Final Degree Project

Degree in Electronic Engineering, Robotics and Mechatronics

BCI interface for word recognition in EEG signals

Author: Pablo Martin Puerto

Tutor: Jose Manuel Quero Reboul

Equation Chapter 1 Section 1

Department of Electronic

Engineering Higher Technical School of

Engineering University of Seville

Sevilla, 2023

Final Degree Project

Degree in Electronic Engineering, Robotics and Mechatronics

BCI interface for word recognition in EEG signals

Author:

Pablo Martin Port

Tutor:

Jose Manuel Quero Reboul

University Professor

Department of Electronic

Engineering Higher Technical School of Engineering

University of Seville

Sevilla, 2023

Final Degree Project: BCI interface for word recognition in EEG signals

Author: Pablo Martin Puerto

Tutor: Jose Manuel Quero Reboul

The court appointed to judge the above-mentioned Project, composed of the following members:

President:

Vocal/s:

Secretary:

They agree to give it the following rating:

The Court Clerk

Date:

To my family: blood and chosen

To Seville and the rocker Silvio

Acknowledgements

These acknowledgements are a sign of my deep and sincere appreciation towards those who have contributed directly and indirectly to my personal and academic development. Without them, I would not have reached the end of this degree.

To all the teachers I had in high school, college and university, your dedication and commitment to public education have been fundamental to my development and growth as a student and citizen.

To my entire family for their unconditional support throughout this academic journey.

To my parents and my brother, for giving me life and for being my greatest inspiration. They have been by my side, giving me all their love and support, and for pushing me to grow and pursue my goals.

To my grandparents for having transmitted to me directly or through the words of other relatives the values that form the basis of my character and my way of seeing the world.

To my aunt Maribel and my partner Abraham, who perhaps are not aware of how much I appreciate them. They have been a source of inspiration and motivation, as well as teaching me the importance of valuing every moment and facing challenges with courage.

To my friends, to Lucía and to football, they are the ones who listen to me, support me and advise me daily.

Thank you all for making me happy.

Pablo Martin Port

Sevilla, 2023

Summary

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

Currently, the electroencephalogram (EEG) stands out as the most widely used method for noninvasive recording of brain activity. This technique offers high temporal resolution, remarkable ease of use and a 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 the subject's posture and concentration on the task to be performed or interference caused by eye and muscle movement. All of 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.

The realization of this work is motivated by the significant potential of EEG signals in the development of BCI, which can be implemented to open a new communication channel for people with motor disabilities in speech 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, in addition to contributing with the contribution of this study to the one carried out by researchers from all over the world in the attempt to decode speech from brain waves.

In this context, a comprehensive study of the state of the art in EEG word recognition will be conducted. 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. In addition, a couple of datasets available in the literature will be analyzed and compared with the BCI system's own implementation.
At the end, solid conclusions will be obtained based on the results and analysis carried out.

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.

Index

INDEX OF TABLES

INDEX OF FIGURES

Notation

BCI	Brain-Computer Interface
CNN	Convolutional Neural Network
TF	Time-Frequency
ERP	Event Related Potential
EOG	Electrooculograma
EMG	Electromyography
ICA	Independent Component Analysis
LSTM	Long Short Term Memory
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
D4	Daubechies-4
%	Percentage
V	Volt
TFG	Final Degree Project
SNR	Signal-to-Noise Ratio

1 INTRODUCTION

L Technological evolution has led to extensive connectivity between electronic devices and the human body. An example of this is the BCI, which is a system that uses the activity brain and EEG signals to create a communication channel between external electronic devices and the human brain [1].

1.1. Brain-Computer Interface (BCI)

BCI systems have been used in various areas. Examples of them are neuromarketing, security, entertainment or control of smart environments, among others. One of the most explored BCI applications is in the medical area, where it has been widely used for the treatment and diagnosis of neurological disorders such as epilepsy, depression, dementia or Alzheimer's [1]. It has also been applied for the recognition and classification of emotions, as well as to identify the different stages of sleep. Furthermore, one of the most challenging and interesting applications of BCI is the recognition of imagined speech, where the aim is to convert brain signals into text, sound or control commands [1]. This is the case of the Direct-Speech BCI (DS-BCI), which acquires brain signals corresponding to imagined speech, processes and decodes them to produce a linguistic output in the form of phonemes, words or phrases [2]. But the acquisition of such signals is not an easy task, since speech production is one of the most complex brain processes performed by human beings, as it requires the interaction of several cortical, basal and subcortical brain regions [3]. On the other hand, it opens the way to several paradigms presented in [3]:

1.1.1 Inner speech

In this paradigm, participants imagine their own voice as if they were giving a direct command to the computer. They do not move their mouth or tongue, and remain as still as possible, without moving their mouth or tongue. The main goal is to detect the brain's electrical activity related to the thought of a particular word.

1.1.2 Pronounced speech

The spoken speech paradigm is proposed with the aim of finding the motor regions involved in pronunciation that are activated during inner speech. In this paradigm, participants pronounce aloud the word corresponding to each visual stimulus, as if they were giving a direct command to the computer.

1.1.3 Visualized condition

Since the selected words have a significant visual and spatial component, and in order to find any visualization-related activity during inner speech, the visualization paradigm is proposed. It is mentioned that the main neural centers related to spatial thinking are located in the occipital and parietal regions. Furthermore, spatial attention has been shown to have a significant impact on the amplitude of visual evoked potentials. In visualization trials, participants are instructed to focus on mentally moving the circle appearing in the center of the screen in the direction indicated by the visual stimulus.

The name of each paradigm will be kept in English from now on.

The structure of a DS-BCI consists of several stages:

1. Signal Acquisition and Conditioning: Recording and preprocessing the user's brain activity.

2. Signal Processing & Features Extraction: the digitalized signal is received, filtered to eliminate noise and different techniques are applied to obtain features.

3. Features classification: using neural networks or machine learning. learning.

4. Application or control interface: functional block that receives the data. control commands and performs the necessary actions.

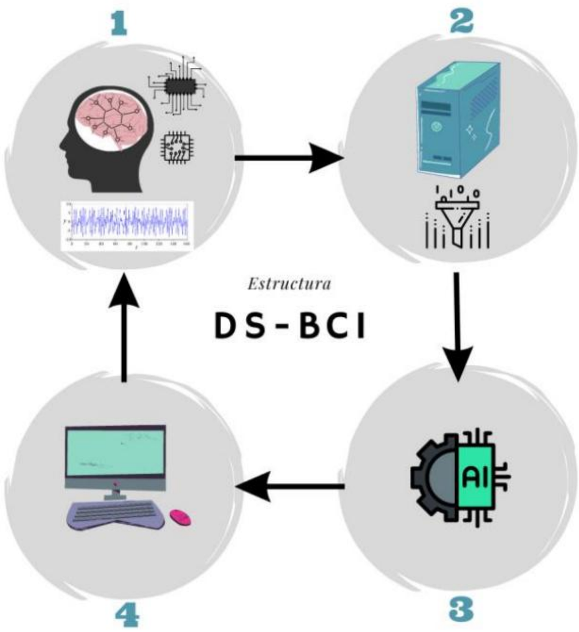

Figure 1. Structure of a DS-BCI

1.2. EEG signals

1.2.1. History of the electroencephalogram

Electroencephalography is the study of the patterns of electrical activity in the human brain, and is a key tool in understanding brain function and diagnosing neurological disorders. Its history dates back to the late 20th century. It began when, in 1875, Richard Caton, a British physiologist, performed the first experiments recording the electrical activity of the brains of animals. His studies showed that the brain emits weak electrical signals and that these signals vary depending on the state of wakefulness and sleep [4].

About 50 years later, Hans Berger, a German psychiatrist, made a significant breakthrough by recording the first EEG signal in a human being [4]. Berger developed an electroencephalograph, a device capable of measuring the electrical activity of the brain through electrodes placed on the scalp [4]. His studies

They demonstrated the existence of different types of brain waves, such as alpha and beta waves, and laid the foundations for modern electroencephalography [5].

Over the following decades, significant advances were made in EEG recording technology and in understanding brain activity patterns. More sophisticated signal amplification and recording systems were developed, allowing for greater accuracy and resolution in capturing brain activity [6].

In the 1960s, digitization of EEG signals was introduced, allowing for more advanced processing and analysis of the data obtained. This led to the development of frequency and time domain analysis techniques for EEG signals, allowing for a better understanding of the different characteristics and patterns present in brain signals [6].

In recent decades, EEG technology has seen significant advances, such as the miniaturization of recording devices and the development of more advanced signal processing techniques, such as machine learning and neural networks. This has enabled greater applicability of EEG signals in diverse areas, such as medicine, cognitive neuroscience, sleep research, and Brain-Computer Interfaces (BCI) [6], which is relevant to our work.

1.2.2. Neural processes and EEG recording

The electroencephalogram records the electrical activity of the brain, which originates from the synchronization of millions or even billions of neurons in the brain [5].

When neurons communicate with each other, they use neurotransmitters, such as sodium (Na+) and potassium (K+) ions, to transmit electrical and chemical signals. These neurotransmitters play a key role in generating action potentials, which are electrical impulses that propagate along neurons [5].

In a neuron, the resting potential is kept negative due to the distribution of ions across the cell membrane. Sodium (Na+) is predominantly found outside the cell, while potassium (K+) is predominantly found inside the cell. This distribution creates a difference in electrical charge and establishes the resting potential [5].

When a neuron receives sufficient excitatory stimulation, such as the release of excitatory neurotransmitters, a change in the permeability of the cell membrane occurs. This allows sodium ions (Na+) to enter the cell and potassium ions (K+) to leave the cell, generating an action potential.
The action potential propagates along the axon of the neuron and can activate other neurons to which it is connected through synapses [5].

In the context of EEG signals, currents generated by neurons propagate through water-containing brain tissues. This is known as volume conduction and allows the effects of the neurons' electrical activity to propagate over relatively long distances, even several centimeters.
These currents generate electric fields that add linearly and propagate instantaneously [5].

At the surface of the scalp, where electrodes are placed to record EEG, the amplitude of the signals is very weak, usually measured in millivolts (uV), while at the membrane the potential is several orders of magnitude higher (mV). This is because the signals at the surface are an indirect representation of the electrical activity of neurons in the brain. Despite their weakness, EEG signals provide valuable information about brain activity and can be analyzed to detect patterns and changes in the brain's electrical activity [5].

1.2.3. Relevant characteristics of EEG signals for word recognition

1.2.3.1. Brodmann, Broca and Wernicke areas

Most models and theories of language agree that speech involves auditory, semantic and syntactic processing, as well as motor planning and articulation processes [3].

Brodmann areas are a classification of regions in the cerebral cortex based on differences in cellular structure and function of each of them. They were defined by the German neurologist Korbinian Brodmann at the beginning of the 20th century, who carried out exhaustive studies of the human brain and divided the cerebral cortex into different areas numbered from 1 to 47 [5] as can be seen in Figure 2. Two regions that play a fundamental role in the production and comprehension of language are: Broca's area (Brodmann areas 44 and 45) and Wernicke's area (Brodmann areas 22, 39, 40).

