



**I  
N  
A  
O  
E**

# Clasificación de palabras no pronunciadas presentes en Electroencefalogramas (EEG)

Por:

Alejandro Antonio Torres García

Supervisada por:

Dr. Carlos Alberto Reyes García

Dr. Luis Villaseñor Pineda

Versión final 27 de octubre de 2011

© Coordinación de Ciencias Computacionales

INAOE

Luis Enrique Erro 1  
Sta. Ma. Tonantzintla,  
72840, Puebla, México.





## Agradecimientos

*Al ETERNO por todo.*

*A mis asesores los Doctores Carlos Alberto Reyes García y Luis Villaseñor Pineda por guiarme en el inmenso mundo de la ciencia, y ayudarme a ser menos tecnólogo y alquimista.*

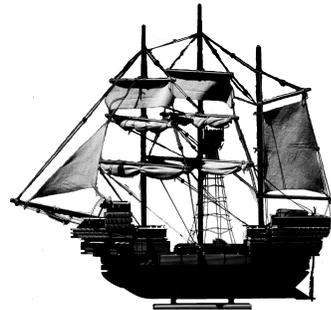
*A quienes creyeron en mí, y me brindaron su invaluable apoyo al inicio de esta aventura: M.C. Walter Torres Robledo, Dr. Jorge Camas Anzueto, M. Francisco Vázquez, Ing. Héctor González Ordoñez, Ing. Daniel Octavio Estrada Paredes, C.P. Ana Utrilla, Lic. Noé Villanueva, Lic. Elisa Jiménez, Ing. Santos Hernández Promotor, Mtra. Marvila Komukai Puga, Profra. Esther Ballinas Culebro y Profra. María de los Ángeles Durán Méndez.*

*A quienes le recordaron a este chiapaneco que aún es útil para muchos pero sobre todo para sí mismo dentro de su propio reto: Norma Tecampo Pérez, Laura Tozqui, Julio Hernández, Lupita Méndez, Alejandro Rosales, Carlos Chang, David Prieto, Viviana Hernández, Maribel Marín Castro, Pedro Tecuanhuehue Vera, Mónica Duarte, Nidia Benitez Palma, Arely Vargas Ortega, Omar Carreño Sánchez, Diana Ahuatzi Reyes, Lupita Martínez Hernández, Gabriel Ramírez y familia, Jessica Cuautle Zacatzi, Carlos Pérez Lara, Iván López Coello, y Ángel Martínez Mijangos.*

*Por su valiosa asesoría a los Maestros en Ciencias Gerardo Rosas y Jorge Morales Cruz, y al Dr. Juan Manuel Ramírez.*

*Por sus atenciones a las Sras. Rosa Tecuapetla y Elena Robles Neri.*

*Al pueblo de México que, mediante el Consejo Nacional de Ciencia y Tecnología, me apoyó con una beca para realizar los estudios de maestría.*





# Dedicatoria

*A mi familia, y de manera muy especial a mi madre (Sra. María Dolores García Albores), a mis hermanitas (Lupita y Margarita), a mi abuelito (Sr. Abelardo García Jiménez), a mis tíos (Sra. Amparo García Albores, Sr. Armando García Santiago, Sra. Bernarda García Albores, y Sr. José García Herrera), y a mi primo (Juan Carlos López García).*

*Al tamal de chipilín, las cajetas, al palmito, a la música de marimba, y principalmente al café de Chiapas remedio infalible para este gélido lugar.*

*A quienes durante la estancia en Tonantzintla me recordaron que el café se bebe a sorbos para la hermandad.*





# Índice general

<b>Resumen</b>	<b>xvii</b>
<b>Abstract</b>	<b>xix</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Problemática . . . . .	3
1.2. Alcances y limitaciones . . . . .	4
1.3. Objetivos . . . . .	5
1.3.1. Objetivo general . . . . .	5
1.3.2. Objetivos específicos . . . . .	5
1.4. Contenido del documento . . . . .	6
<b>2. Fundamento Teórico</b>	<b>7</b>
2.1. El cerebro . . . . .	7
2.1.1. Funciones de las regiones cerebrales . . . . .	8
2.2. Electroencefalograma (EEG) . . . . .	9
2.3. Interfaces cerebro-computadora (BCI) . . . . .	10
2.4. Procesamiento Digital de Señales . . . . .	15
2.4.1. Transformada Wavelet . . . . .	15
2.5. Clasificación . . . . .	17
2.5.1. Máquinas de vectores de soporte . . . . .	18
2.5.2. Naive Bayes . . . . .	20

2.5.3. Random Forests . . . . .	21
2.5.4. Bagging . . . . .	22
2.6. Medidas para evaluar a los clasificadores . . . . .	23
2.7. Validación cruzada con $K$ -pliegues . . . . .	27
<b>3. Estado del Arte</b>	<b>29</b>
3.1. Reconocimiento de habla no pronunciada mediante EEG . . . . .	30
3.1.1. Enfoque sílabas . . . . .	31
3.1.2. Enfoque palabras . . . . .	34
3.2. Discusión . . . . .	37
<b>4. Método propuesto</b>	<b>41</b>
4.1. Materiales y software . . . . .	42
4.2. Adquisición de las señales cerebrales . . . . .	43
4.3. Pre-procesamiento . . . . .	46
4.4. Extracción de características . . . . .	48
4.5. Selección de características . . . . .	50
4.6. Clasificación . . . . .	51
<b>5. Experimentación y resultados</b>	<b>53</b>
5.1. Experimentos con tres individuos . . . . .	53
5.2. Experimentos y resultados con veintiún individuos . . . . .	56
5.3. Análisis de resultados a nivel palabra . . . . .	59
5.4. Discusión . . . . .	64
<b>6. Conclusiones y Trabajo Futuro</b>	<b>67</b>
6.1. Conclusiones . . . . .	68
6.2. Trabajo futuro . . . . .	69
<b>Referencias</b>	<b>75</b>

**A. Corpus de datos**

**77**

**B. Resultados con el corpus completo**

**79**



# Índice de figuras

1.1. BCI basada en EEG de superficie [Millán, 2002]. . . . .	2
2.1. Partes del cerebro [Sagan, 1977] . . . . .	8
2.2. Modelo Geschwind-Wernicke [Wester, 2006] . . . . .	9
2.3. Relación interelectrodos de acuerdo al Sistema Internacional 10-20	11
2.4. Potenciales corticales lentos [Wolpaw <i>et al.</i> , 2002] . . . . .	12
2.5. P300 [Wolpaw <i>et al.</i> , 2002] . . . . .	13
2.6. Ritmos sensoriales motrices [Wolpaw <i>et al.</i> , 2002] . . . . .	14
2.7. Potenciales visuales evocados [Odom <i>et al.</i> , 2004] . . . . .	14
2.8. Búsqueda del hiperplano óptimo en SVM . . . . .	19
3.1. Sincronización para un ejemplo del experimento descrito en [DaSal- la <i>et al.</i> , 2009] . . . . .	32
4.1. Modelo general . . . . .	42
4.2. Metodología seguida en el trabajo . . . . .	43
4.3. Ubicación de los sensores en el Kit EMOTIV . . . . .	44
4.4. Espacio acondicionado para la adquisición de señales EEG. . . . .	45
4.5. Señal EEG del canal F7 del individuo S1 mientras imagina la dic- ción de la palabra “abajo” siguiendo el protocolo de adquisición de datos . . . . .	47
4.6. Ejemplo de señales grabadas con marcadores erróneos . . . . .	48
5.1. Porcentajes de exactitud . . . . .	57

5.2. Gráfica del coeficiente kappa . . . . .	58
5.3. Gráfica del recuerdo promedio por palabra . . . . .	60
5.4. Gráfica de la especificidad promedio por palabra . . . . .	61
5.5. Precisión por palabra de cada uno de los clasificadores . . . . .	62
5.6. Gráfica de la F-Measure promedio por palabra de cada uno de los clasificadores . . . . .	63

# Índice de tablas

2.1. Matriz de confusión para la clase A . . . . .	24
2.2. Valoración del coeficiente kappa [Landis and Koch, 1977] . . . . .	26
3.1. Estado del Arte en el reconocimiento de <i>habla no pronunciada</i> . . . . .	38
4.1. Descripción de los niveles de descomposición de la DWT . . . . .	50
5.1. Porcentajes de exactitud obtenidos por los clasificadores utilizando vectores de características completos (1076 características) . . . . .	54
5.2. Porcentajes de exactitud obtenidos por los clasificadores utilizando vectores de características reducidos (540 características) . . . . .	55
A.1. Distribución del número de ejemplos de cada una de las palabras . . . . .	78
B.1. Porcentajes de exactitud obtenidos por los clasificadores para cada individuo después de aplicar validación cruzada con diez pliegues . . . . .	80
B.2. Coeficientes kappa obtenidos por los clasificadores para cada indi- viduo después de aplicar validación cruzada con diez pliegues . . . . .	81
B.3. Recuerdo para cada palabra de los individuos obtenida por el clasi- ficador Random Forest después de aplicar validación cruzada con diez pliegues . . . . .	82
B.4. Recuerdo para cada palabra de los individuos obtenida por el clasi- ficador Bagging-RF después de aplicar validación cruzada con diez pliegues . . . . .	83

B.5. Especificidad para cada palabra de los individuos obtenida por el clasificador Random Forest después de aplicar validación cruzada con diez pliegues . . . . .	84
B.6. Especificidad para cada palabra de los individuos obtenida por el clasificador Bagging-RF después de aplicar validación cruzada con diez pliegues . . . . .	85
B.7. Precisión para cada palabra de los individuos obtenida por el clasificador Random Forest después de aplicar validación cruzada con diez pliegues . . . . .	86
B.8. Precisión para cada palabra de los individuos obtenida por el clasificador Bagging-RF después de aplicar validación cruzada con diez pliegues . . . . .	87
B.9. F-Measure para cada palabra de los individuos obtenida por el clasificador Random Forest después de aplicar validación cruzada con diez pliegues . . . . .	88
B.10. F-Measure para cada palabra de los individuos obtenida por el clasificador Bagging-RF después de aplicar validación cruzada con diez pliegues . . . . .	89

# Índice de algoritmos

1.	Random Forest . . . . .	21
2.	Entrenamiento del Bagging . . . . .	22
3.	Clasificación del Bagging . . . . .	23



# Resumen

La presente investigación tiene como objetivo desarrollar un método de procesamiento y clasificación de señales electroencefalográficas (EEG) de un individuo para reconocer palabras no pronunciadas de un vocabulario reducido. Específicamente, un reconocimiento preciso permitirá controlar mentalmente el cursor de una computadora, ya que, con ese fin, el vocabulario consta de las palabras del lenguaje español: “arriba”, “abajo”, “izquierda”, “derecha”, y “seleccionar”.

Para iniciar el proceso, las señales EEG se adquieren utilizando un protocolo básico basado en marcadores, para saber a priori en qué parte de la señal la persona imagina la dicción (habla no pronunciada) de la palabra indicada, con los marcadores se forman ventanas. Para cumplir el objetivo, únicamente son de interés las señales de los canales F7, FC5, T7 y P7 (los más cercanos a las áreas cerebrales de Broca y Wernicke). Posteriormente, éstas son filtradas utilizando un FIR pasa-bandas en el rango de 4-25 Hz, y las ventanas son ajustadas a 256 muestras. Además, se extraen características mediante la transformada wavelet discreta que son utilizadas para entrenar y evaluar a cuatro clasificadores: Naive Bayes (NB), Random Forests (RF), Máquina de vectores de soporte (SVM), y Bagging de Random Forest (Bagging-RF). Las medidas de evaluación son la exactitud y el coeficiente kappa obtenidas mediante validación cruzada con 10 pliegues.

Se grabaron señales EEG de 21 individuos, para cada uno de los cuales se tienen alrededor de 30 ventanas correspondientes a cada una de las palabras. Para el caso del mejor clasificador Bagging-RF, los valores promedio tanto del porcentaje de

exactitud como para el coeficiente kappa son 40.48 % y 0.24, respectivamente. Además, se presenta un análisis de Bagging-RF y RF a nivel de la clasificación de las palabras no pronunciadas, donde se evalúa a los clasificadores con el recuerdo, la especificidad, la precisión y la medida-F.

Los resultados aún distan de lo necesario para controlar eficazmente a una BCI sin embargo, debido a que los porcentajes de exactitud están arriba del 20 %, es decir, arriba del azar para cinco clases, y los valores de kappa mayores a 0, se puede concluir que las señales EEG efectivamente contienen información que hace posible la clasificación de las palabras no pronunciadas del vocabulario reducido mediante el método propuesto. Con lo que, el habla no pronunciada resulta ser una fuente electrofisiológica prometedora para controlar BCIs de forma más intuitiva.

# Abstract

The present work aims to the development of a method for the processing and classification of electroencephalographic signals (EEG), belonging of a single subject, to recognize unspoken words belonging to a reduced vocabulary. Specifically, a precise recognition would allow to mentally control a computer screen cursor, thus, with this purpose, the vocabulary is composed by the spanish language words: “arriba”, “abajo”, “izquierda”, “derecha”, and “seleccionar”.

To start the process, the EEG signals were adquired using a protocol based on markers, to know a priori in what part of the EEG signal a subject imagines the pronunciation of an indicated word. With the markers were formed windows, between which the samples could be found. To accomplish the objective, only the signals from channels F7, FC5, T7 y P7 (which are the nearest to the Wernicke and Broca areas) were taken in count. After that, the signals coming from those channels were filtered using a band-pass FIR filter at 4-25 Hz., and the windows were adjusted to 256 samples. Furthermore, features were obtained using discrete wavelet transform (DWT) to train and assess four classifiers: Naive Bayes (NB), Random Forests (RF), Support Vector Machines (SVM) and Bagging-RF. The evaluation measures are accuracy and kappa index obtained using 10-folds cross validation.

EEG signals belonging to 21 subjects were recorded, for each subject around 30 windows for each word were obtained. In the case of the best classifier, Bagging-RF, the accuracy rate and kappa index average were 40.48 % and 0.24, respective-

ly. Furthermore, an analysis at the unspoken words level classification for Bagging-RF and Random Forests was performed, where the classifiers were assessed with recall, specificity, precision, and f-measure.

The results are still far of the precision needed to effectively control a BCI, but they are encouraging because accuracy rates are above 20 %, i.e. above chance for five classes, and kappa values' are greater than 0, which are values comparable to state of the art research. Based on them, it can be concluded that EEG signals actually contain information to allow the classification of the unspoken words belonging to a reduced vocabulary by the proposed method. That is why, unspoken speech turns out to be a promising and intuitive electrophysiologic source to control BCIs.

# Capítulo 1

## Introducción

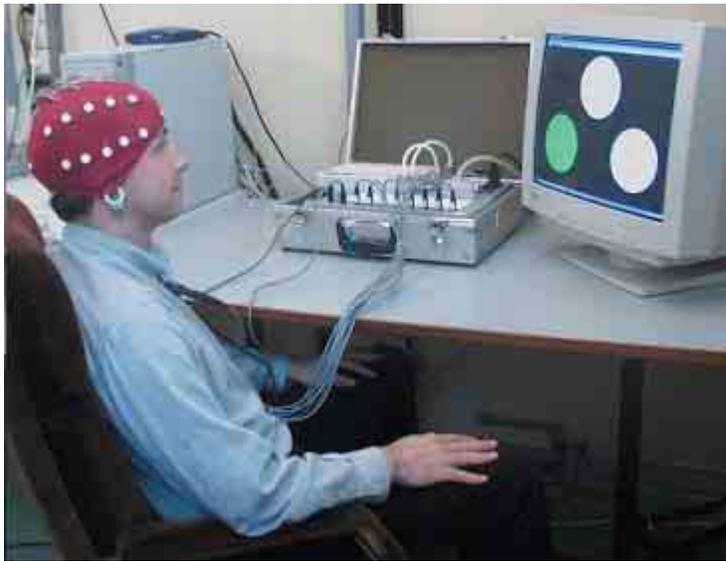
En la República Mexicana, las discapacidades que se presentan con mayor número de ocurrencias son las de tipo motriz con un 58.3%<sup>1</sup>. Dentro de este sector se encuentran discapacidades motrices severas como: la esclerosis lateral amiotrófica (ELA), la embolia (ictus cerebral), las lesiones de médula espinal o cerebral, la parálisis cerebral, las distrofias musculares, la esclerosis múltiple, entre otros padecimientos. Estas discapacidades frecuentemente provocan que la persona no pueda controlar voluntariamente sus movimientos, incluyendo la respiración, el habla y los movimientos de los articuladores del habla [Brumberg *et al.*, 2010; Wolpaw *et al.*, 2002]. En consecuencia, una persona en estas condiciones está prácticamente aislada de su entorno.

Con los avances tecnológicos, recientemente, se han propuesto diversas alternativas que permitan integrar a este sector de la sociedad. Las Interfaces Cerebro-Computadora (BCI, por sus siglas en inglés) son una de estas alternativas ya que intentan proveer al cerebro de un nuevo canal de comunicación y control para transmitir mensajes y comandos al mundo exterior [Brumberg *et al.*, 2010; Wolpaw *et al.*, 2002]. Para adquirir datos acerca de la actividad cerebral, existen varias alternativas. Sin embargo, el electroencefalograma (EEG) es el más uti-

---

<sup>1</sup> XIII Censo General de Población y Vivienda 2010 realizado por el Instituto Nacional de Estadística, Geografía e Informática

lizado debido a que no es invasivo, y requiere un equipo relativamente sencillo y económico (ver figura 1.1) [Wolpaw *et al.*, 2002]. Aunque capta ruido de otras actividades fisiológicas como, por ejemplo, la actividad cardíaca, el movimiento ocular, el movimiento de la lengua, la respiración, los potenciales de la piel entre otras. Además, también capta ruido proveniente de la corriente alterna y de los electrodos [Sánchez de la Rosa, 1993].



**Figura 1.1:** BCI basada en EEG de superficie [Millán, 2002].

En las BCIs basadas en EEG, a los mecanismos neurológicos o procesos empleados por el usuario de la BCI para generar las señales de control, se les denominan fuentes electrofisiológicas. Las más ampliamente utilizadas son: los potenciales corticales lentos (SCP, por sus siglas en inglés), los potenciales P300, los ritmos sensoriales motrices ( $\mu$  y  $\beta$ ) y los potenciales evocados visuales (VEP, por sus siglas en inglés) [Bashashati *et al.*, 2007; Brumberg *et al.*, 2010; Wolpaw *et al.*, 2002]. Con estas fuentes electrofisiológicas, la comunicación se realiza mediante dos paradigmas de control: discreto o continuo. En el paradigma discreto, el usuario puede elegir entre dos o más opciones discretas, por ejemplo elegir una tecla específica de un teclado virtual en el monitor de la computadora. En el paradigma continuo, un pequeño número de variables cinemáticas (por ejemplo,

las coordenadas  $x$  y  $y$  de la posición del cursor en el monitor, o los valores de las primeras dos frecuencias formantes para una prótesis de habla) son controladas por el usuario [Brumberg *et al.*, 2010].

## 1.1. Problemática

Las BCIs descritas previamente, presentan los siguientes dos grandes problemas. El primer problema es el largo periodo de entrenamiento (algunas semanas hasta meses) requerido para que un usuario pueda utilizar la BCI. Lo anterior se debe a que, las fuentes descritas anteriormente (SCP, P300, ritmos  $\mu$  y beta, y VEP) son generadas por el usuario de forma poco consciente [Pfurtscheller, 2004]. Mientras que, el segundo problema son las bajas tasas de comunicación (una sólo palabra procesada, o menos, por minuto) que resultan insuficientes para permitir una interacción natural o casi natural. Este último problema se debe a que cada una de las fuentes electrofisiológicas usadas por las BCIs requieren un “mapeo” o traducción al dominio del habla [Brumberg *et al.*, 2010].

Los problemas descritos anteriormente han motivado una serie de trabajos que tratan de utilizar los potenciales relacionados con la producción del habla, con diversos grados de éxito [Brumberg *et al.*, 2010]. En estos trabajos, la fuente electrofisiológica es el *habla no pronunciada* (*unspoken speech*), también referida como *habla interna* o *habla imaginada*, donde el término *habla no pronunciada* se refiere a la acción de imaginar la dicción de una palabra pero sin emitir ni articular sonidos. Es importante mencionar que, Denby [Denby *et al.*, 2010] incluye a estos trabajos dentro de un área de investigación denominada interfaces de habla silenciosa (SSI, por *Silent Speech Interfaces*) cuya finalidad es desarrollar sistemas capaces de permitir la comunicación “hablada” que toman lugar cuando emitir una señal acústica audible es imposible (por ejemplo, en el caso de las personas con los padecimientos descritos previamente). Los trabajos que utilizan *habla no pronunciada* pueden dividirse en dos enfoques: palabras y sílabas. En

[Calliess, 2006; Porbadnigk, 2008; Suppes *et al.*, 1997; Wand, 2007; Wester, 2006] se explora con palabras. Mientras que, en [Brigham and Kumar, 2010; DaSalla *et al.*, 2009; D’Zmura *et al.*, 2009] únicamente se tratan sílabas.

En el caso específico de los trabajos que exploran palabras, tal y como se describe en la sección 3.2, se requiere de un modelo que permita tratar al reconocimiento de palabras no pronunciadas de manera adecuada.

La presente investigación tiene como objetivo poder interpretar las señales EEG asociadas a las palabras no pronunciadas del lenguaje español “arriba”, “abajo”, “izquierda”, “derecha”, y “seleccionar”. El problema es tratado bajo el enfoque de clasificación, y se conoce *a priori* en qué parte de la señal EEG la persona imagina la pronunciación de las palabras indicadas.

## 1.2. Alcances y limitaciones

En la presente propuesta se propone utilizar un vocabulario reducido de palabras a reconocer en la señal EEG. El vocabulario a utilizar consta de las siguientes palabras que están dirigidas a posteriormente lograr el control de un dispositivo electro-mecánico.

