grandes pesos en la conexión de la capa oculta con la capa de salida, impidiendo un sobreajuste del modelo.

En lo que resta de esta sección, presentaremos dos métodos de podado de ELM propuestos en el contexto de esta tesis, junto con resultados de clasificación en señales de habla imaginada y P300.

### 4.2.1. Podado basado en relevancia de máquinas de aprendizaje extremo

Tal como hemos descripto anteriormente, en las ELMs los pesos de salida se calculan mediante la inversa generalizada de Moore-Penrose. Esto hace que todas las neuronas, incluso aquellas que no



codifican información útil para el problema, tengan que ser tenidas en cuenta por la red para calcular su salida final. Este fenómeno es especialmente importante cuando existen características en los datos de entrada que no son representativas o informativas del problema a resolver (unrepresentative features o junk features, JF). Las señales de EEG presentan una alta cantidad de este tipo de características, siendo unas pocas características las que aportan información relevante. Ahora, ¿Cómo impactan estas características en el comportamiento de estas redes? ¿Las ELMs deben diseñarse y optimizarse de la misma forma que para este tipo de problemas?. En el trabajo presentado en el Anexo II se realizó una descripción y exploración detallada sobre este fenómeno, proponiendo una solución algorítmica acorde. Este trabajo se encuentra publicado en la revista Neuroinformatics y puede accederse mediante el siguiente enlace<sup>1</sup>.

### Impacto de las características no representativas

Las características no representativas son inherentes a casi todos los problemas, siendo esto especialmente importante en el caso de la clasificación de señales de EEG. Para ilustrar este fenómeno, construimos una base de datos sintética con dos características informativas, como se muestra en la Figura 4.1. Utilizando 100 datos se entrenó el clasificador con distinta cantidad de neuronas ocultas y se midió su desempeño de clasificación. Luego, se agregaron distintas cantidades y tipos de características no representativas, repitiendo el mismo procedimiento de entrenamiento y evaluación de la red. En la Figura 4.2 se observan los resultados obtenidos de este experimento, pudiendo distinguirse dos fenómenos principales. El primero, esperable, es que a mayor número de características no representativas, peor el desempeño general del clasificador. En segundo lugar, llegada una determinada cantidad de características agregadas, el desempeño cuando  $M \gg N$  es mayor que cuando  $M \leq N$ . Por otro lado, se observa que la performance de la red llega siempre a un mínimo cuando  $M = N$ . Este mismo comportamiento fue observado también utilizando señales de EEG en el paradigma de habla imaginada (DaSalla *et al.*, 2009) y P300 (Ledesma-Ramirez *et al.*, 2010).

El comportamiento encontrado es similar al efecto de doble descendiente descrito para los perceptrones multicapa (Belkin *et al.*, 2019). Estos resultados demuestran que prácticamente el límite propuesto en (Huang and Babri, 1998) tiene validez sólo en el caso de los problemas que presentan características representativas.

### Método de podado y optimización del número de neuronas ocultas

Para indagar sobre el origen de este fenómeno, un nuevo experimento fue propuesto para analizar los pesos otorgados por la red a cada neurona. En este experimento, una nueva neurona artificial “falsa”, desconectada completamente de la entrada fue agregada en la red. Las activaciones de esta neurona eran realizaciones de una distribución normal, sin ninguna relación con la clase del ejemplo de entrenamiento correspondiente. En la Figura 4.3, se observa el valor absoluto asignado por la red a esta neurona, con respecto a la cantidad de neuronas ocultas. El peso de esta neurona “falsa” aumenta notablemente cuando la cantidad de neuronas ocultas se acerca a la cantidad de ejemplos de entrenamiento, 60 en este experimento, no obstante, se reduce rápidamente conforme aumenta la cantidad de neuronas oculta.

De esta forma, las ELMs se benefician de estructuras donde  $M \gg N$ , al poder ignorar aquellas neuronas que no aportan información útil. El mayor problema de estas estructuras es que utilizado para encontrar el parámetro óptimo de neuronas ocultas puede ser costoso computacionalmente. Es por esto que se propuso un nuevo esquema de entrenamiento, incluyendo un método de podado, que permite evitar el reentrenamiento de la red en busca del hiperparámetro óptimo  $M$ . El método, llamado Relevance-Based Pruning (RBP), permite realizar un único entrenamiento de la red, con  $M \gg N$  y luego eliminar aquellas neuronas cuyo valor absoluto del peso de salida calculado sea menor. De esta forma, se descartan las neuronas menos relevantes para el problema, generando una red más compacta, sin la necesidad de reentrenar la red y sin agregar nuevos hiperparámetros. Al evitar la

<sup>1</sup><https://rdcu.be/cyFP3>

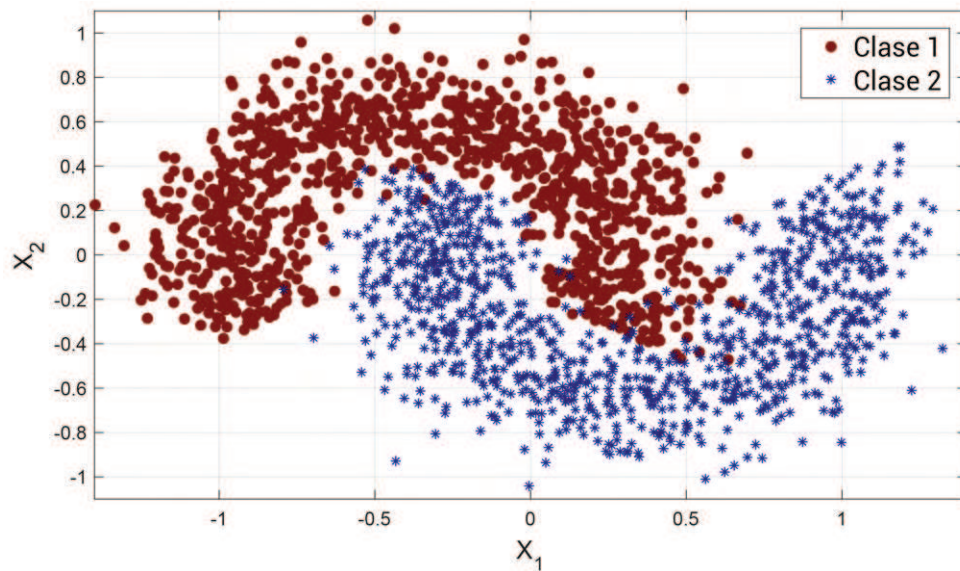


Figura 4.1: Distribución de características informativas

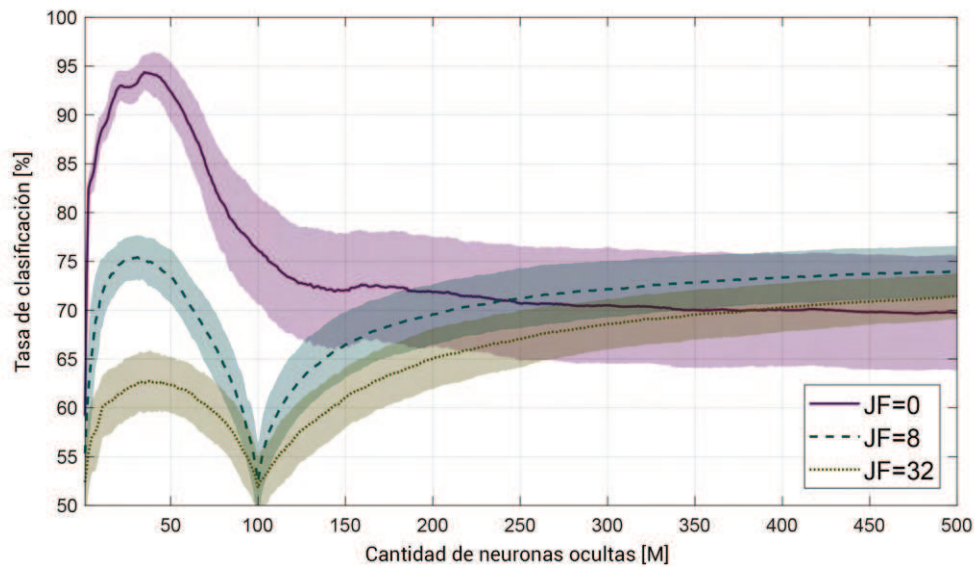


Figura 4.2: Evolución del desempeño de clasificación con respecto a los distinto número de características no representativas y neuronas ocultas

computacionalmente costosa operación de pseudo-inversión, que escala rápidamente con  $N$  y con  $M$ , RBP es significativamente más rápido que el esquema tradicional de entrenamiento de las ELMs, manteniendo su performance de clasificación.

## Resultados

Para demostrar los beneficios del método propuesto para la clasificación de señales de EEG, particularmente en el paradigma habla imaginada (DaSalla *et al.*, 2009) y P300 (Ledesma-Ramirez *et al.*,

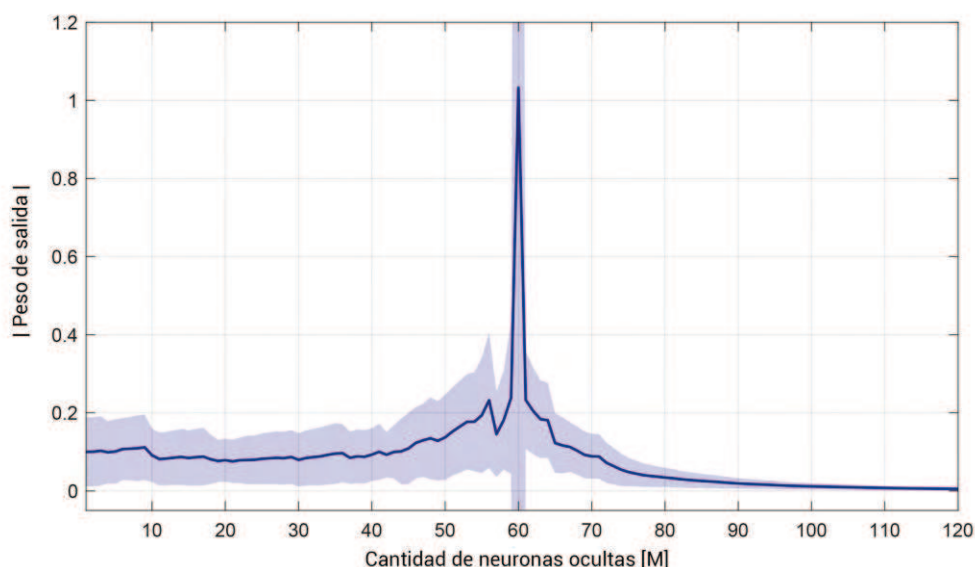


Figura 4.3: Valor absoluto del peso asignado por el modelo a la neurona falsa añadida.

Método	Tasa de clasificación [%]	Tiempo [ms]	$M$
FWD $M^* = 90$	$62,2 \pm 7,3$	$76 \pm 48$	$24 \pm 18$
FWD $M^* = 1000$	$66,9 \pm 5,5$	$195 \pm 21$	$228 \pm 76$
OP-ELM $M^* = 90$	$63,5 \pm 14,8$	$33,9 \pm 9$	$27 \pm 14$
OP-ELM $M^* = 1000$	$63,1 \pm 14,3$	$23471 \pm 54$	$77 \pm 10$
RBP	$68,0 \pm 3,8$	$43 \pm 3$	$209 \pm 151$

Tabla 4.1: Resultados de desempeño obtenidos por cada método en la base de datos de habla imaginada

2010). Para la comparación, se utilizó el método de entrenamiento tradicional de ELM (FWR por Forward ELM) y el método de podado Optimally Pruning (OP-ELM) propuesto en (Miche *et al.*, 2010).

Para RBP y FWR se utilizaron 70 % de datos para entrenamiento, 20 % de validación para encontrar el número óptimo de neuronas y 10 % de datos para la evaluación del método. Por otro lado, ya que OP-ELM utiliza una validación interna, 90 % de los datos fueron usados para entrenamiento y 10 % para evaluación.

Para la primera base de datos se realizaron 20 validaciones cruzadas de los datos y 50 distintas inicializaciones de los pesos aleatorios. Para los métodos FWR y OP-ELM la cantidad máxima de neuronas ( $M^*$ ) fue tomada tanto respetando el límite teórico propuesto en (Huang and Babri, 1998) de  $M \leq N$  como para  $M \gg N$ .

En la Tabla 4.1 se reportan el promedio y el desvío estándar de todas las validaciones cruzadas, inicializaciones, sujetos y condiciones comparadas. Un análisis desglosado por sujeto y condición puede encontrarse en el material suplementario del Anexo II. Dos conclusiones resultan de esta comparación. La primera es que para los métodos que sólo permiten  $M \leq N$ , la performance de clasificación es inferior a los métodos donde  $M \gg N$ . Por otro lado, dentro de estos métodos, RBP es capaz de encontrar el óptimo de neuronas en un tiempo de entrenamiento mucho menor.

Para la segunda base de datos se realizaron 5 validaciones cruzadas de los datos y 20 distintas inicializaciones de los pesos. El método OP-ELM fue omitido en el resto del análisis por tener tiempos de entrenamientos notoriamente mayores a RBP y FWR para  $M \gg N$ . Como la base de datos de P300 presenta un desbalance de clases de 1:6 entre los ensayos que presentan P300 y aquellos que no, el área bajo la curva ROC fue utilizada como medida de desempeño de clasificación. Si bien esta métrica

Método	AUC-ROC	Tiempo [s]	$M[\times 10^3]$
FWD $M^* = 2646$	$0,718 \pm 0,072$	$3 \pm 1$	$0,5 \pm 0,2$
FWD $M^* = 40000$	$0,816 \pm 0,096$	$572 \pm 118$	$28,2 \pm 10,6$
RBP	$0,816 \pm 0,087$	$194 \pm 1$	$14,5 \pm 6,5$

Tabla 4.2: Resultados de desempeño obtenidos por cada método en la base de datos de P300

no es recomendada para grandes desbalances, es menos sensible que la tasa de clasificación.

En la Tabla 4.2 se reportan el promedio y el desvío estándar de todas las validaciones cruzadas, todas las inicializaciones y todos los sujetos. Al igual que con la base de datos anterior, se observa que los métodos que permiten  $M \gg N$  obtienen una performance superior de clasificación. Nuevamente, RBP logra disminuir significativamente el tiempo de entrenamiento, sin disminuir el desempeño de clasificación.

Por último, como el tiempo de entrenamiento está fuertemente ligado al parámetro  $M^*$ , un nuevo experimento se realizó para determinar esta relación. Los resultados de la Figura 4.4 muestran que el método propuesto mejora el tiempo de entrenamiento con respecto al método de entrenamiento tradicional y al método de OP-ELM, sin importar el número inicial de neuronas.

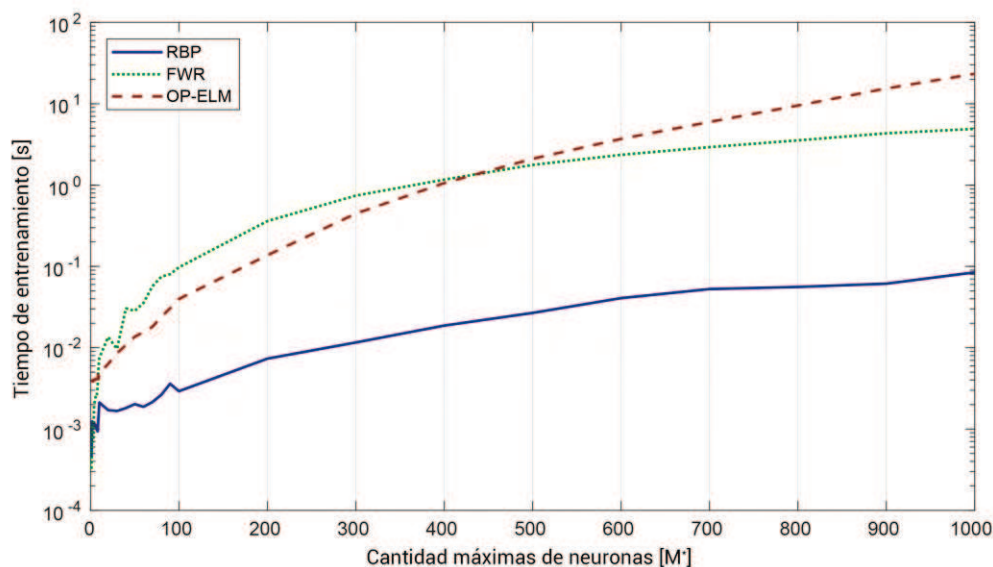


Figura 4.4: Tiempo de entrenamiento de los métodos evaluados con respecto a la cantidad máximas de neuronas.

### Discusiones y trabajos futuros

Buscar el parámetro óptimo del número de neuronas puede ser computacionalmente costoso e ineficiente en grandes ELMs. Sin embargo, este tipo de arquitecturas parecen ser las indicadas para atacar problemas que presenten características no representativas, como las señales de EEG. En respuesta a esto, que antagoniza la principal ventaja de las ELMs, una nueva forma de encontrar este parámetro fue propuesta, basada en un método de podado, que permite buscar eficientemente el número óptimo de neuronas ocultas. En aplicaciones reales, una capa oculta más pequeña reduce el tiempo de respuesta del algoritmo, mejorando la respuesta en tiempo real de las aplicaciones.

La principal limitación del método no es poder saber a priori si el problema cuenta con la cantidad de características no representativas suficientes para que la performance para  $M \gg N$  sea para mayor que para  $M \leq N$ . Esto puede encontrar una solución parcial en la versión regularizada de las

ELMs, evitando el sobreajuste cuando  $M = N$  y generando que la performance de clasificación crezca constantemente mientras crece la cantidad de neuronas ocultas.

Otro de los puntos que se deberá analizar con más detalle a futuro es la selección del parámetro  $M^*$ . Un valor demasiado bajo puede producir una caída en el desempeño de clasificación, mientras que un valor muy grande produciría largos e innecesarios tiempos de entrenamiento.

Por otra parte, sería posible analizar la relación entre los pesos generados aleatoriamente y los pesos obtenidos después del entrenamiento para inferir qué características de la señal que son más relevantes e informativas para el problema.

Por último, las implementaciones del método propuesto, tanto en Matlab como en Python, se encuentran públicamente disponibles en el repositorio de Github<sup>2</sup>, incluyendo un tutorial y ejemplo de uso detallado en la plataforma Collaboratory.

### 4.2.2. Podado estadístico pre-entrenamiento de máquinas de aprendizaje extremo

Si bien crear y fijar los pesos aleatorios en las ELMs genera un gran ahorro en el tiempo de entrenamiento, también trae aparejado serias complicaciones. La primera es la gran variabilidad en el desempeño de clasificación entre las distintas inicializaciones. La segunda es que los pesos de dos neuronas pueden ser lo suficientemente parecidos para codificar la misma información.

Mediante el podado de nodos innecesarios o redundantes, es posible reducir la cantidad de neuronas antes de la etapa de entrenamiento, reduciendo el tiempo de entrenamiento y la variabilidad entre distintas inicializaciones, al mismo tiempo que se mejora el desempeño de clasificación. La pregunta que surge inmediatamente es ¿Cómo medir la utilidad de los nodos de forma eficiente antes de entrenar?.

Esta es sin duda una pregunta abierta en la comunidad y, hasta la fecha, el único trabajo que ataca este problema es el propuesto por (Tavares *et al.*, 2014), donde se eliminan aquellos nodos cuyos pesos sean lo más linealmente independientes posibles. Aunque este método se deshace de neuronas parecidas que codifican la misma información, no hay garantía de que estos nodos sean útiles para la clasificación. Además, los vectores de pesos son comparados uno a la vez, siendo un procedimiento computacionalmente muy costoso.

En la presente sección, un método de podado pre entrenamiento mediante estadísticos de alto orden es introducido. Este método se encuentra en el Anexo III y ha sido publicado y presentado en el congreso de la 47 JAIIO, encontrándose online en el siguiente enlace<sup>3</sup>.

#### Tipos de neuronas

Al fijar los pesos que conectan la capa de entrada con la capa de salida, es posible conocer cómo responderá cada neurona ante los ejemplos de entrenamiento. De esta forma se construye la matriz de activaciones ocultas  $H$ , donde cada columna representa una neurona y cada fila la respuesta a un ejemplo de entrenamiento. Mediante el uso de las etiquetas de entrenamiento es posible agrupar las respuestas para cada clase, generando la distribución de activaciones de cada neurona para cada clase.

Utilizando datos de habla imaginada (DaSalla *et al.*, 2009) se generaron histogramas de activación para cada neurona, como se muestra en la Figura 4.5. Mediante el análisis de estos histogramas, dos tipos de neuronas distintas fueron encontradas, a las cuales se las denominó bimodales y unimodales. Las primeras, responden de forma aleatoria ante los patrones de la misma clase. Por otro lado, las neuronas unimodales responden de forma más consistente ante los ejemplos de la misma clase.

#### Métrica de podado

En principio, las neuronas bimodales aparentan carecer de utilidad para la red, ya que su actividad no se corresponde con las distintas clases, generando salidas variables para la misma clase. La distri-

<sup>2</sup>[https://github.com/N-Nieto/Relevance\\_Based\\_Pruning](https://github.com/N-Nieto/Relevance_Based_Pruning)

<sup>3</sup><http://sedici.unlp.edu.ar/handle/10915/70685>

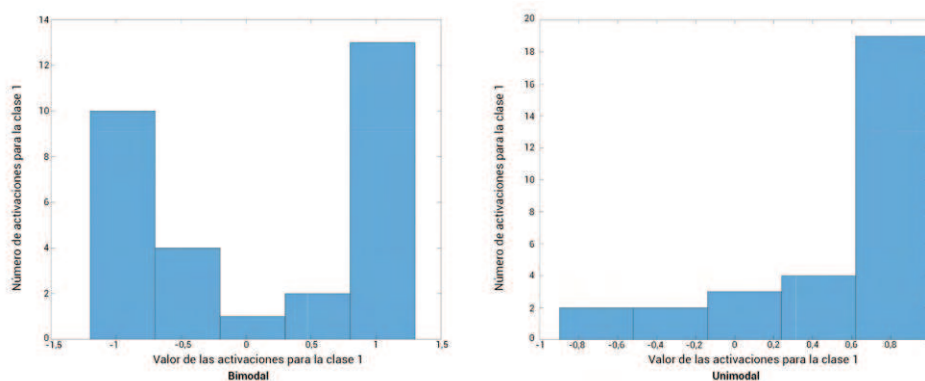


Figura 4.5: Histogramas típicos de activación de neuronas ocultas. Izquierda: Neuronas bimodales. Derecha: Neuronas Unimodales.

bución de este tipo de neuronas posee un desvío estándar grande y una media cercana a cero. Por su parte, las neuronas unimodales tienen un desvío más pequeño y una media distinta de cero.

En el caso extremo, las neuronas bimodales tendrían una actividad completamente aleatoria, generando una distribución plana y uniforme. Por su parte, las neuronas unimodales tendrían todas sus activaciones concentradas cerca de la media.

Para poder caracterizar las distintas formas de las distribuciones, se propuso utilizar la curtosis para ordenar las neuronas de más a menos útiles. La curtosis es un estadístico de cuarto orden que aporta información sobre la forma de la distribución: una curtosis alta representa distribuciones más concentradas, mientras que una menor curtosis se obtiene de distribuciones más dispersas.

### Clasificación de habla imaginada

Para validar la utilidad del método propuesto se realizó una clasificación sobre los datos de habla imaginada introducidos en (DaSalla *et al.*, 2009). Eliminando el 80% de los nodos antes del entrenamiento fue posible mejorar la performance de clasificación, tal como muestra la Figura 4.6. En la misma, se compara el entrenamiento de una ELM sin pre pruning, contra una ELM con la etapa de selección de nodos propuesta. Es importante aclarar que, para el método propuesto, se generan redes más grandes en una primera instancia para igualar luego la cantidad de neuronas después del podado.

Si bien un análisis más detallado será necesario para probar la utilidad práctica del método propuesto, los resultados apuntan a que es posible eliminar aquellas neuronas bimodales que no contienen información relevante sobre el problema.

### Limitaciones y trabajos futuros

Cómo principal limitación, el método propuesto falla en encontrar aquellas neuronas cuya activación fuese idéntica para todas las clases, con iguales distribuciones de activación. Si bien estas neuronas no serían bimodales, no aportan ninguna información discriminativa. Para evitar este problema, es posible explorar otras métricas como la asimetría.

Por otro lado, este método agrega un nuevo hiperparámetro que debe ser optimizado, la cantidad de neuronas descartadas. Un análisis más profundo sobre las activaciones o la característica de los datos podrían arrojar una recomendación práctica para su estimación.

### Combinación de métodos propuestos

Por último, fue posible combinar ambos métodos propuesto, tanto RBP como el método de podado pre entrenamiento, aprovechando la ganancia de tiempo que implica RBP y sumando la estabilidad

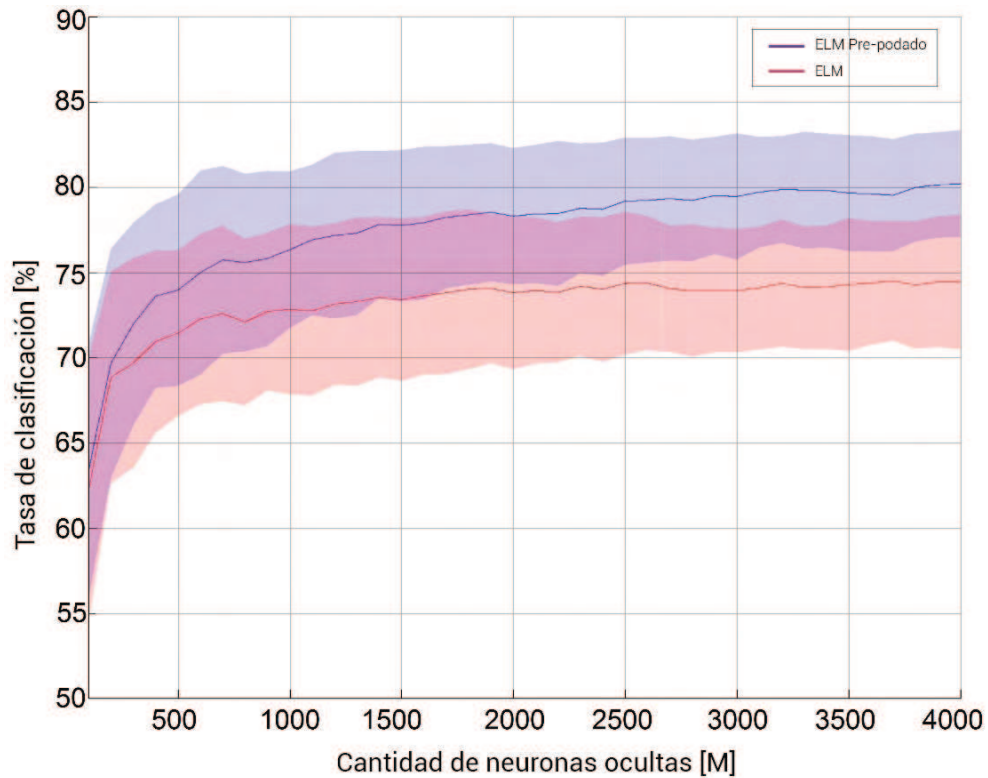


Figura 4.6: Desempeño de clasificación de señales de habla imaginada.

y mejora en clasificación aportada por el método pre entrenamiento. En la Figura 4.7-A se muestra la clasificación sobre los datos sintéticos presentados en la Figura 4.1, contaminados con 100 características no representativas. Con el método de pre-pruning se eliminaron el 80% de los nodos previo al entrenamiento. 10 distintas inicializaciones aleatorias y 3 validaciones cruzadas fueron utilizadas reportándose el desvío estándar de cada método en la Figura 4.7-B. El método de pre pruning disminuye la dispersión de clasificación entre distintas inicializaciones del modelo, produciendo soluciones más robustas y confiables.

Sobre estas bases se deberá construir a futuro un análisis sistemático para demostrar el potencial de combinar ambos métodos propuesto para mejorar el desempeño de las ELMs en el contexto de ICCs.



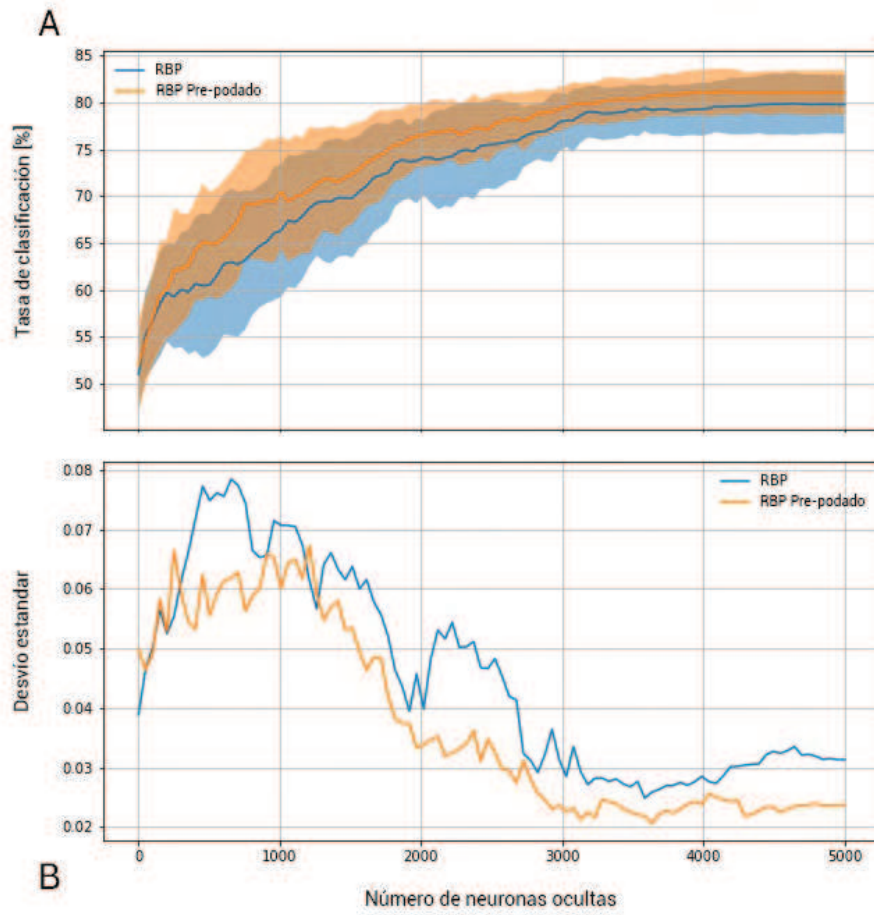


Figura 4.7: A: Desempeño de clasificación sobre el conjunto de prueba para los métodos combinados sobre datos artificialmente generados. B: Desvío estándar de la tasa de clasificación para las distintas inicializaciones del modelo y validaciones cruzada de los datos.



## Sección 5

# Análisis y clasificación del habla interna

### 5.1. Análisis global del habla interna

Para encontrar diferencias y similitudes entre el habla interna y los demás paradigmas adquiridos en la base de datos, se realizó un primer análisis temporal, espectral y espacial de las señales obtenidas. Los análisis presentes en esta sección fueron realizados utilizando todos los trials disponibles para cada paradigma, indistintamente de cada participante o clase.

#### 5.1.1. Potenciales relacionados a eventos

Los potenciales relacionados con eventos (event related potentials, ERPs) son respuestas típicas de la actividad cerebral producida en respuesta a determinados estímulos. Como el protocolo de estimulación propuesto presentaba diferentes estímulos visuales, se realizó un análisis temporal de las señales para encontrar la actividad cerebral modulada por estos estímulos. Es de esperar que esta actividad se encuentre en todos los participantes y que no tenga correlación ni con la condición ni con la clase ejecutada.

En un primer análisis se calculó el promedio de cada canal para todos los trials disponibles, indistintamente de la clase, paradigma o participante, obteniendo un total de 5640 segmentos, para cada uno de los 128 canales. El promedio final de la ventana de tiempo completa, con marcas para cada evento, se muestra en la Figura 5.1. Entre  $t = 0, 1$  s y  $t = 0, 2$  s una onda positiva-negativa-positiva se presenta en respuesta a la aparición del punto de fijación en la pantalla. Un comportamiento similar, pero más pronunciado, se observa entre  $t = 0, 6$  s y  $t = 0, 7$  s, modulado por la aparición del triángulo que indica la clase correspondiente al trial. Unos instantes antes que el triángulo desaparezca comienza un potencial negativo pronunciado, probablemente relacionado con la preparación motora y cognitiva de los participantes para ejecutar la acción solicitada. La señal parece ser mayormente estable durante el resto del intervalo de acción. Finalmente un potencial positivo aparece entre  $t = 3, 8$  s y  $t = 3, 9$  s, en respuesta al cambio de color del círculo.

#### 5.1.2. Análisis espacial y espectral

Con el objetivo de encontrar y analizar más diferencias y similitudes entre las tres condiciones, se calculó una representación de tiempo-frecuencia (Time-Frequency Representation, TFR) mediante una transformada ondita (Wavelet) (Mallat, 1999). Todos los trials disponibles para cada paradigma, indistintamente del participante o de la clase, fueron utilizados para calcular la TFR.

En esta representación se utilizó una ventana Morlet Wavelet, con la cual se obtiene un número complejo, producto de la convolución de la señal y la ventana compleja propuesta. La implementación

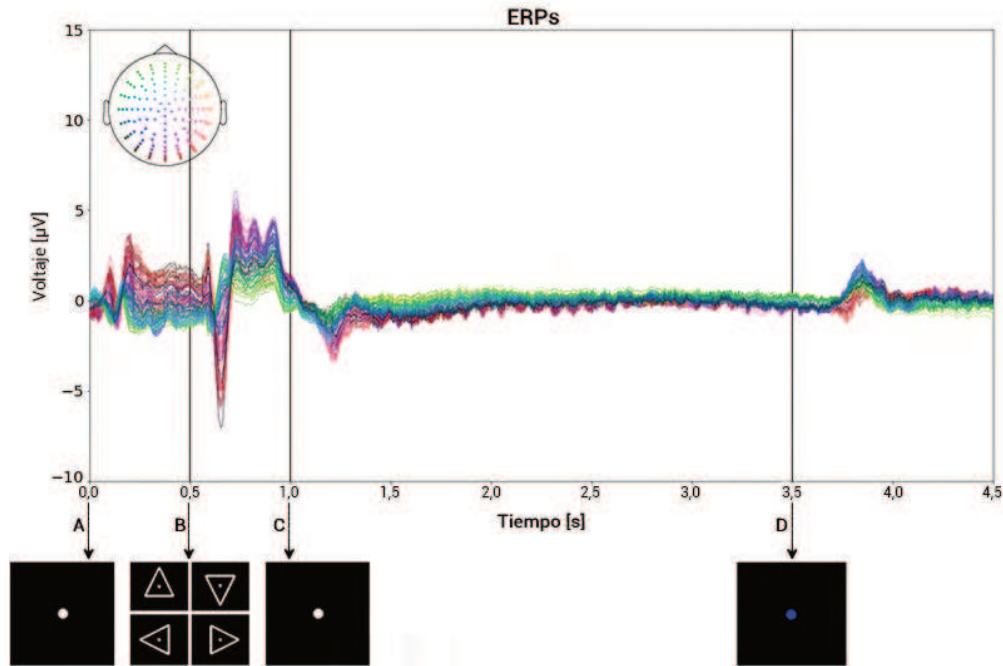


Figura 5.1: Promedio temporal de todos los trials disponibles; todos los canales se trazaron con una ubicación de referencia de color. A-B Intervalo de concentración. B-C Indicación de clase. C-D Intervalo de acción. D-Intervalo de relajación.

de MNE de esta transformación devuelve la potencia de la señal y la coherencia entre ensayos (inter-trial coherence) como la parte real e imaginaria de la TFR, respectivamente.

Para detectar las regiones en las que la actividad neural entre las condiciones es diferente, se restaron las potencias en las principales bandas de frecuencia entre el habla interna y cada uno de los otros dos paradigmas. El objetivo de las sustracciones es minimizar los efectos de las actividades comunes producidas por el protocolo, las cuales deberían debería cancelarse en la resta. En análisis se llevó a cabo en la ventana de tiempo de 1,5 a 3,5 s, para evitar al máximo los potenciales evocados producidos por el protocolo de estimulación. La densidad espectral de potencia (Power Spectral Density, PSD) se añadió al análisis para explorar en profundidad las regiones de interés. Las áreas sombreadas en los gráficos de PSD corresponden a  $\pm 1$  desvío estándar de los diferentes canales utilizados. En caso de no mostrarse ninguna zona sombreada, un solo canal fue utilizado para calcular la PSD.

### Diferencias entre el habla interna y pronunciada

En la Figura 5.2 se muestra la comparación entre el habla interna y pronunciada. En la banda alfa, se aprecia claramente una mayor actividad en los trial de habla interna en la región central occipital/parietal. La PSD presente en el panel IP-1 se calculó utilizando los canales A4, A5, A19, A20 y A32, nomenclatura de BioSemi para su equipo de 128 canales, y muestra una diferencia entre la potencia media de aproximadamente 1 dB a 11 Hz.

Por otro lado, en la banda beta, la distribución espacial de la diferencia de potencia muestra una mayor actividad temporal para la condición pronunciada, consistente con la actividad de la corteza motora. Aquí, la PSD se calculó utilizando los canales B16, B22, B24 y B29 para el panel IP-2 y los canales D10, D19, D21 y D26 para el panel IP-3. El habla pronunciada muestra una mayor potencia en toda la banda beta, con una diferencia más prominente en la zona central izquierda. La referencia de los canales utilizados puede encontrarse en la Figura 3.1.

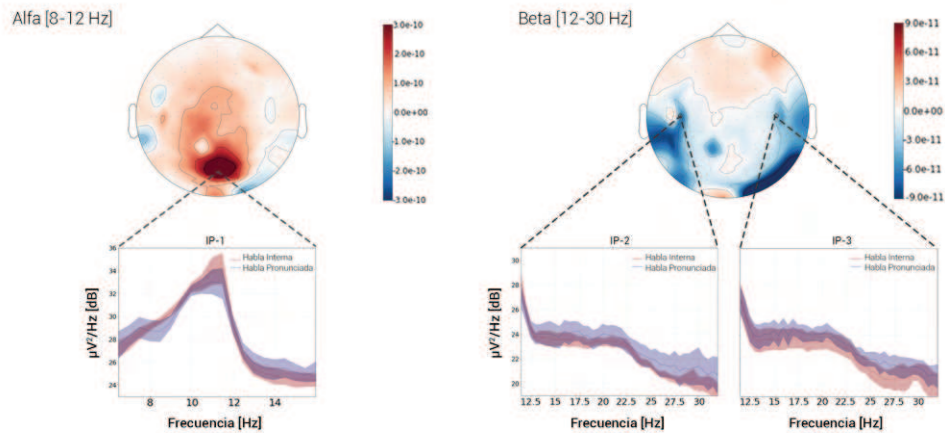


Figura 5.2: Diferencias espaciales y espectrales en la actividad de los paradigmas habla interna y habla pronunciada

### Diferencias entre el habla interna y la condición visualizada

En la Figura 5.3 se muestra la comparación entre el habla interna y la condición visualizada. Los trials de la condición visualizada muestran una mayor actividad en las regiones occipitales laterales, tanto en la banda alfa como beta. Esto puede deberse a que la condición visualizada supondría una componente visual más fuerte, generando una marcada mayor actividad visual occipital. Los trials de habla interna muestran una mayor potencia en la banda alfa en regiones más parietales. Para la banda alfa, los PSDs mostrados en los paneles IV-1 y IV-2 se calcularon utilizando los canales A10 y B7, respectivamente. En ambos gráficos, el pico de potencia correspondiente a la condición de habla interna es mayor que la de la condición de visualización. Para la banda beta, los PSD que se muestran en los paneles IV-3 y IV-4 se calcularon utilizando los canales A13 y A26, respectivamente. La potencia para los trials de la condición visualizada es mayor a la de habla interna en toda la banda.

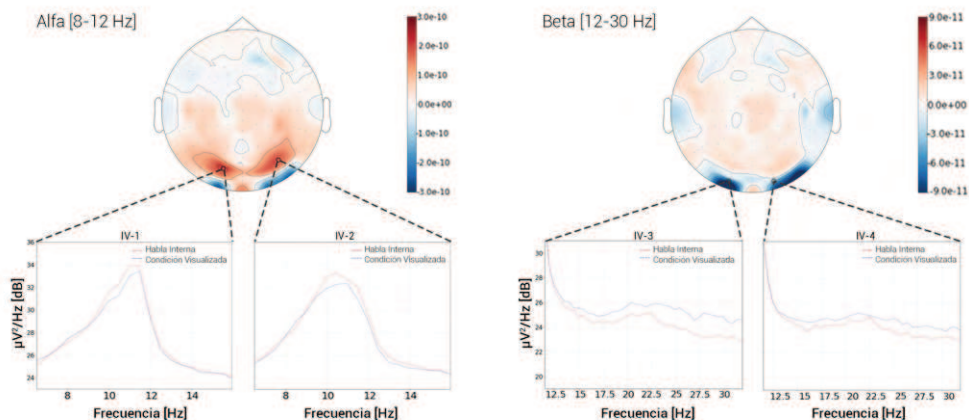


Figura 5.3: Diferencias espaciales y espectrales en la actividad de los paradigmas habla interna y condición visualizada

## 5.2. Clasificación de señales de habla interna

Para analizar la utilidad del habla interna como paradigma de control de una ICC, sería necesario demostrar primero que este paradigma es distinguible de los otros paradigmas registrados. Si bien este análisis no garantizaría que las clases dentro de cada paradigma puedan distinguirse, podría aportar valiosa información sobre cuáles son las características más distintivas de cada paradigma. Parte de los resultados en esta sección forman parte del trabajo presentado en el Anexo IV, aprobado para su publicación en el Simposio Argentino de Inteligencia Artificial en el marco de las Jornadas Argentinas de Informática, 50 JAIIO.

### 5.2.1. Análisis preliminar de factibilidad

En este primer análisis se entrenó un clasificador para distinguir entre las señales de cada uno de los paradigmas adquiridos en la base de datos, comparados de a pares: habla interna con habla pronunciada y habla interna con condición visualizada. Para esto se utilizó un procesamiento estándar de extracción de características en conjunto con el método RBP, propuesto en el Anexo II.

#### Extracción de características

Las señales de EEG fueron utilizadas con el mismo preprocesamiento propuesto en el Anexo I. Para cada trial, solamente los últimos 2 segundos de grabación fueron utilizados, para evitar potenciales evocados por el protocolo de estimulación.

La extracción de características se realizó utilizando un Banco de Filtros de Patrones Espaciales Comunes (Filter Bank Common Spatial Pattern, FBCSP) (Ang *et al.*, 2008). Las señales fueron filtradas en 6 bandas frecuenciales de interés para las señales de EEG: Delta [0.5 - 4 Hz], Theta [4 - 8 Hz], Alfa [8 - 12 Hz], Beta bajo [12 - 20], Beta alto [20 - , 30 Hz] y Gamma [30 - 45 Hz]. Para cada banda frecuencial, se generó un filtro CSP (Wang *et al.*, 2006). Para cada banda de frecuencia, las señales fueron proyectadas en los seis primeros filtros espaciales obtenidos. De cada una de las señales proyectadas se calculó la energía promedio, obteniéndose así un vector de 36 características (dimensiones), correspondientes a 6 filtros espaciales y 6 filtros espectrales.

#### Desempeño de clasificación

La clasificación se realizó para cada uno de los participantes por separado, utilizando el método RBP, propuesto en el Anexo II, con la versión regularizada de las ELM propuesta en (Deng *et al.*, 2009). El parámetro de regularización fue fijado en 1 y no fue optimizado a lo largo del entrenamiento. Como se muestra en la Tabla 3.1, cada participante presenta aproximadamente 100 datos para el paradigma pronunciado y 200 para los otros dos paradigmas. Los datos se separaron en 80%, 20%, 10% para entrenamiento, validación y evaluación, respectivamente. El parámetro de neuronas máximas  $M^*$  fue seteado en 1000 y el paso de búsqueda fue de 50. Para generar una distribución que nos permita comparar el desempeño del clasificador, se realizaron 20 validaciones cruzadas sobre los datos tanto con las etiquetas reales como con las etiquetas aleatorizadas. De esta forma, se obtienen dos distribuciones de desempeño, que representan una comparación más justa y rigurosa que la comparación con el nivel de chance.

En la Figura 5.4 se muestran las distribuciones obtenidas para cada participante en la comparación entre habla interna y condición visualizada. Un test estadístico de Mann–Whitney–Wilcoxon fue realizado entre las distribuciones obtenidas con las etiquetas reales y aleatorizadas de cada participantes. Para todos los participantes las distribuciones fueron significativamente diferentes ( $p < 0.001$ ).

Por otro lado, En la Figura 5.5 se muestran las distribuciones obtenidas para cada participante en la comparación entre habla interna y pronunciada. El mismo test fue realizado, arrojando el mismo resultado para todos los participantes ( $p < 0.001$ ). Como era de esperar, la comparación entre habla interna y pronunciada es más sencilla que entre habla interna y condición visualizada. Esto puede

deberse a que no todo el ruido muscular fue eliminado de las señales de habla pronunciada y el algoritmo es capaz de capturar esta información para discriminar entre señales de cada paradigma.

Para garantizar la reproducibilidad de los resultados, la implementación utilizada en este trabajo está disponible en el siguiente enlace<sup>1</sup>.