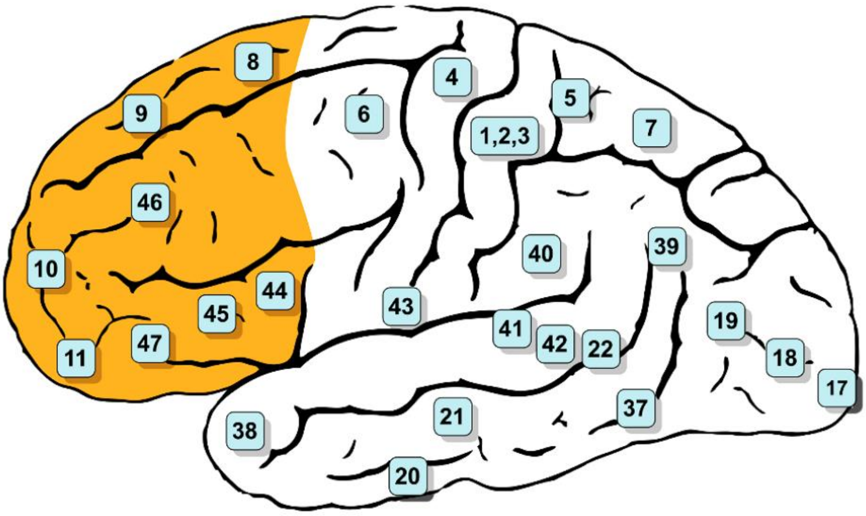

Figure 2. Brodmann areas, image taken from Wikipedia

Broca's area (Brodmann Areas 44 and 45), located in the frontal lobe of the dominant hemisphere (usually the left hemisphere in most people), is responsible for the production of spoken language and the planning of the articulatory movements necessary to articulate words [7].
Lesions in Broca's area can result in a condition known as Broca's aphasia, in which individuals have difficulty expressing themselves verbally, but retain understanding of language [8].

Wernicke's area (Brodmann areas 22, 39, 40), is located in the posterior temporal lobe, adjacent to the primary auditory area [7]. This area is involved in the comprehension and processing of spoken and written language. Lesions in Wernicke's area can result in Wernicke's aphasia, in which individuals have difficulty understanding language and produce incoherent, fluent speech [8].

The relationship between Broca's area and Wernicke's area is crucial for language production and comprehension. These areas are thought to be connected by a neural pathway known as the arcuate fasciculus, which allows communication between them [7].

In short, Broca's and Wernicke's areas are specific regions of the brain involved in language production and comprehension. These areas work together to allow us to communicate verbally and understand the meaning of words.

1.2.3.2. Event Related Potentials (ERP)

When a stimulus is presented, a change in the EEG signal known as an event-related potential (ERP) occurs. The stimulus that gives rise to an ERP represents the activity generated in neurons, i.e. the change in potential caused by the stimulus. The ERP represents the electrical activity of a considerable population of neurons [7].

An ERP is obtained by averaging the EEG signals of a specific type of event [7]. Performing an average in the study of behavior allows to observe and monitor unattended stimuli. In addition,
It also offers another advantage in that it reduces the residual noise in the averaged signal compared to a single trial, although this has its limitations. ERPs have been used to control communication-based BCI systems, and some of the ERPs used are SSVEP and P300.

- **SSVEP**

Steady-state visual evoked potential (SSVEP) is a visual component used in brain-computer interfaces (BCIs). It is based on constant-frequency flashing stimuli and there is a significant correlation between the flashing frequency of the stimulus and the frequency observed in the EEG. This allows the user's desired goal to be determined by matching the EEG activity pattern with the command associated with a particular frequency.

SSVEP offers advantages such as ease of use for most subjects and the possibility of controlling devices with high precision [2]. However, it is important to note that the use of flashing stimuli can lead to user fatigue, especially when low flashing frequencies are used [2]. Furthermore, the SSVEP paradigm is not suitable for visually impaired individuals as it requires visual focus shifts [2].

- **P300**

The P300 is another of the most studied ERPs. The P300 component is generated in response to repetitively presented events and the occasional appearance of a different stimulus, using what is known as an "oddball paradigm." The P300 is a positive spike in the ERP that ranges in size from 5 to 10 microvolts and has a latency between 220 and 500 ms post-event onset. This ERP is defined as an averaged increase in the amplitude of the time series of brain signals, being most significant at midline locations (Pz, Cz, and Fz in the international 10/20 system).

The P300 obtained in any experiment is likely to be the sum of the P300 and other components that are generated by other stimuli or interferences presented before and after each presentation of the stimulus of interest [2].

- Another paradigm related to inner speech is the so-called “auditory comprehension” [3]. In this paradigm, instead of actively producing speech imagination, the individual passively listens to another person's speech. It has already been explored using EEG, ECoG and fMRI. Although this paradigm is not especially useful for real BCI applications, it has contributed to the understanding of the neural processes associated with speech-related paradigms.

ERP analysis in word recognition allows studying processing speed, activation of specific brain areas, and semantic understanding of words. In addition, signal processing techniques and machine learning algorithms can be used to extract relevant features from ERPs and develop word recognition models.

However, it is important to note that the analysis of ERPs in word recognition presents challenges, such as the inherent noise in EEG signals and the interindividual variability in brain responses. Furthermore, the interpretation of ERP components may depend on the context and task used.

1.2.3.3. Frequency bands

The four main brain frequencies are theta (4-8Hz), alpha (8-13 Hz), beta (13-30Hz) and gamma (30-200Hz). These play different roles in speech processing and have been shown to behave differently in different activities, challenging the idea that they have generalized behavior in specific activities. Understanding the brain and its behavior is in its early stages, and phenomena such as brain plasticity make it even more difficult to understand the complexities of the human brain and the brain waves involved [7].

Each frequency has a specific function at different stages of language and speech processing, which helps us understand its importance in this research. Below, we present a review bibliographical review of the frequencies involved in speech processing [7]:

- Theta (ÿ)

  Theta waves react to phonetic features of speech and play a role in phoneme reconstruction and coarticulation signal processing.1 which helps in the construction of words within a speech.

- Alpha(s)

  Event-related desynchronization in the ÿ frequency is considered a sign of feedback to the primary motor cortex during speech production. Phenomena of ÿ synchronization and desynchronization are observed during speech perception, and an increase in ÿ activity is related to concentration tasks and memory workload, which aids in auditory processing and speech production.

- Beta (ÿ)

  Beta waves in speech are related to muscle movement and are speculated to play a role in generating feedback control through the primary motor cortex and internal modeling. These waves are used in speech discrimination and then transferred to the auditory region for processing and speech production.

- Gamma (ÿ)

  During speech production, changes in gamma frequency are observed in several brain areas, including the temporal lobe, supramarginal sulcus, Broca's area, Wernicke's area, premotor cortex, and primary motor cortex. The ÿ frequency has a rapid effect in response to the stimulus presentation interval and the interstimulus interval.

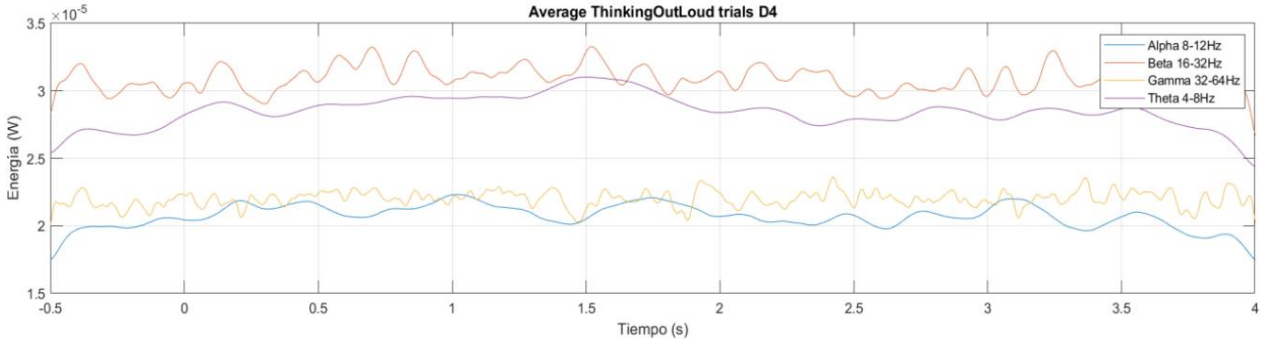

Figure 3. Example of energy in the different frequency bands

In short, each brain frequency plays a specific role in speech processing. Studying them helps to better understand how spoken language is processed and produced in the human brain, so all bands will be used.

1 Coarticulation: Phonetic process in which two individual consonants are pronounced at the same time. [12]

2. STATE OF THE ART

AND The state of the art of EEG signals in BCI covers a broad spectrum of research and developments in the acquisition and preprocessing of signals. These advances seek to overcome the challenges inherent in EEG signals and improve the accuracy and reliability of BCI systems. Knowledge of these advances and state-of-the-art techniques provides a solid foundation for this research and the development of more effective and accessible BCI systems.

2.1. Acquisition and preprocessing of EEG signals in BCI

2.1.1. Electrodes for capturing EEG signals

There are different techniques and electrode placement systems for the acquisition of EEG signals in the context of BCIs.

2.1.1.1. Types of electrodes

Two of the most common types are dry electrodes and wet electrodes:

- Dry electrodes: These do not require conductive gel and are easier to place, making them convenient for portable and home-use applications. However, dry electrodes tend to have inferior signal quality compared to wet electrodes, as the impedance between the electrode and the skin is higher, which can result in higher noise and lower sensitivity in detecting brain signals [10].

- Wet electrodes: These require the application of conductive gel to the scalp to improve electrical conductivity. These electrodes provide better signal quality due to the lower impedance between the electrode and the skin. However, they can be more laborious to place and maintain, and the conductive gel may cause discomfort or skin irritation [10].

2.1.1.2. Electrode placement systems

Regarding electrode placement systems, two of the most commonly used are the 10-10 system and the 10-20 system. These systems are based on measuring relative distances between anatomical reference points on the scalp to determine the precise location of the electrodes.

The 10-10 system is so named because it uses a grid of electrode locations on the head that are separated by a distance of 10% or 10% of the total distance in the anteroposterior and mediolateral plane, as can be seen in Figure 4, resulting in an equidistant distribution of electrodes across the scalp. This system provides broad coverage and is commonly used in research and clinical applications [11].

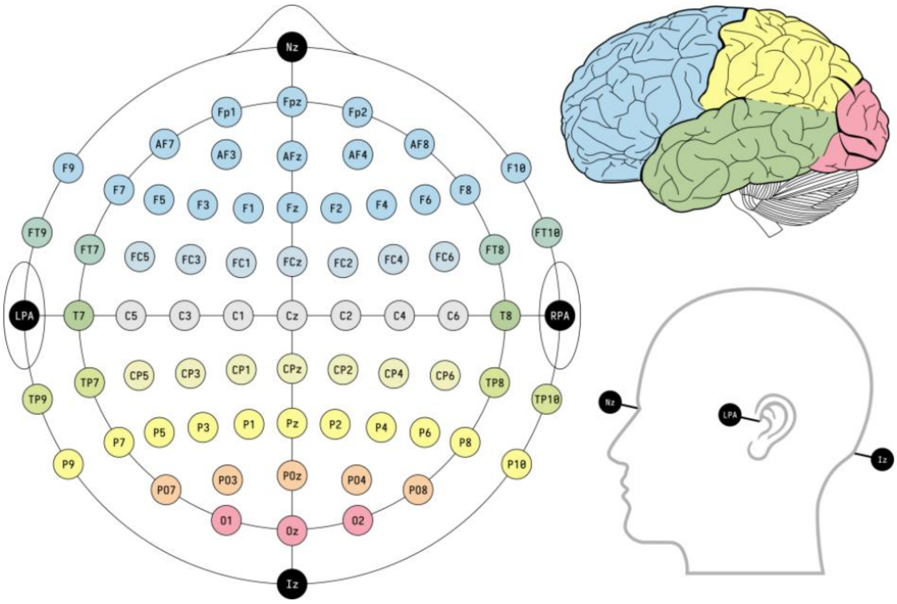

Figure 4. 10-10 system, image taken from Wikipedia

On the other hand, the 10-20 system uses a greater separation between electrodes, which implies a lower spatial resolution in the placement of these devices [11].

