

Trabajo Fin de Grado

Grado en Ingeniería Electrónica, Robótica y
Mecatrónica

Interfaz BCI para reconocimiento de palabras en
señales EEG

Autor: Pablo Martín Puerto

Tutor: José Manuel Quero Reboul

Dpto. Ingeniería Electrónica
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2023



Trabajo Fin de Grado
Grado en Ingeniería Electrónica, Robótica y Mecatrónica

Interfaz BCI para reconocimiento de palabras en señales EEG

Autor:

Pablo Martín Puerto

Tutor:

José Manuel Quero Reboul

Catedrático de Universidad

Dpto. Ingeniería Electrónica
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2023

Trabajo Fin de Grado: Interfaz BCI para reconocimiento de palabras en señales EEG

Autor: Pablo Martín Puerto

Tutor: José Manuel Quero Reboul

El tribunal nombrado para juzgar el Proyecto arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocal/es:

Secretario:

Acuerdan otorgarle la calificación de:

El secretario del Tribunal

Fecha:

A mi familia: de sangre y elegida

A Sevilla y al rockero Silvio

Agradecimientos

Estos agradecimientos son una muestra de mi reconocimiento profundo y sincero hacia aquellos que han contribuido directa e indirectamente a mi desarrollo personal y académico. Sin ellos, no hubiese llegado al final de esta carrera.

A todos los maestros que tuve en el colegio, instituto y Universidad, Su dedicación y compromiso con la educación pública han sido fundamentales para mi formación y crecimiento como estudiante y ciudadano.

A toda mi familia por su apoyo incondicional a lo largo de este viaje académico.

A mis padres y mi hermano, por darme la vida y por ser mi mayor inspiración. Ellos han estado a mi lado brindándome todo su amor y respaldo, y por impulsarme a crecer y perseguir mis metas.

A mis abuelos por haberme transmitido directamente o a través de las palabras de otros familiares los valores que forman la base de mi carácter y mi forma de ver el mundo.

A mi tita Maribel y mi compañero Abraham quienes, quizás, no son conscientes de cuánto les aprecio. Han sido una fuente de inspiración y motivación, además de enseñarme la importancia de valorar cada instante y afrontar los desafíos con valentía.

A mis amigos, a Lucía y al fútbol, ellos son quienes me escuchan, me soportan y me aconsejan diariamente.

Gracias a todos por hacerme feliz.

Pablo Martín Puerto

Sevilla, 2023

Resumen

Una Interfaz Cerebro-Ordenador (o BCI, del inglés Brain-Computer Interface) puede ser definida como un sistema que traduce los patrones de actividad cerebral de un individuo en mensajes o comandos para una aplicación interactiva.

En la actualidad, el electroencefalograma (EEG) se destaca como el método más empleado para el registro no invasivo de la actividad cerebral. Esta técnica ofrece una alta resolución temporal, una notable facilidad de uso y menor riesgo de complicaciones en comparación con otras técnicas invasivas. No obstante, esta técnica también presenta algunas limitaciones, como la presencia de altos niveles de ruido, el efecto de la postura y concentración en la tarea a realizar por parte del sujeto o las interferencias producidas por el movimiento ocular y muscular. Todo ello, junto a la baja intensidad de las señales, resultan en una baja relación señal a ruido (SNR) y, por lo tanto, una alta dificultad para interpretar las señales EEG.

La realización de este trabajo viene motivada por el significativo potencial que protagonizan las señales EEG en el desarrollo BCI, las cuales pueden ser implementadas para abrir un nuevo canal de comunicación para las personas con discapacidades motoras en el habla u otras condiciones neurológicas. Precisamente, se abordará la primera de estas dificultades con el propósito de brindar ayuda a individuos que padecen estas complicaciones, además de contribuir con la aportación de este estudio al realizado por investigadores de todo el mundo en el intento de decodificación del habla a partir de ondas cerebrales.

En este contexto, se realizará un estudio exhaustivo del estado del arte en el reconocimiento de palabras mediante EEG. Se investigarán las diferentes técnicas de adquisición de señales, el procesamiento de las señales EEG y el uso de redes neuronales en este campo. Además, se planteará el diseño y desarrollo de una diadema utilizando tecnología de impresión 3D para la captura de señales EEG. Esta diadema estará equipada con electrodos secos y se evaluará su rendimiento mediante la recopilación de señales reales. Asimismo, se analizarán un par de Datasets disponibles en la literatura y se procederá a la comparación con la realización propia del sistema BCI. Al finalizar, se obtendrán conclusiones sólidas basadas en los resultados y el análisis realizado.

Abstract

A Brain-Computer Interface (BCI) can be defined as a system that translates an individual's brain activity patterns into messages or commands for an interactive application.

Currently, electroencephalogram (EEG) stands out as the most widely used method for non-invasive recording of brain activity. This technique offers high temporal resolution, remarkable ease of use and lower risk of complications compared to other invasive techniques. However, this technique also has some limitations, such as the presence of high noise levels, the effect of posture and concentration on the task to be performed by the subject or the interferences produced by eye and muscle movement. All this, together with the low intensity of the signals, results in a low signal-to-noise ratio (SNR) and, therefore, a high difficulty in interpreting EEG signals.

This work is motivated by the significant potential of EEG signals in BCI development, which can be implemented to open a new communication channel for people with motor speech disabilities or other neurological conditions. Precisely, the first of these difficulties will be addressed with the purpose of providing help to individuals who suffer from these complications as well as contributing to the studies carried out by researchers around the world in the attempt of speech decoding from brain waves.

In this context, an exhaustive study of the state of the art in EEG word recognition will be carried out. The different signal acquisition techniques, EEG signal processing and the use of neural networks in this field will be investigated. In addition, the design and development of a headband using 3D printing technology for EEG signal capture will be proposed. This headband will be equipped with dry electrodes and its performance will be evaluated by collecting real signals. Furthermore, a couple of Datasets available in the literature will be analyzed and compared with the in-house realization of the BCI system. At the end, solid conclusions will be drawn based on the results and analysis performed.

Índice

Agradecimientos	ix
Resumen	xi
Abstract	xiii
Índice	xiv
Índice de Tablas	xvii
Índice de Figuras	xix
Notación	xxi
1 Introducción	1
1.1. <i>Brain-Computer Interface (BCI)</i>	1
1.1.1 Habla interna (Inner speech)	1
1.1.2 Habla pronunciada (Pronounced speech)	1
1.1.3 Condición visualizada (Visualized condition)	1
1.2. <i>Señales EEG</i>	2
1.2.1. Historia del electroencefalograma	2
1.2.2. Procesos neuronales y registro de EEG	3
1.2.3. Características relevante de las señales EEG para reconocimiento de palabras	3
2. Estado del arte	11
2.1. <i>Adquisición y preprocesamiento de señales EEG en BCI</i>	11
2.2.1. Electroodos para captura de señales EEG	11
2.2.2. Desafíos asociados con la adquisición de señales EEG	13
2.2.3. Técnicas de adquisición y preprocesamiento de señales EEG	13
2.2. <i>Procesamiento de señales, extracción y selección de características</i>	16
2.2.1. Morlet Wavelet	17
2.2.2. Daubechies 4 (D4)	18
2.3. <i>Redes neuronales para el reconocimiento de palabras</i>	19
2.4. <i>Resultados de estudios en la literatura</i>	20
3. Diseño e implementación de la BCI	23
3.1. <i>Sistemas de adquisición de señales</i>	23
3.1.1. Detalles de las señales propias	23
3.1.2. Descripción de los datasets en la literatura utilizados	27
3.2. <i>Procesamiento de señales, extracción y selección de características</i>	29
3.2.1. Señales EEG como inputs	29
3.2.2. Bandas de frecuencia y espectrogramas como inputs	30
3.3. <i>Implementación de redes neuronales</i>	30
3.3.1. Aprendizaje profundo para el reconocimiento de palabras mediante audio	30
3.3.2. Arquitectura típica de red LSTM	31

3.3.3.	Red convolucional de tiempo-frecuencia para la clasificación de datos EEG	31
3.3.4.	GoogleNet y AlexNet	31
4.	Resultados y Discusión	34
4.1.	<i>Tabla de resultados obtenidos</i>	34
4.2.	<i>Interpretación de los datos obtenidos</i>	35
5.	Líneas Futuras	37
5.1.	<i>Mejora en el tratamiento del EEG</i>	37
5.2.	<i>Métodos alternativos al EEG</i>	37
	Referencias	39

ÍNDICE DE TABLAS

Tabla 1. Estudios en la literatura para el dataset 'Thinking Out Loud' [3]	20
Tabla 2. Estudios en la literatura para el dataset 'MindBigData' [34]	21
Tabla 3. Estudios en la literatura para otros datasets	22
Tabla 4. Características placa 'EEG Click'	24
Tabla 5. Resultados del aprendizaje mediante diferentes arquitecturas	34
Tabla 6. Comparativa de resultados: métodos propuestos y literatura	35

ÍNDICE DE FIGURAS

Figura 1. Estructura de un DS-BCI	2
Figura 2. Áreas de Brodmann, imagen extraída de Wikipedia	4
Figura 3. Ejemplo de energía en las diferentes bandas de frecuencia	6
Figura 4. Sistema 10-10, imagen extraída de Wikipedia	12
Figura 5. Sistema 10-20, imagen extraída de Wikipedia	12
Figura 6. Sistema 128-Biosemi, imagen extraída de la web de Biosemi [12]	13
Figura 7. Ilustración del concepto de ICA	16
Figura 8. Morlet Wavelet, imagen extraída de la web de MATLAB®	17
Figura 9. Wavelet de Morlet compleja, imagen extraída de la web de MATLAB®	18
Figura 10. Descomposición mediante D4	19
Figura 11. Diseño 3D de la diadema para experimento propio	24
Figura 12. Dispositivo ‘EEG Click’, imagen extraída de su datasheet	25
Figura 13. Atmel® SAM E70 Xplained, imagen extraída de su datasheet	25
Figura 14. Primera interfaz del experimento propio	26
Figura 15. Pantalla inicial y ‘Cue Interval’ del experimento propio	27
Figura 16. Secuencia de cada trial, extraída del artículo original de Thinking Out Loud [3].	28
Figura 17. Dígitos mostrados en ‘MindBigData’, extraído de su artículo original [41].	29
Figura 18. Diagrama de una red típica LSTM	31
Figura 19. Espectrograma del canal A15 para un trial de la palabra ‘abajo’ (Thinking Out Loud)	32
Figura 20. Montaje de 16 espectrogramas para un trial de la palabra ‘abajo’ (Thinking Out Loud)	33
Figura 21. Montaje de los 4 espectrogramas para un trial del dígito ‘0’ (MindBigData)	33

BCI	Brain-Computer Interface
CNN	Convolutional Neural Network
TF	Time-Frecuency
ERP	Event Related Potential
EOG	Electrooculograma
EMG	Electromiografía
ICA	Independent Component Analysis
LSTM	Long Short Term Memory
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
D4	Daubechies-4
%	Tanto por ciento
V	Voltio
TFG	Trabajo Fin de Grado
SNR	Signal-to-Noise Ratio

1 INTRODUCCIÓN

La evolución tecnológica ha provocado una amplia conectividad entre los dispositivos electrónicos cotidianos y el cuerpo humano. Ejemplo de ello es el BCI, que es un sistema que utiliza la actividad cerebral y las señales EEG para crear un canal de comunicación entre dispositivos electrónicos externos y el cerebro humano [1].

1.1. Brain-Computer Interface (BCI)

Los sistemas BCI se han utilizado en diversas áreas. Ejemplos de ellos son neuromarketing, seguridad, entretenimiento o control de entornos inteligentes, entre otros. Una de las aplicaciones de BCI más exploradas es en el área médica, donde ha sido ampliamente utilizada para el tratamiento y diagnóstico de trastornos neurológicos como la epilepsia, la depresión, la demencia o el Alzheimer [1]. También se ha aplicado para el reconocimiento y clasificación de emociones, así como para identificar las diferentes etapas del sueño. Además, una de las aplicaciones más desafiantes e interesantes de la BCI es el reconocimiento del habla imaginada, donde se busca convertir las señales cerebrales en texto, sonido o comandos de control [1]. Este es el caso del Direct-Speech BCI (DS-BCI), el cual adquiere señales cerebrales correspondientes al habla imaginada, las procesa y decodifica para producir una salida lingüística en forma de fonemas, palabras o frases [2]. Pero la adquisición de dichas señales no es tarea fácil, ya que la producción del habla es uno de los procesos cerebrales más complejos que realizan los seres humanos, pues requiere la interacción de varias regiones cerebrales corticales, basales y subcorticales [3]. Por otro lado, se abre paso a varios paradigmas presentados en [3]:

1.1.1 Habla interna (Inner speech)

En este paradigma, los participantes imaginan su propia voz como si estuvieran dando una orden directa al ordenador. No se mueve la boca ni la lengua, manteniéndose lo más quietos posible, sin mover la boca ni la lengua. El objetivo principal es detectar la actividad eléctrica cerebral relacionada con el pensamiento de una palabra en particular.

1.1.2 Habla pronunciada (Pronounced speech)

El paradigma de habla pronunciada se propone con el objetivo de encontrar las regiones motoras involucradas en la pronunciación que se activan durante el habla interna. En este paradigma, los participantes pronuncian en voz alta la palabra correspondiente a cada estímulo visual, como si estuvieran dando una orden directa a la computadora.

1.1.3 Condición visualizada (Visualized condition)

Dado que las palabras seleccionadas tienen un componente visual y espacial importante, y con el objetivo de encontrar cualquier actividad relacionada con la visualización durante el habla interna, se propone el paradigma de visualización. Se menciona que los principales centros neurales relacionados con el pensamiento espacial se encuentran en las regiones occipital y parietal. Además, se ha demostrado que la atención espacial tiene un impacto significativo en la amplitud de los potenciales evocados visuales. En los ensayos de visualización, se indica a los participantes que se concentren en mover mentalmente el círculo que aparece en el centro de la pantalla en la dirección indicada por el estímulo visual.

Se mantendrá el nombre de cada paradigma en inglés a partir de ahora.

La estructura de un DS-BCI consiste de varias etapas:

1. Adquisición y acondicionamiento de la señal (Signal Acquisition): registro y preprocesado de la actividad cerebral del usuario.
2. Procesamiento de señal y extracción de características (Signal Processing & Features extraction): se recibe la señal digitalizada se filtra para eliminar ruidos y se aplican distintas técnicas para la obtención de características.
3. Clasificación de las características (Features classification): mediante redes neuronales o machine learning.
4. Aplicación o Interfaz de control (Application or control interface): bloque funcional que recibe los comandos de control y realiza las acciones necesarias.

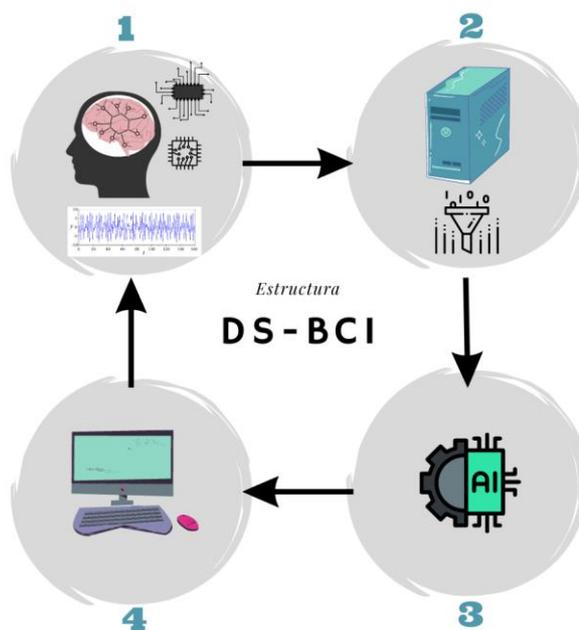


Figura 1. Estructura de un DS-BCI

1.2. Señales EEG

1.2.1. Historia del electroencefalograma

La electroencefalografía es el estudio de los patrones de actividad eléctrica del cerebro humano, y es una herramienta clave en la comprensión de la función cerebral y el diagnóstico de trastornos neurológicos. Su historia se remonta a finales del siglo XX. Esta comienza cuando, en 1875, Richard Caton, un fisiólogo británico, realizó los primeros experimentos registrando la actividad eléctrica del cerebro de animales. Sus estudios mostraron que el cerebro emite señales eléctricas débiles y que estas señales varían según el estado de vigilia y sueño [4].