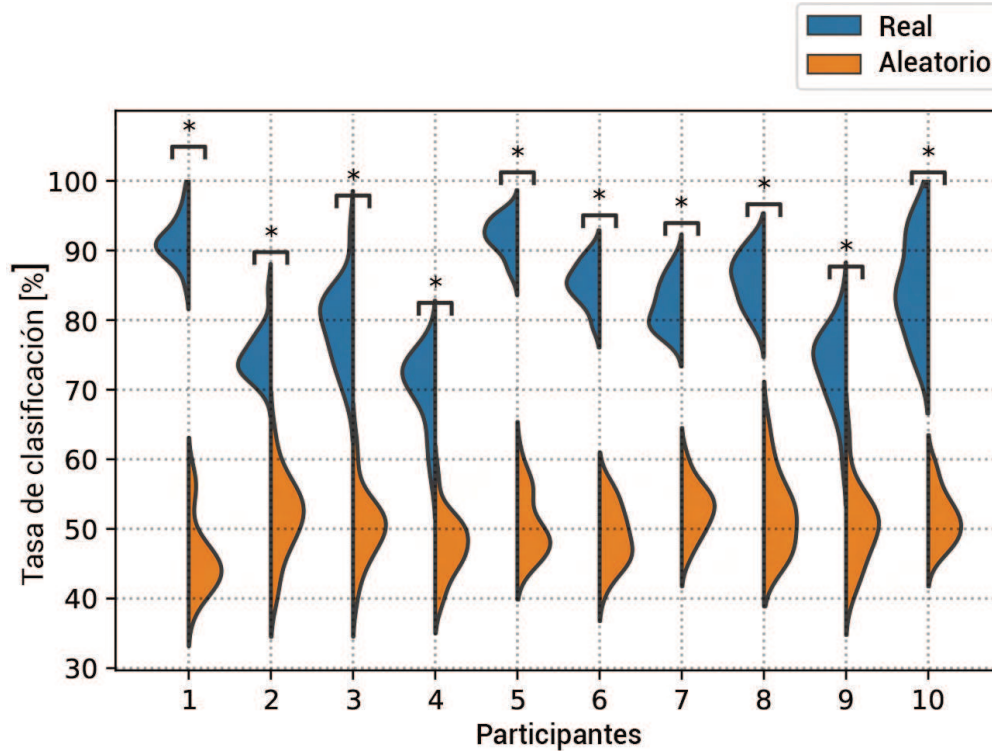


Figura 5.4: Distribución del desempeño de clasificación entre habla interna y condición visualizada. Test de estadístico de Mann–Whitney–Wilcoxon, significancia  $* = p < 0,001$

### 5.2.2. Análisis de las características relevantes

Cómo continuación del análisis presentados en el Anexo IV, un estudio sobre la influencia de cada banda frecuencial en la performance de clasificación fue realizado, siguiendo un procedimiento similar al descrito anteriormente, pero utilizando ahora un vector de solamente 6 características por trial. A diferencia del procedimiento anterior, ahora no se entrenó al clasificador con etiquetas aleatorizadas, ya que el fin de este análisis es medir la diferencia entre distintas características de la señal.

En la Figura 5.6-A se compara el habla interna y pronunciada, mostrando las distribuciones de clasificación para cada una de las bandas frecuenciales analizadas. Cada distribución se compone de las 20 validaciones cruzadas de cada uno de los participantes. En esta comparación es claro que las bandas de alta frecuencia (Beta y Gamma) contienen más información sobre el problema. Si bien una mayor actividad en la banda Beta podría estar relacionada con el ritmo sensoriomotor del habla, un detallado análisis espacial será necesario para determinar si el origen de dicha actividad es la actividad neuronal de la corteza motora o es producida por un artefacto muscular. Por otro lado, si bien las bajas frecuencias presentan una menor performance de clasificación, siguen estando por encima del nivel de chance.

<sup>1</sup> [https://github.com/N-Nieto/Feasibility\\_Analysis\\_Inner\\_Speech](https://github.com/N-Nieto/Feasibility_Analysis_Inner_Speech)

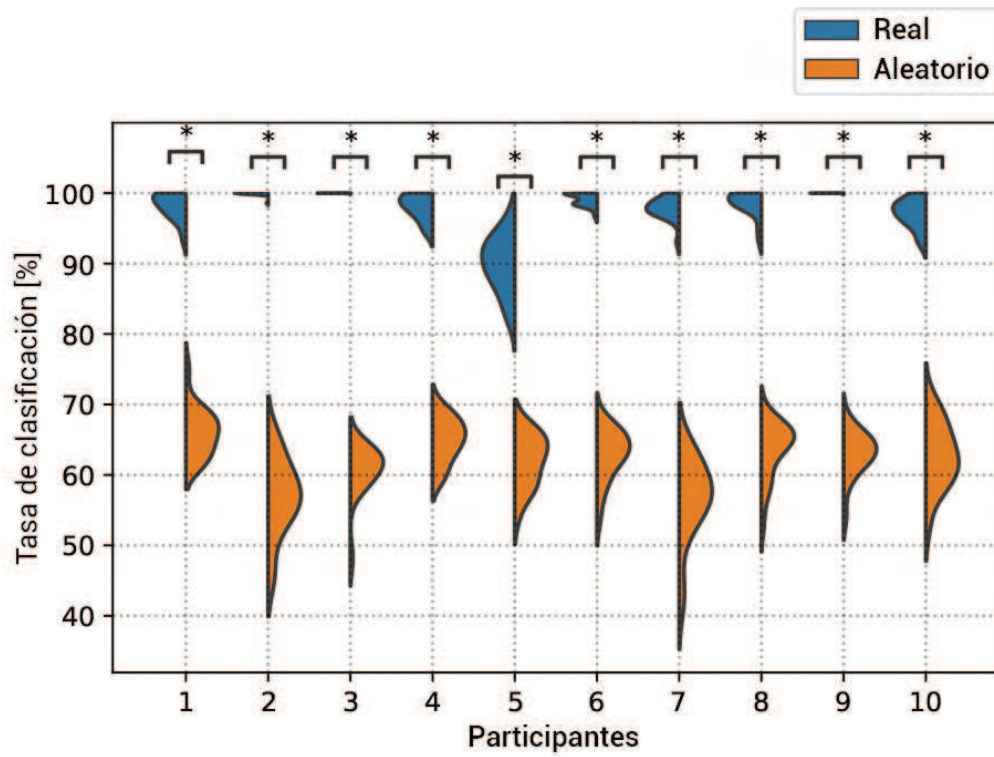


Figura 5.5: Distribución del desempeño de clasificación entre habla interna y habla pronunciada. Test de estadístico de Mann–Whitney–Wilcoxon, significancia  $*$  =  $p < 0,001$

Por su parte, en la Figura 5.6-B se muestra la comparación entre habla interna y condición visualizada. Una tendencia similar a la anterior comparación se ve en esta Figura, donde las bandas de alta frecuencia obtienen mejores performance de clasificación que las de baja frecuencia.

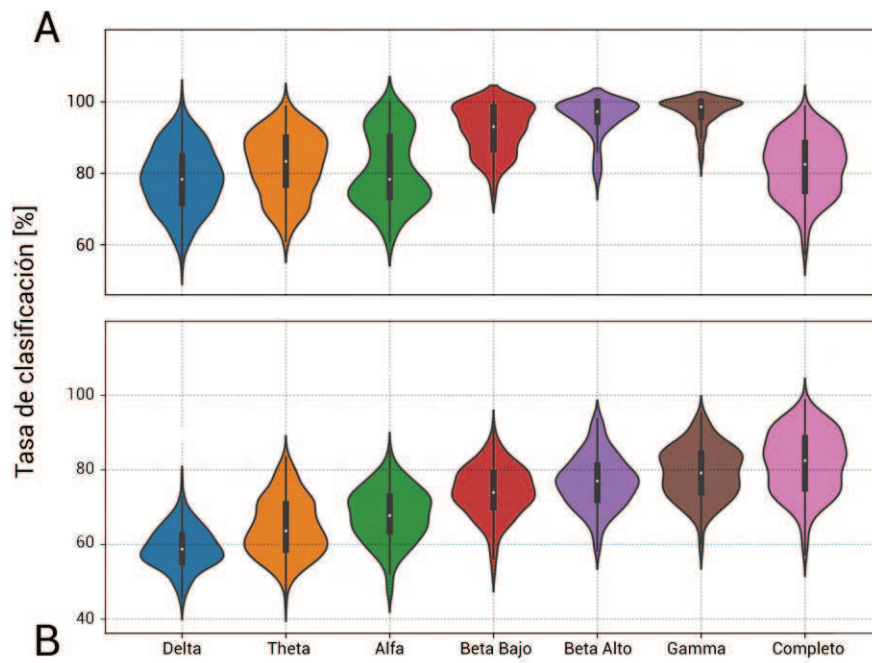


Figura 5.6: A: Distribución del desempeño de clasificación de cada banda de frecuencia entre habla interna y habla pronunciada. B: Distribución del desempeño de clasificación de cada banda de frecuencia entre habla interna y condición visualizada.





## Sección 6

# Transporte óptimo aplicado a imaginología motora

### 6.1. Transporte óptimo y adaptación de dominio aplicado a ICCs

Uno de los principales problemas en las ICCs, indistintamente del paradigma ejecutado, es la alta variabilidad inter e intra sujeto de las señales de EEG. En sus comienzos, la única solución posible era volver a adquirir nuevas señales de entrenamiento cada vez que un paciente requería utilizar una interfaz, desechando todos los datos adquiridos anteriormente. Esto trae como consecuencia la necesidad de adquirir una gran cantidad de datos en cada sesión, reduciendo el tiempo de uso real de la interfaz por parte del paciente.

Afortunadamente, distintas técnicas de alineamiento y adaptación de dominio han sido propuestas para aprovechar al máximo todos los datos disponibles y minimizar el tiempo de reentrenamiento de las interfaces (He and Wu, 2020; Samek *et al.*, 2013; Wu *et al.*, 2020; Devlaminck *et al.*, 2011; Alamgir *et al.*, 2010). Estos métodos transforman los datos para que sus distribuciones sean más similares, permitiendo así utilizar datos de distintas sesiones del mismo paciente, o incluso de distintos sujetos, para minimizar la cantidad de datos de la nueva sesión requeridos para el entrenamiento (Lotte and Guan, 2010). Esto es especialmente importante para aquellos usuarios iniciales y también para los que usarán en forma prolongada o continua una ICC.

Dentro de los métodos de adaptación de dominio, existe un conjunto de técnicas basadas en la teoría de transporte óptimo (Courty *et al.*, 2017). Acorde con este enfoque existen dos dominios: el dominio fuente (source) y el dominio objetivo (target). El primero puede corresponder tanto a datos de una sesión anterior del mismo sujeto o datos de distintos sujetos. El segundo dominio se compone de los datos de recalibración que se deben tomar en la sesión actual de uso.

#### 6.1.1. Transporte óptimo con inversión de dominios

Acorde a la teoría de transporte óptimo, los métodos transforman los datos de la sesión de entrenamiento (source) hacia los datos de recalibración (target). Una vez transportados, un clasificador es entrenado sobre el conjunto de dato de transportados y los datos de recalibración. Este enfoque, que es el tradicionalmente utilizado en el campo, tiene como principal desventaja la necesidad de reentrenar el modelo de clasificación. Esto limita el tiempo de entrenamiento del modelo utilizado, ya que, idealmente, la ICC debería poder comenzar a dar respuesta unos pocos segundos después de adquirir los datos de recalibración (Krauledat *et al.*, 2008).

Para evitar este problema es posible cambiar el sentido del transporte, aprendiendo un plan de transporte desde el dominio objetivo hacia el fuente. Este plan de transporte contiene la información

Método	R1	R2	R3	R4	R5	R6	R7	Promedio
SC	64,0	65,0	66,5	70,5	71,0	70,0	72,0	68,4
SR	68,0	69,0	71,5	81,0	74,0	75,5	71,5	72,9
RPA	67,5	70,5	71,5	75,5	75,0	68,0	76,5	72,14
EA	77,0	73,5	71,5	75,5	73,0	74,0	70,5	73,6
FOTDA-S	73,5	68,5	70,0	72,5	73,5	71,0	70,5	71,5
FOTDA-GL	71,5	68,5	70,5	73,5	72,5	70,5	70,5	71,1
BOTDA-S	73,0	71,5	71,0	73,0	70,5	72,5	69,5	71,5
BOTDA-GL	73,5	66,5	71,5	74,5	71,5	70,5	69,5	71,1

Tabla 6.1: Comparación de métodos de adaptación de dominio en la base de datos presentada en (Peterson *et al.*, 2019)

sobre la transformación que debe realizarse sobre los datos de recalibración para asemejarlos a los datos de entrenamiento. De esta forma, los datos generados en la sesión de recalibración (sesión de uso) son transportados hacia los datos de la sesión de entrenamiento en forma continua. Como los datos seguirán la distribución de los datos de entrenamiento, es posible entrenar un clasificador y extractor de características previo a la sesión, sobre los datos que fueron recolectados anteriormente.

El método, titulado “Backward optimal transport for domain adaptation” (BOTDA), se presenta en detalle en el Anexo V. Este método fue propuesto originalmente para la adaptación de dominio de dos sesiones del mismo sujeto ejecutando un paradigma de imaginería motora para la neurorehabilitación. Sin embargo, esto no es un factor limitante y bien podría ser utilizado en paradigmas relacionados al habla.

### 6.1.2. Resultados y comparaciones con métodos del estado del arte

Para evaluar el método propuesto, se utilizaron las bases de datos de imaginería motora descriptas en la Sección 3. Para la comparación del método, se utilizó el enfoque tradicional de adaptación de dominio “Fordward optimal transport for domain adaptation” (FOTDA) junto con métodos del estado del arte: “Euclidean alignment” (EA) (He and Wu, 2019) y “Riemannian procrustes analysis” (RPA) (Rodrigues *et al.*, 2018). Además, se realizaron comparaciones de clasificación sobre los datos sin adaptación de dominio, permitiendo recalibración (SR) y sin recalibración (SC).

Para los métodos de transporte óptimo, FOTDA y BOTDA, se utilizaron las versiones regularizadas de transporte óptimo, “Sinkhorn” (S) (Cuturi, 2013) y “Grup Lasso” (GL) (Knight, 2008).

En la Tabla 6.1 se presentan los resultados de clasificación para todos los sujetos en los distintos bloques de estimulación de la base de datos presentada en (Peterson *et al.*, 2019). En la Tabla 6.2 se los resultados obtenidos con los datos de (Faller *et al.*, 2012).

El método propuesto obtiene un desempeño de clasificación similar al de todos los métodos comparados, sin la necesidad de reentrenar el clasificador o la etapa de extracción de características. Esto le confiere un tiempo computacional mucho menor, comparado con los otros métodos, como se observa en la Tabla 6.3. Es importante remarcar también que evitar el reentrenamiento de la etapa de clasificación permitiría entrenar modelos más grandes y complejos, ya que este entrenamiento puede ser realizado antes de la sesión donde el paciente utiliza la ICC.

### 6.1.3. Transporte óptimo con información del clasificador

Una de las principales limitaciones de BOTDA es que elige de forma aleatoria, mediante validación cruzada, un subconjunto de muestras del conjunto source para aprender el plan de transporte. De esta forma, el algoritmo depende fuertemente de que las muestras elegidas sean representativas de la distribución del dominio, desechando muestras que podrían contener información útil.

Como continuación de dicho trabajo, se propuso un nuevo método para sortear este problema, utilizando información tanto del clasificador como de la distribución de datos en el dominio fuente y

Método	R1	R2	R3	R4	R5	R6	R7	R8	R9	Promedio
SC	77,9	78,8	74,6	72,5	74,6	75,8	74,6	77,5	74,6	75,6
SR	80,8	81,7	78,8	80,0	79,2	79,6	79,6	85,0	78,8	80,4
RPA	77,5	76,3	72,9	80,8	77,5	79,6	80,0	83,3	79,2	78,6
EA	77,1	75,8	83,3	82,9	78,8	78,8	79,6	79,2	80,0	79,9
FOTDA-S	74,2	80,2	78,3	78,7	75,4	77,5	77,5	77,5	76,7	77,4
FOTDA-GL	77,1	79,6	79,6	71,9	76,25	78,8	78,8	75,8	76,7	77,8
BOTDA-S	78,3	76,3	78,8	77,9	75,4	76,7	80,0	76,7	78,8	77,6
BOTDA-GL	73,8	77,1	77,5	76,7	75,0	76,3	77,9	75,8	75,4	76,2

Tabla 6.2: Comparación de métodos de adaptación de dominio en la base de datos presentada en (Faller *et al.*, 2012)

Método	(Peterson <i>et al.</i> , 2019)	(Faller <i>et al.</i> , 2012)
SC	$0,326 \pm 0,008$	$0,337 \pm 0,015$
RPA	$2,711 \pm 0,557$	$1,099 \pm 0,324$
EA	$0,634 \pm 0,109$	$0,537 \pm 0,089$
BOTDA-GL	$0,034 \pm 0,010$	$0,029 \pm 0,014$

Tabla 6.3: Comparación del tiempo computacional requerido por los métodos de adaptación de dominio para ambas bases de dato

objetivo. Este método aún no fue publicado pero distintas pruebas de concepto y resultados preliminares muestran que es posible mantener, e incluso mejorar, el desempeño de clasificación del clasificador, eliminando la selección del subconjunto del dominio fuente.

### Incorporando información del clasificador

Utilizando el enfoque propuesto en el método BOTDA, el clasificador se entrena sobre los datos del conjunto fuente. De esta forma, es posible conocer la frontera de decisión del clasificador y su desempeño con los datos de entrenamiento. Ahora, la teoría de transporte óptimo exige que cada punto del dominio fuente reciba igual cantidad de masa del dominio objetivo, sin embargo, ¿son todos los puntos de la distribución fuente igualmente útiles para el transporte?. Idealmente quisiéramos que los puntos del dominio objetivo no sean transportados cercanos a la frontera de decisión del clasificador ya entrenado. De esta forma, se propuso un nuevo método de asignación de masas, donde la estas son asignadas a cada punto del dominio de fuente en forma proporcional a su distancia a la frontera de decisión del clasificador. Esto permite que no sea necesaria la elección de un subconjunto de transporte, teniendo entonces información en, mayor o menor medida, de todas las muestras de entrenamiento.

### Prueba de concepto con datos sintéticos

Para corroborar el efecto del método propuesto, en la Figura 6.1-A se muestran los datos artificialmente generados en conjunto con la recta de decisión obtenida por un clasificador LDA, entrenado sobre los mismos datos.

Cómo se observa en la Figura 6.1-A, la distribución de las distintas clases se encuentran parcialmente solapadas, existiendo puntos del conjunto mal clasificados por el modelo. Idealmente, las muestras transportadas a este dominio no deberían parecerse a estos puntos, por lo que la masa de los mismos fue igualada a 0. En el siguiente paso, las masas de los puntos fue redistribuida de forma tal que aquellos que se encuentran a una mayor distancia de la recta de decisión posean una masa proporcionalmente más grande. Este paso es realizado manteniendo las mismas restricciones del problema de optimización de transporte óptimo, donde la suma total de masas de todos los puntos del dominio debe ser igual a 1. Luego de los pasos descritos, se obtiene una distribución de puntos tal como se muestra en la Figura 6.1-B, donde el tamaño de los puntos es proporcional a la masa asignada a los mismos, que

es directamente proporcional también a la distancia de los puntos a la recta de clasificación. Aquellos puntos del conjunto que se encontraban mal clasificados desaparecieron, al no tener masa.

De esta forma, todos los puntos del conjunto de entrenamiento son tenidos en cuenta y aportan, en distinto grado, información sobre la distribución del dominio fuente. Esto evita realizar una validación cruzada para seleccionar el subconjunto de transporte, pero agrega nuevo hiperparámetro que penaliza en distinto grado la distancia de un punto a la frontera de decisión. Cuando este parámetro es igual a 0, la masa de todos los puntos se distribuirá de forma homogénea, tal como en el método tradicional. Por otro lado, cuando el parámetro tiende a infinito, solo un punto por clase, el más alejado de la recta de decisión, tendrá una masa distinta de 0.

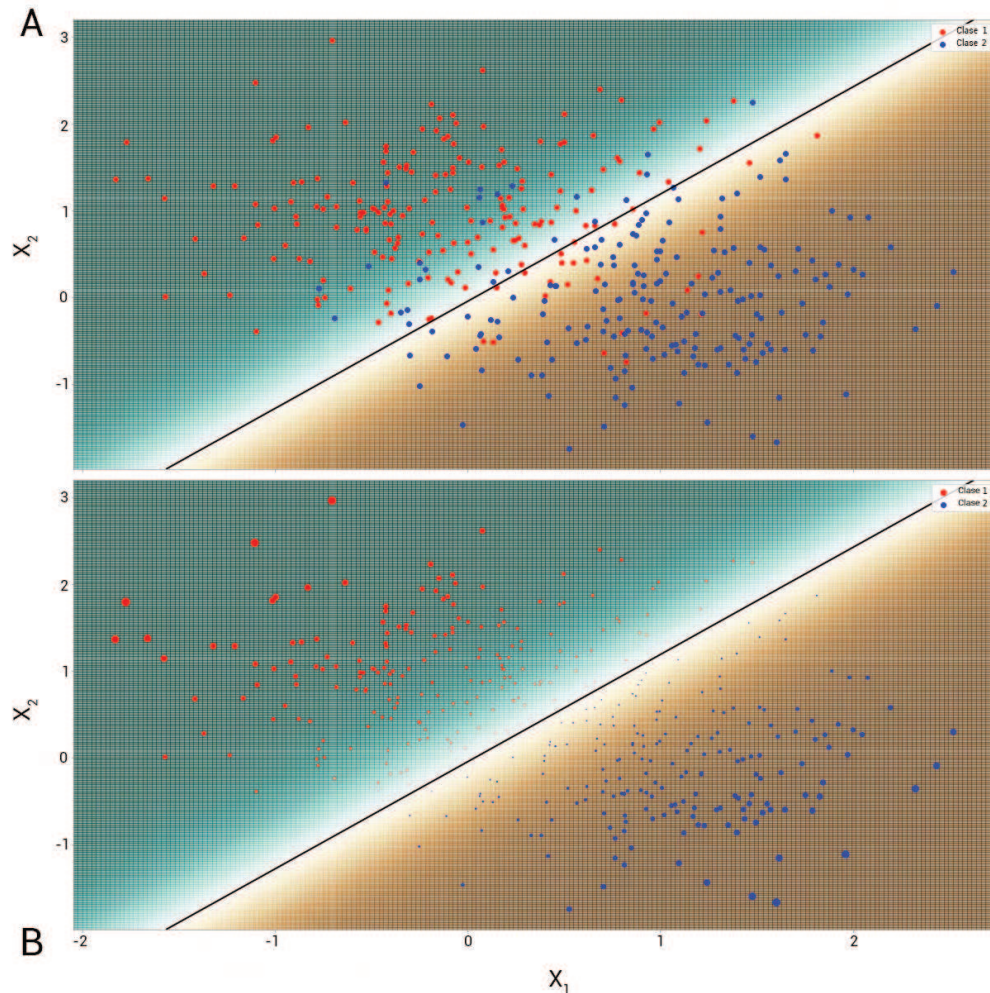


Figura 6.1: Distribución de masas en el conjunto fuente. A: Masas originales, obtenidas con métodos tradicionales de transporte óptimo. B: Distribución de masas obtenidas con el método propuesto

### Trabajos futuros

Cómo primer punto, es necesario realizar un análisis sistemático y detallado del nuevo método propuesto con información del clasificador, principalmente utilizando datos de EEG. En segundo lugar, utilizar este método para ampliar la cantidad de datos disponibles para el entrenamiento en las bases de datos de paradigmas relacionados con el habla.

Por otro lado, es posible también utilizar como alternativa el puntaje de clasificación (o probabilidad) que el clasificador le asigna a cada punto del dominio como medida alternativa a la distancia de los puntos a la frontera de decisión. Esto permitiría utilizar cualquier algoritmo que devuelva una probabilidad de clasificación, asignando masas directamente proporcionales a este puntaje obtenido. De esta forma, puntos donde el clasificador se encuentra muy “seguro”, obtendrán más masa, mientras que aquellos puntos con mayor incertidumbre tendrán una masa menos.

Por último, los métodos de transporte son muy sensibles a los problemas que presentan distintos desbalances de clases en los dominios fuente y objetivo. Esto se debe a que la cantidad de masa de cada clase no es igual, por lo que una clase tendrá preponderancia sobre la otra, forzando a los puntos transportados a adoptar ese desbalance de clases, aún cuando los datos transportados no lo tengan. Esto también puede ser atacado mediante el método propuesto, igualando la cantidad de masa asignada para cada clase.



## Sección 7

# Conclusiones

En la presente tesis se han realizado diferentes contribuciones para avanzar en la utilización de los paradigmas relacionados con el habla, tanto del habla interna como imaginada, como paradigma de control de interfaces cerebro-computadora. Sin lugar a dudas, y continuando en esta misma línea, la clasificación de las distintas clases dentro de los paradigmas obtenidos serán el próximo desafío. El uso de modelos simples de clasificación no fueron suficientes para obtener resultados superiores a la clasificación aleatoria, por lo que un análisis más exhaustivo y detallado, tanto de la etapa de extracción de características como de clasificación deben ser desarrollados.

Los algoritmos de aprendizaje profundo podrían representar una herramienta valiosa para esta aplicación, no obstante, la cantidad de datos disponibles por participante, desglosada por paradigma y clase, no permitiría entrenar grandes estructuras de forma directa. Para enfrentar este problema, las técnicas de transporte óptimo propuestas podrían ayudar significativamente a esta tarea, aumentando la cantidad de datos disponibles para el entrenamiento y permitiendo el uso de redes neuronales profundas. Las relaciones espaciales, que aportan información crucial para el problema de clasificación de señales de EEG, no son completamente capturadas por las redes profundas, por lo que el uso de grafos sería un camino para explotar al máximo esta importante relación.

Por su parte, la base de datos adquirida y publicada es un pilar fundamental para el desarrollo de futuros algoritmos e investigaciones. Sumado a los datos, las librerías desarrolladas facilitan la reutilización de los datos y permiten modificar sencillamente el procesamiento de los datos. Ofrecer las señales tal como fueron grabadas, en conjunto con una detallada descripción del procedimiento, brinda a la comunidad la posibilidad de evaluar nuevas y distintas hipótesis sobre los mismos datos e incluso permitir el desarrollo de nuevas técnicas de eliminación de artefactos. Además, los datos estructurados y preprocesados permiten una rápida y fácil reutilización por parte de la comunidad de inteligencia artificial.

Finalmente, los desarrollos sobre técnicas de podado propuestas han mostrado una utilidad concreta y un potencial aún por explorar, ahorrando no solo tiempo de cómputo, sino también mejorando el desempeño de clasificación y la estabilidad de los algoritmos.





# Bibliografía

- Abiri, R., Borhani, S., Sellers, E. W., Jiang, Y., and Zhao, X. (2019). A comprehensive review of EEG-based brain–computer interface paradigms. *16*(1), 011001.
- Ahn, M. and Jun, S. C. (2015). Performance variation in motor imagery brain–computer interface: a brief review. *Journal of Neuroscience Methods*, **243**, 103–110.
- Alamgir, M., Grosse-Wentrup, M., and Altun, Y. (2010). Multitask learning for brain-computer interfaces. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 17–24. JMLR Workshop and Conference Proceedings.
- Alderson-Day, B. and Fernyhough, C. (2015). Inner speech: development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, **141**(5), 931.
- Alencar, A. S., Neto, A. R. R., and Gomes, J. P. P. (2016). A new pruning method for extreme learning machines via genetic algorithms. *Applied Soft Computing*, **44**, 101–107.
- Allison, B. Z., McFarland, D. J., Schalk, G., Zheng, S. D., Jackson, M. M., and Wolpaw, J. R. (2008). Towards an independent brain–computer interface using steady state visual evoked potentials. *Clinical Neurophysiology*, **119**(2), 399–408.
- Ang, K. K., Chin, Z. Y., Zhang, H., and Guan, C. (2008). Filter bank common spatial pattern (FBCSP) in brain-computer interface. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 2390–2397. IEEE.
- Ayaz, H., Shewokis, P. A., Bunce, S., and Onaral, B. (2011). An optical brain computer interface for environmental control. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6327–6330. IEEE.
- Belkin, M., Hsu, D., Ma, S., and Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, **116**(32), 15849–15854.
- Bell, A. J. and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, **7**(6), 1129–1159.
- Bernat, E. M., Nelson, L. D., Holroyd, C. B., Gehring, W. J., and Patrick, C. J. (2008). Separating cognitive processes with principal components analysis of eeg time-frequency distributions. In *Advanced Signal Processing Algorithms, Architectures, and Implementations XVIII*, volume 7074, page 70740S. International Society for Optics and Photonics.
- Birbaumer, N., Elbert, T., Canavan, A. G., and Rockstroh, B. (1990). Slow potentials of the cerebral cortex and behavior. *Physiological reviews*, **70**(1), 1–41.
- Birbaumer, N., Hinterberger, T., Kubler, A., and Neumann, N. (2003). The thought-translation device (ttdd): neurobehavioral mechanisms and clinical outcome. *IEEE transactions on Neural Systems and rehabilitation engineering*, **11**(2), 120–123.

- Blank, S. C., Scott, S. K., Murphy, K., Warburton, E., and Wise, R. J. (2002). Speech production: Wernicke, broca and beyond. *Brain*, **125**(8), 1829–1838.
- Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Muller, K.-R. (2007). Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal processing magazine*, **25**(1), 41–56.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, **10**(4), 433–436.
- Brigham, K. and Kumar, B. V. (2010). Imagined speech classification with EEG signals for silent communication: a preliminary investigation into synthetic telepathy. In *2010 4th International Conference on Bioinformatics and Biomedical Engineering*, pages 1–4. IEEE.
- Brumberg, J. S., Nieto-Castanon, A., Kennedy, P. R., and Guenther, F. H. (2010). Brain–computer interfaces for speech communication. *Speech communication*, **52**(4), 367–379.
- Buolamwini, J. and Geburu, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *ACM FAccT*, pages 77–91.
- Char, D. S., Shah, N. H., and Magnus, D. (2018). Implementing machine learning in health care—addressing ethical challenges. *The New England journal of medicine*, **378**(11), 981.
- Cooney, C., Folli, R., and Coyle, D. (2018). Neurolinguistics research advancing development of a direct-speech brain-computer interface. *IScience*, **8**, 103–125.
- Cooney, C., Korik, A., Raffaella, F., and Coyle, D. (2019). Classification of imagined spoken word-pairs using convolutional neural networks. In *The 8th Graz BCI Conference, 2019*, pages 338–343.
- Courty, N., Flamary, R., Tuia, D., and Rakotomamonjy, A. (2017). Optimal transport for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**(9), 1853–1865.
- CSIC (2021). Áreas del cerebro. <https://www.dicyt.com/viewItem.php?itemId=40057>.
- Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, **26**, 2292–2300.
- DaSalla, C. S., Kambara, H., Sato, M., and Koike, Y. (2009). Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Networks*, **22**(9), 1334–1339.
- De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O’Donoghue, B., Visentin, D., *et al.* (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature medicine*, **24**(9), 1342.
- Deng, S., Srinivasan, R., Lappas, T., and D’Zmura, M. (2010). EEG classification of imagined syllable rhythm using Hilbert spectrum methods. *Journal of Neural Engineering*, **7**(4), 046006.
- Deng, W., Zheng, Q., and Chen, L. (2009). Regularized extreme learning machine. In *2009 IEEE symposium on computational intelligence and data mining*, pages 389–395. IEEE.
- Devlaminc, D., Wyns, B., Grosse-Wentrup, M., Otte, G., and Santens, P. (2011). Multisubject learning for common spatial patterns in motor-imagery bci. *Computational intelligence and neuroscience*, **2011**.
- Ding, S., Zhang, N., Xu, X., Guo, L., and Zhang, J. (2015). Deep extreme learning machine and its application in eeg classification. *Mathematical Problems in Engineering*, **2015**.
- Ding, S., Guo, L., and Hou, Y. (2017). Extreme learning machine with kernel model based on deep learning. *Neural Computing and Applications*, **28**(8), 1975–1984.

- D’Zmura, M., Deng, S., Lappas, T., Thorpe, S., and Srinivasan, R. (2009). Toward EEG sensing of imagined speech. In *International Conference on Human-Computer Interaction*, pages 40–48. Springer.
- EBMC (2015). Homunculus: Somatosensory and somatomotor cortex. <https://www.ebmconsult.com/articles/homunculus-sensory-motor-cortex>.
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, **542**(7639), 115.
- Faller, J., Vidaurre, C., Solis-Escalante, T., Neuper, C., and Scherer, R. (2012). Autocalibration and recurrent adaptation: Towards a plug and play online erd-bci. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **20**(3), 313–319.
- Farwell, L. A. and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology*, **70**(6), 510–523.
- Fiez, J. A. and Petersen, S. E. (1998). Neuroimaging studies of word reading. *Proceedings of the National Academy of Sciences*, **95**(3), 914–921.
- Frith, C. D. and Done, D. J. (1988). Towards a neuropsychology of schizophrenia. *The British Journal of Psychiatry*, **153**(4), 437–443.
- Gonzalez-Lopez, J. A., Gomez-Alanis, A., Martín-Doñas, J. M., Pérez-Córdoba, J. L., and Gomez, A. M. (2020). Silent speech interfaces for speech restoration: A review. *IEEE Access*.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., and Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *Neuroimage*, **86**, 446–460.
- Guenther, F. H., Brumberg, J. S., Wright, E. J., Nieto-Castanon, A., Tourville, J. A., Panko, M., Law, R., Siebert, S. A., Bartels, J. L., Andreasen, D. S., *et al.* (2009). A wireless brain-machine interface for real-time speech synthesis. *PLoS ONE*, **4**(12).
- He, H. and Wu, D. (2019). Transfer learning for brain–computer interfaces: A euclidean space data alignment approach. *IEEE Transactions on Biomedical Engineering*, **67**(2), 399–410.
- He, H. and Wu, D. (2020). Transfer Learning for Brain-Computer Interfaces: A Euclidean Space Data Alignment Approach. *IEEE Transactions on Biomedical Engineering*, **67**(2), 399–410.
- Hickok, G. and Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, **8**(5), 393–402.
- Hinterberger, T., Schmidt, S., Neumann, N., Mellinger, J., Blankertz, B., Curio, G., and Birbaumer, N. (2004). Brain-computer communication and slow cortical potentials. *IEEE Transactions on Biomedical Engineering*, **51**(6), 1011–1018.
- Hooker, S. (2021). Moving beyond “algorithmic bias is a data problem”. *Patterns*, **2**(4), 100241.
- Huang, G.-B. and Babri, H. A. (1998). Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions. *IEEE Transactions on Neural Networks*, **9**(1), 224–229.
- Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., *et al.* (2004). Extreme learning machine: a new learning scheme of feedforward neural networks. *Neural Networks*, **2**, 985–990.

- Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, **70**(1-3), 489–501.
- Huang, G.-B., Wang, D. H., and Lan, Y. (2011). Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics*, **2**(2), 107–122.
- Indefrey, P. and Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, **92**(1-2), 101–144.
- Iriarte, J., Urrestarazu, E., Valencia, M., Alegre, M., Malanda, A., Viteri, C., and Artieda, J. (2003). Independent component analysis as a tool to eliminate artifacts in eeg: a quantitative study. *Journal of clinical neurophysiology*, **20**(4), 249–257.
- Iversen, I., Ghanayim, N., Kübler, A., Neumann, N., Birbaumer, N., and Kaiser, J. (2008). A brain–computer interface tool to assess cognitive functions in completely paralyzed patients with amyotrophic lateral sclerosis. *Clinical neurophysiology*, **119**(10), 2214–2223.
- Jrad, N., Congedo, M., Phlypo, R., Rousseau, S., Flamary, R., Yger, F., and Rakotomamonjy, A. (2011). sw-svm: sensor weighting support vector machines for eeg-based brain–computer interfaces. *Journal of neural engineering*, **8**(5), 056004.
- Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S., Hudspeth, A. J., and Mack, S. (2000). *Principles of neural science*, volume 5. McGraw-hill New York.
- Knight, P. A. (2008). The sinkhorn–knopp algorithm: convergence and applications. *SIAM Journal on Matrix Analysis and Applications*, **30**(1), 261–275.
- Krauledat, M., Tangermann, M., Blankertz, B., and Müller, K.-R. (2008). Towards zero training for brain–computer interfacing. *PloS one*, **3**(8), e2967.
- Kübler, A., Kotchoubey, B., Hinterberger, T., Ghanayim, N., Perelmouter, J., Schauer, M., Fritsch, C., Taub, E., and Birbaumer, N. (1999). The thought translation device: a neurophysiological approach to communication in total motor paralysis. *Experimental brain research*, **124**(2), 223–232.
- Larrazabal, A. J., Nieto, N., Peterson, V., Milone, D. H., and Ferrante, E. (2020). Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proceedings of the National Academy of Sciences*, **117**(23), 12592–12594.
- Ledesma-Ramirez, C., Bojorges-Valdez, E., Yáñez-Suarez, O., Saavedra, C., Bougrain, L., and Gentiletti, G. G. (2010). An open-access P300 speller database. In *Fourth International Brain-Computer Interface Meeting*, Asilomar, California, USA.
- Leuthardt, E. C., Schalk, G., Wolpaw, J. R., Ojemann, J. G., and Moran, D. W. (2004). A brain–computer interface using electrocorticographic signals in humans. *Journal of Neural Engineering*, **1**(2), 63.
- Liang, N.-Y., Saratchandran, P., Huang, G.-B., and Sundararajan, N. (2006). Classification of mental tasks from EEG signals using extreme learning machine. *International Journal of Neural Systems*, **16**(01), 29–38.
- Lieberman, P. (2007). The evolution of human speech: Its anatomical and neural bases. *Current anthropology*, **48**(1), 39–66.
- Lotte, F. and Guan, C. (2010). Learning from other subjects helps reducing brain–computer interface calibration time. In *2010 IEEE International conference on acoustics, speech and signal processing*, pages 614–617. IEEE.

- Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., and Yger, F. (2018). A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update. *Journal of neural engineering*, **15**(3), 031005.
- Mallat, S. (1999). *A wavelet tour of signal processing*. Elsevier.
- MATLAB (2010). *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts.
- McCane, L. M., Heckman, S. M., McFarland, D. J., Townsend, G., Mak, J. N., Sellers, E. W., Zeitlin, D., Tenteromano, L. M., Wolpaw, J. R., and Vaughan, T. M. (2015). P300-based brain-computer interface (BCI) event-related potentials (ERPs): People with amyotrophic lateral sclerosis (ALS) vs. age-matched controls. *Clinical Neurophysiology*, **126**(11), 2124–2131.
- McFarland, D. J., McCane, L. M., David, S. V., and Wolpaw, J. R. (1997). Spatial filter selection for eeg-based communication. *Electroencephalography and clinical Neurophysiology*, **103**(3), 386–394.
- McGuire, P., Silbersweig, D., Murray, R., David, A., Frackowiak, R., and Frith, C. (1996). Functional anatomy of inner speech and auditory verbal imagery. *Psychological Medicine*, **26**(1), 29–38.
- Miche, Y., Sorjamaa, A., Bas, P., Simula, O., Jutten, C., and Lendasse, A. (2010). OP-ELM: optimally pruned extreme learning machine. *IEEE Transactions on Neural Networks*, **21**(1), 158–162.
- Morgan, S., Hansen, J., and Hillyard, S. (1996). Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proceedings of the National Academy of Sciences*, **93**(10), 4770–4774.
- Murugavel, A. M. and Ramakrishnan, S. (2016). Hierarchical multi-class SVM with ELM kernel for epileptic EEG signal classification. *Medical & Biological Engineering & Computing*, **54**(1), 149–161.
- Nicolas-Alonso, L. F. and Gomez-Gil, J. (2012). Brain computer interfaces, a review. *Sensors*, **12**(2), 1211–1279.
- Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottureau, B. R., and Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of vision*, **15**(6), 4–4.
- Oliphant, T. E. (2006). *A guide to NumPy*, volume 1. Trelgol Publishing USA.
- Opie, N. L. and Oxley, T. J. (2019). Removing the need for invasive brain surgery: the potential of stent electrodes.
- Oxley, T. J., Yoo, P. E., Rind, G. S., Ronayne, S. M., Lee, C. S., Bird, C., Hampshire, V., Sharma, R. P., Morokoff, A., Williams, D. L., *et al.* (2021). Motor neuroprosthesis implanted with neuro-interventional surgery improves capacity for activities of daily living tasks in severe paralysis: first in-human experience. *Journal of neurointerventional surgery*, **13**(2), 102–108.
- Pei, X., Barbour, D. L., Leuthardt, E. C., and Schalk, G. (2011). Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. *Journal of Neural Engineering*, **8**(4), 046028.
- Pels, E. G., Aarnoutse, E. J., Leinders, S., Freudenburg, Z. V., Branco, M. P., van der Vijgh, B. H., Snijders, T. J., Denison, T., Vansteensel, M. J., and Ramsey, N. F. (2019). Stability of a chronic implanted brain-computer interface in late-stage amyotrophic lateral sclerosis. *Clinical Neurophysiology*, **130**(10), 1798–1803.
- Peterson, V., Wyser, D., Lamercy, O., Spies, R., and Gassert, R. (2019). A penalized time-frequency band feature selection and classification procedure for improved motor intention decoding in multi-channel eeg. *Journal of neural engineering*, **16**(1), 016019.

- Pressel-Coreto, G., Gareis, I. E., and Rufiner, H. L. (2016). Open access database of EEG signals recorded during imagined speech. In *12th International Symposium on Medical Information Processing and Analysis (SIPAIM)*.
- Price, C. J. (2000). The anatomy of language: contributions from functional neuroimaging. *The Journal of Anatomy*, **197**(3), 335–359.
- Ravden, D. and Polich, J. (1998). Habituation of p300 from visual stimuli. *International Journal of Psychophysiology*, **30**(3), 359–365.
- Raza, S. A., Opie, N. L., Morokoff, A., Sharma, R. P., Mitchell, P. J., and Oxley, T. J. (2020). Endovascular neuromodulation: Safety profile and future directions. *Frontiers in neurology*, **11**, 351.
- Rodrigues, P. L. C., Jutten, C., and Congedo, M. (2018). Riemannian procrustes analysis: transfer learning for brain–computer interfaces. *IEEE Transactions on Biomedical Engineering*, **66**(8), 2390–2401.
- Rong, H.-J., Ong, Y.-S., Tan, A.-H., and Zhu, Z. (2008). A fast pruned-extreme learning machine for classification problem. *Neurocomputing*, **72**(1-3), 359–366.
- Rousseau, M.-C., Baumstarck, K., Alessandrini, M., Blandin, V., De Villemeur, T. B., and Auquier, P. (2015). Quality of life in patients with locked-in syndrome: Evolution over a 6-year period. *Orphanet journal of rare diseases*, **10**(1), 1–8.
- Saha, S., Mamun, K. A., Ahmed, K. I. U., Mostafa, R., Naik, G. R., Darvishi, S., Khandoker, A. H., and Baumert, M. (2021). Progress in brain computer interface: Challenges and potentials. *Frontiers in Systems Neuroscience*, **15**, 4.
- Samek, W., Meinecke, F. C., and Müller, K.-R. (2013). Transferring subspaces between subjects in brain–computer interfacing. *IEEE Transactions on Biomedical Engineering*, **60**(8), 2289–2298.
- Schalk, G., Kubanek, J., Miller, K., Anderson, N., Leuthardt, E., Ojemann, J., Limbrick, D., Moran, D., Gerhardt, L., and Wolpaw, J. (2007). Decoding two-dimensional movement trajectories using electrocorticographic signals in humans. *Journal of neural engineering*, **4**(3), 264.
- Schmidt, W. F., Kraaijveld, M. A., Duin, R. P., *et al.* (1992). Feed forward neural networks with random weights. In *International Conference on Pattern Recognition*, pages 1–1. IEEE Computer Society Press.
- Schultz, T., Wand, M., Hueber, T., Krusienski, D. J., Herff, C., and Brumberg, J. S. (2017). Biosignal-based spoken communication: A survey. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **25**(12), 2257–2271.
- Sereshkeh, A. R., Trott, R., Bricout, A., and Chau, T. (2017). Online EEG classification of covert speech for brain-computer interfacing. *International Journal of Neural Systems*, **27**(08), 1750033.
- Shergill, S. S., Brammer, M. J., Fukuda, R., Williams, S. C., Murray, R. M., and McGuire, P. K. (2003). Engagement of brain areas implicated in processing inner speech in people with auditory hallucinations. *The British Journal of Psychiatry*, **182**(6), 525–531.
- Shi, L.-C. and Lu, B.-L. (2013). EEG-based vigilance estimation using extreme learning machines. *Neurocomputing*, **102**, 135–143.
- Shih, J. J., Krusienski, D. J., and Wolpaw, J. R. (2012). Brain-computer interfaces in medicine. In *Mayo Clinic Proceedings*, volume 87, pages 268–279. Elsevier.

- Sitaram, R., Zhang, H., Guan, C., Thulasidas, M., Hoshi, Y., Ishikawa, A., Shimizu, K., and Birbaumer, N. (2007). Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface. *NeuroImage*, **34**(4), 1416–1427.
- Song, Y. and Zhang, J. (2013). Automatic recognition of epileptic EEG patterns via extreme learning machine and multiresolution feature extraction. *Expert Systems with Applications*, **40**(14), 5477–5489.
- Suppes, P., Lu, Z.-L., and Han, B. (1997). Brain wave recognition of words. *Proceedings of the National Academy of Sciences*, **94**(26), 14965–14969.
- Tavares, L. D., Saldanha, R. R., Vieira, D. A., and Lisboa, A. C. (2014). A comparative study of extreme learning machine pruning based on detection of linear independence. In *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*, pages 63–69. IEEE.
- Torres-García, A., Reyes-García, C., Villaseñor-Pineda, L., and Ramírez-Cortés, J. (2013). Análisis de senales electroencefalográficas para la clasificación de habla imaginada. *Revista Mexicana de Ingeniería Biomédica*, **34**(1), 23–39.
- Van Rossum, G. (2020). *The Python Library Reference, release 3.8.2*. Python Software Foundation.
- Van Rossum, G. and Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.
- Vansteensel, M., Kristo, G., Aarnoutse, E., and Ramsey, N. (2017). The brain-computer interface researcher’s questionnaire: from research to application. *Brain-Computer Interfaces*, **4**(4), 236–247.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E. W., Vand erPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and Contributors, S. . . (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, **17**, 261–272.
- Vlek, R. J., Steines, D., Szibbo, D., Kübler, A., Schneider, M.-J., Haselager, P., and Nijboer, F. (2012). Ethical issues in brain-computer interface research, development, and dissemination. *Journal of neurologic physical therapy*, **36**(2), 94–99.
- Voulodimos, A., Doulamis, N., Doulamis, A., and Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, **2018**.
- Wang, Y., Gao, S., and Gao, X. (2006). Common spatial pattern method for channel selection in motor imagery based brain-computer interface. In *2005 IEEE engineering in medicine and biology 27th annual conference*, pages 5392–5395. IEEE.
- Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., *et al.* (2019). Do no harm: a roadmap for responsible machine learning for health care. *Nature medicine*, pages 1–4.
- Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, **113**(6), 767–791.
- Wu, D., Xu, Y., and Lu, B. (2020). Transfer learning for eeg-based brain-computer interfaces: A review of progresses since 2016. *arXiv preprint arXiv:2004.06286*.
- Yuan, Q., Zhou, W., Li, S., and Cai, D. (2011). Epileptic EEG classification based on extreme learning machine and nonlinear features. *Epilepsy Research*, **96**(1-2), 29–38.