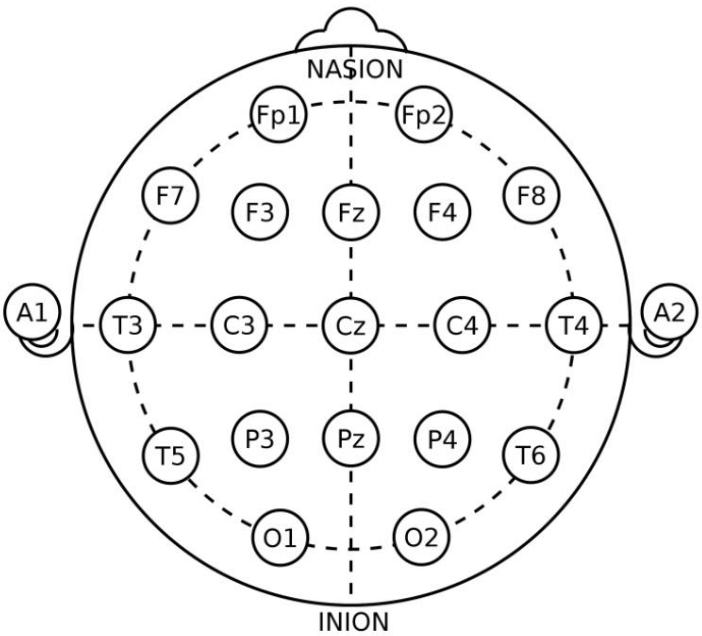

Figure 5. 10-20 system, image taken from Wikipedia

In addition to these systems, there is the 128-Biosemi system, which uses a radially distributed electrode array with 128 channels. This system provides high spatial resolution and allows for more detailed coverage of brain activity compared to the 10-10 and 10-20 systems, in addition to a different nomenclature [12].

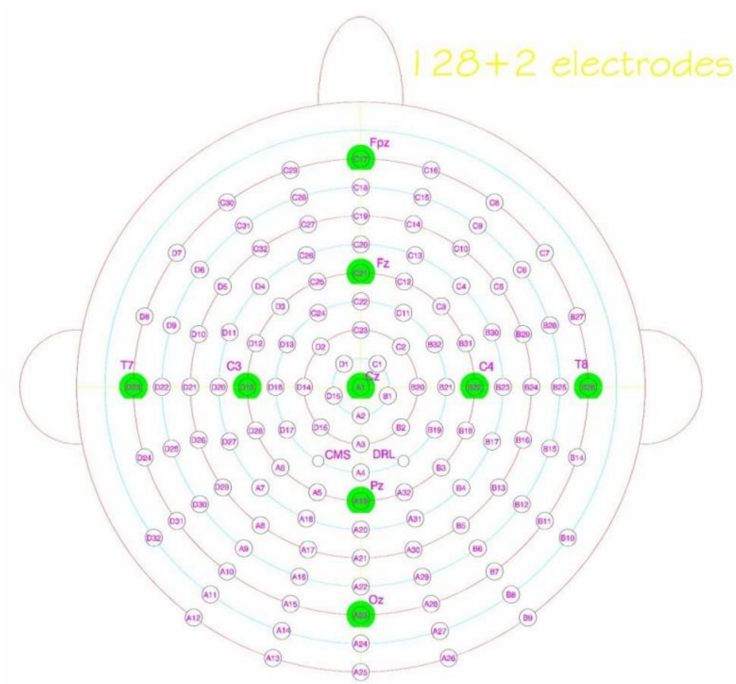

Figure 6. 128-Biosemi system, image taken from the Biosemi website [12]

2.1.2. Challenges associated with EEG signal acquisition

EEG signal acquisition presents several challenges due to the presence of noise and interindividual variability. These challenges can affect the quality and reliability of the measurements obtained. Each of these challenges is briefly described below:

2.1.2.1. Artifacts in EEG signals

During EEG signal acquisition, various types of noise may occur that can interfere with the neural signal of interest. Noise can originate from both internal and external sources, such as environmental noise, muscle movements, and electrical interference. Processing and analysis of EEG signals requires appropriate techniques to reduce and eliminate noise and improve signal quality [13].

2.1.2.2. Interindividual variability

Each individual has a unique brain configuration, which can lead to differences in the amplitude, shape, and location of EEG signals. This interindividual variability can make it difficult to compare results across different subjects and to generalize findings. It is important to take these individual differences into account when interpreting the results of EEG studies and to consider strategies to minimize their impact, such as the use of appropriate statistical analyses [14].

2.1.3. EEG signal acquisition and preprocessing techniques

The EEG signal acquisition and preprocessing techniques that will be mentioned in later sections are detailed below.

2.1.3.1. Re-reference

The idea behind re-referencing is to select a suitable reference or reference point that represents minimal or artifact-free brain activity. The choice of reference may vary depending on the specific objectives of the study and the electrode configuration used.

There are different methods of re-referencing, but two of the most common approaches are:

- Bipolar re-referencing: In this approach, the signal from each electrode is referenced relative to the signal from another nearby electrode. By subtracting the signals from two adjacent electrodes, some of the noise and interference common to both electrodes is removed.

- Reference averaging: In this method, the signal from each electrode is averaged relative to the signal from all other electrodes. This helps to eliminate unwanted references and accentuates the features of interest common to all electrodes.

The re-referencing process can be performed during real-time acquisition or as a subsequent step in the analysis of EEG signals. It is important to note that the choice of reference can affect the results and interpretations of EEG studies, and should be carefully considered depending on the specific context.

2.1.3.2. Digital filtering

In EEG signal pre-processing, digital filtering is a fundamental technique to improve signal quality and remove unwanted noise [14]. Digital filters are used to select frequencies of interest and remove those that are not relevant for the analysis.

There are different types of filters used in EEG, including:

- High-pass filter [13]: This type of filter allows the passage of frequencies above a certain threshold, eliminating lower frequencies. The high-pass filter is useful for removing low-frequency components, such as basal activity and motion artifacts.

- Low-pass filter [13]: On the contrary, the low-pass filter allows the passage of frequencies below a given threshold, attenuating higher frequencies. It is used to eliminate high-frequency noise, such as ambient electrical noise.

- Band-pass filter [14]: This type of filter allows frequencies within a specific range to pass through, while rejecting frequencies outside that range. It is used to select a range of frequencies of interest, such as those related to specific brain events.

- Notch filter [14]: This filter is designed to remove a specific frequency, such as a 50 Hz or 60 Hz power line frequency. It helps to remove electrical noise from the network and other specific artifacts [15].

Digital filtering is performed using mathematical algorithms that process the EEG signal efficiently and accurately. It is important to note that filtering must be applied carefully so as not to affect the features of interest in the EEG signals.

2.1.3.3. Epoching

Epoching in EEG refers to the process of dividing the continuous EEG signal into smaller, discrete segments called "epochs." Each epoch is a time window spanning a specific interval of duration, typically a few hundred milliseconds to a few seconds.

It is performed to facilitate the analysis of the EEG signal, as it allows the examination of specific characteristics of interest in defined time intervals. By dividing the signal into epochs, signal processing techniques and statistical analysis can be applied to each epoch individually.

The key points of EEG epoching are [15]:

- Temporal segmentation: The continuous EEG signal is divided into discrete intervals called epochs.

- Epoch duration: Each epoch has a specific duration, which can vary depending on the study and the type of analysis performed.

- Individual epoch analysis: Each epoch is considered an independent unit of analysis, allowing specific signal characteristics to be examined at defined time intervals.

EEG epoching is a crucial stage in EEG signal processing and is widely used in various studies and analyses such as ERP analysis and brain oscillation analysis. It allows for greater flexibility and accuracy in the analysis of EEG-recorded brain activity.

2.1.3.4. Decimation

Decimation in EEG (or signal processing) refers to the process of reducing the sampling rate of an EEG signal to decrease the amount of data without losing important information. This can be useful to reduce the size of data files and facilitate further processing [16].

When applying decimation, it is important to consider the appropriate cutoff frequency to avoid aliasing, which can distort the EEG signal. Appropriate filters should be applied prior to decimation to remove unwanted frequencies and ensure signal integrity [17].

Decimation may introduce some loss of information into the EEG signal due to discarding samples. However, this impact can be minimized if performed properly and an appropriate decimation rate is selected [18]. This technique can be especially useful in situations where a lower sampling rate is needed, such as in real-time applications where lower latency is required in EEG signal processing [18].

It is important to consider the effects of decimation on the frequency spectrum of the EEG signal. Reducing the sampling rate may affect the spectral representation and frequency resolution, which should be considered in the subsequent analysis [19].

2.1.3.5. Control of ocular and muscular activity

It is crucial to control and mitigate artifacts caused by eye movement (EOG) and muscle activity (EMG). These artifacts can introduce interference into the EEG signal and hinder accurate interpretation of the data.

- Eye movement artifacts (EOG) [20]: Eye movement generates electrical signals that can contaminate the EEG signal. These artifacts can be eliminated or reduced by simultaneously recording eye movements, usually using electrooculograms (EOG) or eye-tracking sensors. These eye recordings are then used as reference signals to remove eye movement-related artifacts from the EEG signal.

- Muscle activity artifacts (EMG): Muscle activity, especially in facial and neck muscles, can introduce artifacts into the EEG signal. These artifacts can be the result of voluntary or involuntary movements, such as blinking, chewing, or muscle tension.
  To mitigate these artifacts, electromyography (EMG) recording techniques can be used to monitor and detect concurrent muscle activity. By identifying the moments of muscle activity and applying artifact removal methods, the effects of muscle activity on the EEG signal can be minimized [21].

-

2.1.3.6. ICA

ICA, or Independent Component Analysis, is a statistical method used to separate mixed signals into independent components [22]. In EEG signal processing, ICA is used to separate brain signals into independent sources, such as signals generated by different areas of interest in the brain and artifacts caused by eye movements, blinks, muscle activity, the heart, and other unwanted noise.

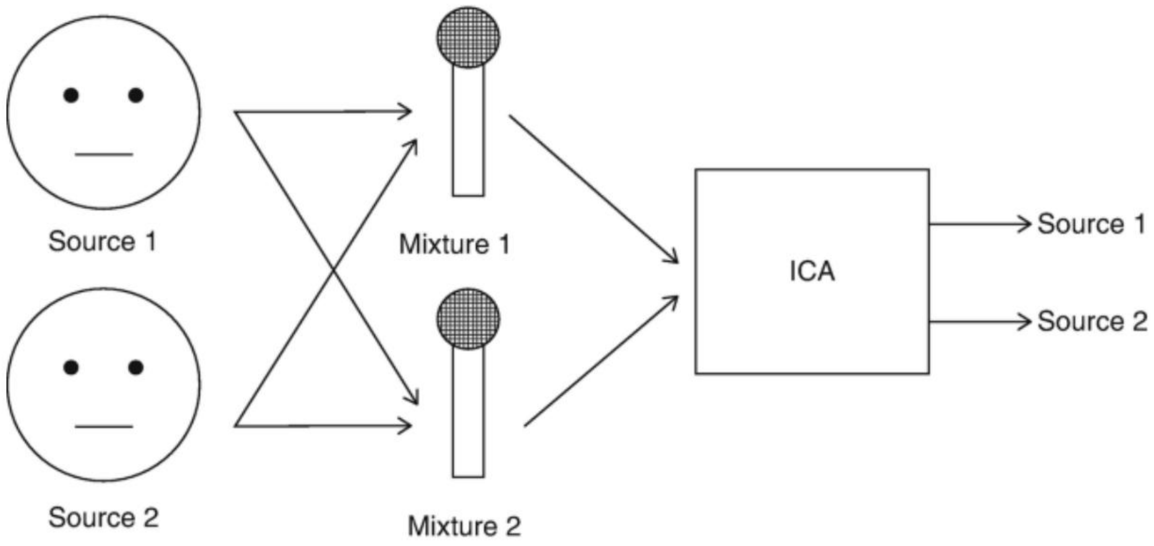

Figure 7. Illustration of the ICA[2] concept

In the linear mixture model used in ICA[3] . The EEG signal recorded on multiple channels is represented as a vector = [1 2 …] of dimensions, where each element represents the signal on a specific channel. This mixed signal is generated by independent sources represented by the vector = [1 2 …] dimensions:

=

The relationship between the mixed signal and the original sources can be described by a mixing matrix. The task of the ICA is of dimensions to find a version = of the original sources, where is a square matrix specifying spatial filters that reverse the linear mixing process.