Unos 50 años más tarde, Hans Berger, un psiquiatra alemán, realizó un avance significativo al registrar la primera señal EEG en un ser humano [4]. Berger desarrolló un electroencefalógrafo, un dispositivo capaz de medir la actividad eléctrica del cerebro a través de electrodos colocados en el cuero cabelludo [4]. Sus estudios

demonstraron la existencia de diferentes tipos de ondas cerebrales, como las ondas alfa y beta, y sentaron las bases de la electroencefalografía moderna [5].

A lo largo de las décadas siguientes, se realizaron importantes avances en la tecnología de registro EEG y en la comprensión de los patrones de actividad cerebral. Se desarrollaron sistemas más sofisticados de amplificación y registro de señales, lo que permitió una mayor precisión y resolución en la captura de la actividad cerebral [6].

En la década de 1960, se introdujo la digitalización de las señales EEG, lo que permitió un procesamiento más avanzado y análisis de los datos obtenidos. Esto llevó al desarrollo de técnicas de análisis de frecuencia y tiempo en el dominio de las señales EEG, lo que permitió una mejor comprensión de las diferentes características y patrones presentes en las señales cerebrales [6].

En las últimas décadas, la tecnología EEG ha experimentado avances significativos, como la miniaturización de los dispositivos de registro y el desarrollo de técnicas de procesamiento de señales más avanzadas, como el aprendizaje automático y las redes neuronales. Esto ha permitido una mayor aplicabilidad de las señales EEG en diversas áreas, como la medicina, la neurociencia cognitiva, la investigación del sueño y las Interfaces Cerebro-Computadora (BCI) [6], que es relevante para nuestro trabajo.

1.2.2. Procesos neuronales y registro de EEG

El electroencefalograma registra la actividad eléctrica del cerebro, la cual se origina en la sincronización de millones o incluso miles de millones de neuronas en el cerebro [5].

Cuando las neuronas se comunican entre sí, utilizan neurotransmisores, como los iones de sodio (Na^+) y potasio (K^+), para transmitir señales eléctricas y químicas. Estos neurotransmisores desempeñan un papel fundamental en la generación de potenciales de acción, que son los impulsos eléctricos que se propagan a lo largo de las neuronas [5].

En una neurona, el potencial de reposo se mantiene negativo debido a la distribución de iones a través de la membrana celular. El sodio (Na^+) se encuentra predominantemente fuera de la célula, mientras que el potasio (K^+) se encuentra predominantemente dentro de la célula. Esta distribución crea una diferencia de carga eléctrica y establece el potencial de reposo [5].

Cuando una neurona recibe suficiente estimulación excitatoria, como la liberación de neurotransmisores excitatorios, se produce un cambio en la permeabilidad de la membrana celular. Esto permite que los iones de sodio (Na^+) ingresen a la célula y los iones de potasio (K^+) salgan de la célula, generando un potencial de acción. El potencial de acción se propaga a lo largo del axón de la neurona y puede activar a otras neuronas con las que está conectada a través de sinapsis [5].

En el contexto de las señales EEG, las corrientes generadas por las neuronas se propagan a través de los tejidos cerebrales que contienen agua. Esto se conoce como conducción de volumen y permite que los efectos de la actividad eléctrica de las neuronas se propaguen a distancias relativamente largas, incluso varios centímetros. Estas corrientes generan campos eléctricos que se suman linealmente y se propagan instantáneamente [5].

En la superficie del cuero cabelludo, donde se colocan los electrodos para registrar el EEG, la amplitud de las señales es muy débil, generalmente medida en milivoltios (μV), mientras que en la membrana el potencial es de varios órdenes de magnitud mayores (mV). Esto se debe a que las señales en la superficie son una representación indirecta de la actividad eléctrica de las neuronas en el cerebro. A pesar de su debilidad, las señales EEG proporcionan información valiosa sobre la actividad cerebral y se pueden analizar para detectar patrones y cambios en la actividad eléctrica del cerebro [5].

1.2.3. Características relevante de las señales EEG para reconocimiento de palabras

1.2.3.1. Áreas de Brodmann, Broca y Wernicke

La mayoría de los modelos y teorías del lenguaje coinciden en que el habla implica procesamiento auditivo, semántico y sintáctico, así como planificación motora y procesos de articulación [3].

Las áreas de Brodmann son una clasificación de regiones en la corteza cerebral basada en diferencias en la estructura celular y la función de cada una de ellas. Fueron definidas por el neurólogo alemán Korbinian Brodmann a principios del siglo XX, quien realizó estudios exhaustivos del cerebro humano y dividió la corteza cerebral en diferentes áreas numeradas del 1 al 47 [5] como se puede observar en la Figura 2. Dos regiones que desempeñan un papel fundamental en la producción y comprensión del lenguaje son: El área de Broca (Áreas de Brodmann 44 y 45) y el área de Wernicke (Áreas de Brodmann 22, 39, 40).

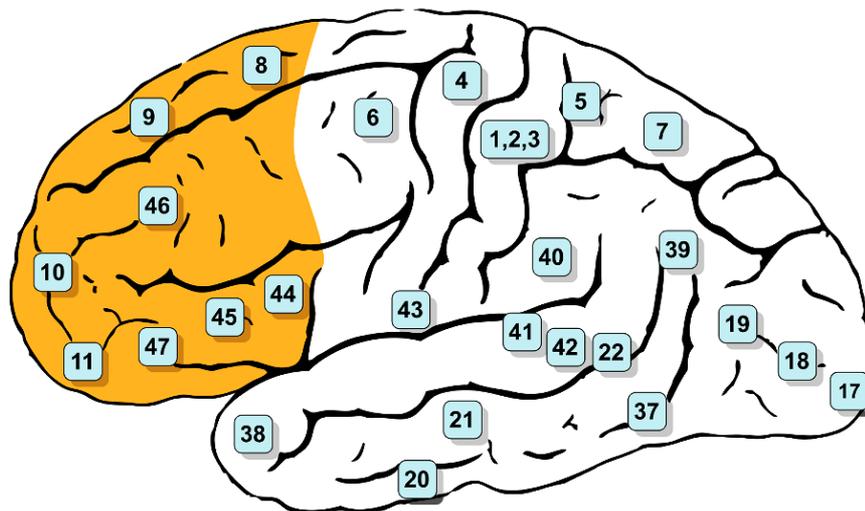


Figura 2. Áreas de Brodmann, imagen extraída de Wikipedia

El área de Broca (Áreas de Brodmann 44 y 45), ubicada en el lóbulo frontal del hemisferio dominante (generalmente el hemisferio izquierdo en la mayoría de las personas), es responsable de la producción del lenguaje hablado y la planificación de los movimientos articulatorios necesarios para articular las palabras [7]. Lesiones en el área de Broca pueden dar lugar a una condición conocida como afasia de Broca, en la cual los individuos tienen dificultades para expresarse verbalmente, pero conservan la comprensión del lenguaje [8].

El área de Wernicke (Áreas de Brodmann 22, 39, 40), se encuentra en el lóbulo temporal posterior, adyacente al área auditiva primaria [7]. Esta área está involucrada en la comprensión y procesamiento del lenguaje hablado y escrito. Las lesiones en el área de Wernicke pueden dar lugar a una afasia de Wernicke, en la cual los individuos tienen dificultades para comprender el lenguaje y producen un discurso incoherente y fluente [8].

La relación entre el área de Broca y el área de Wernicke es crucial para la producción y comprensión del lenguaje. Se cree que estas áreas están conectadas por una vía neuronal conocida como fascículo arqueado, que permite la comunicación entre ellas [7].

En resumen, las áreas de Broca y de Wernicke, son regiones específicas del cerebro involucradas en la producción y comprensión del lenguaje. Estas áreas trabajan en conjunto para permitirnos comunicarnos verbalmente y comprender el significado de las palabras.

1.2.3.2. Event Related Potentials (ERP)

Cuando se presenta un estímulo, se produce un cambio en la señal de EEG conocido como potencial relacionado con eventos (ERP, por sus siglas en inglés). El estímulo que da origen a un ERP representa la actividad generada en las neuronas, es decir, el cambio de potencial causado por el estímulo. El ERP representa la actividad eléctrica de una población considerable de neuronas [7].

Un ERP se obtiene promediando las señales de EEG de un tipo específico de evento [7]. La realización de un promedio en el estudio del comportamiento permite observar y monitorear los estímulos desatendidos. Además, también ofrece otra ventaja, ya que reduce el ruido residual en la señal promediada en comparación con un solo ensayo, aunque esto tiene sus limitaciones. Los ERPs se han utilizado para controlar sistemas BCI basados en la comunicación, y algunos de los ERPs utilizados son SSVEP y P300.

- **SSVEP**

El potencial evocado visual de estado estable (SSVEP) es un componente visual utilizado en las interfaces cerebro-computadora (BCI). Se basa en estímulos intermitentes de frecuencia constante y existe una correlación significativa entre la frecuencia de intermitencia del estímulo y la frecuencia observada en el EEG. Esto permite determinar el objetivo deseado por el usuario al coincidir el patrón de actividad del EEG con el comando asociado a una frecuencia particular.

El SSVEP ofrece ventajas como la facilidad de uso para la mayoría de los sujetos y la posibilidad de controlar dispositivos con alta precisión [2]. Sin embargo, es importante destacar que el uso de estímulos intermitentes puede generar fatiga en el usuario, especialmente cuando se utilizan frecuencias de intermitencia bajas [2]. Además, el paradigma del SSVEP no es adecuado para personas con discapacidades visuales, ya que requiere cambios de enfoque visual [2].

- **P300**

El P300 es otro de los ERP más estudiados. El componente P300 se genera en respuesta a eventos presentados repetitivamente y la aparición ocasional de un estímulo diferente, utilizando lo que se conoce como un "paradigma oddball". El P300 es un pico positivo en el ERP que varía en tamaño de 5 a 10 microvoltios y tiene una latencia entre 220 y 500 ms posterior al inicio del evento. Este ERP se define como un aumento promediado en la amplitud de las series temporales de las señales cerebrales, siendo más significativo en ubicaciones de la línea media (Pz, Cz y Fz en el sistema internacional 10/20). Es probable que el P300 obtenido en cualquier experimento sea la suma del P300 y otros componentes que son generados por otros estímulos o interferencias presentados antes y después de cada presentación del estímulo de interés [2].

- Otro paradigma relacionado con el habla interior es la denominada "comprensión auditiva" [3]. En este paradigma, en lugar de producir activamente la imaginación del habla, el individuo escucha pasivamente el habla de otra persona. Ya se ha explorado utilizando EEG, ECoG y fMRI. Aunque este paradigma no es especialmente útil para aplicaciones reales de BCI, ha contribuido a la comprensión de los procesos neuronales asociados con paradigmas relacionados con el habla.

El análisis de los ERP en el reconocimiento de palabras permite estudiar la velocidad de procesamiento, la activación de áreas cerebrales específicas y la comprensión semántica de las palabras. Además, se pueden utilizar técnicas de procesamiento de señales y algoritmos de aprendizaje automático para extraer características relevantes de los ERP y desarrollar modelos de reconocimiento de palabras.

Sin embargo, es importante tener en cuenta que el análisis de los ERP en el reconocimiento de palabras presenta desafíos, como el ruido inherente en las señales EEG y la variabilidad interindividual en las respuestas cerebrales. Además, la interpretación de los componentes ERP puede depender del contexto y de la tarea utilizada.

1.2.3.3. Bandas de frecuencias

Las cuatro frecuencias cerebrales principales son theta (4-8Hz), alpha (8-13 Hz), beta (13-30Hz) y gamma (30-200Hz). Estas desempeñan diferentes roles en el procesamiento del habla y se ha demostrado que se comportan de manera distinta en diferentes actividades, lo que desafía la idea de que tienen un comportamiento generalizado en actividades específicas. La comprensión del cerebro y su comportamiento está en sus primeras etapas, y fenómenos como la plasticidad cerebral hacen que sea aún más difícil comprender las complejidades del cerebro humano y las ondas cerebrales involucradas [7].

Cada frecuencia tiene una función específica en diferentes etapas del procesamiento del lenguaje y el habla, lo que nos ayuda a comprender su importancia en esta investigación. A continuación, presentamos una revisión

bibliográfica de las frecuencias involucradas en el procesamiento del habla [7]:

- Theta (θ)

Las ondas theta reaccionan a las características fonéticas del habla y desempeñan un papel en la reconstrucción de fonemas y el procesamiento de señales de coarticulación¹, lo que ayuda en la construcción de palabras dentro de un discurso.

- Alpha (α)

La desincronización relacionada con eventos en la frecuencia α se considera un signo de retroalimentación al córtex motor primario durante la producción del habla. Se observan fenómenos de sincronización y desincronización de α durante la percepción del habla, y un aumento en la actividad de esta se relaciona con tareas de concentración y carga de trabajo de memoria, lo que ayuda en el procesamiento auditivo y la producción del habla.

- Beta (β)

Las ondas beta en el habla están relacionadas con el movimiento muscular y se especula que juegan un papel en la generación de control de retroalimentación a través del córtex motor primario y la modelización interna. Estas ondas se utilizan en la discriminación del habla y luego se transfieren a la región auditiva para su procesamiento y producción del habla.

- Gamma (γ)

Durante la producción del habla, se observan cambios en la frecuencia gamma en diversas áreas del cerebro, como el lóbulo temporal, el surco supramarginal, el área de Broca, el área de Wernicke, el córtex premotor y el córtex motor primario. La frecuencia γ tiene un efecto rápido en respuesta al intervalo de presentación del estímulo y al intervalo entre estímulos.

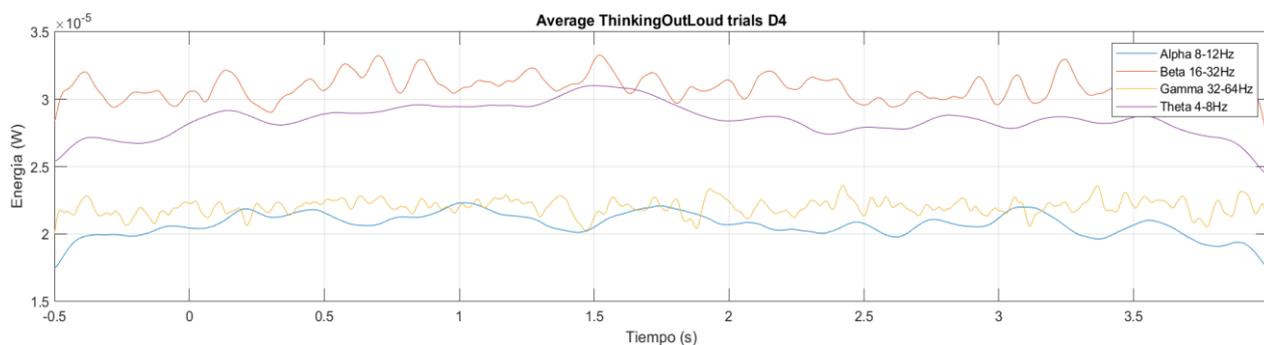


Figura 3. Ejemplo de energía en las diferentes bandas de frecuencia

En resumen, cada frecuencia cerebral desempeña un papel específico en el procesamiento del habla. Su estudio ayuda a comprender mejor cómo se procesa y produce el lenguaje hablado en el cerebro humano, por lo que se utilizarán todas las bandas.