- Zhang, R., Xu, P., Guo, L., Zhang, Y., Li, P., and Yao, D. (2013). Z-score linear discriminant analysis for eeg based brain-computer interfaces. *PloS one*, **8**(9), e74433.
- Zhao, H., Guo, X., Wang, M., Li, T., Pang, C., and Georgakopoulos, D. (2018). Analyze EEG signals with extreme learning machine based on pmis feature selection. *International Journal of Machine Learning and Cybernetics*, **9**(2), 243–249.
- Zhao, S. and Rudzicz, F. (2015). Classifying phonological categories in imagined and articulated speech. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 992–996. IEEE.
- Zhukov, L., Weinstein, D., and Johnson, C. (2000). Independent component analysis for eeg source localization. *IEEE Engineering in Medicine and Biology Magazine*, **19**(3), 87–96.



# Anexos



# Contribuciones

## **Thinking out loud, an open-access EEG-based BCI dataset for inner speech recognition EEG-BCI**

En este trabajo se presenta la primera base de datos pública de habla interna. La publicación responde al segundo y tercer objetivo particular propuestos en esta Tesis. En el artículo, se describe detalladamente el procedimiento seguido para adquirir los datos y su posterior procesamiento y análisis. Se dejó público, junto con las grabaciones originales, una base de datos preprocesada y estructurada, lista para ser utilizada por la comunidad. Mi contribución en este trabajo fue planificar y diseñar el protocolo de estimulación, adquirir los datos de los 10 participantes, desarrollar íntegramente la librería de procesamiento, el curado y análisis de los datos, la escritura y posterior corrección del manuscrito.

## **Extreme Learning Machine design for dealing with unrepresentative features**

En este artículo se describe y demuestra el impacto de las características no representativas en el desempeño de clasificación de las Extreme Learning Machines. Además, se comprobó empíricamente el origen del fenómeno. Mi aporte en este trabajo fue la idea original, el diseño, programación y ejecución de cada experimento, en conjunto con el análisis de los resultados obtenidos y la escritura del manuscrito.

## **Podado pre-entrenamiento de máquinas de aprendizaje extremo**

En este artículo se detalla un nuevo método de podado pre-entrenamiento de Extreme Learning Machines. Se caracterizan e identifican exitosamente los distintos tipos de neuronas que perjudican la performance del clasificador. En este trabajo mi aporte fue la idea original, el diseño, programación y ejecución de cada experimento, en conjunto con el análisis de los resultados obtenidos y la escritura del manuscrito.

## **Preliminary feasibility analysis of inner speech as a control paradigm for brain-computer interfaces**

En este artículo se realiza la primera prueba de clasificación sobre los datos de habla interna obtenidos. La distinción entre los paradigmas apunta a que los paradigmas generan actividades neuronales distintas, y estas se pueden distinguir a partir de las señales de electroencefalografía. Mi aporte fue la idea original, junto con el diseño y ejecución de los experimentos realizados. Por otro lado, también se realizó la visualización y discusión de los resultados obtenidos y la escritura del manuscrito.

## **Transfer learning based on optimal transport for motor imagery brain-computer interfaces**

En este trabajo se presentó un algoritmo de transporte óptimo que evita el costoso re-entrenamiento necesario en las técnicas tradicionales. Mi aporte en este trabajo fue principalmente en la discusión del método, los experimentos y los resultados. Desde la parte algorítmica, mi aporte fue principalmente en la legibilidad y reproducibilidad de los códigos presentes en el repositorio asociado.



# Anexo I: Thinking out loud, an open-access EEG-based BCI dataset for inner speech recognition EEG-BCI

# Thinking out loud, an open-access EEG-based BCI dataset for inner speech recognition

Nicolás Nieto<sup>1,2,\*</sup>, Victoria Peterson<sup>2</sup>, Hugo Leonardo Rufiner<sup>1,3</sup>, Juan Esteban Kamienkowski<sup>4</sup>, and Ruben Spies<sup>2</sup>

<sup>1</sup>Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(i), FICH-UNL / CONICET, Santa Fe, Argentina

<sup>2</sup>Instituto de Matemática Aplicada del Litoral, IMAL-UNL / CONICET, Santa Fe, Argentina

<sup>3</sup>Laboratorio de Cibernética, Universidad Nacional de Entre Ríos, FI-UNER, Oro Verde, Argentina

<sup>4</sup>Laboratorio de Inteligencia Artificial Aplicada, Instituto de Ciencias de la Computación, Universidad de Buenos Aires - CONICET, Ciudad Autónoma de Buenos Aires, Argentina

\*Corresponding author: Nicolás Nieto (nnieto@sinc.unl.edu.ar)

## ABSTRACT

Surface electroencephalography is a standard and noninvasive way to measure electrical brain activity. Recent advances in artificial intelligence led to significant improvements in the automatic detection of brain patterns, allowing increasingly faster, more reliable and accessible Brain-Computer Interfaces. Different paradigms have been used to enable the human-machine interaction and the last few years have broad a mark increase in the interest for interpreting and characterizing the "inner voice" phenomenon. This paradigm, called inner speech, raises the possibility of executing an order just by thinking about it, allowing a "natural" way of controlling external devices. Unfortunately, the lack of publicly available electroencephalography datasets, restricts the development of new techniques for inner speech recognition. A ten-participant dataset acquired under this and two others related paradigms, recorded with an acquisition system of 136 channels, is presented. The main purpose of this work is to provide the scientific community with an open-access multiclass electroencephalography database of inner speech commands that could be used for better understanding of the related brain mechanisms.

## Background & Summary

Brain-Computer Interfaces (BCIs) are a promising technology for improving the quality of life of people who have lost the capability to either communicate or interact with their environment<sup>1</sup>. A BCI provides an alternative way of interaction to such individuals, by decoding the neural activity and transforming it into control commands for triggering wheelchairs, prosthesis, spellers or any other virtual interface device<sup>2,3</sup>. In BCI applications, neural activity is typically measured by electroencephalography (EEG), since it is a non-invasive technique, the measuring devices can be easily portable and the EEG signals have high time resolution<sup>1,2</sup>.

Different paradigms have been used in order to establish communication between a user and a device. Some of the most widely adopted paradigms are P300<sup>4</sup>, steady-state visual evoked potentials (SSVP)<sup>5</sup> and motor imagery<sup>6</sup>. Although the use of these paradigms have resulted in great advances in EEG-based BCI systems, for some applications, they are still unable to lead to efficient ways for controlling devices. This is so mainly because they turned out to be too slow or they required a large effort from the users, restricting the applicability of BCIs in real-life and long-term applications.

In this context, BCIs based on speech-related paradigms, silent, imagined or inner speech, seek to find a solution to the aforementioned limitations, as they provide a more natural way for controlling external devices. Speech production is one of the most complex brain processes performed by human beings since it requires the interaction of several cortical, basal and subcortical brain regions<sup>7,8</sup>. Most language models and theories agree on the fact that speech involves auditory processing, semantic and syntactic processing as well as motor planning and articulation processes<sup>8-11</sup>. Although major and clear theoretical differences exist between the three paradigms mentioned above, they are quite often inconsistently and misleadingly referred to in the literature. In this article, we present a description of the main characteristics of each one of those three paradigms.

- **Silent speech** refers to the articulation produced during normal speech, but with no sound emitted. It is usually measured using motion-capturing devices, imaging techniques or by measuring the activity of muscles, and not only from brain signals<sup>12-14</sup>. In addition, Cooney et al.<sup>15</sup> proposed a similar paradigm called "Intended speech", where participants not having the capability to emit sound, are asked to perform speech.

- **Imagined speech** is similar to silent speech but it is produced without any articulatory movements, just like in motor imagery of speaking. Here, the participant must feel as if he/she is producing speech<sup>13</sup>, mainly focusing on the different articulatory gestures. This paradigm was widely explored using EEG<sup>16–21</sup>, electrocorticography (ECoG) signals<sup>22–24</sup> and magnetoencephalography<sup>25</sup>. In DaSalla et al.<sup>16</sup>, the vowels /a/ and /u/ were carefully selected, as they have the most different articulation processes. The requested actions were: “imagine mouth opening and imagine vocalization” for the vowel /a/ and “imagine lip rounding and imagine vocalization” for the vowel /u/. In Pressel et al.<sup>17</sup> the requested action was: “imagine the pronunciation or pronounce the word given as cue.”. Finally, the classes (prompts) used in Zhao et al.<sup>18</sup> were selected to have different articulations, as the authors mention: “These prompts were chosen to maintain a relatively even number of nasals, plosives, and vowels, as well as voiced and unvoiced phonemes”.
- **Inner speech** is defined as the internalized process in which the person thinks in pure meanings, generally associated with an auditory imagery of own inner “voice”. It is also referred to as verbal thinking, inner speaking, covert self-talk, internal monologue, internal speech and internal dialogue. Unlike imagined and silent speech, no phonological properties and turn-taking qualities of an external dialogue are retained<sup>13,26</sup>. Compared to brain signals in the motor system, language processing appears to be more complex and involves neural networks of distinct cortical areas engaged in phonological or semantic analysis, speech production and other processes<sup>23,27</sup>. Different studies investigate inner speech, as they benefit some aspects of the paradigm, using EEG<sup>28–30</sup>, ECoG<sup>23</sup>, functional Magnetic Resonance Imaging (fMRI) and positron emission tomography scan<sup>31–34</sup>. In D’Zmura et al.<sup>28</sup>, participants were instructed to “think the imagined speech without any vocal or subvocal activity”. Furthermore, as the cues were given with auditory stimulus, auditory processes of inner speech may be beneficial: “During this initial period, subjects heard through Stax electrostatic earphones either a spoken “ba” or a spoken “ku” followed by a train of clicks”. In Deng et al.<sup>29</sup> participants were instructed to “imagine speaking the cued syllable”, and the cues were also presented with auditory stimulus. Furthermore, no motor activity analysis was carried out in those papers. Finally, in Suppes et al.<sup>30</sup>, participants performed both auditory comprehension: were participants were instructed to “passively but carefully listen to the spoken words and try to comprehend them” and inner speech (called “Internal speech” in the paper), were participants were asked to “to silently “say” the word immediately after seeing it”.

Another paradigm related to the inner speech is the so-called “auditory comprehension”<sup>30,35,36</sup>. In this paradigm, instead of actively producing the speech imagination, the individual passively listens to someone else’s speech. It has already been explored using EEG<sup>30,37</sup>, ECoG<sup>38,39</sup> and fMRI<sup>40,41</sup>. Although this paradigm is not particularly useful for real BCI applications, it has contributed to the understanding of neural processes associated with speech-related paradigms.

While publicly available datasets for imagined speech<sup>17,18</sup> and for motor imagery<sup>42–46</sup> do exist, to the best of our knowledge there is not a single publicly available EEG dataset for the inner speech paradigm. In addition, the dataset presented by Zhao et al.<sup>18</sup> was recorded with a 64 channels acquisition system and all participants were native or advanced English speakers. On the other hand, in the dataset presented in Pressel et al.<sup>17</sup>, where all participants were native Spanish speakers, the acquisition system had only six channels, highly restricting spatial analysis. As mentioned before, both datasets were focused on the imagined speech paradigm rather than on inner speech. In order to improve the understanding of inner speech and its applications in real BCIs systems, we have built a multi speech-related BCI dataset consisting of EEG recordings from ten naive BCI users, performing four mental tasks in three different conditions: inner speech, pronounced speech and visualized condition. All paradigms and the requested actions are explained in detail in the BCI Interaction Conditions Section. This dataset will allow future users to explore whether inner speech activates similar mechanisms as pronounced speech or whether it is closer to visualizing a spatial location or movement. Each participant performed between 475 and 570 trials in a single day recording, obtaining a dataset with more than 9 hours of continuous EEG data recording, with over 5600 trials.

## Methods

### Participants

The experimental protocol was approved by the “Comité Asesor de Ética y Seguridad en el Trabajo Experimental” (CEySTE, CCT-CONICET, Santa Fe, Argentina, <https://santafe.conicet.gov.ar/ceyste/>). Ten healthy right-handed participants, four females and six males with **mean age = 34 (std = 10 years)**, with no hearing loss, no speech loss, and with no neurological, movement, or psychiatric disorders, joined the experiment and gave their written informed consent. All participants were native Spanish speakers. None of the individuals had any previous BCI experience, and participated in approximately two hours of recording. In this work, the participants are identified by aliases “sub-01” through “sub-10”. Detailed information about the participants can be found in Table 1.

## 87 Experimental Procedures

88 The study was conducted in an electrically shielded room. The participants were seated in a comfortable chair in front of a  
89 computer screen where the visual cues were presented. In order to familiarize the participant with the experimental procedure  
90 and the room environment, all steps of the experiment were explained, while the EEG headcap and the external electrodes were  
91 placed. The setup process took approximately 45 minutes. Figure 1 shows the main experiment setup.

92 The stimulation protocol was designed using Psychtoolbox-3<sup>47</sup> running in MatLab<sup>48</sup> and was executed on a computer,  
93 referred to as PC1 in Figure 1. The protocol displayed the visual cues to the participants in the Graphic User Interface (GUI).  
94 The screen's background was light-grey coloured in order to prevent dazzling and eye fatigue.

95 Each individual participated in one single recording day comprising three consecutive sessions, as shown in Figure 2. A  
96 self-selected break period between sessions, to prevent boredom and fatigue, was given (inter-session break). At the beginning  
97 of each session, a fifteen seconds baseline was recorded where the participant was instructed to relax and stay as still as  
98 possible. Within each session, five stimulation runs were presented. Those runs correspond to the different proposed conditions:  
99 pronounced speech, inner speech and visualized condition (see Section BCI Interaction Conditions). At the beginning of each  
100 run, the condition was announced in the computer screen for a period of 3 seconds. In all cases, the order of the runs was: one  
101 pronounced speech, two inner speech and two visualized conditions. A one minute break between runs was given (inter-run  
102 break).

103 The classes were specifically selected considering a natural BCI control application with the Spanish words: "arriba",  
104 "abajo", "derecha", "izquierda" (i.e. "up", "down", "right", "left", respectively). The trial's class (word) was randomly presented.  
105 Each participant had 200 trials in both the first and the second sessions. Nevertheless, depending on the willingness and  
106 tiredness, not all participants performed the same number of trials in the third session.

107 Figure 3 describes the composition of each trial, together with the relative and cumulative times. Each trial began at time  
108  $t = 0$  s with a concentration interval of 0.5 s. The participant had been informed that a new visual cue would soon be presented.  
109 A white circle appeared in the middle of the screen and the participant had been instructed to fix his/her gaze on it and not to  
110 blink, until it disappeared at the end of the trial. At time  $t = 0.5$  s the cue interval started. A white triangle pointing to either  
111 right, left, up or down was presented. The pointing direction of the cue corresponded to each class. After 0.5 s, i.e. at  $t = 1$  s,  
112 the triangle disappeared from the screen, moment at which the action interval started. The participants were instructed to start  
113 performing the indicated task right after the visual cues disappeared and the screen showed only the white circle. After 2.5 s of  
114 action interval, i.e. at  $t = 3.5$  s, the white circle turned blue, and the relax interval began. The participant had been previously  
115 instructed to stop performing the activity at this moment, but not to blink until the blue circle disappears. At  $t = 4.5$  s the blue  
116 circle vanished, meaning that the trial has ended. A rest interval, with a variable duration of between 1.5 s and 2 s, was given  
117 between trials.

118 To evaluate each participant's attention, a concentration control was randomly added to the inner speech and the visualized  
119 condition runs. The control task consisted of asking the participant, after some randomly selected trials, which was the direction  
120 of the last class shown. The participant had to select the direction using the keyboard arrows. No time limit was given to reply  
121 to these questions and the protocol continued after the participant pressed any of the four arrow keys. Visual feedback was  
122 provided indicating whether the question was correctly or incorrectly answered.

## 123 Data Acquisition

124 Electroencephalography (EEG), Electrooculography (EOG) and Electromyography (EMG) data were acquired using a BioSemi  
125 ActiveTwo high resolution biopotential measuring system (<https://www.biosemi.com/products.htm>). For data  
126 acquisition, 128 active EEG channels and 8 external active EOG/EMG channels with a 24 bits resolution and a sampling rate of  
127 1024 Hz were used. BioSemi also provides standard EEG head caps of different sizes with pre-fixed electrode positions. A cap  
128 of appropriate size was chosen for each participant by measuring the head circumference with a measuring tape. Each EEG  
129 electrode was placed in the corresponding marked position in the cap and the gap between the scalp and the electrodes was  
130 filled with a conductive SIGNAGEL® gel.

131 Signals in the EOG/EMG channels were recorded using a flat-type active electrode, filled with the same conductive gel and  
132 taped with a disposable adhesive disk. External electrodes are referred from "EXG1" to "EXG8". Electrodes EXG1 and EXG2  
133 were both used as a no-neural activity reference channels, and were placed in the left and right lobe of each ear, respectively.  
134 Electrodes EXG3 and EXG4 were located over the participant's left and right temples, respectively, and were intended to  
135 capture horizontal eye movement. Electrodes EXG5 and EXG6 aimed to capture vertical eye movement, mainly blinking  
136 movements. Those electrodes were placed above and below the right eye, respectively. Finally, electrodes EXG7 and EXG8  
137 were placed over the superior and inferior right orbicularis oris, respectively. Those electrodes were aimed to capture mouth  
138 movement in the pronounced speech and to provide a way for controlling that no movement was made during the inner speech  
139 and visualization condition runs.

140 The software used for recording was ActiView, developed also by BioSemi. It provides a way of checking the electrode



141 impedance and the general quality of the incoming data. It was carefully checked that the impedance of each electrode was less  
142 than 40  $\Omega$  before starting any recording session. Only a digital 208 Hz low-pass filter was used during acquisition time (no  
143 high-pass filter was used).

144 Once the recording of each session was finished, a .bdf file was created and stored in computer PC2. This file contains the  
145 continuous recording of the 128 EEG channels, the 8 external channels and the tagged events.

## 146 **BCI Interaction Conditions**

147 The design of the dataset was made having in mind as main objectives the decoding and understanding of the processes  
148 involved in the generation of inner speech, as well as the analysis of its potential use in BCI applications. As described in the  
149 “Background & Summary” Section, the generation of inner speech involves several complex neural networks interactions. With  
150 the objective of localizing the main activation sources and analyzing their connections, we asked the participants to perform the  
151 experiment under three different conditions: inner speech, pronounced speech and visualized condition.

### 152 ***Inner speech***

153 Inner speech is the main condition in the dataset and it is aimed to detect the brain’s electrical activity related to a participant’s  
154 thought about a particular word. In the inner speech runs, each participant was indicated to imagine his/her own voice as if  
155 he/she was giving a direct order to the computer, repeating the corresponding word until the white circle turned blue. Each  
156 participant was explicitly asked not to focus on the articulation gestures. Also, each participant was instructed to stay as still as  
157 possible and not to move the mouth nor the tongue. For the sake of natural imagination, no rhythm cue was provided.

### 158 ***Pronounced speech***

159 Although motor activity is mainly related to the imagined speech paradigm, inner speech may also show activity in the motor  
160 regions. The pronounced speech condition was proposed with the purpose of finding motor regions involved in the pronunciation  
161 matching those activated during the inner speech condition. In the pronounced speech runs, each participant was indicated to  
162 repeatedly pronounce aloud the word corresponding to each visual cue, as if he/she was giving a direct order to the computer.  
163 Like in the inner speech runs, no rhythm cue was provided.

### 164 ***Visualized condition***

165 Since the selected words have a high visual and spatial component, and with the objective of finding any activity related to that  
166 being produced during inner speech, the visualized condition was proposed. It is timely to mention that the main neural centers  
167 related to spatial thinking are located in the occipital and parietal regions<sup>49</sup>. Also, it has been demonstrated that the spatial  
168 attention have a significant impact on the SSVP amplitude<sup>50</sup>. In the visualized condition runs, the participants were indicated to  
169 focus on mentally moving the circle appearing at the center of the screen in the direction indicated by the visual cue.

## 170 **Data Processing**

171 In order to recast the continuous raw data into a more compact dataset and to facilitate their use, a transformation procedure  
172 was proposed. Such processing was implemented in Python, mainly using the MNE library<sup>51</sup>, and the code along with the raw  
173 data are available, so any interested reader can easily change the processing setup as desired (see Code Availability Section).

### 174 ***Raw data loading***

175 A function that rapidly allows loading of the raw data corresponding to a particular participant and session, was developed. The  
176 raw data stored in the .bdf file contains records of the complete EEG and external electrodes signals as well as the tagged events.

### 177 ***Events checking and correction***

178 The first step of the signal processing procedure was checking for correct tagging of events in the signals. Missing tags  
179 were detected and a correction method was proposed. The method detects and completes the sequences of events. After the  
180 correction, no tags were missing and all the events matched those sent from PC1.

### 181 ***Re-reference***

182 As BioSemi is a "reference free" acquisition system, the Common-Mode (CM) voltage is recorded in all channels, thus a  
183 re-reference step is needed. This procedure was made by means of the specific MNE re-reference function, using channels  
184 EXG1 and EXG2. The aforementioned function creates a virtual channel, averaging both EXG1 and EXG2, and then subtracting  
185 the virtual channel from each one of the rest of the acquired channels. This step eliminates the CM voltage, and helps to reduce  
186 both line noise (50 Hz) and body potential drifts.

### 187 ***Digital filtering***

188 The data were filtered with a zero-phase bandpass finite impulse response filter using the corresponding MNE function. The  
189 lower and upper bounds were set to 0.5 and 100 Hz, respectively. This broad band filter aims to keep the data as raw as possible,  
190 allowing future users the possibility of filtering the data in their desired bands. A Notch filter in 50Hz was also applied.

### 191 **Epoching and decimation**

192 The data were decimated four times, obtaining a final sampling rate of 254 Hz. Then, the continuous recorded data were  
193 epoched, keeping only the 4.5s length signals corresponding to the time window between the beginning of the concentration  
194 interval and the end of the relaxation interval. The matrices of dimension [channels × samples] corresponding to each trial,  
195 were stacked in a final tensor of size [Trials × channels × samples].

### 196 **Independent Components Analysis**

197 Independent Components Analysis (ICA) is a standard and widely used blind source separation method for removing artifacts  
198 from EEG signals<sup>52–54</sup>. For our dataset, ICA processing was performed only on the EEG channels, using the MNE implementa-  
199 tion of the infomax ICA<sup>55</sup>. No Principal Component Analysis (PCA) was applied and 128 sources were captured. Correlation  
200 with the EXG channels was used to determine the sources related to blink, gaze and mouth movement, which were neglected in  
201 the process of reconstructing the EEG signals, for obtaining the final dataset.

### 202 **Electromyography control**

203 The EMG control aims to determine whether a participant moved his/her mouth during the inner speech or visualized condition  
204 runs. The simplest approach to find EMG activity is the single threshold method<sup>56</sup>. The baseline period was used as a basal  
205 activity. The signals coming from the EXG7 and EXG8 channels were rectified and bandpass filtered between 1 and 20 Hz<sup>57–59</sup>.  
206 The power in a sliding window of 0.5 s length with a time step of 0.05 s was calculated as implemented in Peterson et al<sup>60</sup>. The  
207 power values were obtained by the following equation,

$$Pwr = \frac{1}{S - s + 1} \sum_{n=s}^S x[n]^2, \quad (1)$$

208 where  $x[\cdot]$  denotes the signal being considered, and  $s, S$  are the initial and final samples of the window, respectively. For every  
209 window, a scalar value was obtained after computing the power. Those scalar values were then stacked in a vector called  $V_{pb}$ .  
210 Finally the mean and standard deviations of those values were calculated and used to construct a decision threshold defined as  
211 follows:

$$th = mean(V_{pb}) + \gamma * std(V_{pb}). \quad (2)$$

212 In Equation 2,  $\gamma$  is an appropriately chosen parameter. According to Micera et al.<sup>61</sup>  $\gamma = 3$  is a reasonable choice. The same  
213 procedure was repeated for both channels and the mean power in the action interval of every trial was calculated. Then, if  
214 one of those values, for either the EXG7 or EXG8 channels was above the threshold, the corresponding trial was tagged as  
215 “contaminated”.

216 A total of 115 trials were tagged as contaminated, which represents a 2.5% of the inspected trials. The number of tagged  
217 trials is shown in Table 2. The tagged trials and their mean power corresponding to EXG7 and EXG8 were also stored in  
218 a report file. In order to reproduce the decision threshold, the mean and standard deviation power for the baseline for the  
219 corresponding session were also stored in the same report file.

220 The developed script performing the described control is publicly available and interested readers can use it to conduct  
221 different analyses.

### 222 **Ad-hoc tags correction**

223 After session 1, participant sub-03 claimed that, due to a misinterpretation, he performed only one inner speech run and  
224 three visualized condition runs. The condition tag was appropriately corrected. In addition, in session 3, the same participant  
225 performed three inner speech runs and only one visualized condition, aiming to balance the uneven number of trials by  
226 condition.

### 227 **Data Records**

228 All data files, including the raw recordings, can be accessed at the OpenNeuro repository<sup>62</sup>. All files are contained in a main  
229 folder called “Inner Speech Dataset”, structured as depicted in Figure 4, organized and named using the EEG data extension  
230 of BIDS recommendations<sup>63,64</sup>. The final dataset folder is composed of ten subfolders containing the session raw data, each  
231 one corresponding to a different participant. There is an additional folder, containing five files obtained after the proposed  
232 processing: EEG data, Baseline data, External electrodes data, Events data and a Report file. We now proceed to describe the  
233 contents of each one of these five files along with the raw data.

### 234 **Raw data**

235 The raw data file contains the continuous recording of the entire session for all 136 channels. The mean duration of the  
236 recordings is 1554 seconds. The .bdf file contains all the EEG/EXG data and the tagged events with further information about  
237 the recording sampling rate, the names of the channels and the recording filters, among other information. The raw events are  
238 obtained from the raw data file and contain the tags sent by PC1, synchronized with the recorded signals. Each event code, its  
239 ID and description are depicted in Table 3. A spurious event, of unknown origin, with ID 65536 appeared at the beginning of  
240 the recording and also it randomly appeared within some sessions. This event has no correlation with any sent tag and it was  
241 removed in the “Events Check” step of the processing. The raw events are stored in a three column matrix, where the first  
242 column contains the time stamp information, the second has the trigger information, and the third column contains the event ID.

### 243 **Electroencephalography data**

244 Each EEG data file, stored in .fif format, contains the acquired data for each participant and session, after processing as  
245 described above. Each one of these files contains an MNE Epoched object, with the EEG data information of all trials in the  
246 corresponding session. The dimension of the corresponding tensor data is [Trials  $\times$  128  $\times$  1154]. The number of trials changed  
247 among participants in each session, as explained in the “Data Aquisition” Section. The number of channels used for recording  
248 was 128 while the number of samples was 1154, each one of them corresponding to 4.5 s of signal acquisition with a final  
249 sampling rate of 256 Hz. A total of 1128 pronounced speech trials, 2236 inner speech trials and 2276 visualization condition  
250 trials, were acquired, distributed as shown in Table 4. Also, the detailed number of trials acquired in each block, divided by  
251 runs and participants, is included in Tables SI-1, SI-2 and SI-3 of the Supplementary Information Section.

### 252 **External electrodes data**

253 Each one of the EXG data files contains the data acquired by the external electrodes after the described processing was applied,  
254 with the exception of the ICA processing. They were saved in .fif format. The corresponding data tensor has dimension [Trials  
255  $\times$  8  $\times$  1154]. Here, the number of EXG trials equals the number of EEG data trials, 8 corresponds to the number of external  
256 electrodes used, while 1154 corresponds to the number samples of 4.5 s of signal recording at a final sampling rate of 256 Hz.

### 257 **Events data**

258 Each event data file (in .dat format) contains a four column matrix where each row corresponds to one trial. The first two  
259 columns were obtained from the raw events, by deleting the trigger column (second column of the raw events) and renumbering  
260 the classes 31, 32, 33, 34 as 0, 1, 2, 3, respectively. Finally, the last two columns correspond to condition and session number,  
261 respectively. Thus, the resulting final structure of the events data file is as depicted in Table 5.

### 262 **Baseline data**

263 Each baseline data file (in .fif format) contains a data tensor of dimension [1  $\times$  136  $\times$  3841]. Here, 1 corresponds to the  
264 only recorded baseline in each session, 136 corresponds to the total number of EEG + EXG channels (128+8), while 3841  
265 corresponds to the numbers of seconds of signal recording (15) times the final sampling rate (256 Hz). Through a visual  
266 inspection it was observed that the recorded baselines of participant sub-03 in session 3 and participant sub-08 in session 2,  
267 were highly contaminated.

### 268 **Report**

269 The report file (in .pkl format) contains general information about the participant and the particular results of the session  
270 processing. Its structure is depicted in Table 6.

## 271 **Technical Validation**

### 272 **Attentional Monitoring**

273 The evaluation of the participant’s attention was performed on the inner speech and the visualized condition runs. It was aimed  
274 to monitor their concentration on the requested activity. The results of the evaluation showed that participants correctly followed  
275 the task, as they performed very few mistakes (Table 7; mean = 0.5, std 0.62). participants sub-01 and sub-10 claimed that  
276 they had accidentally pressed the keyboard while answering the first two questions in session 1. Also, after the first session,  
277 participant sub-01 indicated that he/she felt that the questions were too many, reason for which, for the subsequent participants,  
278 the number of questions was reduced, in order to prevent participants from getting tired.

### 279 **Event Related Potentials**

280 It is well known that [Event Related Potentials \(ERPs\)](#) are manifestations of typical brain activity produced in response to  
281 certain stimuli. As different visual cues were presented during our stimulation protocol, we expected to find brain activity  
282 modulated by those cues. Moreover, we expected this activity to have no correlation with the condition nor with the class

283 and to be found across all participants. In order to show the existence of ERPs, an average over all participants, for each one  
284 of the channels at each instant of time, was computed using all the available trials ( $N_{ave} = 5640$ ), for each one of the 128  
285 channels. The complete time window average, with marks for each described event is shown in Figure 5. Between  $t = 0.1$  s  
286 and  $t = 0.2$  s a positive-negative-positive wave appears, as it is clearly shown in Figure [5-a]. A similar behavior is observed  
287 between  $t = 0.6$  s and  $t = 0.7$  s, but now with a more pronounced potential deflection, reflecting the fact that the white triangle  
288 (visual cue) appeared at  $t = 0.5$  s (see Figure [5-b]). At time  $t = 1$  s, the triangle disappeared and only the white fixation circle  
289 remained. As shown in Figure [5-c], a pronounced negative potential started just a few milliseconds before the triangle vanished.  
290 It is reasonable to believe that this ERP is related to the motor and cognitive preparation of the participants for executing the  
291 requested action. The signal appears to be mostly stable for the rest of the action interval. As seen in Figure [5-d], a positive  
292 peak appears between  $t = 3.8$  s and  $t = 3.9$  s, in response to the white circle turning blue, instant at which the relax interval  
293 begins.

294 For a more detailed analysis, instead of taking the time average of all trials (all together) we calculated the time average for  
295 all the trials belonging to each one of the three conditions, separately. All of the previously described ERPs remained present  
296 in the individual plots of Figure 6. The pronounced speech signals seem to have more noise in the action interval than both  
297 the inner speech and the visualized condition signals. Also, pronounce speech signals present two moments of more intense  
298 activity, approximately at 1.8 and 2.6 s. This is most likely due to the fact that, despite the signal pre-processing, some EMG  
299 noise remains.

### 300 **Time-Frequency Representation**

301 With the objective of finding and analyzing further differences and similarities between the three conditions, a Time-Frequency  
302 Representation (TFR) was obtained by means of a wavelet transform<sup>65</sup>. This representation is performed via a Morlet Wavelet  
303 transformation which returns a complex number. The MNE implementation returns the signal power and the inter trial  
304 coherence as the real and imaginary part of the TFR, respectively. This kind of analysis allows to detect changes in the signal  
305 characteristics across frequency and time. The corresponding implementation is available in the file "TFR\_representations.py",  
306 at our GitHub repository (see Code Availability Section).

### 307 **Inter Trial Coherence**

308 By means of the TFR, the Inter Trial Coherence (ITC) was calculated for all 5640 trials (all together). A stronger coherence  
309 was found within the concentration, cue and relax intervals, mainly at lower frequencies (see Figure 7). Also, the beginning of  
310 the action interval presents a strong coherence. This could be a result of the modulated activity generated by the disappearance  
311 of the cue.

312 Now, as in the "Event Related Potentials" Section, instead of taking the ITC of all trials (all together) we calculated the ITC  
313 for all the trials belonging to each one of the three conditions, separately. Of the three conditions, pronounced speech appears  
314 to have a slightly more intense global coherence at around 1.8 s and 2.6 s, mainly at lower frequencies (0.5 to 3 Hz). A careful  
315 observation of the time average results of the Pronounced speech condition presented in Figure 6, allows us to conjecture that  
316 such an increase in activity is the result of a quite natural pace in the articulation of generated sounds, and the aforementioned  
317 moments represent the times at which the participants, on average, match their pronunciation. Inner speech and visualized  
318 condition show consistently lower coherence during the action interval (see Figure 7-a and Figure 7-b).

### 319 **Averaged Power Spectral Density**

320 Using all available trials for each condition, the Averaged Power Spectral Density (APSD) between 0.5 and 100 Hz was  
321 computed. This APSD is defined as the average between all PSDs of the 128 channels. Figure 8 shows all APSD plots, in  
322 which shaded areas correspond to  $\pm 1$  std of all channels. As shown in the Inter Trial Coherence Section, all trials have a strong  
323 coherence up to  $t = 1.5$  s. Therefore, comparisons were made only in the action interval between 1.5 and 3.5 s. As it can  
324 be seen, the plots in Figure 8 show a peak in the alpha band [8 - 12 Hz] for all conditions, as it was to be expected, with a  
325 second peak in the beta band [12 - 30 Hz]. Also, from the stacked plot it is possible to see that at high frequencies (beta-gamma),  
326 pronounced speech shows higher power than both the inner speech and the visualized condition. That higher activity is most  
327 likely related to the brain motor activity and muscular artifacts. Finally, a narrow depression at 50 Hz appears, corresponding  
328 to the Notch filter applied during data processing.

### 329 **Spatial Distribution**

330 In order to detect regions where neural activity between conditions is markedly different, the differences in the power in the  
331 main frequency bands between each pair of conditions were computed. The aim of the subtractions is to minimize the effects of  
332 protocol activity, which are common to every paradigm, boosting the activity produced exclusively by the requested mental  
333 tasks. Like in the Averaged Power Spectral Density section, the time window used was 1.5 - 3.5 s. The Power Spectral Density  
334 (PSD) was added to the analysis to further explore regions of interest. Shaded areas on the PSD graphics in Figure 9 corresponds  
335 to  $\pm 1$  std of the different channels used. No shaded area is shown when only one channel was used to compute the PSD.

336 The top row of Figure 9 shows a comparison between inner and pronounced speech. The plots show the difference between  
337 the computed activity of inner speech and pronounced speech trials. In the alpha band, a major activity in the inner speech trials  
338 can be clearly seen in the central occipital/parietal region. The PSD shown in panel "a" was calculated using channels A4,  
339 A5, A19, A20 and A32 (BioSemi nomenclature for a head cap with 128 channels - [https://www.biosemi.com/pics/  
340 cap\\_128\\_layout\\_medium.jpg](https://www.biosemi.com/pics/cap_128_layout_medium.jpg)) and shows a difference of approximately 1 dB at 11 Hz. On the other hand, in the beta  
341 band, the spatial distribution of the power differences shows an increased temporal activity for the pronounced condition,  
342 consistent with muscular activity artifacts. Here, the PSD was calculated using channels B16, B22, B24 and B29 for panel "b"  
343 and channels D10, D19, D21 and D26 for panel "c". Pronounced speech shows higher power in the whole beta band with a  
344 more prominent difference in the central left area.

345 The middle row of Figure 9 shows a comparison of the visualized condition and pronounced speech. Here, the plots show  
346 the difference between the computed activity for trials corresponding to the first paradigm and those corresponding to the  
347 second paradigm. In the alpha band, the visualized condition trials present a larger difference in the central parietal regions and  
348 a more subtle difference in the lateral occipital regions. The PSD was calculated using channels A17, A20, A21, A22 and A30,  
349 panel "d". Here again, a difference of about 1 dB at 11 Hz can be observed. In the beta band, an intense activity in the central  
350 laterals regions appears for the pronounced condition. For this band, the PSD shown in panels "e" and "f" were calculated using  
351 the same channels as those for "b" and "c", respectively. As seen, power for pronounced speech trials is higher than for the  
352 visualized condition ones in the whole beta band, mainly in the left central region. This result is consistent with the fact that the  
353 occipital region is related to the visual activity while the central lateral region is related to the motor activity.

354 Finally, a comparison of the inner speech with the visualized condition is shown in the bottom row of Figure 9, where the  
355 difference between the computed activity of inner speech and visualized condition trials is plotted. Visualized condition trials  
356 exhibit a stronger activity in the laterals occipital regions in both the alpha and beta bands. This was to be expected since the  
357 visualized condition, containing a stronger visual component, generates marked occipital activity. Interesting, inner speech  
358 trials show a broad although subtle higher power in the alpha band in a more parietal region. For the alpha band, the PSDs  
359 showed in panel "g" and "h" were computed using channels A10 and B7 for the left and right plots respectively. In both plots,  
360 the peak corresponding to the inner speech condition is markedly higher than the one corresponding to the visualized condition.  
361 For the beta band, the PSD showed in panel "i" and "j" was calculated using channels A13 and A26 for the left and right PSD  
362 plots, respectively. As it can be observed, the power for the visualized condition trials in the whole beta band is higher than the  
363 inner speech power. It is timely to mention that no significant activity was presented in the central regions for neither of both  
364 conditions.

### 365 **Limitations and final remarks**

366 In this work we aim to provide a novel EEG dataset, acquired in three different speech related conditions, accounting for 5640  
367 total trials and more than 9 hours of continuous recording. To the best of our knowledge, there is no other dataset recorded with  
368 native Spanish speakers and acquired with a high dense (128+8) acquisition system.

369 In regard to the limitations of this dataset, it is important to mention that even though the participants were specifically  
370 instructed to imagine their own voice and explicitly requested not to focus on the muscular activity, as in any other endogenous  
371 BCI paradigm, there is no guarantee at all that the mental activity executed by a participant was in fact the correct one.  
372 Furthermore, as all the participants were naive BCI users, it may not had been possible for them to clearly differentiate between  
373 different components of the speech. In the same direction it is worth mentioning that, even though all participants received  
374 the same instructions, their interpretation and posterior executed mental activity may vary across individuals. It is also  
375 important to remark that, until today, there is not enough evidence to support that imagined speech and inner speech generate a  
376 distinguishable brain processes. And even if they in fact do, it is not clear that distinguishable features could be captured with  
377 current Electroencephalography technologies. Nonetheless, inner speech is clearly a more natural way of controlling a BCI,  
378 since the participant does not need to focus on the articulatory gestures.

379 Although the pronounced speech trials were most likely contaminated with muscular activity, providing the raw EEG  
380 data along with the EMG recordings will allow to develop and test EMG denoising algorithms. As mentioned before, no  
381 trials were discarded from the dataset in the "Electromyography control" Section nor in the "Attentional Monitoring" Section,  
382 so letting future users decide whether or not to include them in their analyses. Not providing rhythmic cues to benefit the  
383 participant's natural pace, could also produce a variation in the the timing of the pronunciation and imagination across trials.  
384 This phenomenon is likely to happen, since it is not natural to recite or think about one word at the time. Nevertheless, this is an  
385 issue that will also appear in any real BCI application. Much further analysis needs to be performed in order to determine if the  
386 participants changed their pace over the trials, and if they did, what the impact of this phenomenon in the decoding performance  
387 could be. It is important to remark that the references to the related papers in the field in regard to their executed paradigms,  
388 constitutes just a first classification attempt and further analysis needs to be done. This is so mainly because the description of  
389 the requested actions is not always clearly detailed and the difference between imagined and inner speech are not always taken  
390 into account in those articles.



391 Finally, we strongly believe that providing not just the raw data, but also the processed data, along with all the implementation  
392 codes will be of great benefit to the scientific community.

## 393 Usage Notes

394 The processing script was developed in Python 3.7<sup>66</sup>, using the MNE-python package v0.21.0<sup>51</sup>, NumPy v1.19.2<sup>67</sup>, Scipy  
395 v.1.5.2<sup>68</sup>, Pandas v1.1.2<sup>69</sup> and Pickle v4.0<sup>70</sup>. The main script, “InnerSpeech\_processing.py”, contains all the described  
396 processing steps and it can be modified to obtain different processing results, as desired. In order to facilitate data loading and  
397 processing, six more scripts defining functions are also provided.

398 The stimulation protocols were developed using Psychtoolbox-3<sup>47</sup> in MatLab R2017b<sup>48</sup>. The auxiliary functions, including  
399 the parallel port communication needed to send the tags from PC1 to BioSemi Active 2, were also developed in MatLab. The  
400 execution of the main script, called “Stimulation\_protocol.m”, shows the visual cue in the screen to the participant, and sends,  
401 via parallel port, the event being shown. The parallel port communication was implemented in the function “send\_value\_pp.m”.  
402 The main parameter that has to be controlled in the parallel communication is the delay needed after sending each value. This  
403 delay allows the port to send and receive the send value. Although we used a delay of 0.01 s, it can be changed as desired for  
404 other implementations.

## 405 Code Availability

406 In line with reproducible research philosophy, all codes used in this paper are publicly available and can be accessed at  
407 [https://github.com/N-Nieto/Inner\\_Speech\\_Dataset](https://github.com/N-Nieto/Inner_Speech_Dataset). The stimulation protocol and the auxiliary MatLab  
408 functions are also available. The code was run in PC1, and shows the stimulation protocol to the participants while sending the  
409 event information to PC2, via parallel port. The processing Python scripts are also available. The repository contains all the  
410 auxiliary functions to facilitate the load, use and processing of the data, as described above. By changing a few parameters in  
411 the main processing script, a completely different process can be obtained, allowing any interested user to easily build his/her  
412 own processing code. Additionally, all scripts for generating the Time-Frequency Representations and the plots here presented,  
413 are also available.

## 414 References

- 415 1. Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G. & Vaughan, T. M. Brain-computer interfaces for  
416 communication and control. *Clin. Neurophysiol.* **113**, 767–791 (2002).
- 417 2. Nicolas-Alonso, L. F. & Gomez-Gil, J. Brain computer interfaces, a review. *Sensors* **12**, 1211–1279 (2012).
- 418 3. Holz, E. M., Botrel, L., Kaufmann, T. & Kübler, A. Long-term independent brain-computer interface home use improves  
419 quality of life of a patient in the locked-in state: a case study. *Arch. Phys. Medicine Rehabil.* **96**, S16–S26 (2015).
- 420 4. McCane, L. M. *et al.* P300-based brain-computer interface (BCI) event-related potentials (ERPs): People with amyotrophic  
421 lateral sclerosis (ALS) vs. age-matched controls. *Clin. Neurophysiol.* **126**, 2124–2131 (2015).
- 422 5. Allison, B. Z. *et al.* Towards an independent brain–computer interface using steady state visual evoked potentials. *Clin.*  
423 *Neurophysiol.* **119**, 399–408 (2008).
- 424 6. Ahn, M. & Jun, S. C. Performance variation in motor imagery brain–computer interface: a brief review. *J. Neurosci.*  
425 *Methods* **243**, 103–110 (2015).
- 426 7. Blank, S. C., Scott, S. K., Murphy, K., Warburton, E. & Wise, R. J. Speech production: Wernicke, broca and beyond. *Brain*  
427 **125**, 1829–1838 (2002).
- 428 8. Lieberman, P. The evolution of human speech: Its anatomical and neural bases. *Curr. anthropology* **48**, 39–66 (2007).
- 429 9. Timmers, I., van den Hurk, J., Di Salle, F., Rubio-Gozalbo, M. E. & Jansma, B. M. Language production and working  
430 memory in classic galactosemia from a cognitive neuroscience perspective: future research directions. *J. inherited*  
431 *metabolic disease* **34**, 367–376 (2011).
- 432 10. Hickok, G. Computational neuroanatomy of speech production. *Nat. reviews neuroscience* **13**, 135–145 (2012).
- 433 11. Timmers, I., Jansma, B. M. & Rubio-Gozalbo, M. E. From mind to mouth: event related potentials of sentence production  
434 in classic galactosemia. *PLoS One* **7**, e52826 (2012).
- 435 12. Denby, B. *et al.* Silent speech interfaces. *Speech Commun.* **52**, 270–287 (2010).
- 436 13. Schultz, T. *et al.* Biosignal-based spoken communication: A survey. *IEEE/ACM Transactions on Audio, Speech, Lang.*  
437 *Process.* **25**, 2257–2271 (2017).