Mathematically, the ICA seeks to identify the matrix that approximates the inverse of the mixing matrix such that the separated ÿ1 . signals are similar to the original sources, except for the scale and the permutation. This can be represented as:

ÿ ÿ1

Therefore, the resulting matrix can be obtained by:

= = ( ) = ÿ ÿ

It is assumed that the original signals s are statistically independent at any time instant , and that the mixing of these sources is a linear process. The objective of the ICA is to find the spatial filters represented by the matrix that allow undoing the mixing and recovering the original sources with the greatest possible independence.

2.2. Signal processing, feature extraction and selection

Wavelet transforms are mathematical tools used in signal and image analysis to capture features that vary at different scales [23]. Unlike the Fourier transform, which is based on sinusoids of fixed frequencies, wavelet transforms decompose the signal into wavelets, which are localized oscillating functions in both the time and frequency domains [23].

The continuous wavelet transform (CWT) is well suited for analyzing non-stationary signals, i.e. those whose properties vary over time. The CWT maps the time-frequency plane using variable-sized windows, allowing it to capture both high-frequency and low-frequency components accurately. This makes it useful for analyzing signals with transients, changing frequencies, and trends.

2 Image taken from https://link.springer.com/referenceworkentry/10.1007/978-1-4939-7131-2_147

3 ICA concept, extracted and modified from [22]

slow [23], an example of which are Morlet wavelets.

On the other hand, the discrete wavelet transform (DWT) performs a decomposition at discrete scales, making it an effective tool for multi-resolution analysis and compression of signals and images. DWT allows signal components to be separated at different levels of detail, facilitating noise removal and preserving important features [23]. This will be the case for Daubechies-4 wavelets.

That is, wavelet transforms offer a more versatile and flexible representation than the Fourier transform, since they can capture local features in time and frequency.

2.2.1. Morlet Wavelet

Time-Frequency analysis uses a sinusoidal wave that varies in time and is called wavelets. These are oscillations whose amplitude starts at zero, increases and then decreases. Morlet wavelets are the most commonly used in this context [24]. The Morlet wavelet transform is a complex function that combines a modulated sinusoidal function (carrier) by a Gaussian function (envelope).

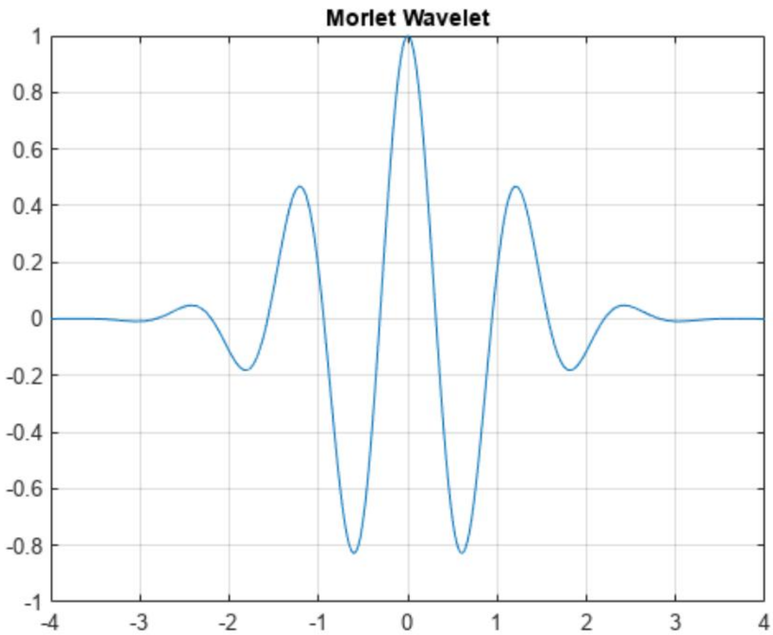

Figure 8. Morlet Wavelet, image taken from the MATLAB® website

It is defined by its mathematical form [25], which involves two main parameters: the central frequency (carrier frequency) and the temporal standard deviation . These parameters determine the frequency and time resolution of the wavelet, respectively. In [25], spectral estimates are obtained using Morlet wavelets (,):

$$(,) = (ÿ) - \frac{1}{2} \quad \frac{1}{2 \quad 2}^{2} \quad ÿ 2$$

Then, the time-frequency estimate (,) of a signal ( ) is calculated by convolution with (,):

$$(,) = ( ) ÿ (,)$$

Time shifting of the Morlet wavelet allows different parts of the signal to be analyzed at different times. By shifting the wavelet in time, different representations of the signal are obtained at different time scales.

The Morlet function provides a localized representation in time and frequency, which allows to detect brief and transient events in an EEG signal, and is therefore especially useful for analyzing non-stationary signals, such as signals that vary in time, as is the case with many brain signals. However, this estimate will not be used in our development in MATLAB® (which will be discussed later) but will be implemented with the help of the *cwtfilterbank()* function to calculate both the frequency bands (ÿ, ÿ, ÿ, ÿ) and the frequency-time scalograms of the EEG signals, taking into account that a complex wavelet is used.

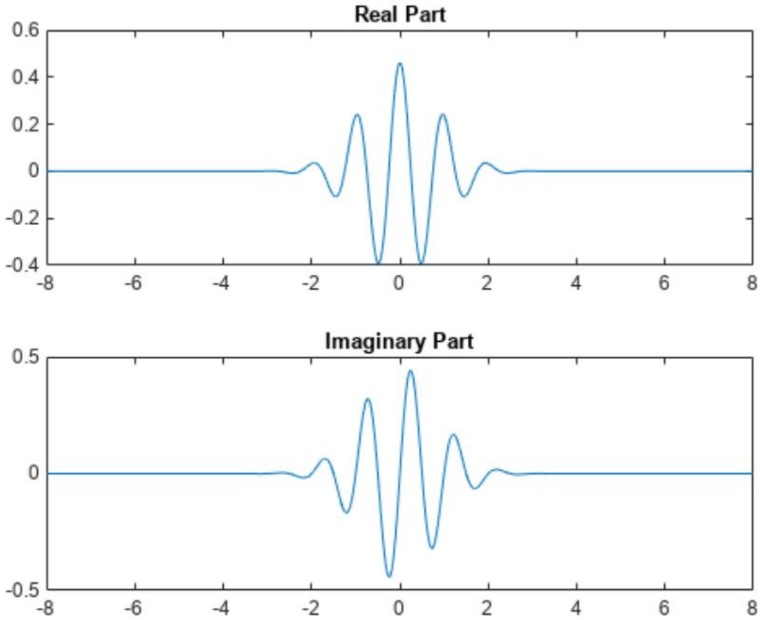

Figure 9. Complex Morlet wavelet, image taken from the MATLAB® website

2.2.2. Daubechies 4 (D4)

Daubechies wavelets are a family of wavelets that are widely used in signal processing and data compression. Daubechies 4 (D4), named for its 4 coefficients, is one of the simplest and most popular wavelets in this family.

The D4 wavelet is defined in terms of a scaling function and a wavelet function [26]. These functions are used in the decomposition and reconstruction of a signal at different levels of detail.

The scaling function ÿ(t) is used to approximate the low frequency components of the original signal [27]. This function is a dilated and shifted version of itself, meaning that it is spread out over time and shifted to cover different parts of the signal. The scaling function has the property of being orthogonal, implying that the energy of the original signal is conserved at each level of decomposition.

The wavelet function ÿ(t) is used to capture the high-frequency details of the original signal [27]. Like the scaling function, the wavelet function is also dilated and shifted, but it has oscillatory characteristics that allow detecting rapid changes in the signal. The wavelet function has the property of being orthogonal to the scaling function, which means that the signal details are captured independently of the low-frequency approximation [26].

Decomposition using the D4 wavelet involves passing the original signal through a filtering and downsampling process. At each decomposition level, convolution filters are applied to both the low-frequency approximation (AC) and high-frequency detail (DC) obtained at the previous level [28]. This allows the different components of the signal to be separated into different levels of detail. The decomposition is repeated iteratively until the desired decomposition level is reached [28].

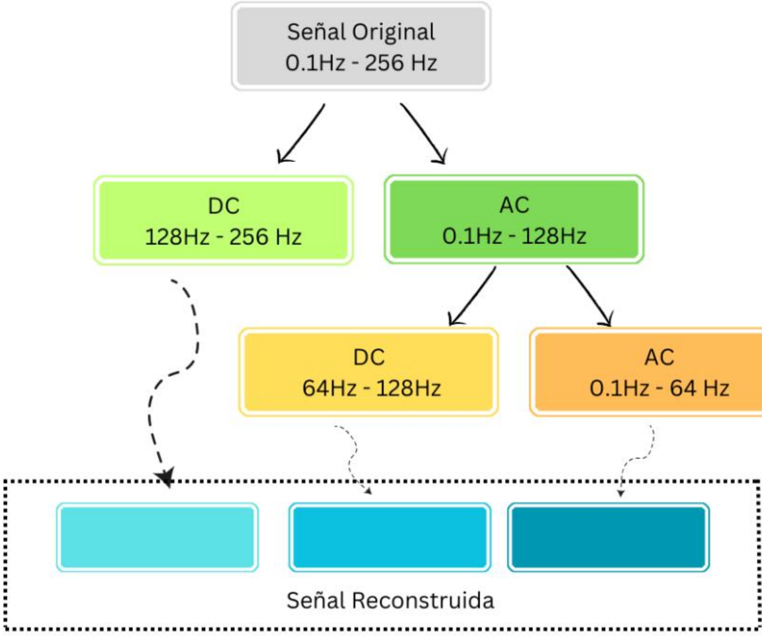

Figure 10. Decomposition using D4

Signal reconstruction involves the inverse process, where the detail coefficients and the low-frequency approximation obtained in the decomposition are used to reconstruct the original signal. This is achieved by applying inverse convolution filters and combining the different approximations and details at each level. It will be especially useful in the feature extraction of the MindBigData dataset, since the experiment performed in [28] will be replicated, where an energy threshold is applied to each coefficient, eliminating those that are small, essentially considering them as noise.

2.3. Neural networks for word recognition

To classify the features extracted from the EEG signal, researchers have used both classical Machine Learning and Deep Learning algorithms. Both are methods that provide computers with the ability to learn and recognize patterns. In the case of BCI, the patterns to be recognized are the features extracted from the EEG waves, and then, based on what the computer has learned, some predictions are made to classify the signals.

Thanks to the study carried out in [1], it can be summarized that:

Several classical machine learning techniques have been used to address imaginary speech decoding for EEG-based BCI systems. Some of the most common algorithms include Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), Random Forests (RF), k-Nearest-Neighbors (kNN), and Naive Bayes.

Furthermore, Deep Learning approaches have recently gained prominence in imagined speech recognition. Some of these techniques are Extreme Learning Machine (ELM), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and parallel CNN+RNN.

2.4. Results of studies in the literature

Based on everything mentioned in this section 2, there are several studies that try to achieve high performance of imagined speech by using BCI systems, which involves various feature extraction and classification methods, as well as different tasks. Table 1, Table 2 and Table 3 summarize the most relevant experiments found in the literature for 'Thinking Out Loud' [3], 'MindBigData' and others. datasets, respectively.

From this point on, the terms "accuracy" and "precision" will be used interchangeably to refer to the same concept in the context of this study. Both terms will be used to describe the measure of accuracy or success in the classification tasks performed, defined as:

$$= \frac{+}{+\ +\ +}$$

where:

- TP (True Positives) represents the number of correctly classified positive cases (true positive).
- TN (True Negative) represents the number of correctly classified negative cases (true negative).
- FP (False Positive) represents the number of negative cases incorrectly classified as positives (false positives).
- FN (False Negative) represents the number of positive cases incorrectly classified as negative (false negatives).