¹ Coarticulación: Proceso fonético en que dos consonantes individuales se pronuncian al mismo tiempo. [12]

2. ESTADO DEL ARTE

El estado del arte de las señales EEG en BCI abarca un amplio espectro de investigaciones y desarrollos en la adquisición y preprocesamiento de las señales. Estos avances buscan superar los desafíos inherentes a las señales EEG y mejorar la precisión y confiabilidad de los sistemas BCI. El conocimiento de estos avances y técnicas en el estado del arte proporciona una base sólida para esta investigación y el desarrollo de sistemas BCI más efectivos y accesibles.

2.1. Adquisición y preprocesamiento de señales EEG en BCI

2.1.1. Electrodo para captura de señales EEG

Existen diferentes técnicas y sistemas de colocación de electrodos para la adquisición de señales EEG en el contexto de las BCI.

2.1.1.1. Tipos de electrodos

Dos de los tipos más comunes son los electrodos secos y los electrodos húmedos:

- Electrodo seco: no requieren gel conductor y son más fáciles de colocar, lo que los hace convenientes para aplicaciones portátiles y de uso doméstico. Sin embargo, los electrodos secos tienden a tener una calidad de señal inferior en comparación con los electrodos húmedos, ya que la impedancia entre el electrodo y la piel es mayor, lo que puede resultar en un mayor ruido y menor sensibilidad en la detección de señales cerebrales [10].
- Electrodo húmedo: requieren la aplicación de gel conductor en el cuero cabelludo para mejorar la conductividad eléctrica. Estos electrodos proporcionan una mejor calidad de señal debido a la menor impedancia entre el electrodo y la piel. Sin embargo, su colocación y mantenimiento pueden ser más laboriosos, y el gel conductor puede causar molestias o irritación en la piel [10].

2.1.1.2. Sistemas de colocación de electrodos

En cuanto a los sistemas de colocación de electrodos, dos de los más utilizados son el sistema 10-10 y el sistema 10-20. Estos sistemas se basan en la medición de distancias relativas entre puntos de referencia anatómicos en el cuero cabelludo para determinar la ubicación precisa de los electrodos.

El sistema 10-10 se llama así porque utiliza una cuadrícula de ubicaciones de electrodos en la cabeza que están separados por una distancia de 10% o 10% de la distancia total en el plano anteroposterior y mediolateral, como se puede observar en la Figura 4, lo que resulta en una distribución equidistante de los electrodos en todo el cuero cabelludo. Este sistema proporciona una cobertura amplia y es comúnmente utilizado en investigaciones y aplicaciones clínicas [11].

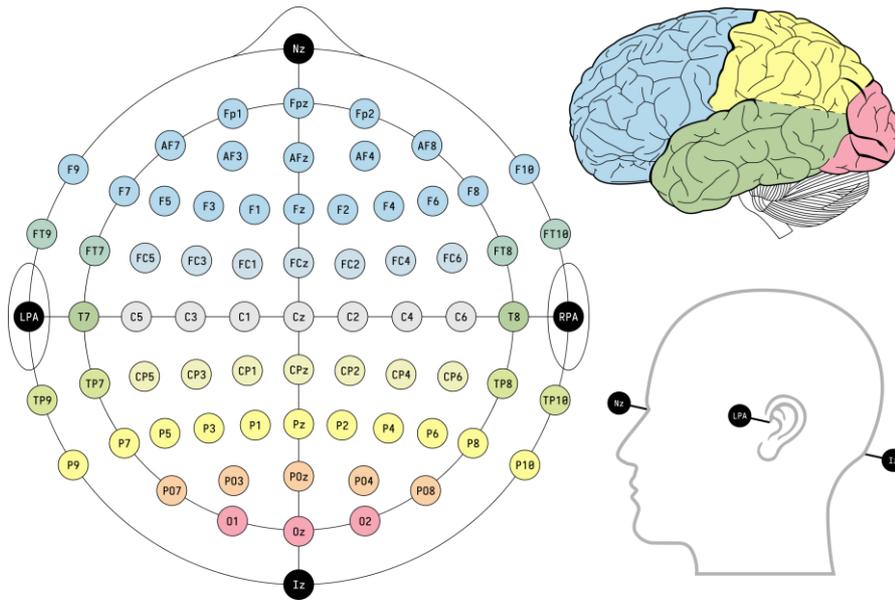


Figura 4. Sistema 10-10, imagen extraída de Wikipedia

Por otro lado, el sistema 10-20 utiliza una separación entre electrodos mayor, lo que implica una menor resolución espacial en la colocación de estos dispositivos [11].

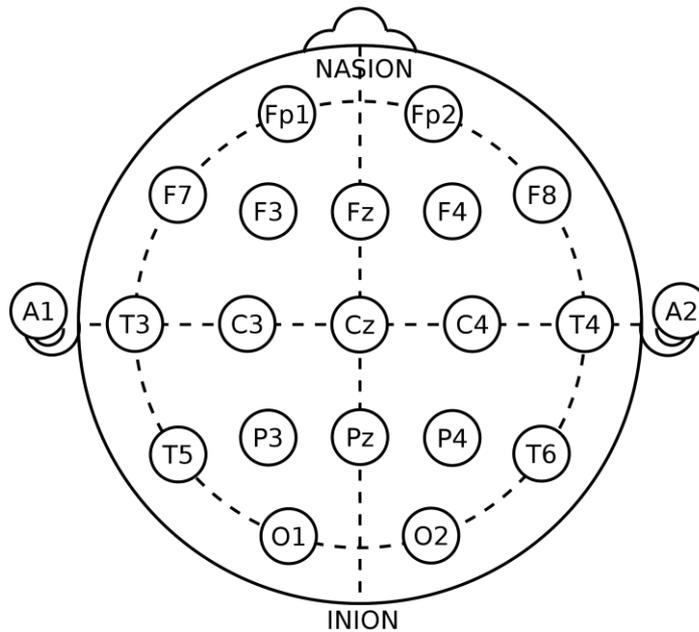


Figura 5. Sistema 10-20, imagen extraída de Wikipedia

Además de estos sistemas, existe el sistema 128-Biosemi, que utiliza una disposición de electrodos distribuidos de forma radial con 128 canales. Este sistema proporciona una alta resolución espacial y permite una cobertura más detallada de la actividad cerebral en comparación con los sistemas 10-10 y 10-20, además de una nomenclatura diferente [12].

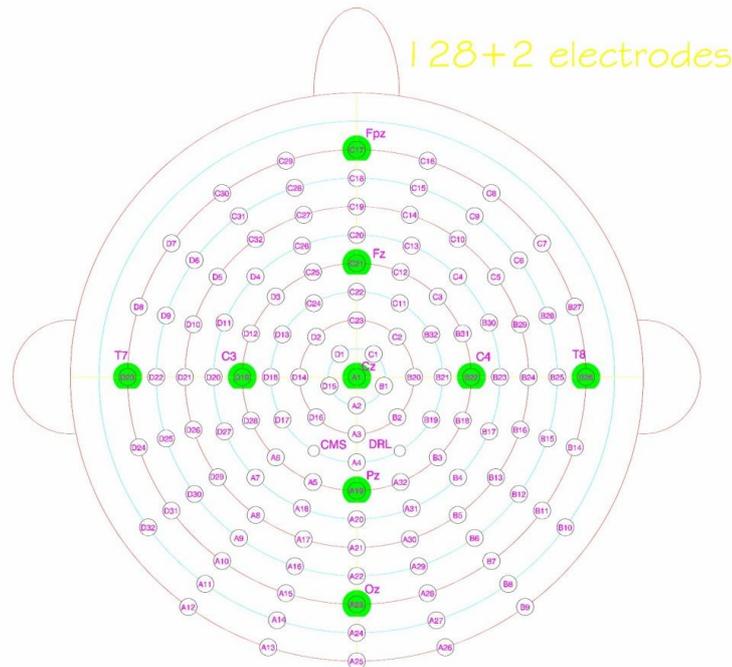


Figura 6. Sistema 128-Biosemi, imagen extraída de la web de Biosemi [12]

2.1.2. Desafíos asociados con la adquisición de señales EEG

La adquisición de señales EEG presenta varios desafíos debido a la presencia de ruido y la variabilidad interindividual. Estos desafíos pueden afectar la calidad y confiabilidad de las mediciones obtenidas. A continuación, se describen brevemente cada uno de estos desafíos:

2.1.2.1. Artefactos en las señales EEG

Durante la adquisición de señales EEG, se pueden producir diversos tipos de ruido que pueden interferir con la señal neuronal de interés. El ruido puede originarse tanto de fuentes internas como externas, como el ruido ambiental, los movimientos musculares y las interferencias eléctricas. El procesamiento y análisis de las señales EEG requieren técnicas adecuadas para reducir y eliminar el ruido y mejorar la calidad de la señal [13].

2.1.2.2. Variabilidad interindividual

Cada individuo tiene una configuración cerebral única, lo que puede dar lugar a diferencias en la amplitud, forma y ubicación de las señales EEG. Esta variabilidad interindividual puede dificultar la comparación de los resultados entre diferentes sujetos y la generalización de los hallazgos. Es importante tener en cuenta estas diferencias individuales al interpretar los resultados de los estudios EEG y considerar estrategias para minimizar su impacto, como el uso de análisis estadísticos adecuados [14].

2.1.3. Técnicas de adquisición y preprocesamiento de señales EEG

A continuación, se detallan las técnicas de adquisición y preprocesamiento de señales EEG que se mencionarán en apartados posteriores.

2.1.3.1. Re-referenciar

La idea detrás de la re-referencia es seleccionar una referencia o punto de referencia adecuado que represente una actividad cerebral mínima o libre de artefactos. La elección de la referencia puede variar dependiendo de los objetivos específicos del estudio y la configuración del electrodo utilizada.

Existen diferentes métodos de re-referencia, pero dos de los enfoques más comunes son:

- **Bipolar re-referencing:** En este enfoque, la señal de cada electrodo se referencia con respecto a la señal de otro electrodo cercano. Al restar las señales de dos electrodos adyacentes, se elimina parte del ruido y las interferencias comunes a ambos electrodos.
- **Promedio de referencia:** En este método, la señal de cada electrodo se promedia con respecto a la señal de todos los demás electrodos. Esto ayuda a eliminar las referencias no deseadas y acentúa las características de interés comunes a todos los electrodos.

El proceso de re-referencia puede realizarse durante la adquisición en tiempo real o como un paso posterior en el análisis de las señales EEG. Es importante tener en cuenta que la elección de la referencia puede afectar los resultados y las interpretaciones de los estudios EEG, y se debe considerar cuidadosamente según el contexto específico.

2.1.3.2. Filtrado digital

En el pre-procesamiento de señales EEG, el filtrado digital es una técnica fundamental para mejorar la calidad de las señales y eliminar el ruido no deseado [14]. Los filtros digitales se utilizan para seleccionar las frecuencias de interés y eliminar aquellas que no son relevantes para el análisis .

Existen diferentes tipos de filtros utilizados en EEG, entre ellos se encuentran:

- **Filtro paso alto (high-pass filter) [13]:** Este tipo de filtro permite el paso de frecuencias superiores a un umbral determinado, eliminando las frecuencias más bajas. El filtro paso alto es útil para eliminar componentes de baja frecuencia, como la actividad basal y los artefactos de movimiento.
- **Filtro paso bajo (low-pass filter) [13]:** Por el contrario, el filtro paso bajo permite el paso de frecuencias inferiores a un umbral dado, atenuando las frecuencias más altas. Se utiliza para eliminar el ruido de alta frecuencia, como el ruido eléctrico ambiental.
- **Filtro de banda (band-pass filter) [14]:** Este tipo de filtro permite el paso de frecuencias dentro de un rango específico, rechazando las frecuencias fuera de ese rango. Se utiliza para seleccionar un rango de frecuencias de interés, como las relacionadas con eventos cerebrales específicos.
- **Filtro de rechazo de banda (notch filter) [14]:** Este filtro está diseñado para eliminar una frecuencia específica, como una frecuencia de línea eléctrica de 50 Hz o 60 Hz. Ayuda a eliminar el ruido eléctrico de la red y otros artefactos específicos [15].

El filtrado digital se realiza mediante algoritmos matemáticos que procesan la señal EEG de manera eficiente y precisa. Es importante tener en cuenta que el filtrado debe aplicarse cuidadosamente para no afectar las características de interés en las señales EEG.

2.1.3.3. Epoching

El epoching en EEG se refiere al proceso de dividir la señal continua de EEG en segmentos más pequeños y discretos llamados "epochs" o "épocas". Cada epoch es una ventana de tiempo que abarca un intervalo específico de duración, generalmente unos pocos cientos de milisegundos a unos pocos segundos.

Se realiza para facilitar el análisis de la señal de EEG, ya que permite examinar características específicas de interés en intervalos de tiempo definidos. Al dividir la señal en epochs, se pueden aplicar técnicas de procesamiento de señales y análisis estadístico a cada epoch individualmente.

Los puntos clave del epoching en EEG son [15]:

- **Segmentación temporal:** La señal continua de EEG se divide en intervalos discretos llamados epochs.
- **Duración del epoch:** Cada epoch tiene una duración específica, que puede variar dependiendo del estudio y el tipo de análisis realizado.

- Análisis individual de epochs: Cada epoch se considera una unidad de análisis independiente, lo que permite examinar características específicas de la señal en intervalos de tiempo definidos.

El epoching en EEG es una etapa crucial en el procesamiento de señales EEG y se utiliza ampliamente en diversos estudios y análisis, como el análisis de ERP y el análisis de oscilaciones cerebrales. Permite una mayor flexibilidad y precisión en el análisis de la actividad cerebral registrada por el EEG.

2.1.3.4. Decimation

La decimación en EEG (o en procesamiento de señales) se refiere al proceso de reducir la tasa de muestreo de una señal de EEG para disminuir la cantidad de datos sin perder información importante. Esto puede ser útil para reducir el tamaño de los archivos de datos y facilitar el procesamiento posterior [16].

Al aplicar la decimación, es importante considerar la frecuencia de corte adecuada para evitar el aliasing, que puede distorsionar la señal de EEG. Se deben aplicar filtros adecuados antes de la decimación para eliminar las frecuencias no deseadas y garantizar la integridad de la señal [17].

La decimación puede introducir cierta pérdida de información en la señal de EEG debido al descarte de muestras. Sin embargo, este impacto puede minimizarse si se realiza de manera adecuada y se selecciona una tasa de decimación apropiada [18]. Esta técnica puede ser especialmente útil en situaciones donde se necesita una menor tasa de muestreo, como en aplicaciones en tiempo real donde se requiere una menor latencia en el procesamiento de la señal de EEG [18].

Es importante considerar los efectos de la decimación en el espectro de frecuencia de la señal de EEG. La reducción en la tasa de muestreo puede afectar la representación espectral y la resolución en frecuencia, lo que debe ser considerado en el análisis posterior [19].

2.1.3.5. Control de actividad ocular y muscular

Es crucial controlar y mitigar los artefactos causados por el movimiento ocular (EOG) y la actividad muscular (EMG). Estos artefactos pueden introducir interferencias en la señal de EEG y dificultar la interpretación precisa de los datos.