- arriba
- abajo
- izquierda
- derecha
- seleccionar

Por otra parte, es importante recalcar que el uso del electroencefalograma para monitorear la actividad cerebral se debe a su bajo costo, y su relativa sencillez con respecto a otros medios de adquisición de dicha actividad. Sin embargo, la

naturaleza no estacionaria del EEG (ver sección 2.2) hace necesario que el enfoque de clasificación sea de manera individual. Además, por su utilidad personal, cada BCI debe ajustarse a un usuario en particular, lo que hace importante el proceso de entrenamiento y prueba con muestras de un sólo individuo. No obstante, en este trabajo se experimenta con muestras de múltiples individuos con el fin de obtener mayor información sobre el potencial de la clasificación independiente del sujeto.

De igual manera, también se propone usar únicamente los canales EEG más relacionados con el habla. Específicamente, los canales EEG más cercanos al modelo Geschwind-Wernicke (ver sección 2.1.1). Asimismo, para cada uno de estos canales, se exploran las frecuencias del EEG más relacionadas con el habla no pronunciada (ver sección 4.5). Ambas acciones con las ideas de: utilizar los potenciales más relacionados con la producción del lenguaje, y disminuir el impacto del problema conocido como *maldición de la dimensionalidad* (descrito al final de la sección 2.5), que se presenta en clasificadores para BCIs.

## 1.3. Objetivos

### 1.3.1. Objetivo general

Desarrollar e implementar un método de procesamiento y clasificación para interpretar palabras no pronunciadas en señales EEG, que obtenga resultados comparables a los descritos en el estado del arte, utilizando un menor número de canales, y características orientadas al habla no pronunciada.

### 1.3.2. Objetivos específicos

- Caracterizar las señales EEG asociadas a palabras no pronunciadas dentro de un vocabulario reducido.
- Seleccionar los canales del EEG adecuados para señales lingüísticas.

- Definir el proceso de extracción de características de la señal EEG.
- Adecuar la dimensionalidad del vector (o vectores) de características de la señal EEG.
- Implementar un modelo de clasificación de señales EEG.

## **1.4. Contenido del documento**

El resto del documento se compone de 5 capítulos. A continuación se describe brevemente cada uno de ellos. En el capítulo 2 “Fundamento Teórico” se describen los principales conceptos que dan soporte a la tesis. En el capítulo 3 “Estado del Arte” se describe una serie de trabajos que están relacionados con la actual propuesta. En el capítulo 4 “Método propuesto” se describen los componentes principales de la solución propuesta para tratar la problemática. En el capítulo 5 “Experimentos y resultados” se presentan los experimentos más relevantes llevados a cabo durante el trabajo así como sus resultados. Por último, en el capítulo 6 “Conclusiones y trabajo futuro” se presentan las conclusiones del trabajo, y se da una perspectiva acerca de las posibilidades que se pueden explorar posteriormente para tratar de refinar la solución propuesta.

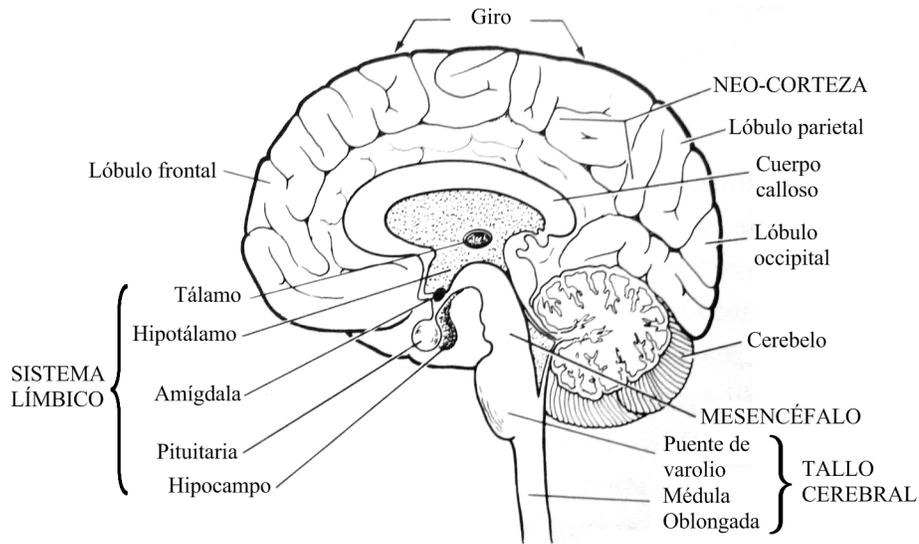
# Capítulo 2

## Fundamento Teórico

Este capítulo da una descripción de los conceptos más relevantes que son tratados en la tesis, y que le dan soporte teórico a los trabajos desarrollados.

### 2.1. El cerebro

El cerebro es un órgano cuyas partes principales son: neo-corteza, el cerebelo, el tallo cerebral (incluyendo el mesencéfalo, el puente de Varolio y la formación reticular), y el tálamo (entre el mesencéfalo y los hemisferios). El cerebro está dividido en dos hemisferios, separados por una fisura longitudinal a través de la cual hay una gran banda conectiva de fibras conocida como el cuerpo calloso. La superficie externa de los hemisferios cerebrales, es conocida como la corteza cerebral, se compone de neuronas (materia gris) dentro de patrones circunvolucionados, y dividido en regiones por las fisuras (surcos). Bajo la corteza se encuentran las fibras nerviosas que se dirigen a otras partes del cerebro y del cuerpo (materia blanca) [Rangayyan, 2002]. En el caso específico de la materia gris, de acuerdo con [Uruchurtu, 2010], el cerebro está constituido por una intrincada red de unos 100 000 millones de neuronas —células nerviosas— en constante comunicación. La figura 2.1 ilustra mejor lo anterior.

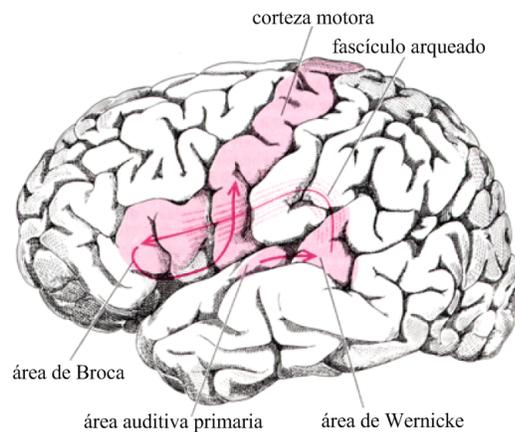


**Figura 2.1:** Partes del cerebro [Sagan, 1977]

### 2.1.1. Funciones de las regiones cerebrales

El hemisferio izquierdo del cerebro es el centro de la actividad de las tareas que impliquen lenguaje, números y lógica. Mientras que el hemisferio derecho es más activo durante las relaciones espaciales y la imaginación de movimientos [Vallabhaneni *et al.*, 2005].

En el caso específico del lenguaje, el modelo Geschwind-Wernicke trata de explicar el funcionamiento del lenguaje en el cerebro (ver figura 2.2) [Geschwind, 1972; Restak, 1979]. El área de Broca es fundamental para la producción de frases armadas sintácticamente, el área de Wernicke es fundamental para la producción de un discurso con significado (a la vez que para la comprensión) y el fascículo arqueado (o según la teoría de Luria, el lóbulo parietal) es necesario para encadenar los fonemas que componen las palabras [Obler and Gjerlow, 2001].



**Figura 2.2:** Modelo Geschwind-Wernicke [Wester, 2006]

## 2.2. Electroencefalograma (EEG)

De acuerdo con [Gutiérrez, 2001; Muñoz, 2001], la señal del EEG es resultado de la interacción de un sinnúmero de procesos entre cientos de millones de neuronas organizadas en determinados grupos neuronales. Además, en [Lotte *et al.*, 2007] se menciona que las señales EEG están clasificadas como no estacionarias, es decir, que son señales que presentan una frecuencia variable en el tiempo o su contenido de frecuencia es variable en el tiempo [Boashash, 2003]. También, se conoce que las amplitudes de la señal del EEG están normalmente en un rango de 30 a 100  $\mu V$  y, la actividad rítmica del cerebro en un sujeto sano está caracterizada principalmente por los siguientes tipos de onda:

- Onda Delta ( $\delta$ ), rango de frecuencia de 0 a 4 Hz. Estas ondas se presentan durante la fase III del sueño profundo en el adulto y en el lactante en estado de vigilia. Cuando se presentan en el adulto despierto indican lesión cerebral.
- Onda Thetha ( $\theta$ ), rango de frecuencia de 4 a 8 Hz. Estas ondas suelen ocurrir en niños y en adultos en la fase I y II del sueño fisiológico y, cuando se experimenta tensión emocional y fatiga.
- Onda Alfa ( $\alpha$ ), rango de frecuencia de 8 a 12 Hz. Se observa en los sujetos

sanos si están despiertos y, en reposo se bloquea con la apertura ocular y la concentración.

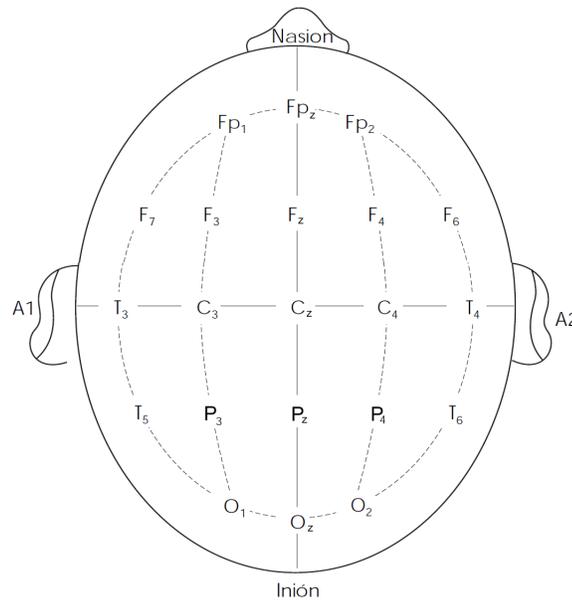
- Onda Beta ( $\beta$ ), las frecuencias de este tipo de onda varían de 12 a 60 Hz. Sin embargo, algunos autores como en el caso de [Vallabhaneni *et al.*, 2005] clasifican a las señales que exceden 30 Hz. como ondas gamma ( $\gamma$ ). Estas ondas se presentan con la actividad intensa del sistema nervioso, es decir, durante los periodos de actividad sensorial y mental.

El Electroencefalograma de superficie consiste en colocar electrodos sobre la superficie del cuero cabelludo según las posiciones y mediciones establecidas de acuerdo al Sistema Internacional 10-20, como se muestra en la figura 2.3 [Gutiérrez, 2001]. El nombre 10-20 indica el hecho de que los electrodos son ubicados a lo largo de la línea media en 10, 20, 20, 20, 20, y 10% del total de la distancia nasión - inión (el nasión es la unión de los huesos de la nariz y el frontal, y el inión es la parte más prominente del occipital) [Cooper *et al.*, 1980].

Por otra parte, es importante mencionar que el EEG es sensible al ruido provocado por otras actividades fisiológicas como, por ejemplo, la actividad cardíaca, el movimiento ocular, el movimiento de la lengua, la respiración, los potenciales de la piel, entre otras. Incluso, también capta ruido proveniente de la corriente alterna y de los electrodos [Sánchez de la Rosa, 1993].

### 2.3. Interfaces cerebro-computadora (BCI)

Las Interfaces Cerebro-Computadora (BCI, por sus siglas en inglés) intentan proveer al cerebro de un nuevo canal de comunicación y control para transmitir mensajes y comandos al mundo exterior [Brumberg *et al.*, 2010; Wolpaw *et al.*, 2002]. La actividad cerebral puede ser obtenida de medios no invasivos e invasivos. Dentro de los medios no invasivos están los electroencefalogramas (EEG) y los magnetoencefalogramas (MEG). Mientras que, en los invasivos se encuentran



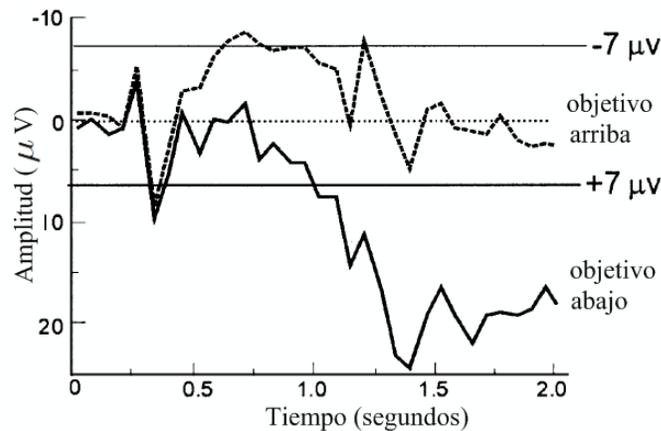
**Figura 2.3:** Relación interelectrodos de acuerdo al Sistema Internacional 10-20 [Gutiérrez, 2001]. Acotaciones de las etiquetas de los canales: fp- pre-frontal, f- frontal, p- parietal, c- central, o- occipital, t- temporal, z- línea media, a-auricular; los números impares están sobre el lado izquierdo y, los números pares están sobre el lado derecho del individuo

los electrocortigramas (ECoG), los potenciales de campo locales (LFP, por sus siglas en inglés) o las unidades simples de actividad (SUA, por sus siglas en inglés). Además, los medios invasivos requieren un procedimiento quirúrgico. Es importante mencionar que, las BCIs basadas en EEG son las más utilizadas debido a que el equipo para adquirir EEG, además de no ser invasivo, es relativamente más simple y más económico.

En las BCIs basadas en EEG, a los mecanismos neurológicos o procesos empleados por el usuario de la BCI para generar las señales de control, se les denomina fuentes electrofisiológicas. Las más ampliamente utilizadas son: los potenciales corticales lentos (SCP, por sus siglas en inglés), los potenciales P300, los ritmos sensoriales motrices ( $\mu$  y  $\beta$ ) y los potenciales evocados visuales (VEP, por sus siglas en inglés) [Bashashati *et al.*, 2007; Brumberg *et al.*, 2010; Wolpaw *et*

*al.*, 2002]. A continuación se describe brevemente cada una de estas señales.

- Potenciales corticales lentos.** Entre las frecuencias características más bajas del EEG grabado de superficie están cambios lentos de voltaje generados en la corteza. Estos cambios de potencial ocurren entre 0.5 a 10 segundos. Los SCPs negativos están típicamente asociados con movimientos y otras funciones que impliquen activación cortical, mientras que los SCPs positivos están generalmente asociados con la activación cortical reducida. La figura 2.4 muestra los SCPs presentes en la señal EEG generados por un usuario de una BCI cuando desea mover el cursor de una computadora hacia un objetivo en el monitor. Cuando el usuario desea mover el cursor hacia un objetivo en la parte inferior genera los SCPs más positivos, y cuando desea mover el cursor hacia la parte superior genera los SCPs más negativos.



**Figura 2.4:** Potenciales corticales lentos [Wolpaw *et al.*, 2002]

- Potenciales P300.** Estímulos visuales, auditivos o somatosensoriales infrecuentes o particularmente significativos, cuando intercalados con estímulos rutinarios o frecuentes, típicamente evocan en el EEG sobre la corteza parietal un pico positivo en alrededor de 300 milisegundos. En la figura 2.5 se aprecia una P300 en la señal EEG grabada cuando el usuario de una BCI desea seleccionar una opción que destella dentro de una matriz de opciones

posibles en la pantalla de la computadora.

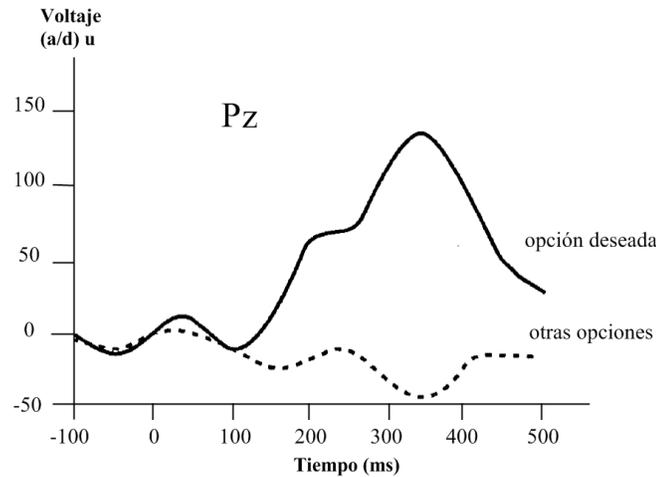
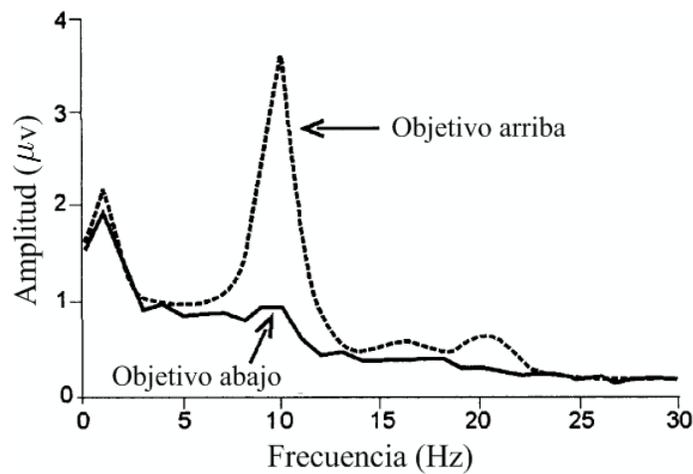


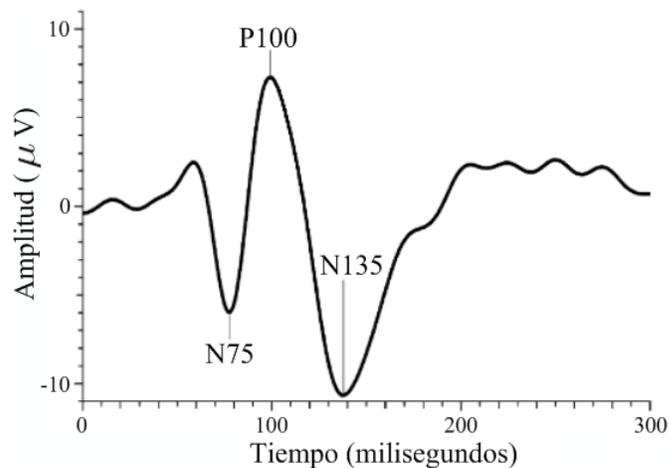
Figura 2.5: P300 [Wolpaw *et al.*, 2002]

- **Ritmos sensoriales motrices (Mu y Beta).** En la gente despierta, las principales áreas sensoriales o motrices frecuentemente exhiben actividad electroencefalográfica en el rango de 8-12 Hz. cuando ellas no están ocupadas en el procesamiento de una entrada sensorial o produciendo una salida motriz. Esta actividad es llamada ritmo mu, y generalmente está asociada con ritmos beta de 18-26 Hz. La figura 2.6 muestra los ritmos sensoriales motrices generados por un usuario de una BCI para controlar el movimiento del cursor de una computadora. En ella, se observa que el usuario aumenta la amplitud de los ritmos mu (8-12 Hz.) para mover el cursor a un objetivo en la parte superior del monitor, o la decrece para moverlo a un objetivo en la parte inferior.
- **Potenciales visuales evocados.** Son pequeños cambios en la señal cerebral en curso. Son generados en respuesta a un estímulo visual, tal como luces intermitentes y sus propiedades dependen del tipo de los estímulos visuales. Estos potenciales son más prominentes en la zona occipital (ver figura 2.7). Si un estímulo visual se presenta repetidamente a un ritmo de 5 a 6 Hz.



**Figura 2.6:** Ritmos sensoriales motrices [Wolpaw *et al.*, 2002]

o mayor, una respuesta eléctrica oscilatoria continua se suscita en las vías visuales. Esta respuesta se denomina potenciales evocados visuales de estado estable (SSVEP, por sus siglas en inglés). La diferencia entre los VEP y los SSVEP depende de la tasa de repetición de la estimulación.



**Figura 2.7:** Potenciales visuales evocados [Odom *et al.*, 2004]. Las letras N y P se refieren a picos negativos y positivos respectivamente, mientras que los números que acompañan a dichas letras se refieren a las latencias promedio después del estímulo.

## 2.4. Procesamiento Digital de Señales

El Procesamiento Digital de Señales (DSP, por sus siglas en inglés) se distingue de otras áreas en Ciencias Computacionales por el único tipo de datos que utiliza: las señales. En la mayoría de los casos, estas señales se originan como datos capturados del mundo real: las vibraciones sísmicas, las imágenes visuales, ondas de sonido, etc. DSP son las matemáticas, los algoritmos y las técnicas utilizadas para manipular estas señales después de que se han convertido a una forma digital. El DSP incluye una amplia variedad de objetivos, tales como: mejora de imágenes visuales, reconocimiento y generación de voz, compresión de datos para el almacenamiento y transmisión, entre otros [Smith, 1999].

Dentro de las técnicas matemáticas se encuentran las transformadas: de Fourier, de Fourier de tiempo breve (STFT, por sus siglas en inglés), wavelet entre otras. De especial interés para el presente trabajo es la transformada wavelet, por lo que en la sección 2.4.1 se trata detalladamente el tema.

### 2.4.1. Transformada Wavelet

En [Lotte *et al.*, 2007] se menciona que las características utilizadas en las BCI son no estacionarias ya que las señales EEG pueden rápidamente variar con el tiempo. Además, estas características deben contener información del tiempo debido a que los patrones de actividad cerebral están generalmente relacionados a variaciones específicas del EEG en el tiempo. Lo anterior, hace necesaria una representación que considere eso. Una técnica que permite modelar dichas variaciones es la transformada wavelet. A continuación se describen las versiones continua y discreta de dicha transformada.