This equivalence in terminology is intended to avoid confusion and facilitate understanding of the results and conclusions presented in this work.

Author(s)	**Brain Area/s**	**Neural Network**	**Task**	**Accuracy**
[29]	All areas (128 Biosemi)	2D CNN based on the architecture EEGNet	Classification of 4 words in Spanish: 'up', 'down', 'left', 'right'	29.7%
F. Gasparini et al. [30]	All areas (128 Biosemi)	LSTM	Classification of 4 words in Spanish: 'up', 'down', 'left', 'right'	30.4%
F. Gasparini et al. [30]	All areas (128 Biosemi)	BiLSTM	Classification of 4 words in Spanish: 'up', 'down', 'left', 'right'	36.1%

Table 1. Studies in the literature for the 'Thinking Out Loud' dataset [3]

Author(s)	Brain Area/s	Neural Network	Task	Accuracy
Bird et al. [31]	FP1, FP2, TP9, TP10 (MUSE)	Attribute selection + MLP	10-digit classification: 0-9	30%
Jolly et al. [32]	FP1, FP2, TP9, TP10 (MUSE)	Unidirectional-GRU	10-digit classification: 0-9	33.8%
Pratama et al. [33]	FP1, FP2, TP9, TP10 (MUSE)	FFT + k-NN	10-digit classification: 0-9	31%
NC Mahapatra & Bhuyan P [28]	FP1, FP2, TP9, TP10 (MUSE)	D4 + bidirectional-LSTM	10-digit classification: 0-9	96.2%

Table 2. Studies in the literature for the 'MindBigData' dataset [34]

Author(s)	Brain Area/s	Neural Network	Task	Accuracy
García et al. [34]	F7, FC5, T7, P7 (Wernicke's area)	Naive Baye classifier, Random Forest, SVM and Bagging-RF	Classification of 5 words in Spanish	>20%
Sarmiento et al. [35]	Broca's areas and Wernicke	SVM	Classification of 5 vowels: /a/, /o/ (open), /e/, /i/, /u/ (closed)	84-94%
Pawar & Dhage [36]	Prefrontal cortex, Broca's area, Wernicke's area	Kernel Based extreme machine learning	Classification of 4 English words: 'left', 'right', 'up', 'down'	multi-class: 49.8%, binary: 85.5%
Cooney et al. [37]	All areas	1. CNN 2. ICA/LDA	1. Classification of 5 vowels: /a/, /e/, /i/, /o/, /u/ 2. Classification of 6 words in Spanish: 'left', 'right', 'up', 'down', 'forward', 'back'	1. 35% 2. 30%

Tamm et al. [38]	F3, F4, C3, C4, P3, P4	CNN	Classification of 5 vowels and 6 words	24%

Table 3. Studies in the literature for other datasets

3. BCI DESIGN AND IMPLEMENTATION

The true method of knowledge is experiment.

- Willian Blake -

In the framework of this study, two types of brain signal analysis were performed. First,
AND makes use of a signal acquisition specific to word recognition through
EEG, studying the feasibility of a relatively simple 3D headband design that seeks to facilitate this capture. These signals were recorded under controlled conditions (such as the use of an anechoic chamber), which ensured consistency in the way they were obtained and processed.

Second, previously collected datasets available in the scientific literature were used. These datasets provided an additional source of brain signals recorded in different contexts and by different research groups. By using these datasets, the data sample was expanded and the performance of the techniques and models was evaluated under more varied conditions and scenarios.

3.1. Signal acquisition systems

3.1.1. Details of own signals

One of the objectives of this work is to study the feasibility of a relatively simple design of a 3D headband that seeks to facilitate the acquisition of signals specific to word recognition using EEG. The headband is held on the subject's head by an adjustable elastic band with a hook and loop closure.

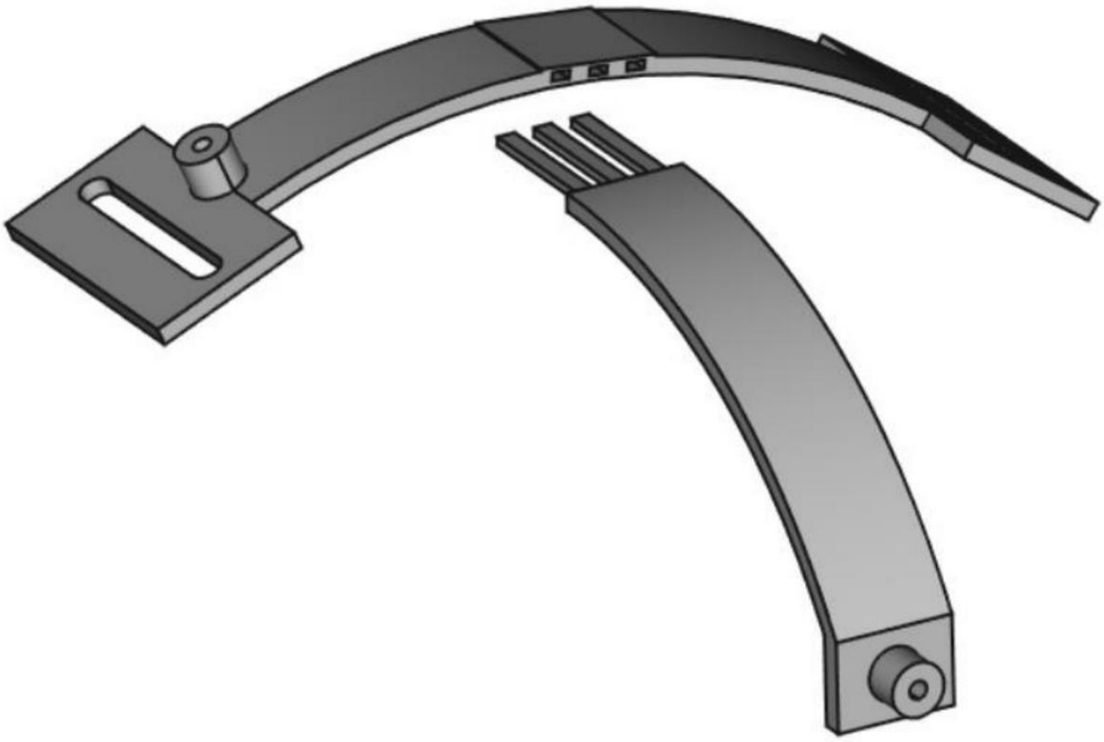

Figure 11. 3D design of the headband for own experiment

3.1.1.1. Experiment design

Therefore, the strategic choice of positions for a reduced number of electrodes will be key. By using a limited number of strategically placed electrodes, the complexity of the acquisition system is reduced and interference from unwanted signals is minimized. This would contribute to improving the quality of the recorded EEG signals and simplify the process of word analysis and recognition.

In particular, the Fpz and F3 positions have been selected for the differential EEG measurement, and A1 as a reference. This decision was made based on Broca's area and its relevance in word formation in language processing, as explained in section 1.2.3.1. The signals acquired through these three electrodes will be monitored by Mikrobus® 'EEG click', a device based on the INA114 instrumentation amplifier from Texas Instruments.

Revenue	Up to 10.000 (adjustable) by potentiometer)
GND virtual	2.048 V
Connector	Jack 3.5mm

Table 4. Features of the 'EEG Click' board

Figure 12. 'EEG Click' device, image taken from its datasheet

After this amplification stage, the differential signal is received by the Atmel® SAM E70 Xplained board, whose microcontroller is the powerful Atmel ATSAME70Q21. The signal is digitized with a sampling frequency of 256 Hz and sent to the MATLAB software via serial port, where a Notch filter at 50 Hz is applied to eliminate the power supply noise and Morlet wavelets are applied in real time to avoid the well-known edge effect [39].

Figure 13. Atmel® SAM E70 Xplained, image taken from its datasheet

The experiment carried out will be based on the one designed in 'Thinking Out Loud'4 [3], where it is intended to distinguish four words in Spanish: 'left', 'right', 'up', 'down'. However, only the 'inner speech' paradigm will be executed. The recordings have been recorded both in the Master's Laboratory of the department and in an anechoic chamber available at the School, designed to fully absorb the reflections produced by acoustic or electromagnetic waves. The purpose of making the recordings in this room is to eliminate any external noise, since the work is done with very small wave amplitudes.

4 See section 3.1.2.1 Thinking Out Loud

Machine Translated by Google

3.1.1.2. Implementation of the experiment in MATLAB

The experiment is designed using the MATLAB App Designer. It consists of a first interface (see Figure 14) where the correct reception of data via serial port is checked and the Notch filter is activated at 50Hz. Once this has been checked, the experiment is started using the 'Start' button.

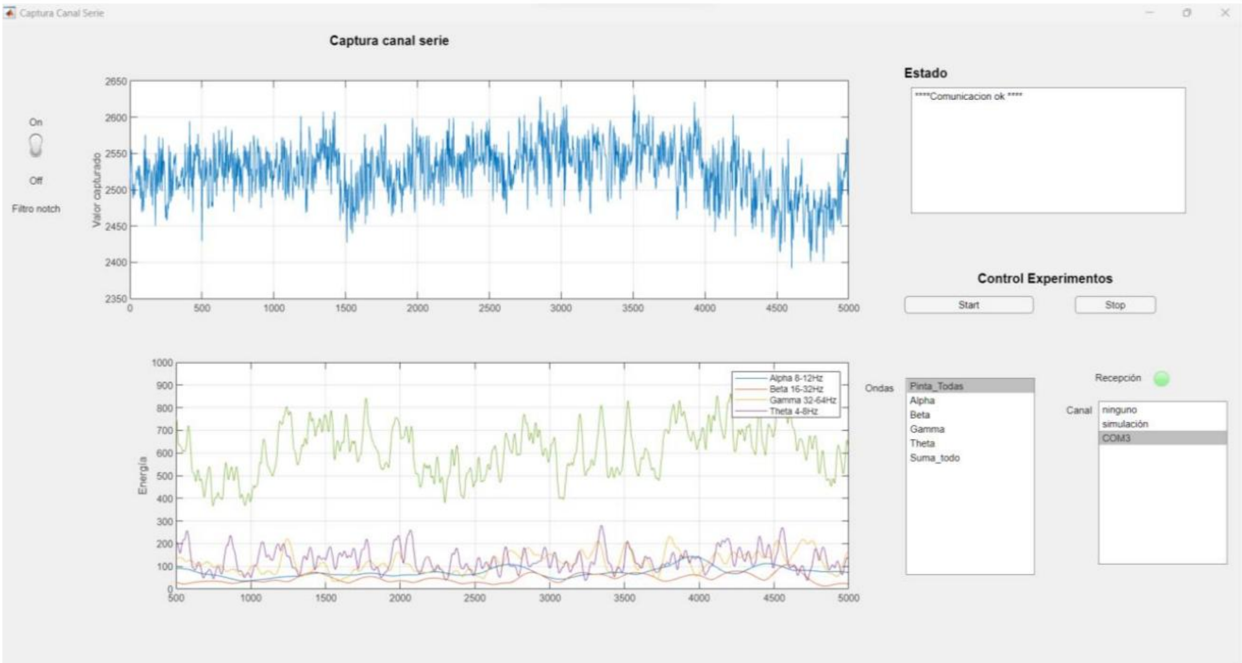

Figure 14. First interface of the own experiment

Once the experiment has been launched, the first concentration interval begins, with a screen appearing with the caption shown in Figure 15, so that the subject making the recording can concentrate on the task. The sequence shown in Figure 16 is then followed.