- Artefactos de movimiento ocular (EOG) [20]: El movimiento de los ojos genera señales eléctricas que pueden contaminar la señal de EEG. Estos artefactos pueden ser eliminados o reducidos mediante el registro simultáneo de los movimientos oculares, generalmente utilizando electrooculogramas (EOG) o sensores de seguimiento ocular. Estos registros oculares se utilizan luego como señales de referencia para eliminar los artefactos relacionados con el movimiento ocular de la señal de EEG.
- Artefactos de actividad muscular (EMG): La actividad muscular, especialmente en los músculos faciales y del cuello, puede introducir artefactos en la señal de EEG. Estos artefactos pueden ser el resultado de movimientos voluntarios o involuntarios, como parpadeo, masticado o tensión muscular. Para mitigar estos artefactos, se pueden utilizar técnicas de registro de electromiografía (EMG) para monitorear y detectar la actividad muscular concurrente. Al identificar los momentos de actividad muscular y aplicar métodos de eliminación de artefactos, se pueden minimizar los efectos de la actividad muscular en la señal de EEG [21].

-

2.1.3.6. ICA

El ICA, o Análisis de Componentes Independientes, es un método estadístico que se utiliza para separar señales mezcladas en componentes independientes [22]. En el procesamiento de señales EEG, el ICA se utiliza para separar las señales del cerebro en fuentes independientes, como las señales generadas por las diferentes áreas de interés del cerebro y los artefactos causados por movimientos de los ojos, parpadeos, actividad muscular, el corazón y otros ruidos no deseados.

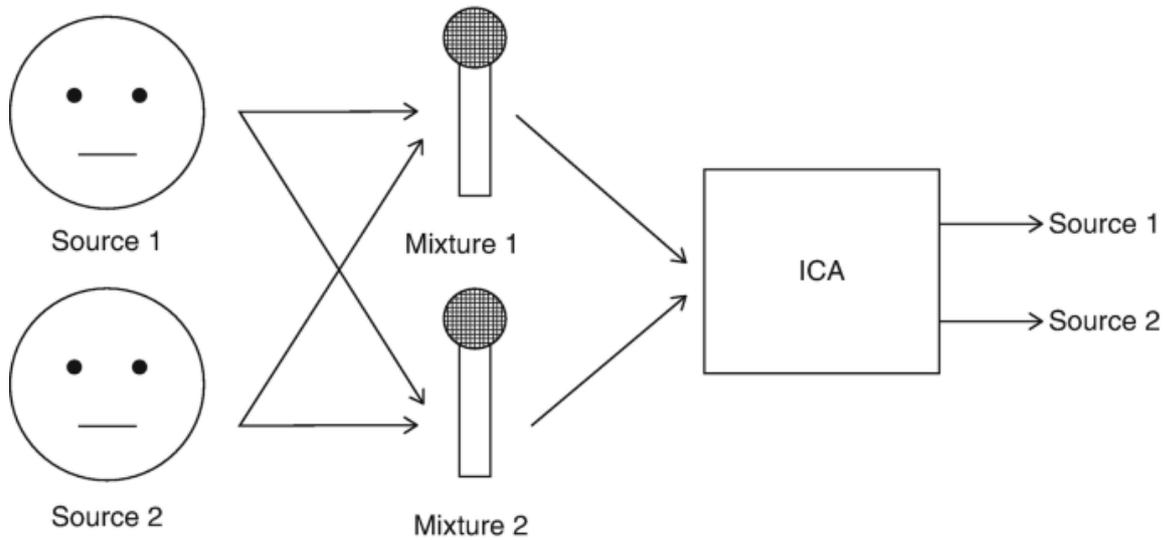


Figura 7. Ilustración del concepto de ICA²

En el modelo de mezcla lineal utilizado en el ICA³, la señal EEG registrada en varios canales se representa como un vector $x = [x_1 x_2 \dots x_n]$ de n dimensiones, donde cada elemento x_i representa la señal en un canal específico. Esta señal mezclada es generada por m fuentes independientes representadas por el vector $s = [s_1 s_2 \dots s_m]$ de m dimensiones:

$$x = As$$

La relación entre la señal mezclada x y las fuentes originales s se puede describir mediante una matriz de mezcla A de dimensiones $n \times m$. La tarea del ICA es encontrar una versión $u = Wx$ de las fuentes originales s , donde W es una matriz cuadrada que especifica filtros espaciales que invierten el proceso de mezcla lineal.

Matemáticamente, el ICA busca identificar la matriz W que se aproxima a la inversa de la matriz de mezcla A^{-1} , de modo que las señales separadas u sean similares a las fuentes originales s , excepto por la escala y la permutación. Esto se puede representar como:

$$W \approx A^{-1}$$

Por lo tanto, la matriz resultante u se puede obtener mediante:

$$u = Wx = W(As) = s' \approx s$$

Se asume que las señales originales s son estadísticamente independientes en cualquier instante de tiempo t , y que la mezcla de estas fuentes es un proceso lineal. El objetivo del ICA es encontrar los filtros espaciales representados por la matriz W que permitan deshacer la mezcla y recuperar las fuentes originales con la mayor independencia posible.

2.2. Procesamiento de señales, extracción y selección de características

Las transformadas wavelets son herramientas matemáticas utilizadas en el análisis de señales e imágenes para capturar características que varían en diferentes escalas [23]. A diferencia de la transformada de Fourier, que se basa en sinusoides de frecuencias fijas, las transformadas wavelets descomponen la señal en ondículas, que son funciones oscilantes localizadas tanto en el dominio del tiempo como en el de la frecuencia [23].

La transformada wavelet continua (CWT) es adecuada para analizar señales no estacionarias, es decir, aquellas cuyas propiedades varían con el tiempo. La CWT mapea el plano tiempo-frecuencia utilizando ventanas de tamaño variable, lo que permite capturar tanto componentes de alta frecuencia como de baja frecuencia de manera precisa. Esto la hace útil para analizar señales con transitorios, frecuencias cambiantes y tendencias

² Imagen extraída de https://link.springer.com/referenceworkentry/10.1007/978-1-4939-7131-2_147

³ Concepto de ICA, extraído y modificado de [22]

lentas [23], ejemplo de ello son las wavelets de Morlet.

Por otro lado, la transformada wavelet discreta (DWT) realiza una descomposición en escalas discretas, lo que la convierte en una herramienta efectiva para el análisis multirresolución y la compresión de señales e imágenes. La DWT permite separar los componentes de la señal en diferentes niveles de detalle, lo que facilita la eliminación de ruido y la preservación de características importantes [23]. Será el caso de las wavelets de Daubechies-4.

Es decir, las transformadas wavelets ofrecen una representación más versátil y flexible que la transformada de Fourier, ya que pueden capturar características locales en el tiempo y la frecuencia.

2.2.1. Morlet Wavelet

Los análisis Tiempo-Frecuencia utilizan una onda sinusoidal que varía en el tiempo y se denomina wavelets. Estas son oscilaciones cuya amplitud comienza en cero, aumenta y luego disminuye. Las wavelets de Morlet son los más utilizados en este contexto [24]. La transformada wavelet de Morlet es una función compleja que combina una función sinusoidal modulada (portadora) por una función gaussiana (envolvente).

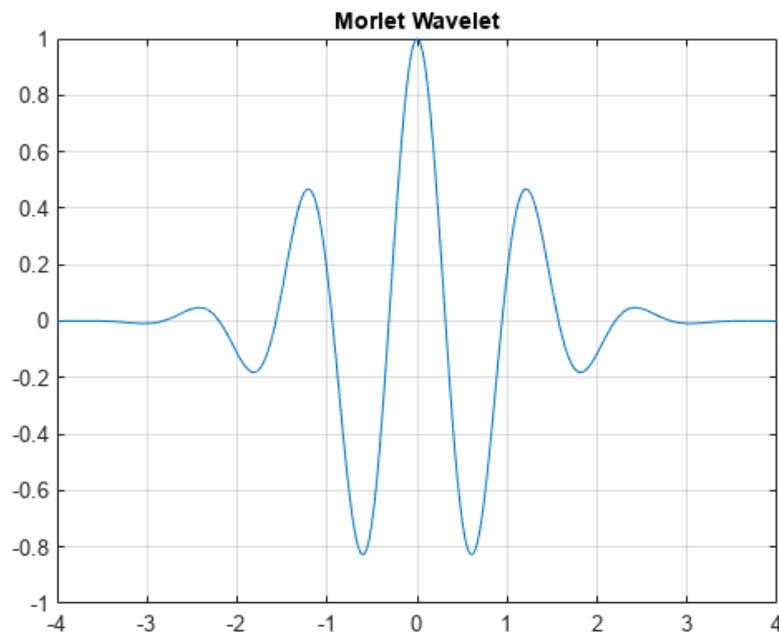


Figura 8. Morlet Wavelet, imagen extraída de la web de MATLAB®

Se define por su forma matemática [25], que involucra dos parámetros principales: la frecuencia central (frecuencia portadora) f y la desviación estandar temporal σ_t . Estos parámetros determinan la resolución en frecuencia y tiempo de la wavelet, respectivamente. En [25], se obtienen estimaciones espectrales utilizando las wavelets de Morlet $w(t, f)$:

$$w(t, f) = (\sigma_t \sqrt{\pi})^{-\frac{1}{2}} e^{-\frac{1t^2}{2\sigma_t^2}} e^{-i2\pi ft}$$

Luego, se calcula la estimación de tiempo-frecuencia $X(t, f)$ de una señal $x(t)$ mediante la convolución con $w(t, f)$:

$$X(t, f) = x(t) * w(t, f)$$

La traslación en el tiempo de la wavelet de Morlet permite analizar diferentes partes de la señal en diferentes momentos. Al desplazar la wavelet en el tiempo, se obtienen diferentes representaciones de la señal en diferentes escalas temporales.

La función de Morlet proporciona una representación localizada en el tiempo y en la frecuencia, lo que permite detectar eventos breves y transitorios en una señal EEG, siendo así es especialmente útil para analizar señales no estacionarias, como señales que varían en el tiempo, como es el caso de muchas señales cerebrales. Sin embargo, esta estimación no será utilizada en nuestro desarrollo en MATLAB® (de lo que se hablará posteriormente) sino que se implementará con la ayuda de la función *cwtfilterbank()* para calcular tanto las bandas de frecuencias (θ , α , β , γ) como los escalogramas frecuencia-tiempo de las señales EEG, teniendo en cuenta que se emplea una ondícula compleja.

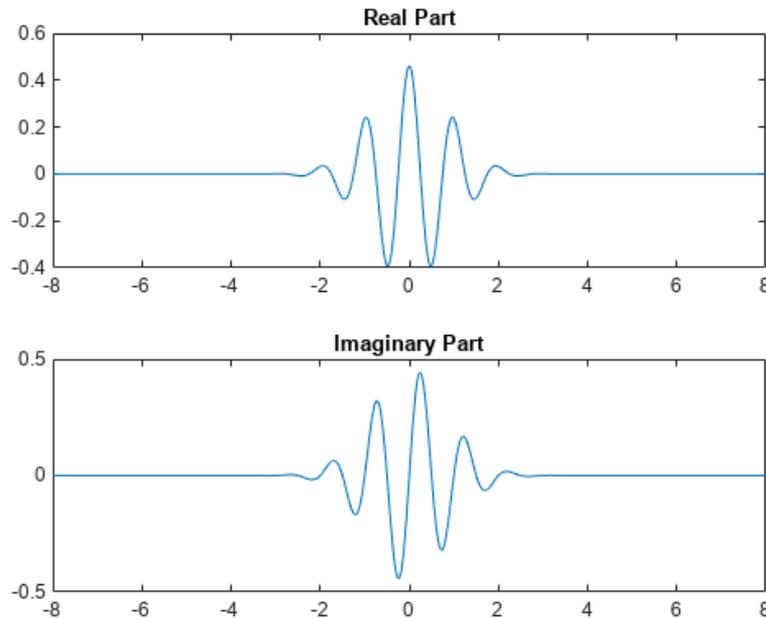


Figura 9. Wavelet de Morlet compleja, imagen extraída de la web de MATLAB®

2.2.2. Daubechies 4 (D4)

Las wavelets de Daubechies son una familia de wavelets que se utilizan ampliamente en el procesamiento de señales y la compresión de datos. La Daubechies 4 (D4), llamada así por formarse con 4 coeficientes, es una de las wavelets más simples y populares de esta familia.

La wavelet D4 se define en términos de una función de escalado y una función de wavelet [26]. Estas funciones son utilizadas en la descomposición y reconstrucción de una señal en diferentes niveles de detalle.

La función de escalado $\phi(t)$ se utiliza para aproximar los componentes de baja frecuencia de la señal original [27]. Esta función es una versión dilatada y desplazada de sí misma, lo que significa que se extiende a lo largo del tiempo y se desplaza para cubrir diferentes partes de la señal. La función de escalado tiene la propiedad de ser ortogonal, lo que implica que la energía de la señal original se conserva en cada nivel de descomposición.

La función de wavelet $\psi(t)$ se utiliza para capturar los detalles de alta frecuencia de la señal original [27]. Al igual que la función de escalado, la función de wavelet también está dilatada y desplazada, pero tiene características oscilatorias que permiten detectar cambios rápidos en la señal. La función de wavelet tiene la propiedad de ser ortogonal a la función de escalado, lo que significa que los detalles de la señal se capturan de forma independiente de la aproximación de baja frecuencia [26].

La descomposición utilizando la wavelet D4 implica pasar la señal original a través de un proceso de filtrado y submuestreo. En cada nivel de descomposición, se aplican filtros de convolución tanto a la aproximación de baja frecuencia (AC, del inglés Approximation Coefficients) como a los detalles de alta frecuencia (DC, del inglés Detail Coefficients) obtenidos en el nivel anterior [28]. Esto permite separar los diferentes componentes de la señal en diferentes niveles de detalle. La descomposición se repite iterativamente hasta alcanzar el nivel de descomposición deseado [28].

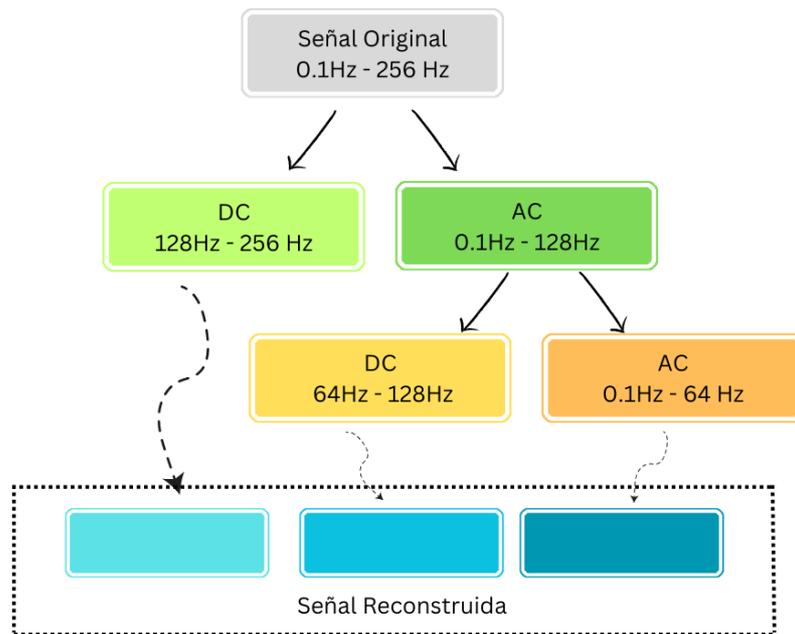


Figura 10. Descomposición mediante D4

La reconstrucción de la señal implica el proceso inverso, donde se utilizan los coeficientes de detalle y la aproximación de baja frecuencia obtenidos en la descomposición para reconstruir la señal original. Esto se logra mediante la aplicación de filtros de convolución inversa y la combinación de las diferentes aproximaciones y detalles en cada nivel. Será especialmente útil en la extracción de características del dataset de MindBigData, ya que se replicará el experimento realizado en [28], donde se aplica un umbral de energía a cada coeficiente, eliminando aquellos que sean pequeños, considerándose esencialmente ruido.