De acuerdo con [Semmlow, 2004; Sherwood and Derakhshani, 2009], la transformada wavelet continua (CWT, por sus siglas en inglés) de una señal  $x(t)$  está dada por:

$$W(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \psi * \left( \frac{t-b}{a} \right) dt, \quad (2.1)$$

donde  $\psi$  es denominada función wavelet madre,  $b$  sirve para trasladar la función  $\psi$  a través de  $x(t)$ , y la variable  $a$  sirve para variar la escala de tiempo de la función  $\psi$ . El  $*$  indica la operación de conjugación compleja, y el factor de normalización  $\frac{1}{\sqrt{a}}$  asegura que la energía sea la misma para todos los valores de  $a$ .

Los coeficientes wavelet,  $W(a, b)$  representan la similaridad de la señal original con cada una de las muchas versiones trasladadas y escaladas de la función wavelet madre,  $\psi$ . Sin embargo, la CWT tiene el inconveniente de ser altamente redundante debido a que se sobremuestra la señal original  $x(t)$ , es decir, se generan muchos más coeficientes de los que se necesitan para especificar únicamente la señal. Esta redundancia provoca más tiempo de cómputo.

La Transformada wavelet discreta (DWT, por su siglas en inglés) provee una representación wavelet altamente eficiente mediante la restricción de la variación en la traslación y la escala, usualmente a potencias de dos. En ese caso, la DWT es algunas veces llamada transformada wavelet diádica [Wu *et al.*, 2009].

$$DWT(k, \ell) = \frac{1}{\sqrt{2^k}} \int_{-\infty}^{\infty} x(t) \psi * \left( \frac{t - 2^k \ell}{2^k} \right) dt, \quad (2.2)$$

donde  $a$  está relacionada con  $k$  como  $a = 2^k$ ;  $b$  está relacionada a  $\ell$  como  $b = 2^k \ell$ ; y  $DWT(k, \ell)$  es una muestra de  $W(a, b)$  en los puntos discretos  $k$  y  $\ell$ .

Además, la DWT puede ser implementada con un simple esquema recursivo de filtrado, y la señal original reconstruida mediante un filtrado inverso [Güler and Übeyli, 2005; Sherwood and Derakhshani, 2009; Xu and Song, 2008]. Este esquema de DWT puede ser representado con el concepto de filtros complementarios. Los filtros complementarios están compuestos de un filtro pasa altas y un filtro pasa bajas, con los que se obtienen los coeficientes wavelet de altas frecuencias (detalles,  $D_j$ ) y bajas frecuencias (aproximaciones,  $A_j$ ), respectivamente. El proceso de análisis puede ser una descomposición iterada, tal que la señal  $x(t)$  puede ser

expresada como:

$$x(t) = A_J + \sum_{j \leq J} D_j \quad (2.3)$$

donde  $A_J$  y  $D_J$  representan los coeficientes de aproximación y detalle del  $J$ -ésimo nivel.

## 2.5. Clasificación

De acuerdo con [Michie *et al.*, 1994], la clasificación cubre cualquier contexto en el que alguna decisión o pronóstico es hecho sobre la base de información histórica disponible.

Un problema de clasificación tiene una base de datos de la forma:

$$D = \{\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle, \dots, \langle x_m, y_m \rangle\} = \langle X, Y \rangle \quad (2.4)$$

donde los valores  $x_i \in X$  son típicamente vectores multi-dimensionales de la forma:  $x_i = \{z_1, z_2, \dots, z_n\}$  cuyos elementos pueden tomar valores reales o discretos. Estos componentes se denominan atributos (o características). El objetivo es inferir una función (o relación)  $f$ .

$$f : X \rightarrow Y. \quad (2.5)$$

donde los valores de  $Y$  están contenidos en un conjunto finito de clases  $C = \{C_1, \dots, C_k\}$  que caracterizan los datos dados. Los modelos aprendidos de los datos de entrenamiento son, entonces, evaluados con un conjunto de prueba distinto para determinar si los modelos pueden ser generalizados a nuevos casos [Jensen and Shen, 2008].

En [Lotte *et al.*, 2007] se realiza un análisis exhaustivo acerca de los diversos problemas que afectan a los clasificadores utilizados en las BCIs basadas en EEG. Particularmente, se hace énfasis en dos problemas. El primer problema es el denominado maldición de la dimensionalidad (*curse of dimensionality*), se debe a que

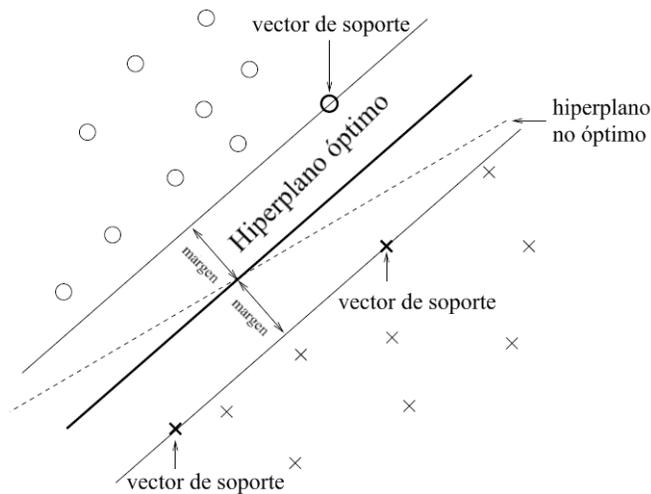
la dimensión del vector de características es mucho mayor que el número de ejemplos para el entrenamiento del clasificador. Mientras que, el segundo problema es el compromiso existente entre la varianza y el sesgo (*bias-variance trade-off*), donde la varianza representa la sensibilidad del clasificador al conjunto de datos que se usa para entrenarlo y, el sesgo representa la divergencia entre la salida estimada y la mejor salida.

La maldición de dimensionalidad se puede tratar mediante selección de características. La selección de características implica seleccionar un subconjunto mínimo, con  $M$  características,  $S = (S_1, \dots, S_M)$  del conjunto de características original  $F = (F_1, \dots, F_N)$ , donde  $M \leq N$  y  $S \subseteq F$ , de manera que el espacio de características sea óptimamente reducido y el desempeño de la clasificación sea mejorada o no decremente significativamente [Zhu *et al.*, 2010].

### 2.5.1. Máquinas de vectores de soporte

Una Máquina de vectores de soporte (SVM, por sus siglas en inglés) utiliza un hiperplano discriminante para identificar las clases [40] [41]. Sin embargo, en el caso de SVM, el hiperplano seleccionado es el que maximiza los márgenes, es decir, la distancia entre los puntos de entrenamiento más cercanos (ver figura 2.8). Maximizar los márgenes se sabe que aumenta la capacidades de generalización [Burges, 1998; Bennett and Campbell, 2000]. Además, SVM utiliza un parámetro de regularización  $C$  que es capaz de acomodar “outliers” y permite errores en el conjunto de entrenamiento.

Un SVM que permite la clasificación utilizando fronteras de decisión lineal es conocido como SVM lineal. Sin embargo, es posible crear límites no lineales de decisión, con sólo un pequeño aumento de la complejidad del clasificador, utilizando el “truco del kernel”. Consiste en la asignación implícita de los datos a otro espacio, generalmente de dimensión mucho mayor, usando una función del núcleo  $K(x, y)$ .



**Figura 2.8:** Búsqueda del hiperplano óptimo en SVM

Algunos de los núcleos que más destacan son los siguientes:

1. Lineal:  $K(x, y) = x^t y$
2. Funciones de base radial (*Radial Basis Functions*):  
gaussiana  $K(x, y) = \exp\{-\|x - y\|^2 / 2\sigma^2\}$   
exponencial  $K(x, y) = \exp\{-\|x - y\| / 2\sigma^2\}$
3. Polinomios:  $K(x, y) = (1 + \langle x, y \rangle)^d$
4. Función sigmoide:  $K(x, y) = \tanh(\rho \langle x, y \rangle + \gamma)^d$

SVM ha sido aplicado a problemas multiclase utilizando la estrategia conocida como OVR (*One Versus the Rest*), que consiste en separar cada clase de las otras [Schlögl *et al.*, 2005].

Por otra parte, SVM tiene muchas ventajas como: buenas propiedades de generalización, es insensible al sobre-entrenamiento y a la maldición de la dimensionalidad. Además, necesita que pocos hiper-parámetros sean definidos manualmente [Jain *et al.*, 2000; Burges, 1998; Bennett and Campbell, 2000].

### 2.5.2. Naive Bayes

Aunque simples, frecuentemente exhiben alta exactitud en la clasificación, comparable en rendimiento con los mejores árboles de decisión y las redes neuronales. Está basado en las probabilidades determinadas de los datos, los nuevos objetos pueden ser determinados para pertenecer a las clases con diversos grados de probabilidad [Jensen and Shen, 2008].

Naive Bayes asume que los atributos son independientes, es decir, el efecto de un atributo sobre una clase específica es independiente de los valores de los otros atributos (variables o características).

Naive Bayes está basado en el teorema de Bayes,

$$P(h|D) = \frac{P(D|h) \cdot P(h)}{P(D)}, \quad (2.6)$$

donde  $P(h)$  es la probabilidad *a priori* de la hipótesis  $h$ ,  $P(D)$  es la probabilidad *a priori* del dato de entrenamiento  $D$ ,  $P(h|D)$  es la probabilidad de  $h$  dado  $D$ , y  $P(D|h)$  es la probabilidad de  $D$  dado  $h$ . En general, la hipótesis más probable dado el dato de entrenamiento se calcula, la hipótesis máxima *a posteriori*  $h_{MAP}$ :

$$\begin{aligned} h_{MAP} &= \arg \max_{h \in H} P(h|D) \\ &= \arg \max_{h \in H} \frac{P(D|h) \cdot P(h)}{P(D)} \\ &= \arg \max_{h \in H} P(D|h) \cdot P(h) \end{aligned} \quad (2.7)$$

Si se asume que  $P(h_i) = P(h_j)$ , entonces una mayor simplificación puede ser llevada a cabo, seleccionando la hipótesis con máxima probabilidad, esto es:

$$h_{ML} = \arg \max_{h_i \in H} P(D|h_i) \quad (2.8)$$

Para la clasificación, el modelo de probabilidad Naive Bayes es combinado con una regla de decisión. Una regla común para este propósito es elegir la hipótesis más probable de un vector no etiquetado, vía MAP o ML.

### 2.5.3. Random Forests

Random Forests (RF) son una combinación de árboles predictores tal que cada uno de los árboles depende de los valores de un vector aleatorio muestreado independientemente y con la misma distribución para todo los árboles en el bosque. Cada árbol arroja un único voto para la clase más popular para una entrada  $x$  dada, y al final la salida de RF se realiza usando voto mayoritario. Los árboles individuales son construidos usando el algoritmo 1. Mientras que, el error de generalización por bosques converge casi seguramente a un límite cuando el número de árboles en el bosque llega a ser grande [Breiman, 2001].

---

#### Algoritmo 1 Random Forest

---

**Require:**  $IDT$  (un árbol de decisión),  $T$  (el número de iteraciones),  $S$  (el conjunto de entrenamiento),  $\mu$  (el tamaño de la submuestra),  $N$  (número de atributos usados en cada nodo)

**Asegurar:**  $M_t; t = 1, \dots, T$

**for**  $t \leftarrow 1$  hasta  $T$  **do**

$S_t \leftarrow$  muestra con  $\mu$  instancias de  $S$  con remplazo

Construir el clasificador  $M_t$  usando  $IDT(N)$  en  $S_t$ .

**end for**

---

De acuerdo con [Rokach, 2009], en el algoritmo 1,  $IDT$  representa un árbol de decisión con las siguientes modificaciones: el árbol de decisión no se poda, y en cada nodo, en vez de seleccionar la mejor división entre todos los atributos, el inductor de manera aleatoria muestrea  $N$  atributos y selecciona entre ellos la mejor división.

Algunas de las características de Random Forest son: su rapidez, y su capacidad para manejar, fácilmente, un gran número de atributos de entrada.

### 2.5.4. Bagging

Bagging (*Bootstrap AGGregatING*) es un método simple pero efectivo para generar un ensamble de clasificadores. El clasificador ensamblado, que es creado con este método, consolida las salidas de varios clasificadores entrenados en una única clasificación. Esto se traduce en un clasificador cuya exactitud es mayor que la exactitud de cada clasificador individual. Específicamente, cada clasificador en el ensamble es entrenado con una muestra de instancias, tomadas con reemplazo (permitiendo repeticiones), del conjunto de entrenamiento. Todos los clasificadores son entrenados usando el mismo algoritmo de aprendizaje (*inductor*).

Para asegurar que hay un número suficiente de instancias de entrenamiento en cada muestra, es común definir el tamaño de cada muestra al tamaño del conjunto de entrenamiento original. El algoritmo 2 muestra como construir un ensamble usando Bagging.

---

#### Algoritmo 2 Entrenamiento del Bagging

---

**Sean:**  $I$  (un algoritmo de aprendizaje base),  $T$  (número de iteraciones),  $S$  (el conjunto de entrenamiento original),  $\mu$  (el tamaño de la muestra).

**for**  $t \leftarrow 1$  hasta  $T$  **do**

$S_t \leftarrow$  una muestra de  $\mu$  instancias de  $S$  con reemplazo.

Construir un clasificador  $\mu_t$  usando  $I$  con  $S_t$  como el conjunto de entrenamiento.

**end for**

---

Con el fin de garantizar la diversidad entre los miembros del ensamble, un inductor relativamente inestable debería ser usado. Esto dará como resultado un ensamble de clasificadores suficientemente diferentes que pueden ser adquiridos mediante la aplicación de pequeñas perturbaciones al conjunto de entrenamiento. Si un inductor estable es utilizado, el ensamble estará compuesto de un conjunto de clasificadores que producen clasificaciones casi similares, y por lo tanto será poco probable que mejore el desempeño en exactitud.

Por otra parte, para clasificar una nueva instancia, cada clasificador retorna la predicción de la clase para la instancia desconocida. El ensamble basado en Bagging devuelve la clase con el mayor número de predicciones (también conocido como voto mayoritario). Lo anterior, se puede resumir como se muestra en el algoritmo 3.

---

**Algoritmo 3** Clasificación del Bagging
 

---

**Sea:**  $x$  (una instancia a ser clasificada)

**Asegurar:**  $C$  (predicción de la clase)

$Counter_1, \dots, Counter_{|dom(y)|} \leftarrow 0$  {inicializar los contadores de votos para las clases}

**for**  $i \leftarrow 1$  hasta  $T$  **do**

$vote_i \leftarrow M_i(x)$  {obtener la clase a la que se asigno el miembro  $i$ }.

$Counter_{vote_i} \leftarrow Counter_{vote_i} + 1$  { incrementar uno al contador de la clase correspondiente}

**end for**

$C \leftarrow$  la clase con el mayor número de votos.

Retornar  $C$ .

---

Frecuentemente, Bagging produce un modelo combinado que supera al modelo que se construye utilizando una única instancia de los datos originales. De acuerdo con [Breiman, 1996], lo anterior, es especialmente cierto para los algoritmos de aprendizaje inestables debido a que Bagging puede reducir su inestabilidad.

## 2.6. Medidas para evaluar a los clasificadores

En [Rokach, 2009] se describen algunas medidas para evaluar a los clasificadores. La más ampliamente utilizada es la exactitud (*accuracy*). Ésta se define como el número de decisiones correctas tomadas por el clasificador entre el número total de casos presentados al clasificador. Sin embargo, en muchos casos

la exactitud no es suficiente o aporta poca información respecto al desempeño del clasificador para cada una de las clases. En esta sección se describen algunas alternativas.

Antes de introducir las medidas, se parte de explicar una matriz de confusión para una clase dada. Dicha matriz contiene el número de elementos que han sido correcta o incorrectamente clasificados. La tabla 2.1 muestra la matriz de confusión para dos clases, donde la clase de interés es A. En ese caso, los verdaderos positivos  $VP$  son los ejemplos etiquetados como A que son clasificados como A, los falsos negativos  $FN$  son aquellos ejemplos cuya etiqueta es A que son clasificados como B, los falsos positivos  $FP$  son los ejemplos cuya etiqueta es B que son clasificados como A, y los verdaderos negativos  $VN$  son los ejemplos cuya etiqueta es B que son clasificados como B.

**Tabla 2.1:** Matriz de confusión para la clase A

		Clasificado como	
		A	B
clase	A	VP	FN
	B	FP	VN

La sensibilidad, también conocida como “recuerdo”, evalúa qué tan bien puede el clasificador reconocer ejemplos positivos, y se define como:

$$sensibilidad = \frac{VP}{P} = \frac{VP}{VP + FN}, \quad (2.9)$$

donde  $VP$  corresponde al número de ejemplos verdaderamente positivos, y  $P$  es el número de ejemplos positivos. Además, el número de ejemplos positivos  $P$  se compone de los verdaderos positivos  $VP$  y de los falsos negativos  $FN$ .

La especificidad mide qué tan bien puede el clasificador reconocer los ejemplos negativos. Ésta se define como:

$$especificidad = \frac{VN}{N} = \frac{VN}{VN + FP}, \quad (2.10)$$

donde  $VN$  corresponde al número de ejemplos verdaderamente negativos, y  $N$  es el número de ejemplos negativos. Además, el número de ejemplos negativos  $N$  está compuesto de los verdaderos negativos  $VN$ , y los falsos positivos  $FP$ .

Otra medida de rendimiento bien conocida es la precisión. La precisión mide cuantos ejemplos clasificados como clase “positiva” son efectivamente “positiva”. Esta medida es útil para evaluar clasificadores *crisp* que son usados para clasificar un conjunto de datos. La precisión puede ser definida como:

$$precisión = \frac{VP}{VP + FP}, \quad (2.11)$$

donde  $VP$  son los verdaderos positivos, y  $FP$  son los falsos positivos.

Por lo general, existe un compromiso entre las medidas de precisión y recuerdo. Tratando de mejorar una medida, frecuentemente, se deteriora la otra medida. Por lo que una manera de combinar ambas medidas es mediante la medida  $F$  (*F-Measure*) que resulta ser la media armónica de las dos medidas, y está dada como sigue:

$$F_{measure} = \frac{2 \cdot Precision \cdot Recuerdo}{Precision + Recuerdo} = \frac{2 \cdot VP}{FP + FN + 2 \cdot VP}, \quad (2.12)$$

donde  $FN$  son los falsos negativos,  $FP$  son los falsos positivos, y  $VP$  son los verdaderos positivos.

Por último, otra medida relevante es el coeficiente kappa ( $\kappa$ ) que determina el grado de concordancia entre dos observadores, es decir, hasta qué punto ambos coinciden en su medición. El coeficiente kappa se define como la proporción de concordancias observadas sobre el total de observaciones, donde se excluyen la concordancias atribuibles al azar. El coeficiente kappa toma valores entre  $-1$  y  $+1$ ; mientras más cercano a  $+1$ , mayor es el grado de concordancia inter-observador. Por el contrario un valor de  $\kappa = 0$  refleja que la concordancia observada es precisamente la que se espera a causa del azar [Cerdeira and Villaroel del P., 2008]. Además, de acuerdo con [Argimon Pallás and J., 2004], el coeficiente kappa tendrá un va-

lor negativo, si el porcentaje observado es menor que el esperado por azar. La ecuación 2.13 expresa matemáticamente cómo calcular el coeficiente kappa:

$$\kappa = \frac{\sum CO - \sum CAA}{N - \sum CAA}, \quad (2.13)$$

donde  $CO$  son las concordancias observadas,  $CAA$  son las concordancias atribuibles al azar, y  $N$  es el número total de observaciones.

Las concordancias atribuibles al azar se calculan mediante la sumatoria de los productos del número total de instancias clasificadas como clase  $i$  por cada uno de los clasificadores divididos entre el número total de decisiones  $N$ . Lo anterior, se expresa matemáticamente como sigue:

$$CAA = \frac{1}{N} \sum_{i=1}^C NI_i(ML_1) * NI_i(ML_2), \quad (2.14)$$

donde  $C$  es el número de clases,  $NI_i(ML_1)$  significa el número de instancias que son clasificadas como clase  $i$  por el clasificador  $ML_1$ , y  $NI_i(ML_2)$  representa al número de instancias clasificadas como clase  $i$  por el clasificador  $ML_2$ .

Una manera de interpretar los valores obtenidos mediante el coeficiente kappa se puede observar en la tabla 2.2.

**Tabla 2.2:** Valoración del coeficiente kappa [Landis and Koch, 1977]

Coeficiente kappa	Fuerza de la concordancia
$\leq 0.00$	Pobre ( <i>Poor</i> )
0.01 - 0.20	Leve ( <i>Slight</i> )
0.21 - 0.40	Aceptable ( <i>Fair</i> )
0.41 - 0.60	Moderada ( <i>Moderate</i> )
0.61 - 0.80	Considerable ( <i>Substantial</i> )
0.81 - 1.00	Casi perfecta ( <i>Almost perfect</i> )

## 2.7. Validación cruzada con $K$ -pliegues

En la validación cruzada con  $K$  pliegues ( *$K$ -fold cross-validation*), la muestra original es aleatoriamente dividida en  $K$  submuestras. De las  $K$  submuestras, una submuestra es retenida como datos de validación para probar al modelo, y los restantes  $K - 1$  submuestras son usadas como datos de entrenamiento. El proceso de validación cruzada se repite  $K$  veces (pliegues), con cada una de las  $K$  submuestras usadas, exactamente, una vez como datos de validación. Los  $K$  resultados de los pliegues entonces pueden ser promediados (o combinados). La validación cruzada con 10 pliegues es la más comúnmente usada.