- 438 **14.** Gonzalez-Lopez, J. A., Gomez-Alanis, A., Martín-Doñas, J. M., Pérez-Córdoba, J. L. & Gomez, A. M. Silent speech  
439 interfaces for speech restoration: A review. *IEEE Access* (2020).
- 440 **15.** Cooney, C., Folli, R. & Coyle, D. Neurolinguistics research advancing development of a direct-speech brain-computer  
441 interface. *IScience* **8**, 103–125 (2018).
- 442 **16.** DaSalla, C. S., Kambara, H., Sato, M. & Koike, Y. Single-trial classification of vowel speech imagery using common  
443 spatial patterns. *Neural Networks* **22**, 1334–1339 (2009).
- 444 **17.** Pressel-Coreto, G., Gareis, I. E. & Rufiner, H. L. Open access database of EEG signals recorded during imagined speech.  
445 In *12th International Symposium on Medical Information Processing and Analysis (SIPAIM)*, [https://doi.org/10.1117/12.](https://doi.org/10.1117/12.2255697)  
446 [2255697](https://doi.org/10.1117/12.2255697) (2016).
- 447 **18.** Zhao, S. & Rudzicz, F. Classifying phonological categories in imagined and articulated speech. In *2015 IEEE International*  
448 *Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 992–996, <https://doi.org/10.1109/ICASSP.2015.7178118>  
449 (IEEE, 2015).
- 450 **19.** Brigham, K. & Kumar, B. V. Imagined speech classification with EEG signals for silent communication: a preliminary  
451 investigation into synthetic telepathy. In *2010 4th International Conference on Bioinformatics and Biomedical Engineering*,  
452 1–4 (IEEE, 2010).
- 453 **20.** Sereshkeh, A. R., Trott, R., Bricout, A. & Chau, T. Online EEG classification of covert speech for brain-computer  
454 interfacing. *Int. J. Neural Syst.* **27**, 1750033 (2017).
- 455 **21.** Cooney, C., Korik, A., Raffaella, F. & Coyle, D. Classification of imagined spoken word-pairs using convolutional neural  
456 networks. In *The 8th Graz BCI Conference, 2019*, 338–343 (2019).
- 457 **22.** Leuthardt, E. C., Schalk, G., Wolpaw, J. R., Ojemann, J. G. & Moran, D. W. A brain–computer interface using  
458 electrocorticographic signals in humans. *J. Neural Eng.* **1**, 63 (2004).
- 459 **23.** Pei, X., Barbour, D. L., Leuthardt, E. C. & Schalk, G. Decoding vowels and consonants in spoken and imagined words  
460 using electrocorticographic signals in humans. *J. Neural Eng.* **8**, 046028 (2011).
- 461 **24.** Guenther, F. H. *et al.* A wireless brain-machine interface for real-time speech synthesis. *PLoS ONE* **4** (2009).
- 462 **25.** Dash, D., Ferrari, P. & Wang, J. Decoding imagined and spoken phrases from non-invasive neural (meg) signals. *Front.*  
463 *neuroscience* **14**, 290 (2020).
- 464 **26.** Alderson-Day, B. & Fernyhough, C. Inner speech: development, cognitive functions, phenomenology, and neurobiology.  
465 *Psychol. Bull.* **141**, 931 (2015).
- 466 **27.** Indefrey, P. & Levelt, W. J. The spatial and temporal signatures of word production components. *Cognition* **92**, 101–144  
467 (2004).
- 468 **28.** D’Zmura, M., Deng, S., Lappas, T., Thorpe, S. & Srinivasan, R. Toward EEG sensing of imagined speech. In *International*  
469 *Conference on Human-Computer Interaction*, 40–48 (Springer, 2009).
- 470 **29.** Deng, S., Srinivasan, R., Lappas, T. & D’Zmura, M. EEG classification of imagined syllable rhythm using Hilbert spectrum  
471 methods. *J. Neural Eng.* **7**, 046006 (2010).
- 472 **30.** Suppes, P., Lu, Z.-L. & Han, B. Brain wave recognition of words. *Proc. Natl. Acad. Sci.* **94**, 14965–14969 (1997).
- 473 **31.** Fiez, J. A. & Petersen, S. E. Neuroimaging studies of word reading. *Proc. Natl. Acad. Sci.* **95**, 914–921 (1998).
- 474 **32.** Price, C. J. The anatomy of language: contributions from functional neuroimaging. *The J. Anat.* **197**, 335–359 (2000).
- 475 **33.** Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402 (2007).
- 476 **34.** McGuire, P. *et al.* Functional anatomy of inner speech and auditory verbal imagery. *Psychol. Medicine* **26**, 29–38 (1996).
- 477 **35.** Hubbard, T. L. Auditory imagery: empirical findings. *Psychol. Bull.* **136**, 302 (2010).
- 478 **36.** Martin, S. *et al.* Decoding spectrotemporal features of overt and covert speech from the human cortex. *Front. Neuroengi-*  
479 *neering* **7**, 14 (2014).
- 480 **37.** Suppes, P., Han, B. & Lu, Z.-L. Brain-wave recognition of sentences. *Proc. Natl. Acad. Sci.* **95**, 15861–15866 (1998).
- 481 **38.** Pasley, B. N. *et al.* Reconstructing speech from human auditory cortex. *PLoS Biol.* **10** (2012).
- 482 **39.** Cheung, C., Hamilton, L. S., Johnson, K. & Chang, E. F. The auditory representation of speech sounds in human motor  
483 cortex. *eLife* **5**, e12577 (2016).

- 484 **40.** Mitchell, T. M. *et al.* Predicting human brain activity associated with the meanings of nouns. *Science* **320**, 1191–1195  
485 (2008).
- 486 **41.** Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E. & Gallant, J. L. Natural speech reveals the semantic maps  
487 that tile human cerebral cortex. *Nature* **532**, 453–458 (2016).
- 488 **42.** Kaya, M., Binli, M. K., Ozbay, E., Yanar, H. & Mishchenko, Y. A large electroencephalographic motor imagery dataset for  
489 electroencephalographic brain computer interfaces. *Sci. Data* **5**, 180211 (2018).
- 490 **43.** Ofner, P. *et al.* Attempted arm and hand movements can be decoded from low-frequency EEG from persons with spinal  
491 cord injury. *Sci. Reports* **9**, 1–15 (2019).
- 492 **44.** Ofner, P., Schwarz, A., Pereira, J. & Müller-Putz, G. R. Upper limb movements can be decoded from the time-domain of  
493 low-frequency EEG. *PLoS ONE* **12**, e0182578 (2017).
- 494 **45.** Tangermann, M. *et al.* Review of the BCI competition IV. *Front. Neurosci.* **6**, 55 (2012).
- 495 **46.** Höhne, J. *et al.* Motor imagery for severely motor-impaired patients: Evidence for brain-computer interfacing as superior  
496 control solution. *PLoS ONE* **9**, 1–11, [10.1371/journal.pone.0104854](https://doi.org/10.1371/journal.pone.0104854) (2014).
- 497 **47.** Brainard, D. H. The psychophysics toolbox. *Spatial vision* **10**, 433–436 (1997).
- 498 **48.** MATLAB. *version 7.10.0 (R2010a)* (The MathWorks Inc., Natick, Massachusetts, 2010).
- 499 **49.** Kandel, E. R. *et al.* *Principles of neural science*, vol. 5 (McGraw-hill New York, 2000).
- 500 **50.** Morgan, S., Hansen, J. & Hillyard, S. Selective attention to stimulus location modulates the steady-state visual evoked  
501 potential. *Proc. Natl. Acad. Sci.* **93**, 4770–4774 (1996).
- 502 **51.** Gramfort, A. *et al.* MNE software for processing MEG and EEG data. *Neuroimage* **86**, 446–460 (2014).
- 503 **52.** Jung, T.-P. *et al.* Extended ICA removes artifacts from electroencephalographic recordings. *Adv. Neural Inf. Process. Syst.*  
504 894–900 (1998).
- 505 **53.** Vorobyov, S. & Cichocki, A. Blind noise reduction for multisensory signals using ICA and subspace filtering, with  
506 application to EEG analysis. *Biol. Cybern.* **86**, 293–303 (2002).
- 507 **54.** Makeig, S., Bell, A. J., Jung, T.-P. & Sejnowski, T. J. Independent component analysis of electroencephalographic data. In  
508 *Advances in Neural Information Processing Systems*, 145–151 (1996).
- 509 **55.** Bell, A. J. & Sejnowski, T. J. An information-maximization approach to blind separation and blind deconvolution. *Neural*  
510 *Comput.* **7**, 1129–1159 (1995).
- 511 **56.** Theetion, A. A randomisation method for discriminating between signal and noise in recordings of rhythmic electromyo-  
512 graphic activity. *J. Neurosci. Methods* **66**, 93–98 (1996).
- 513 **57.** Porcaro, C., Medaglia, M. T. & Krott, A. Removing speech artifacts from electroencephalographic recordings during overt  
514 picture naming. *NeuroImage* **105**, 171–180 (2015).
- 515 **58.** Laganaro, M. & Perret, C. Comparing electrophysiological correlates of word production in immediate and delayed naming  
516 through the analysis of word age of acquisition effects. *Brain Topogr.* **24**, 19–29 (2011).
- 517 **59.** Ganushchak, L. Y. & Schiller, N. O. Motivation and semantic context affect brain error-monitoring activity: an event-related  
518 brain potentials study. *Neuroimage* **39**, 395–405 (2008).
- 519 **60.** Peterson, V., Galván, C., Hernández, H. & Spies, R. A feasibility study of a complete low-cost consumer-grade brain-  
520 computer interface system. *Heliyon* **6**, e03425 (2020).
- 521 **61.** Micera, S., Vannozzi, G., Sabatini, A. & Dario, P. Improving detection of muscle activation intervals. *IEEE Eng. Medicine*  
522 *Biol. Mag.* **20**, 38–46 (2001).
- 523 **62.** Nieto, N., Peterson, V., Rufiner, H., Kamienskowski, J. & Spies, R. "Inner Speech", [https://doi.org/10.18112/openneuro.](https://doi.org/10.18112/openneuro.ds003626.v2.0.0)  
524 [ds003626.v2.0.0](https://doi.org/10.18112/openneuro.ds003626.v2.0.0) (2021).
- 525 **63.** Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging  
526 experiments. *Sci. Data* **3**, 1–9 (2016).
- 527 **64.** Pernet, C. R. *et al.* EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Sci. Data* **6**,  
528 1–5 (2019).
- 529 **65.** Mallat, S. *A wavelet tour of signal processing* (Elsevier, 1999).
- 530 **66.** Van Rossum, G. & Drake, F. L. *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).



**Table 1.** Participants information.

Participant	Self-declared gender	Age	Dominance	Native language
sub-01	Female	56	Right	Spanish
sub-02	Male	50	Right	Spanish
sub-03	Male	34	Right	Spanish
sub-04	Female	24	Right	Spanish
sub-05	Female	30	Right	Spanish
sub-06	Male	29	Right	Spanish
sub-07	Male	26	Right	Spanish
sub-08	Female	28	Right	Spanish
sub-09	Male	35	Right	Spanish
sub-10	Male	31	Right	Spanish

- 531 **67.** Oliphant, T. E. *A guide to NumPy*, vol. 1 (Trelgol Publishing USA, 2006).
- 532 **68.** Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **17**, 261–272,  
533 <https://doi.org/10.1038/s41592-019-0686-2> (2020).
- 534 **69.** McKinney, W. *et al.* Data structures for statistical computing in Python. In *Proceedings of the 9th Python in Science*  
535 *Conference*, vol. 445, 51–56 (Austin, TX, 2010).
- 536 **70.** Van Rossum, G. *The Python Library Reference, release 3.8.2* (Python Software Foundation, 2020).

## 537 **Acknowledgements**

538 This research was funded in part by Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, Argentina,  
539 through PIP 2014-2016 No. 11220130100216-CO, the Agencia Nacional de Promoción Científica y Tecnológica through  
540 PICT-2017-4596 and by Universidad Nacional del Litoral, UNL, through CAI+D-UNL 2016 PIC No.50420150100036LI and  
541 CAI+D 2020, number 50620190100069LI. We would like to thank the Laboratorio de Neurociencia, Universidad Torcuato Di  
542 Tella (Buenos Aires, Argentina) for giving us access to the facilities where the experiments were performed. Also, we will like  
543 to thank all the participants for their willingness and time, and María José Schmidt for produce and edit the images presented in  
544 this article.

## 545 **Author contributions statement**

546 NN acquired the data, developed all codes, ran all the experiments and wrote the manuscript. VP helped to acquire the data,  
547 provided technical feedback for designing the experiments, analyzed results and reviewed the manuscript. HR provided  
548 technical feedback for designing the experiments, analyzed results and reviewed the manuscript. JK acquired the data, provided  
549 technical feedback for designing the experiments, analyzed results and reviewed the manuscript. RS analyzed results, wrote  
550 and reviewed the manuscript.

## 551 **Competing interests**

552 The authors declare no competing interests.

## 553 **Figures & Tables**

**Table 2.** Number of tagged trials by participant and session.

Participant	Session 1	Session 2	Session 3	Total
1	6	6	1	13
2	38	0	4	42
3	0	1	0	1
4	0	0	1	1
5	0	0	1	1
6	11	0	11	22
7	0	0	0	0
8	0	0	0	0
9	8	4	15	27
10	7	0	1	8

**Table 3.** Raw data event tags number and meanings.

Event ID	Description
1	Start of protocol
12	End of protocol
13	Start of baseline
14	End of baseline
15	Start of run
16	End of run
17	Cognitive control - question posing
21	Start of pronounced speech run
22	Start of inner speech run
23	Start of Visualized condition run
31	“Arriba/Up” trial - start of cue interval
32	“Abajo/Down” trial - start of cue interval
33	“Derecha/Right” trial - start of cue interval
34	“Izquierda/Left” trial - start of cue interval
42	Start of concentration interval
44	Start of action interval
45	Start of relax interval
46	Start of rest interval
51	Start of inter runs rest interval
61	Answer to cognitive control: “Arriba/Up”
62	Answer to cognitive control: “Abajo/Down”
63	Answer to cognitive control: “Derecha/Right”
64	Answer to cognitive control: “Izquierda/Left”

**Table 4.** Final number of trials divided by participant, class and condition.

Participant	Pronounced Speech				Inner Speech				Visualized Condition			
	Up	Down	Right	Left	Up	Down	Right	Left	Up	Down	Right	Left
sub-01	25	25	25	25	50	50	50	50	50	50	50	50
sub-02	30	30	30	30	60	60	60	60	60	60	60	60
sub-03	25	25	25	25	45	45	45	45	55	55	55	55
sub-04	30	30	30	30	60	60	60	60	60	60	60	60
sub-05	30	30	30	30	60	60	60	60	60	60	60	60
sub-06	27	27	27	27	54	54	54	54	54	54	54	54
sub-07	30	30	30	30	60	60	60	60	60	60	60	60
sub-08	25	25	25	25	50	50	50	50	50	50	50	50
sub-09	30	30	30	30	60	60	60	60	60	60	60	60
sub-10	30	30	30	30	60	60	60	60	60	60	60	60
Sub Total	282	282	282	282	559	559	559	559	569	569	569	569
<b>Total</b>	<b>1128</b>				<b>2236</b>				<b>2276</b>			

**Table 5.** Events data format and tag meaning.

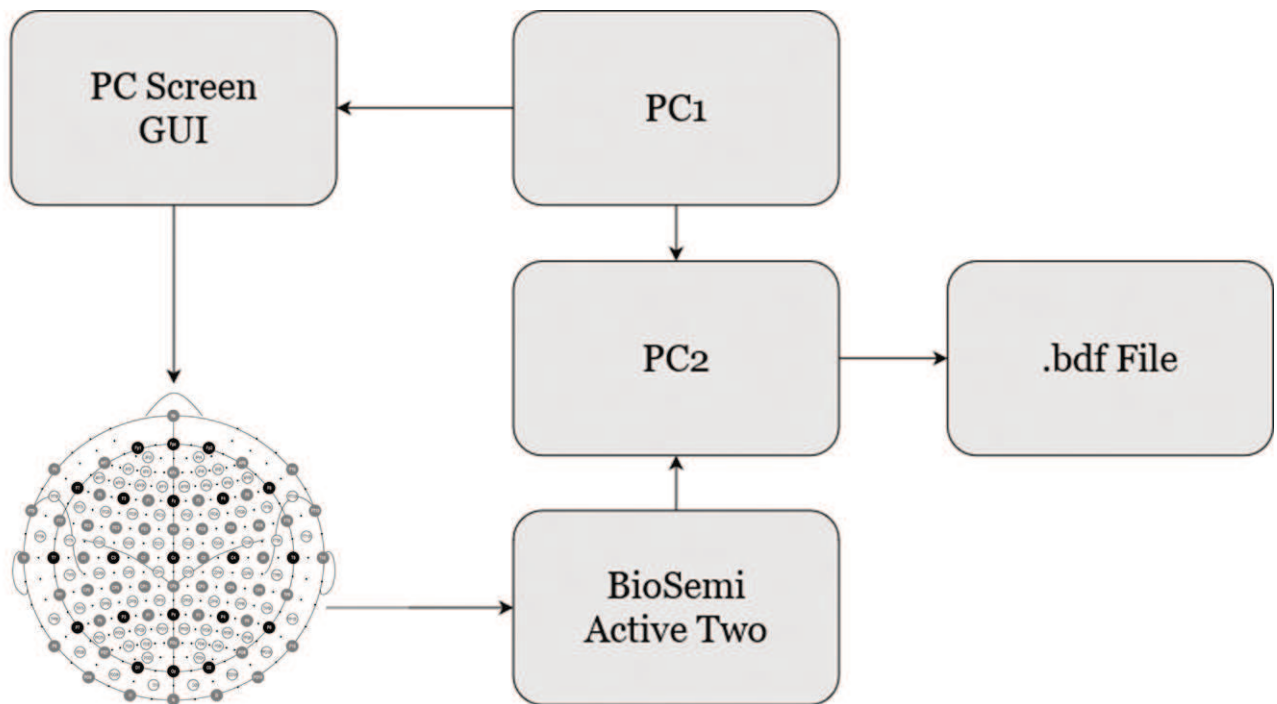
Sample	Trial's class	Trials' condition	Trials' session
Sample at which the event occurred (Numbered starting at n=0, corresponding to the beginning of the recording)	0 = "Arriba" (up) 1 = "Abajo" (down) 2 = "Derecha" (right) 3 = "Izquierda" (left)	0 = Pronounced speech 1 = Inner speech 2 = Visualized condition	1 = session 1 2 = session 2 3 = session 3

**Table 6.** Report file fields

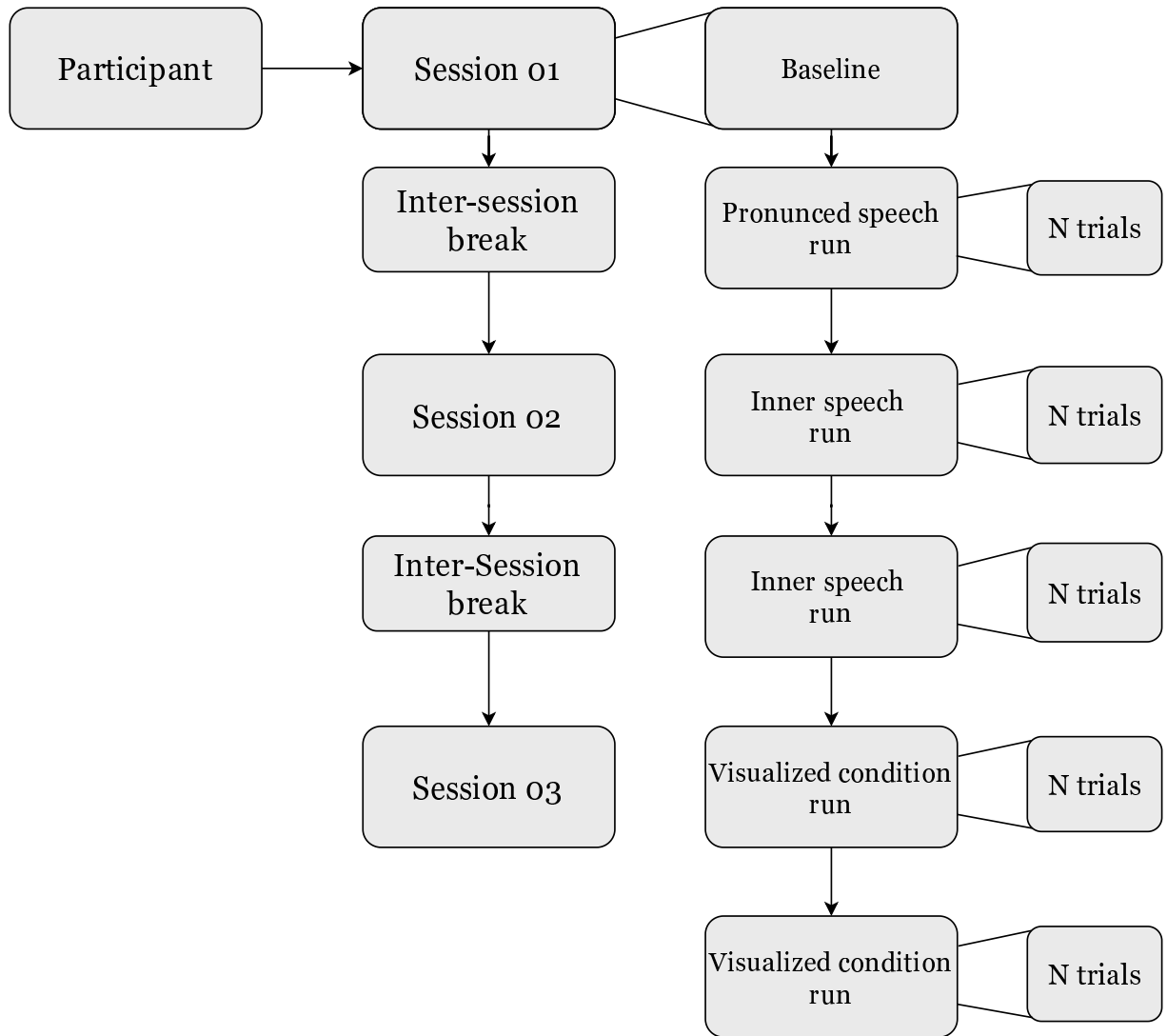
Listed field	Content
Age	Participant's age.
Self-declared gender	Participant's Self-declared gender: 'F' for female, 'M' for male.
Recording_time	Length of the complete session recording in seconds.
Ans_R	Number of times the participant correctly answered the cognitive control questions.
Ans_W	Number of times the participant incorrectly answered the cognitive control questions.
EMG_trials	Position of the contaminated trials.
Power_EXG7	Mean power for channel EXG7 of the contaminated trials. Array with the same dimension as EMG_trials.
Power_EXG8	Mean power for channel EXG8 of the contaminated trials. Array with the same dimension as EMG_trials.
Baseline_EXG7_mean	Mean power for channel EXG7 in the Baseline.
Baseline_EXG8_mean	Mean power for channel EXG8 in the Baseline.
Baseline_EXG7_std	Standard deviation of the power for channel EXG7 in the Baseline.
Baseline_EXG8_std	Standard deviation of the power for channel EXG8 in the Baseline.

**Table 7.** Result of attention monitoring. Note that the maximum number of incorrect answers is 2. The large variability in the number of questions in session 3 is due to the different number of trials for each one of the participants.

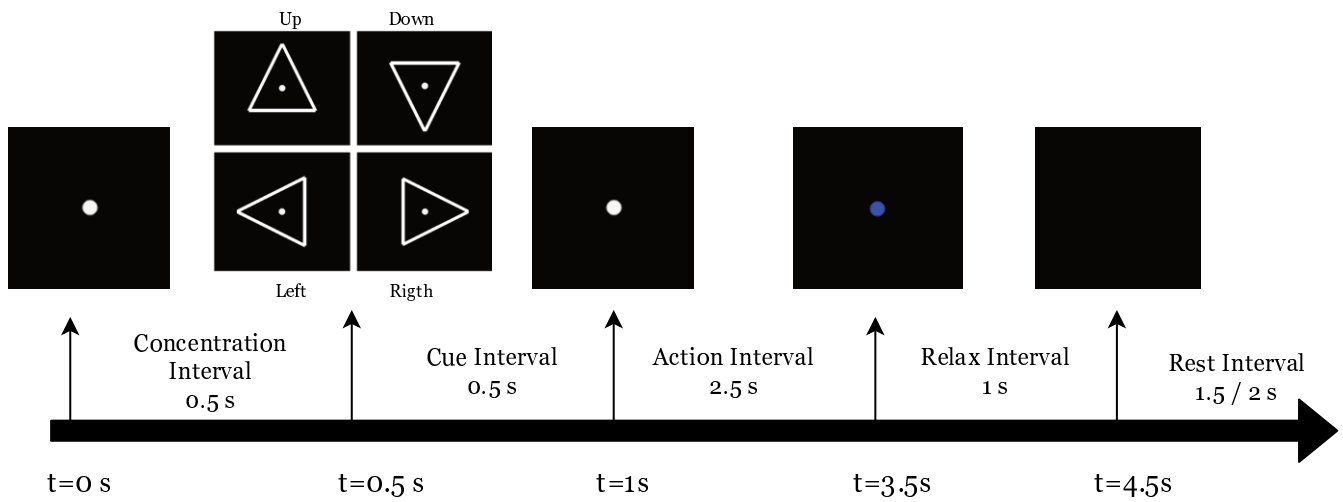
Participant	Session	Questions	Wrong
1	1	45	2
	2	12	2
	3	4	0
2	1	11	0
	2	16	1
	3	16	1
3	1	12	1
	2	10	1
	3	8	0
4	1	12	1
	2	14	0
	3	14	1
5	1	12	0
	2	10	0
	3	13	1
6	1	12	1
	2	12	0
	3	9	0
7	1	12	0
	2	11	0
	3	16	1
8	1	12	0
	2	11	0
	3	8	1
9	1	12	0
	2	10	0
	3	13	0
10	1	11	1
	2	11	0
	3	10	0



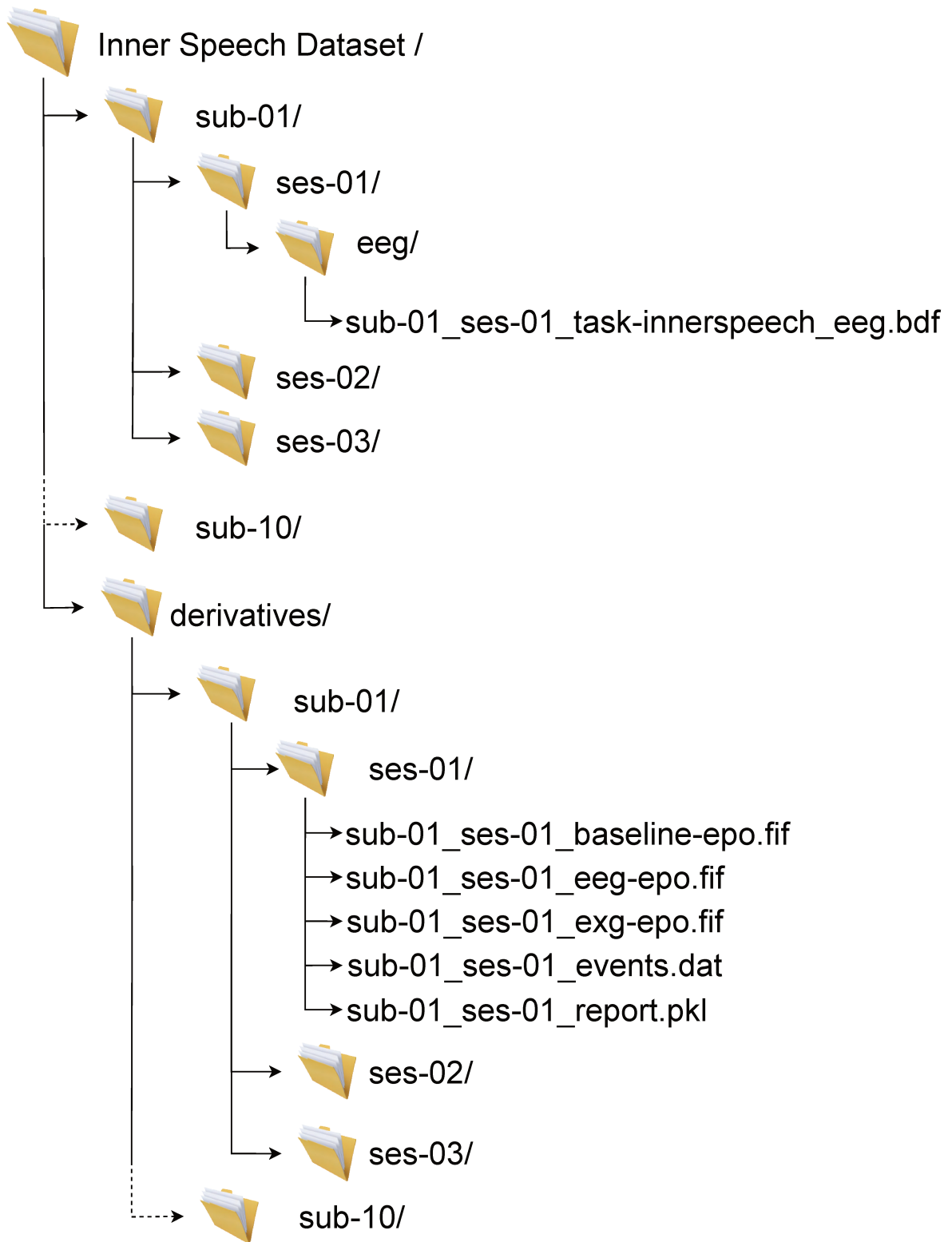
**Figure 1.** Experiment setup. Both computers, PC1 and PC2, were located outside the acquisition room. PC1 runs the stimulation protocol while communicating to PC2 every cue displayed. PC2 received the sampled EEG data from the acquisition system and tagged the events with the information received from PC1. At the end of the recording, a .bdf file was created and saved.



**Figure 2.** Organization of the recording day for each participant.

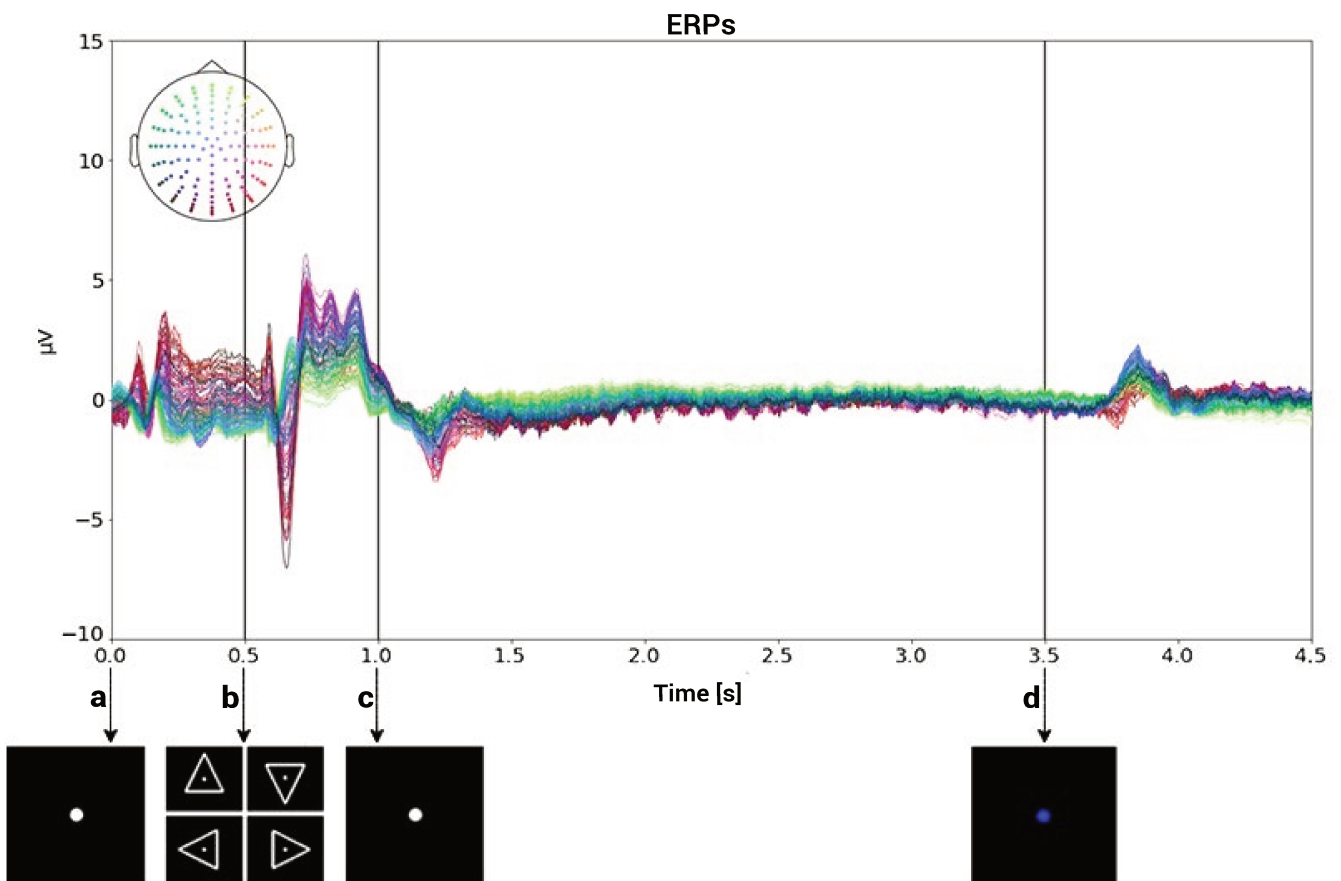


**Figure 3.** Trial workflow. The screen presented to the participant in each time interval was plotted on the top arrow of the figure. Relative and global time were plotted above and below the arrow, respectively.

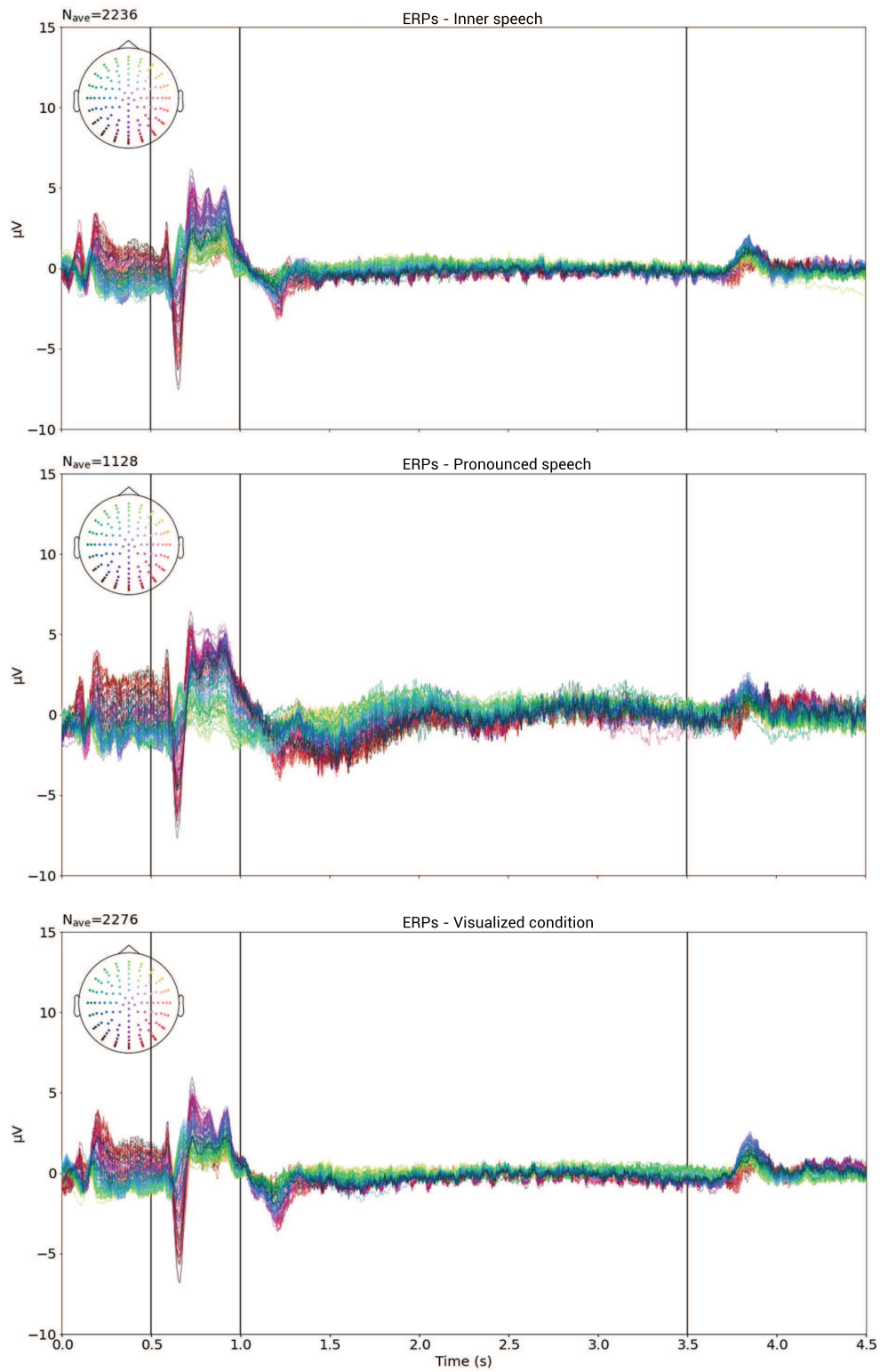


**Figure 4.** Final dataset structure, files, and naming.

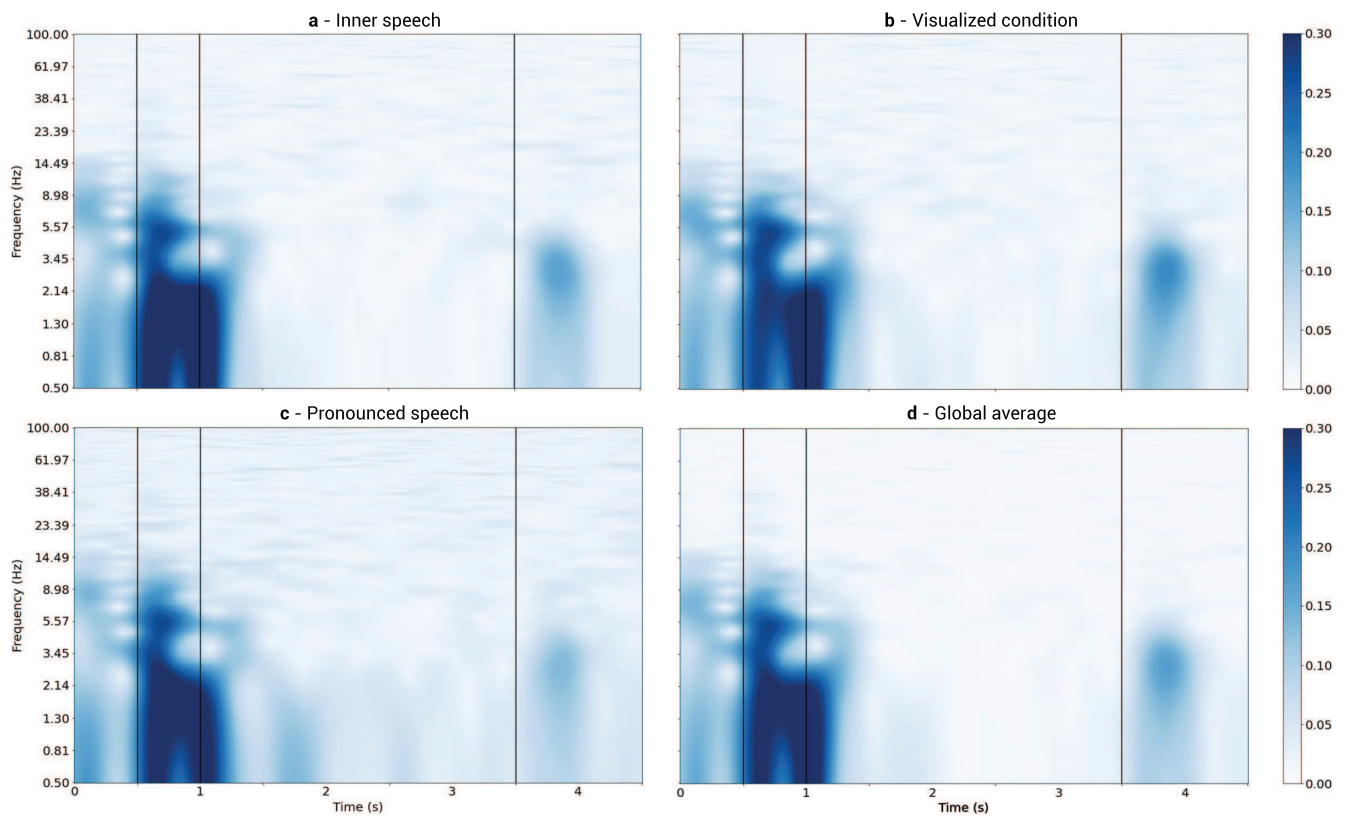




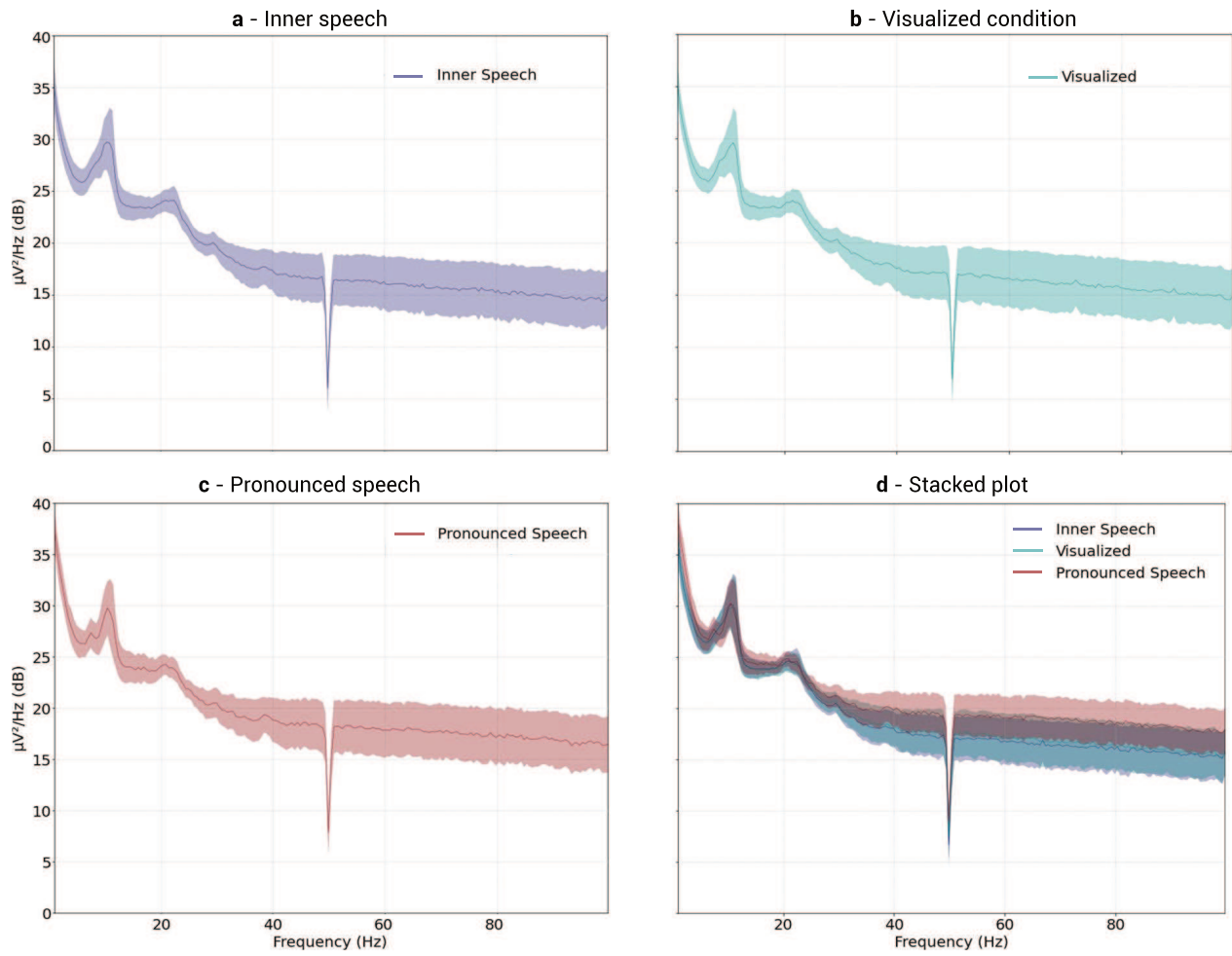
**Figure 5.** Global participant average trial and interval plots; all the channels were plotted with **a** color reference location. **a-b** Concentration interval. **b-c** Cue interval. **c-d** Action interval. **d-end** Relax interval.