2.3. Redes neuronales para el reconocimiento de palabras

Para clasificar las características extraídas de la señal EEG, los investigadores han utilizado tanto algoritmos clásicos de Machine Learning como de Deep Learning. Ambos son métodos que proporcionan a los ordenadores la capacidad de aprender y reconocer patrones. En el caso de la BCI, los patrones que hay que reconocer son las características extraídas de las ondas del EEG y, a continuación, basándose en lo aprendido por el ordenador, se realizan algunas predicciones para clasificar las señales.

Gracias al estudio realizado en [1], se puede resumir que:

Se han utilizado varias técnicas clásicas de Machine Learning para abordar la descodificación imaginaria del habla para sistemas BCI basados en EEG. Algunos de los algoritmos más comunes incluyen el Análisis Discriminante Lineal (LDA), Máquinas de Vectores Soporte (SVM), Random Forests (RF), k-Nearest-Neighbors (kNN) y Naive Bayes.

Además, los enfoques de Deep Learning han adquirido recientemente un gran protagonismo en el reconocimiento del habla imaginada. Algunas de estas técnicas son la Máquina de Aprendizaje Extremo (ELM), Redes Neuronales Convolucionales (CNN), Redes Neuronales Recurrentes (RNN), y CNN+RNN paralelas.

2.4. Resultados de estudios en la literatura

En base a todo lo mencionado en esta sección 2, existen diversos estudios que intentan lograr un alto rendimiento del habla imaginaria mediante el uso de sistemas BCI, lo que involucra a diversos métodos de extracción y clasificación de características, así como diferentes tareas. En Tabla 1, Tabla 2 y Tabla 3 se resume los experimentos más relevantes encontrados en la literatura para 'Thinking Out Loud' [3], 'MindBigData' y otros datasets, respectivamente.

A partir de este punto en adelante, se utilizarán indistintamente los términos "Accuracy" y "precisión" para referirse al mismo concepto en el contexto de este estudio. Ambos términos se emplearán para describir la medida de la exactitud o acierto en las tareas de clasificación realizadas, definida como:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

donde:

- TP (True Positives) representa el número de casos positivos correctamente clasificados (verdaderos positivos).
- TN (True Negative) representa el número de casos negativos correctamente clasificados (verdaderos negativos).
- FP (False Positive) representa el número de casos negativos incorrectamente clasificados como positivos (falsos positivos).
- FN (False Negative) representa el número de casos positivos incorrectamente clasificados como negativos (falsos negativos).

Esta equivalencia en la terminología tiene como objetivo evitar confusiones y facilitar la comprensión de los resultados y conclusiones presentados en el presente trabajo.

Autor/es	Área/s Cerebral/es	Red Neuronal	Tarea	Accuracy
Bram van den Berg et al. [29]	Todas las áreas (128 Biosemi)	2D CNN basada en la arquitectura EEGNet	Clasificación de 4 palabras en castellano: 'arriba', 'abajo', 'izquierda', 'derecha'	29.7%
F. Gasparini et al. [30]	Todas las áreas (128 Biosemi)	LSTM	Clasificación de 4 palabras en castellano: 'arriba', 'abajo', 'izquierda', 'derecha'	30.4%
F. Gasparini et al. [30]	Todas las áreas (128 Biosemi)	BiLSTM	Clasificación de 4 palabras en castellano: 'arriba', 'abajo', 'izquierda', 'derecha'	36.1%

Tabla 1. Estudios en la literatura para el dataset 'Thinking Out Loud' [3]

Autor/es	Área/s Cerebral/es	Red Neuronal	Tarea	Accuracy
Bird et al. [31]	FP1, FP2, TP9, TP10 (MUSE)	Attribute selection + MLP	Clasificación de 10 dígitos: 0-9	30%
Jolly et al. [32]	FP1, FP2, TP9, TP10 (MUSE)	Unidirectional-GRU	Clasificación de 10 dígitos: 0-9	33.8%
Pratama et al. [33]	FP1, FP2, TP9, TP10 (MUSE)	FFT + k-NN	Clasificación de 10 dígitos: 0-9	31%
Mahapatra NC & Bhuyan P [28]	FP1, FP2, TP9, TP10 (MUSE)	D4 + bidirectional-LSTM	Clasificación de 10 dígitos: 0-9	96.2%

Tabla 2. Estudios en la literatura para el dataset 'MindBigData' [34]

Autor/es	Área/s Cerebral/es	Red Neuronal	Tarea	Accuracy
García et al. [34]	F7, FC5, T7, P7 (Área de Wernicke)	Naive Baye classifier, Random Forest, SVM and Bagging-RF	Clasificación de 5 palabras en castellano	>20%
Sarmiento et al. [35]	Áreas de Broca y Wernicke	SVM	Clasificación de 5 vocales: /a/, /o/ (abierta), /e/, /i/, /u/ (cerrada)	84-94%
Pawar & Dhage [36]	Cortex prefrontal, Área de Broca, Área de Wernicke	Kernel Based extreme machine learning	Clasificación de 4 palabras en inglés: 'left', 'right', 'up', 'down'	multi-class: 49.8%, binary: 85.5%
Cooney et al. [37]	Todas las áreas	1. CNN 2. ICA/LDA	1. Clasificación de 5 vocales: /a/, /e/, /i/, /o/, /u/ 2. Clasificación de 6 palabras en castellano: 'izquierda', 'derecha', 'arriba', 'abajo', 'adelante', 'atrás'	1. 35% 2. 30%

Tamm et al. [38]	F3, F4, C3, C4, P3, P4	CNN	Clasificación de 5 vocales y 6 palabras	24%
-------------------------	---------------------------	-----	--	-----

Tabla 3. Estudios en la literatura para otros datasets

3. DISEÑO E IMPLEMENTACIÓN DE LA BCI

El verdadero método de conocimiento es el experimento.

- William Blake -

En el marco de este estudio, se realizaron dos tipos de análisis de señales cerebrales. En primer lugar, se hace uso de una adquisición de señales propia y específica para el reconocimiento de palabras mediante EEG, estudiando la viabilidad de un diseño relativamente sencillo de una diadema 3D que busque facilitar esta captura. Estas señales fueron registradas bajo condiciones controladas (como el uso de una cámara anecoica), lo que aseguró una consistencia en la forma en que se obtuvieron y procesaron.

En segundo lugar, se emplearon conjuntos de datos previamente recopilados y disponibles en la literatura científica. Estos conjuntos de datos proporcionaron una fuente adicional de señales cerebrales registradas en distintos contextos y por diferentes grupos de investigación. Al utilizar estos conjuntos de datos, se amplió la muestra de datos y se evaluó el rendimiento de las técnicas y modelos en condiciones y escenarios más variados.

3.1. Sistemas de adquisición de señales

3.1.1. Detalles de las señales propias

El presente trabajo tiene como uno de sus objetivos estudiar la viabilidad de un diseño relativamente sencillo de una diadema 3D que busque facilitar una adquisición de señales propia y específica para el reconocimiento de palabras mediante EEG. Esta se retiene a la cabeza del sujeto mediante una cinta elástica ajustable mediante cierre de gancho y bucle.

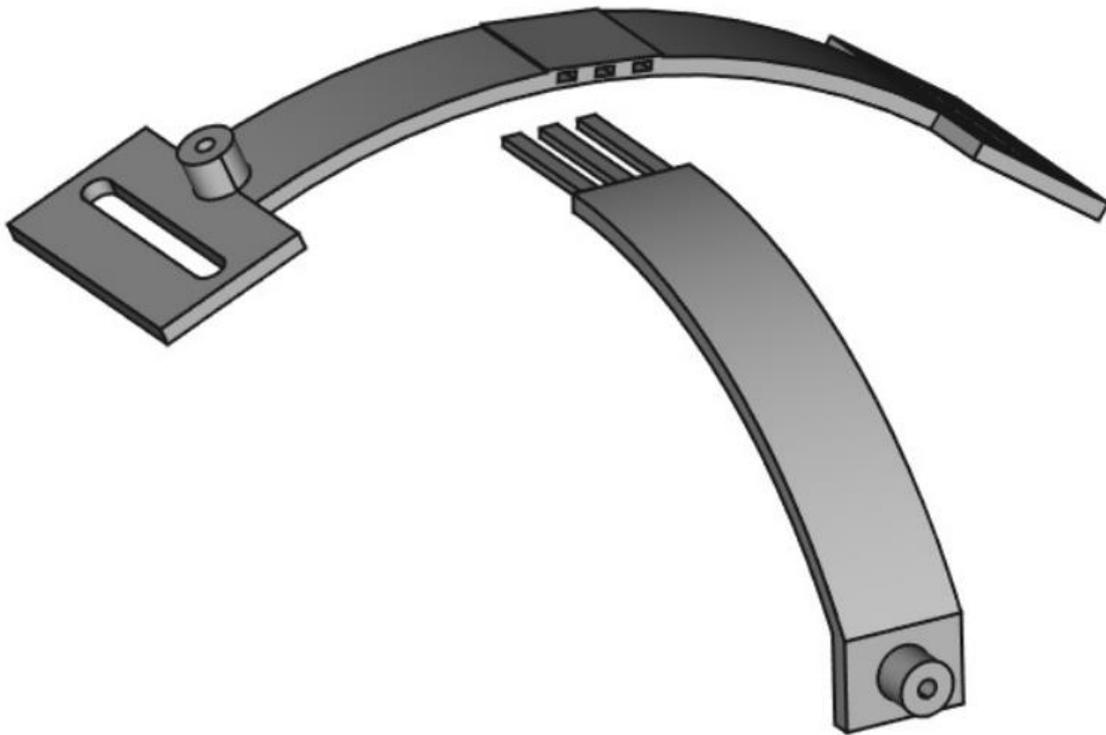


Figura 11. Diseño 3D de la diadema para experimento propio

3.1.1.1. Diseño del experimento

Por ello, será clave la elección estratégica de las posiciones de un número reducido de electrodos. Al emplear un número limitado de electrodos estratégicamente ubicados, se reduce la complejidad del sistema de adquisición y se minimiza la interferencia de señales no deseadas. Esto contribuiría a mejorar la calidad de las señales EEG registradas y simplifica el proceso de análisis y reconocimiento de palabras.

En particular, se han seleccionado las posiciones Fpz y F3 para la medida diferencial de EEG, y A1 como referencia. Se ha decidido así en base al área de Broca y su relevancia en la formación de palabras en el procesamiento del lenguaje, como se explica en el apartado 1.2.3.1. Las señales adquiridas mediante estos tres electrodos serán monitorizadas por ‘EEG click’ de Mikrobus®, un dispositivo basado en el amplificador de instrumentación INA114, de Texas Instruments.

Ganancia	Hasta 10.000 (ajustable mediante potenciómetro)
GND virtual	2,048 V
Conector	Jack 3.5mm

Tabla 4. Características placa ‘EEG Click’



Figura 12. Dispositivo ‘EEG Click’, imagen extraída de su datasheet

Tras esta etapa de amplificación, la señal diferencial es recibida por la placa Atmel® SAM E70 Xplained, cuyo microcontrolador es el potente Atmel ATSAME70Q21. En ella se digitaliza la señal con una frecuencia de muestreo de 256Hz y es enviada al software de MATLAB mediante puerto serie, donde se aplicará un filtro Notch a 50Hz para eliminar el ruido de alimentación y se aplicarán las wavelets de Morlet en tiempo real para evitar el conocido efecto de borde [39].



Figura 13. Atmel® SAM E70 Xplained, imagen extraída de su datasheet

El experimento realizado se basará en el diseñado en ‘Thinking Out Loud’⁴ [3], donde se pretende distinguir cuatro palabras en castellano: ‘izquierda’, ‘derecha’, ‘arriba’, ‘abajo’. Sin embargo, únicamente se ejecutará el paradigma ‘inner speech’. Las grabaciones han sido registradas tanto en el Laboratorio de Máster del departamento como en una cámara anecoica disponible en la Escuela, diseñada para absorber en su totalidad las reflexiones producidas por ondas acústicas o electromagnéticas. El fin de la realización de las grabaciones en esta sala es el de eliminar cualquier ruido externo, pues se trabaja con amplitudes de onda muy pequeñas.

⁴ Ver la sección 3.1.2.1 Thinking Out Loud

3.1.1.2. Implementación del experimento en MATLAB

El experimento está diseñado mediante el MATLAB App Designer. Consta de una primera interfaz (ver Figura 14) donde se comprueba la correcta recepción de datos por puerto serie y se activa el filtro Notch a 50Hz. Una vez chequeado lo anterior, se inicia el experimento mediante el botón 'Start'.

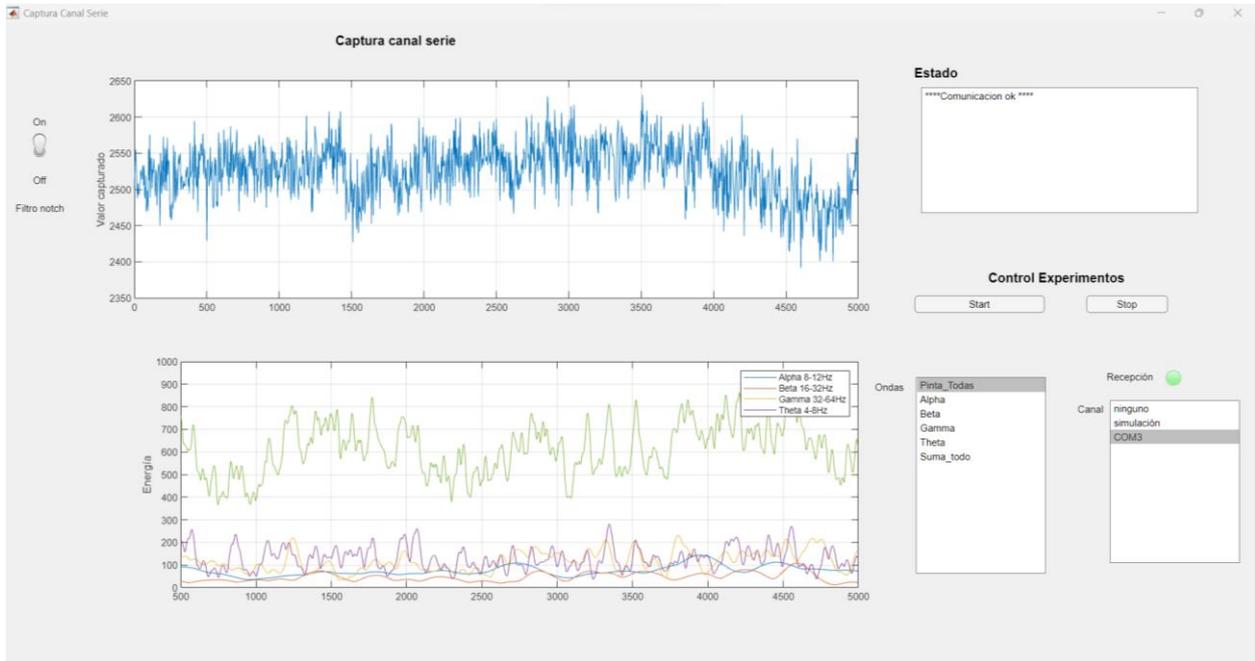


Figura 14. Primera interfaz del experimento propio

Una vez lanzado el experimento, se inicia el primer intervalo de concentración, apareciendo una pantalla inicia con el epígrafe mostrado en la Figura 15, con el fin de que el sujeto que realiza la grabación se concentre en la tarea. A continuación, se sigue la secuencia representada en la Figura 16.