# Capítulo 3

## Estado del Arte

En las BCIs basadas en EEG, a los mecanismos neurológicos o procesos empleados por el usuario de la BCI para generar las señales de control, se les denomina fuentes electrofisiológicas. Las más ampliamente utilizadas son: los potenciales corticales lentos (SCP, por sus siglas en inglés), los potenciales P300, los ritmos sensoriales motrices (mu y beta) y los potenciales evocados visuales (VEP, por sus siglas en inglés) [Bashashati *et al.*, 2007; Brumberg *et al.*, 2010; Wolpaw *et al.*, 2002]. En la sección 2.3 se describen algunos ejemplos de uso de BCIs que utilizan estas fuentes electrofisiológicas.

Las fuentes electrofisiológicas anteriores tienen como primer inconveniente que, el generarlas no es sencillo. Esto repercute en el periodo de entrenamiento requerido para que un usuario pueda operar una BCI basada en ellas, los cuales van desde algunas semanas hasta inclusive meses [Pfurtscheller, 2004]. Otra restricción de estas fuentes electrofisiológicas se debe a que sólo se pueden tomar decisiones binarias con ellas, por ejemplo si se presenta la P300 o no, si existen SCPs positivos o negativos. Lo anterior se remarca en [Desain *et al.*, 2008], en dicho trabajo se menciona que el estado del arte de BCIs basadas en las fuentes electrofisiológicas previamente mencionadas, aunque es prometedor, aún está lejos de lo necesario para controlar rápida y fiablemente juegos e interfaces. Para ejemplificar esto,

Desain [Desain *et al.*, 2008] describe dos ejemplos de BCIs maduras basadas en P300 y SCP, respectivamente. En el primer caso, la BCI que utiliza señales P300 alcanza porcentajes de exactitud altos pero la tasa de datos máxima es todavía muy lenta. Además, esta BCI puede no ser independiente porque requiere la dirección y la atención de la mirada, y no puede ser controlada sólo con actividad mental. Por otra parte, la BCI basada en SCPs sólo logra una modesta tasa de datos de control. Esta misma problemática se presenta en las BCIs que buscan permitir comunicación debido a que las cuatro fuentes electrofisiológicas requieren de un “mapeo” o traducción al dominio del lenguaje [Brumberg *et al.*, 2010], con lo que las tasas de comunicación no exceden una palabra procesada por minuto [Brumberg *et al.*, 2010].

Con esa panorámica, Desain [Desain *et al.*, 2008] cuestiona porqué no se ha logrado mayor avance a pesar de tantos años de desarrollo. Dentro de las directrices que da como respuesta está, el hecho de que la investigación ha girado principalmente en la imaginación de movimientos, y han explorado poco otras tareas mentales para ser utilizadas como fuentes electrofisiológicas. Ante esta situación, Desain [Desain *et al.*, 2008] menciona que existen muchas acciones que el ser humano puede imaginar o pensar, y que pueden ser usadas como fuentes electrofisiológicas para el control de BCIs. Dentro de ellas se encuentra el *habla no pronunciada*, la cual es el centro de interés para la presente investigación. A continuación se presentan los trabajos realizados en el reconocimiento de habla no pronunciada.

### **3.1. Reconocimiento de habla no pronunciada mediante EEG**

Los problemas en BCIs derivados del uso de las fuentes electrofisiológicas, descritas en la sección 2.3, han motivado trabajos que tratan de utilizar los po-

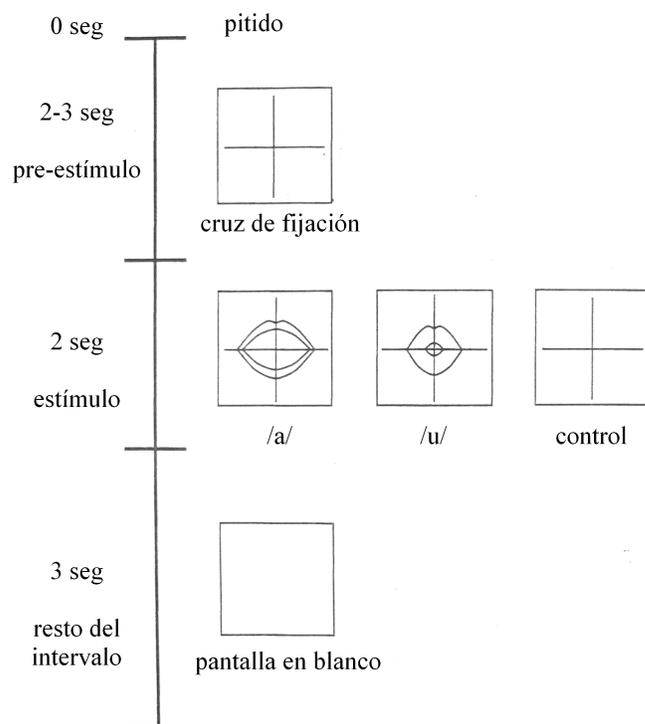
tenciales más relacionados con la producción del habla. Estos trabajos exploran el uso del habla no pronunciada, también referida como habla imaginada, para controlar BCIs. El término habla no pronunciada se refiere a la acción de imaginar la dicción de una palabra pero sin emitir ni articular sonidos. Es de gran importancia mencionar que, el habla no pronunciada se genera de manera más consciente que las cuatro fuentes electrofisiológicas usadas en BCIs (VEP, P300, ritmos mu y beta, SCP). Con lo anterior, el usuario no tiene que entrenarse para generar la fuente electrofisiológica. Además, no requiere de traducción al dominio del lenguaje para permitir comunicación.

Por otra parte, el estado del arte de los trabajos basados en habla no pronunciada permite ver que se han seguido dos enfoques: sílabas y palabras. En [Calliess, 2006; Porbadnigk, 2008; Suppes *et al.*, 1997; Wand, 2007; Wester, 2006] se explora con palabras. Mientras que, en [Brigham and Kumar, 2010; DaSalla *et al.*, 2009; D'Zmura *et al.*, 2009] únicamente se tratan sílabas. En las siguientes secciones se describen a detalle los trabajos encontrados en ambos casos, pero se tiene que hacer hincapié que el centro de atención está en aquellos que utilizan palabras.

### 3.1.1. Enfoque sílabas

El primer trabajo dentro del enfoque sílabas se describe en [DaSalla *et al.*, 2009], donde se propone un esquema de control para una BCI usando actividad neuronal asociada con la pronunciación imaginada de vocales. Las señales EEG con las que se trabaja son capturadas de tres sujetos sanos durante la realización de las siguientes tres actividades: imaginar la pronunciación de las vocales del idioma inglés /a/ y /u/, y un estado sin acción como control. El experimento para capturar las señales EEG consiste en mostrar en una pantalla, ubicada a un metro de distancia del individuo, un estímulo visual para cada una de las actividades. Para la vocal /a/ se muestra una imagen con una boca abierta, para

la vocal /u/ se muestra una imagen con los labios formando un pequeño círculo, mientras que para el control se presenta una cruz fija (ver figura 3.1). Para adquirir las señales EEG se utiliza un kit que consta de 64 electrodos cuya frecuencia de muestreo es de 2048 Hz. Posteriormente, se submuestran (*downsampled*) las señales a 256 Hz. A cada una de las señales EEG se les aplica un filtro pasa-bandas en el rango de 1-45 Hz. Cada ejemplo (*epoch*) tiene una duración de 3 segundos. Cada individuo realiza 50 muestras para cada tarea, en consecuencia se tienen 150 muestras para cada individuo. En lo que a extracción de características se refiere, se realiza utilizando el método de patrones espaciales comunes (CSP, por sus siglas en inglés). Mientras que en la etapa de clasificación se utiliza una Máquina de Vectores de Soporte (SVM, por sus siglas en inglés) no lineal para clasificar a los vectores de características. El rango de precisión está entre el 68 y el 78 % para clasificación entre dos opciones.



**Figura 3.1:** Sincronización para un ejemplo del experimento descrito en [DaSalla *et al.*, 2009]

El segundo trabajo dentro del enfoque sílabas está descrito en [D’Zmura *et al.*, 2009]. Para experimentar se grabaron las señales EEG de cuatro individuos imaginando dos sílabas , /ba/ y /ku/, sin hablar ni realizar algún movimiento. Cada una de las sílabas se imaginaron 120 veces para tres distintos ritmos. Los ritmos fueron indicados por una señal de audio. En consecuencia se tienen seis distintas condiciones.

Este trabajo tiene una etapa de pre-procesamiento que a continuación se describe. Primeramente, se segmentaron las señales EEG para proveer ondas dependientes del tiempo en cada condición, electrodo y ejemplo. En seguida, se removieron los 18 electrodos más sensibles a artefactos electromiográficos (posiciones bajas de la cabeza, y cercanas a los ojos, la sien, las orejas y la protuberancia occipital externa). Los artefactos electromiográficos (EMG) son aquellas señales eléctricas presentes en el EEG pero que son producidas por los músculos durante una contracción muscular. También, se removieron la media y la tendencia lineal de cada señal segmentada. Posteriormente, se filtraron estas señales con un filtro pasa bajas. Por último, se utilizaron umbrales para identificar y remover las ondas que no cumplan con dichos umbrales.

Terminando la etapa de pre-procesamiento, se calcularon y utilizaron para construir filtros emparejados (*matched filters*) a las envolturas de Hilbert obtenidas para cada electrodo en las bandas de frecuencias alfa, beta y teta. Además, estos filtros se calcularon para cada una de las seis condiciones con la finalidad de clasificar. Este trabajo alcanzó porcentajes de clasificación que van desde 38 % hasta el 87 %.

Para concluir esta sección se describe el trabajo presentado en [Brigham and Kumar, 2010], donde se utilizó el mismo corpus de datos del trabajo descrito en [D’Zmura *et al.*, 2009] pero se orientó la clasificación únicamente a las dos sílabas sin importar los distintos ritmos. La metodología del trabajo incluye, antes de la etapa de extracción de características, una etapa de pre-procesamiento. En esta etapa, primero se descartaron las señales provenientes de 18 de los 128 electro-

dos que están cercanos al cuello, a los ojos y a la sien debido a que son los más propensos a contener artefactos electromiográficos. Además, se eliminaron aquellos ejemplos que tenían un gran número de electrodos que excedían el umbral de  $\pm 30\mu V$  debido a que están fuertemente contaminados por artefactos EMG. Por último, filtraron la señal en un rango de frecuencias de 4-25 Hz para remover los artefactos EMG y el ruido de la corriente alterna de 60 Hz. Después de lo anterior, aplicaron análisis de componentes independientes para remover artefactos adicionales. Los componentes independientes resultantes fueron evaluados mediante el coeficiente de Hurst para rechazar a todos los que no caigan en el rango de 0.7- 0.76. Posteriormente, las señales que contengan al menos un componente independiente fueron nuevamente combinadas para restaurar las señales de los sensores. Por último, se aplicó filtrado Wiener basado en subespacios.

En la etapa de extracción de características utilizaron seis coeficientes autorregresivos (AR) mientras que, en la etapa de clasificación, un clasificador de  $k$ -vecinos más cercanos. Con su estrategia, lograron una tasa de clasificación de 61 % para 85 ejemplos correspondientes a los siete individuos.

### 3.1.2. Enfoque palabras

El primer trabajo realizado al respecto, descrito en [Suppes *et al.*, 1997], utilizó y grabó simultáneamente las señales EEG y MEG con el propósito de reconocer que palabra fue procesada de un conjunto de siete palabras ( “first”, “second”, “third”, “yes”, “no”, “right”, “left”), y en algunos experimentos se utilizan las palabras homófonas del idioma inglés (“two”, “to”, “too”) y (“hear”, “here”). Las señales EEG fueron grabadas utilizando distintas configuraciones en el número de los electrodos siendo dieciséis(F7, T3, T5, FP1, F3, C3, P3, Fz, Cz, FP2, F4, C4, P4, F8, T4, y T6), el máximo. También, se experimentaron tres situaciones: comprensión auditiva, habla interna (no pronunciada) y habla normal. De especial interés es el caso de habla interna donde únicamente se utilizaron 5 sujetos. A

cada sujeto se le mostró en la pantalla de una computadora la palabra a “decir” internamente. Cada palabra del vocabulario se grabó 100 veces.

A partir de los datos disponibles de cada palabra se formaron dos grupos compuestos de 50 ejemplos. Con el primer grupo se formó un prototipo base. Mientras que, con los 50 restantes se formaron 10 sub-grupos compuestos de 5 ejemplos para formar 10 prototipos a ser clasificados con los prototipos base creados para cada palabra. Con lo anterior, para cada individuo resultaron 7 prototipos base (uno por cada palabra), y 70 prototipos a ser clasificados (10 por cada palabra). Para ambos casos, primeramente se calculó la transformada rápida de Fourier (FFT, por sus siglas en inglés). Posteriormente, se filtró el resultado mediante un filtro pasabanda Butterworth de cuarto orden seleccionado óptimamente para cada sujeto. Después del filtrado se calculó la transformada inversa de Fourier para obtener una señal filtrada en el dominio del tiempo. En la etapa de clasificación se utilizó el criterio de mínimos cuadrados, y tomando como base la tabla 3 [Suppes *et al.*, 1997], el porcentaje de reconocimiento obtenido en el caso de habla no pronunciada es de 52.57%.

Por otra parte, en la universidad de Karlsruhe, Alemania en colaboración con la universidad de Carnegie Mellon, Estados Unidos de América se han realizado cuatro trabajos de tesis bajo la asesoría de la Dra. Tanja Schultz [Wester, 2006; Calliess, 2006; Wand, 2007; Porbadnigk, 2008]. Estas tesis tienen en común que siguen el enfoque palabras, toman como base a la señal EEG para procesarla con un sistema desarrollado para el reconocimiento de habla convencional denominado JANUS, y que la fase de clasificación es realizada mediante modelos ocultos de Markov (HMM, por sus siglas en inglés). A continuación describimos algunos detalles de cada una de estas tesis.

La primera tesis es la de Wester [Wester, 2006], quien exploró las siguientes modalidades de habla: normal, susurrada, murmurada, silenciosa (gesticulada), no pronunciada. De especial interés es el caso de habla no pronunciada. En la adquisición de las señales EEG se utilizaron 16 canales cuya frecuencia de muestreo

es de 300 Hz. Los nombres de los canales, de acuerdo al sistema internacional 10-20, son: Fp1, Fp2, F2, F3, F4, F7, C3, C4, Cz, T3, T4, T5, T6, P3, P4, Pz.

En la fase de extracción de características se utilizó la transformada de Fourier de tiempo breve (STFT, por sus siglas en inglés) con un tamaño de ventana de 26.6 milisegundos y un desplazamiento de ventana (*window shift*) de 4 milisegundos. Además, se obtuvieron los coeficientes delta, delta de los delta y la media de ambos deltas. En consecuencia, resultaron doce características para cada uno de los 16 canales, éstas se concatenaron para formar un vector de características cuya dimensión es 192. Posteriormente, se realiza reducción de la dimensión del vector de características mediante Análisis del Discriminante Lineal (LDA, por sus siglas en inglés), con lo que se redujo la dimensión de dicho vector a 35. En la etapa de clasificación, tal como se describió anteriormente, se utilizó un HMM con el que se obtuvo un 47.27 % de reconocimiento en la clasificación para un sólo individuo y un vocabulario de cinco palabras. Además, las áreas de Broca, Wernicke y la corteza motriz fueron las más importantes para el reconocimiento de habla no pronunciada.

La segunda tesis la realizó Callies [Callies, 2006], y en ella se continua el trabajo descrito en [Wester, 2006]. Esta tesis utiliza un EEG de alta densidad con 128 electrodos. Sin embargo se ven restringidos a utilizar únicamente 16 electrodos debido a que utilizan un convertidor analógico-digital de 16 canales. Además, se hace una modificación al protocolo de adquisición de datos descrito en [Wester, 2006] que consiste en utilizar a los parpadeos como marcadores para delimitar el inicio y el fin de la imaginación de la pronunciación de la palabra indicada. Lo anterior se logra mediante la implementación de un algoritmo basado en correlación cruzada. En la fase de clasificación se utiliza un HMM. Otra modificación con respecto al trabajo descrito en [Wester, 2006] es el uso de Modelos de Mezcla de Gaussianas (GMM, por sus siglas en inglés) para entrenar a los HMM.

La tercera tesis realizada por Wand [Wand, 2007] es de gran ayuda para la cuarta, a cargo de Porbadnigk [Porbadnigk, 2008] debido a que Wand determinó

la factibilidad de utilizar, en la fase de extracción de características, a la transformada wavelet compleja de árbol doble (DTCWT, por sus siglas en inglés) con tres niveles de descomposición en lugar de la STFT. En consecuencia, en [Porbadnigk, 2008] se utiliza a la DTCWT para obtener un vector con 96 características. Posteriormente el vector se reduce a 35 características mediante LDA. Mientras que en la etapa de clasificación utilizando HMM logró un reconocimiento de 45.50%. Este porcentaje de clasificación se obtiene a partir de las señales EEG de 18 sujetos grabando 16 canales mientras imaginan la pronunciación de alguna de las cinco palabras del vocabulario. Cada una de las palabras se grabó 20 veces.

## 3.2. Discusión

El estado del arte se puede resumir tal y como se muestra en la tabla 3.1. Como se puede observar en ella, los trabajos relacionados con el habla no pronunciada son pocos y relativamente recientes. Además, aún no se ha explorado un lenguaje distinto al inglés. También, es importante remarcar que los trabajos que utilizan *habla no pronunciada* pueden dividirse en dos enfoques: palabras y sílabas. El primer enfoque es seguido en [Calliess, 2006; Porbadnigk, 2008; Suppes *et al.*, 1997; Wand, 2007; Wester, 2006]. Mientras que, únicamente se tratan sílabas en [Brigham and Kumar, 2010; DaSalla *et al.*, 2009; D’Zmura *et al.*, 2009].

En el caso específico de los trabajos que exploran palabras -donde cae el presente trabajo-, se han identificado los siguientes problemas. En [Suppes *et al.*, 1997], se crearon prototipos en el dominio del tiempo, sin embargo para crear tan sólo 11 prototipos (uno base y 10 a ser clasificados) se requieren de 100 ejemplos de cómo un usuario imagina la dicción de cada una de las palabras del vocabulario, con lo que el método resulta inadecuado para llevarse a procesamiento en tiempo real y ayuda poco a mejorar las bajas tasas de comunicación. Mientras en [Calliess, 2006; Porbadnigk, 2008; Wand, 2007; Wester, 2006] se asume que las características extraídas pueden ser reconocidas con los modelos existentes

**Tabla 3.1:** Estado del Arte en el reconocimiento de *habla no pronunciada*

Trabajo	Palabras no pronunciadas	Núm. de canales EEG	Pre-procesamiento	Extracción de características	Reducción o selección de características	Clasificación
[Suppes <i>et al.</i> , 1997]	Sí	16	Sí	No	No	Mínimos cuadrados (52.57%)
[Wester, 2006]	Sí	16	No	192 coef. con STFT	35 coef. con LDA	HMM (47.27%)
[Porbadnigk, 2008]	Sí	16	No	96 coef. con DTCWT	35 coef. con LDA	HMM (45.5%)
[DaSalla <i>et al.</i> , 2009]	No(a, u, /ninguna acción/)	64	Sí	CSP	No	SVM (68%-78%)
[D’Zmura <i>et al.</i> , 2009]	No(ba, ku)	118	Sí	No	No	Filtros emparejados (38%-87%)
[Brigham and Kumar, 2010]	No(ba, ku)	118	Sí	6 coef. AR por canal	No	KNN (61%)

para reconocimiento de habla común, no obstante las señales del habla y EEG presentan características muy diferentes. Por ejemplo, las frecuencias del EEG llegan hasta 60 Hz, mientras que el habla humana está en el rango de 125 a 4000 Hz. Además, la señal del habla se capta en un sólo canal, mientras que la señal de EEG puede ser hasta inclusive 128 canales. En consecuencia, se requiere de un modelo que permita tratar al reconocimiento de palabras no pronunciadas de manera adecuada.

Otro factor a considerar es naturaleza no estacionaria de la señal EEG, por lo que se requiere un modelo que pueda ubicar variaciones repentinas en las frecuencias del mismo. De aquí, tal como se menciona en la sección 2.4.1, la necesidad

de una técnica como la transformada wavelet que es sensible a dichos cambios inherentes de las señales EEG.

Por otra parte, tal como se menciona en la sección 2.5, uno de los mayores problemas en la etapa de clasificación es la maldición de la dimensionalidad. Para tratar de reducir el impacto de éste, en los trabajos previos [Calliess, 2006; Porbadnigk, 2008; Wand, 2007; Wester, 2006] han optado por utilizar la técnica de reducción de características LDA. Sin embargo, esta técnica reduce los vectores de características pero hace difícil la interpretación de las características elegidas. En consecuencia, se requiere de un enfoque que permita ambas cosas: disminuir la dimensión de los vectores de características, y que los vectores reducidos resultantes sean más interpretables.

Otro punto a recalcar es que, las posibilidades de exploración de las señales EEG también dependen del sistema de adquisición de estas señales. Estos sistemas, por lo general, tienen distinto número de canales para monitorear. El impacto de lo anterior, se puede ver reflejado en el número de canales utilizados en los trabajos previos (ver tabla 3.1). No obstante, independiente del número de canales del sistema de adquisición de las señales EEG a ser usado; para la presente investigación, únicamente serán de interés aquellos más relacionados con las áreas del modelo Geschwind-Wernicke (ver sección 2.1.1). Con esto, también se disminuye el impacto de la maldición de la dimensionalidad en la etapa de clasificación, y se reduce un número importante de artefactos presentes en la señal EEG.