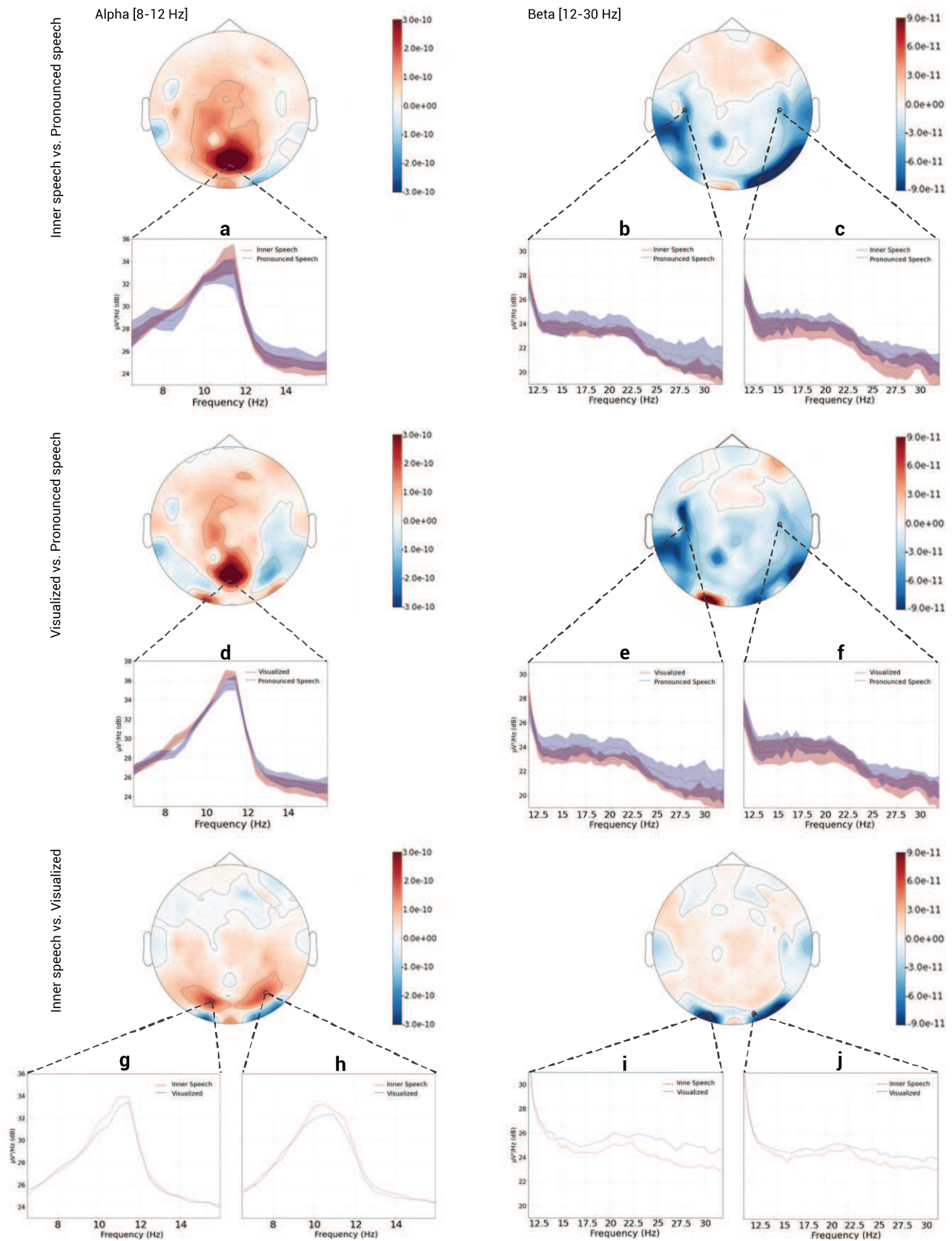
**Figure 6.** Global average trial for each class. Vertical black lines correspond to interval boundaries, as in Figure 5. Top row: Inner speech, Middle row: Pronounced speech. Bottom row: Visualized condition



**Figure 7.** Inter Trials Coherence. Vertical black lines correspond to interval boundaries, as in Figure 5. **a:** Inner speech trials. **b:** Pronounced speech trials. **c:** Visualized condition trials. **d:** Global Average.



**Figure 8.** Power spectral density for all conditions. **a** - Top Left: Inner speech. **b** - Top Right: Visualized condition. **c** - Bottom left: Pronounced speech. **d** - Bottom right: Staked plot



**Figure 9.** Power difference between conditions. Left column: alpha band comparisons. Right column: beta band comparison. Time window used in all comparisons to compute the PSD: 1.5 to 3.5 s. Channels used to compute the PSD: **a:** A4, A5, A19, A20 and A32. **b:** B16, B22, B24 and B29. **c:** D10, D19, D21 and D26. **d:** A17, A20, A21, A22 and A30. **e:** B16, B22, B24 and B29. **f:** D10, D19, D21 and D26. **g:** A10. **h:** B7. **i:** A13. **j:** A26.

**Supplementary Table 1.** Distribution of trials in session 1. \* Check the "Ad-Hoc Correction" Section for further details of the uneven number of trials in Participant sub-03.

	Run 1 - Pronounced Speech							
Participant	Up	Down	Right	Left				
sub-01	10	10	10	10				
sub-02	10	10	10	10				
sub-03	10	10	10	10				
sub-04	10	10	10	10				
sub-05	10	10	10	10				
sub-06	10	10	10	10				
sub-07	10	10	10	10				
sub-08	10	10	10	10				
sub-09	10	10	10	10				
sub-10	10	10	10	10				
	Run 2 - Inner Speech				Run 3 - Inner Speech			
Participant	Up	Down	Right	Left	Up	Down	Right	Left
sub-01	10	10	10	10	10	10	10	10
sub-02	10	10	10	10	10	10	10	10
sub-03	10	10	10	10	0*	0*	0*	0*
sub-04	10	10	10	10	10	10	10	10
sub-05	10	10	10	10	10	10	10	10
sub-06	10	10	10	10	10	10	10	10
sub-07	10	10	10	10	10	10	10	10
sub-08	10	10	10	10	10	10	10	10
sub-09	10	10	10	10	10	10	10	10
sub-10	10	10	10	10	10	10	10	10
	Run 4 - Visualized Condition				Run 5 - Visualized Condition			
Participant	Up	Down	Right	Left	Up	Down	Right	Left
sub-01	10	10	10	10	10	10	10	10
sub-02	10	10	10	10	10	10	10	10
sub-03	20*	20*	20*	20*	10	10	10	10
sub-04	10	10	10	10	10	10	10	10
sub-05	10	10	10	10	10	10	10	10
sub-06	10	10	10	10	10	10	10	10
sub-07	10	10	10	10	10	10	10	10
sub-08	10	10	10	10	10	10	10	10
sub-09	10	10	10	10	10	10	10	10
sub-10	10	10	10	10	10	10	10	10

**Supplementary Table 2.** Distribution of trials in session 2.

	Run 1 - Pronounced Speech							
Participant	Up	Down	Right	Left				
sub-01	10	10	10	10				
sub-02	10	10	10	10				
sub-03	10	10	10	10				
sub-04	10	10	10	10				
sub-05	10	10	10	10				
sub-06	10	10	10	10				
sub-07	10	10	10	10				
sub-08	10	10	10	10				
sub-09	10	10	10	10				
sub-10	10	10	10	10				
	Run 2 - Inner Speech				Run 3 - Inner Speech			
Participant	Up	Down	Right	Left	Up	Down	Right	Left
sub-01	10	10	10	10	10	10	10	10
sub-02	10	10	10	10	10	10	10	10
sub-03	10	10	10	10	10	10	10	10
sub-04	10	10	10	10	10	10	10	10
sub-05	10	10	10	10	10	10	10	10
sub-06	10	10	10	10	10	10	10	10
sub-07	10	10	10	10	10	10	10	10
sub-08	10	10	10	10	10	10	10	10
sub-09	10	10	10	10	10	10	10	10
sub-10	10	10	10	10	10	10	10	10
	Run 4 - Visualized Condition				Run 5 - Visualized Condition			
Participant	Up	Down	Right	Left	Up	Down	Right	Left
sub-01	10	10	10	10	10	10	10	10
sub-02	10	10	10	10	10	10	10	10
sub-03	10	10	10	10	10	10	10	10
sub-04	10	10	10	10	10	10	10	10
sub-05	10	10	10	10	10	10	10	10
sub-06	10	10	10	10	10	10	10	10
sub-07	10	10	10	10	10	10	10	10
sub-08	10	10	10	10	10	10	10	10
sub-09	10	10	10	10	10	10	10	10
sub-10	10	10	10	10	10	10	10	10

**Supplementary Table 3.** Distribution of trials in session 3. \* Check the "Ad-Hoc Correction" Section for further details of the uneven number of trials in Participant sub-03.

	Run 1 - Pronounced Speech							
Participant	Up	Down	Right	Left				
sub-01	5	5	5	5				
sub-02	10	10	10	10				
sub-03	5	5	5	5				
sub-04	10	10	10	10				
sub-05	10	10	10	10				
sub-06	7	7	7	7				
sub-07	10	10	10	10				
sub-08	5	5	5	5				
sub-09	10	10	10	10				
sub-10	10	10	10	10				
	Run 2 - Inner Speech				Run 3 - Inner Speech			
Participant	Up	Down	Right	Left	Up	Down	Right	Left
sub-01	5	5	5	5	5	5	5	5
sub-02	10	10	10	10	10	10	10	10
sub-03	5	5	5	5	10*	10*	10*	10*
sub-04	10	10	10	10	10	10	10	10
sub-05	10	10	10	10	10	10	10	10
sub-06	7	7	7	7	7	7	7	7
sub-07	10	10	10	10	10	10	10	10
sub-08	5	5	5	5	5	5	5	5
sub-09	10	10	10	10	10	10	10	10
sub-10	10	10	10	10	10	10	10	10
	Run 4 - Visualized Condition				Run 5 - Visualized Condition			
Participant	Up	Down	Right	Left	Up	Down	Right	Left
sub-01	5	5	5	5	5	5	5	5
sub-02	10	10	10	10	10	10	10	10
sub-03	0*	0*	0*	0*	5	5	5	5
sub-04	10	10	10	10	10	10	10	10
sub-05	10	10	10	10	10	10	10	10
sub-06	7	7	7	7	7	7	7	7
sub-07	10	10	10	10	10	10	10	10
sub-08	5	5	5	5	5	5	5	5
sub-09	10	10	10	10	10	10	10	10
sub-10	10	10	10	10	10	10	10	10



# Anexo II: Extreme Learning Machine design for dealing with unrepresentative features

# Extreme Learning Machine design for dealing with unrepresentative features

2 Nicolás Nieto<sup>a,b</sup>, Francisco J. Ibarrola<sup>a</sup>, Victoria Peterson<sup>b</sup>, Hugo L. Rufiner<sup>a</sup>, Ruben Spies<sup>b</sup>

4 <sup>a</sup>*Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(i), UNL-CONICET, FICH, Ciudad  
Universitaria, CC 217, Ruta Nac. 168, km 472.4, (3000) Santa Fe, Argentina.*

6 <sup>b</sup>*Instituto de Matemática Aplicada del Litoral, IMAL, UNL-CONICET, Centro Científico Tecnológico CONICET Santa Fe,  
Colectora Ruta Nac. 168, km 472, Paraje "El Pozo", (3000), Santa Fe, Argentina.*

---

*Email addresses:* [nnieto@sinc.unl.edu.ar](mailto:nnieto@sinc.unl.edu.ar) - Argentina (Nicolás Nieto), [fiбарrola@sinc.unl.edu.ar](mailto:fiбарrola@sinc.unl.edu.ar) - Argentina (Francisco J. Ibarrola), [vpeterson@santafe-conicet.gov.ar](mailto:vpeterson@santafe-conicet.gov.ar) - Argentina (Victoria Peterson), [lrufiner@sinc.unl.edu.ar](mailto:lrufiner@sinc.unl.edu.ar) - Argentina (Hugo L. Rufiner), [rspies@santafe-conicet.gov.ar](mailto:rspies@santafe-conicet.gov.ar) - Argentina (Ruben Spies)

## Abstract

8 Extreme Learning Machines (ELMs) have become a popular tool for the classification of electroencephalography (EEG) signals for Brain Computer Interfaces. This is so mainly due to their very high training speed and  
10 generalization capabilities. Another important advantage is that they have only one hyperparameter that must be calibrated: the number of hidden nodes. While most traditional approaches dictate that this parameter should be  
12 chosen smaller than the number of available training examples, in this article we argue that, in the case of problems in which the data contain unrepresentative features, such as in EEG classification problems, it is beneficial to choose  
14 a much larger number of hidden nodes. We characterize this phenomenon, explain why this happens and exhibit several concrete examples to illustrate how ELMs behave. Furthermore, as searching for the optimal number of  
16 hidden nodes could be time consuming in enlarged ELMs, we propose a new training scheme, including a novel pruning method. This scheme provides an efficient way of finding the optimal number of nodes, making ELMs more  
18 suitable for dealing with real time EEG classification problems. Experimental results using synthetic data and real EEG data show a major improvement in the training time with respect to most traditional and state of the art  
20 ELM approaches, without jeopardising classification performance and resulting in more compact networks.

*Keywords:* Brain Pattern Recognition, Brain Computer Interfaces, Pruning, Unrepresentative Features,  
22 Electroencephalography  
*2010 MSC:* 68T05

## 24 1. Introduction

Brain-Computer Interfaces (BCIs) have become attractive as they provide alternative ways of communication for  
26 people who have lost the capability to interact with their environment (Wolpaw et al., 2002). By means of a BCI, the neural activity of a person is decoded and transformed into commands, which are then used for controlling a device  
28 (Nicolas-Alonso & Gomez-Gil, 2012; Holz et al., 2015). In BCI applications, neural activity is classically measured by electroencephalography (EEG), since it is a non-invasive technique, measurement devices can be portable and  
30 the EEG signals have high time resolution (Wolpaw et al., 2002; Nicolas-Alonso & Gomez-Gil, 2012). However, those signals present a low spatial resolution, high redundancy and the information is usually encoded only in a  
32 small subset of the features. Once the EEG signal is obtained, machine learning techniques are typically used to classify them for their posterior use as inputs for control commands. This classification process presents several  
34 challenges, as it needs to yield good performance, its training time must be short and it must be applicable even when only a few training samples are available. In this context, Extreme Learning Machines (ELMs) had become a  
36 popular tool for EEG signal classification in BCIs applications (Liang et al., 2006; Duan et al., 2016; Zhang et al., 2018; Kong et al., 2018; Tan et al., 2016; Jin et al., 2020).

38 The use of random weights in neural networks was first proposed in (Schmidt et al., 1992). The idea was then reintroduced under the concept of Extreme Learning Machines by (Huang et al., 2004, 2006, 2011), and it has been

40 extensively used since then, not only for BCI applications, but also in the whole Neuroscience field (Murugavel  
& Ramakrishnan, 2016; Song & Zhang, 2013; Yuan et al., 2011; Zhao et al., 2018; Shi & Lu, 2013). Moreover,  
42 ELMs have been recently incorporated into Deep Learning frameworks for EEG classification, resulting in deep  
architectures with short training times (Ding et al., 2015, 2017). The main reasons for their wide use are that  
44 the training process is very fast and the networks yield good generalization capabilities. Another quite appealing  
aspect of ELMs is that, unlike most machine learning techniques, they have only one hyperparameter that must  
46 be tuned up: the number of hidden nodes. In (Huang & Babri, 1998) it has been proved that given a training  
dataset consisting of  $N$  samples, the model can learn them exactly, with probability 1 (w.r.t. the random parameter  
48 initialization), using  $N$  hidden nodes. Nevertheless, perfect classification over a training set often entails a loss of  
generalization capabilities. In fact, for the case of  $N$  training samples, according to (Huang & Babri, 1998),  $N$  is an  
50 upper bound for the number of nodes in the hidden layer. However, for the particular case of problems in which  
the data are contaminated by unrepresentative features, like EEG signals, choosing the number of hidden nodes  
52  $M \leq N$  turns out to be suboptimal, since  $M \gg N$  will almost surely result in better classification performances.  
In Sections 2 and 3, we use synthetic data to show this phenomenon and describe the reason for its occurrence.

54 Enlarged architecture ( $M \gg N$ ) are most often unfeasible in real time applications with traditional ELM training  
schemes. This is so because training time rapidly increases with the number of nodes, and hence hyperparameter  
56 search can become time consuming, thus antagonizing one of the most important ELMs advantages. To cope with  
the problem of finding the optimal number of hidden nodes, different pruning methods have been developed (Rong  
58 et al., 2008; Miche et al., 2010; Luo et al., 2013; Tavares et al., 2014; Alencar et al., 2016). In (Rong et al., 2008)  
the Information Gain and Chi-Square are computed and used to determine the relation between the hidden nodes  
60 and the target labels. In (Miche et al., 2010) the authors propose a method named Optimally Pruned ELM (OP-  
ELM). This algorithm ranks the nodes using multiresponse sparse regression (MRSR) (Similä & Tikka, 2005), and  
62 then prunes the hidden layer using the leave-one-out (LOO) validation error. In (Luo et al., 2013) the authors  
propose to find a sparse representation of the output weights, effectively pruning the hidden layer. However, this  
64 approach uses Bayesian inference instead of the generalized inverse to obtain the output weights, and therefore it  
can hardly be considered an ELM. In (Tavares et al., 2014) the authors propose to pre-prune linearly dependent  
66 nodes, disregarding any other information such as discriminative power. More recently, a pruning method based on  
genetic algorithms (GA) was introduced in (Alencar et al., 2016). Here, a fitness function simultaneously minimizes  
68 the LOO validation error and the number of hidden nodes. However, as this method has to train a different ELM  
for each individual in every generation, it ends up being computationally expensive.

70 All previously described pruning methods need to retrain the ELMs for searching their internal hyperparameters  
and/or have to retrain the whole network after the optimal number of hidden nodes is reached. Furthermore, the  
72 methods focus their attention on improving classification performance, with complete disregard for the computa-  
tional burden of tuning the hyperparameters, which can be very heavy in enlarged ELMs. In order to cope with this  
74 drawback, in this article we introduce a novel post-training pruning method. Our proposal changes the traditional  
way of searching for the number of hidden nodes in enlarged ELMs, eliminating the need for the computational

76 expensive retraining and without adding extra hyperparameters. The method is described in Section 4. In Section  
 5, we compare our scheme with the traditional ELM approach with two different EEG databases, and show its  
 78 potential in real-life problems. Finally, a detailed comparison between our proposed algorithm and the OP-ELM is  
 also presented.

## 80 2. Extreme Learning Machines

For simplicity, we shall consider an ELM within the context of a binary classification problem. We point out,  
 82 however, that all results presented in this work remain valid for multi-class problems.

Given an arbitrary vector  $x \in \mathbb{R}^D$ , an ELM classification output is given by

$$z = \beta^T g(Wx + b), \quad (1)$$

where  $W \in \mathbb{R}^{M \times D}$  is the matrix associated to the hidden layer,  $g : \mathbb{R} \rightarrow \mathbb{R}$  is an activation function,  $b \in \mathbb{R}^M$  is the  
 84 bias vector, and  $\beta \in \mathbb{R}^M$  is the weight vector connecting the hidden layer to the output. Here and in the sequel,  
 the action of  $g$  on a vector or on a matrix is meant to be its components-wise evaluation.

86 The training process of an ELM consists of two main steps. First, the entries of  $W$  and  $b$  are randomly generated  
 as independent realizations of an absolutely continuous random variable (usually with uniform distribution in  
 88  $[-1, 1]$ ). The second step consists of finding an appropriate vector  $\beta$ . This can be done as described below.

Let us consider a dataset consisting of  $N$  training samples  $x_n \in \mathbb{R}^D$ ,  $n = 1, \dots, N$ , stacked as the columns of a  
 90 matrix  $X \in \mathbb{R}^{D \times N}$ . Let  $y \in \{-1, 1\}^N$  be the desired output vector, where  $y_n$ , gives account for the class of  $x_n$ ,  $\forall n$ .

Let us define

$$H \doteq [g(WX + b \mathbf{1}_{(1,N)})]^T, \quad (2)$$

where  $\mathbf{1}_{(1,N)}$  is an  $N$ -dimensional row vector with all its elements equal to 1. Then, training the ELM weight vector  
 $\beta$  amounts to “solving” the linear system

$$H\beta = y. \quad (3)$$

Note that  $H \in \mathbb{R}^{N \times M}$ , and therefore if  $M < N$ , (*i.e.* if the number of nodes is less than the number of training  
 examples) the linear system (3) is under-determined. In this case, an appropriate way to deal with equation (3),  
 is to resort to least squares solutions. We then define the vector  $\hat{\beta}$  as the best approximate solution (*i.e.* the  
 minimal-norm least-squares solution) of (3) (Heinz et al., 1996). That is

$$\hat{\beta} \doteq H^\dagger y, \quad (4)$$

where  $H^\dagger$  is the Moore-Penrose generalized inverse of  $H$ .

92 When  $M < N$ , the ELM’s generalization capability is associated with the fact that least squares solutions assign  
 larger weights to the columns of  $H$  which are most relevant for classification purposes. However, when  $M = N$ ,

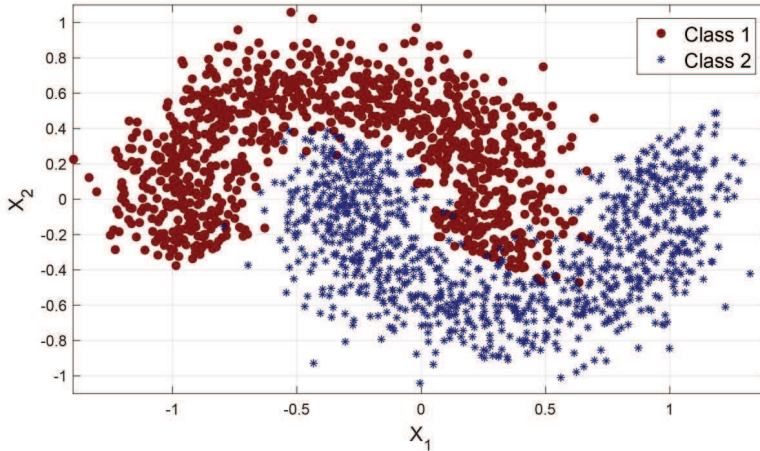


Figure 1: Synthetic data distribution, differentiated by class. 500 data points for each class are plotted

94 system (3) has a unique solution (with probability one), as shown in (Huang & Babri, 1998). Hence, in this case,  
the solution  $\hat{\beta}$  is forced to take all columns of  $H$  into account, even those which are irrelevant for classification.  
96 This constitutes a classic case of overfitting.

Although it is theoretically true that  $M \geq N$  implies that the matrix  $H$  has  $N$  independent column vectors  
98 (with probability one), this does not necessarily imply that the feature space be well represented. To illustrate  
this, let us consider the following example. Suppose that there are two columns of  $H$ , say  $h$  and  $h' \in \mathbb{R}^N$ , such  
100 that  $h_1 = h'_1 + \epsilon$  (for a small  $\epsilon \in \mathbb{R}$ ) and  $h_n = h'_n, \forall n \geq 2$  (here  $h_i, h'_i$  denotes the  $i^{\text{th}}$  components of the vectors  
 $h$  and  $h'$ , respectively). Although these two vectors are strictly different and linearly independent, for all practical  
102 purposes they clearly encode the same feature information. Formally, while  $M = N$  ensures that  $H$  be invertible, it  
can still present very small singular values (and therefore a high condition number), which is a reflection of a poor  
104 representation of the feature space. As shown in (Horn & Johnson, 1990), as the smaller non-zero singular values  
increase, in a context of normalized data, the feature space becomes better represented.

### 106 3. Changing the network size

In light of the above discussion, we argue that the proposed upper bound  $M \leq N$  is suboptimal in the case of  
108 problems where the data have unrepresentative features. Thus, in this kind of problems, such as classification of  
EEG signals,  $M$  should be chosen and calibrated differently. We perform an experiment using artificially generated  
110 data in order to characterize the type of problems where choosing  $M \gg N$  can be highly beneficial.

Let us consider a binary classification problem in which the data points corresponding to the two classes are  
112 distributed as depicted in Figure 1. It is clear that the displayed coordinates  $x_1, x_2$  are enough for classification.  
In an ideal situation, any additional coordinate added to the data points taking random values independently of  
114 the class, should not be taken into account by any classifier. In a real problem, the data points might be highly  
contaminated with this kind of unrepresentative “junk features” (JF) which are, *a-priori*, indistinguishable from  
116 the representative ones. Hence a question that immediately arises is: should the same kind of ELM architecture be

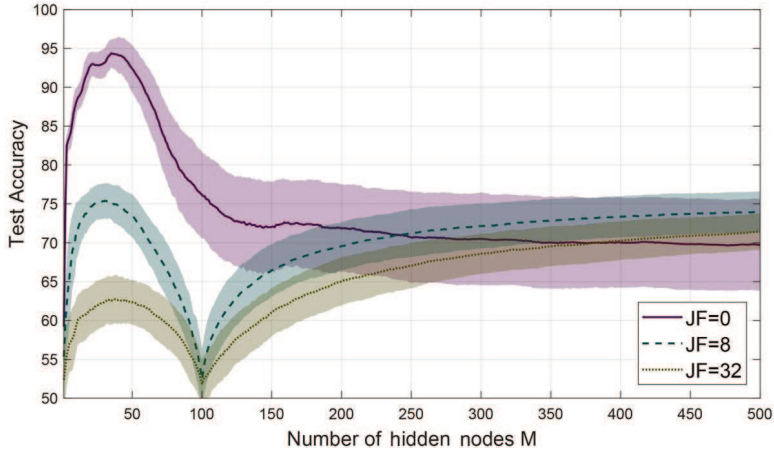


Figure 2: Mean test accuracy for 300 realizations (30 random initializations of parameters  $W$  and  $b$  and 10 cross validations of the data) in synthetic data adding Junk Features (JF) generated from a uniform distribution in the interval  $[0,1]$ . 100 data points are used for training. Shading illustrates  $\pm 1$  standard deviation.

used for this kind of problems?

118 Let us take a look at Figure 2, where for several choices of the number of neurons  $M$ , the average test accuracy  
 120 of an ELM is plotted. The purple solid line corresponds to the test accuracy obtained using the data as displayed  
 in Figure 1, while the other two dashed lines correspond to those obtained using the data points contaminated with  
 different numbers of junk features. Those features were generated as random realizations of a uniform distribution  
 122 in the interval  $(0, 1)$ . One can immediately see that  $M < N$  is an optimal choice for the two-coordinates case.  
 However, when the data contain unrepresentative features, the choice of  $M \gg N$  could be more convenient. As  
 124 the number of unrepresentative features increases, this convenience becomes more notorious. Different distributions  
 of the junk features have different impacts in the described phenomenon. This can be clearly observed in Figures  
 126 SI-1, SI-2 and SI-3 of the Supplementary Information Section, where analogous results are shown for different  
 distributions of the unrepresentative features.

128 One might wonder, if this is a scenario one might often expect in practical problems. As pointed out before,  
 when working with EEG signals, the extracted features will, in most cases, not be representative of the problem  
 130 at hand, resulting in many junk features. We show that this happens to be the case for an experiment using the  
 datasets introduced in (DaSalla et al., 2009) and (Ledesma-Ramirez et al., 2010), which contain EEG signals. Both  
 132 datasets are explained in detail in Section 4.1).

The experiments were performed using 70% of the data for training and 30% for testing, with 50 random  
 134 initializations of the parameters  $W$  and  $b$  and 20 cross validations. Figure 3 illustrates the train and test accuracy  
 of the ELM as a function of the parameter  $M$  for the dataset introduced in (DaSalla et al., 2009). As it can be seen,  
 136 at first the accuracy grows with  $M$ , until reaching a local maximum, after which it starts to decay up to the global  
 minimum, reached at  $M = N$ . However, the test accuracy starts to increase again as  $M$  further increases. Similar  
 138 results are shown in Figure SI-4 of the Supplementary Information Section, obtained for the dataset introduced  
 in (Ledesma-Ramirez et al., 2010). These results are consistent with the analyzed synthetic data and also with

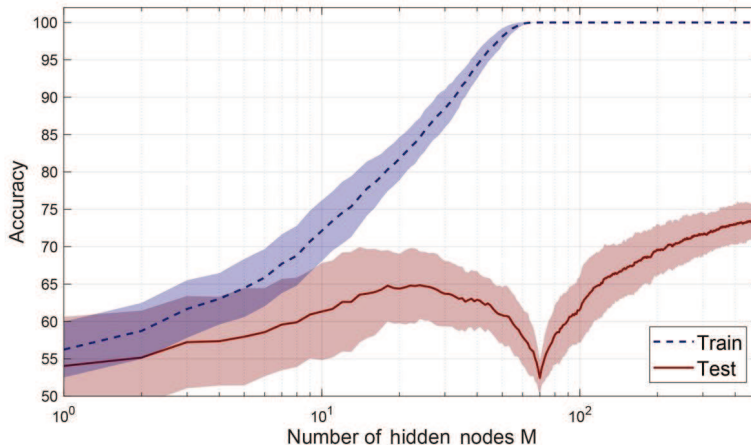


Figure 3: Mean accuracy for 1000 realizations (50 random initializations and 20 cross validations) obtained for training and testing data, as a function of the hidden layer size,  $M$ . 70 training examples for DaSalla dataset were used. Shading illustrates  $\pm 1$  standard deviation of the results.

140 the findings in (Belkin et al., 2019), where a similar behaviour is reported in the context of traditional multi-layer  
 perceptrons (MLP). Although this phenomenon is well described and known for MLP, to our knowledge, it has  
 142 never been previously studied in the context of ELMs.

The overfitting observed in Figures 2 and 3 when  $M \approx N$  can be explained by the fact that when training under  
 144 this condition, we are forcing the network to take into account all the nodes, even those that are irrelevant for  
 classification purposes. To corroborate that this is in fact the reason, we have performed an experiment consisting  
 146 of adding a disconnected “fake” neuron to the ELM (before training using the DaSalla dataset). That is, a column  
 of random elements having no correlation with the classes was stacked to the right of the matrix  $H$ . Figure 4 depicts  
 148 the absolute value of the weight that the model assigns to this disconnected neuron as a function of the number  
 of neurons  $M$ . As seen, this weight remains small until  $M$  approaches  $N$ , which supports our previous hypothesis.  
 150 It is timely to observe, however, that the weight assigned to the fake neuron starts to decay again after this point.  
 This means that for  $M > N$ , the ELM becomes capable of neglecting the value of the disconnected neuron, since  
 152 it is irrelevant for classification purposes.

#### 4. Relevance-based pruning

154 As we have shown, in the cases of unrepresentative features, an ELM can benefit from choosing  $M \gg N$ . Yet this  
 has the downside of increasing the network size. In order to find a compromise between ELM size and classification  
 156 performance, one could use a traditional validation method and retrain the network increasing the number of  
 neurons  $M$  until the change in performance is small enough to be neglected. Nevertheless, the computational  
 158 burden associated with solving (4) increases with  $M$ , and the sensitivity of the method with respect to the random  
 values might require a few initializations, making this idea unfeasible in practice. Hence, in order to make ELMs  
 160 of practical use for real data in real applications, like EEG signals classification in BCI devices, we propose a novel



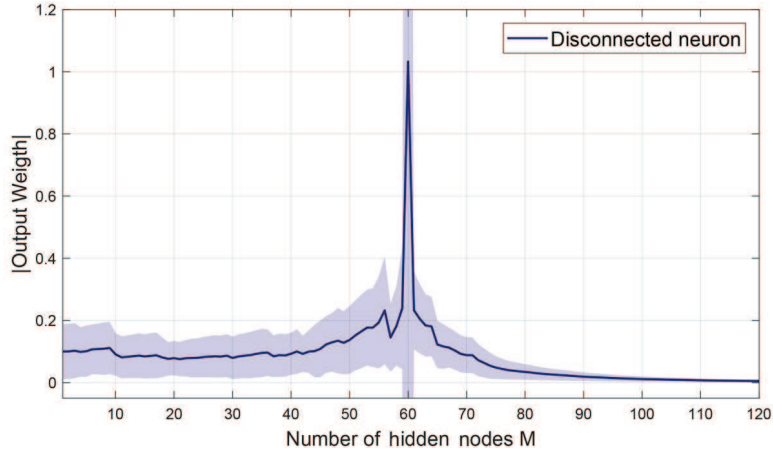


Figure 4: Mean absolute value for 1000 realizations (50 random initializations and 20 cross validations) of the weight  $\beta$  assigned to a “fake” neuron of the hidden layer. Shading illustrates  $\pm 1$  standard deviation of the results.

pruning method which allows for an efficient search of the parameter  $M$ , by appropriately reducing the network size after just a single computation of (4) for  $M \gg N$ .

As observed in the description of Figure 4, the weight of a neuron is proportional to the relevance of the corresponding node. Hence, it is reasonable to discard the neurons whose associated weights  $\beta$  are small enough. Given that the process of discarding a neuron and testing the performance of the resulting ELM is computationally inexpensive, the proposed pruning method begins using a large initial number  $M^*$  of hidden nodes, and then discard one (or a few) at a time, just until before the performance exhibits a significant drop. We shall refer to the resulting method as Relevance-Based Pruning (RBP). The steps for performing RBP are shown in Algorithm 1. In Section 4.1 we present different experiments in order to validate the pruning criterion and in Section 5, comparisons between this pruning method, the standard  $M < N$  setting, the traditional validation forward scheme and the OP-ELM, in terms of classification performance and computational costs, are also shown.

#### 4.1. Pruning criterion validation experiments

##### 4.1.1. Experimental setting

Two EEG datasets were used for the experiments. As previously described, EEG data typically presents high levels of noise, and the relevant classification information is mostly encoded in a particularly small subset of features. The first dataset is the imagined speech dataset, introduced in (DaSalla et al., 2009). It contains EEG signals recorded from three different subjects using 64 electrodes with a sampling frequency of 256 Hz. Here, we used the spatially filtered EEG signals related to imagination of mouth movement involved in the pronunciation of two vowels (/a/ and /u/) and a resting state, in pairwise comparisons (C1 = /a/ vs. rest; C2 = /u/ vs. rest; C3 = /a/ vs. /u/). 100 trials were presented for every comparison, with 512 features each. The second dataset is a P300-based BCI dataset introduced in (Ledesma-Ramirez et al., 2010), consisting of 3780 EEG trials of one second (630 trials with P300 activity), acquired from 25 subjects using 10 channels at 256 Hz. Each trial has 2560 features (256 samples x 10 channels) and each trial label corresponds to the presence or absence of P300 activity.

---

**Algorithm 1** : Relevance-Based Pruning (RBP)

---

Set  $M^* \gg N$  and  $\delta > 0$ .

Initialize the elements of  $W \in \mathbb{R}^{M^* \times D}$  and  $b \in \mathbb{R}^{M^*}$  randomly as realization of a distribution  $\mathcal{U}[-1, 1]$ ,  $M^*$  being the maximum number of hidden nodes and  $D$  the data dimension.

$$H = [g(WX_{train} + b \mathbf{1}_{(1,N)})]^T.$$

$$\hat{\beta} = H^\dagger y_{train}.$$

$$H = [g(WX_{val} + b \mathbf{1}_{(1,N)})]^T.$$

Permute  $\hat{\beta}$  so that  $|\hat{\beta}_m| \geq |\hat{\beta}_{m+1}|, \forall m = 1, \dots, M^*$ .

Perform the same permutation on the columns of  $H$ .

$$\text{Let } H' = H, \hat{\beta}' = \hat{\beta}$$

**while**  $\text{mean}(H\hat{\beta} - y_{val}) < \text{mean}(H'\hat{\beta}' - y_{val}) + \delta$  **do**

$$\text{Let } H' = H, \hat{\beta}' = \hat{\beta}$$

Remove the last element from  $\hat{\beta}$ .

Remove the last column from  $H$ .

**end while**

---

---

**Algorithm 2** : Standard ELM forward scheme

---

Set  $M = 1$  and  $\delta > 0$ .

Set  $acc_{val} = 0$  and  $acc'_{val} = 0$ .

**while**  $M < M^*$  and  $acc_{val} > acc'_{val} + \delta$  **do**

Initialize the elements of  $W \in \mathbb{R}^{M^* \times D}$  and  $b \in \mathbb{R}^{M^*}$  randomly as realization of a distribution  $\mathcal{U}[-1, 1]$ .

$$H = [g(WX_{train} + b \mathbf{1}_{(1,N)})]^T.$$

$$\hat{\beta} = H^\dagger y_{train}$$

$$acc'_{val} = acc_{val}$$

Compute  $acc_{val}$  using  $X_{val}$  and  $y_{val}$ .

$$M \leftarrow M + 1.$$

**end while**

---

#### 184 4.1.2. Classification performance analysis

In order to compare RBP against the standard “forward” ELM approach, we propose the following experimental  
186 setting: consider the datasets  $X_{train} \in \mathbb{R}^{D \times N}$  and  $X_{val} \in \mathbb{R}^{D \times N'}$  and a maximum ELM size  $M^*$ . For the traditional  
method, the steps followed are shown in Algorithm 2. It is worth mentioning that in these experiments, we aim  
188 to validate the pruning criterion rather than to find the optimal parameter  $M$  of the network. Therefore, the data  
were only split in train and test sets. Also, the Algorithms 2 and 1 ignore the stop parameter  $\delta$ .

190 Using 50 random initializations for  $W$  and  $b$ , for every value of  $M \in \{1, \dots, M^* = 1000\}$ , we run the proposed  
experiment for a randomly chosen subject of the DaSalla dataset, over 20 cross validations with a 70/30 train/test  
192 scheme. The resulting average validation accuracy for both the standard forward method (Algorithm 2) and the  
RBP method are depicted in Figure 5 (top). The results obtained using a random pruning scheme (RND) are used  
194 to compare and validate the proposed pruning criterion.

An analogous test was made using a (randomly chosen) subject of the P300 database. Results are shown in  
196 Figure 5 (bottom). Since in this case the dataset is much larger, instead of taking unitary increments on the values  
of  $M$ , a logarithmic grid was used. Also, as the dataset is unbalanced, the area under the ROC curve (AUC)  
198 was used to evaluate the classification performance. The results correspond to 5 cross validations over 20 random  
initializations, with a choice of  $M^* = 40000$ .

200 As seen in Figure 5, RBP performs at least as good as the forward scheme. This means that choosing a large  
value of  $M^*$  followed by a reduction of the network size by means of RBP (until just before a significant decay in  
202 validation performance is observed) will yield a result as good as using the forward scheme, and that can be done in  
a very short time, as the pruning procedure is computationally inexpensive. Thus, we can choose  $M^*$  large enough  
204 (e.g. as determined by the computational cost we are willing to pay), and then run RBP to reduce the network size  
and quickly find the optimal parameter  $M$  without losing classification performance, so ending up with a compact  
206 network, which is specially important in real time applications. Finally, the random pruning method validates our  
hypothesis that the less relevant nodes can be pruned with little or no impact in the classification performance.

## 208 5. Analysis of RBP global performance on real data

In order to validate our proposal for finding the optimal number of hidden nodes  $M$  in real applications, we  
210 made comparisons with four different schemes: STD, FWD, RBP, OP-ELM, which are described below :

**STD:** (traditional approach) best result obtained using  $M < N$ .

212 **FWD:** start with  $k = 0$  and  $M_k = 1$ . Increase  $M_k$  over a logarithmic grid until the validation performance levels  
up or the maximum value  $M^*$  is reached (see Algorithm 2).

214 **RBP:** start with  $k = 0$  and  $M_k = M^*$  and prune  $M_k$  over a logarithmic grid until just before the first time  
classification performance shows a significant (prescribed) reduction. The stopping criterion, grid and the  
216 value of  $M^*$  are set equal to those in FWD.

**OP-ELM:** the publicly available implementation of OP-ELM was used with  $M = M^*$  and  $M = N$ .

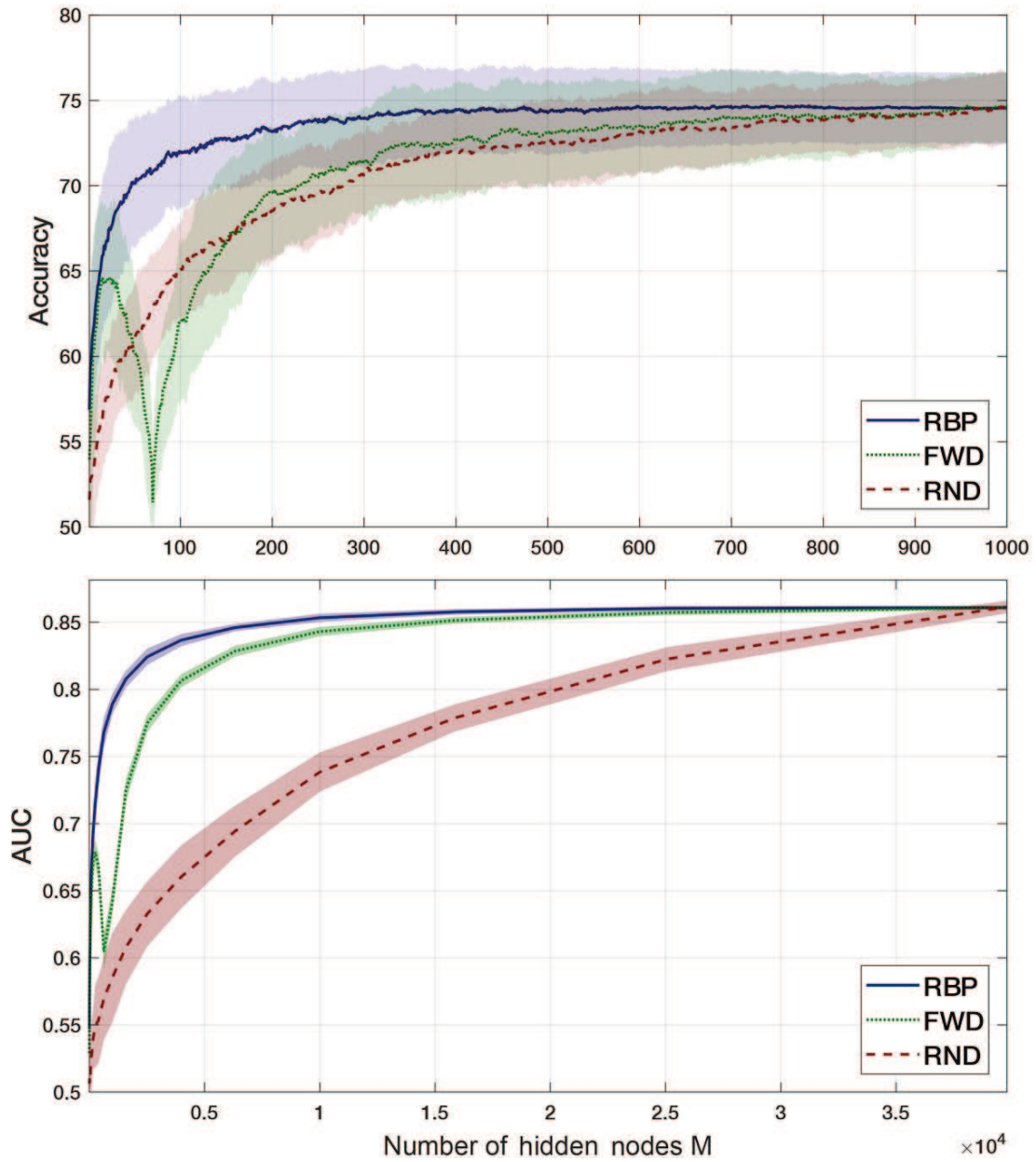


Figure 5: Average classification performance obtained with the traditional forward scheme (FWD), Random Pruning (RND) and Relevance-Based Pruning (RBP) methods, for one subject of the DaSalla dataset (top) and one subject of the P300 database (bottom). 70 and 2646 training examples were used for DaSalla and P300 experiment respectively. Shading accounts for  $\pm$  one standard deviation.

Method	Accuracy (%)	Time [ms]	$M$
STD	$62.2 \pm 7.3$	$76 \pm 48$	$24 \pm 18$
FWD	$66.9 \pm 5.5$	$195 \pm 21$	$228 \pm 76$
OP-ELM $M^* = 90$	$63.5 \pm 14.8$	$33.9 \pm 9$	$27 \pm 14$
OP-ELM $M^* = 1000$	$63.1 \pm 14.3$	$23471 \pm 54$	$77 \pm 10$
RBP	$68.0 \pm 3.8$	$43 \pm 3$	$209 \pm 151$

Table 1: Overall performance results yielded by each tested method for the DaSalla database.

Method	AUC	Time [s]	$M[\times 10^3]$
STD	$0.718 \pm 0.072$	$3 \pm 1$	$0.5 \pm 0.2$
FWD	$0.816 \pm 0.096$	$572 \pm 118$	$28.2 \pm 10.6$
RBP	$0.816 \pm 0.087$	$194 \pm 1$	$14.5 \pm 6.5$

Table 2: Overall performance results yielded by each tested method for the P300-based BCI dataset. Figure 6 also illustrates the performance of the five first subjects

218 For our first experiment the full DaSalla dataset (three subjects in three comparisons) was used. For all  
methods, 5 random initializations over 20 cross validations were run. While for STD, FWD and RBP the data  
220 were split in 70%, 20% and 10% for training, validation and testing, respectively, for OP-ELM, 90% training and  
10% testing were used, as the validation split is automatically made within the implementation. As the dataset  
222 is balanced, accuracy was used to measure classification performances. For FWD and RBP, the parameters were  
set to  $M^* = 1000$  and  $\delta = 0.02$ . Table 1 shows the obtained results, from which three main conclusions can be  
224 derived. First, both methods allowing an ELM with a large number of hidden nodes result in better performances,  
in terms of accuracy. Secondly, between those two approaches, RBP is observed to require much less computing  
226 time with a comparable network size. Finally, OP-ELM is not able to handle this type of problems, as the required  
time is several orders of magnitude higher and it does not yield better classification performance. As the OP-ELM  
228 method is not designed nor suggested for the case  $M > N$ , nor for the case of data with high redundancy, like EEG  
data, in the sequel, OP-ELM was neglected in all the analyses. A subject by subject analysis is presented in the  
230 Supplementary Information Section in Figures SI-5 and SI-6.

The same experiment was performed, but now using the P300-based BCI dataset. The data was also split into  
232 70%, 20% and 10% for training, validation and testing. For each one of the 25 subjects, five cross validations  
over 20 random initializations were performed. The parameters were set to  $M^* = 40000$  and  $\delta = 0.001$ . As  
234 previously mentioned, the P300-based BCI dataset is unbalanced, so the AUC was used for measuring classification  
performance. Overall results are shown in Table 2, and those obtained for the first five subjects are illustrated in  
236 Figure 6, along with overall performance.