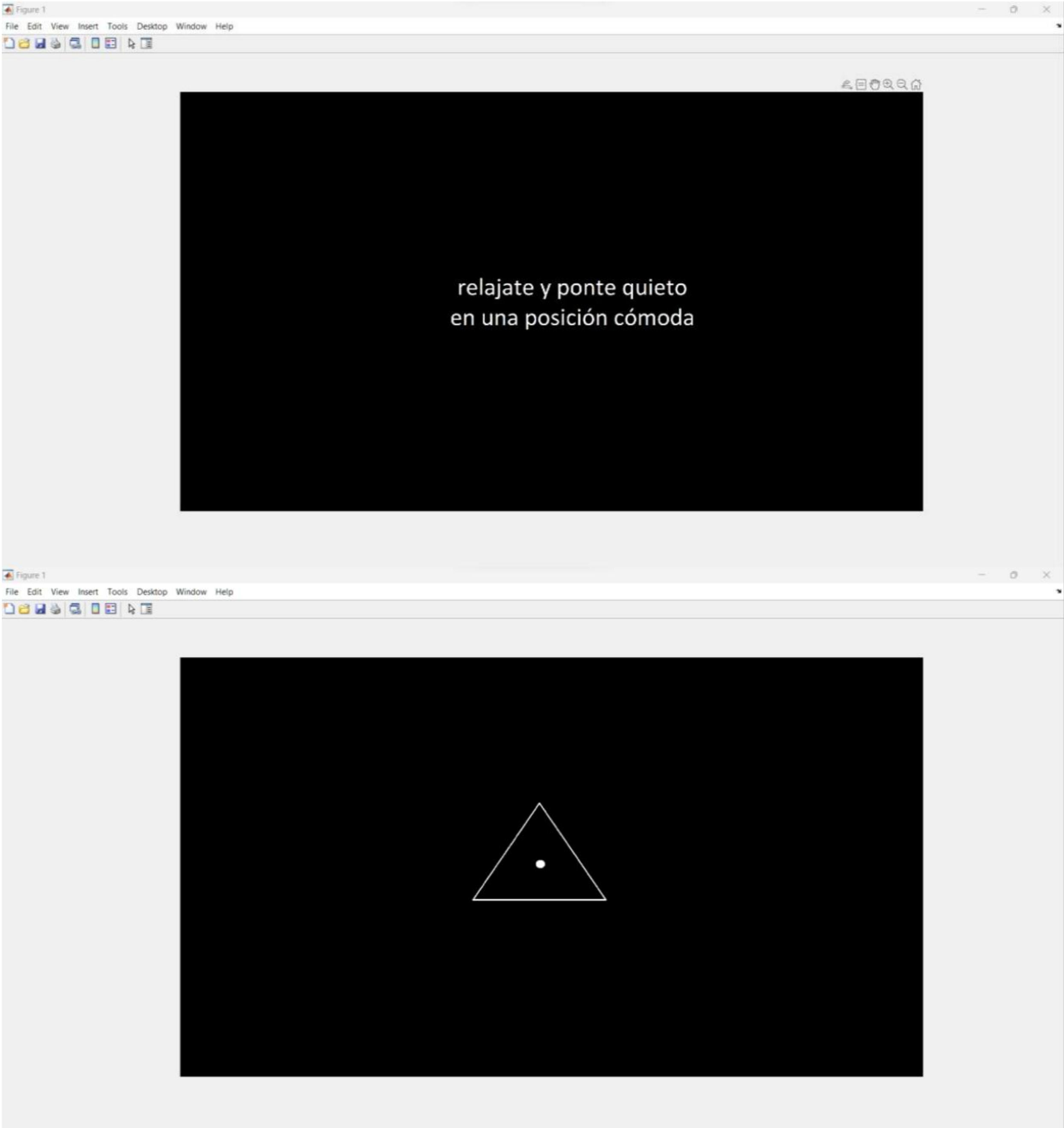

Figure 15. Initial screen and 'Cue Interval' of the experiment itself

3.1.2. Description of the datasets used in the literature

The first bibliographic dataset used comes from the "Thinking Out Loud" study [3]. On the other hand, the "MindBigData" dataset [2] will also be analyzed, specifically the EEG signals acquired with the commercial device Interaxon Muse [40].

3.1.2.1. Thinking Out Loud

In this experiment, ten healthy subjects with no hearing or speech loss participated, with no previous experience in BCI [3]. The experiment consisted of performing the three paradigms described in section 2.a: "inner speech", "pronounced speech" and "visualized condition".

viewed).

Each subject participates in three consecutive sessions, each structured by a 15-second baseline recording, followed by one execution of the "pronounced speech" paradigm, two executions of the "inner speech" paradigm, and two executions of the "visualized condition" paradigm.

It is important to note that the order of the selected Spanish words ("up", "down", "right" and "left") is presented randomly during the experiments.

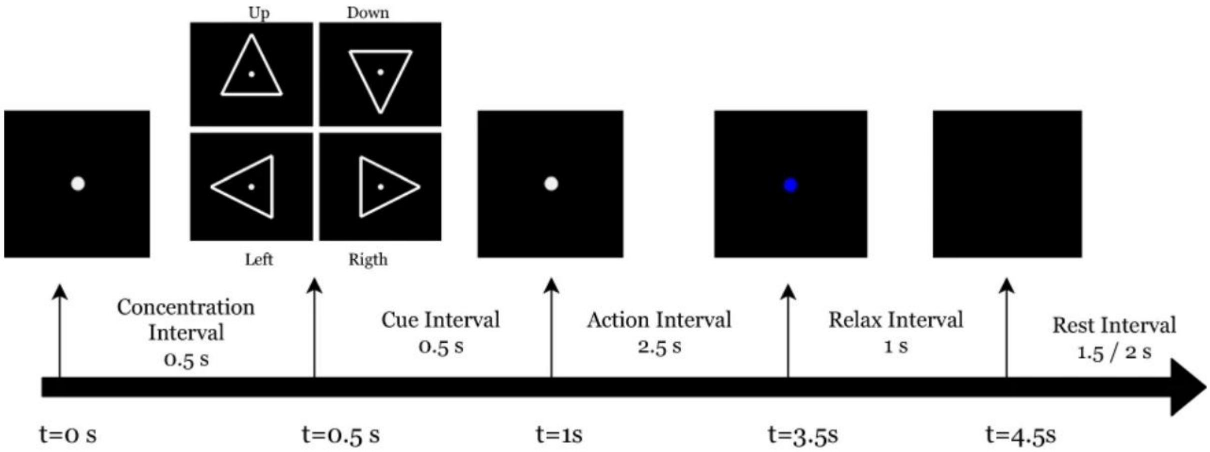

Figure 16. Sequence of each trial, taken from the original Thinking Out Loud article [3].

During the experiment, a white circle is shown in the center of the screen and the subject is asked to stare at it without blinking. Then, a white triangle is presented pointing in one of four directions corresponding to the selected Spanish words. When the triangle disappears and the white circle is presented again, the subject must perform the indicated task. Task execution is stopped when the white circle turns blue. The subject is instructed to control eye blinking until the circle disappears.

To assess participants' attention, they are asked to indicate the direction of the last stimulus after a random number of trials. Responses are given using the arrow keys on the keyboard and the answer is displayed on the screen.

Data are acquired using a system of 128 active wet electrodes for EEG (Biosemi 128 system) and 8 external ones for EOG/EMG. A resolution of 24 bits and a sampling frequency of 1024 Hz are used. Prior to analysis, the authors of the experiment preprocessed the EEG signals as follows [3]:

- Reference: Because the BioSemi acquisition system is “reference free”, a re-referencing step was performed using channels EXG1 and EXG2 to eliminate common mode voltage and reduce line noise and body potential drifts.
- Digital filtering: A finite impulse response bandpass filter with frequency cutoffs from 0.5 to 100 Hz was applied to keep the data as raw as possible, allowing future users to filter the data to their needs. A notch filter at 50 Hz was also applied to remove line noise.
- Segmentation and decimation: The data are decimated four times to obtain a final sampling frequency of 254 Hz. The recorded continuous data are then segmented, retaining only the 4.5 second signals corresponding to the concentration and relaxation interval.
- ICA Analysis: Independent Component Analysis (ICA) is applied to the EEG channels using the MNE implementation of the infomax ICA. 128 sources are captured and correlation with the EXG channels is used to identify and exclude sources related to blinking, gaze and mouth movement, thus obtaining the final dataset.
- EMG control: The simplest approach to detect electromyographic activity is to establish an energy threshold for these signals. This control is used to identify and discard trials in which unwanted muscle activity is detected, which is important to ensure the quality of the results.

data in the study.

3.1.2.2. MindBigData

In this experiment, a single healthy subject participates. Brain signals are captured while the subject views for 2 seconds a single digit (displayed on a 65" television screen with white font on a completely black background) from 0 to 9, as shown in Figure 17.

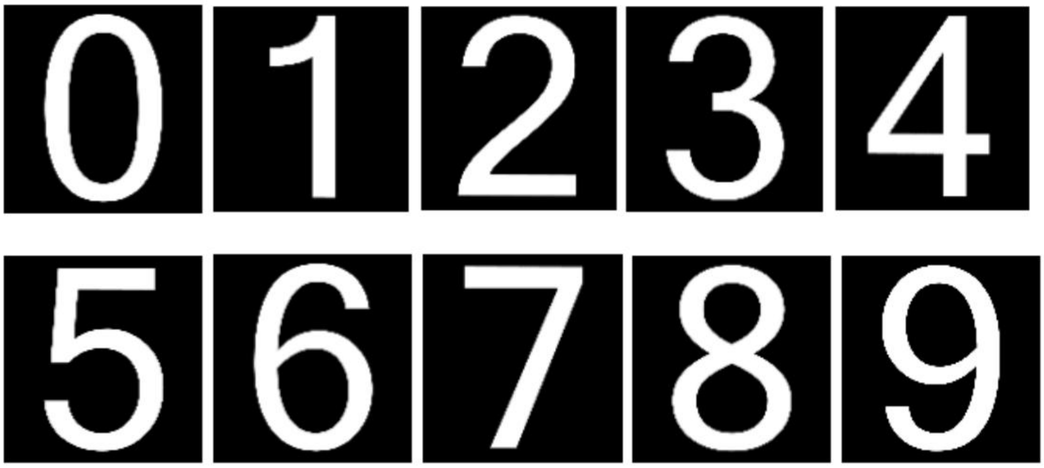

Figure 17. Digits shown in 'MindBigData', extracted from its original article [41].

The appearance of the digits is random, with a rest interval between them (black screen). A total of 163,250 brain signals of 2 seconds each are recorded, labeled with the corresponding digit. 59,339 signals are also captured while the subject was performing other mental activities unrelated to the digits, with no limitation on movements, blinking, or other actions, and these are labeled as "-1" [41].

The captured EEG channels are four (FP1, FP2, TP9 and TP10, of the 10-10 system) and a theoretical sampling frequency of 220 Hz is used [41].

The signal measures the variation in voltages of brain neural activities and is provided in "raw" form, as captured from the device. The author does not apply any preprocessing, although it is decided to apply Daubechies-4 and an energy threshold to each decomposition to remove irrelevant components.
(turning them into null if they do not exceed the threshold), as proposed in [28]. It should be noted that, in the present study, only a binary classification is performed between the digits '0' and '1'.

3.2. Signal processing, feature extraction and selection

The purpose of this stage is to prepare the data, extracting features, for later analysis and classification in a neural network, to which we will introduce the aforementioned features, 'inputs'. To do this, we will use EEG signals, frequency bands or spectrograms.

2.4.1. EEG signals as inputs

EEG (electroencephalography) signals will be used directly as input data for word recognition. Both for the EEG signals themselves and for signals from other sources such as "Thinking Out Loud" [3] and "MindBigData", pre-processing is performed. This processing includes filtering, re-referencing, segmentation and other techniques to improve the quality of the signals and highlight the relevant features for word recognition explained in section 3.1.