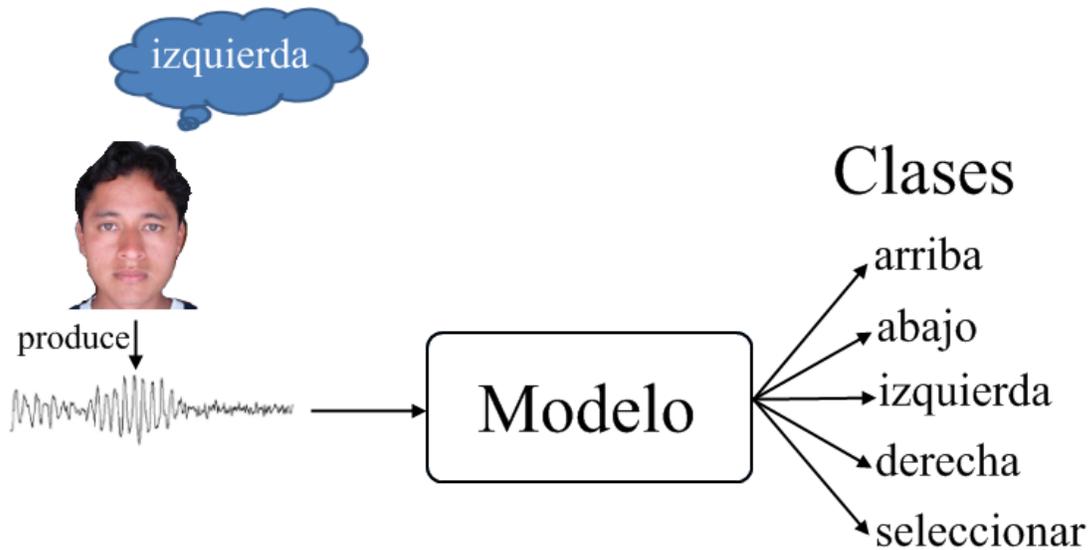


# Capítulo 4

## Método propuesto

En la sección 1.1 se menciona que el problema se trata bajo un enfoque de clasificación. Con lo anterior, se pretende hallar un modelo que pueda, a partir de una señal EEG -grabada mientras un individuo imagina la pronunciación de una palabra dada-, identificar correctamente cuál de las 5 palabras del vocabulario propuesto fue pensada. La figura 4.1 muestra el modelo general de la propuesta, en ella se observa a un individuo mientras imagina la pronunciación de la palabra “izquierda”, la cual se ve reflejada como actividad eléctrica en su cerebro que mediante el modelo se debe traducir a una de las palabras del vocabulario reducido propuesto. Es importante remarcar que, el modelo se orienta a clasificar señales EEG de un individuo en particular debido a la naturaleza no estacionaria de las señales EEG, y al uso personal de una BCI.

Para poder llevar a cabo lo anterior, es imprescindible contar con ejemplos, es decir, señales cerebrales grabadas mientras los individuos imaginan la pronunciación de las palabras. En consecuencia, se requiere una etapa de adquisición de señales cerebrales. Además, se requieren otras etapas que permitan refinar al modelo de la figura 4.1. Este refinamiento se realiza con base en [Bashashati *et al.*, 2007], donde se describen algunas etapas que están presentes en la mayoría de las BCIs. De especial interés, dado el enfoque de clasificación seguido en el



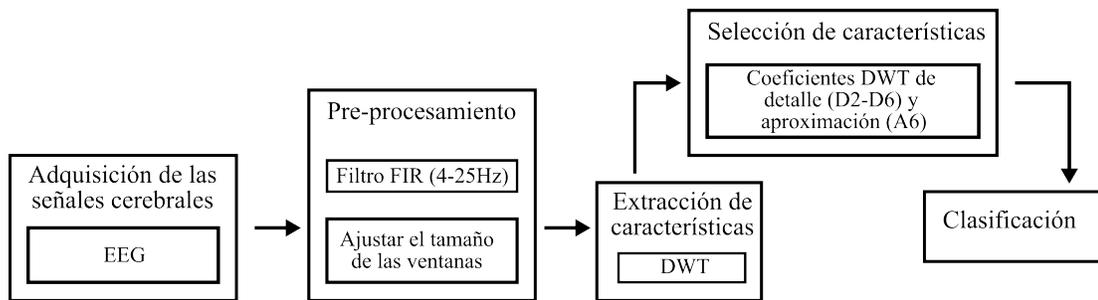
**Figura 4.1:** Modelo general del método propuesto

presente trabajo, son las etapas de: pre-procesamiento, extracción de características, selección de características/reducción de la dimensionalidad, y clasificación. Bashashati [Bashashati *et al.*, 2007] hace hincapié que cada una de estas etapas persigue un fin específico.

De acuerdo con lo anterior, las etapas del método propuesto en este trabajo son: Adquisición de las señales cerebrales, Pre-procesamiento (Mejoramiento de las señales cerebrales), Extracción de características, Selección de características, y Clasificación (ver figura 4.2). En resumen, a las señales cerebrales grabadas mientras un individuo imagina la pronunciación de una palabra, se les mejora, se les extrae características con las que se selecciona un subconjunto de ellas para entrenar y probar a los modelos de clasificación. A partir de la sección 4.2 y hasta la sección 4.6 se describen a detalle cada una de las etapas.

## 4.1. Materiales y software

Los materiales y software utilizados en el trabajo se mencionan a continuación. En lo que se refiere al monitoreo de la actividad cerebral, como se describe en la



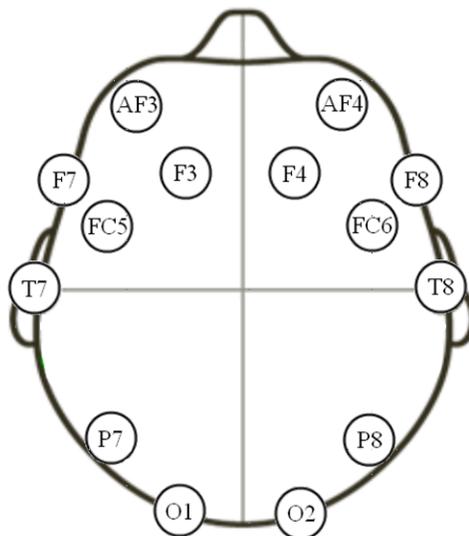
**Figura 4.2:** Método propuesto y sus correspondientes etapas

sección 2.3, existen diversas alternativas pero, en este trabajo, se opta por el EEG debido a su sencillez, y a que es relativamente económico. Específicamente, para la adquisición de señales EEG se utiliza un kit de la compañía EMOTIV<sup>©</sup>. Este kit es inalámbrico y consta de catorce electrodos (canales) de alta resolución (más las referencias CMS/DRL, P3/P4) cuya frecuencia de muestreo es de 128 Hz. Los nombres de los canales, de acuerdo con el sistema internacional 10-20, son: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 (ver figura 4.3). Además, el kit está acompañado de un software que indica el estado en el que se encuentran los sensores. Algunas de estas indicaciones son: sin señal, señal muy pobre, señal pobre, señal regular, buena señal. Este software, también, permite grabar las señales EEG que se captan en tiempo real para que posteriormente se puedan procesar.

Por otra parte, se utiliza Matlab 2009b para la implementación de programas para las etapas de preprocesamiento, extracción de características, y selección de características. Mientras que, para la etapa de clasificación se utiliza Weka 3.6.2.

## 4.2. Adquisición de las señales cerebrales

Esta etapa es de gran importancia debido a que en ella se define desde qué canales del EEG son los que se deben procesar para poder reconocer habla no pronunciada, hasta definir un protocolo que delimite cómo y dónde grabar las

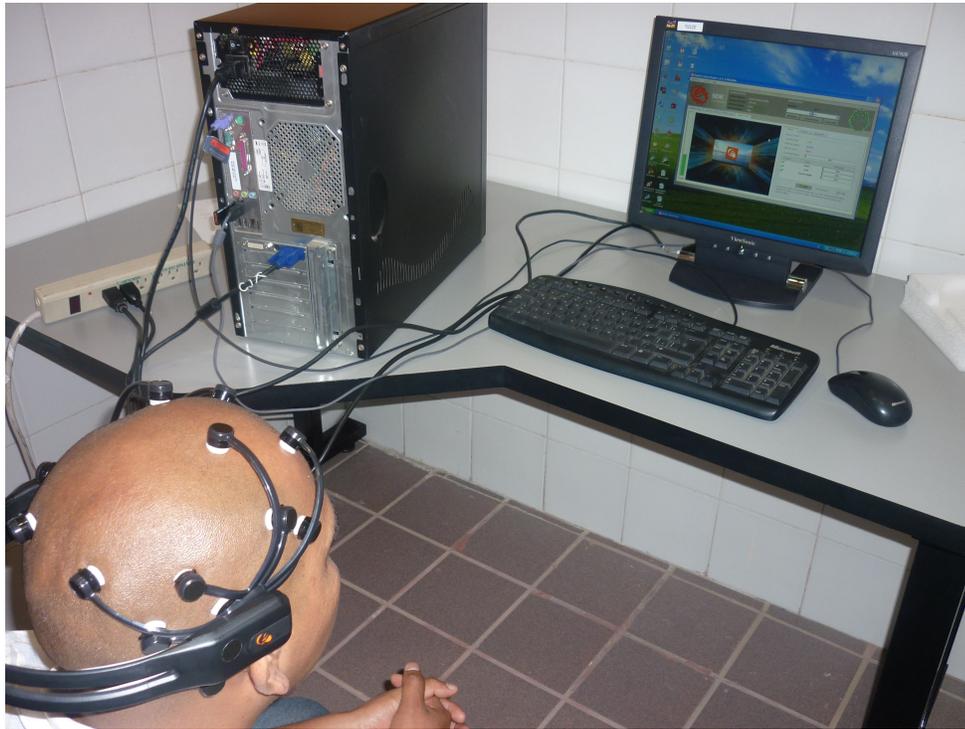


**Figura 4.3:** Ubicación de los sensores en el Kit EMOTIV

señales EEG. Así como, también, las condiciones en las cuales deben presentarse los individuos el día del experimento.

La importancia de elegir un lugar para la adquisición de las señales EEG radica en evitar el mayor número de distracciones al individuo de tal manera que éste pueda concentrarse en el experimento. Algunas distracciones que se han tomado en cuenta son: el ruido audible externo, y ruido visual (por ejemplo ver a gente pasar, notar la transición entre el día y la noche, entre otros). En la figura 4.4 se muestra el espacio acondicionado en el Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), dicho espacio está aislado de las distracciones mencionadas previamente, consta de una silla cómoda para que el usuario se sienta, y una computadora para grabar las señales EEG captadas por el kit de adquisición. También, se muestra a un individuo con el kit de adquisición de señales EEG sobre su cuero cabelludo.

Como se menciona en la sección 4.1, se utiliza un kit de adquisición que consta de catorce canales. Sin embargo, tal como se menciona en la sección 2.1.1, las señales EEG relacionadas con la producción del habla afectan diferentes áreas en la parte izquierda del cerebro. Particularmente, las áreas correspondientes a los



**Figura 4.4:** Espacio acondicionado para la adquisición de señales EEG.

canales F7, FC5, T7 y P7. Por lo tanto, en el presente trabajo únicamente son de interés las señales EEG provenientes de estos canales que son los más cercanos a las regiones del modelo Geschwind-Wernicke.

Por otra parte, los individuos a quienes se les graban sus señales EEG deben estar perfectamente descansados (sin desvelarse la noche previa), sin aplicarse gel o alguna otra sustancia en el cabello. Además, es importante que las personas no ingieran ni drogas, ni alcohol antes de la prueba. En los experimentos únicamente participan individuos diestros, y la mayoría son varones debido a que en el caso de algunas mujeres es complicado lograr que todos los sensores hagan el correcto contacto con el cuero cabelludo.

En lo referente a cómo adquirir las señales EEG de cada individuo mientras imaginan que pronuncian las palabras “arriba”, “abajo”, “izquierda”, “derecha”, y “seleccionar”; se utiliza un protocolo básico. El protocolo consiste en colocar a la persona cómodamente sentada con los ojos abiertos cerca del monitor de la

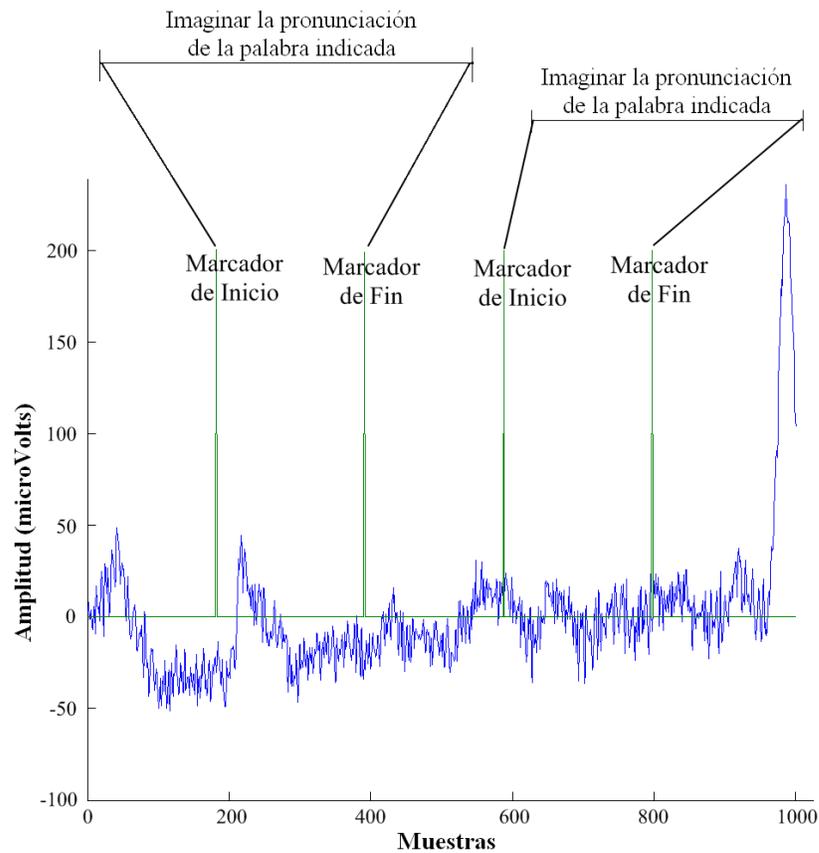
computadora, y con la mano derecha sobre el mouse. Con el mouse se envían marcadores al software que se encarga de la adquisición de las señales EEG, para delimitar el inicio y fin de la imaginación de la pronunciación de alguna de las cinco palabras indicadas (ver figura 4.5). Al conjunto de muestras que se encuentran entre los marcadores de inicio y fin, se les denomina ventanas (ejemplos). Además, al inicio de la grabación de las señales EEG, se le indica a la persona que evite parpadear o realizar movimientos corporales mientras imagina la pronunciación de la palabra indicada, ya que después de cada marcador de fin puede tomarse un tiempo de descanso para dichos movimientos.

Con la finalidad de que el individuo no sepa cuantas veces se repetirá una palabra, en la sala de experimentos otra persona, alejada del campo visual y guardando el debido silencio, se encarga de realizar el conteo de repeticiones e indica cuando el individuo debe concluir. Esto con la finalidad de que el individuo no se distraiga contando el número de repeticiones ni se predisponga a saber que le falta poco o mucho para concluir el experimento.

Los ejemplos de cómo las personas imaginan la pronunciación de la palabra indicada se guardan en un sólo archivo, es decir, se grabaran cinco archivos pertenecientes a cada individuo (un archivo por palabra). De aquí la importancia de saber *a priori* en qué parte de la señal EEG buscar.

### 4.3. Pre-procesamiento

Para extraer los segmentos de las señales EEG relevantes se puede realizar un programa con la idea de que cada dos marcadores forman una ventana. Sin embargo, existe una problemática que hace necesario que el procedimiento sea semi-automático. La figura 4.6 muestra mejor lo anterior, en ella se observan flechas sobre los marcadores en color rojo. Esto se debe a que el software de adquisición de las señales EEG recibe más de un clic casi simultáneamente producto de una falla en el sistema mecánico del botón clic izquierdo del mouse. Lo anterior, provo-

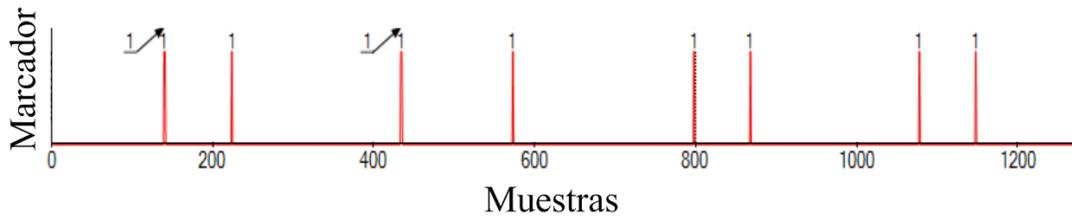


**Figura 4.5:** Señal EEG del canal F7 del individuo S1 mientras imagina la dicción de la palabra “abajo” siguiendo el protocolo de adquisición de datos

ca que, de aplicarse la estrategia descrita al inicio, se puedan tener ventanas con una sólo muestra como en el caso de los marcadores que encima tienen una flecha de la figura 4.6, y se descarten las ventanas que son verdaderamente de interés. Para corregir este tipo de errores y posteriormente aplicar la estrategia descrita al inicio de esta sección, se realiza una inspección visual sobre los marcadores de los archivos grabados con la finalidad de remover aquellos marcadores erróneos.

Las señales EEG obtenidas de los canales de interés (F7, FC5, T7 y P7) son filtradas utilizando un filtro de respuesta finita al impulso (FIR, por sus siglas en inglés) pasa-bandas en el rango de 4 a 25 Hz.

Es importante mencionar que, de manera similar al habla convencional, la duración de las ventanas de *habla no pronunciada* de una palabra es variable



**Figura 4.6:** Ejemplo de señales grabadas con marcadores erróneos

tanto en ventanas de un sólo individuo como en ventanas de individuos distintos. Por lo que, se hace necesario ajustar a todas las ventanas a un mismo tamaño. Este tamaño se determina eligiendo a la ventana de mayor tamaño de entre todos los individuos para buscar a la potencia de dos ( $2^k$  para  $k \geq 0$ ) que sea mayor y más cercana a ella. En consecuencia, todas las ventanas se rellenan con ceros al final de ellas hasta llegar a esta potencia de dos. Al final de esta etapa se tienen ventanas con 256 muestras y con un rango de frecuencias entre 4 y 25 Hz.

## 4.4. Extracción de características

Esta etapa tiene como finalidad transformar las ventanas a nuevo espacio de variables donde el problema de clasificación sea más fácil de tratar. Es aquí donde la DWT, descrita en la sección 2.4.1, toma importancia, ya que es el dominio en el cual se intentan clasificar a las señales EEG asociadas al habla no pronunciada.

La DWT precisa de una función wavelet madre  $\psi$ . En el presente trabajo se elige a la Daubechies de segundo orden (db2) debido a que ha dado buenos resultados en problemas similares descritos en [Güler and Übeyli, 2005; Jahankhani *et al.*, 2006; Orhan *et al.*, 2011]. Además, tal como se describe al final de la sección 2.4.1, la DWT se implementa mediante un esquema recursivo de filtrado donde el número de coeficientes del filtro dependen del orden de la wavelet madre, *OrdenWavelet*. En el caso de la Daubechies, el número de coeficientes de los filtros pasa bajas y pasa altas es  $2 * \text{OrdenWavelet}$ . Otro parámetro importante

de la DWT es el número de niveles de descomposición. Este parámetro se fija en el número máximo de niveles de descomposición posibles para la wavelet madre usada, y el tamaño de la señal de entrada. Esto con la finalidad de obtener las máximas resoluciones en tiempo y frecuencia, que se presentan en el primer y en el último nivel de descomposición, respectivamente. De aquí que, el número de niveles de descomposición utilizado en el presente trabajo sea seis.

Para describir la descomposición con la DWT se parte de conocer la frecuencia de muestreo ( $f_N$ ) del kit de adquisición, la cual es de 128 Hz. De acuerdo al teorema de Nyquist, la frecuencia de muestreo debe ser cuando menos dos veces mayor a la frecuencia de la señal a muestrear. Lo anterior se expresa matemáticamente como sigue:

$$f_N \geq 2f_s, \quad (4.1)$$

donde  $f_s$  es la frecuencia de la señal muestreada.

Despejando  $f_s$  en 4.1 se tiene que:

$$f_s \approx \frac{128}{2} \approx 64Hz. \quad (4.2)$$

Ahora, se toma como referencia a  $f_s$  aproximada a 64 Hz. para cada una de las ventanas, y se aplica la DWT como se ha descrito previamente. De lo anterior, se tiene que cada ventana es descompuesta en 6 niveles de detalle (D1, D2, D3, D4, D5, D6) y uno de aproximación (A6). En cada uno de los niveles hay distinto número de coeficientes y distinto rango de frecuencias.

En el caso del software usado para esta etapa, MatLab implementa la DWT mediante un esquema de filtrado recursivo, como el descrito en la sección 2.4.1, donde el número de coeficientes del  $j$ -ésimo nivel de descomposición ( $NumCoe_f_j$ ), tanto de aproximación  $cA_j$  como de detalle  $cD_j$ , se obtiene mediante la siguiente ecuación:

$$NumCoe_f_j = \left\lfloor \frac{|cA_{j-1}| - 1}{2} \right\rfloor + OrdenWavelet, \quad (4.3)$$

donde  $|cA_{j-1}|$  es el número de coeficientes de aproximación del nivel anterior, y  $OrdenWavelet$  es el orden de la wavelet madre, debido a que se utiliza una

Daubechies-2 *OrdenWavelet* = 2. Cuando  $j = 1$  se toma que  $cA_0$  es la señal de entrada, es decir, las ventanas con 256 muestras cuya frecuencia es  $f_s \approx 64$  Hz.

La tabla 4.1 muestra la descomposición wavelet que se obtiene de cada una de las ventanas. También, se observan los rangos de frecuencias en cada uno de los niveles y el número de coeficientes perteneciente a cada nivel.