As it can be seen, the AUC values obtained by FWD and RBP, *i.e.* the methods allowing for  $M > N$ , perform  
238 considerably better than those obtained with STD. While FWD and RBP do not account for practical differences in

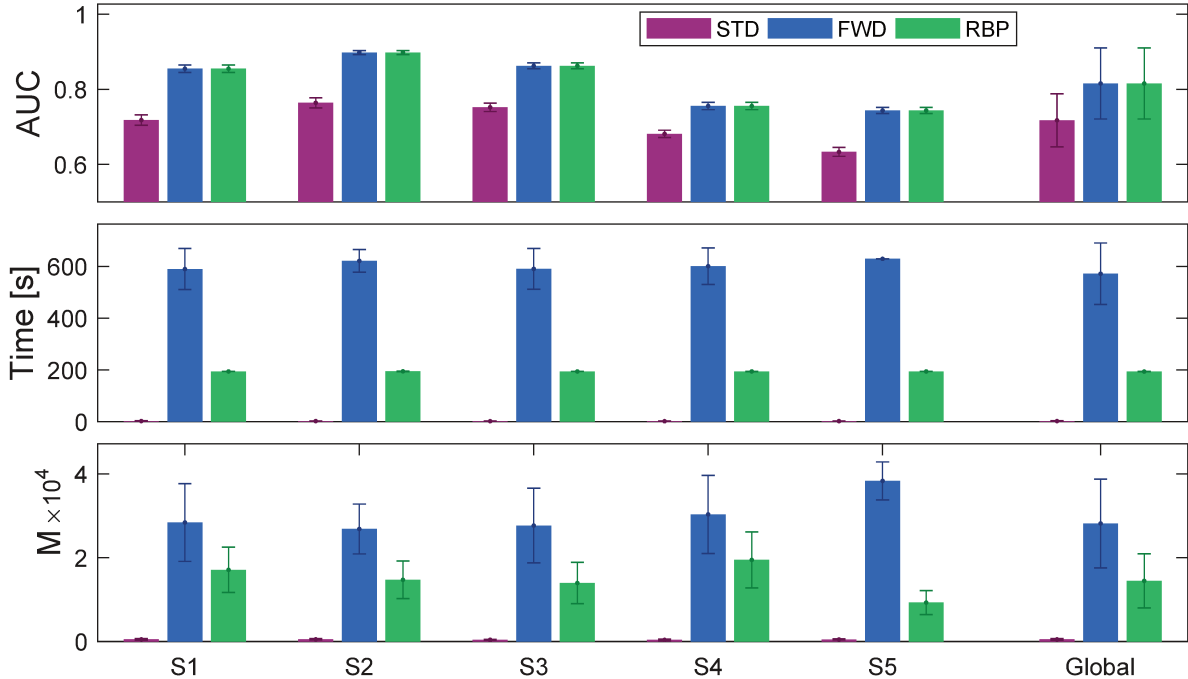


Figure 6: Results obtained with standard (STD), forward (FWD) and Relevance-Based Pruning (RBP) schemes on the P300-based BCI dataset.

AUC, RBP requires much less computing time than the former. Additionally, RBP yields a much smaller network size resulting in a more compact ELM.

From a practical point of view, if the number of hidden nodes is a restriction, then RBP is the best choice because it will yield better classification performance with the same number of neurons. On the other hand, if only classification is relevant, then RBP is more appropriate since it will yield the same performance in much less training time than FWD with a more compact network.

An experiment was also conducted for measuring the impact of the election of  $M^*$  in the training time, as the time performance could strongly depend on its choice. Comparison results for the FWR, RBP and OP-ELM methods are presented in Figures SI-7 and SI-8 of the Supplementary Information Section for the DaSalla and the P300 datasets, respectively. These results clearly show that the proposed RBP method is faster than both the traditional FWD and the OP-ELM algorithm, regardless of the  $M^*$  initialization. It is also timely to point out that time gain becomes larger for larger values of  $M^*$ . Therefore, the advantage of using the RBP method is greater in the cases of datasets containing a high number of  $N$  training examples.

## 6. Conclusions and future work

In this work we have shown that when dealing with classification problems in the context of data with unrepresentative features, such as EEG signals, ELMs benefit from choosing  $M \gg N$ . A detailed analysis on why standard approaches are suboptimal for this type of data was provided.

256 Results using real EEG data show that choosing a large number of hidden neurons is beneficial and that our  
proposed Relevance-Based Pruning method provides a time-efficient way to search for the optimal number of hidden  
258 nodes without jeopardizing classification performance. Furthermore, its implementation is very simple and it can  
result in great benefits in real application.

260 In regard to the limitations or constraints of the proposed RBP method we must mention that its use will certainly  
not lead to any advantage with respect to the other methods, when the data contain just a few unrepresentative  
262 features or are completely free of them. This could pose a problem in cases where we have absolutely no information  
about the existence of junk features in the data.

264 In the future, we shall tackle the problem of choosing an appropriate maximum ELM size ( $M^*$ ) depending on the  
problem. Also, we intend to incorporate this method in the context of regularized ELMs and explore its potential  
266 as a feature selection tool.

### Conflict of interest

268 The authors declare that they have no conflict of interest.

### Information Sharing Statement

270 The P300 based BCI dataset is publicly available at <https://akimpech.izt.uam.mx/p300db/>. The DaSalla  
imagined speech dataset was originally available at [http://www.brainliner.jp/data/brainliner-admin/Speech\\_Imagery\\_Dataset](http://www.brainliner.jp/data/brainliner-admin/Speech_Imagery_Dataset).  
272 The synthetic data, along with a Python and MatLab implementation of the Relevance-  
Based Pruned method are also publicly available to encourage reproducible research and can be accessed at  
274 [https://github.com/N-Nieto/Relevance\\_Base\\_Pruning](https://github.com/N-Nieto/Relevance_Base_Pruning). The OP-ELM implementation was downloaded from  
<https://research.cs.aalto.fi/aml/software.shtml>.

### 276 Acknowledgments

This research was funded in part by Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET,  
278 Argentina, through PIP 2014-2016 No. 11220130100216-CO, the Agencia Nacional de Promoción Científica y Tecno-  
lógica through PICT-2017-4596 and by Universidad Nacional del Litoral, UNL, through CAI+D-UNL 2016 PIC No.  
280 50420150100036LI.

### Bibliography

- 282 Alencar, A. S., Neto, A. R. R., & Gomes, J. P. P. (2016). A new pruning method for extreme learning machines  
via genetic algorithms. *Applied Soft Computing*, *44*, 101–107.
- 284 Belkin, M., Hsu, D., Ma, S., & Mandal, S. (2019). Reconciling modern machine-learning practice and the classical  
bias–variance trade-off. *Proceedings of the National Academy of Sciences*, *116*, 15849–15854.

- 286 DaSalla, C. S., Kambara, H., Sato, M., & Koike, Y. (2009). Single-trial classification of vowel speech imagery using  
common spatial patterns. *Neural Networks*, *22*, 1334–1339.
- 288 Ding, S., Guo, L., & Hou, Y. (2017). Extreme learning machine with kernel model based on deep learning. *Neural  
Computing and Applications*, *28*, 1975–1984.
- 290 Ding, S., Zhang, N., Xu, X., Guo, L., & Zhang, J. (2015). Deep extreme learning machine and its application in  
eeg classification. *Mathematical Problems in Engineering*, *2015*.
- 292 Duan, L., Bao, M., Miao, J., Xu, Y., & Chen, J. (2016). Classification based on multilayer extreme learning machine  
for motor imagery task from EEG signals. *Procedia Computer Science*, *88*, 176–184.
- 294 Heinz, W., Engl, M. H., & Neubauer., A. (1996). Regularization of inverse problems. *Mathematics and its Appli-  
cations (Dordrecht)*, *375*.
- 296 Holz, E. M., Botrel, L., Kaufmann, T., & Kübler, A. (2015). Long-term independent brain-computer interface home  
use improves quality of life of a patient in the locked-in state: a case study. *Archives of Physical Medicine and  
298 Rehabilitation*, *96*, S16–S26.
- Horn, R. A., & Johnson, C. R. (1990). *Topics in matrix analysis*. Cambridge University Press.
- 300 Huang, G.-B., & Babri, H. A. (1998). Upper bounds on the number of hidden neurons in feedforward networks  
with arbitrary bounded nonlinear activation functions. *IEEE Transactions on Neural Networks*, *9*, 224–229.
- 302 Huang, G.-B., Wang, D. H., & Lan, Y. (2011). Extreme learning machines: a survey. *International Journal of  
Machine Learning and Cybernetics*, *2*, 107–122.
- 304 Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: theory and applications. *Neurocom-  
puting*, *70*, 489–501.
- 306 Huang, G.-B., Zhu, Q.-Y., Siew, C.-K. et al. (2004). Extreme learning machine: a new learning scheme of feedforward  
neural networks. *Neural Networks*, *2*, 985–990.
- 308 Jin, Z., Zhou, G., Gao, D., & Zhang, Y. (2020). EEG classification using sparse Bayesian extreme learning machine  
for brain-computer interface. *Neural Computing and Applications*, *32*, 6601–6609.
- 310 Kong, W., Guo, S., Long, Y., Peng, Y., Zeng, H., Zhang, X., & Zhang, J. (2018). Weighted extreme learning  
machine for P300 detection with application to brain computer interface. *Journal of Ambient Intelligence and  
312 Humanized Computing*, (pp. 1–11).
- Ledesma-Ramirez, C., Bojorges-Valdez, E., Yáñez-Suarez, O., Saavedra, C., Bougrain, L., & Gentiletti, G. G.  
314 (2010). An open-access P300 speller database. In *Fourth International Brain-Computer Interface Meeting*.  
Asilomar, California, USA.



- 316 Liang, N.-Y., Saratchandran, P., Huang, G.-B., & Sundararajan, N. (2006). Classification of mental tasks from EEG signals using extreme learning machine. *International Journal of Neural Systems*, *16*, 29–38.
- 318 Luo, J., Vong, C.-M., & Wong, P.-K. (2013). Sparse bayesian extreme learning machine for multi-classification. *IEEE Transactions on Neural Networks and Learning Systems*, *25*, 836–843.
- 320 Miche, Y., Sorjamaa, A., Bas, P., Simula, O., Jutten, C., & Lendasse, A. (2010). OP-ELM: optimally pruned extreme learning machine. *IEEE Transactions on Neural Networks*, *21*, 158–162.
- 322 Murugavel, A. M., & Ramakrishnan, S. (2016). Hierarchical multi-class SVM with ELM kernel for epileptic EEG signal classification. *Medical & Biological Engineering & Computing*, *54*, 149–161.
- 324 Nicolas-Alonso, L. F., & Gomez-Gil, J. (2012). Brain computer interfaces, a review. *Sensors*, *12*, 1211–1279.
- Rong, H.-J., Ong, Y.-S., Tan, A.-H., & Zhu, Z. (2008). A fast pruned-extreme learning machine for classification  
326 problem. *Neurocomputing*, *72*, 359–366.
- Schmidt, W. F., Kraaijveld, M. A., Duin, R. P. et al. (1992). Feed forward neural networks with random weights.  
328 In *International Conference on Pattern Recognition* (pp. 1–1). IEEE Computer Society Press.
- Shi, L.-C., & Lu, B.-L. (2013). EEG-based vigilance estimation using extreme learning machines. *Neurocomputing*,  
330 *102*, 135–143.
- Similä, T., & Tikka, J. (2005). Multiresponse sparse regression with application to multidimensional scaling. In  
332 *International Conference on Artificial Neural Networks* (pp. 97–102). Springer.
- Song, Y., & Zhang, J. (2013). Automatic recognition of epileptic EEG patterns via extreme learning machine and  
334 multiresolution feature extraction. *Expert Systems with Applications*, *40*, 5477–5489.
- Tan, P., Sa, W., & Yu, L. (2016). Applying extreme learning machine to classification of EEG BCI. In *2016 IEEE  
336 International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)* (pp. 228–232). doi:10.1109/CYBER.2016.7574827.
- 338 Tavares, L. D., Saldanha, R. R., Vieira, D. A., & Lisboa, A. C. (2014). A comparative study of extreme learning machine pruning based on detection of linear independence. In *2014 IEEE 26th International Conference on Tools with Artificial Intelligence* (pp. 63–69). IEEE.
- 340 Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, *113*, 767–791.
- Yuan, Q., Zhou, W., Li, S., & Cai, D. (2011). Epileptic EEG classification based on extreme learning machine and  
344 nonlinear features. *Epilepsy Research*, *96*, 29–38.

Zhang, Y., Wang, Y., Zhou, G., Jin, J., Wang, B., Wang, X., & Cichocki, A. (2018). Multi-kernel extreme learning  
346 machine for EEG classification in brain-computer interfaces. *Expert Systems with Applications*, *96*, 302–310.

Zhao, H., Guo, X., Wang, M., Li, T., Pang, C., & Georgakopoulos, D. (2018). Analyze EEG signals with extreme  
348 learning machine based on pmis feature selection. *International Journal of Machine Learning and Cybernetics*,  
*9*, 243–249.

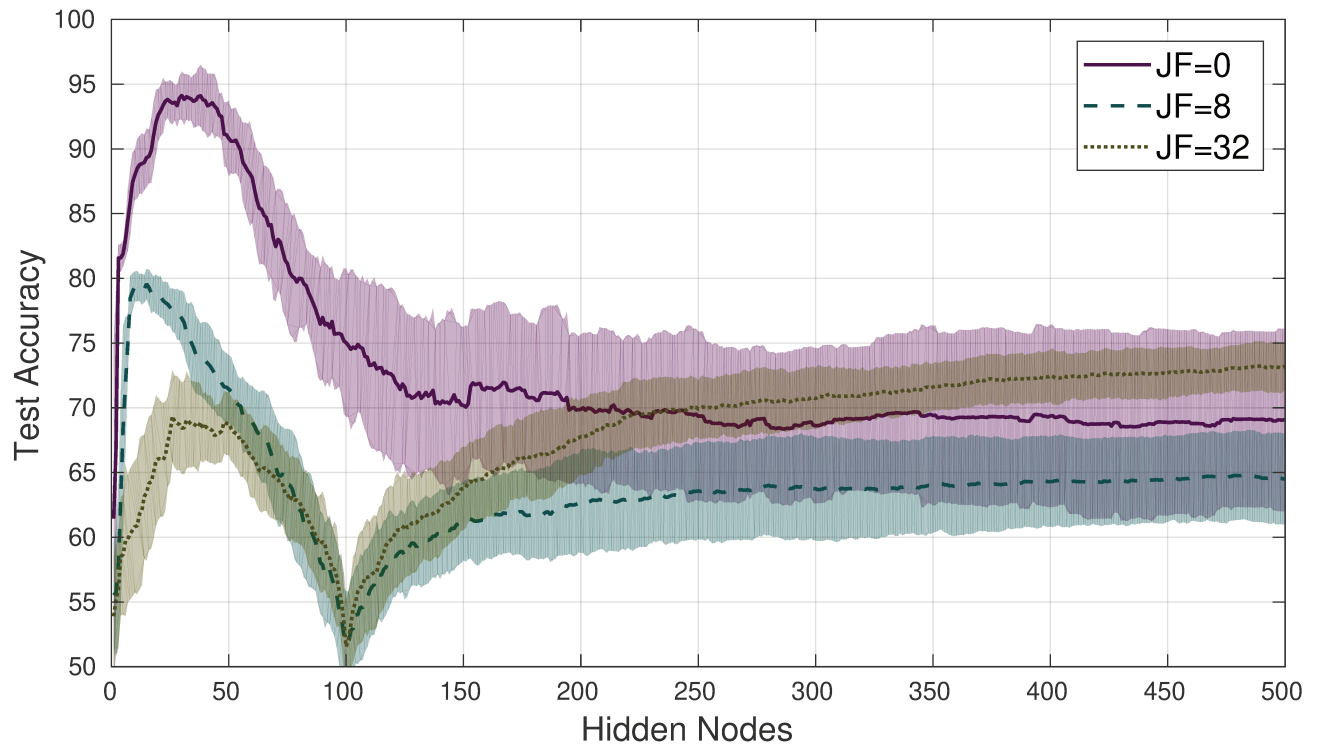


Figure 1: Mean test accuracy for 25 realizations (5 random initializations of parameters  $W$  and  $b$  and 5 cross validations of the data) in synthetic data adding Junk Features (JF) generated from a uniform distribution in the interval  $[-1,1]$ . 50 examples by class are used for training. Shading illustrates  $\pm 1$  standard deviation.

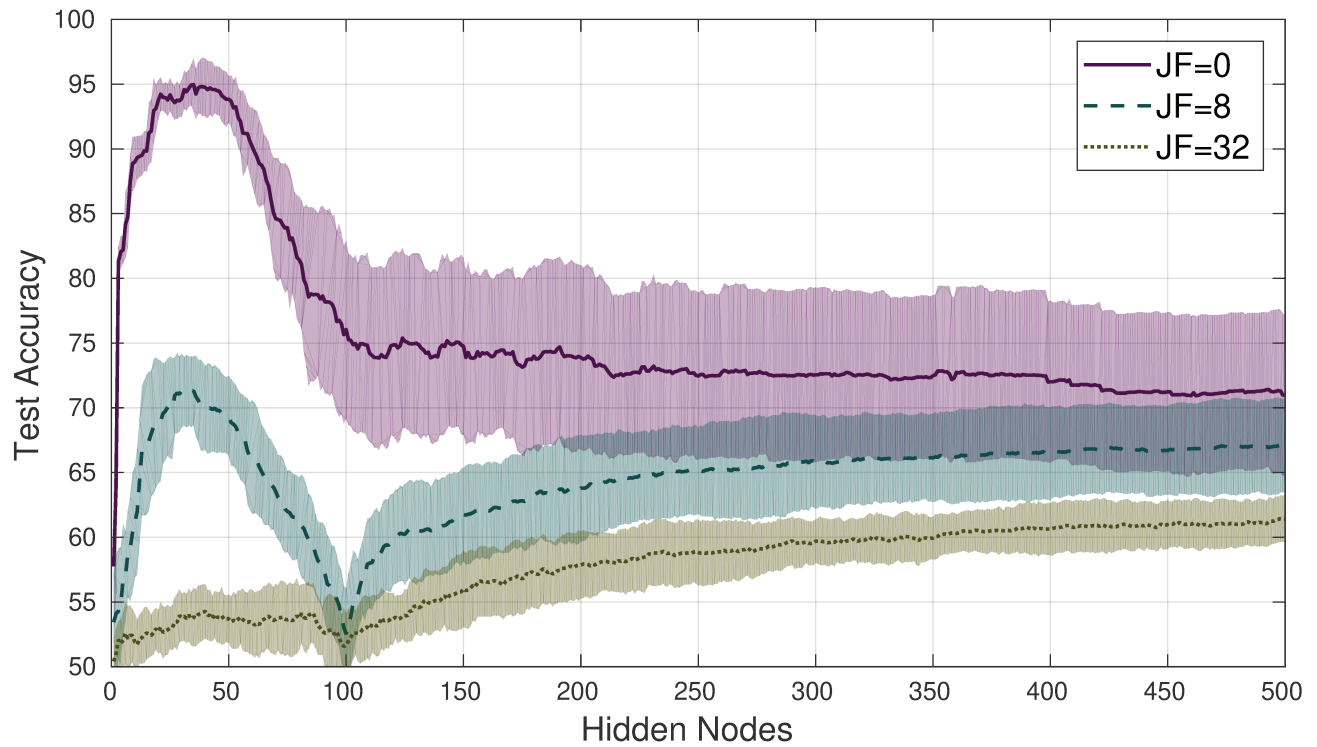


Figure 2: Mean test accuracy for 25 realizations (5 random initializations of parameters  $W$  and  $b$  and 5 cross validations of the data) in synthetic data adding Junk Features (JF) generated from a normal distribution in the interval  $[0, 1]$ . 50 examples by class are used for training. Shading illustrates  $\pm 1$  standard deviation.

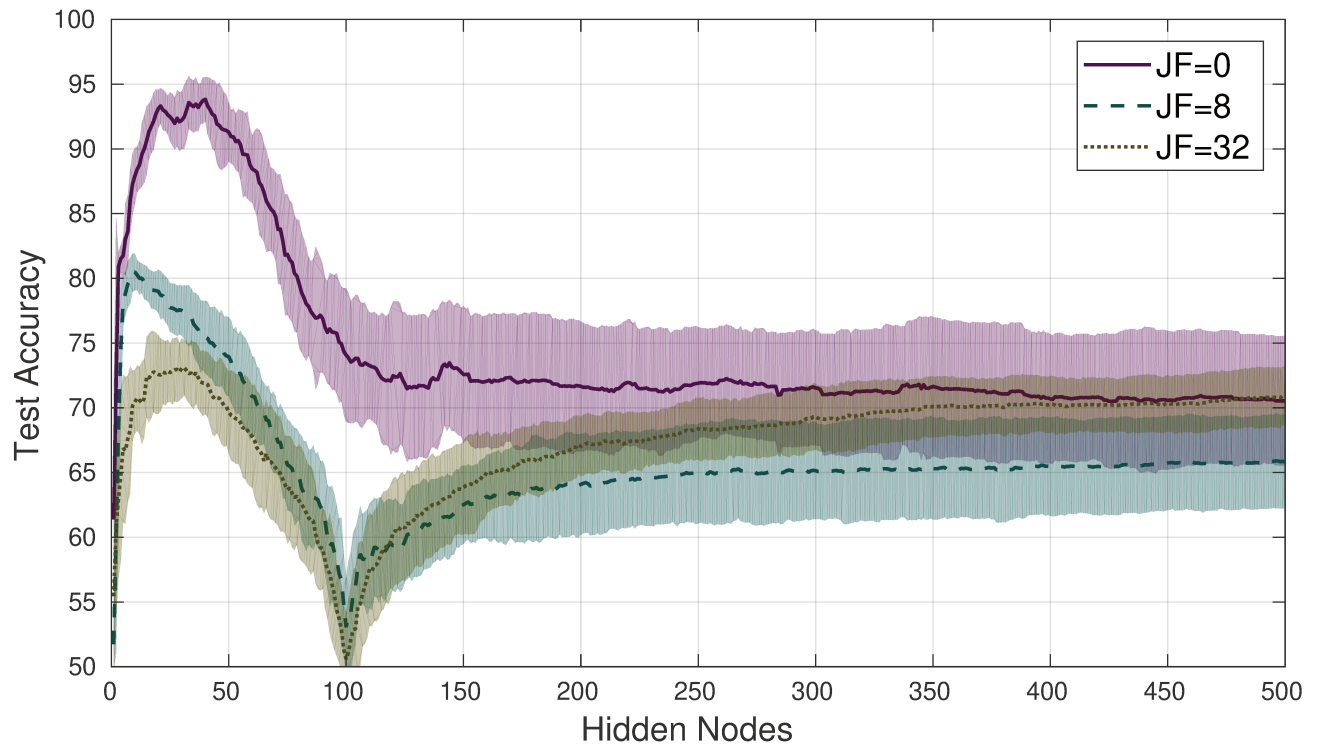


Figure 3: Mean test accuracy for 25 realizations (5 random initializations of parameters  $W$  and  $b$  and 5 cross validations of the data) in synthetic data adding Junk Features (JF) generated from a exponential distribution. 50 examples by class are used for training. Shading illustrates  $\pm 1$  standard deviation.

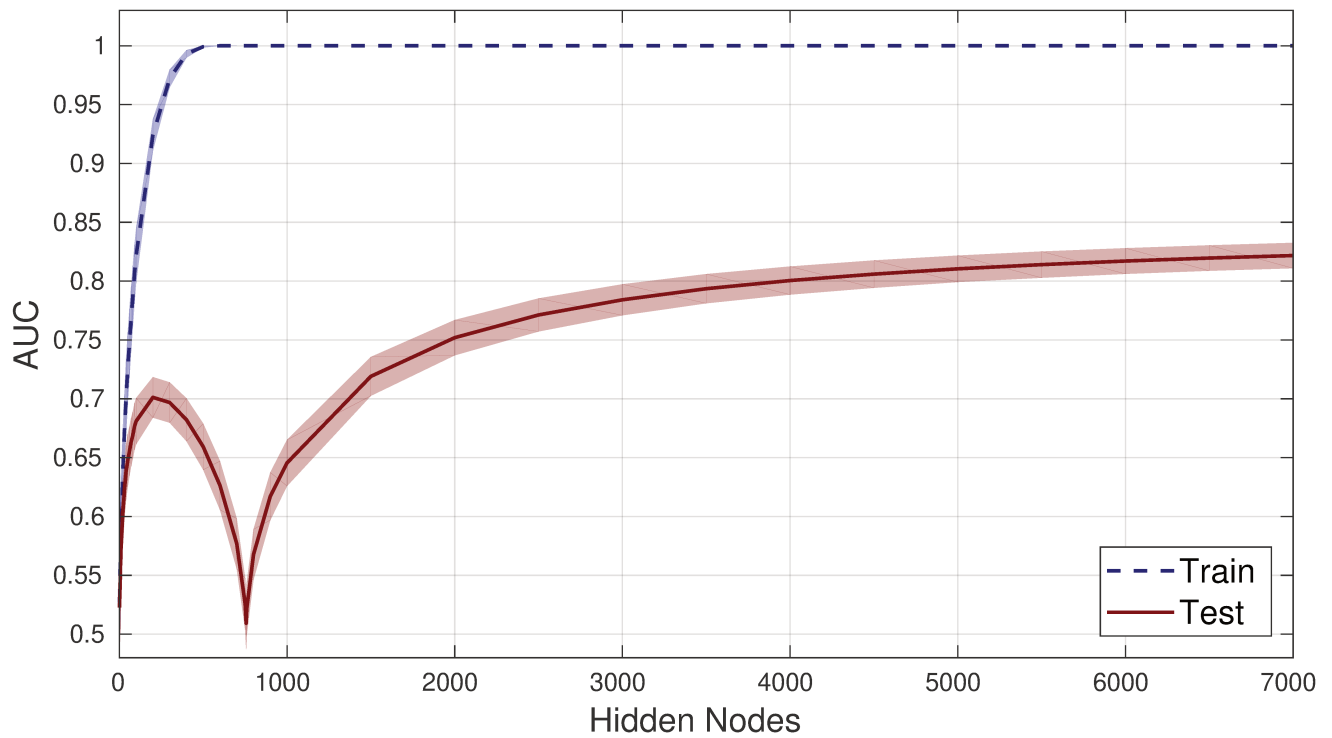


Figure 4: Mean accuracy for 1000 realizations (50 random initializations and 20 cross validations) obtained for training and testing data, as a function of the hidden layer size,  $M$ . 756 training examples for P300 dataset were used. Shading illustrates  $\pm 1$  standard deviation of the results.

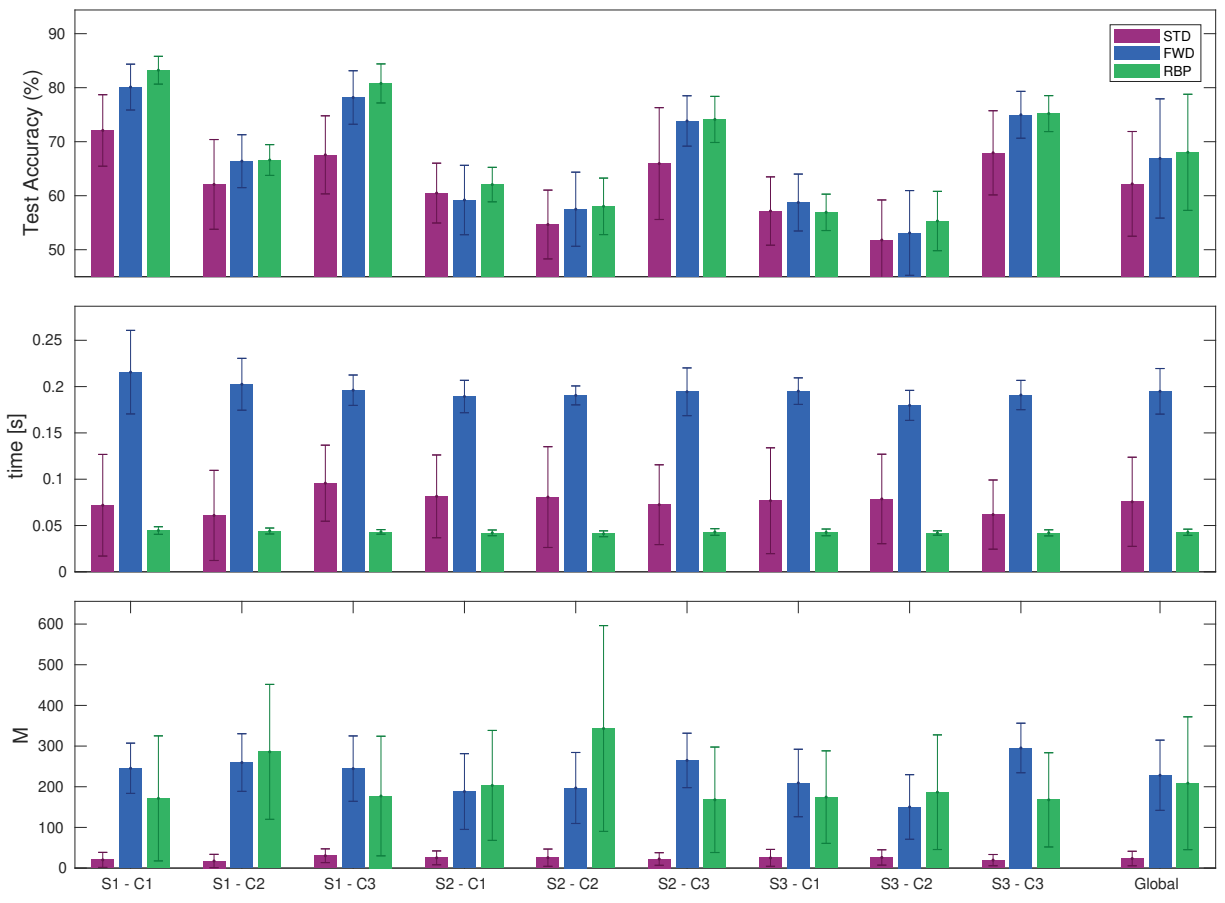


Figure 5: Results obtained with standard (STD), forward (FWD) and Relevance-Based Pruning (RBP) schemes on the complete DaSalla dataset. 3 Subjects (S) in 3 comparisons (C)

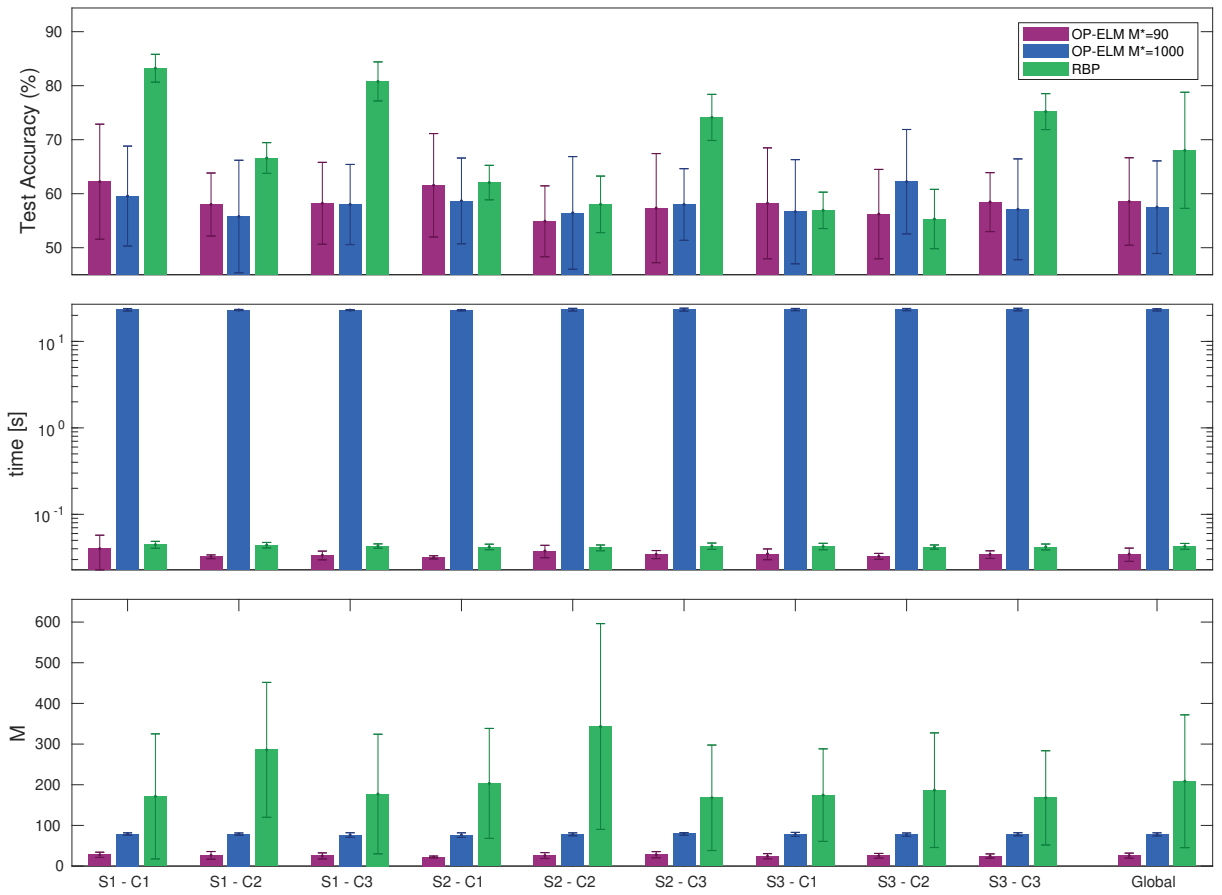


Figure 6: Results obtained with OP-ELM with the initialization of  $M^* = 90$ ,  $M^* = 1000$  and the Relevance-Based Pruning (RBP) method on the complete DaSalla dataset. 3 Subjects (S) in 3 comparisons (C)



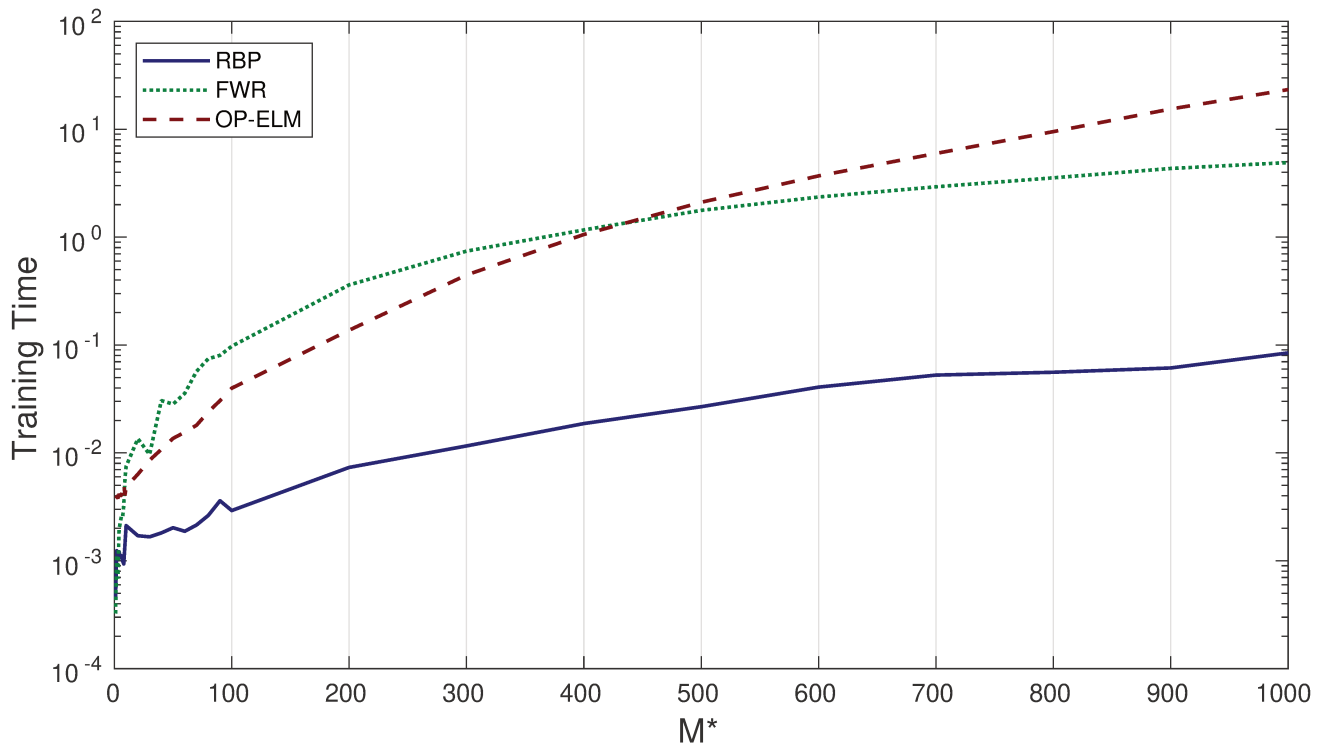


Figure 7: Training time results obtained for different initializations of  $M^*$  with forward ELM (FWR), Relevance-Based Pruning (RBP) and OP-ELM on subject 1 comparison 1 of the DaSalla dataset. 80 training examples were used

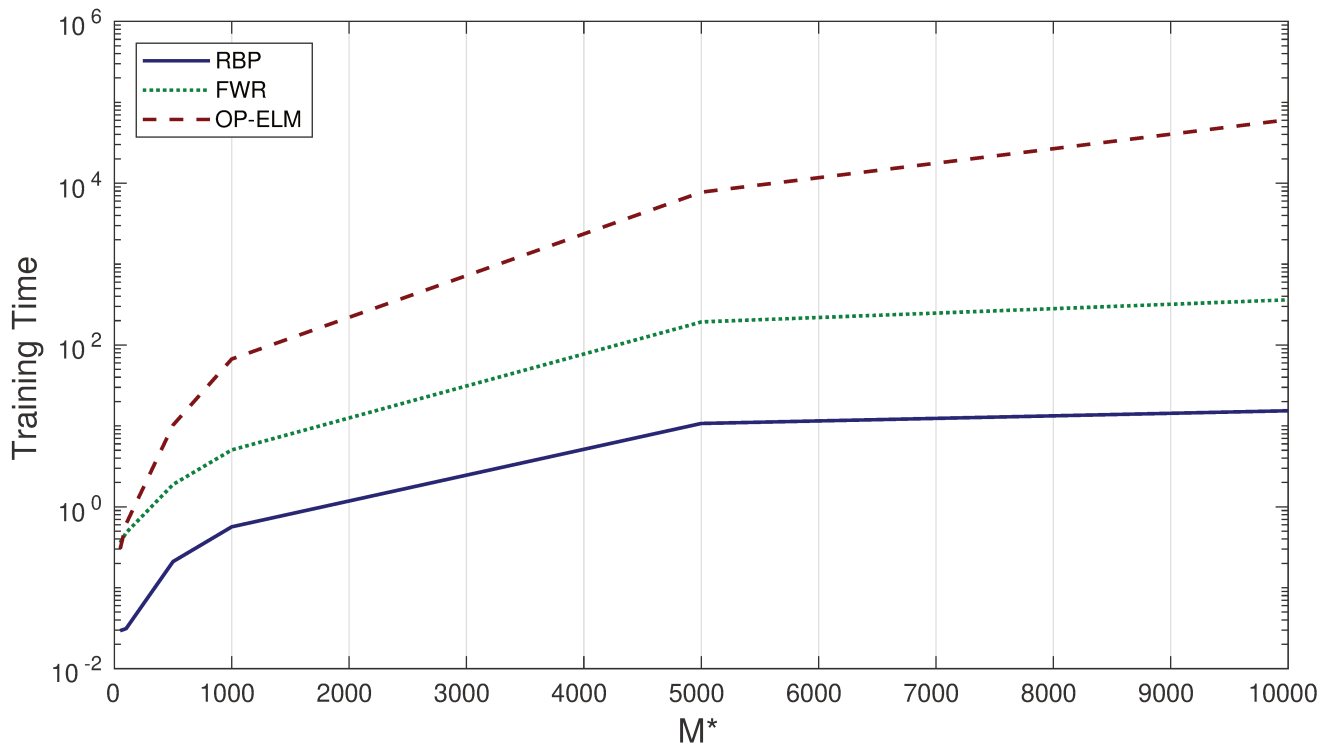


Figure 8: Training time results obtained for different initializations of  $M^*$  with forward ELM (FWD), Relevance-Based Pruning (RBP) and OP-ELM on Subject 1 of the P300 dataset. 3024 training examples were used

# Anexo III: Podado pre-entrenamiento de máquinas de aprendizaje extremo

# Podado Pre-Entrenamiento de Máquinas de Aprendizaje Extremo

Nicolás Nieto<sup>1</sup>, Guido Bracalenti<sup>1</sup>, Iván Gareis<sup>1,2</sup>, and Hugo L. Rufiner<sup>1,2</sup>

<sup>1</sup> Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(*i*), FICH–UNL/CONICET, Argentina  
 nnieto@sinc.unl.edu.ar,

<sup>2</sup> Laboratorio de Cibernética, Fac. de Ing., Univ. Nacional de Entre Ríos, Argentina

**Resumen** Las Máquinas de Aprendizaje Extremo son una eficiente herramienta de aprendizaje maquina, con tiempos de entrenamiento reducidos y buena capacidad de generalización. Este tipo de redes suele poseer una gran cantidad de neuronas en su capa oculta, no siendo necesariamente todas de utilidad. En este trabajo se propone un nuevo método de podado pre-entrenamiento que utiliza información de las distribuciones de probabilidad asociadas a las activaciones de cada neurona. Para la evaluación del método propuesto se utiliza como ejemplo de aplicación la clasificación de señales de electroencefalografía, registradas durante tareas de habla imaginada. Se muestran resultados preliminares que evidencian la potencialidad del método propuesto.

**Keywords:** Máquinas de aprendizaje extremo, Podado de redes neuronales, Estadística de alto orden, Habla Imaginada

## 1. Introducción

Las Máquinas de Aprendizaje Extremo (ELM por sus siglas en inglés) son redes neuronales de una única capa oculta que se caracterizan por inicializar de forma aleatoria los pesos de la capa de entrada y por determinar de forma analítica las conexiones de la capa oculta con la capa de salida [1]. Esto les permite realizar un entrenamiento más rápido respecto a los perceptrones multicapa *tradicionales*, que poseen una estructura similar pero con un aprendizaje iterativo. Muchas veces para lograr un buen desempeño con las ELM se requiere un número relativamente elevado de neuronas aleatorias en la capa oculta. Sin embargo, no todas las neuronas resultan útiles e incrementan innecesariamente el tiempo de entrenamiento. Es por ello que, a fin de disminuir la cantidad de neuronas, se aplican métodos de podado que permitan eliminar las neuronas que resultan menos útiles. El principal desafío de estas técnicas es encontrar la manera de *medir* la utilidad de cada neurona. Este problema ha sido abordado anteriormente por distintos autores [2,3,4,5]. En este trabajo se propone caracterizar el comportamiento de las activaciones de las neuronas de la capa oculta mediante estadística de alto orden, que podría indicar cuales aportan información valiosa.

Se desarrolla un nuevo método de podado pre-entrenamiento que utiliza la estimación de la curtosis de las funciones de distribución de probabilidad (FDP) de dichas activaciones. Para evaluar el desempeño del método se utilizan datos reales de electroencefalografía (EEG), registradas en la tarea de habla imaginada [6]. La clasificación de señales de EEG ha sido uno de los problemas donde se han utilizado las ELM [7,8,9].

## 2. Máquinas de Aprendizaje Extremo

Las ELM son redes neuronales de aprendizaje supervisado, propuestas inicialmente por Huang et. al., que cuentan teóricamente con capacidad de aproximación universal [1]. Sean  $(\mathbf{x}_i, t_i)$  los  $N$  ejemplos disponibles para un problema de clasificación, donde  $\mathbf{x}_i \in \mathbb{R}^d$  es el  $i$ -ésimo patrón ( $d$  es la dimensión de los patrones de entrada) y  $t_i \in \mathbb{R}$  su correspondiente etiqueta. Sea también  $g_j : \mathbb{R}^d \times \mathbb{R}^L \times \mathbb{R} \rightarrow \mathbb{R}$  la función de activación de la  $j$ -ésima neurona oculta, donde  $L$  es la cantidad de neuronas de la capa oculta. Utilizando esta notación la salida de la  $j$ -ésima neurona oculta a la  $i$ -ésima entrada puede escribirse como  $h_{i,j} = g_j(\mathbf{x}_i; \mathbf{w}_j, b_j)$ , donde  $\mathbf{w}_j \in \mathbb{R}^d$  y  $b_j \in \mathbb{R}$  son los parámetros asociados a la neurona  $j$ , fijados aleatoriamente. En el caso particular de este trabajo todas las neuronas ocultas fueron aditivas con funciones de activación sigmoideas. Teniendo en cuenta estas consideraciones se puede escribir  $h_{i,j} = S(\mathbf{x}_i \cdot \mathbf{w}_j + b_j)$ , donde  $S(x) = (1 - e^{-x}) / (1 + e^{-x})$ . Definiendo la matriz de activaciones ocultas como  $\mathbf{H} = (h_{i,j}) \in \mathbb{R}^{N \times L}$ , el vector de salidas de la red puede calcularse como  $\mathbf{o} = \mathbf{H}\hat{\boldsymbol{\beta}}$ , donde  $\hat{\boldsymbol{\beta}}$  es el vector de pesos de la capa de salida. Sea  $\mathbf{t} = [t_1, \dots, t_N]^T$  el vector de salidas deseadas óptimo, el  $\hat{\boldsymbol{\beta}}$  es el que minimiza la expresión  $\|\mathbf{o} - \mathbf{t}\|$ . Típicamente el conjunto óptimo de pesos de salida se obtiene en forma analítica como  $\hat{\boldsymbol{\beta}} = \mathbf{H}^\dagger \mathbf{t}$ , siendo  $\mathbf{H}^\dagger$  la inversa generalizada de Moore-Penrose.