Signal acquisition systems.

2.4.2. Frequency bands and spectrograms as inputs

In this study, spectrograms generated using, as previously mentioned, one of the most widely used techniques in EEG signal processing: the Morlet transform, will be used as input for the neural networks.

The Morlet wavelet is particularly suitable for calculating the energy of the frequency bands used in our experiment, since it avoids the edge effect (distortion at the ends of the signal) by applying the transform in real time, thus using the part of the signal that is outside the interval of interest. This means that a more accurate representation of the frequency bands present in EEG signals (ÿ, ÿ, ÿ, ÿ) can be obtained. It will also be applied to the existing datasets in the literature that are studied in this work, 'Thinking Out Loud' [3] and 'Mindbigdata', taking into account that the edge effect will exist when using already segmented signals [39].

In addition to frequency bands, spectrograms generated from EEG signals will also be used as input data [42]. These spectrograms provide visual information about the energy distribution at different frequencies over time.

3.3. Implementation of neural networks

This section presents the different approaches and architectures explored in this project with the aim of developing an efficient and accurate system for word recognition from EEG signals. It is worth highlighting the use of the technique known as 'Transfer Learning' [43] for the adaptation of existing neural networks in the literature.

3.3.1. Deep learning for word recognition from audio

The neural network proposed in [43] is used for word recognition from audio signals, adapted to EEG signal processing.

Within this approach, three variants are presented:

3.3.1.1. The dataset is close

In this variant, we will use exclusively EEG signals from the anechoic room for training and validation.

Each test consists of one channel and six bands, including five frequency bands and the EEG signal.

3.3.1.2. Thinking Out Loud

In this variant, both the EEG signals are used directly (with 128 channels and one band per trial) and the frequency bands in a separate channel (one channel and five bands per trial). Tests are performed with the three paradigms both interchangeably and separately, with the 'Inner Speech' paradigm being the most relevant for our analysis.

3.3.1.3. MindBigData

In this variant, both the EEG signals will be used directly (with 4 channels and one band per test) and the frequency bands in a separate channel (one channel and five bands per test). Only one test will be performed. preliminary study with binary classification between the digits '0' and '1'.

3.3.2. Typical LSTM network architecture

A Long Short-Term Memory (LSTM) network is a type of recurrent neural network (RNN) that is capable of learning long-term dependencies between time units in sequential data.

The architecture of an LSTM network consists of a sequence input layer and an LSTM layer. The sequence input layer allows sequential or time series data to be fed into the network, while the LSTM layer learns the long-term dependencies between time units in the sequential data.

Figure 18 shows the architecture of a simple LSTM network for classification. The network starts with a sequence input layer followed by an LSTM layer. To predict class labels, the network ends with a fully connected layer, a softmax layer, and an output classification layer.

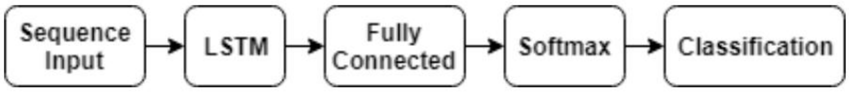

Figure 18. Diagram of a typical LSTM[5] network

Within this architecture, the same data are presented as in section 3.3.1. Deep learning for word recognition from audio.

3.3.3. Time-frequency convolutional network for EEG data classification

A network is adapted to classify electroencephalogram (EEG) time series from people with and without epilepsy in the time-frequency domain. The convolutional network (TF CNN) uses the continuous wavelet transform (CWT) to represent the EEG data in the time-frequency domain. Unlike other networks that use the spectrogram as a preprocessing step, this network uses a differentiable spectrogram layer, allowing to apply learnable operations before and after the spectrogram. This layer extends the architectural possibilities by working with time-frequency transforms.

Within this architecture, the same data as in section 3.3.1 are presented. Deep learning for word recognition through audio, excluding frequency bands. This is because the network makes use of the CWT in one of its layers and it would be pointless to apply it on a Time-Frequency transform.

3.3.4. GoogleNet y AlexNet

GoogleNet and AlexNet are two widely recognized convolutional neural network (CNN) architectures in the field of image recognition.

By using the spectrograms generated from EEG signals as input, we will be able to leverage the deep learning capabilities of GoogleNet and AlexNet to extract meaningful features and perform word classification based on the patterns identified in the spectrograms.

This choice of input representation could allow us to take advantage of the ability of convolutional neural networks to recognize spatial and frequency patterns in EEG data, which may improve the accuracy of word recognition in our specific context.

It is important to note that both GoogleNet and AlexNet have been widely studied and used in the scientific community, and have demonstrated good performance in various image classification tasks on datasets such as ImageNet. These architectures are adapted in the MATLAB Deep Network Toolbox for word recognition from EEG signal spectrograms.

5 Image taken from https://es.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html

3.3.4.1. The dataset is close

No training is performed for this dataset.

3.3.4.2. Thinking Out Loud

In this variant, both spectrograms for one channel and for the assembly of 16 channels are used. In both cases, the images will be introduced into the neural network without titles or legends. The frequency range (0.1Hz-88Hz) of the Morlet transform will be complete (i.e. the vertical axis) compared to the samples of the complete interval in time (horizontal axis) in the case of a single image (shown in Figure 19). However, for the assembly it is decided to cut the vertical axis, thus leaving out many frequencies. The frequency band used will be 4-64Hz.

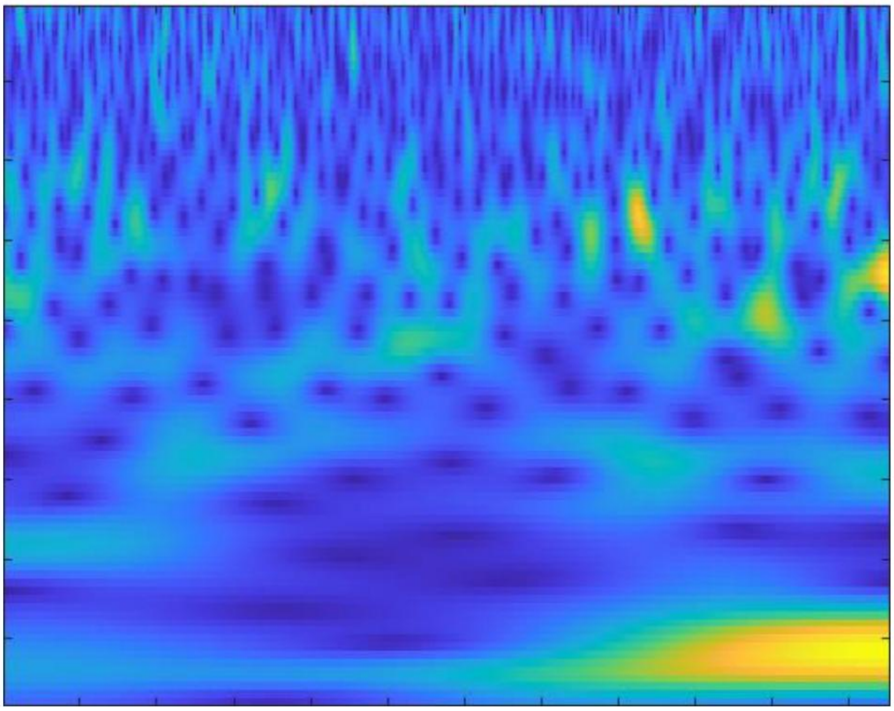

Figure 19. Spectrogram of channel A156 for a trial of the word 'down' (Thinking Out Loud)

As can be seen in Figure 20, the spectrogram of each channel is converted into a rectangular shape so that the set of all of them is square (a necessary condition for the correct functioning of these latter neural networks). Finally, the images will be resized to the input size corresponding to GoogleNet or AlexNet.

6 A15: Nomenclature for Biosemi 128 system electrode

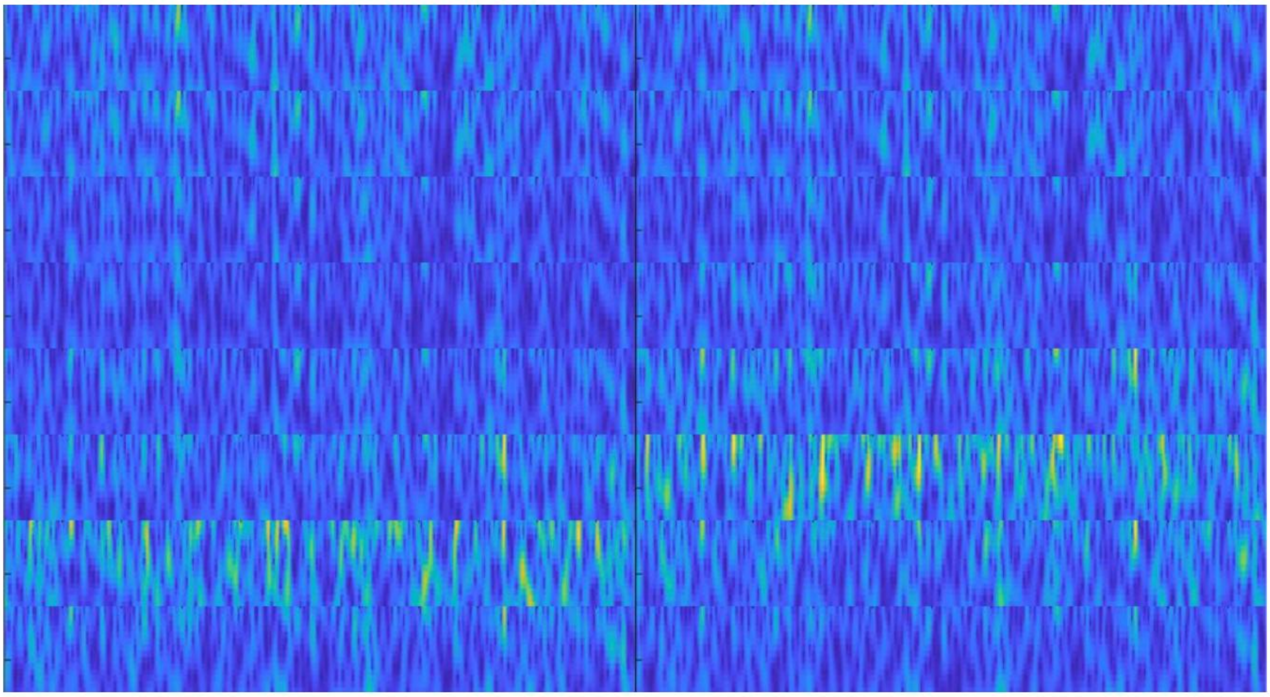

Figure 20. Montage of 16 spectrograms for a trial of the word 'down' (Thinking Out Loud)

3.3.4.3. MindBigData

In this variant, you will follow the same steps as for the previous dataset, but taking into account that we only have 4 channels.