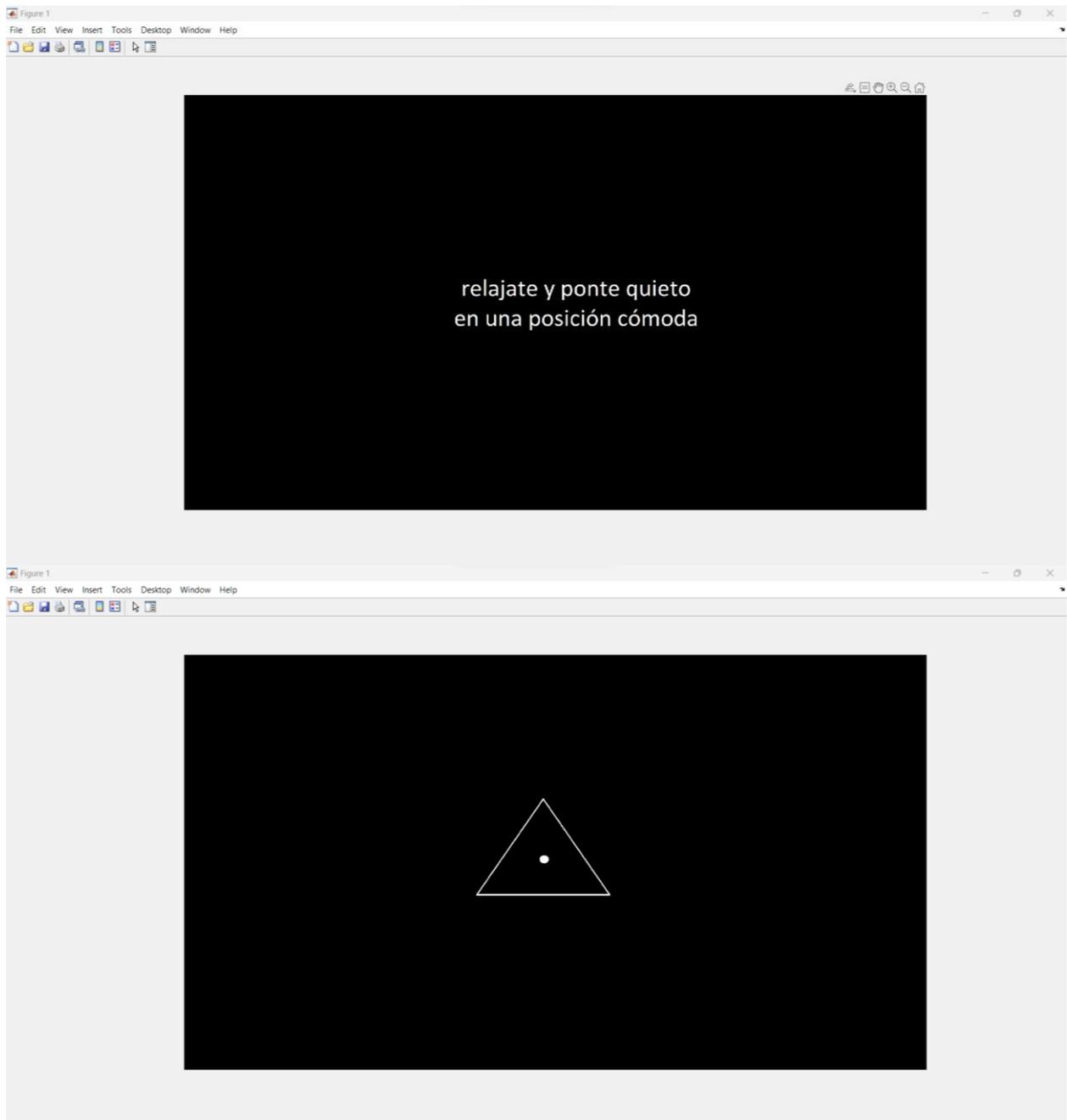


Figura 15. Pantalla inicial y 'Cue Interval' del experimento propio

3.1.2. Descripción de los datasets en la literatura utilizados

El primer conjunto de datos bibliográficos utilizado proviene del estudio "Thinking Out Loud" [3]. Por otro lado, también se analizará el dataset de "MindBigData" [2], concretamente las señales EEG adquiridas con el dispositivo comercial Interaxon Muse [40].

3.1.2.1. Thinking Out Loud

En este experimento, participaron diez sujetos sanos sin pérdida auditiva ni del habla, sin experiencia previa en BCI [3]. El experimento consistió en la realización de los tres paradigmas descritos en la sección 2.a: "inner speech" (habla interna), "pronounced speech" (habla pronunciada) y "visualized condition" (condición

visualizada).

Cada sujeto participa en tres sesiones consecutivas, cada una de ellas estructurada por un registro de línea base de 15 segundos, seguido de la ejecución del paradigma de "pronounced speech", dos ejecuciones del paradigma de "inner speech" y dos ejecuciones del paradigma de "visualized condition".

Es importante destacar que el orden de las palabras en español seleccionadas ("arriba", "abajo", "derecha" e "izquierda") se presenta de manera aleatoria durante los experimentos.

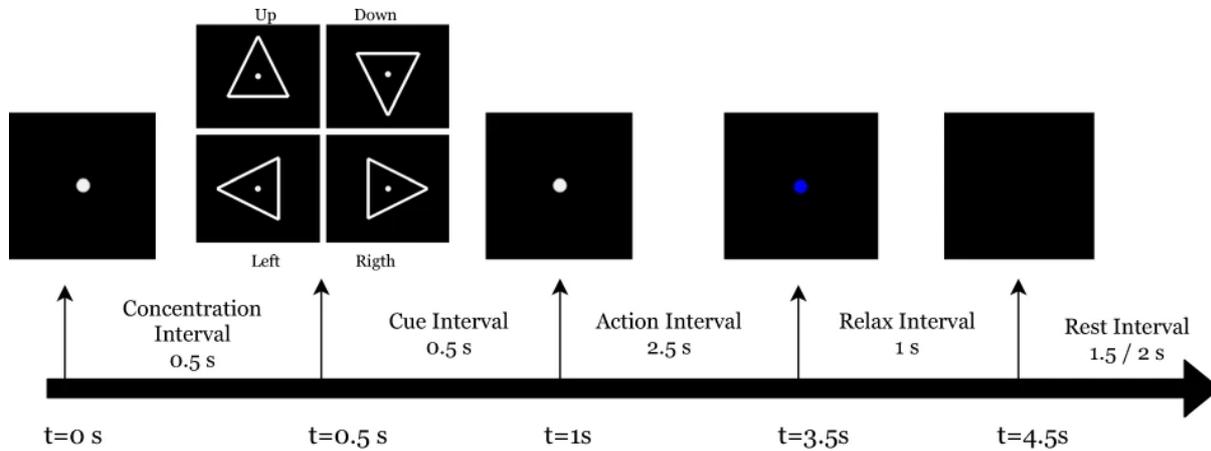


Figura 16. Secuencia de cada trial, extraída del artículo original de Thinking Out Loud [3].

Durante el experimento, se muestra un círculo blanco en el centro de la pantalla y se le pide al sujeto que lo mire fijamente sin parpadear. Luego, se presenta un triángulo blanco apuntando en una de las cuatro direcciones correspondientes a las palabras españolas seleccionadas. Cuando el triángulo desaparece y se vuelve a presentar el círculo blanco, el sujeto debe realizar la tarea indicada. La ejecución de la tarea se detiene cuando el círculo blanco se vuelve azul. Se instruye al sujeto a controlar el parpadeo de los ojos hasta que el círculo desaparezca.

Para evaluar la atención de los participantes, se les solicita que indiquen la dirección del último estímulo después de un número aleatorio de ensayos. Las respuestas se darán utilizando las flechas del teclado y la respuesta se muestra en la pantalla.

Los datos se adquieren utilizando un sistema de 128 electrodos húmedos activos para EEG (sistema 128 Biosemi) y otros 8 externos para EOG/EMG. Se emplea una resolución de 24 bits y una frecuencia de muestreo de 1024 Hz. Previo a su análisis, los autores del experimento realizaron el preprocesamiento de las señales EEG de la siguiente forma [3]:

- Referencia: Debido a que el sistema de adquisición BioSemi es "libre de referencia", se realizó un paso de re-referencia utilizando los canales EXG1 y EXG2 para eliminar el voltaje de modo común y reducir el ruido de línea y las derivas del potencial corporal.
- Filtrado digital: Se aplicó un filtro pasa banda de respuesta finita al impulso con límites de frecuencia de 0.5 a 100 Hz para mantener los datos lo más sin procesar posible, permitiendo a los usuarios futuros filtrar los datos según sus necesidades. También se aplicó un filtro Notch en 50 Hz para eliminar el ruido de línea.
- Segmentación y decimación: Los datos se deciman cuatro veces para obtener una frecuencia de muestreo final de 254 Hz. Luego, se segmentan los datos continuos registrados, conservando solo las señales de 4.5 segundos correspondientes al intervalo de concentración y relajación.
- Análisis ICA: Se aplica el Análisis de Componentes Independientes (ICA) a los canales de EEG utilizando la implementación de MNE del ICA infomax. Se capturan 128 fuentes y se utiliza la correlación con los canales EXG para identificar y excluir las fuentes relacionadas con el parpadeo, la mirada y el movimiento de la boca, obteniendo así el conjunto de datos final.
- Control EMG: El enfoque más simple para detectar la actividad electromiográfica es estableciendo un umbral de energía para estas señales. Este control se utiliza para identificar y descartar los ensayos en los que se detecta actividad muscular no deseada, lo cual es importante para garantizar la calidad de los

datos en el estudio.

3.1.2.2. MindBigData

En este experimento, participa un único sujeto sano. Se capturan las señales cerebrales mientras el sujeto ve durante 2 segundos un solo dígito (mostrado en una pantalla de televisión de 65" con fuente blanca sobre un fondo negro completo) del 0 al 9, como se muestra en la Figura 17.

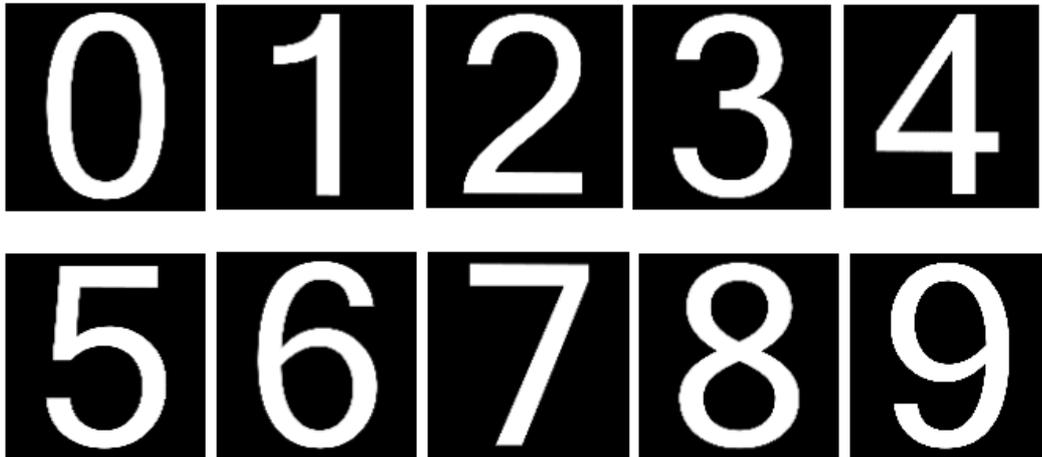


Figura 17. Dígitos mostrados en 'MindBigData', extraído de su artículo original [41].

La aparición de los dígitos es aleatoria, con un intervalo de descanso entre ellos (pantalla negra). Se graban un total de 163,250 señales cerebrales de 2 segundos cada una, etiquetadas con el dígito correspondiente. También se capturan 59,339 señales mientras el sujeto realizaba otras actividades mentales no relacionadas con los dígitos, sin limitación de movimientos, parpadeos u otras acciones, y estas se etiquetan como "-1" [41].

Los canales de EEG capturados son cuatro (FP1, FP2, TP9 y TP10, del sistema 10-10) y se emplea una frecuencia de muestreo teórica de 220 Hz [41].

La señal mide la variación de los voltajes de las actividades neuronales cerebrales y se proporcionan en "bruto", tal y como se captura del dispositivo. El autor no aplica ningún preprocesamiento, aunque se decide aplicar Daubechies-4 y un umbral de energía a cada descomposición para eliminar componentes irrelevantes (convirtiéndolos en nulos si no superan el umbral), como se propone en [28]. Cabe destacar que, en el presente estudio, solamente se realiza una clasificación binaria entre los dígitos '0' y '1'.

3.2. Procesamiento de señales, extracción y selección de características

Esta etapa tiene como finalidad preparar los datos, extrayendo características, para su posterior análisis y clasificación en una red neuronal, a la que introduciremos las mencionadas características, 'inputs'. Para ello, emplearemos las señales EEG, bandas de frecuencias o espectrogramas.

2.4.1. Señales EEG como inputs

Las señales EEG (electroencefalografía) serán utilizadas directamente como datos de entrada para el reconocimiento de palabras. Tanto para las señales EEG propias como para las señales provenientes de otras fuentes como "Thinking Out Loud" [3] y "MindBigData", se realiza un procesamiento previo. Este procesamiento incluye el filtrado, re-referenciación, segmentación y otras técnicas para mejorar la calidad de las señales y resaltar las características relevantes para el reconocimiento de palabras explicados en el apartado 3.1. Sistemas de adquisición de señales.

2.4.2. Bandas de frecuencia y espectrogramas como inputs

En este estudio, se emplearán como entrada para las redes neuronales espectrogramas generados mediante, como ya se ha comentado anteriormente, una de las técnicas más utilizadas en el procesamiento de las señales EEG: la transformada de Morlet.

La wavelet Morlet es especialmente adecuada para calcular la energía de las bandas de frecuencia utilizadas en el experimento propio, ya que evita el efecto de borde (distorsión en los extremos de la señal) al aplicar la transformada en tiempo real, utilizando así la parte de la señal que está fuera del intervalo de interés. Esto significa que se puede obtener una representación más precisa de las bandas de frecuencia presentes en las señales EEG (θ , α , β , γ). También se aplicará a los datasets existentes en la literatura que se estudian en este trabajo, ‘Thinking Out Loud’ [3] y ‘Mindbigdata’, teniendo en cuenta que sí existirá el efecto de borde al usar señales ya segmentadas [39].

Además de las bandas de frecuencia, también se utilizarán espectrogramas generados a partir de las señales EEG como datos de entrada [42]. Estos espectrogramas proporcionan información visual sobre la distribución de energía en diferentes frecuencias a lo largo del tiempo.

3.3. Implementación de redes neuronales

En este apartado se muestran los distintos enfoques y arquitecturas exploradas en este proyecto con el objetivo de desarrollar un sistema eficiente y preciso para el reconocimiento de palabras a partir de señales EEG. Cabe destacar el uso de la técnica conocida como ‘Transfer Learning’ [43] para la adaptación de las redes neuronales existentes en la literatura.

3.3.1. Aprendizaje profundo para el reconocimiento de palabras mediante audio

Se utiliza la red neuronal propuesta en [43] para el reconocimiento de palabras a partir de señales de audio, adaptada al procesamiento de señales EEG.

Dentro de este enfoque, se presentan tres variantes:

3.3.1.1. Dataset propio

En esta variante, utilizaremos exclusivamente señales EEG provenientes de la sala anecoica para el entrenamiento y validación.

Cada ensayo consiste en un canal y seis bandas, que incluyen cinco bandas de frecuencias y la señal EEG.

3.3.1.2. Thinking Out Loud

En esta variante, se utilizan tanto las señales EEG directamente (con 128 canales y una banda por ensayo) como las bandas de frecuencia en un canal separado (un canal y cinco bandas por ensayo). Se realizan pruebas con los tres paradigmas tanto indistintamente como por separado, resultando el paradigma de ‘Inner Speech’ el más relevante para nuestro análisis.