**Tabla 4.1:** Descripción de los niveles de descomposición de la DWT

Nivel	Rango de frecuencias	Núm. de Coeficientes
D1	32-64	129
D2	16-32	66
D3	8-16	34
D4	4-8	18
D5	2-4	10
D6	1-2	6
A6	0-1	6

Con lo anterior, se obtiene un vector con 269 coeficientes wavelet para cada una de las ventanas en cada uno de los canales de interés. Posteriormente, se concatenan los coeficientes de los cuatro canales de interés en el orden F7-FC5-T7-P7 que están en el mismo intervalo de tiempo, es decir la misma ventana. Al final de esta etapa resulta un vector con 1076 características y su respectiva etiqueta de clase.

## 4.5. Selección de características

Esta etapa tiene como objetivo determinar un subconjunto de coeficientes wavelet que permita tener mejores resultados en la etapa de clasificación o que los resultados no decrezcan significativamente, tal como se describe en la sección 2.5. Lo anterior, es una estrategia que busca disminuir la dimensión de los vectores

de características, y reducir el impacto de la maldición de la dimensionalidad en la etapa de clasificación.

En la selección del subconjunto de coeficientes wavelet se toma como base que, de acuerdo con [Brigham and Kumar, 2010], las frecuencias de la señal EEG que son mayores a 25 Hz. están más relacionadas a actividad electromiográfica (EMG). Con lo anterior, y tomando como referencia a la tabla 4.1, se forma un subconjunto de coeficientes wavelet con los coeficientes de detalle (D2-D6) y el de aproximación (A6), que son los coeficientes con rango de frecuencia que descartan la actividad EMG. Con lo anterior, cada una de las ventanas de cada uno de los canales está representada con 140 coeficientes wavelet. Posteriormente, se concatenan los coeficientes de los cuatro canales de interés en el orden F7-FC5-T7-P7 que están en el mismo intervalo de tiempo para formar un vector con 560 características con su respectiva etiqueta de clase.

## 4.6. Clasificación

Esta etapa es, también, de gran importancia debido a que el enfoque que se le da a la tesis es de clasificación. Para esta etapa se parte de tener, para cada individuo, un conjunto de vectores de características (coeficientes DWT) que describen a cada ventana correspondiente a cómo imagina cada individuo la pronunciación de las palabras del vocabulario. Cada vector de características está asociado a una etiqueta de clase, es decir, a una palabra del vocabulario. La etiqueta de clase sirve en primera instancia durante el entrenamiento de los clasificadores, y en segunda instancia, una vez entrenados los clasificadores, sirve como parámetro para comparar la salida del clasificador. Si la salida del clasificador coincide con la etiqueta de clase (salida deseada), entonces el ejemplo fue correctamente clasificado, de lo contrario se toma como un error. El número de ejemplos correctamente e incorrectamente clasificados permiten evaluar a los clasificadores con las medidas descritas en la sección 2.6.

Por otra parte, la clasificación se puede llevar a cabo utilizando los vectores de características completos, es decir, los coeficientes DWT que se obtienen después de la etapa de extracción de características; o con los vectores reducidos que se obtienen de la etapa de selección de características.

Con estos vectores se entrenan y prueban los siguientes tres clasificadores de naturaleza distinta: Naive Bayes (NB), Random Forests (RF) y Máquina de vectores de soporte (SVM, por sus siglas en inglés). A continuación se describen algunas de las características principales de estos clasificadores. SVM tiene buenas propiedades de generalización, y son prácticamente insensibles al sobre-entrenamiento [Lotte *et al.*, 2007]. RF es un ensamble de árboles clasificadores rápido y con capacidad para manejar una gran cantidad de atributos de entrada [Rokach, 2009]. NB es un clasificador que, aunque simple, frecuentemente tiene altos porcentajes de exactitud en la clasificación, comparables a los mejores clasificadores basados en árboles y redes neuronales [Jensen and Shen, 2008]. Por último, después de evaluar a los clasificadores individuales, se elige al que tenga mejor porcentaje de exactitud para utilizarlo como clasificador base en el ensamble conocido como Bagging (ver la sección 2.5.4).

# Capítulo 5

## Experimentación y resultados

El presente capítulo presenta los experimentos y resultados que se obtienen durante la investigación. En los experimentos se evalúa a los clasificadores con: la exactitud, el coeficiente kappa, el recuerdo, la especificidad, la precisión, y la medida F (*F-Measure*). La primera sección presenta experimentos con 3 individuos utilizando *vectores de características completos*, y *vectores de características reducidos*. Los vectores de características completos se obtienen de las señales EEG que pasan por las etapas de adquisición de las señales cerebrales, pre-procesamiento, y extracción de características. Mientras que, los vectores de características reducidos se obtienen a partir de las etapas de adquisición de las señales cerebrales, pre-procesamiento, extracción de características, y selección de características. Por otra parte, la segunda sección presenta experimentos utilizando un corpus de datos que consta de veintidós individuos. Además, se muestran resultados obtenidos por los clasificadores a nivel palabra.

### 5.1. Experimentos con tres individuos

Al iniciar esta etapa del trabajo de investigación, y antes de procesar a todos los individuos, se realiza un experimento que consiste en entrenar a tres clasificadores (NB, RF, y SVM) con los datos de las señales EEG de tres individuos

(S1, S2, y S3). Los clasificadores son entrenados, y probados, utilizando los vectores completos y los vectores reducidos. Este experimento tiene la finalidad de ver la pertinencia de utilizar los vectores completos o los vectores reducidos, y de seleccionar al clasificador a ser usado como base en el ensamble conocido como *Bagging*. Para tales fines, la medida para evaluar a los clasificadores es la exactitud. La exactitud en la clasificación se obtiene mediante validación cruzada con 10 pliegues. Ésta se realiza dividiendo el corpus de datos disponible en diez particiones. Posteriormente, se utilizan nueve particiones para entrenar y una para probar al clasificador. Lo anterior se repite diez veces, de tal manera que cada una de las diez particiones se utilice únicamente una vez para probar. Los diez resultados obtenidos al probar los pliegues son promediados para dar una estimación de la exactitud.

En el caso de los vectores de características completos, las señales EEG pasan por las etapas de adquisición de la señal EEG, pre-procesamiento, y extracción de características descrito en el método propuesto (capítulo 4). Al final cada ejemplo de cada una de las palabras está descrito con 1076 características obtenidas mediante la DWT asociado con su respectiva etiqueta de clase.

Los porcentajes de exactitud de cada uno de los tres clasificadores, cuando se utilizan vectores de características completos, después de aplicar validación cruzada con 10 pliegues se pueden observar en la tabla 5.1.

**Tabla 5.1:** Porcentajes de exactitud obtenidos por los clasificadores utilizando vectores de características completos (1076 características)

	Porcentaje de exactitud		
Individuo	NB	RF	SVM
S1	23.35	24.08	23.35
S2	17.09	31.63	24.78
S3	35.75	41.21	18.18

Posteriormente, los vectores de características completos de los tres individuos (S1-S3) pasan por la etapa de selección de características (descrita en la sección 4.5), con lo que los vectores de características reducidos únicamente constan de los coeficientes wavelet de menor frecuencia, es decir, desde el segundo nivel de detalle hasta el sexto (D2, D3, D4, D5 y D6), y utilizando, también, a los coeficientes de aproximación en el sexto nivel (A6). Con lo anterior, se obtienen vectores reducidos con 560 características para describir a cada ejemplo.

La tabla 5.2 muestra los porcentajes de exactitud de cada uno de los clasificadores, utilizando vectores de características reducidos, después de aplicar validación cruzada con 10 pliegues.

**Tabla 5.2:** Porcentajes de exactitud obtenidos por los clasificadores utilizando vectores de características reducidos (540 características)

Individuo	Porcentaje de exactitud		
	NB	RF	SVM
S1	24.08	43.78	21.9
S2	18.8	38.46	21.37
S3	33.94	43.64	19.39

Los resultados obtenidos en las tablas 5.1 y 5.2 permiten observar que por lo general, cuando se seleccionan los coeficientes de detalle (D2-D6) y de aproximación (A6) se obtienen mejores resultados que cuando se utilizan vectores con todos los coeficientes wavelet como características. Es decir, en más casos, con los vectores reducidos se obtienen mayores porcentajes de exactitud en la clasificación que con los vectores completos.

Este primer experimento permite primero ver la conveniencia de utilizar únicamente vectores con 560 características, con ello se beneficia el tiempo de entrenamiento de los modelos de clasificación, y reducir también el efecto de la maldición de la dimensionalidad. Además, las tablas 5.1 y 5.2 hacen notar que

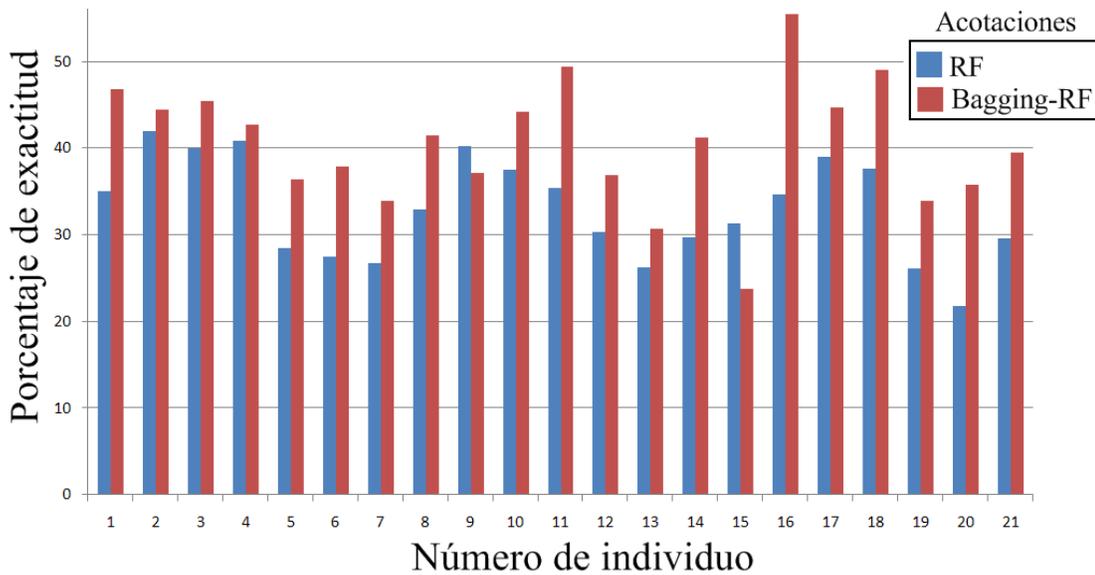
el clasificador que obtiene los mejores porcentajes de exactitud es el RF, por lo que se utiliza en los siguientes experimentos como clasificador base del *Bagging*. A partir de aquí, a este ensamble se le denota como Bagging-RF.

## 5.2. Experimentos y resultados con veintiún individuos

Para estos experimentos participan veintiún individuos (S1-S21), a cada uno se le grabaron 33 ejemplos de cómo imagina la pronunciación de cada una de las cinco palabras (“arriba”, “abajo”, “izquierda”, “derecha”, “seleccionar”). Sin embargo, aquellos ejemplos que están compuestos de más de 256 muestras (2 segundos de duración) no son tomados en cuenta en la fase de experimentación. Después de hacer esta selección resulta la distribución de ejemplos que se muestran en la tabla A.1 del apéndice A.

Los ejemplos de cada individuo pasan por todas las etapas descritas en el método propuesto, es decir, adquisición de las señales EEG, pre-procesamiento, extracción de características, selección de características y clasificación. Con lo anterior, cada ejemplo está descrito por un vector que consta de 540 características obtenidas mediante la DWT.

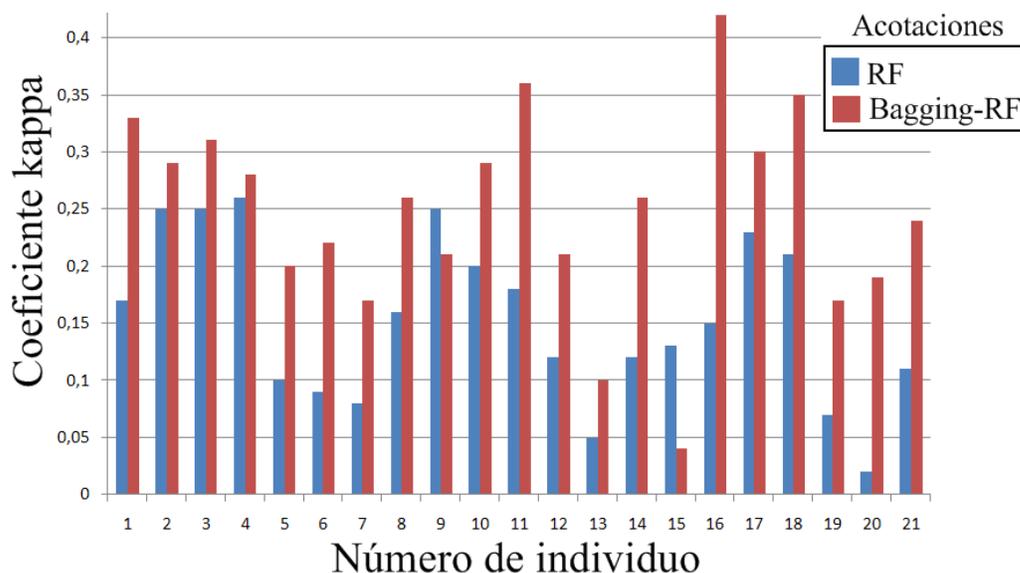
En el primer experimento se utilizan los datos de cada uno de los veintiún individuos (S1-S21) de manera separada para entrenar y probar a los cuatro clasificadores (RF, SVM, NB, Bagging-RF). Los porcentajes de exactitud obtenidos después de realizar validación cruzada con diez pliegues se pueden observar en la tabla B.1 del apéndice B. En ella se observa que los mejores clasificadores son RF y Bagging-RF. Los resultados obtenidos para ambos clasificadores se ilustran mejor en la figura 5.1.



**Figura 5.1:** Porcentajes de exactitud obtenidos por cada clasificador después de aplicar validación cruzada con diez pliegues a los datos de cada uno de los individuos

La figura 5.1 muestra que a pesar de que los porcentajes de exactitud aún son bajos (para ambos clasificadores: RF y Bagging-RF), éstos se encuentran, para todos los individuos, por encima del azar para cinco clases, el cual es del 20%. Este porcentaje de exactitud se toma como cota inferior debido a que, de acuerdo con [Dietterich, 2000], un clasificador exacto es aquel que tiene una tasa de error mejor que el azar en la etapa de generalización (prueba). También, es importante remarcar que, el clasificador Bagging-RF obtiene mejores porcentajes de exactitud que RF para casi todos los individuos, salvo en los casos de los individuos S9 y S15.

Por otra parte, también se evalúa el coeficiente kappa que obtiene cada uno de los cuatro clasificadores para cada uno de los individuos después de aplicar validación cruzada con diez pliegues. En la tabla 2.2 del apéndice B se presentan los resultados obtenidos para los cuatro clasificadores. Los individuos con peor promedio de coeficientes kappa son: S5, S6, S12, S13, S15, S19, S20, y S21. En el caso de los clasificadores, RF y Bagging-RF son los que obtienen más veces un coeficiente kappa en el rango de  $[0,2 - 0,4]$  que, de acuerdo a la tabla 2.2, quiere decir que ambos clasificadores tienen una fuerza de concordancia aceptable. Lo anterior, se ilustra de mejor manera en la figura 5.2.



**Figura 5.2:** Gráfica del coeficiente kappa obtenida por cada clasificador después de aplicar validación cruzada con diez pliegues a los datos de cada uno de los individuos

Estos resultados demuestran que a pesar de la complejidad de la tarea, existe información relevante en la señal EEG para realizar el reconocimiento de habla no pronunciada.

### 5.3. Análisis de resultados a nivel palabra

A continuación, se presentan resultados a nivel de palabra de los clasificadores Random Forest y Bagging-RF utilizando las medidas de recuerdo (sensibilidad), especificidad, precisión, y F-Measure.

Para calcular las gráficas 5.3, 5.4, 5.5, y 5.6 se parte de realizar algunas definiciones. Los datos con los que se calculan dichas gráficas se encuentran en el anexo B desde la tabla B.3 hasta la tabla B.10. De hecho, en el penúltimo renglón de las tablas, se tienen los valores promedio resultantes de las ecuaciones que a continuación se describen.

Sean:

$Palabra_1$  = “arriba”,  $Palabra_2$  = “abajo”,  $Palabra_3$  = “izquierda”,  
 $Palabra_4$  = “derecha”,  $Palabra_5$  = “seleccionar”,  $N$  el número de individuos participantes en el experimento, en este caso  $N = 21$ , y  $S_k$  es el  $k$ -ésimo individuo. Mientras que,  $RecuerdoPalabra_i(S_k)$ ,  $EspecificidadPalabra_i(S_k)$ ,  $PrecisionPalabra_i(S_k)$ ,  $FMeasurePalabra_i(S_k)$  son el recuerdo, la especificidad, la precisión y la F-Measure, respectivamente, obtenidos en los ejemplos del  $k$ -ésimo individuo para la  $i$ -ésima palabra.

Entonces, para la  $i$ -ésima palabra se tiene que el recuerdo promedio está dado como:

$$RecuerdoPromedioPalabra_i = \frac{1}{N} \sum_{k=1}^N RecuerdoPalabra_i(S_k), \quad (5.1)$$

la especificidad promedio como:

$$EspecificidadPromedioPalabra_i = \frac{1}{N} \sum_{k=1}^N EspecificidadPalabra_i(S_k), \quad (5.2)$$

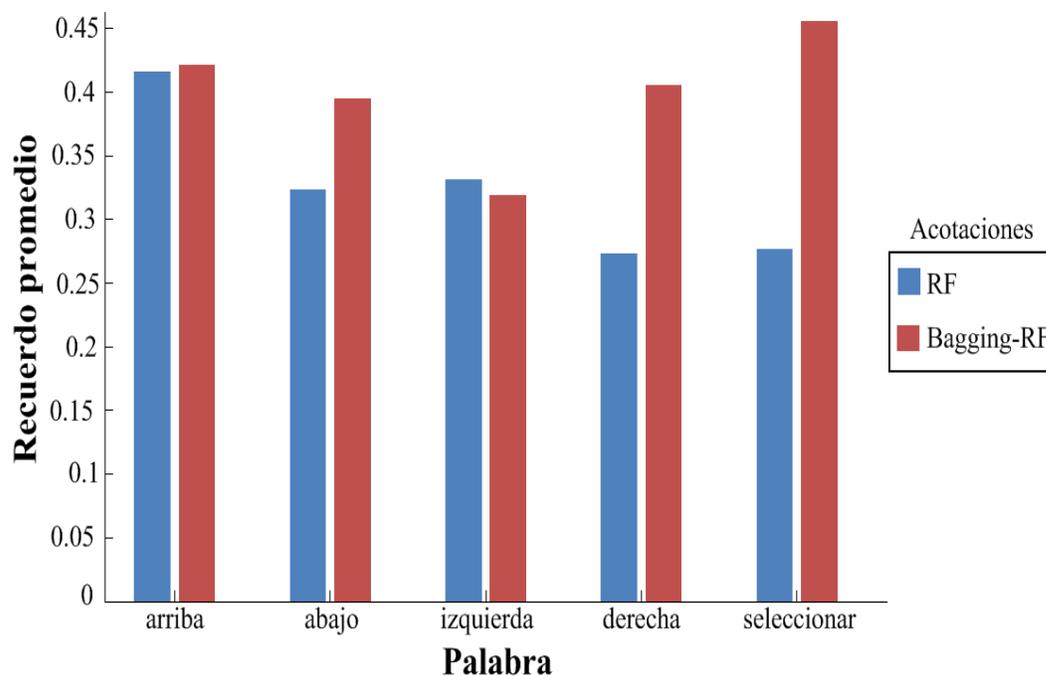
la precisión promedio de la siguiente manera:

$$PrecisionPromedioPalabra_i = \frac{1}{N} \sum_{k=1}^N PrecisionPalabra_i(S_k), \quad (5.3)$$

y la F-Measure promedio está dada como:

$$FMeasurePromedioPalabra_i = \frac{1}{N} \sum_{k=1}^N FMeasurePalabra_i(S_k). \quad (5.4)$$

La figura 5.3 muestra la tasa de recuerdo promedio obtenida por los clasificadores mediante la ecuación 5.1. El recuerdo, descrito en la sección 2.6, permite ver qué tan bien pueden los clasificadores reconocer, verbigracia, a los ejemplos etiquetados como “arriba”, y asignarlos con su predicción a la clase “arriba”. En el caso del clasificador RF, el orden de las palabras, de mayor a menor tasa de recuerdo, es: “arriba”, “izquierda”, “abajo”, “seleccionar” y “derecha”. Mientras que, en el caso de Bagging-RF, el orden de las palabras, de mayor a menor recuerdo, es: “seleccionar”, “arriba”, “derecha”, “abajo” e “izquierda”.