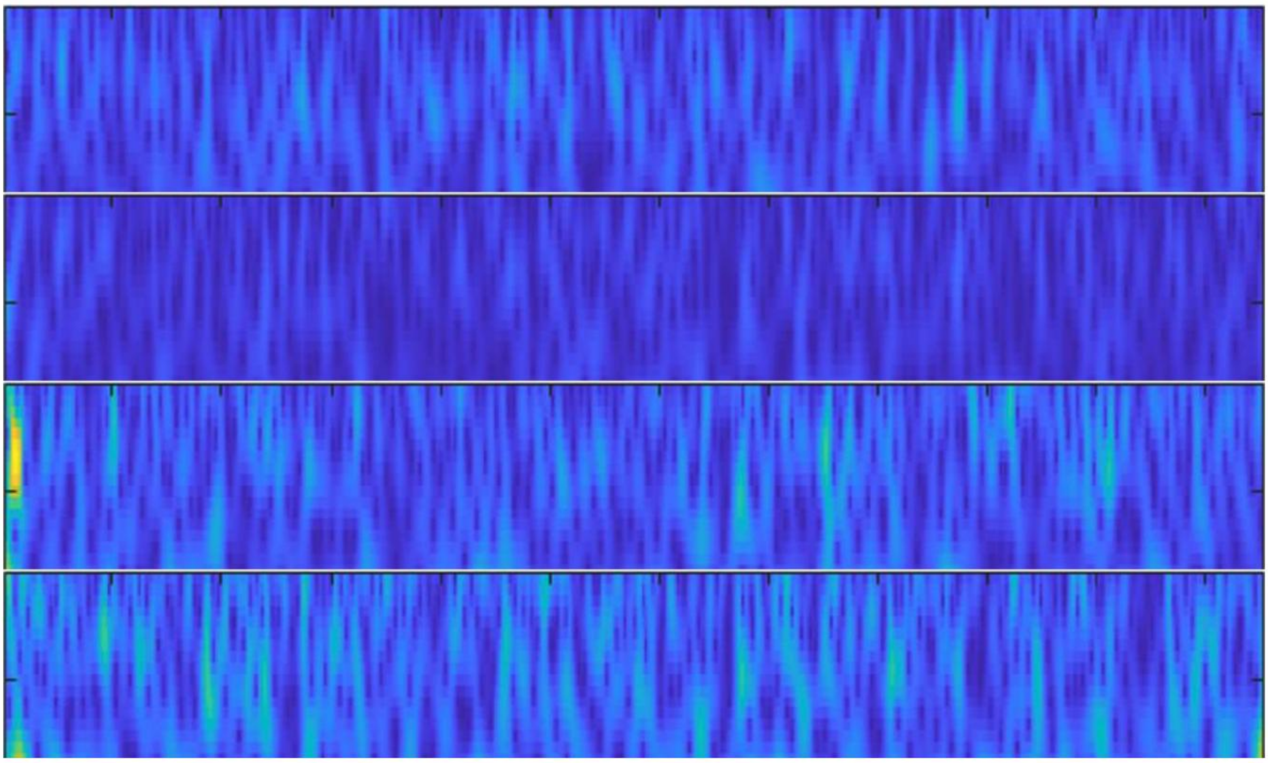

Figure 21. Assembly of the 4 spectrograms for a trial of the digit '0' (MindBigData)

Machine Translated by Google

4. RESULTS AND DISCUSSION

L The following section presents the results obtained within the framework of this TFG. The data are shown collected during the investigation, which have been analyzed to provide information relevant to how to address the stated objectives.

4.1. Table of results obtained

Through this results table, we seek to offer a clear and concise view of the information obtained with the research carried out in this study, showing the combinations between architectures and approaches explained in 3.3. Implementation of neural networks.

	Learning Simple		**LSTM**		**TF CNN**	**GoogleNet**		**AlexNet**	
	EEG	Frequency bands	EEG	Frequency bands	EEG 1 canal	Spectrograms		Spectrograms	
						1 canal	Various channels	1 canal	Various channels
Signs Own	29.0%	**35.5%**	~25%	27.3%	**36.4%**	-	-	-	-
Thinking Out Loud [3]	~25%	~25%	~25%	~25%	~25%	~25%	**29.1%**	~25%	25.3%
MindBigData[7]	~50%	~50%	~50%	**85.8%**[8]	~50%	~50%	~50%	~50%	~50%

Table 5. Learning results using different architectures

7 Binary classification. The frequency bands are from the TP9 electrode.

8, 7 Strange behavior of the neural network is noted, review the interpretation of the data obtained.

4.2. Interpretation of the data obtained

The levels of precision achieved generally barely exceed chance probability. Therefore, the aim is to carry out an analysis of limitations and possible sources of error.

The first thing that stands out is a higher 'Accuracy' for the own signals both with the 'Simple Learning' network and for 'TF CNN' with respect to all the classifications made with the 'Thinking Out Loud' dataset.
This could be due to the fact that the positions of the electrodes placed in the experiment itself are of special interest, while the respective neural networks are not able to extract relevant features from the second dataset, even having this better spatial resolution (including the area studied in the first dataset). The number of trials belonging to the validation set would also have an influence, since there is a figure
relatively low number of recordings in the anechoic room. This makes it more likely that the number of correct answers per number of trials in the validation set will shoot up and remain above the probability of correct answers.
randomly.

On the other hand, we observed a similar behaviour between the classification of the own signals in 'Simple Learning' when the data to be introduced are the frequency bands plus the EEG signal and 'TF CNN', which calculates the energy level of the different frequencies within the network. Both architectures raise the level of precision by around 10% compared to the other approaches studied. It can be stated that the frequency bands as a whole (ÿ, ÿ, ÿ, ÿ, total energy) have a relevant importance in the recognition of words using EEG.

It has also been observed that the MindBigData dataset has difficulties in achieving high levels of accuracy in the classification tasks performed. Specifically, it has been identified that when a high percentage of Accuracy is achieved in the validation set, unusual behavior occurs in the neural network, since the percentage of Accuracy in the training set is lower than that of the validation set.
This finding raises questions about the performance and generalizability of the model on the MindBigData dataset, suggesting potential additional challenges in accurately classifying the data. These results highlight the importance of further understanding the characteristics and particularities of the dataset used, as well as the need to investigate additional strategies to improve the quality of predictions and address the unexpected behavior observed in the results.

Finally, when comparing the methods proposed here with those described in the scientific literature, it is observed that the levels of precision obtained are below those previously reported. It is particularly relevant to mention that, despite having performed a binary classification with MindBigData in this study and a strange behavior in learning, the results still present lower levels of precision than those reported in the literature for more complex pattern recognition tasks, i.e., the distinction of 10 digits.

	Proposed method	**Literature**
Own signs	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%**

Table 6. Comparison of results: proposed methods and literature

Taking into account all of the above, it is important to acknowledge a limitation in word recognition in this study. This is believed to be inherent to the EEG technique, in relation to the spatial resolution of the acquired signals, which implies that the precision to accurately locate the active neural sources and to understand in detail the patterns of brain activity is compromised. This can be attributed to several factors.

- The way in which signals are recorded on the surface of the scalp using electrodes limits the ability to obtain an accurate image of brain activity in terms of spatial location, since, in one electrode, the integration of the activities of hundreds of millions of neurons occurs, each of which has its own purpose and significance in the complex functioning of the brain. In addition, the conductivity of the tissues and skull introduces distortions and attenuations into the signals, which can affect the precise localization of the underlying neural sources.

- It is vitally important to consider the complexity of the activities that occur simultaneously in the brain. This organ is constantly involved in a multitude of cognitive, emotional and perceptual processes, which generates a highly dynamic and complex environment and a 'cognitive interference' that is reflected in the recorded signals and makes it difficult to accurately interpret the brain activity that is sought to be analyzed. In addition, this cognitive interference can also be present at the level of neuronal activity itself. Different brain regions are interconnected and in constant communication, which means that activity in one region can influence activity in other regions. This interaction between regions can generate complex and difficult-to-unravel patterns, especially when seeking to identify the contribution of a specific region to a given cognitive phenomenon.

5. FUTURE LINES

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

Unknown author9

A Yes, all the data shown in section '4. Results and Discussion' suggest that further research is needed. refinement and improvement in the approaches used for word recognition in the different datasets. Therefore, several lines of research are proposed for future investigations.

5.1. Improvement in EEG treatment

These results indicate the importance of exploring and fine-tuning the models, preprocessing techniques, feature extraction, and architectures used to obtain significantly better than chance performance in this task. More sophisticated approaches, such as combining different techniques or using more complex models, may be required to improve the accuracy of word classification in these specific EEG signal contexts. Examples of this are [28] and [7], where neural networks are designed specifically for this context. In the first, different LSTM, GRU, or RNN cells are created to form 'bidirectional recurrent layers', which in turn will constitute a bidirectional LSTM network with a high level of accuracy according to its writers, while in the second, different CNNs are created to discriminate between imagined speech and other thoughts, capture the subject's moment of attention, or separate brain regions to distinguish the grammatical class of the imagined word. These authors highlight in their articles that they achieved accuracies of 96.18% in the classification of 10 digits and 76.6% in the classification of 3 pairs of words, respectively.

5.2. Alternative methods to EEG

In view of the limitation of lack of spatial resolution, it is pertinent to explore and consider the use of techniques other than EEG in order to obtain better spatial resolution in future research.
There are various approaches that could complement or replace the exclusive use of EEG, such as functional magnetic resonance imaging (fMRI), the exploration of which has already begun [42], magnetoencephalography or transcranial magnetic stimulation, among others.

These techniques offer greater spatial precision when mapping brain activity, facilitating a more accurate identification of brain regions involved in cognitive processes and understanding the spatial organization of brain activity. Furthermore, the combination of multiple neuroimaging techniques can provide a more complete and holistic view of brain activity, thus complementing the results obtained through EEG.

However, it is important to note that each technique has its own advantages and limitations, and the choice of the most appropriate technique will depend on the research objectives, resource constraints and

9 No reliable source has been found that affirms the origin of this quote, although it has been attributed on numerous occasions to Thomas Alva Edison, 1847-1931.

the specific characteristics of the study in question. A thorough analysis and careful evaluation of the different options available is recommended before selecting the most appropriate method or combination of methods for each investigation.

In conclusion, it is suggested that alternative methods to EEG be explored and considered in order to overcome the limitation in spatial resolution presented by this technique. The incorporation of complementary neuroimaging techniques will allow obtaining a more precise and detailed view of brain activity, thus enriching the understanding of the cognitive and neurophysiological phenomena addressed in the study.

REFERENCES

[1] B. D. P. P. a. M. A. Lopez-Bernal D, «A State-of-the-Art Review of EEG-Based Imagined Speech nº 16:867281, 2022. Decoding,» *Front. Hum. Neurosci.*

[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),» [Online]. https:// Available: en.wikipedia.org/wiki/Electroencephalography. [Last accessed: 30 5 2023].

[5] LA y. F. Cincotti, «Apuntes Neuroengineering,» *Sapienza University of Rome,* 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 Institue on Deafness and Other Communication Disorders,» [En línea]. Available: https://www.nidcd.nih.gov/es/espanol/afasia. [Last accessed: 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,» [Online]. 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. DPCP-VEI a. JMA Belén López, «APPLICATION OF ICA FOR THE ELIMINATION OF OCULAR AND MANDIBULAR ARTIFACTS IN EEG SIGNALS DURING THE PERFORMANCE OF MENTAL TASKS,» *Miguel Hernández University,* 2014.

[23] «Wavelet transforms. Mathworks,» [En línea]. Available: https://es.mathworks.com/discovery/wavelet-transforms.html. [Last accessed: 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». wavelet Wikipedia,» [In Available: https://en.wikipedia.org/wiki/Daubechies_wavelet. [Last accessed: 1 6 2023].

[27] «Introduction to the wavelet families. Mathworks,» [Online]. Available: https://es.mathworks.com/help/wavelet/gs/introduction-to-the-wavelet-families.html. [Last accessed: 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] KMWA a. GIGA Pratama IWP, «Frequency band and PCA feature comparison for EEG signal classification,» *Lontar Komputer: Scientific Journal of Information Technology,* 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. Latvian,* 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] «Boundatry effects and the cone of influence. Mathworks,» [En línea]. Available: [Último https://es.mathworks.com/help/wavelet/ boundary-effects-and-the-cone-of-influence.html. acceso: 5 6 2023]. ug/

[40] «Web Muse,» [Online]. 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 Recognition. Speech Mathworks, » online]. [https:// Available: 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] Author, «This is an example of a quote,» *Doctoral Thesis,* vol. 2, no. 13, 2012.

[47] O. Author, «Another different quote,» *magazine,* p. 12, 2001.

[48] «Coarticulation. Wikipedia,» [Online]. https:// Available: es.wiktionary.org/wiki/coarticulaci%C3%B3n. [Last accessed: 1 6 2023].