3.3.1.3. MindBigData

En esta variante, se utilizarán tanto las señales EEG directamente (con 4 canales y una banda por ensayo) como las bandas de frecuencia en un canal separado (un canal y cinco bandas por ensayo). Se realizará únicamente un estudio preliminar con la clasificación binaria entre los dígitos ‘0’ y ‘1’.

3.3.2. Arquitectura típica de red LSTM

Una red LSTM (Long Short-Term Memory) es un tipo de red neuronal recurrente (RNN) que es capaz de aprender dependencias a largo plazo entre unidades de tiempo en datos secuenciales.

La arquitectura de una red LSTM consta de una capa de entrada de secuencias y una capa LSTM. La capa de entrada de secuencias permite introducir datos secuenciales o de series de tiempo en la red, mientras que la capa LSTM aprende las dependencias a largo plazo entre las unidades de tiempo de los datos secuenciales.

En la Figura 18, se muestra la arquitectura de una red LSTM sencilla para clasificación. La red comienza con una capa de entrada de secuencias seguida de una capa LSTM. Para predecir las etiquetas de clase, la red finaliza con una capa totalmente conectada, una capa softmax y una capa de clasificación de salida.

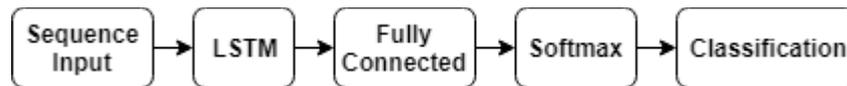


Figura 18. Diagrama de una red típica LSTM⁵

Dentro de esta arquitectura, se presentan los mismos datos que en el apartado 3.3.1. Aprendizaje profundo para el reconocimiento de palabras mediante audio.

3.3.3. Red convolucional de tiempo-frecuencia para la clasificación de datos EEG

Se adapta una red para clasificar series temporales de electroencefalogramas (EEG) de personas con y sin epilepsia en el dominio tiempo-frecuencia. La red convolucional (TF CNN) utiliza la transformada continua de wavelet (CWT) para representar los datos EEG en el dominio tiempo-frecuencia. A diferencia de otras redes que utilizan el espectrograma como paso de preprocesamiento, esta red utiliza una capa de espectrograma diferenciable, lo que permite aplicar operaciones aprendibles antes y después del espectrograma. Esta capa amplía las posibilidades arquitectónicas al trabajar con transformadas tiempo-frecuencia.

Dentro de esta arquitectura, se presentan los mismos datos que en el apartado 3.3.1. Aprendizaje profundo para el reconocimiento de palabras mediante audio, excluyendo las bandas de frecuencia. Esto se debe a que la red hace uso de la CWT en una de sus capas y carecería de sentido aplicarla sobre una transformada Tiempo-Frecuencia.

3.3.4. GoogleNet y AlexNet

GoogleNet y AlexNet son dos arquitecturas de redes neuronales convolucionales (CNN) ampliamente reconocidas en el campo del reconocimiento de imágenes.

Al utilizar los espectrogramas generados a partir de señales de EEG como input, podremos aprovechar las capacidades de aprendizaje profundo de GoogleNet y AlexNet para extraer características significativas y realizar la clasificación de palabras basada en los patrones identificados en los espectrogramas.

Esta elección de representación de entrada podría permitir aprovechar la capacidad de las redes neuronales convolucionales para reconocer patrones espaciales y frecuenciales en los datos de EEG, lo cual puede mejorar la precisión del reconocimiento de palabras en nuestro contexto específico.

Es importante destacar que tanto GoogleNet como AlexNet han sido ampliamente estudiadas y utilizadas en la comunidad científica, y han demostrado buen rendimiento en diversas tareas de clasificación de imágenes en conjuntos de datos como ImageNet. Estas arquitecturas son adaptadas en el Deep Network Toolbox de MATLAB para el reconocimiento de palabras a partir de espectrogramas de señales de EEG.

⁵ Imagen extraída de <https://es.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html>

3.3.4.1. Dataset propio

No se realiza el entrenamiento para este dataset.

3.3.4.2. Thinking Out Loud

En esta variante, se utilizan tanto espectrogramas para 1 canal como para el montaje de 16 de ellos. En ambos casos se introducirán las imágenes en la red neuronal sin títulos ni leyendas. El rango de frecuencias (0.1Hz-88Hz) de la transformada de Morlet será completo (esto es el eje vertical) frente a las muestras del intervalo completo en el tiempo (eje horizontal) en el caso de una única imagen (mostrado en la Figura 19). Sin embargo, para el montaje se decide recortar el eje vertical, prescindiendo así de numerosas frecuencias. La banda de frecuencia utilizada será la de 4-64Hz.

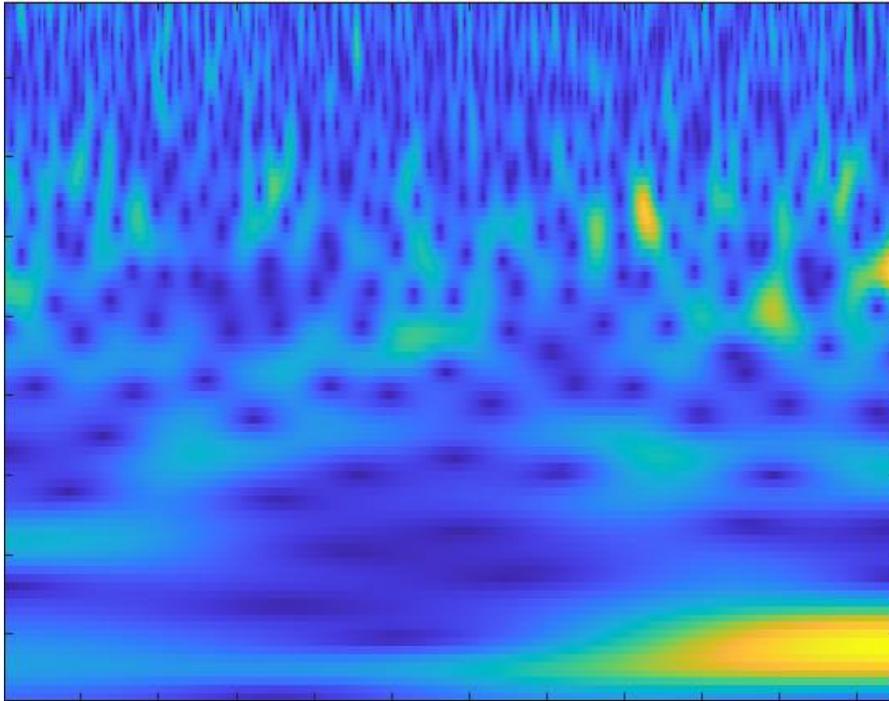


Figura 19. Espectrograma del canal A15⁶ para un trial de la palabra ‘abajo’ (Thinking Out Loud)

Como se puede observar en la Figura 20, el espectrograma de cada canal se convierte en rectangular para que el conjunto de todos ellos sea cuadrado (condición necesaria para el correcto funcionamiento de estas últimas redes neuronales). Por último, las imágenes serán redimensionadas al tamaño de entrada correspondiente a GoogleNet o AlexNet.

⁶ A15: Nomenclatura para electrodo del sistema 128 Biosemi

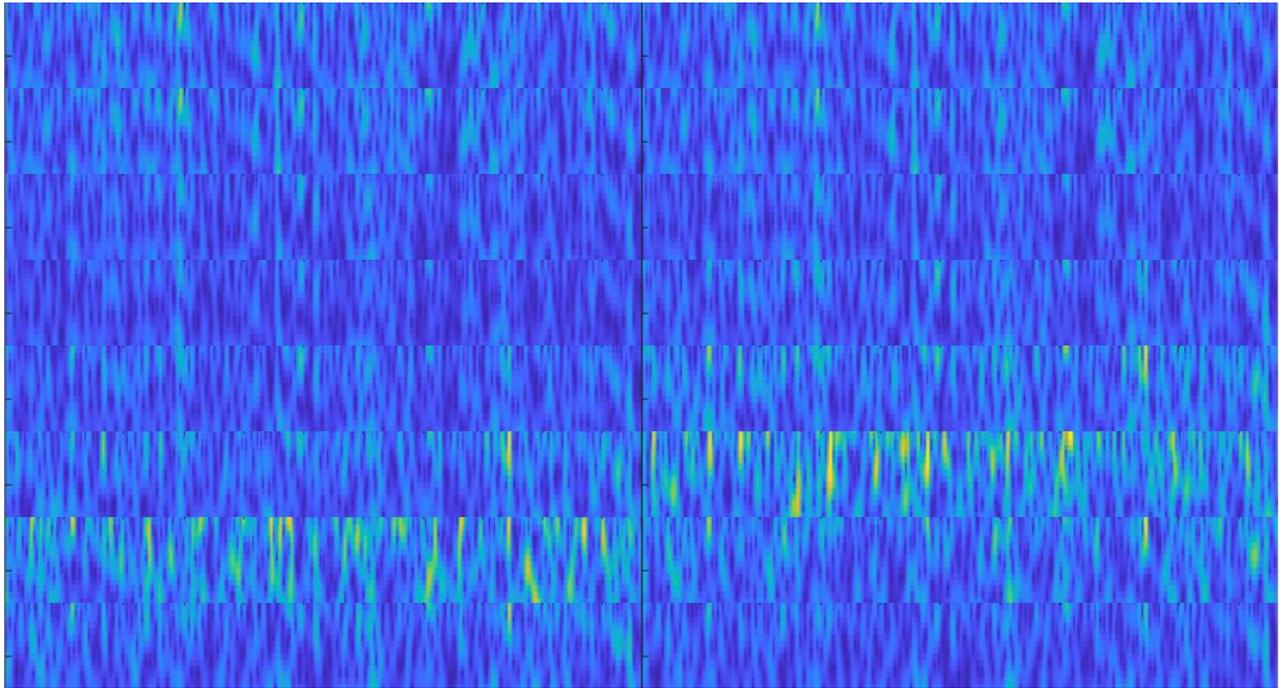


Figura 20. Montaje de 16 espectrogramas para un trial de la palabra 'abajo' (Thinking Out Loud)

3.3.4.3. MindBigData

En esta variante, seguirá los mismos pasos que para el dataset anterior, pero teniendo en cuenta que solamente disponemos de 4 canales.

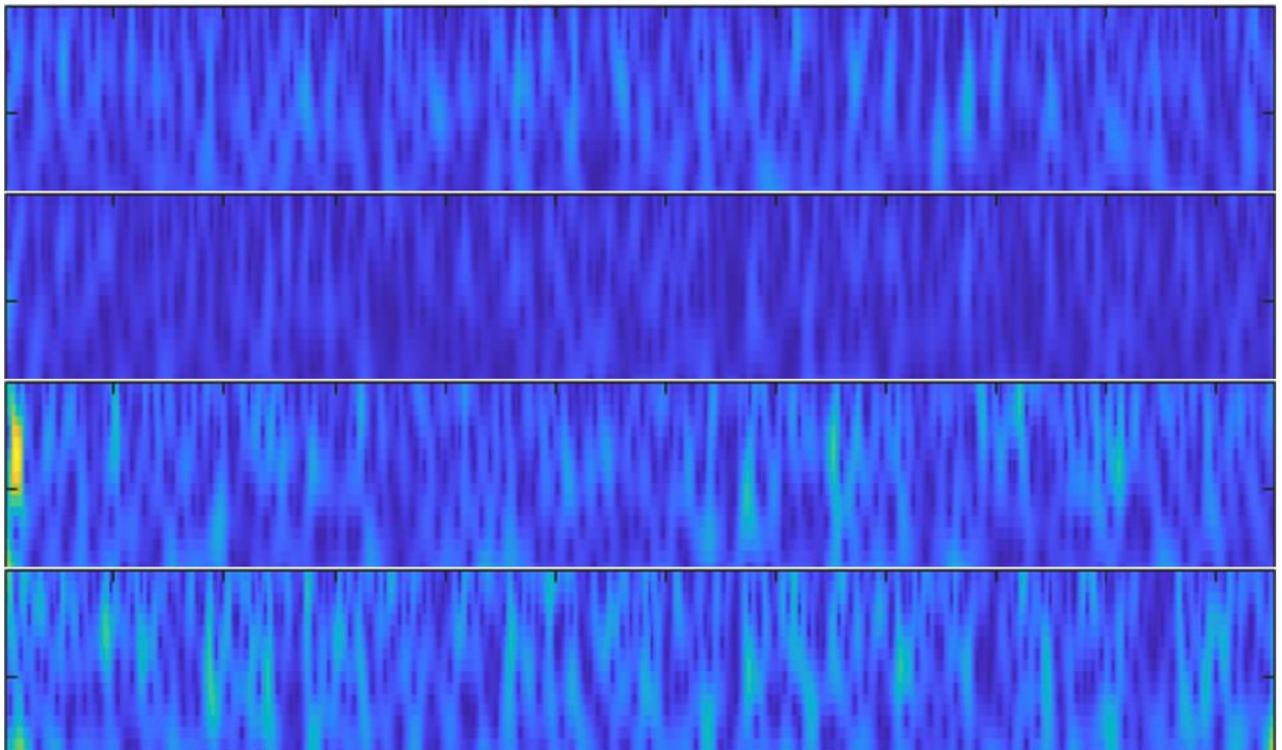


Figura 21. Montaje de los 4 espectrogramas para un trial del dígito '0' (MindBigData)

4. RESULTADOS Y DISCUSIÓN

La siguiente sección presenta los resultados obtenidos en el marco de este TFG. Se muestran los datos recopilados durante la investigación, los cuales han sido analizados para proporcionar información relevante acerca de cómo abordar los objetivos planteados.

4.1. Tabla de resultados obtenidos

A través de esta tabla de resultados, se busca ofrecer una visión clara y concisa de la información obtenida con la investigación realizada en este estudio, mostrando las combinaciones entre arquitecturas y enfoques explicadas en 3.3. Implementación de redes neuronales.

	Aprendizaje Simple		LSTM		TF CNN	GoogleNet		AlexNet	
	EEG	Bandas frecuencia	EEG	Bandas frecuencia	EEG 1 canal	Espectrogramas		Espectrogramas	
						1 canal	Varios canales	1 canal	Varios canales
Señales Propias	29.0%	35.5%	~25%	27.3%	36.4%	-	-	-	-
Thinking Out Loud [3]	~25%	~25%	~25%	~25%	~25%	~25%	29.1%	~25%	25.3%
MindBigData⁷	~50%	~50%	~50%	85.8%⁸	~50%	~50%	~50%	~50%	~50%

Tabla 5. Resultados del aprendizaje mediante diferentes arquitecturas

⁷ Clasificación binaria. Las bandas de frecuencias son del electrodo TP9.

^{8,7} Se denota un comportamiento extraño de la red neuronal, revisar la interpretación de los datos obtenidos.

4.2. Interpretación de los datos obtenidos

Los niveles de precisión alcanzados, por lo general, apenas superan la probabilidad del azar. Siendo así, se pretende realizar un análisis de limitaciones y posibles fuentes de error.

Lo primero que destaca es un mayor ‘Accuracy’ para las señales propias tanto con la red ‘Aprendizaje Simple’ como para ‘TF CNN’ con respecto a todas las clasificaciones realizadas con el dataset ‘Thinking Out Loud’. Esto podría deberse a que las posiciones de los electrodos colocados en el experimento propio son de especial interés, mientras que las respectivas redes neuronales no son capaces de extraer características relevantes del segundo dataset, aun teniendo esta mejor resolución espacial (incluyendo la zona estudiada en el primer conjunto de datos). También influiría el número de trials perteneciente al conjunto de validación, pues se tiene una cifra relativamente baja de grabaciones en la sala anecoica. Siendo así, es más probable que el número de aciertos por número de trials en el conjunto de validación se dispare y quede por encima de la probabilidad de acertar aleatoriamente.