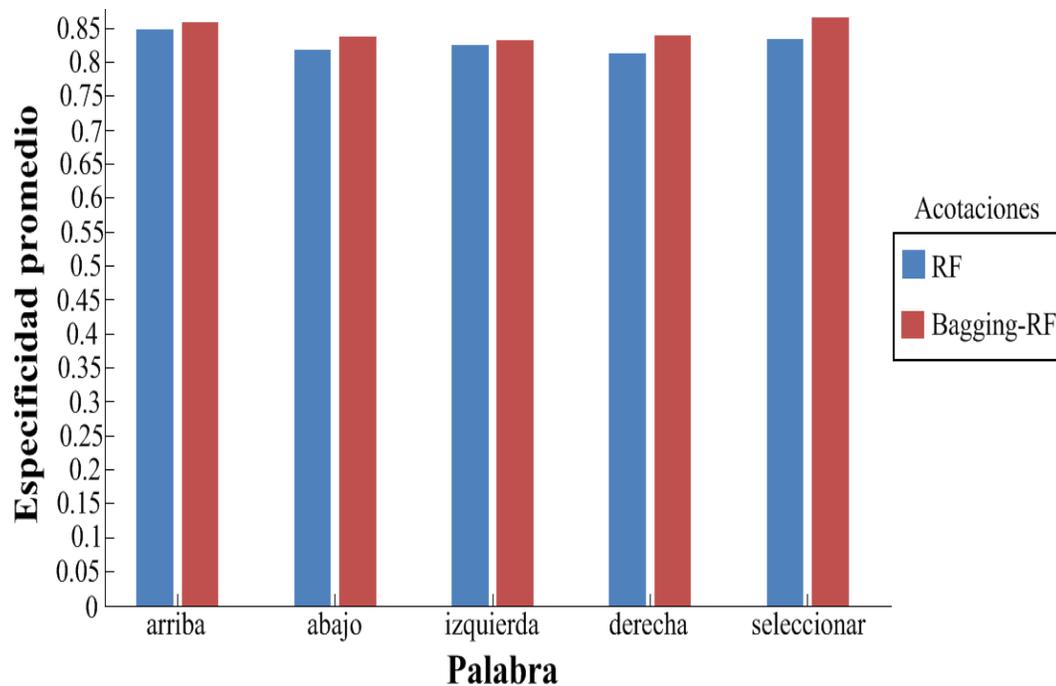


**Figura 5.3:** Gráfica del recuerdo promedio por palabra obtenida por cada clasificador después de aplicar validación cruzada con diez pliegues a los datos de cada uno de los individuos

Es importante recalcar que, el recuerdo para ambos clasificadores es mayor que el valor del azar para cinco clases en todas las palabras del vocabulario. Además, independientemente del clasificador utilizado, la palabra “arriba” se reconoce con un recuerdo mayor al doble del azar para cinco clases.

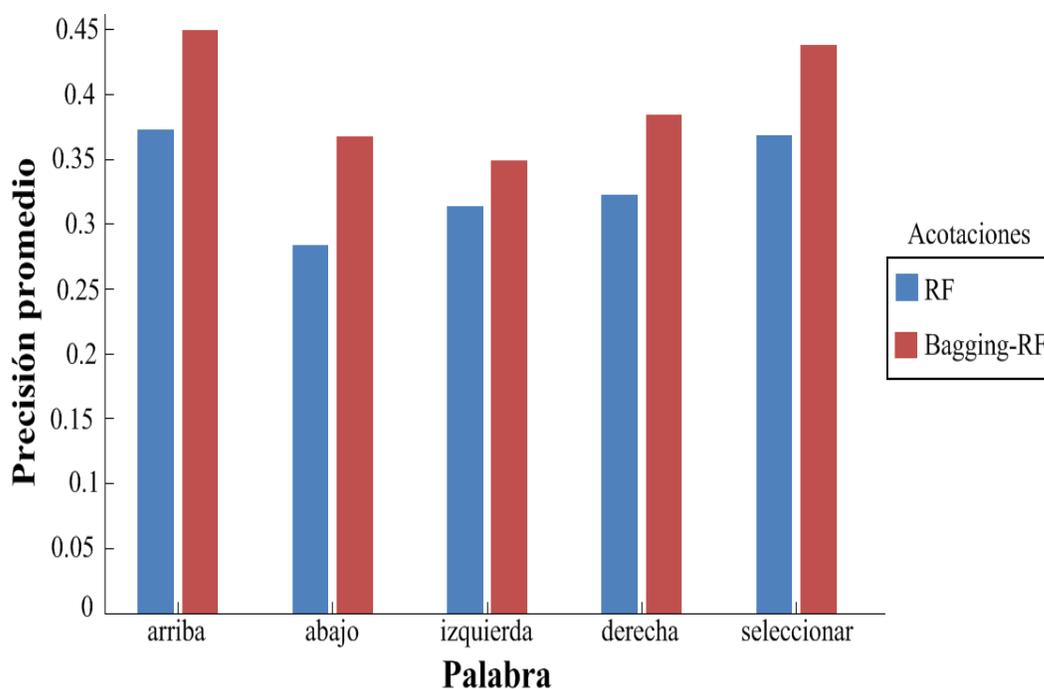
La figura 5.4 muestra la especificidad promedio obtenida mediante la ecuación 5.2 por ambos clasificadores para cada una de las cinco palabras. La especificidad, descrita en la sección 2.6, permite ver la capacidad del clasificador para descartar, verbigracia, aquellos ejemplos que no pertenecen a una clase dada. En el caso de la palabra “arriba”, esto quiere decir, clasificar con etiquetas distintas a “arriba” a aquellos ejemplos que no estén etiquetados como “arriba”.

Para ambos clasificadores, todas las palabras tienen especificidad mayor a 0.8, y no se aprecian diferencias significativas entre palabras ni entre clasificadores.



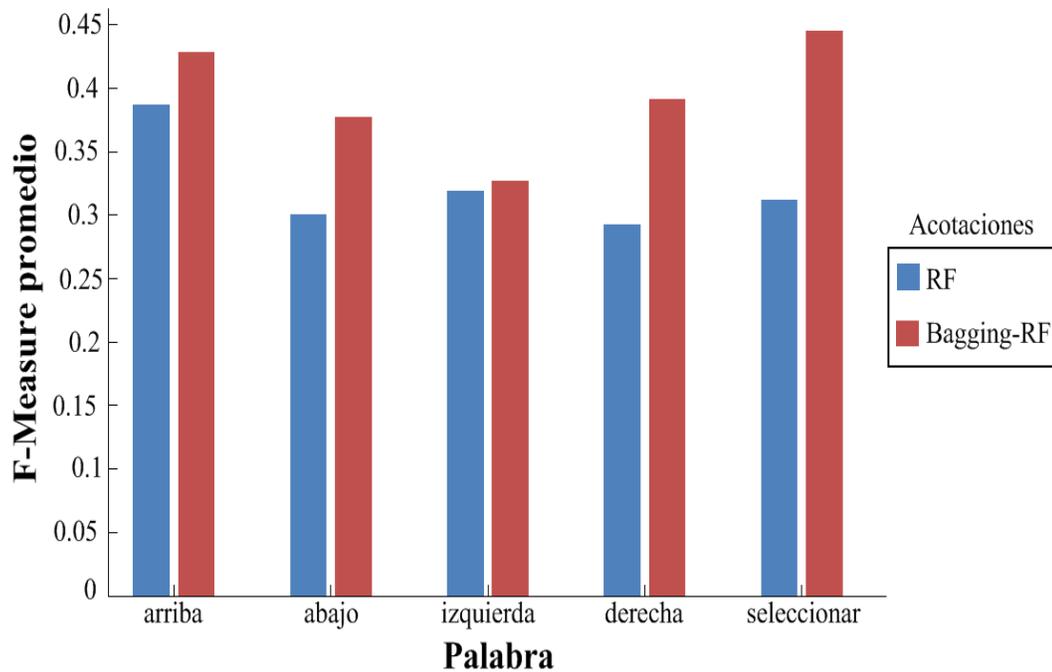
**Figura 5.4:** Gráfica de la especificidad promedio por palabra obtenida por cada clasificador después de aplicar validación cruzada con diez pliegues a los datos de cada uno de los individuos

La precisión promedio obtenida, mediante la ecuación 5.3, por cada uno de los clasificadores para cada una de las palabras se puede observar en la figura 5.5. La precisión mide qué tan cierto es que, verbigracia, un ejemplo clasificado como “arriba” efectivamente, tenga como etiqueta de clase “arriba”. En el caso del clasificador RF, el orden de las palabras, de mayor a menor tasa de precisión, es: “arriba”, “seleccionar”, “derecha”, “izquierda” y “abajo”. Mientras que, para Bagging-RF, el orden de las palabras, de mayor a menor tasa de precisión, es: “arriba”, “seleccionar”, “derecha”, “abajo” e “izquierda”.



**Figura 5.5:** Gráfica de la precisión promedio por palabra obtenida por cada clasificador después de aplicar validación cruzada con diez pliegues a los datos de cada uno de los individuos

La figura 5.6 muestra la F-Measure promedio obtenida, mediante la ecuación 5.4, por los clasificadores para cada una de las palabras. En el caso del clasificador RF, el orden de las palabras, de mayor a menor F-Measure, es: “arriba”, “izquierda”, “seleccionar”, “abajo”, y “derecha”. Mientras que, en el caso de Bagging-RF, el orden de las palabras, de mayor a menor F-Measure, es: “seleccionar”, “arriba”, “derecha”, “abajo”, e “izquierda”.



**Figura 5.6:** Gráfica de la F-Measure promedio por palabra obtenida por cada clasificador después de aplicar validación cruzada con diez pliegues a los datos de cada uno de los individuos

Es importante hacer notar, que las palabras “seleccionar” y “arriba” clasificadas con Bagging-RF tienen una F-Measure superior 0.4, es decir, dos veces mayor al azar para cinco clases.

## 5.4. Discusión

Los resultados obtenidos con el método propuesto son similares a los obtenidos por el trabajo más reciente que utilizó el enfoque de palabras, reportado en [Porbadnigk, 2008] (ver la sección 3.1.2). Específicamente, el método propuesto utilizando Bagging-RF obtiene un porcentaje de exactitud en la clasificación de 40.48 %, mientras que, en [Porbadnigk, 2008] obtuvieron un 45.50 %. Sin embargo, es necesario poner en contexto a la comparación anterior con algunas de las principales diferencias entre ambos métodos. Primeramente, se exploraron dos vocabularios distintos compuestos de 5 palabras, el vocabulario usado en [Porbadnigk, 2008] está compuesto de las palabras del idioma inglés: “alpha”, “bravo”, “charlie”, “delta” y “echo”. Mientras que, en la propuesta actual, se usaron las palabras del idioma español: “arriba”, “abajo”, “izquierda”, “derecha”, y “seleccionar”. Otra diferencia importante radica en el hecho de que, en [Porbadnigk, 2008], el idioma del vocabulario era distinto al idioma materno de los individuos. Mientras que en el trabajo actual, tanto el idioma materno de los individuos como el idioma del vocabulario es el mismo (idioma español). De aquí que, las palabras utilizadas en el presente trabajo sean semánticamente más significativas que las usadas en [Porbadnigk, 2008]. Esto también origina que el número de ideas que puedan asociarse a la pronunciación imaginada de las palabras sea mayor, y que por ende, clasificarlas con porcentajes alrededor de dos veces del valor azar para cinco clases sea destacable.

Otras diferencias entre ambos trabajos se encuentran en, el número de individuos a los que se les procesaron sus señales EEG, y en el número de repeticiones por palabra realizado por cada individuo. En [Porbadnigk, 2008] participaron 18 individuos, cada uno de los cuales repitió 20 veces cada una de las palabras del vocabulario. Mientras que, en el presente trabajo se utilizaron señales EEG de 21 individuos, cada uno de los cuales repitió alrededor de 30 veces cada una de las palabras del vocabulario propuesto. En ambos casos, la propuesta utilizó tanto

más individuos como más repeticiones de cada palabra por individuo.

También, otro factor a remarcar es la cantidad de canales utilizados. En [Porbadnigk, 2008] usaron 16 canales EEG. Mientras que en la presente propuesta únicamente fueron de interés los 4 canales más cercanos a las áreas del modelo Geschwind-Wernicke: Broca y Wernicke. Además, en el trabajo actual se seleccionaron y procesaron solamente los coeficientes obtenidos mediante la DWT que están más relacionadas al habla no pronunciada, con lo que se hace más interpretable a las características utilizadas en la clasificación a diferencia de las características que con LDA se obtuvieron en [Porbadnigk, 2008]. Con lo anterior, se reduce el impacto de artefactos electromiográficos en la clasificación, asimismo de la maldición de la dimensionalidad.

Por último, no está de más recalcar que, [Porbadnigk, 2008] procesó a la señal EEG con modelos existentes de reconocimiento de habla. Sin embargo, como ya se ha descrito, la señal de habla y la EEG son de características distintas. Asimismo, un trabajo menos comparable a la actual propuesta se describe en [Suppes *et al.*, 1997], debido a que su método resulta inadecuado para llevarse a procesamiento de tiempo real, y no ayuda en la mejora de las tasas de comunicación y control de las BCIs.



# Capítulo 6

## Conclusiones y Trabajo Futuro

En la presente investigación se desarrolló un método de procesamiento y clasificación de las señales EEG asociadas al habla no pronunciada pertenecientes a un individuo. Éste difiere de los anteriores, en el hecho de no utilizar técnicas de reconocimiento de habla convencional para tratar a la señal EEG debido a que las características del EEG y del habla son distintas. Además, se utilizaron menos canales del EEG, el vocabulario estuvo compuesto de 5 palabras en el idioma español, las palabras son semánticamente más significativas que las descritas en el trabajo más reciente del estado del arte, y los individuos tienen el mismo idioma que el del vocabulario usado. En cuanto a la manera de reducir las características, el estado del arte sólo reportaba el uso de LDA para obtener un conjunto pequeño de características a clasificar pero cuya interpretación era difícil. Ante lo anterior, en el presente trabajo se exploró el uso de los coeficientes DWT de más bajas frecuencias para entrenar y probar a los clasificadores.

El enfoque seguido a lo largo del trabajo fue de clasificación de forma individual, es decir, se utilizan datos pertenecientes a un sólo individuo para entrenar y probar a los clasificadores. Para esta tarea se emplearon cuatro distintos clasificadores (RF, NB, SVM, Bagging-RF), y se grabaron las señales EEG de veintiún individuos (S1-S21). Las señales EEG fueron procesadas con el método propuesto

para obtener coeficientes DWT que fueron utilizados en la etapa de clasificación.

## 6.1. Conclusiones

Aunque los resultados obtenidos aún distan de lo necesario para el control de BCIs, se puede observar en la figura 5.1 que, generalmente, los porcentajes de exactitud en la clasificación son superiores al 20 %, es decir, están arriba del azar para cinco clases. Con lo que se puede conjeturar que las señales EEG efectivamente contienen información que hace posible la clasificación de las palabras no pronunciadas “arriba”, “abajo”, “izquierda”, “derecha”, y “seleccionar” del idioma español. En este sentido, el uso de habla no pronunciada como fuente electrofisiológica para controlar a las BCIs es prometedor.

Por otra parte, también se evaluó a los clasificadores con el coeficiente kappa donde los clasificadores RF y Bagging-RF obtuvieron los mejores resultados, logrando para varios individuos un coeficiente kappa en el nivel de concordancia aceptable (ver figura 5.2). A nivel palabra se presentaron gráficas con distintas medidas que se pueden resumir con la medida-F que, en el caso de Random Forest la palabra no pronunciada de los individuos que se reconoce mejor es “arriba”. Mientras que, en el caso de Bagging-RF, es “seleccionar” seguida de “arriba”.

En lo referente a las palabras del vocabulario utilizado, éstas ciertamente son semánticamente más significativas que las usadas en [Porbadnigk, 2008]. Sin embargo, es importante mencionar que, las palabras “arriba”, “abajo”, “izquierda” y “derecha” están asociadas con imágenes visuales muy ambiguas. Esto hace de la tarea de clasificación de habla no pronunciada una tarea compleja, y que se requieran de estudios más específicos para evaluar el impacto de dichas imágenes en la etapa de clasificación.

Por otra parte, la elección de los canales más relacionados con el habla no pronunciada (F7, FC5, T7,P7), que cubren el área del modelo de Geschwind-Wernicke, ayudó a mitigar efectos de ruido proveniente de otros movimientos y

actividades fisiológicas. Además, en la búsqueda de disminuir el impacto de la maldición de la dimensionalidad en la clasificación, se seleccionaron coeficientes wavelet en los niveles de detalle (D2, D3, D4, D5 y D6), y en nivel de aproximación (A6) que generalmente obtuvieron porcentajes de clasificación mayores a cuando se utilizaban los vectores de características completos.

Por otra parte, también es importante remarcar que una de las contribuciones del presente trabajo estuvo en la generación de un corpus de señales EEG asociadas con el habla no pronunciada de 21 individuos diestros. Este corpus de datos está compuesto de cinco palabras en el lenguaje español, cada palabra grabada alrededor de 30 veces para cada individuo. Para lograr lo anterior, se dedicaron aproximadamente seis meses (octubre de 2010 a marzo de 2011). En estos meses, además de la adquisición de las señales EEG, se realizó el filtrado de dichas señales, y las correcciones de los marcadores erróneos.

## 6.2. Trabajo futuro

Como se describió en el trabajo, al igual que el habla convencional, la duración en tiempo del habla no pronunciada es variable. Ante esto, el trabajo actual se limita a utilizar ventanas con duración máxima de dos segundos, por lo que queda como trabajo futuro, encontrar medidas que permitan hacer más comparables a las ventanas independientemente de su duración.

Por otra parte, los resultados de esta investigación podrían ser mejorados mediante alguna de las siguientes propuestas para cada una de las etapas. En la etapa de pre-procesamiento, en búsqueda de señales más representativas que las señales EEG adquiridas de los individuos, se puede utilizar análisis de componentes independientes (ICA, por sus siglas en inglés). Además, con la finalidad de eliminar artefactos como el parpadeo o el latido de corazón, se puede evaluar a estos componentes mediante el exponente de Hurst y eliminar a aquellos que, de acuerdo a la evidencia experimental, caen en un rango de posibles artefactos.

También en esta misma etapa, se pueden explorar estrategias que permitan tratar con señales de baja intensidad como las EEG, y/o combinar las señales obtenidas de los cuatro canales de interés para formar una única señal a la que se le extraerán características usando la DWT u otro método para dicho fin. En lo que a extracción de características se refiere, seleccionar otro método para extraer características (por ejemplo, la transformada de Hilbert-Huang) o seleccionar a otra familia wavelet (Symlet, Mexican Hat, Morlet entre otras) como madre, y otro orden (actualmente se utiliza un Daubechies de segundo orden) también podría ayudar. Además, se puede explorar el uso de un esquema mixto donde, además de los coeficientes de la DWT, se puedan utilizar otros coeficientes simultáneamente como los autorregresivos. En lo que se refiere, a la etapa de clasificación quedan abiertas las posibilidades de utilizar modelos híbridos inteligentes (por ejemplo redes neuro-difusas), y otros esquemas de ensamble. Un ensamble que se podría explorar es entrenar clasificadores distintos para cada uno de los canales del EEG, y combinar la salida de todos los clasificadores mediante voto mayoritario.

# Referencias

- Argimon Pallás, J. and J., J. V. (2004). *Métodos de investigación clínica y epidemiológica*. Elsevier.
- Bashashati, A., Fatourechi, M., Ward, R., and Birch, G. (2007). A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals. *Journal of Neural engineering*, 4:R32–R57.
- Bennett, K. and Campbell, C. (2000). Support vector machines: hype or hallelujah? *ACM SIGKDD Explorations Newsletter*, 2(2):1–13.
- Boashash, B. (2003). *Time frequency signal analysis and processing: a comprehensive reference*. Elsevier Science.
- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- Brigham, K. and Kumar, B. (2010). Imagined Speech Classification with EEG Signals for Silent Communication: A Preliminary Investigation into Synthetic Telepathy. In *Bioinformatics and Biomedical Engineering (iCBBE), 2010 4th International Conference on*, pages 1–4. IEEE.
- Brumberg, J., Nieto-Castanon, A., Kennedy, P., and Guenther, F. (2010). Brain-computer interfaces for speech communication. *Speech communication*, 52(4):367–379.
- Burges, C. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167.

- Calliess, J. (2006). Further investigations on unspoken speech. Master's thesis, Institut für Theoretische Informatik Universität Karlsruhe (TH), Karlsruhe, Germany.
- Cerda, J. and Villaroel del P., L. (2008). Evaluación de la concordancia inter-observador en investigación pediátrica: Coeficiente de Kappa. *Rev. chil. pediatr*, 79(1):54–58.
- Cooper, R., Osselton, J., and Shaw, J. (1980). *EEG Technology*. Butterworth & Co Publishers Ltd, London, United Kingdom.
- DaSalla, C., Kambara, H., Sato, M., and Koike, Y. (2009). 2009 Special Issue: Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Networks*, 22(9):1334–1339.
- Denby, B., Schultz, T., Honda, K., Hueber, T., Gilbert, J., and Brumberg, J. (2010). Silent speech interfaces. *Speech Communication*, 52(4):270–287.
- Desain, P., Farquhar, J., Haselager, P., Hesse, C., and Schaefer, R. (2008). What BCI research needs. In *Proc. ACM CHI 2008 Conf. on Human Factors in Computing Systems (Venice, Italy)*.
- Dietterich, T. (2000). Ensemble methods in machine learning. *Multiple classifier systems*, pages 1–15.
- D’Zmura, M., Deng, S., Lappas, T., Thorpe, S., and Srinivasan, R. (2009). Toward EEG sensing of imagined speech. *Human-Computer Interaction. New Trends*, pages 40–48.
- Geschwind, N. (1972). Language and the brain. *Scientific American*.
- Güler, I. and Übeyli, E. (2005). Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients. *Journal of neuroscience methods*, 148(2):113–121.
- Gutiérrez, J. (2001). Análisis de señales en el neuromonitoreo. *Revista Mexicana de Ingeniería Biomédica*, 22(2):66–77.