## 3. Materiales y Método Propuesto

La base de datos utilizada cuenta con señales de EEG de tres sujetos en la tarea de imaginar el movimiento de los labios para pronunciar los fonemas /a/, /u/ y un estado de *control* sin acción. Tras un filtrado espacial [10] se obtienen series temporales que maximizan la varianza entre cada uno de los pares de clases: /a/ vs control, /a/ vs /u/ y /u/ vs control.

A fin de poder caracterizar el comportamiento de las neuronas ocultas frente a los datos, en la Fig. 1 se presentan los histogramas de activación condicional por clase típicos. La Fig.1a corresponde a un comportamiento de tipo bimodal. En este caso las activaciones parecen no tener información relevante respecto de los datos de entrada, dando un número similar de activaciones positivas y negativas para la misma clase. En la Fig.1b se presenta un comportamiento de tipo unimodal, en el cual sí parece apreciarse una correlación entre las activaciones y los datos de entrada para una clase dada. El hecho de que estos histogramas sean los condicionales por clase, implica que una misma neurona podría tener

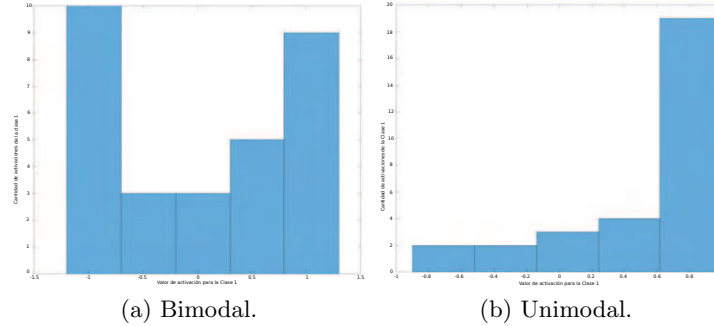


Figura 1: Histogramas de activación condicional típicos de las neuronas aleatorias de la capa oculta para una clase determinada.

una FDP bimodal para ambas clases, bimodal para una clase y unimodal para la otra, o bien unimodal para ambas clases. La hipótesis consiste entonces en suponer que las neuronas que presenten un comportamiento bimodal pueden ser eliminadas. Con el fin de poder identificar este tipo de neuronas se estimó la curtosis de las FDP de las activaciones para cada clase. Este momento estadístico brinda información sobre la forma de la distribución. Un mayor valor implica una concentración de activaciones alrededor de la media (FDP más *picuda*). Sea  $\hat{h}_j^{(\ell)}$  la media de las activaciones de la neurona  $j$  tomando como entrada los patrones correspondientes a la clase  $\ell$ . La curtosis de la FDP asociada a las activaciones de la neurona  $j$  y la clase  $\ell$  puede calcularse como,

$$K_j^{(\ell)} = \frac{\frac{1}{n_\ell} \sum_{i|t_i=c_\ell} (h_{i,j} - \hat{h}_j^{(\ell)})^4}{\left(\frac{1}{n_\ell} \sum_{i|t_i=c_\ell} (h_{i,j} - \hat{h}_j^{(\ell)})^2\right)^2}, \quad (1)$$

donde  $c_\ell$  es la etiqueta y  $n_\ell$  es el número de patrones de entrenamiento asociados a la clase  $\ell$ . Utilizando la Ec. (1) podemos definir la figura de mérito utilizada para evaluar la neurona  $j$  como  $M_j = K_j^{(c_1)} + K_j^{(c_2)}$ . De esta forma se penaliza aquellas neuronas con un comportamiento del tipo bimodal, cuya curtosis es pequeña, frente a las de activación unimodal.

#### 4. Resultados y Discusión

En la Fig. 2 se compara el ELM clásico (rojo) y con podado (azul) en la tarea de clasificación entre */a/ vs control* para el sujeto 1 de [6]. Se utilizaron 60 ejemplos de entrenamiento y se dejaron 40 ejemplos para el conjunto de prueba. Los experimentos se corrieron con 100 inicializaciones de  $\mathbf{w}_j$  y  $b_j$  distintas, reportándose media y desvío finales para diferentes cantidades de neuronas en la capa oculta. El algoritmo selecciona el 20% de las mejores neuronas previo a la etapa de entrenamiento. Para analizar qué tipo de neuronas resultaron elimina-

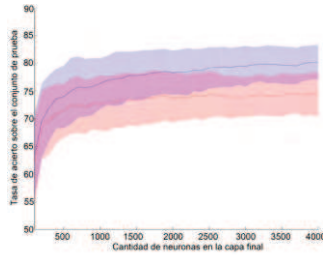


Figura 2: Tasa de acierto *vs* cantidad de neuronas finales en la capa oculta.

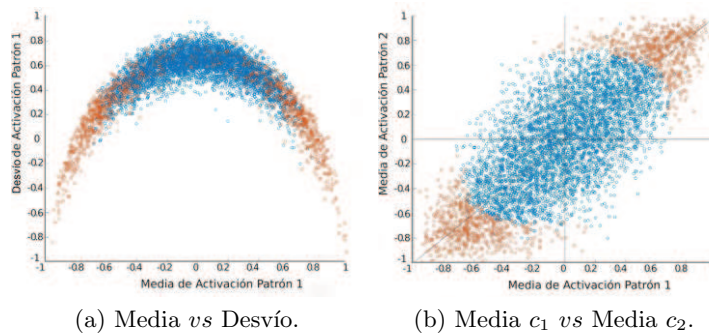


Figura 3: Neuronas eliminadas/podadas (azul), seleccionadas (naranja)

das se realizó un gráfico que muestra la relación entre el desvío y la media de la distribución para una clase (Fig.3a) y otro gráfico donde se muestran la media de las activaciones para cada clase (Fig. 3b). Se puede apreciar que el método propuesto elimina aquellas neuronas que poseen una media cercana a cero para ambas distribuciones, característica de las neuronas con una FDP bimodal.

### 5. Conclusiones y Trabajo Futuro

Se ha mostrado que con el método propuesto es posible identificar en forma exitosa las neuronas más útiles a partir de información estadística de la matriz de activaciones  $H$ . Además, gracias al podado de las neuronas con FDP bimodal, el método mejora la capacidad de clasificación de las ELM. Resulta relevante destacar que a priori las neuronas con FDP condicionales unimodales similares para ambas clases no aportarían información discriminativa, sin embargo el algoritmo propuesto no las eliminaría. Si bien este problema escapa al alcance de este trabajo, ha sido contemplado y se proyecta abordarlo con métodos de doble podado. Por otra parte, se están evaluando otros estadísticos, como la asimetría, para agregar información adicional a la curtosis que pueda resultar de utilidad para el podado.

## Referencias

1. G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006.
2. Y. Miche, A. Sorjamaa, P. Bas, O. Simula, C. Jutten, and A. Lendasse, "Op-elm: optimally pruned extreme learning machine," *IEEE transactions on neural networks*, vol. 21, no. 1, pp. 158–162, 2010.
3. A. S. Alencar, A. R. R. Neto, and J. P. P. Gomes, "A new pruning method for extreme learning machines via genetic algorithms," *Applied Soft Computing*, vol. 44, pp. 101–107, 2016.
4. L. D. Tavares, R. R. Saldanha, D. A. Vieira, and A. C. Lisboa, "A comparative study of extreme learning machine pruning based on detection of linear independence," in *Tools with Artificial Intelligence (ICTAI), 2014 IEEE 26th International Conference on*, pp. 63–69, IEEE, 2014.
5. M. Sánchez-Gutiérrez, E. M. Albornoz, H. L. Rufiner, and J. G. Close, "Post-training discriminative pruning for rbms," *Soft Computing*, pp. 1–15, 2017.
6. C. S. DaSalla, H. Kambara, M. Sato, and Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns," *Neural networks*, vol. 22, no. 9, pp. 1334–1339, 2009.
7. S. Ding, N. Zhang, X. Xu, L. Guo, and J. Zhang, "Deep extreme learning machine and its application in eeg classification," *Mathematical Problems in Engineering*, vol. 2015, 2015.
8. L. Gao, W. Cheng, J. Zhang, and J. Wang, "Eeg classification for motor imagery and resting state in bci applications using multi-class adaboost extreme learning machine," *Review of Scientific Instruments*, vol. 87, no. 8, p. 085110, 2016.
9. B. Min, J. Kim, H.-j. Park, and B. Lee, "Vowel imagery decoding toward silent speech bci using extreme learning machine with electroencephalogram," *BioMed research international*, vol. 2016, 2016.
10. J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial eeg classification in a movement task," *Clinical neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.



Anexo IV: Preliminary feasibility  
analysis of inner speech as a control  
paradigm for brain-computer  
interfaces

# Preliminary feasibility analysis of inner speech as a control paradigm for brain-computer interfaces

Nicolás Nieto<sup>1,2,\*</sup>, Hugo L. Rufiner<sup>1</sup> and Ruben Spies<sup>2</sup>

<sup>1</sup> Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(i), UNL-CONICET

<sup>2</sup> Instituto de Matemática Aplicada del Litoral, IMAL, UNL-CONICET.  
`*nnieto@sinc.unl.edu.ar`

**Abstract.** Brain Computer Interfaces (BCIs) are useful devices that provide new ways of communication to people who have lost the capability of interacting with their environment. Although several paradigms have resulted in large improvements in the construction of BCIs, quite often they require great efforts from the patient or they are not able to generate natural and efficient interfaces. In that scenario, inner speech appears as a promising paradigm for tackling those problems. Nevertheless, the lack of publicly available databases largely precludes the analysis and development of methods for using this paradigm. In this work we use a recently released database to show that it is possible to classify and differentiate inner speech signals from signals acquired within other two well known paradigms. This is undoubtedly a first step in the search and construction of an inner speech based BCI.

**Keywords:** Electroencephalography · Machine Learning · Brain Pattern Recognition · Extreme Learning Machines

## 1 Introduction

Spinal injuries, strokes, cerebral palsy, amyotrophic lateral sclerosis, among other diseases, can interrupt the normal pathways that the brain uses for muscle control. For such patients, Brain Computer Interfaces (BCIs) provide an alternative way of interaction with the environment, offering great benefits [28,12]. In a BCI, the brain activity is usually measured by surface electroencephalography (EEG), as it is a standard and noninvasive technique [21]. EEG provides signals with good time resolution but with a poor spatial resolution and low signal-to-noise ratio. Once the signals are obtained, they are typically classified by machine learning techniques. These classifiers use the EEG signals to generate outputs for controlling external devices (wheelchairs, computers, etc.)

The so called “inner speech” paradigm has been studied using EEG [3,8,26], electrocorticography [24], functional magnetic resonance imaging and positron emission tomography scan [10,25,11,20]. The potential advantages of using inner speech as a control paradigm are clear, as it can generate more natural interfaces, allowing patients to execute an order, literally by just thinking about it. Nevertheless, compared to other paradigms, inner speech involves more complex neural networks of different cortical areas engaged in phonological and semantic analysis, speech production and other processes [24,16,1].

Although the understanding of inner speech has largely increased in the last few years, there is still not enough evidence to support the conjecture that this paradigm can in fact be used for efficiently controlling a BCI. The aim of this work is to show that it is possible to classify and differentiate inner speech signals from those acquired within other two well known paradigms. This is undoubtedly a first step in the search and construction of an inner speech-based BCI.

## 2 Materials and Methods

### 2.1 Data description

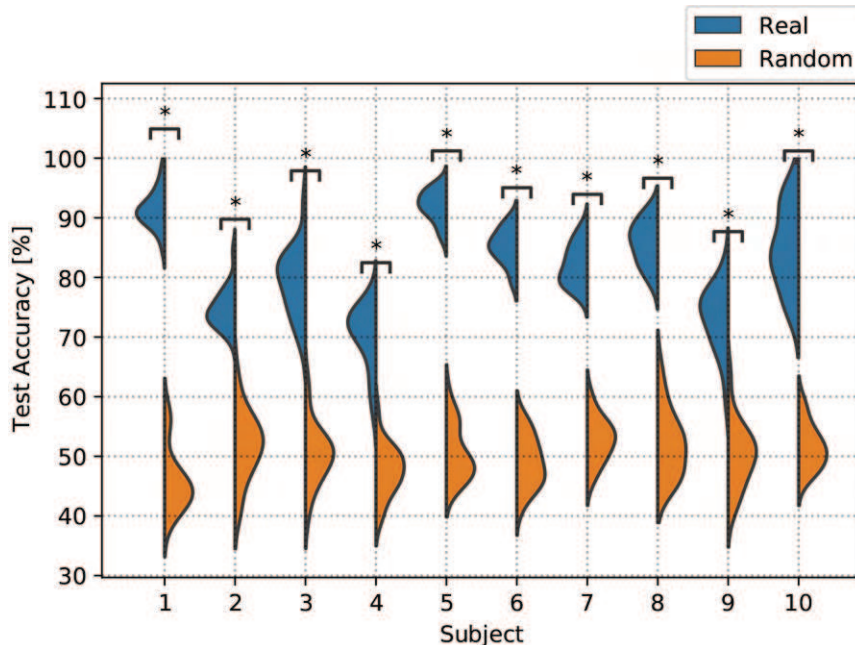
**Tasks and participants.** For the experiments, the dataset presented in [23] was used. This dataset contains EEG signals from ten healthy participants, all right-handed and native Spanish speakers. The participants were requested to perform three different conditions (paradigms): pronounced speech, inner speech and visualized condition. The data were acquired with a BioSemi ActiveTwo acquisition system of 128+8 channels at 1024 Hz. The number of trials varied among subjects. A more detailed description of the acquisition procedures and the number of trials for each subject can be found in [23].

**EEG processing.** In [23], the data were filtered between 0.5 and 100 Hz, and a notch filter was applied at 50 Hz. An Independent Component Analysis was applied in order to detect and remove noisy components, mainly contaminated with ocular and muscular artifacts. Finally, the continuous recording were split in 2.5 seconds trials. In this work, only the final two seconds of each trial were used to avoid possible evoked potentials produced by the stimulation protocol.

A similar approach to the Filter Bank Common Spatial Pattern (FBCSP) proposed in [2] was used for generating the spectral and spatial features. The band-pass filter frequencies used, in Hz, were: [0.5, 4.0], [4.0, 8.0], [8.0, 12.0], [12.0, 20.0], [20.0, 30.0] and [30.0, 45.0]. From each band, a Common Spatial Pattern filter was learned and the average power in the first six spatial components were calculated. These six features for each one of the six bands generated the 36-dimensional feature vector used for classification. Finally, each feature was scaled between 0 and 1. A twenty-fold cross-validation was used, splitting the data in 80% and 20% for training and testing, respectively, for each subject.

### 2.2 Classification algorithm

**Extreme Learning Machines.** Extreme Learning Machines (ELMs) are single hidden layer neural networks, originally proposed in [15,14,13]. ELMs have been widely used in EEG signals classification problems [6,5,19,7,29,18,27,17]. The training process of an ELM consists of two steps. First, the matrix of input weights  $W$  and the vector of bias weights  $b$  are randomly set as independent realizations, usually of a uniform distribution. The second step consists of finding an appropriate output weight  $\beta$ . This is done by means of the Moore-Penrose generalized inverse [9]. One of the most appealing aspects of the ELMs is that they only have one hyperparameter that must be calibrated: the number of hidden nodes  $M$ . In this work, we use the regularized version of ELM formulated in [4], setting the regularization parameter  $\lambda = 1$ . The regularized ELM does not suffer from overfitting, which commonly appears when the number of hidden nodes is close to the number of training examples.



**Fig. 1.** Violin plots for the first experiment. Test accuracy distribution obtained with real and random labels, for each subject. Statistical significance according to Mann-Whitney-Wilcoxon test is marked with “\*” ( $p \leq 0.001$ ).

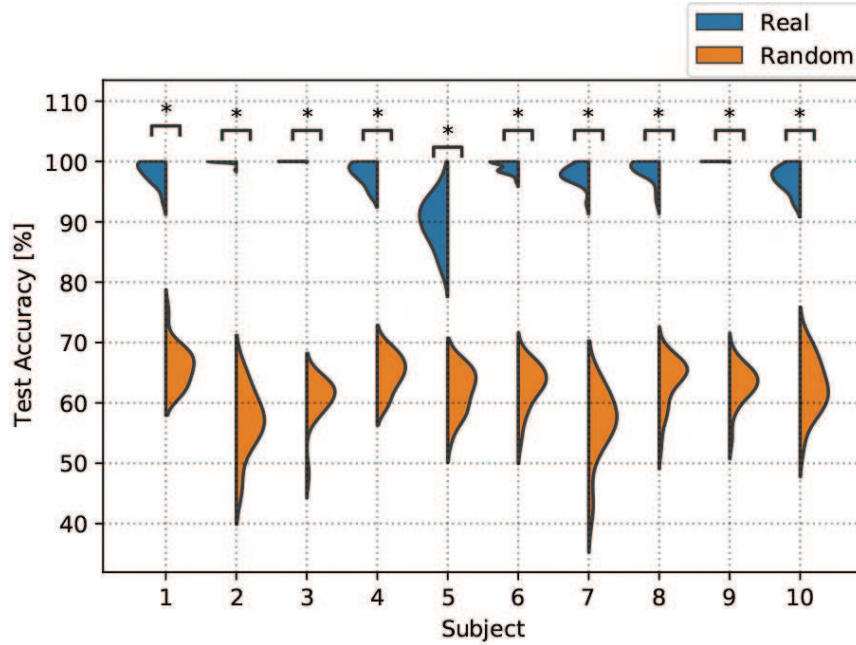
**Relevance-Based Pruning.** ELMs usually contain nodes that ought to be pruned out from the network in order to improve classification accuracy and to generate a more compact network. Here, the Relevance-Based Pruning (RBP) method proposed in [22] was used, as it gets rid of the need for computational expensive retraining while searching for the optimal number of neurons.

For finding the hyperparameter  $M$ , a three-fold cross-validation was performed within the training set, splitting the training data in 80% and 20% for train and validation subsets, respectively. In each fold, ten different random initializations of the parameter vectors  $W$  and  $b$  were generated. For each initialization in each validation fold, a grid search between 1000 and 50, with decremental steps of 50 nodes, was used to find the minimum number of nodes that maximizes the validation accuracy. Then,  $M$  was set as the average of the best number of hidden nodes obtained in each search. Once the appropriate number of nodes was obtained, the FBCSP and the ELM were trained over the whole training set and the testing accuracy was computed for each fold.

Finally, the same experiment was performed randomizing the condition label of the trials, allowing the comparison of the accuracy distribution obtained with the real and the randomized labels.

### 3 Results

In the first experiment, all the available trials in the inner speech and the visualized conditions were used, for each subject. The distribution of the test accuracy



**Fig. 2.** Violin plots for the second experiment. Test accuracy distribution obtained with real and random labels, for each subject. Statistical significance according to Mann–Whitney–Wilcoxon test is marked with “\*” ( $p \leq 0.001$ ).

with the real and the randomized labels for the 20 folds is shown in Figure 1. A Mann–Whitney–Wilcoxon two-sided test was performed between the accuracy obtained with each kind of label. The significance level was set to 0.001 and, for all subjects, significant differences were found.

For the second experiment, the inner speech and the pronounced speech trials were used. The distribution of the test accuracy with the real and the randomized labels for the 20 folds is shown in Figure 2. The same statistical test was performed and significant differences were found for all subjects. Moreover, the accuracy obtained in this experiment is consistently higher than the one obtained in the first experiment.

## 4 Conclusions

In an effort to substantiate the feasibility of an inner speech-based BCI, a comparison between conditions was made, showing encouraging results. This comparison allows us to state beyond any reasonable doubt that inner speech is clearly distinguishable from the other two examined conditions. Moreover, the brain mechanisms that generate each condition can be recognized by means of the EEG signal analysis. This can be thought of as a first milestone in the continuous working efforts for the construction of a more natural BCI. Needless to say, much further work has to be done to separate different classes within each condition. To encourage reproducible science, the code used in this work is publicly available at [https://github.com/N-Nieto/Feasibility\\_Analysis\\_Inner\\_Speech](https://github.com/N-Nieto/Feasibility_Analysis_Inner_Speech).

## References

1. Alderson-Day, B., Fernyhough, C.: Inner speech: development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin* **141**(5), 931 (2015)
2. Ang, K.K., Chin, Z.Y., Zhang, H., Guan, C.: Filter bank common spatial pattern (FBCSP) in brain-computer interface. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). pp. 2390–2397. IEEE (2008)
3. Deng, S., Srinivasan, R., Lappas, T., D’Zmura, M.: EEG classification of imagined syllable rhythm using Hilbert spectrum methods. *Journal of Neural Engineering* **7**(4), 046006 (2010)
4. Deng, W., Zheng, Q., Chen, L.: Regularized extreme learning machine. In: 2009 IEEE symposium on computational intelligence and data mining. pp. 389–395. IEEE (2009)
5. Ding, S., Guo, L., Hou, Y.: Extreme learning machine with kernel model based on deep learning. *Neural Computing and Applications* **28**(8), 1975–1984 (2017)
6. Ding, S., Zhang, N., Xu, X., Guo, L., Zhang, J.: Deep extreme learning machine and its application in EEG classification. *Mathematical Problems in Engineering* **2015** (2015)
7. Duan, L., Bao, M., Miao, J., Xu, Y., Chen, J.: Classification based on multilayer extreme learning machine for motor imagery task from EEG signals. *Procedia Computer Science* **88**, 176–184 (2016)
8. D’Zmura, M., Deng, S., Lappas, T., Thorpe, S., Srinivasan, R.: Toward EEG sensing of imagined speech. In: International Conference on Human-Computer Interaction. pp. 40–48. Springer (2009)
9. Engl, H., Hanke, M., Neubauer, A.: Regularization of inverse problems. *Mathematics and its Applications (Dordrecht)* **375** (1996)
10. Fiez, J.A., Petersen, S.E.: Neuroimaging studies of word reading. *Proceedings of the National Academy of Sciences* **95**(3), 914–921 (1998)
11. Hickok, G., Poeppel, D.: The cortical organization of speech processing. *Nature Reviews Neuroscience* **8**(5), 393–402 (2007)
12. Holz, E.M., Botrel, L., Kaufmann, T., Kübler, A.: Long-term independent brain-computer interface home use improves quality of life of a patient in the locked-in state: a case study. *Archives of Physical Medicine and Rehabilitation* **96**(3), S16–S26 (2015)
13. Huang, G.B., Wang, D.H., Lan, Y.: Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics* **2**(2), 107–122 (2011)
14. Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: theory and applications. *Neurocomputing* **70**(1-3), 489–501 (2006)
15. Huang, G.B., Zhu, Q.Y., Siew, C.K., et al.: Extreme learning machine: a new learning scheme of feedforward neural networks. *Neural Networks* **2**, 985–990 (2004)
16. Indefrey, P., Levelt, W.J.: The spatial and temporal signatures of word production components. *Cognition* **92**(1-2), 101–144 (2004)
17. Jin, Z., Zhou, G., Gao, D., Zhang, Y.: EEG classification using sparse Bayesian extreme learning machine for brain-computer interface. *Neural Computing and Applications* **32**(11), 6601–6609 (2020)
18. Kong, W., Guo, S., Long, Y., Peng, Y., Zeng, H., Zhang, X., Zhang, J.: Weighted extreme learning machine for P300 detection with application to brain computer interface. *Journal of Ambient Intelligence and Humanized Computing* pp. 1–11 (2018)

19. Liang, N.Y., Saratchandran, P., Huang, G.B., Sundararajan, N.: Classification of mental tasks from EEG signals using extreme learning machine. *International Journal of Neural Systems* **16**(01), 29–38 (2006)
20. McGuire, P., Silbersweig, D., Murray, R., David, A., Frackowiak, R., Frith, C.: Functional anatomy of inner speech and auditory verbal imagery. *Psychological Medicine* **26**(1), 29–38 (1996)
21. Nicolas-Alonso, L.F., Gomez-Gil, J.: Brain computer interfaces, a review. *Sensors* **12**(2), 1211–1279 (2012)
22. Nieto, N., Ibarrola, F., Peterson, V., Rufiner, H., Spies, R.: Extreme learning machine design for dealing with unrepresentative features. *arXiv preprint arXiv:1912.02154* (2019)
23. Nieto, N., Peterson, V., Rufiner, H.L., Kamienkoski, J., Spies, R.: “Thinking out loud”: an open-access EEG-based BCI dataset for inner speech recognition. *bioRxiv* (2021). <https://doi.org/10.1101/2021.04.19.440473>, <https://www.biorxiv.org/content/early/2021/04/20/2021.04.19.440473>
24. Pei, X., Barbour, D.L., Leuthardt, E.C., Schalk, G.: Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. *Journal of Neural Engineering* **8**(4), 046028 (2011)
25. Price, C.J.: The anatomy of language: contributions from functional neuroimaging. *The Journal of Anatomy* **197**(3), 335–359 (2000)
26. Suppes, P., Lu, Z.L., Han, B.: Brain wave recognition of words. *Proceedings of the National Academy of Sciences* **94**(26), 14965–14969 (1997)
27. Tan, P., Sa, W., Yu, L.: Applying extreme learning machine to classification of EEG BCI. In: 2016 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER). pp. 228–232 (2016). <https://doi.org/10.1109/CYBER.2016.7574827>
28. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain-computer interfaces for communication and control. *Clinical Neurophysiology* **113**(6), 767–791 (2002)
29. Zhang, Y., Wang, Y., Zhou, G., Jin, J., Wang, B., Wang, X., Cichocki, A.: Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces. *Expert Systems with Applications* **96**, 302–310 (2018)





# Anexo V: Transfer learning based on optimal transport for motor imagery brain-computer interfaces

# Transfer Learning based on Optimal Transport for Motor Imagery Brain-Computer Interfaces

Victoria Peterson, Nicolás Nieto, Dominik Wyser, Olivier Lambercy, Roger Gassert, Diego H. Milone and Rubén D. Spies

**Abstract—Objective:** This paper tackles the cross-sessions variability of electroencephalography-based brain-computer interfaces (BCIs) in order to avoid the lengthy recalibration step of the decoding method before every use. **Methods:** We develop a new approach of domain adaptation based on optimal transport to tackle brain signal variability between sessions of motor imagery BCIs. We propose a backward method where, unlike the original formulation, the data from a new session are transported to a calibration session, and thereby avoiding model retraining. Several domain adaptation approaches are evaluated and compared. We simulated two possible online scenarios: i) block-wise adaptation and ii) sample-wise adaptation. In this study, we collect a dataset of 10 subjects performing a hand motor imagery task in 2 sessions. A publicly available dataset is also used. **Results:** For the first scenario, results indicate that classifier retraining can be avoided by means of our backward formulation yielding to equivalent classification performance as compared to retraining solutions. In the second scenario, classification performance rises up to 90.23% overall accuracy when the label of the indicated mental task is used to learn the transport. Adaptive time is between 10 and 80 times faster than the other methods. **Conclusions:** The proposed method is able to mitigate the cross-session variability in motor imagery BCIs. **Significance:** The backward formulation is an efficient retraining-free approach built to avoid lengthy calibration times. Thus, the BCI can be actively used after just a few minutes of setup. This is important for practical applications such as BCI-based motor rehabilitation.

**Index Terms—**Brain-Computer Interfaces, Domain Adaptation, Motor Imagery, Optimal Transport, Transfer Learning.

Manuscript received MONTH DAY, YEAR; revised MONTH DAY, YEAR and MONTH DAY, 2020; accepted MONTH DAY, YEAR. Date of publication MONTH DAY, YEAR; date of current version MONTH DAY, YEAR. This work was supported in part by the Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, through PUE-IMAL No. 22920180100041-CO, Universidad Nacional del Litoral, UNL, through CAI+D-UNL 2016 PIC No. 50420150100036LI, and by the ETH Zurich Foundation in collaboration with Hocoma AG. (Corresponding author: V. Peterson)

V. Peterson and R. D. Spies are at the Instituto de Matemática Aplicada del Litoral, IMAL, UNL, CONICET, Santa Fe, Argentina (e-mail: vpeterson@santafe-conicet.gov.ar; rspies@santafe-conicet.gov.ar)

D. H. Milone and N. Nieto are at the Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(i), FICH-UNL, CONICET, Santa Fe, Argentina. (e-mail: dmilone@sinc.unl.edu.ar; nni-etc@sinc.unl.edu.ar)

D. Wyser, O. Lambercy and R. Gassert are at the Rehabilitation Engineering Laboratory, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland (e-mail: dominik.wyser@hest.ethz.ch; olivier.lambercy@hest.ethz.ch; roger.gassert@hest.ethz.ch).

## I. INTRODUCTION

Brain-computer interfaces (BCI) based on electroencephalography (EEG) can be used as a rehabilitation approach to improve the functional ability of people with severe sensorimotor impairments [1]–[4]. In this context, the brain activity associated to a mental motor task, or an actual motor attempt, must be decoded to control an external device, such as a virtual reality avatar providing visual feedback [5], a robotic device providing direct physical support [6], or the combination of both [7]. In order to implement and validate a robust BCI, two phases need to be completed: i) a *calibration* phase, in which the subject-specific decoding model is learned, mapping electrical brain activity patterns to output commands, and ii) a *testing* phase, in which the learned model is applied and the actual command is generated [8]. However, different sources of variability in the EEG signal affect the ability to robustly detect the intended mental task. Changes at the level of the subject, the electrode position or the type of feedback presented, are potential sources of variability. The lack of stationarity of the EEG can be observed between subjects but also within and between sessions of the same subject [9]. Thus, in view of real-life EEG-based BCI applications, the decoding algorithm must be able to deal with such variability in a non time-consuming manner.

Within the BCI community, a standard approach to tackle EEG variability consists of learning a new decoding algorithm before each and every use (recalibration). Although such an approach can lead to good classification performance (accuracy > 70%), it is not only time-consuming (generally a large amount of data must be recorded) but it also neglects all the information from previously collected data. Another strategy consists of finding the shared structure in the feature space across training data of multiple sessions and subjects [10].

The EEG's nonstationarity can be thought of as a data distribution drift between calibration and testing recordings [11]. Domain adaptation, a particular case of transfer learning (TL), provides a way of devising models that can cope with such data shift [12]. In the context of BCIs, domain adaptation aims at addressing the within or between-subjects variability by adapting features and/or classifier parameters from one domain, e.g. one subject or session, to another domain, e.g. another subject or session of the same subject.

The feature extraction method most widely used in rehabilitative BCIs is the common spatial pattern (CSP) algorithm,

first used to distinguish between the mental imagination (motor imagery, MI) of opening and closing the right vs. left hand [13]. In short, CSP spans band-pass filtered EEG data into a discriminative subspace by applying a linear mapping which maximizes the variance of one class and minimizes it for the other class [14]. Despite its popularity, CSP is sensitive to both data variations and small training datasets [15]. Different solutions have been proposed to tackle these drawbacks, ranging from adaptations at the level of the method itself [15], or changes in the classifier being used [16], to transformations of the feature space [9], [17].

In the specific context of neurorehabilitation applications, a subject-specific MI-BCI system is expected to be used across several sessions (days). The MI-decoding model will be subject to not only physiological noise, but potentially also to the brain pattern changes induced with recovery. In this context, in order to provide accurate feedback to the user as soon as the session begins, the decoding algorithm can be either *domain-invariant* or *domain-adaptive* to cross-sessions distribution drifts. Whereas in the former the decoding algorithm is designed so as to find common components between domains, in the latter the data distribution drift is learned so as to make the calibration and testing distributions more similar. This approach considers that samples (EEG trials) from two different domains (sessions) may have different data distributions, but have the same conditional distributions of the mental tasks with respect to the data [18].

Every domain-adaptive method seeks to “match” the probability distributions coming from two different domains: the *source* and the *target*. A recent approach within the BCI community for matching those distributions consists of domain alignments based on simple geometrical transformations of the data. In particular, the authors in [19] proposed what is called Riemannian alignment, a TL framework that consists of re-centering the covariance matrices of both domains in order to make them comparable. The Riemannian procrustes analysis (RPA) [11], which can be viewed as an improvement of the aforementioned TL method, also re-centers data as the first step of a series of geometrical transformations. More recently, the Euclidean alignment (EA) [20] was proposed with the main advantage of transforming the EEG trials in the Euclidean space, and thus any machine learning pipeline can then be applied to the aligned trials. Another relevant difference of EA as compared to RPA is that while the latter makes use of the available label information, the former is a completely unsupervised TL method. Regardless of how the alignment is made, all these methods transform data from both domains, and then a new classifier must be learned with the transformed training data.

Within the machine learning community attention has been paid to the use of optimal transport (OT) [21] as a TL approach. The use of optimal transport for domain adaptation (OTDA) was first proposed in [22], where the authors make use of OT theory to compute the optimal coupling between two probability distributions in a cost-effective manner. Once the optimal coupling is computed, the optimal transport is constructed in terms of it, and the data from the source domain are then transformed to make its probability distribution more

“similar” to the distribution in the target domain. A new classifier is then learned using the transported source data such that the label prediction in the target domain improves. The use of OTDA for transfer learning in BCI has already been evaluated in [23] for P300-BCI applications. Although the authors show that data distribution drift can potentially be addressed by keeping the classifier fixed and transforming the testing data, the transportation plan was learned by using testing data, an approach that is not applicable to online scenarios.

In this work, we provide a novel framework of domain adaptation based on OT for addressing the cross-session variability in online MI-BCI. Considering real-time applications of BCI systems for motor rehabilitation, we are interested in finding a solution to the problem of long calibration times and classifier retraining between sessions. In order to do so, we propose a new backward method, called backward optimal transport for domain adaptation (BOTDA), which transforms target samples to boost the performance of the already trained classifier. Unlike the approach followed in [23], the transport operator is directly calculated in an optimal way, without inverting the forward operator. By using BOTDA, both classifier retraining and long calibration data before each session can be avoided. In addition, a complete online compatible workflow for applying TL based on OTDA is presented here. This work is a contribution towards optimal strategies for the development of robust MI-BCI systems. It opens new avenues for increasing time-efficiency and performance of BCI-based neurorehabilitation approaches by better addressing EEG variability.

The organization of this article is as follows. Section II describes the problem, introduces OTDA as well as our backward approach and explains the use of OT in the context of BCI. Section III describes the two real multiple-sessions MI-BCI datasets used throughout this work. Experiments and results are presented in Section IV whereas discussions and conclusions, including future works, are in Sections V and VI, respectively.

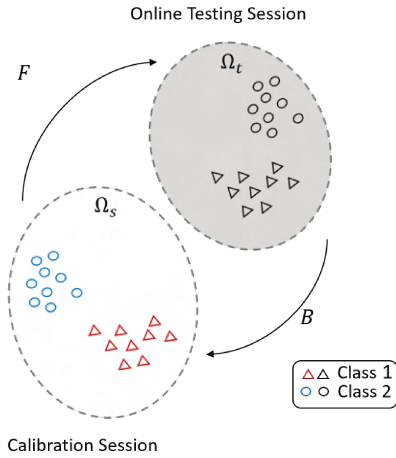
## II. TRANSFER LEARNING AND DOMAIN ADAPTATION BASED ON OPTIMAL TRANSPORT

This section introduces some of the mathematical assumptions and foundations of domain adaptation based on optimal transport. The traditional as well as the proposed backward formulation are detailed.

### A. Assumptions and main definitions

Throughout this work, we consider that a domain is a BCI session, and thus, the *source* domain  $\Omega_s$  refers to the calibration session, whereas the *target* domain  $\Omega_t$  refers to the testing session. We analyze different regularized OTDA versions proposed by Courty et al. [22] from two perspectives: i) the forward approach (original), in which a forward transport mapping  $F$  is learned from the source to the target domain and ii) the backward approach, our proposed new alternative, which learns a backward transport mapping  $B$  directly from the target to the source domain. Fig. 1 illustrates

the two aforementioned OTDA alternatives: i) forward OTDA (FOTDA) and ii) backward OTDA (BOTDA). For learning the transportation plan by either of the two OTDA alternatives, data from the new session is needed. With real-life applications in mind, we provide a domain adaptation algorithm based on optimal transport for MI-BCI that can be completely implemented online. In this context, we simulate two possible online scenarios: block-wise adaptation and sample-wise adaptation. In the first scenario we hypothesize that the data distribution drift remains unchanged from one data-block (testing run) to another, and therefore it can be learned from the previous available data. On the contrary, in the second scenario, we assume that the data distribution is continuously changing and thus the current testing trial should be considered in the transport learning process. Although experiments were conducted as online simulations, throughout this work we shall indistinguishably refer to them as online testing session since this is how our workflow was designed. That is, all the proposed methods can be applied in fully online scenarios.



**Fig. 1:** Illustration of two different domain adaptation approaches: forward OTDA, where the distribution drift is learned from *source* to *target* domain and backward OTDA, where the transportation mapping is learned from *target* to *source* domain. Here  $\Omega_s$  and  $\Omega_t$  denote the source and target domains, respectively, and  $F$  and  $B$  denote the forward and backward transport mappings, respectively.

## B. Problem formulation

Consider two datasets, the *source* ( $\mathcal{S}$ ) and the *target* ( $\mathcal{T}$ ) dataset, coming from two different MI-BCI sessions (e.g., days). Each dataset is composed of EEG trials belonging to one and only one of two MI classes. Let  $N_s$  and  $N_t$  denote the number of trials in the source and target domains, respectively, and  $\mathbf{x}_i^s \in \Omega_s, \mathbf{x}_i^t \in \Omega_t$  the feature vectors, where  $\Omega_s, \Omega_t \subset \mathbb{R}^d$ ,  $\mathcal{K} \doteq \{k_1, k_2\}$  and  $y_i^s, y_i^t \in \mathcal{K}$  are class labels. Hence our datasets are

$$\begin{aligned} \mathcal{S} &\doteq \{(\mathbf{x}_i^s, y_i^s), i = 1, \dots, N_s\} \subset \Omega_s \times \mathcal{K}, \\ \mathcal{T} &\doteq \{(\mathbf{x}_i^t, y_i^t), i = 1, \dots, N_t\} \subset \Omega_t \times \mathcal{K}. \end{aligned}$$

We shall denote with  $X_s \in \mathbb{R}^{N_s \times d}$  and  $X_t \in \mathbb{R}^{N_t \times d}$  the feature matrices whose  $i^{\text{th}}$  rows are  $\mathbf{x}_i^s$  and  $\mathbf{x}_i^t$ , respectively.

In every standard learning paradigm, the discriminative model is constructed using a training set (with label information) and then evaluated on the unseen testing set. This learning-evaluation framework assumes that training and testing sets are drawn from the same distribution. However, in multiple-sessions EEG-based BCIs this assumption cannot be guaranteed since training and testing sets are collected at different times by different system conditions, which may lead to poor classification performances. This issue can be modeled as a covariate shift problem, in which changes in the distributions in the two domains are considered ( $\mathbf{P}_s(\mathbf{x}^s) \neq \mathbf{P}_t(\mathbf{x}^t)$ ), but it is assumed that the conditional distributions of the labels with respect to the data remain the same ( $\mathbf{P}_s(y|\mathbf{x}^s) = \mathbf{P}_t(y|\mathbf{x}^t)$ ). Assuming that the domain drift is given by a certain transformation  $F : \Omega_s \rightarrow \Omega_t$ , if such a transformation is known, one can then “adapt” the domains so as to make the distributions similar, and thus prevent classification failure. The domain adaptation approach based on optimal transport (OTDA) consists of estimating such a transformation  $F$  in a cost-effective manner [22], as described in the following subsection.

## C. A brief introduction to OTDA

OT theory studies a problem known as the Monge-Kantorovich transportation problem [21], which, roughly speaking, seeks to find a cost-effective way to transport mass between two probability distributions. In this direction, it is said that OT solves and optimization problem which minimizes what is called transportation cost. For  $\Omega \subset \mathbb{R}^d$  let  $\mathcal{P}(\Omega)$  denote the space of all probability measures with support in  $\Omega$ . Given  $\Omega_s, \Omega_t \subset \mathbb{R}^d$ , a measurable mapping  $F : \Omega_s \rightarrow \Omega_t$ , as above, and a measure  $\alpha \in \mathcal{P}(\Omega_s)$ , the measure  $\sigma \in \mathcal{P}(\Omega_t)$  defined by  $\sigma(A) \doteq \alpha(F^{-1}(A))$  for every  $\sigma$ -measurable set  $A \subset \Omega_t$ , is denoted by  $F\#\alpha$ . The mapping  $F$  is said to be a push-forward or a transport map of  $\alpha$  in  $\sigma$ , and  $\sigma$  is referred to as the *push-forward* of  $\alpha$  by  $F$ .

In this context, addressing the data distribution drift by the forward OTDA problem consists of finding a transport map  $F$  of the source data distribution  $\mu_s \in \mathcal{P}(\Omega_s)$  in the target data distribution  $\mu_t \in \mathcal{P}(\Omega_t)$ , i.e. such that  $F\#\mu_s = \mu_t$ . This adaptation problem can be tackled by following the next three steps: first, estimate the measures  $\mu_s$  and  $\mu_t$  from the feature matrices  $X_s$  and  $X_t$ , then find a transport map  $F$  from  $\mu_s$  in  $\mu_t$ , and finally use  $F$  to transport  $X_s$ . Although that sounds simple, searching for  $F$  in the space of all possible transformations turns out to be an absolutely unmanageable problem and appropriate restrictions on  $F$  must be imposed. In this line of work, the Monge approach [24] to the optimal transport problem consists of finding, among all the possible transports  $F$  from  $\mu_s$  in  $\mu_t$ , an optimal transformation  $F_0$  as the solution of the minimization problem

$$F_0 \doteq \underset{F \text{ s.t. } F\#\mu_s = \mu_t}{\operatorname{argmin}} \int_{\Omega_s} c(\mathbf{x}, F(\mathbf{x})) d\mu_s(\mathbf{x}),$$

where  $c : \Omega_s \times \Omega_t \rightarrow \mathbb{R}^+$  is a given cost function.

A convex relaxation of the OT problem was introduced by Kantorovich et al. [24]. Given  $\mu_s \in \mathcal{P}(\Omega_s)$  and  $\mu_t \in \mathcal{P}(\Omega_t)$ , let  $\Upsilon = \Upsilon(\mu_s, \mu_t)$  be the set of all couplings between  $\mu_s$  and  $\mu_t$ , i.e. all joint probability measures  $\gamma$  in  $\mathcal{P}(\Omega_s \times \Omega_t)$  whose respective marginals are  $\mu_s$  and  $\mu_t$ . The Kantorovich approach to OT consists of finding the optimal transportation plan, which is defined as the coupling  $\gamma_0$  given by

$$\gamma_0 \doteq \operatorname{argmin}_{\gamma \in \Upsilon} \int_{\Omega_s \times \Omega_t} c(\mathbf{x}^s, \mathbf{x}^t) d\gamma(\mathbf{x}^s, \mathbf{x}^t). \quad (1)$$

When the measures  $\mu_s$  and  $\mu_t$  are to be estimated from the feature matrices  $X_s$  and  $X_t$  respectively, as in our case, the corresponding sample distributions take the form

$$m_s = \sum_{i=1}^{N_s} p_i^s \delta_{\mathbf{x}_i^s} \quad \text{and} \quad m_t = \sum_{j=1}^{N_t} p_j^t \delta_{\mathbf{x}_j^t}, \quad (2)$$

where  $p_i^s$  and  $p_j^t$  are the probability masses at the points  $\mathbf{x}_i^s$  and  $\mathbf{x}_j^t$ , respectively, and  $\delta_x \in \mathcal{P}(\Omega)$  denotes the unit Dirac delta measure at the point  $\mathbf{x} \in \Omega$ . Usually  $p_i^s = \frac{1}{N_s} \forall i$  and  $p_j^t = \frac{1}{N_t} \forall j$ . The continuous Kantorovich formulation of OT (1) has an immediate correlate in this discrete case. In fact, if we now denote by  $A_s \doteq \Pi_1 \mathcal{S}$  and  $A_t \doteq \Pi_1 \mathcal{T}$  the projections of the datasets  $\mathcal{S}$  and  $\mathcal{T}$  into  $\Omega_s$  and  $\Omega_t$ , respectively, and with  $\Gamma$  the set of all discrete probabilistic couplings between the discrete measures  $m_s$  and  $m_t$ , i.e.

$$\Gamma = \left\{ \gamma = (\gamma_{ij}) \in \mathcal{P}(A_s \times A_t) \text{ s.t. } \sum_j \gamma_{ij} = m_s, \sum_i \gamma_{ij} = m_t \right\},$$

then the discrete Kantorovich formulation of OT boils down to finding the optimal discrete transportation plan, defined as

$$\gamma_0 \doteq \operatorname{argmin}_{\gamma \in \Gamma} \langle \gamma, C \rangle_F, \quad (3)$$

where  $\langle \cdot, \cdot \rangle_F$  denotes the Frobenius inner product, and  $C = (c_{ij})$ ,  $c_{ij} \geq 0$  is a cost function matrix with  $c_{ij} = c(\mathbf{x}_i^s, \mathbf{x}_j^t)$  representing the cost of moving a unit probability mass from  $\mathbf{x}_i^s \in \Omega_s$  to  $\mathbf{x}_j^t \in \Omega_t$ . Quite often  $c$  is simply the square  $L_2$  norm or the Euclidean distance.

Several regularized versions of the discrete OT problem (3) exist. In [25], the authors proposed to add a regularizer to the formulation in (3) in order to reduce the sparsity of the transportation plan. The optimal coupling was then defined as:

$$\gamma_0 \doteq \operatorname{argmin}_{\gamma \in \Gamma} \langle \gamma, C \rangle_F + \lambda W_e(\gamma), \quad (4)$$

where  $\lambda$  is a positive constant called regularization parameter, and  $W_e(\gamma) \doteq \sum_{ij} \gamma_{ij} \log(\gamma_{ij})$  is the negentropy of  $\gamma$ . This formulation of OT not only favors smoother versions of the transport by reducing its sparsity, but it also allows for the use of computationally efficient algorithms based on the Sinkhorn-Knopp's scaling matrix approach [26]. In the sequel, we shall refer to this regularized OT version as OT-S.