Por otro lado, observamos un comportamiento parecido entre la clasificación de las señales propias en ‘Aprendizaje Simple’ cuando los datos a introducir son las bandas de frecuencias más la señal EEG y ‘TF CNN’, que calcula el nivel de energía de las distintas frecuencias dentro de la red. Ambas arquitecturas elevan el nivel de precisión alrededor de un 10% en comparación con los demás enfoques estudiados. Se puede afirmar que las bandas de frecuencia en su conjunto (θ , α , β , γ , energía total) tienen una relevante importancia en el reconocimiento de palabras mediante EEG.

También se ha observado que el dataset de MindBigData presenta dificultades para alcanzar altos niveles de precisión en las tareas de clasificación realizadas. Específicamente, se ha identificado que cuando se logra obtener un alto porcentaje de Accuracy en el conjunto de validación, se presenta un comportamiento inusual en la red neuronal, ya que el porcentaje de Accuracy en el conjunto de entrenamiento es inferior al de validación. Este hallazgo plantea interrogantes sobre el desempeño y la generalización del modelo en el dataset de MindBigData, lo que sugiere posibles desafíos adicionales en la clasificación precisa de los datos. Estos resultados resaltan la importancia de profundizar en la comprensión de las características y particularidades del dataset utilizado, así como la necesidad de investigar estrategias adicionales para mejorar la calidad de las predicciones y abordar el comportamiento inesperado observado en los resultados.

Por último, al comparar los métodos aquí propuestos con los descritos en la literatura científica, se observa que los niveles de precisión obtenidos se encuentran por debajo de los reportados anteriormente. Es particularmente relevante mencionar que, a pesar de haber realizado una clasificación binaria con MindBigData en este estudio y un extraño comportamiento en el aprendizaje, los resultados aún presentan niveles de precisión inferiores a los informados en la literatura para tareas de reconocimiento de patrones más complejas, es decir, la distinción de 10 dígitos.

	Método propuesto	Literatura
Señales propias	TF CNN: 36.4%	Pawar & Dhage [36]: 49.8%
Thinking Out Loud [3]	GoogleNet: 29.1%	F. Gasparini et al [30]: 36.1%
MindBigData	LSTM: 85.8%	Mahapatra NC & Bhuyan P [28]: 96.2%

Tabla 6. Comparativa de resultados: métodos propuestos y literatura

Teniendo en cuenta todo lo anterior, es importante reconocer una limitación en el reconocimiento de palabras en este estudio. Se cree que este sea inherente a la técnica de EEG, en relación con la resolución espacial de las señales adquiridas, lo cual implica que la precisión para localizar con exactitud las fuentes neuronales activas y comprender en detalle los patrones de actividad cerebral es comprometida. Puede ser atribuida a varios factores.

- La forma en que las señales se registran en la superficie del cuero cabelludo mediante electrodos limita la capacidad de obtener una imagen precisa de la actividad cerebral en términos de ubicación espacial, ya que, en un electrodo, se produce la integración de las actividades de cientos de millones de neuronas,

cada una de las cuales posee su propio propósito y significado en el complejo funcionamiento del cerebro. Además, la conductividad de los tejidos y el cráneo introduce distorsiones y atenuaciones en las señales, lo cual puede afectar la localización precisa de las fuentes neuronales subyacentes.

- Es de vital importancia considerar la complejidad de las actividades que ocurren simultáneamente en el cerebro. Este órgano está constantemente involucrado en una multitud de procesos cognitivos, emocionales y perceptuales, lo que genera un entorno altamente dinámico y complejo y una ‘interferencia cognitiva’ que se refleja en las señales registradas y dificulta la interpretación precisa de la actividad cerebral que se busca analizar. Además, esta interferencia cognitiva también puede estar presente a nivel de la actividad neuronal en sí misma. Las diferentes regiones cerebrales están interconectadas y en constante comunicación, lo que significa que la actividad en una región puede influir en la actividad en otras regiones. Esta interacción entre regiones puede generar patrones complejos y difíciles de desentrañar, especialmente cuando se busca identificar la contribución de una región específica a un fenómeno cognitivo determinado.

5. LÍNEAS FUTURAS

Just because something doesn't do what you planned it to do doesn't mean it's useless.

Autor desconocido⁹

A sí todo, los datos mostrados en el apartado ‘4. Resultados y Discusión’ sugieren que se necesita un mayor refinamiento y mejora en los enfoques utilizados para el reconocimiento de palabras en los diferentes datasets. Por lo tanto, se proponen varias líneas de investigación para investigaciones futuras.

5.1. Mejora en el tratamiento del EEG

Estos resultados indican la importancia de explorar y ajustar los modelos, las técnicas de preprocesamiento, la extracción de características y las arquitecturas utilizadas para obtener un rendimiento significativamente mejor que el azar en esta tarea. Es posible que se requieran enfoques más sofisticados, como la combinación de diferentes técnicas o la utilización de modelos más complejos, para mejorar la precisión en la clasificación de palabras en estos contextos específicos de señales EEG. Ejemplo de ello son [28] y [7], donde se diseñan redes neuronales para este contexto específicamente. En el primero, se crean diferentes células LSTM, GRU o RNN para formar ‘capas bidireccionales recurrentes’, que a su vez constituirán una red LSTM bidireccional con alto nivel de precisión según sus redactores, mientras que en el segundo se crean diferentes CNN para discriminar entre habla imaginada y otros pensamientos, captar el momento de atención del sujeto o separar regiones del cerebro para distinguir la clase gramatical de la palabra imaginada. Estos autores destacan en sus artículos que alcanzaron las precisiones de 96.18% en la clasificación de 10 dígitos y de 76.6% en la clasificación de 3 pares de palabras, respectivamente.

5.2. Métodos alternativos al EEG

En vista de la limitación por falta de resolución espacial, resulta pertinente explorar y considerar el uso de técnicas diferentes a la EEG con el fin de obtener una mejor resolución espacial en futuras investigaciones. Existen diversos enfoques que podrían complementar o reemplazar el uso exclusivo del EEG, como la resonancia magnética funcional (fMRI), cuya exploración ya ha sido iniciada [42], la magnetoencefalografía o la estimulación magnética transcraneal, entre otras.

Estas técnicas ofrecen una mayor precisión espacial al mapear la actividad cerebral, lo que facilita una identificación más precisa de las regiones cerebrales involucradas en los procesos cognitivos y la comprensión de la organización espacial de la actividad cerebral. Además, la combinación de múltiples técnicas de neuroimagen puede proporcionar una visión más completa y holística de la actividad cerebral, complementando así los resultados obtenidos a través de la EEG.

No obstante, es importante tener en cuenta que cada técnica tiene sus propias ventajas y limitaciones, y la elección de la técnica más adecuada dependerá de los objetivos de investigación, las restricciones de recursos y

⁹No se ha encontrado una fuente confiable que afirme el origen de esta cita, aunque en numerosas ocasiones ha sido atribuida a Thomas Alva Edison, 1847-1931.

las características específicas del estudio en cuestión. Se recomienda realizar un análisis exhaustivo y una cuidadosa evaluación de las diferentes opciones disponibles antes de seleccionar el método o la combinación de métodos más apropiados para cada investigación.

En conclusión, se sugiere explorar y considerar la utilización de métodos alternativos a la EEG con el objetivo de superar la limitación en la resolución espacial que presenta esta técnica. La incorporación de técnicas complementarias de neuroimagen permitirá obtener una visión más precisa y detallada de la actividad cerebral, enriqueciendo así la comprensión de los fenómenos cognitivos y neurofisiológicos abordados en el estudio.

REFERENCIAS

- [1] B. D. P. P. a. M. A. Lopez-Bernal D, «A State-of-the-Art Review of EEG-Based Imagined Speech Decoding,» *Front. Hum. Neurosci.* , nº 16:867281, 2022.
- [2] R. F. a. D. C. Ciaran Cooney, «Neurolinguistics Research Advancing Development of a Direct-Speech Brain-Computer Interface,» *iScience*, 2018.
- [3] N. P. V. R. H. e. a. Nieto, «Thinking out loud, an open-access EEG-based BCI dataset for inner speech recognition,» *Sci Data*, 2022.
- [4] «Electroencephalography (Wikipedia),» [En línea]. Available: <https://en.wikipedia.org/wiki/Electroencephalography>. [Último acceso: 30 5 2023].
- [5] L. A. y. F. Cincotti, «Apuntes Neuroengineering,» *Sapienza Università di Roma*, 2022.
- [6] J. Sorkin, «Getting into Brain Waves: History and Resources,» *NeuWrite West*, 2016 .
- [7] S. Datta, «Brain signal recognition using deep learning,» *Brunel University London*, 2022.
- [8] «Afasia. National Institute on Deafness and Other Communication Disorders,» [En línea]. Available: <https://www.nidcd.nih.gov/es/espanol/afasia>. [Último acceso: 1 6 2023].
- [9] S. B. ., E. W. S. Y. J. a. X. Z. Reza Abiri, «A comprehensive review of EEG-based brain-computer interface paradigms,» *J. Neural Eng.*, 2019.
- [10] R. T. B. R. J. C. R. D. R. J. F. N. & N. M. R. Pivik, «Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts,» *Psychophysiology*, 1993.
- [11] R. & P. P. Oostenveld, «The five percent electrode system for high-resolution EEG and ERP measurements.,» *Clinical Neurophysiology*, 2001.
- [12] «128 channel layout. Biosemi,» [En línea]. Available: <https://www.biosemi.com>.
- [13] P. L. & S. R. Nunez, «Electric fields of the brain: the neurophysics of EEG,» *Oxford University Press*, 2006 .
- [14] S. J. Luck, « An introduction to the event-related potential technique,» *MIT Press*, 2014.
- [15] L. L. F. L. Q. C. Y. C. Y. W. X. & W. W. Dong, «A review of methods for removing power line interference from EEGs.,» *IEEE Access*, 2019.
- [16] T. W. B. S. B. P. D. E. H. S. A. J. J. R. .. & T. M. J. Picton, «Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria.,» *Psychophysiology*, 2000.
- [17] G. B. P. v. d. B. R. V. M. & L. d. S. F. H. Baldewijns, «Data reduction in EEG research: An overview on

- the essentials,» *Journal of Neuroscience Methods*, 2019.
- [18] C. L. L. & W. F. Rusu, «Decimation and interpolation techniques for reducing size of bio-potential signal datasets,» *Biomedical Signal Processing and Control*, 2016.
- [19] X. Z. Y. J. X. & G. X. Liu, «The influence of sampling rate on the characterization of EEG oscillations,» *NeuroImage*, 2015.
- [20] G. C. M. G. & D. E. Gratton, «A new method for off-line removal of ocular artifact,» *Electroencephalography and Clinical Neurophysiology*, 1983.
- [21] B. W. S. A. J. M. J. S. B. D. R. K. A. M. G. L. L. & D. R. J. McMenamin, «Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG,» *Neuroimage*, 2010.
- [22] A. Ú. D. P. C. P.-V. E. I. a. J. M. A. Belén López, «APLICACIÓN DE ICA PARA LA ELIMINACIÓN DE ARTEFACTOS OCULARES Y MANDIBULARES EN SEÑALES EEG DURANTE LA REALIZACIÓN DE TAREAS MENTALES,» *Universidad Miguel Hernández*, 2014.
- [23] «Wavelet transforms. Mathworks,» [En línea]. Available: <https://es.mathworks.com/discovery/wavelet-transforms.html>. [Último acceso: 2023].
- [24] S. & B. M. Morales, « Time-frequency analysis methods and their application in developmental EEG data.,» *Developmental Cognitive Neuroscience*, 2022.
- [25] H. D. C. M. S. M. E. A. Hipp JF, « Large-scale cortical correlation structure of spontaneous oscillatory activity,» *Nat Neurosci*, 2012.
- [26] «Daubechies wavelet. Wikipedia,» [En línea]. Available: https://en.wikipedia.org/wiki/Daubechies_wavelet. [Último acceso: 1 6 2023].
- [27] «Introduction to the wavelet families. Mathworks,» [En línea]. Available: <https://es.mathworks.com/help/wavelet/gs/introduction-to-the-wavelet-families.html>. [Último acceso: 1 6 2023].
- [28] B. P. Mahapatra NC, « EEG-based classification of imagined digits using a recurrent neural network,» *J Neural Eng.*, 2023.
- [29] B. & v. D. S. & A. M. Berg, «Inner Speech Classification using EEG Signals: A Deep Learning Approach,» 2021.
- [30] E. C. & A. S. Francesca Gasparini, «Inner speech recognition through electroencephalographic signals,» *University of Milano-Bicocca & NeuroMI*, 2022.
- [31] F. D. R. M. L. J. E. A. a. B. C. D. Bird J J, «A deep evolutionary approach to bioinspired classifier optimisation for brain-machine interaction,» *Complexity*, 2019.
- [32] A. P. N. S. S. G. V. G. M. S. a. S. R. R. Jolly B L K, «Universal EEG encoder for learning diverse intelligent tasks,» *IEEE 5th Int. Conf. on Multimedia Big Data (BigMM) (Singapore: IEEE)*, 2019.
- [33] K. M. W. A. a. G. I. G. A. Pratama I W P, «Frequency band and PCA feature comparison for EEG signal classification,» *Lontar Komputer : Jurnal Ilmiah Teknologi Informatika*, 2021.

- [34] A. A. T. G. C. A. R. a. P. L. V. García, «Toward a silent speech interface based on unspoken speech,» *Biosignals*, 2012.
- [35] L. L. P. C. C. A. W. B. J. a. T. A. Sarmiento, «Brain computer interface (BCI) with EEG signals for automatic vowel recognition based on articulation model,» *5th ISSNIP-IEEE Biosignals and Biorobotics Conference, Biosignals and Robotics for Better and Safer Living (BRC)*, (Salvador: IEEE), 2014.
- [36] D. a. D. S. Pawar, « Multiclass covert speech classification using extreme learning machine,» *Biomed. Eng. Lett*, 2020.
- [37] C. K. A. F. R. a. C. D. Cooney, « Evaluation of hyperparameter optimization in machine and deep learning methods for decoding imagined speech EEG,» *Sensors*, 2020.
- [38] M.-O. M. Y. a. M. N. Tamm, «Classification of vowels from imagined speech with convolutional neural networks.,» *Computers*, 2020.
- [39] «Boundary effects and the cone of influence. Mathworks,» [En línea]. Available: <https://es.mathworks.com/help/wavelet/ug/boundary-effects-and-the-cone-of-influence.html>. [Último acceso: 5 6 2023].
- [40] «Web Muse,» [En línea]. Available: <https://choosemuse.com/>.
- [41] D. V. & F. Cuesta, «MindBigData 2022: A Large Dataset of Brain Signals,» 2022.
- [42] B. C. & M. M, «Brain Computer Interface -EEG Based Imagined Word Prediction Using Convolutional Neural Network Visual Stimuli for Speech Disability,» *ResearchSquare*, 2021.
- [43] «Transfer Learning. Mathworks,» [En línea]. Available: <https://es.mathworks.com/discovery/transfer-learning.html>.
- [44] «Deep Learning Speech Recognition. Mathworks,» [En línea]. Available: <https://es.mathworks.com/help/deeplearning/ug/deep-learning-speech-recognition.html>.
- [45] V. G. R. S. K. D. N. A. S. R. S. W. H. W. M. L. J. E.] Foteini Simistira Liwicki, «Bimodal pilot study on inner speech decoding reveals the potential of combining EEG and fMRI,» *bioRxiv*, 2022.
- [46] Autor, «Este es el ejemplo de una cita,» *Tesis Doctoral*, vol. 2, nº 13, 2012.
- [47] O. Autor, «Otra cita distinta,» *revista*, p. 12, 2001.
- [48] «Coarticulación. Wikipedia,» [En línea]. Available: <https://es.wiktionary.org/wiki/coarticulaci%C3%B3n>. [Último acceso: 1 6 2023].