- Jahankhani, P., Kodogiannis, V., and Revett, K. (2006). Eeg signal classification using wavelet feature extraction and neural networks. *Computing*, pages 52–57.
- Jain, A., Duin, R., and Mao, J. (2000). Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(1):4–37.
- Jensen, R. and Shen, Q. (2008). *Computational intelligence and feature selection: rough and fuzzy approaches*. Ieee Press Series On Computational Intelligence.
- Landis, J. and Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159.
- Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnald, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 4:r1–r13.
- Michie, D., Spiegelhalter, D., and Taylor, C. (1994). *Machine Learning, Neural and Statistical Classification*. Prentice Hall.
- Millán, J. (2002). *Brain-Computer Interfaces*. Handbook of Brain Theory and Neural Networks. MIT Press, Cambridge, Massachusetts, second edition.
- Muñoz, C. (2001). Potenciales evocados 1: Introducción. *Revista Mexicana de Ingeniería Biomédica*, 22(2):101–106.
- Obler, L. and Gjerlow, K. (2001). *El lenguaje y el cerebro*. Cambridge University Press, Cambridge, United Kingdom.
- Odom, J., Bach, M., Barber, C., Brigell, M., Marmor, M., Tormene, A., and Holder, G. (2004). Visual evoked potentials standard (2004). *Documenta ophthalmologica*, 108(2):115–123.
- Orhan, U., Hekim, M., Ozer, M., and Provaznik, I. (2011). Epilepsy diagnosis using probability density functions of eeg signals. In *Innovations in Intelli-*

- gent Systems and Applications (INISTA), 2011 International Symposium on*, pages 626–630. IEEE.
- Pfurtscheller, G. (2004). Brain-computer interfaces: State of the art and future prospects. In *Proceedings of the 12th European Signal Processing Conference: EUROSIPCO 04*, pages 509–510.
- Porbadnigk, A. (2008). EEG-based Speech Recognition: Impact of Experimental Design on Performance. Master’s thesis, Institut für Theoretische Informatik Universität Karlsruhe (TH), Karlsruhe, Germany.
- Rangayyan, R. (2002). *Biomedical signal analysis*. IEEE press.
- Restak, R. (1979). *The brain: The last frontier*. Doubleday.
- Rokach, L. (2009). *Pattern Classification Using Ensemble Methods*. World Scientific.
- Sagan, C. (1977). *The dragons of Eden: speculations on the evolution of human intelligence*. Random House, New York, USA.
- Schlögl, A., Lee, F., Bischof, H., and Pfurtscheller, G. (2005). Characterization of four-class motor imagery EEG data for the BCI-competition 2005. *Journal of neural engineering*, 2:L14.
- Semmlow, J. (2004). *Biosignal and biomedical image processing: MATLAB-based applications*. Marcel Dekker.
- Sherwood, J. and Derakhshani, R. (2009). On classifiability of wavelet features for EEG-based brain-computer interfaces. In *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*, pages 2895–2902. IEEE.
- Smith, S. W. (1999). *The Scientist and Engineer’s Guide to Digital Signal Processing*. California Technical Publishing, San Diego, California, second edition.
- Sánchez de la Rosa, J. L. (1993). *Métodos para el procesamiento y análisis estadístico multivariante de señales multicanal: aplicación al estudio del EEG*. PhD thesis, Universidad de La Laguna, España.

- Suppes, P., Lu, Z., and Han, B. (1997). Brain wave recognition of words. *Proceedings of the National Academy of Sciences of the United States of America*, 94(26):14965.
- Uruchurtu, G. (2010). La vida de un cerebro - De la gestación a la senectud. *¿Cómo ves?*, (142):10–14.
- Vallabhaneni, A., Wang, T., and He, B. (2005). Brain—Computer Interface. *Neural Engineering*, pages 85–121.
- Wand, M. (2007). Wavelet-based Preprocessing of Electroencephalographic and Electromyographic Signals for Speech Recognition.
- Wester, M. (2006). Unspoken Speech - Speech Recognition Based On Electroencephalography. Master's thesis, Institut für Theoretische Informatik Universität Karlsruhe (TH), Karlsruhe, Germany.
- Wolpaw, J., Birbaumer, N., McFarland, D., Pfurtscheller, G., and Vaughan, T. (2002). Brain-computer interfaces for communication and control. *Clinical neurophysiology*, 113(6):767–791.
- Wu, J.-D., Hsu, C.-C., and Wu, G.-Z. (2009). Fault gear identification and classification using discrete wavelet transform and adaptive neuro-fuzzy inference. *Expert Systems with Applications*, 36(3, Part 2):6244 – 6255.
- Xu, B. and Song, A. (2008). Pattern recognition of motor imagery EEG using wavelet transform. *J. Biomedical Science and Engineering*, 1:64–67.
- Zhu, Z., Jia, S., and Ji, Z. (2010). Towards a Memetic Feature Selection Paradigm [Application Notes]. *Computational Intelligence Magazine, IEEE*, 5(2):41–53.



# Anexo A

## Corpus de datos

El corpus de datos está formado por las señales EEG de veintiún individuos capturadas mientras imaginan la pronunciación de las cinco palabras del idioma español. Al conjunto de muestras de la señal EEG grabadas mientras la persona imagina la dicción de una palabra se le denomina ventana. La tabla A.1 sólo toma en cuenta a aquellas ventanas que no exceden las 256 muestras, y muestra la distribución del número de ejemplos de cada una de las palabras para cada uno de los individuos.

**Tabla A.1:** Distribución del número de ejemplos de cada una de las palabras

Individuo	Género	Número de ejemplos				
		Arriba	Abajo	Derecha	Izquierda	Seleccionar
S1	F	24	31	28	33	21
S2	M	31	22	23	18	23
S3	M	33	33	33	33	33
S4	F	32	33	33	33	33
S5	F	33	33	33	33	33
S6	F	33	33	33	32	33
S7	M	33	33	33	33	33
S8	M	33	33	32	33	33
S9	M	29	31	33	33	33
S10	M	23	30	27	33	23
S11	M	33	33	33	32	19
S12	M	30	28	33	30	31
S13	M	14	26	32	33	32
S14	M	33	33	33	33	33
S15	F	33	33	30	31	33
S16	F	33	30	33	33	16
S17	M	31	33	32	23	21
S18	M	17	33	33	33	33
S19	M	33	33	33	33	33
S20	M	33	33	33	33	33
S21	M	33	32	33	33	31

# Anexo B

## Resultados con el corpus completo

En la tabla B.1 se presentan los porcentajes de exactitud obtenidos por los clasificadores para cada uno de los individuos después de aplicar validación cruzada con diez pliegues. La tabla B.2 presenta los coeficientes kappa obtenidos por los clasificadores para cada uno de los individuos después de aplicar validación cruzada con diez pliegues.

Por otra parte, en las tablas B.3, B.5, B.7 y B.9 se muestran el recuerdo, la especificidad, la precisión y la medida-F, respectivamente, obtenida por el clasificador Random Forest para cada una de las palabras de vocabulario reducido. Mientras que, en las tablas B.4, B.6, B.8 y B.10 muestran el recuerdo, la especificidad, la precisión y la medida-F, respectivamente, obtenida por el clasificador Bagging-RF para cada una de las palabras de vocabulario reducido.

**Tabla B.1:** Porcentajes de exactitud obtenidos por los clasificadores para cada individuo después de aplicar validación cruzada con diez pliegues

Individuo	Porcentaje de exactitud			
	RF	SVM	NB	Bagging-RF
S1	35.03	21.16	24.81	46.71
S2	41.88	20.51	17.94	44.44
S3	40	17.57	36.36	45.45
S4	40.85	21.95	35.97	42.68
S5	28.48	20.6	26.06	36.36
S6	27.43	13.41	29.26	37.8
S7	26.66	23.63	29.09	33.93
S8	32.92	25	28.04	41.46
S9	40.25	22.01	21.38	37.1
S10	37.5	24.26	19.85	44.11
S11	35.33	24.66	20.66	49.33
S12	30.26	17.76	27.63	36.84
S13	26.27	26.27	37.95	30.65
S14	29.69	21.21	24.84	41.21
S15	31.25	23.75	22.5	23.75
S16	34.61	26.15	26.92	55.38
S17	39	17.73	29.78	44.68
S18	37.58	28.85	36.24	48.99
S19	26.06	25.45	23.63	33.93
S20	21.81	16.36	26.06	35.75
S21	29.62	23.45	19.75	39.5
Media	32.975	21.988	26.891	40.479
Desviación estándar	± 5.808	±3.809	± 5.883	± 7.154

**Tabla B.2:** Coeficientes kappa obtenidos por los clasificadores para cada individuo después de aplicar validación cruzada con diez pliegues

Individuo	Coeficiente kappa			
	RF	SVM	NB	Bagging-RF
S1	0.17	-0.02	0.08	0.33
S2	0.25	-0.05	-0.02	0.29
S3	0.25	-0.03	0.2	0.31
S4	0.26	0.02	0.19	0.28
S5	0.1	0	0.07	0.2
S6	0.09	-0.08	0.11	0.22
S7	0.08	0.04	0.11	0.17
S8	0.16	0.06	0.1	0.26
S9	0.25	0.02	0.02	0.21
S10	0.2	0.02	0	0.29
S11	0.18	0.03	0.04	0.36
S12	0.12	-0.03	0.09	0.21
S13	0.05	0.03	0.2	0.1
S14	0.12	0.01	0.06	0.26
S15	0.13	0.04	0.03	0.04
S16	0.15	0.01	0.09	0.42
S17	0.23	-0.05	0.12	0.3
S18	0.21	0.09	0.19	0.35
S19	0.07	0.06	0.04	0.17
S20	0.02	-0.04	0.07	0.19
S21	0.11	0.04	0	0.24
Media	0.152	0.008	0.085	0.248
Desviación estándar	±0.072	±0.043	±0.067	±0.089

**Tabla B.3:** Recuerdo para cada palabra de los individuos obtenida por el clasificador Random Forest después de aplicar validación cruzada con diez pliegues

	Recuerdo para cada palabra				
Individuo	arriba	abajo	izquierda	derecha	seleccionar
S1	0.29	0.25	0.57	0.17	0.42
S2	0.67	0.27	0.38	0.3	0.34
S3	0.93	0.15	0.33	0.18	0.39
S4	0.43	0.51	0.45	0.3	0.33
S5	0.45	0.18	0.27	0.18	0.33
S6	0.21	0.18	0.28	0.48	0.21
S7	0.27	0.33	0.18	0.33	0.21
S8	0.24	0.39	0.48	0.21	0.3
S9	0.48	0.35	0.42	0.57	0.18
S10	0.43	0.43	0.48	0.25	0.21
S11	0.33	0.6	0.34	0.18	0.26
S12	0.36	0.46	0.23	0.33	0.12
S13	0.35	0.15	0.27	0.18	0.37
S14	0.36	0.24	0.3	0.27	0.3
S15	0.48	0.24	0.25	0.26	0.3
S16	0.6	0.36	0.27	0.12	0.31
S17	0.38	0.66	0.26	0.25	0.31
S18	0.58	0.57	0.33	0.24	0.24
S19	0.03	0	0.42	0.57	0.27
S20	0.39	0.21	0.15	0.12	0.21
S21	0.48	0.25	0.3	0.24	0.19
Media	0.416	0.323	0.331	0.273	0.276
Desviación estándar	±0.185	±0.169	±0.107	±0.128	±0.076

**Tabla B.4:** Recuerdo para cada palabra de los individuos obtenida por el clasificador Bagging-RF después de aplicar validación cruzada con diez pliegues

Individuo	Recuerdo para cada palabra				
	arriba	abajo	izquierda	derecha	seleccionar
S1	0.54	0.29	0.51	0.28	0.8
S2	0.54	0.4	0.11	0.6	0.43
S3	0.96	0.33	0.3	0.36	0.3
S4	0.5	0.51	0.33	0.48	0.3
S5	0.48	0.36	0.24	0.21	0.51
S6	0.27	0.36	0.28	0.57	0.39
S7	0.27	0.66	0.21	0.24	0.3
S8	0.33	0.33	0.45	0.43	0.51
S9	0.17	0.29	0.27	0.72	0.36
S10	0.52	0.33	0.75	0.29	0.21
S11	0.36	0.42	0.62	0.42	0.73
S12	0.56	0.67	0.1	0.3	0.22
S13	0.21	0.23	0.21	0.31	0.5
S14	0.3	0.3	0.3	0.45	0.69
S15	0.36	0.09	0.12	0.23	0.36
S16	0.63	0.56	0.16	0.63	0.62
S17	0.29	0.6	0.47	0.37	0.5
S18	0.47	0.69	0.36	0.45	0.45
S19	0.12	0.06	0.42	0.48	0.6
S20	0.24	0.3	0.21	0.48	0.54
S21	0.72	0.5	0.27	0.21	0.25
Media	0.421	0.394	0.319	0.405	0.456
Desviación estándar	±0.202	±0.175	±0.169	±0.145	±0.168

**Tabla B.5:** Especificidad para cada palabra de los individuos obtenida por el clasificador Random Forest después de aplicar validación cruzada con diez pliegues

Individuo	Especificidad para cada palabra				
	arriba	abajo	izquierda	derecha	seleccionar
S1	0.85	0.77	0.85	0.79	0.9
S2	0.85	0.82	0.89	0.84	0.85
S3	0.98	0.79	0.83	0.8	0.84
S4	0.86	0.87	0.85	0.83	0.83
S5	0.85	0.79	0.81	0.8	0.84
S6	0.79	0.79	0.81	0.87	0.81
S7	0.81	0.83	0.79	0.83	0.8
S8	0.79	0.84	0.86	0.82	0.82
S9	0.88	0.83	0.84	0.88	0.8
S10	0.88	0.83	0.81	0.82	0.85
S11	0.8	0.87	0.82	0.79	0.89
S12	0.84	0.87	0.8	0.8	0.79
S13	0.93	0.79	0.74	0.75	0.81
S14	0.82	0.8	0.82	0.82	0.84
S15	0.85	0.79	0.82	0.83	0.82
S16	0.81	0.8	0.88	0.74	0.9
S17	0.83	0.89	0.85	0.79	0.87
S18	0.94	0.86	0.81	0.78	0.8
S19	0.75	0.75	0.84	0.89	0.83
S20	0.83	0.79	0.79	0.78	0.81
S21	0.86	0.81	0.81	0.8	0.81
Media	0.848	0.818	0.825	0.812	0.834
Desviación estándar	±0.054	±0.038	±0.033	±0.038	±0.033

**Tabla B.6:** Especificidad para cada palabra de los individuos obtenida por el clasificador Bagging-RF después de aplicar validación cruzada con diez pliegues

Individuo	Especificidad para cada palabra				
	arriba	abajo	izquierda	derecha	seleccionar
S1	0.9	0.79	0.85	0.81	0.96
S2	0.83	0.85	0.85	0.89	0.86
S3	0.99	0.83	0.83	0.83	0.82
S4	0.87	0.87	0.82	0.87	0.82
S5	0.86	0.83	0.81	0.8	0.87
S6	0.82	0.82	0.82	0.89	0.84
S7	0.82	0.9	0.8	0.81	0.83
S8	0.83	0.83	0.86	0.86	0.87
S9	0.82	0.82	0.8	0.92	0.83
S10	0.9	0.81	0.91	0.82	0.84
S11	0.83	0.83	0.89	0.83	0.96
S12	0.89	0.92	0.78	0.8	0.8
S13	0.91	0.82	0.75	0.77	0.84
S14	0.82	0.83	0.82	0.86	0.91
S15	0.82	0.77	0.79	0.82	0.83
S16	0.87	0.87	0.87	0.85	0.94
S17	0.8	0.86	0.89	0.81	0.91
S18	0.93	0.9	0.82	0.84	0.84
S19	0.77	0.76	0.86	0.87	0.89
S20	0.82	0.81	0.82	0.86	0.87
S21	0.92	0.87	0.81	0.81	0.82
Media	0.858	0.838	0.831	0.839	0.864
Desviación estándar	±0.053	±0.041	±0.040	±0.037	±0.047

**Tabla B.7:** Precisión para cada palabra de los individuos obtenida por el clasificador Random Forest después de aplicar validación cruzada con diez pliegues

Individuo	Precisión para cada palabra				
	arriba	abajo	izquierda	derecha	seleccionar
S1	0.31	0.24	0.46	0.19	0.6
S2	0.42	0.26	0.43	0.5	0.53
S3	0.79	0.16	0.34	0.25	0.32
S4	0.41	0.47	0.37	0.45	0.34
S5	0.34	0.17	0.28	0.2	0.45
S6	0.17	0.18	0.23	0.53	0.29
S7	0.24	0.35	0.18	0.33	0.22
S8	0.19	0.38	0.47	0.29	0.33
S9	0.43	0.29	0.38	0.57	0.28
S10	0.45	0.36	0.38	0.33	0.33
S11	0.27	0.41	0.39	0.3	0.35
S12	0.34	0.41	0.22	0.29	0.19
S13	0.71	0.14	0.2	0.18	0.46
S14	0.27	0.22	0.27	0.3	0.5
S15	0.41	0.2	0.26	0.32	0.37
S16	0.34	0.34	0.38	0.22	0.55
S17	0.41	0.53	0.27	0.33	0.28
S18	0.55	0.42	0.33	0.26	0.34
S19	0.02	0	0.3	0.55	0.52
S20	0.3	0.19	0.18	0.12	0.25
S21	0.45	0.23	0.27	0.25	0.24
Media	0.372	0.283	0.314	0.322	0.369
Desviación estándar	±0.172	±0.129	±0.089	±0.128	±0.119

**Tabla B.8:** Precisión para cada palabra de los individuos obtenida por el clasificador Bagging-RF después de aplicar validación cruzada con diez pliegues

Individuo	Precisión para cada palabra				
	arriba	abajo	izquierda	derecha	seleccionar
S1	0.68	0.29	0.56	0.26	0.62
S2	0.54	0.34	0.25	0.5	0.41
S3	0.8	0.34	0.35	0.34	0.33
S4	0.48	0.53	0.31	0.5	0.31
S5	0.42	0.35	0.27	0.24	0.48
S6	0.33	0.29	0.31	0.57	0.38
S7	0.29	0.48	0.22	0.26	0.35
S8	0.35	0.35	0.44	0.4	0.51
S9	0.22	0.29	0.26	0.54	0.42
S10	0.52	0.35	0.67	0.26	0.27
S11	0.48	0.4	0.55	0.42	0.66
S12	0.6	0.48	0.12	0.29	0.26
S13	0.5	0.24	0.22	0.24	0.47
S14	0.31	0.37	0.33	0.44	0.54
S15	0.3	0.11	0.13	0.25	0.33
S16	0.61	0.68	0.33	0.46	0.58
S17	0.32	0.47	0.47	0.4	0.61
S18	0.72	0.58	0.4	0.44	0.42
S19	0.1	0.05	0.5	0.55	0.55
S20	0.3	0.25	0.36	0.43	0.41
S21	0.57	0.47	0.27	0.28	0.28
Media	0.450	0.367	0.349	0.384	0.438
Desviación estándar	±0.178	±0.147	±0.142	±0.113	±0.123

**Tabla B.9:** F-Measure para cada palabra de los individuos obtenida por el clasificador Random Forest después de aplicar validación cruzada con diez pliegues

	F-Measure para cada palabra				
Individuo	arriba	abajo	izquierda	derecha	seleccionar
S1	0.3	0.25	0.51	0.18	0.5
S2	0.52	0.26	0.41	0.37	0.42
S3	0.86	0.15	0.33	0.21	0.35
S4	0.42	0.49	0.41	0.36	0.33
S5	0.38	0.17	0.27	0.19	0.38
S6	0.19	0.18	0.25	0.5	0.24
S7	0.25	0.34	0.18	0.33	0.21
S8	0.21	0.38	0.47	0.25	0.31
S9	0.45	0.32	0.4	0.57	0.22
S10	0.44	0.39	0.42	0.29	0.26
S11	0.3	0.49	0.36	0.22	0.3
S12	0.35	0.44	0.22	0.31	0.15
S13	0.47	0.14	0.23	0.18	0.41
S14	0.31	0.23	0.28	0.28	0.37
S15	0.44	0.22	0.26	0.29	0.33
S16	0.43	0.35	0.32	0.15	0.4
S17	0.4	0.59	0.26	0.28	0.29
S18	0.57	0.48	0.33	0.25	0.28
S19	0.02	0	0.35	0.56	0.36
S20	0.34	0.2	0.16	0.12	0.23
S21	0.47	0.24	0.28	0.25	0.21
Media	0.387	0.300	0.319	0.292	0.312
Desviación estándar	±0.166	±0.146	±0.094	±0.124	±0.086

**Tabla B.10:** F-Measure para cada palabra de los individuos obtenida por el clasificador Bagging-RF después de aplicar validación cruzada con diez pliegues

Individuo	F-Measure para cada palabra				
	arriba	abajo	izquierda	derecha	seleccionar
S1	0.6	0.29	0.53	0.27	0.7
S2	0.54	0.37	0.15	0.54	0.42
S3	0.87	0.33	0.32	0.35	0.31
S4	0.49	0.52	0.32	0.49	0.3
S5	0.45	0.35	0.25	0.22	0.5
S6	0.3	0.32	0.29	0.57	0.38
S7	0.28	0.56	0.21	0.25	0.32
S8	0.34	0.34	0.44	0.41	0.51
S9	0.19	0.29	0.26	0.62	0.39
S10	0.52	0.34	0.71	0.28	0.24
S11	0.41	0.41	0.58	0.42	0.7
S12	0.58	0.56	0.1	0.29	0.24
S13	0.3	0.23	0.21	0.27	0.48
S14	0.3	0.33	0.31	0.44	0.61
S15	0.32	0.1	0.13	0.24	0.34
S16	0.62	0.61	0.22	0.53	0.6
S17	0.3	0.53	0.47	0.38	0.55
S18	0.57	0.63	0.38	0.44	0.44
S19	0.11	0.05	0.45	0.51	0.57
S20	0.27	0.27	0.26	0.45	0.47
S21	0.64	0.48	0.27	0.24	0.27
Media	0.429	0.377	0.327	0.391	0.445
Desviación estándar	±0.182	±0.156	±0.155	±0.124	±0.142