In addition, to take advantage of the label information available in the source domain, an extra penalizer term can be added to (4) in the following way:

$$\gamma_0 \doteq \operatorname{argmin}_{\gamma \in \Gamma} \langle \gamma, C \rangle_F + \lambda W_e(\gamma) + \eta W_c(\gamma), \quad (5)$$

where  $\eta > 0$  is also a regularization parameter and  $W_c(\gamma)$  is a regularization term which introduces label information into the OT formulation. For instance, the group-lasso regularizer [27] on the columns of  $\gamma$  can be used to penalize couplings that take any two samples in the source domain having different labels to the same sample in the target domain, as follows:

$$W_c(\gamma) \doteq \sum_j \sum_k \|\gamma(\mathcal{I}_k, j)\|_2,$$

where  $\mathcal{I}_k$  denotes the set of indices of all the rows of  $\gamma$  corresponding to all the source domain samples of class  $k \in \mathcal{K}$  [22].

For domain adaptation, once the optimal transportation plan  $\gamma_0$  is found, source samples have to be transported to the target domain. For that, we need to define a transport map  $F$  in terms of  $\gamma_0$ . This mapping can be conveniently expressed for each  $\mathbf{x}_i^s$  as the following  $c$ -based barycentric mapping:

$$\hat{\mathbf{x}}_i^s = F_{\gamma_0}(\mathbf{x}_i^s) \doteq \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^d} \sum_{j=1}^{N_t} \gamma_0(i, j) c(\mathbf{x}, \mathbf{x}_j^t), \quad (6)$$

for all  $i = 1, \dots, N_s$ .

In the discrete, regularized with class-labeled information formulation of OT (5), to which we shall refer in the sequel as OT-GL, it is expected that each target sample receives masses from source samples in the source domain that have the same label. For domain adaptation it is thus expected that the prior distributions of the labels are preserved in both domains. Fortunately, this assumption is not a problem for the current MI-BCI application considered throughout this work, as it will be explained in Subsection II-E.

#### D. Backward OTDA

The forward OTDA maps source samples into the target domain by the transport map  $F_{\gamma_0}$  and then a new classifier is trained with those transformed source samples. If classifier retraining is to be avoided, rather than transforming the source samples, target samples can be transported to fit the source data distribution. However, we can not guarantee that the transport map  $F_{\gamma_0}$  defined in (6) be invertible. Although clearly a transport map can be defined in terms of  $\gamma_0^T$  as  $\hat{\mathbf{x}}_i^t = F_{\gamma_0^T}(\mathbf{x}_i^t)$ , it might happen that  $\gamma_0^T$  as a coupling between  $m_t$  and  $m_s$  is not optimal. With this in mind, we provide an alternative way to learn the mapping from the target to the source domain, in which the inversion of the  $F$  operator is avoided, ensuring that the learned mapping is optimal.

Given that in BCI applications it is common to acquire a set of data at the beginning of each new session for updating the model (recalibration), the transportation plan can be learned from the target to the source domains (see Fig. 1). This novel approach addresses the data drift from the target domain (new session) into the source domain (old session), reason for which we shall refer to it as backward OTDA (BOTDA). With  $\zeta_0$  and  $B_{\zeta_0}$  we will denote the optimal transport plan as obtained by (4) or (5) when the backward approach is used, and the corresponding transport map, respectively. Hence, target



samples can be transformed as follows:

$$\hat{\mathbf{x}}_i^t = B_{\zeta_0}(\mathbf{x}_i^t) \doteq \underset{\mathbf{x} \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{j=1}^{N_s} \zeta_0(i, j) c(\mathbf{x}, \mathbf{x}_j^s), \quad (7)$$

for all  $i = 1, \dots, N_t$ .

In summary, BOTDA learns a mapping from the target to the source domain in a cost-effective manner, and the classifier (already trained with the source data) is tested with the transformed target data. Since the implementation of BOTDA relies on the same optimizer as FOTDA [25], [28], from the algorithmic point of view the novelty relies on how to build the source and the target datasets.

### E. OTDA applied to BCI

In the present work it will always be assumed that the discriminative model is learned using data from the source domain (calibration session) and then evaluated in the target domain, representing a new BCI session. In spite of the OTDA approach that we investigate within this paper, in order to learn the transportation plan, data from the target domain are always needed. In this work, we shall refer as *transportation set*,  $\mathcal{V} = \{(\mathbf{x}_i^v, y_i^v)\}_{i=1}^{N_v} \subset \mathcal{T}$ ,  $N_v \ll N_t$ , to such portion of data coming from the target domain which is used to learn the transportation plan. With  $X_v \in \mathbb{R}^{N_v \times d}$  we shall denote the transportation feature matrix whose  $i^{\text{th}}$  row is  $\mathbf{x}_i^v$ .

Unlike other classification problems in which target data are fully available, for online brain signal detection the target data become available one trial at a time. This is undoubtedly a restriction to the way the data distribution drift can be learned. With this in mind, we simulate two possible online scenarios to learn the mapping: block-wise adaptation and sample-wise adaptation. In the first scenario we assume that the new BCI session is divided into data blocks of  $n_t$  trials each, which we call runs. With the aim of keeping the recalibration time as short as possible while having enough data to learn the mapping, we set  $n_t = 20$ , which comprises around 10 trials per class<sup>1</sup>. In addition, we hypothesize that the distribution drift of a testing run is equal to the distribution drift of the previous runs. We also assume here that the first session is meant for calibration whereas the first run of the new session (target domain) is meant for recalibration purposes,  $R_0 = \{(\mathbf{x}_i^t, y_i^t)\}_{i=1}^{n_t}$ . The remaining  $n_r$  runs are considered as online testing runs ( $R_1, \dots, R_{n_r}$ ). In this context, the prior available data in the target domain must be wisely used to learn the OT mapping. In this direction, for every OTDA alternative, we decide to learn the mapping using all the available testing data up to the current  $r^{\text{th}}$  run, as follows:

$$\mathcal{V}_r \doteq R_0 \cup \bigcup_{j=1}^{r-1} R_j, \quad r = 1, 2, \dots, n_r. \quad (8)$$

Now, how can we make sample-wise learning of EEG distribution drift? In rehabilitative BCIs, a synchronous paradigm is used for both the calibration and the testing-with-feedback

phases. In a synchronous BCI, generally by means of a visual cue, the system indicates the user when to start making a certain mental task [29], and thus the feedback provided to the user in the evaluation phase is based on what he/she did and can be compared to what he/she should have done. Given that in such cue-based environments the indicated mental task of a trial is always known, we provide a way of continuously adapting the data by considering the current  $i^{\text{th}}$  trial as part of the transportation set, i.e. we define:

$$\mathcal{V}_i \doteq R_0 \cup \bigcup_{j=n_t+1}^i \{(\mathbf{x}_j^t, y_j^t)\} \subset \mathcal{T}, \quad \forall i = n_t+1, \dots, N_t \quad (9)$$

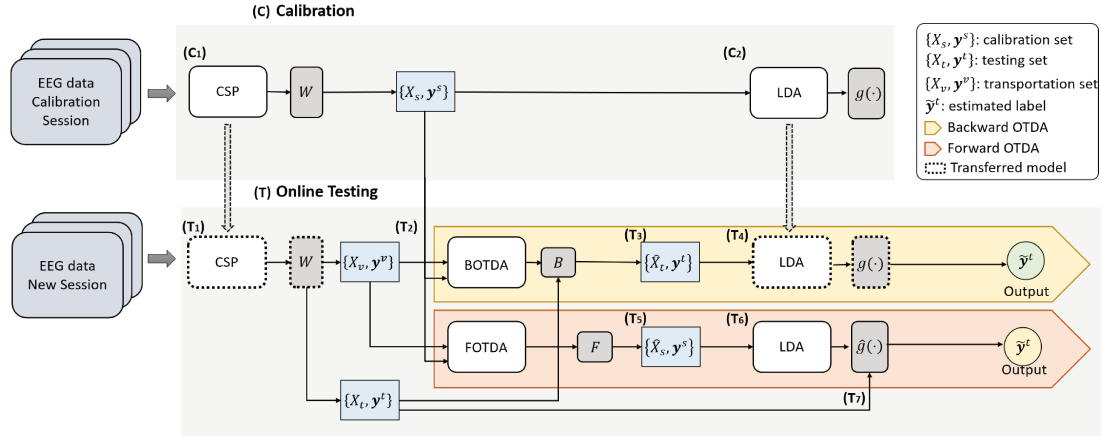
It is important to highlight here that the current label information included in  $\mathcal{V}_i$  is only used in the BOTDA with group-lasso penalty.

For MI-BCI, a simple but effective discriminative model can be built based on CSP and a linear classifier, such as linear discriminative analysis (LDA) [30]. Within this standard setup, applying OTDA involves four main steps (calibration, mapping learning, data transformation and classifier evaluation), as shown in Fig. 2. In the calibration phase (**C**), the calibration data from the source domain are used for learning the CSP features (**C**<sub>1</sub>) and an initial LDA classifier (**C**<sub>2</sub>). In the online testing session (**T**), the trained CSP model (represented here by the matrix  $W$ ) is applied at the target domain (**T**<sub>1</sub>) in both the transportation and the testing sets. Both OTDA mappings (forward and backward) are learned using the transportation and the calibration sets (**T**<sub>2</sub>). While in the BOTDA alternative (light yellow arrow) testing features are transformed (**T**<sub>3</sub>) to feed the already trained classifier (**T**<sub>4</sub>), in the case of FOTDA (light orange arrow) the transformation is applied on the calibration feature set (**T**<sub>5</sub>) and then a new classifier  $\hat{g}(\cdot)$  is trained using the transformed data matrix  $\hat{X}_s$  (**T**<sub>6</sub>). This classifier is evaluated with the original testing feature set  $X_t$  (**T**<sub>7</sub>). Note that the label information will only be used for learning the mapping when the OT-GL formulation is used in either of the two OTDA alternatives, as explained below.

In this work four OTDA alternatives are tested. Although for each method the training or testing features can be different, the discriminative model is always CSP+LDA. We compare the classification performances of each OTDA approach against: a BCI model without transfer learning, a recalibration procedure already used in rehabilitative BCI [31], and two state-of-the-art data alignment methods. Subsequently, all investigated processing methods are summarized:

- **Standard with calibration (SC)**: train the model with source data (calibration) and do the classification on the target dataset (new session) with no transformation whatsoever.
- **Standard with recalibration (SR)**: employ model re-training (CSP+LDA) by updating the training data with the previous testing run and eliminating the oldest one, as proposed by Ang et al. [31]. Note that this strategy is made on the raw data space, before feature extraction, and thus a new CSP model is constructed at each data updating step.
- **Riemannian procrustes analysis (RPA)**: the sample

<sup>1</sup>Larger  $n_t$  values were tested, namely  $n_t$  equals to 40 and 80, yielding to similar overall classification results as when  $n_t = 20$  but with more associated computational time.



**Fig. 2:** Schematic representation of the OTDA pipeline from calibration (C) to online testing (T) when either the backward (light yellow arrow) or the forward approach (light orange arrow) is being used. Note that: already trained models from Calibration Session to the New Session are drawn with dashed lines; methods are depicted in white boxes; the outputs of each model are shown in gray boxes; and feature sets are drawn in light blue rectangles. Since the outputs (predicted class) of the forward and the backward alternative may not be the same, two different colors were used to show the output label in each path.

covariance matrices are geometrically transformed (translated, scaled and rotated) in a semi-supervised way [11]. The approach is performed in the Riemannian manifold. After the source and the target data are transformed, the Minimum-Distance to Mean (MDM) classifier [32] is trained and tested on the transformed target dataset.

- **Euclidean alignment (EA):** performs an unsupervised data alignment using the arithmetic mean of all covariance matrices to transform the data in the Euclidean space [20]. Like RPA, both source and target data are adapted. The decoding model is then learned in the transformed source samples and applied to the transformed target dataset.
- **FOTDA-S:** employ Sinkhorn OTDA formulation (4) to learn the mapping between the *source* and *target* domains. Source samples are then transformed to train a new classifier in which target samples are tested.
- **FOTDA-GL:** employ group-lasso OTDA formulation (5) to learn the mapping between the *source* and *target* domains. Source samples are then transformed to train a new classifier in which target samples are tested.
- **BOTDA-S:** employ Sinkhorn OTDA formulation (4) in the backward approach so as to learn the mapping from the *target* to the *source* domains. Target samples are transformed to feed the already trained classifier with the source data.
- **BOTDA-GL:** employ group-lasso OTDA formulation (5) in the backward approach so as to learn the mapping from *target* to *source* domains. Target samples are transformed to feed the already trained classifier with the source data.

### III. DATABASES

Two different EEG-based BCI datasets were used in this study, each comprising at least two MI-BCI sessions of multiple participants. Since we aimed at providing an adaptive

strategy for online testing in a motor rehabilitation scenario, where typically only binary decisions are required (e.g., to trigger an external device), both datasets comprised two MI classes.

**Dataset-1:** this dataset was collected from 10 naive able-bodied BCI users (3 females, 4 left-handed, mean age  $\pm$  SD =  $25.45 \pm 2.50$  years) on two different days (sessions) with a session-to-session separation of 5 days maximum. The experiment was approved by the local ethics committee (BASEC-Nr. Req-2017-00631, Cantonal Ethics Commission, Zurich, Switzerland). A portable 64-channel EEG system (eegort Ant Neuro, Netherlands) was used for brain signal recording, with a sampling frequency of 512 Hz. Surface electrodes were placed in accordance with the international 10-20 system, using CPz as reference and AFz as ground electrodes. EEG signals were band-pass filtered between 0.5 Hz and 40 Hz. Two mental tasks were performed by the subjects: i) the kinesthetic imagination of movement of their dominant hand (grasping movement) and ii) a rest/relax condition. Each session was composed of four runs separated by short breaks. Each run consisted of 40 trials (20 for each condition), yielding a total of 160 trials at the end of each session. No feedback was provided to the subject during the sessions. For more information refer to [33]. For the experiments detailed below, we downsampled the EEG signals to 128 Hz, extracted EEG segments from 0.5 to 2.5 s after the onset of the visual cue, and selected 28 electrodes covering the sensorimotor areas, in accordance to [34].

**Dataset-2:** known in the literature as BNCI2015001. It is a publicly available dataset<sup>2</sup> by Faller et al. [35]. The EEG data were obtained from 12 able-bodied BCI-naive volunteers (5 female, age  $24.8 \pm 3.0$  years) which participated in at least two MI-BCI sessions (maximum day time frame was 5 days).

<sup>2</sup>Available at <http://bnci-horizon-2020.eu/database/data-sets>

The EEG recordings were acquired with the g.GAMMAsys active electrode system (electrode positions at FC3, FCz, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CPz and CP4) along with a g.USBamp amplifier (g.tec, Guger Technologies OEG, Graz, Austria) at a sampling frequency of 512 Hz. During acquisition, the signals were bandpass filtered between 0.5 and 100 Hz with an additional notch filter at 50 Hz. In each session, the participants performed 5 runs of 40 trials (i.e. 200 trials) of hand MI vs. feet MI. Only the first run of each session was without providing visual feedback. As before, the EEG signals were windowed from 0.5 to 2.5 s after the onset of the visual cue. Signals were band-pass filtered between 0.5 and 40 Hz, and then downsampled to 128 Hz. Although for some participants three BCI sessions were acquired, for fair comparison purposes, only the first two sessions were used for the experiments.

#### IV. EXPERIMENTS AND RESULTS

In order to keep the focus on the transfer learning rather than on the classifier, we used the traditional CSP and LDA frameworks with diagonal loading [15]. Thus, after filtering each EEG trial between 8 and 30 Hz, six spatial filters were used for feature learning with CSP. A subject-specific CSP+LDA model using calibration data was built to analyze the impact of OTDA. Each OTDA alternative was applied at the CSP feature level space, as shown in Fig. 2. Two different online scenarios were tested: i) block-wise adaptation and ii) sample-wise (continuous) adaptation, as described in subsections IV-A and IV-B, respectively. The simulations as well as the corresponding source codes implemented in this work are publicly available<sup>3</sup>. We used the POT library [28] for learning the regularized discrete transportation plans, the MNE library [36] for implementing CSP and filtering the EEG data, the Scikit-learn library [37] for learning the linear classifier and the open Python code of RPA [11].

##### A. Data drift learned from previous data

As explained in Section II-E, for this part of the study the new BCI session was divided into runs of 20 trials each. Given the number of trials of each dataset, we ended up having 7 testing runs for Dataset-1 and 9 for Dataset-2. We tested the four OTDA alternatives (FOTDA-S, FOTDA-GL, BOTDA-S and BOTDA-GL) against the two not domain-adaptive methods (SC and SR) as well as the two data alignment methods (RPA and EA). For the OTDA alternatives, to learn the transportation plan at each  $r^{th}$  testing run, we used the transportation set  $\mathcal{V}_r$  as defined in (8). In order to boost the transport learning process, we selected a small number of source samples to learn the mapping. Since it has been shown that the proportions of samples between the source and the target distributions influence the transport [38], this source subset consists of  $M = N_v$  trials randomly selected from the source dataset. This process was repeated 20 times, and the best subset of samples was selected based on classification accuracy. Within

this validation procedure, the regularization parameters were also selected by means of a grid search process ( $\lambda, \eta \in \{0.1, 0.5, 1, 2, 5, 10, 20\}$ ). In order to statistically analyze the differences between the performances yielded by each method, the non-parametric Friedman test and the post-hoc Nemenyi test at level of significance  $\alpha = 0.05$  were applied.

For Dataset-1, the overall accuracy across subjects is shown in Table I. Columns correspond to runs, organized in increasing order, while the last column shows the average performance across runs. Rows are organized in three subgroups: standard, data alignment and OTDA methods. It can be seen that all domain adaptive methods despite EA achieved similar classification performance between each other ( $p$ -value  $> 0.11$ ), whereas all transfer learning method are significantly different than the non-adaptive one ( $p$ -value  $< 0.02$ ). Significant differences were found between EA and all OTDA alternatives ( $p$ -value  $< 0.01$ ). For the first testing run, accuracy improvements of 9.5%, 5.5% and 6.5% as compared to SC, SR and RPA, respectively, can be achieved by means of the free re-training model based on BOTDA. A similar tendency between the full retraining model (SR) and the different domain-adaptive alternatives is observed in most of the testing runs.

Table II shows the classification results achieved on average across subjects for Dataset-2. It is organized as Table I, where rows and columns separate methods and testing runs, respectively. As before, there is not a clear winner between the OTDA alternatives. In addition, note BOTDA-S present competitive classification performance as compared to the RPA and EA method along the different testing runs ( $p$ -value  $> 0.07$ ). Detailed plots of the results as well as the statistics made can be found in the supplementary material.

**TABLE I:** Overall classification results (accuracy in %) yielded by each tested method for Dataset-1 in the block-wise adaptation scenario. Columns correspond to testing runs. Last column aggregates the average across runs.

	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	R <sub>7</sub>	av
SC	64.0	65.0	66.5	70.5	71.0	70.0	72.0	68.4
SR	68.0	69.0	71.5	81.0	74.0	75.5	71.5	72.9
RPA	67.5	70.5	71.5	75.5	75.0	68.0	76.5	72.14
EA	77.0	73.5	71.5	75.5	73.0	74.0	70.5	73.6
FOTDA-S	73.5	68.5	70.0	72.5	73.5	71.0	70.5	71.4
FOTDA-GL	71.5	68.5	70.5	73.5	72.5	70.5	70.5	71.1
BOTDA-S	73.0	71.5	71.0	73.0	70.5	72.5	69.0	71.5
BOTDA-GL	73.5	66.5	71.5	74.5	71.5	70.5	69.5	71.1

##### B. Data drift learned from current and prior data

In the previous experiment we considered that the estimated distribution drift of prior runs was similar to the data variation of a posterior set of trials. Although this assumption allows to learn the mapping for online data adaptation, it may not always be true given the high nonstationarity of the EEG signals, and thus continuous drift learning should be performed.

Since for learning the transportation plan a set of trials from the target domain is always needed, as before, we used the first 20 trials of the new session for recalibration purposes.

<sup>3</sup><https://github.com/vpeterson/otda-mibci>. The provided examples were run over Subject S9 of Database-1.



**TABLE II:** Overall classification results (accuracy in %) yielded by each tested method for Dataset-2 in the block-wise adaptation scenario. Columns correspond to testing runs. Last column aggregates the average across runs.

	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	R <sub>7</sub>	R <sub>8</sub>	R <sub>9</sub>	av
<b>SC</b>	77.9	78.8	74.6	72.5	74.6	75.8	74.6	77.5	74.6	75.6
<b>SR</b>	80.8	81.7	78.8	80.0	79.2	79.6	79.6	85.0	78.8	80.4
<b>RPA</b>	77.5	76.3	72.9	80.8	77.5	79.6	80.0	83.3	79.2	78.6
<b>EA</b>	77.1	75.8	83.3	82.9	78.8	78.8	79.6	79.2	80.0	79.9
<b>FOTDA-S</b>	74.2	80.2	78.3	78.7	75.4	77.5	77.5	77.5	76.7	77.4
<b>FOTDA-GL</b>	77.1	79.6	79.6	77.9	76.25	78.8	78.8	75.8	76.7	77.8
<b>BOTDA-S</b>	78.3	76.3	78.8	77.9	75.4	76.7	80.0	76.7	78.8	77.6
<b>BOTDA-GL</b>	73.8	77.1	77.5	76.7	75.0	76.3	77.9	75.8	75.4	76.2

For every testing trial, we learned the mapping by using not only all prior available data, but also the current trial, as defined in (9). Following the same methodology described in Subsection IV-A, the transportation plan was learned by using the corresponding transportation set  $\mathcal{V}_i$  and a small subset of  $M$  trials of the source dataset. Considering the online time processing restrictions, this small subset of  $M$  trials was selected following the same selection process explained in Subsection IV-A but using only  $R_0$ , thus  $M = 20$ . Here we simulated a continuous online session. At the end of this experiment, the classification accuracy, per each subject and processing method, was evaluated based upon the label assigned to each trial during the testing phase. For completeness we also evaluated the performance of the calibrated classifier with no adaptation (SC), the standard with recalibration (SR) method and the two data-alignment methods (RPA and EA). Note that the information of the indicated mental task is actually used only by SR, RPA and BOTDA-GL. Statistical analyses were performed also here by means the Friedman test and the post-hoc Nemenyi test at level of significance  $\alpha = 0.05$ .

Tables III and IV show the accuracy reached in the testing phase by each tested method for each subject of Dataset-1 and Dataset-2, respectively. In both tables the average accuracy across subjects (last row denoted as ‘av’) shows that BOTDA-GL performs better ( $p$ -value  $< 0.05$ ) than the rest of the OTDA alternatives as well as the SC method, reaching mean accuracy levels of 88.85% and 90.23% for Dataset-1 and Dataset-2, respectively. These values correspond to relative improvements of 20% and 14% with respect to the method without adaptation (SC) for each dataset, respectively. Although, no significant differences were found between BOTDA-GL and SR, RPA and EA, for most of the cases BOTDA-GL achieved the highest accuracy values (7/10 for Dataset-1 and 9/12 for Dataset-2). The rest of the OTDA alternatives achieve similar classification results as the standard calibration method (SC).

It is also interesting to compare the computational cost associated to running each adaptive method. We ran our experiments on an Intel<sup>®</sup> Core<sup>™</sup> i7-6700K CPU @ 4.00 GHz  $\times$  8 with 56 GB of RAM. For the sake of comparison, the optimal hyper-parameters search was not included for these time measurements. Therefore, Table V shows the computational time of adapting each new trial in this sample-wise scenario

for the SR, RPA, EA and BOTDA-GL methods. For Dataset-1, our proposed BOTDA-GL method is 9.6, 79.9 and 18.7 times faster than the SR, RPA and EA methods, respectively. Those values are 11.65, 38.04 and 18.6 for Dataset-2. Although here we only show the computational time of the BOTDA-GL method, it is worth mentioning that all the other OTDA alternatives result in similar computational times.

## V. DISCUSSIONS

In this work we investigated the use of optimal transport as a transfer learning approach for addressing EEG variability between sessions of MI-BCI, aiming at avoiding classifier re-training. We have provided a complete framework for applying OTDA, taking into account the constraints of rehabilitative MI-BCI applications. Moreover, we have also proposed a new backward alternative which allows for learning the distribution matching from the target domain (new session) to the source domain (calibration session). In addition, two online simulated adaptive scenarios were tested: i) block-wise adaptation and ii) sample-wise adaptation, with the main difference between these two scenarios being the way that the distribution mapping is learned. Finally, we have also explored two regularized OT formulations for learning the transportation plan ((4) and (5)).

In the first part of this work, we assumed that data distribution drift of a block of trials (run) can be estimated using previous available data. The results showed that for most of the cases there are no significant differences between any of the OTDA alternatives and both SR and RPA. It is important to underline here that BOTDA has the advantage of avoiding classifier retraining, and thus, from the machine learning viewpoint, BOTDA presents a more challenging problem than either of the other adaptive strategies. The lack of significant difference between BOTDA and both SR and RPA indicates that similar classification performance can be achieved avoiding classifier retraining when BOTDA is applied. With regard to the EA approach, we observed that it outperforms all other methods, being as good as SR. This method, as well as all the other considered adaptive methods, requires the training of a new classifier before a new testing block starts. Furthermore, in this scenario, we also note that adding the group-lasso penalty to the BOTDA formulation does not contribute to increasing classification performance. This can be explained by the fact that without classifier retraining, BOTDA-GL only relies on the optimal transportation plan learned with label information from a set of previous trials different from the current testing run. Finally, the fact that after a certain point, domain adaptation does not contribute towards classification improvement can be explained by the ‘‘saturation’’ effect in transfer learning [39], which establishes that when a sufficient amount of data from the target domain is available, a good classifier can already be trained without transfer learning. However, for MI-BCIs applications in motor rehabilitation, therapy session time is typically limited. Thus, having a machine-learning solution for avoiding long calibration times before each session is of utmost importance. In fact, for such practical application, it could allow dedicating more therapy

**TABLE III:** Accuracy reached by each subject for Dataset-1 (S1-S10) for each one of the evaluated methods in the sample-wise adaptation scenario. The average across subjects is aggregated at the last column. Best results appear in bold.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	av
<b>SC</b>	77.86	82.14	48.57	57.86	65.71	87.86	52.14	87.86	72.14	52.14	68.43
<b>SR</b>	85.00	87.86	81.43	77.86	78.57	89.29	81.43	<b>92.86</b>	82.14	71.43	82.79
<b>RPA</b>	97.14	89.29	<b>80.71</b>	<b>81.43</b>	84.29	95.71	87.86	90.71	88.57	87.86	88.357
<b>EA</b>	81.43	84.29	70.00	77.14	77.86	89.29	82.14	<b>92.86</b>	80.71	65.71	80.14
<b>FOTDA-S</b>	78.57	82.86	47.86	55.00	70.71	84.29	76.43	82.86	78.57	54.29	71.14
<b>FOTDA-GL</b>	78.57	82.14	47.86	55.00	71.43	85.00	75.71	83.57	78.57	54.29	71.21
<b>BOTDA-S</b>	68.57	78.57	50.00	52.86	73.57	87.86	72.14	88.57	75.71	55.00	70.29
<b>BOTDA-GL</b>	<b>100.00</b>	<b>95.00</b>	55.71	79.29	<b>85.00</b>	<b>97.14</b>	<b>88.57</b>	91.43	<b>98.57</b>	<b>97.86</b>	<b>88.86</b>

**TABLE IV:** Accuracy reached by each subject for Dataset-2 (S1 - S12) for each one of the evaluated methods in the sample-wise adaptation scenario. The average across subjects is aggregated at the last column. Best results appear in bold.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	av
<b>SC</b>	96.67	96.67	90.00	88.33	86.11	71.11	87.22	57.22	71.11	60.56	53.89	48.89	75.65
<b>SR</b>	98.89	97.78	91.67	89.44	92.78	84.44	94.44	68.89	83.89	70.00	73.89	66.67	84.40
<b>RPA</b>	98.33	97.22	92.78	90.00	93.89	<b>85.56</b>	<b>93.33</b>	84.44	86.67	75.56	79.44	<b>76.67</b>	87.83
<b>EA</b>	98.89	97.22	91.11	90.56	88.89	81.67	91.67	68.33	82.78	68.33	69.44	72.78	83.47
<b>FOTDA-S</b>	98.33	96.67	88.89	82.78	62.22	67.78	88.33	60.00	68.33	65.00	63.33	64.44	75.51
<b>FOTDA-GL</b>	99.44	96.67	89.44	85.00	85.56	74.44	90.00	62.22	66.11	65.00	63.33	64.44	78.47
<b>BOTDA-S</b>	98.33	96.11	87.22	88.33	86.67	72.22	86.67	62.78	71.67	65.00	64.44	52.78	77.69
<b>BOTDA-GL</b>	<b>100.00</b>	<b>98.89</b>	<b>96.11</b>	<b>91.67</b>	<b>100.00</b>	85.00	88.33	<b>93.33</b>	<b>100.00</b>	<b>76.11</b>	<b>100.00</b>	53.33	<b>90.23</b>

**TABLE V:** Overall computational time (seconds) of the SR, RPA, EA and BOTDA-GL methods for online adaptation.

	Dataset-1	Dataset-1
<b>SR</b>	0.326 ± 0.008	0.337 ± 0.015
<b>RPA</b>	2.711 ± 0.557	1.099 ± 0.324
<b>EA</b>	0.634 ± 0.109	0.537 ± 0.089
<b>BOTDA-GL</b>	0.034 ± 0.010	0.029 ± 0.014

time to provide the biofeedback to the patient rather than to re-train the decoding model, where no feedback can be provided. The improvements found by using BOTDA at the first testing runs without classifier retraining not only mitigates the well-known decrease in classification performance at the beginning of a new session, but it also constitutes a step forward in the search for more efficient and robust methodologies for applying BCIs to neurorehabilitation.

In our second experiment we analyzed a way to implement OTDA in online sequential scenarios where labeled data are available after every experimental trial. The classification results shown in tables III and IV clearly indicate the advantages of using the current trial in the BOTDA-GL configuration, whereas for the other tested alternatives of OTDA, the incorporation of the current trial to the transportation set does not improve classification performance. Interestingly, although in this context both SR and RPA strategies use the indicated mental task of the current trial to make the adaptation and learn the decoding model, overall, it is better to use the classifier already trained in the source domain and transport the current trial by means of BOTDA-GL. Additionally, we note here that there are few cases in which classification performance is not improved by BOTDA-GL. These cases are in fact shedding light on the idea that using the indicated mental task (known by the system) does not always represent the mental task

truly performed by the user. Thus, only good classification results will be achieved if the patterns provided by the user are discriminative enough for the corresponding mental task. In the case of stroke rehabilitation, patients may exhibit mental fatigue after repetitively performing the intended mental tasks [40]. In such situation the use of BOTDA could be a valuable solution to cope with the lack of ability of stroke survivors to maintain the same mental state throughout the whole BCI session.

Approaches that use the current label information for adaptive BCI have already been evaluated [16], [41], in which the testing class information was used to update the weights of a linear classifier. When BOTDA is being used, the classifier trained with data from the calibration session is kept fixed. Given that BOTDA-GL transformed testing data based on label information, i.e. penalizing the learned mapping so as to transport samples of the same class together, this approach will fail if the features do not match with the indicated mental task. This could explain the reduced decoding performance observed in some subjects (S3 an S4 in Dataset-1, and S7 and S12 in Dataset-2). On the contrary, although RPA also makes use of the indicated mental task, the classifier is re-trained, and thus this issue becomes less evident. As discussed by the authors in [42], if a user is not able to produce stable and distinctive EEG patterns, then the algorithm fails in matching the target sample to its supposed label class. In this direction, the classification reached by each subject in the new session could be interpreted as a quantification of how well the subject performed the indicated mental task. Although more research is needed, we believe that this supervised transfer learning based on our proposed backward OT could be a valuable alternative for avoiding the current pitfalls on how the user is trained and how the feedback is presented to the user during MI-BCI [43].

The computational time required for adapting each trial in the sample-wise scenario revealed that BOTDA is about 10 times faster than the fastest conventional method tested (SR). This result is important if we extrapolate the associated computational time of more complex decoding algorithms. The low computational time can be explained by two reasons: i) the decoding model is kept fixed, without any retraining needs, and ii) once the transportation plan is learned, the sample transportation (see Eq.(6)) consists in simple matrix multiplications [22]. The fact that by means of BOTDA the adaption is less time-consuming than re-training a simple decoding model reveals a huge advantage when considering real-time decoding, where the feedback presenting time should be kept as short as possible [44], [45]. Such an advantage, together with the ability to transport discriminative trials and boost the classification performance of an already trained decoding model, makes BOTDA-GL a promising, effective and efficient method for online applications.

In the last years the use of TL based on deep learning to tackle the variability in BCI has gained more relevance. Although different challenges exist for using deep learning methods to successfully train models in neuroscience, recent works have shown that deep learning architectures can be used for such applications [46]–[48]. However, most of the existing works aim to address the cross-subject variability and not the cross-session variability, the former being easier to implement since the cost of producing sufficient high-quality data and annotations is lower with respect to multi-subjects multi-sessions datasets. Nevertheless, given the rapid advance in the deep learning community, future research directions should explore the use of such techniques.

In this work, we have presented evidence that OT can be used for TL in MI-BCI in the context of a motor rehabilitation application. When data becomes available in blocks (i.e., group of multiple trials), we showed that classification performance can be improved for the first testing runs. For such scenarios, we suggest to use the BOTDA-S formulation. On the contrary, when data is available after each experimental trial, together with the indicated mental task information, BOTDA-GL should be preferred. Previous works have also shown the impact of using OT for domain adaptation in BCIs. In particular, the work in [23] focused on the use of OTDA for multi-subject P300 based-BCIs. Although the beforementioned work described a simulated online scenario, our approach cannot be directly compared, since, as opposed to our work, all available target data was used to learn the transport. Here, we have provided a workflow for using OTDA in real-life MI-BCI applications. Additionally, the authors in [11] introduce the RPA method and compared its performance against OTDA, applied in the Riemannian manifold. Unfortunately, there is no enough information on how the method is modified and implemented for running it in such domains. In regard to this, an interesting paper showing how to use OTDA in the Riemannian manifold has recently appeared [49]. Future work should explore further analyses in this direction.

Although we have focused on synchronized MI-BCI for motor rehabilitation, the proposed BOTDA method is not restricted to such applications. For asynchronous BCI, where

the user himself/herself chooses which task he/she wants to do and when to do it, the block-wise adaptation scenario can still be used by relying on a short calibration block of few trials to learn the transportation map. For such asynchronous BCI applications, BOTDA-S can still be used without any limitation. This method yields comparable classification results as compared to BOTDA-GL in the block-wise adaptation. Moreover, since BOTDA is applied at the level of the feature space, the workflow presented here can be easily adapted for any other BCI paradigm with either two or more classes.

## VI. CONCLUSIONS

In this paper we described and studied different strategies based on optimal transport for tackling cross-session variability in MI-BCI. Our results show that when data from the new BCI session is available in blocks (runs), classification accuracy can be increased by up to 9.5% by BOTDA as compared to the traditional recalibration approach, with only 20 trials of the new session required for the adaptation. In addition, when considering the current trial with its mental task information within the OT learning procedure, we found that BOTDA-GL improves up to perfect classification when the user is able to produce discriminative brain patterns, trimming down the computational time to only 30 ms. These findings indicate that BOTDA-GL is a promising approach for providing accurate and rapid feedback in cue-based environments. Further research is needed to demonstrate the applicability of BOTDA-GL as a user performance metric, which is part of our future research plans.

The use of transfer learning for tackling the EEG non-stationarity is gaining more and more attention within the BCI community. Methods based on data alignment and deep learning are at the top of the list. In the future, we plan to combine the advantages of these transfer learning strategies to try to further improve classifier robustness. In addition we plan to conduct multiple-session MI-BCI experiments for further validating the proposed strategies in real online scenarios.

## REFERENCES

- [1] K. K. Ang, C. Guan, K. S. G. Chua, B. T. Ang, C. W. K. Kuah, C. Wang, K. S. Phua, Z. Y. Chin, and H. Zhang, "A large clinical study on the ability of stroke patients to use an EEG-based motor imagery brain-computer interface," *Clinical EEG and Neuroscience*, vol. 42, no. 4, pp. 253–258, 2011.
- [2] L. Van Dokkum, T. Ward, and I. Laffont, "Brain computer interfaces for neurorehabilitation—its current status as a rehabilitation strategy post-stroke," *Annals of Physical and Rehabilitation Medicine*, vol. 58, no. 1, pp. 3–8, 2015.
- [3] I. Lazarou, S. Nikolopoulos, P. C. Petrantonakis, I. Kompatsiaris, and M. Tsolaki, "EEG-based brain–computer interfaces for communication and rehabilitation of people with motor impairment: a novel approach of the 21st century," *Frontiers in Human Neuroscience*, vol. 12, p. 14, 2018.
- [4] R. Mane, T. Chouhan, and C. Guan, "BCI for stroke rehabilitation: motor and beyond," *Journal of Neural Engineering*, vol. 17, no. 4, p. 041001, 2020.
- [5] A. Vourvopoulos, O. M. Pardo, S. Lefebvre, M. Neureither, D. Saldana, E. Jahng, and S.-L. Liew, "Effects of a brain-computer interface with virtual reality (VR) neurofeedback: A pilot study in chronic stroke patients," *Frontiers in Human Neuroscience*, vol. 13, p. 210, 2019.



- [6] K. K. Ang, K. S. G. Chua, K. S. Phua, C. Wang, Z. Y. Chin, C. W. K. Kuah, W. Low, and C. Guan, "A randomized controlled trial of EEG-based motor imagery brain-computer interface robotic rehabilitation for stroke," *Clinical EEG and Neuroscience*, vol. 46, no. 4, pp. 310–320, 2015.
- [7] D. Rathee, A. Chowdhury, Y. K. Meena, A. Dutta, S. McDonough, and G. Prasad, "Brain-machine interface-driven post-stroke upper-limb functional recovery correlates with beta-band mediated cortical networks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 1020–1031, 2019.
- [8] F. Lotte, L. Bougrain, and M. Clerc, "Electroencephalography (EEG)-based brain-computer interfaces," *Wiley Encyclopedia of Electrical and Electronics Engineering*, pp. 1–20, 1999.
- [9] W. Samek, M. Kawanabe, and K.-R. Müller, "Divergence-based framework for common spatial patterns algorithms," *IEEE Reviews in Biomedical Engineering*, vol. 7, pp. 50–72, 2013.
- [10] P. Gaur, K. McCreddie, R. B. Pachori, H. Wang, and G. Prasad, "Tangent space features-based transfer learning classification model for two-class motor imagery brain-computer interface," *International journal of neural systems*, vol. 29, no. 10, p. 1950025, 2019.
- [11] P. L. C. Rodrigues, C. Jutten, and M. Congedo, "Riemannian procrustes analysis: Transfer learning for Brain-Computer Interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 8, pp. 2390–2401, 2018.
- [12] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.
- [13] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.
- [14] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2007.
- [15] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2010.
- [16] P. Shenoy, M. Krauledat, B. Blankertz, R. P. Rao, and K.-R. Müller, "Towards adaptive classification for BCI," *Journal of Neural Engineering*, vol. 3, no. 1, p. R13, 2006.
- [17] X. Li, C. Guan, K. K. Ang, H. Zhang, and S. H. Ong, "Spatial filter adaptation based on the divergence framework for motor imagery EEG classification," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 1847–1850.
- [18] W. M. Kouw and M. Loog, "An introduction to domain adaptation and transfer learning," *arXiv preprint arXiv:1812.11806*, 2018.
- [19] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu, "Transfer Learning: A Riemannian Geometry Framework with Applications to Brain-Computer Interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 5, pp. 1107–1116, 2018.
- [20] H. He and D. Wu, "Transfer Learning for Brain-Computer Interfaces: A Euclidean Space Data Alignment Approach," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 399–410, 2020.
- [21] C. Villani, *Optimal transport: old and new*. Springer Science & Business Media, 2008, vol. 338.
- [22] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy, "Optimal transport for domain adaptation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 9, pp. 1853–1865, 2017.
- [23] N. T. Gayraud, A. Rakotomamonjy, and M. Clerc, "Optimal transport applied to transfer learning for P300 detection," in *BCI 2017 - 7th Graz Brain-Computer Interface Conference, Sep 2017, Graz, Austria*, 2017, p. 6.
- [24] L. V. Kantorovich, "On the translocation of masses," *Dokl. Akad. Nauk SSSR*, 37, vol. 37, no. 7-8, pp. 227–229, 1942.
- [25] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," in *Advances in Neural Information Processing Systems*, 2013, pp. 2292–2300.
- [26] P. A. Knight, "The Sinkhorn-Knopp algorithm: convergence and applications," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 1, pp. 261–275, 2008.
- [27] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A sparse-group lasso," *Journal of Computational and Graphical Statistics*, vol. 22, no. 2, pp. 231–245, 2013.
- [28] R. Flamary and N. Courty, "POT python optimal transport library," 2017. [Online]. Available: <https://github.com/rflamary/POT>
- [29] J. Wolpaw and E. W. Wolpaw, *Brain-computer interfaces: principles and practice*. OUP USA, 2012.
- [30] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain-computer interfaces," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 871–890, 2015.
- [31] K. K. Ang and C. Guan, "EEG-based strategies to detect motor imagery for control and rehabilitation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 4, pp. 392–401, 2016.
- [32] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain-computer interface classification by Riemannian geometry," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 920–928, 2011.
- [33] V. Peterson, D. Wyser, O. Lambercy, R. Spies, and R. Gassert, "A penalized time-frequency band feature selection and classification procedure for improved motor intention decoding in multichannel EEG," *Journal of Neural Engineering*, vol. 16, no. 1, p. 016019, 2019.
- [34] S. Marchesotti, R. Martuzzi, A. Schurger, M. L. Blefari, J. R. del Millán, H. Bleuler, and O. Blanke, "Cortical and subcortical mechanisms of brain-machine interfaces," *Human brain mapping*, vol. 38, no. 6, pp. 2971–2989, 2017.
- [35] J. Faller, C. Vidaurre, T. Solis-Escalante, C. Neuper, and R. Scherer, "Autocalibration and recurrent adaptation: Towards a plug and play online ERD-BCI," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 3, pp. 313–319, 2012.
- [36] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen *et al.*, "MEG and EEG data analysis with MNE-Python," *Frontiers in Neuroscience*, vol. 7, p. 267, 2013.
- [37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [38] N. Courty, R. Flamary, and D. Tuia, "Domain adaptation with regularized optimal transport," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2014, pp. 274–289.
- [39] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine Learning*, vol. 79, no. 1-2, pp. 151–175, 2010.
- [40] R. Foong, N. Tang, E. Chew, K. S. G. Chua, K. K. Ang, C. Quek, C. Guan, K. S. Phua, C. W. K. Kuah, V. A. Deshmukh, L. H. L. Yam, and D. K. Rajeswaran, "Assessment of the Efficacy of EEG-Based MI-BCI with Visual Feedback and EEG Correlates of Mental Fatigue for Upper-Limb Stroke Rehabilitation," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 3, pp. 786–795, 2020.
- [41] C. Vidaurre, M. Kawanabe, P. von Büna, B. Blankertz, and K.-R. Müller, "Toward unsupervised adaptation of LDA for brain-computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 587–597, 2010.
- [42] F. Lotte, F. Larue, and C. Mühl, "Flaws in current human training protocols for spontaneous brain-computer interfaces: lessons learned from instructional design," *Frontiers in human neuroscience*, vol. 7, p. 568, 2013.
- [43] F. Lotte, C. Jeunet, B. Mladenovi, Jelena abd Nkaoua, and L. Pillette, *A BCI challenge for the signal processing community: considering the user in the loop*. IET, 2018, pp. 1–33.
- [44] E. F. Oblak, J. A. Lewis-Peacock, and J. S. Sulzer, "Self-regulation strategy, feedback timing and hemodynamic properties modulate learning in a simulated fmri neurofeedback environment," *PLoS computational biology*, vol. 13, no. 7, p. e1005681, 2017.
- [45] A. Belinskaya, N. Smetanin, M. Lebedev, and A. Ossadtchi, "Short-delay neurofeedback facilitates training of the parietal alpha rhythm," *Journal of Neural Engineering*, 2020.
- [46] D. Wu, Y. Xu, and B. Lu, "Transfer learning for eeg-based brain-computer interfaces: A review of progresses since 2016," *arXiv preprint arXiv:2004.06286*, 2020.
- [47] N. A. Alzhab, L. Apollonio, A. Di Iorio, M. Alshalak, S. Iarlori, F. Ferracuti, A. Monteriù, and C. Porcaro, "Hybrid Deep Learning (hDL)-Based Brain-Computer Interface (BCI) Systems: A Systematic Review," *Brain Sciences*, vol. 11, no. 1, p. 75, 2021.
- [48] S. Roy, A. Chowdhury, K. McCreddie, and G. Prasad, "Deep learning based inter-subject continuous decoding of motor imagery for practical brain-computer interfaces," *Frontiers in Neuroscience*, vol. 14, 2020.
- [49] O. Yair, F. Dietrich, R. Talmon, and I. G. Kevrekidis, "Domain Adaptation with Optimal Transport on the Manifold of SPD matrices," *arXiv*, 2019.

**Doctorado en Ingeniería**  
**Mención en Inteligencia Computacional, Señales y Sistemas**

Título de la obra:

**Algoritmos para interfaces**  
**Cerebro-Computadora en**  
**paradigmas relacionados con el habla**

Autor: Nicolás Nieto

Lugar: Santa Fe, Argentina

Palabras Claves:

Habla interna, habla imaginada, aprendizaje maquina, base de datos pública,  
Máquinas de aprendizaje extremo, transporte óptimo.