

UNIVERSIDAD NACIONAL DEL LITORAL



Representación Óptima de Señales de Potenciales Evocados Relacionados a Eventos para la Implementación en Interfaces Cerebro Computadoras

Rubén Carlos Acevedo

FICH

FACULTAD DE INGENIERIA Y CIENCIAS HIDRICAS

sinc(i)

INSTITUTO DE INVESTIGACIÓN EN SEÑALES, SISTEMAS E INTELIGENCIA
ARTIFICIAL

Tesis de Doctorado **2017**



UNIVERSIDAD NACIONAL DEL LITORAL
Facultad de Ingeniería y Ciencias Hídricas

Santa Fe, 30 de marzo de 2017.

Como miembros del Jurado Evaluador de la Tesis de Doctorado en Ingeniería titulada ***“Representación óptima de señales de potenciales evocados relacionados a eventos para la implementación en interfaces cerebro computadoras”***, desarrollada por el Bioing. Rubén Carlos ACEVEDO, en el marco de la Mención **“Inteligencia Computacional, Señales y Sistemas”**, certificamos que hemos evaluado la Tesis y recomendamos que sea aceptada como parte de los requisitos para la obtención del título de Doctor en Ingeniería.

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



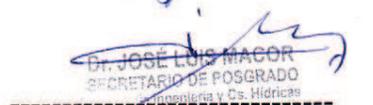
 Dr. Eric Laciari Leber



 Dr. Pablo Diez



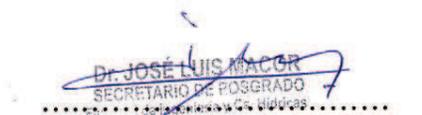
 Dr. Diego Tomassi



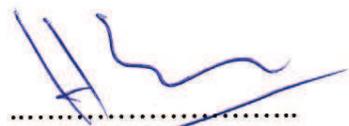
 Dr. Enrique Spinelli (*)

Santa Fe, 30 de marzo de 2017.

Certifico haber leído la Tesis, preparada bajo mi dirección en el marco de la Mención **“Inteligencia Computacional, Señales y Sistemas”** y recomiendo que sea aceptada como parte de los requisitos para la obtención del título de Doctor en Ingeniería.

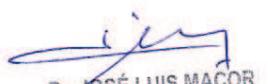


 Dra. Verónica Medina Bañuelos
 Codirectora de Tesis (*)



 Dr. Leonardo Rufiner
 Director de Tesis

* El Dr. Spinelli y la Dra. Verónica Medina Bañuelos participaron por video conferencia



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

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

Ciudad Universitaria
 C.C. 217
 Ruta Nacional N° 168 - Km. 472,4
 (3000) Santa Fe
 Tel: (54) (0342) 4575 229
 Fax: (54) (0342) 4575 224
 E-mail: posgrado@fich.unl.edu.ar

REPRESENTACIÓN ÓPTIMA DE
SEÑALES DE POTENCIALES
EVOCADOS RELACIONADOS A
EVENTOS PARA LA
IMPLEMENTACIÓN EN INTERFACES
CEREBRO COMPUTADORAS

TESIS PRESENTADA POR RUBÉN CARLOS ACEVEDO
PARA OBTENER EL GRADO DE DOCTOR EN INGENIERÍA - MENCIÓN
SEÑALES, SISTEMAS E INTELIGENCIA COMPUTACIONAL

DIRECTORES: DR. LEONARDO RUFINER, DRA. VERÓNICA MEDINA
BAÑUELOS



UNIVERSIDAD NACIONAL DEL LITORAL

FACULTAD DE INGENIERÍA Y CIENCIAS HÍDRICAS

2018

A Paola

Federico, Juan Pablo y Valentín

Agradecimientos

Son muchas las personas que han colaborado, de una forma u otra, para poder hacer realidad esta tesis y a las cuales quiero agradecer.

A la comunidad de la Facultad de Ingeniería de la Universidad Nacional de Entre Ríos (FIUNER), docentes y administrativos de la casa, y en particular a mis compañeros del equipo de gestión de la FIUNER representados en la persona del Decano Dr. Gerardo G. Gentiletti, por el permanente apoyo y aliento para culminar con esta etapa académica de mi vida.

A mis colegas de trabajo del Laboratorio de Ingeniería en Rehabilitación e Investigaciones Neuromusculares y Sensoriales (LIRINS) de la FIUNER, por el apoyo recibido, en particular a Yanina Atum, Iván Gareis y José Biurrun Manresa, por la colaboración y el tiempo dedicado a fructíferas discusiones.

A los alumnos que han realizado adscripciones, pasantías, proyectos finales y tesis de maestría en el LIRINS, en particular a Victoria Peterson, Florencia Jauregui, Agostina Larrazábal, José Pentácolo y Cristian Arjona.

A los integrantes del Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional (*sinc(i)*) de la FICH-UNL, quienes me hicieron sentir como uno más de ellos durante mis estancias en el instituto.

A mis directores, la Dra. Verónica Medina Bañuelos, quien siempre estuvo para aconsejarme cuando lo necesité, y el Dr. Leonardo Rufiner, quien además de guiarme me brindó su amistad y fue un apoyo fundamental en los momentos de duda.

A mis hijos Federico, Juan Pablo y Valentín, por la paciencia y comprensión cuando no podía dedicarles el tiempo que merecían y sin embargo siempre conté con su apoyo y amor.

A mi esposa Paola, quien durante estos años, y sobre todo en los momentos en los cuales parecía que no podía cumplir el objetivo, siempre estuvo a mi lado para apoyarme y alentarme, y si no fuera por ella esta tesis no sería una realidad.

Gracias a todos.

Índice general

Agradecimientos	3
Resumen	15
Abstract	18
Introducción	21
1. Descripción del problema	25
1.1. Conceptos de electrofisiología	25
1.1.1. Electroencefalograma	25
1.1.2. Potenciales evocados relacionados con eventos	27
1.2. Interfaces cerebro computadoras	30
1.2.1. Interfaces cerebro computadoras basadas en ERP	33
1.3. Planteo conceptual de la solución propuesta	35
1.3.1. Generación de características	36
1.3.2. Selección de características	38
1.3.3. Clasificación	39
1.4. Comentarios de cierre de capítulo	39
2. Generación de características mediante bases ortogonales	40
2.1. Conceptos generales	40
2.2. Transformada Discreta de Fourier	42
2.3. Transformada Wavelet Diádica Discreta	43
2.4. Transformada Wavelet Packet	47
2.4.1. Algoritmo Best Orthogonal Basis	48
2.4.2. Algoritmo Local Discriminant Basis	49
2.5. Comentarios de cierre de capítulo	51

3. Selección de características	52
3.1. Conceptos generales	52
3.2. Algoritmos genéticos	55
3.3. Eliminación recursiva de características	57
3.4. Comentarios de cierre de capítulo	58
4. Clasificación	59
4.1. Conceptos generales	59
4.2. Discriminante lineal	60
4.3. Ensamble de clasificadores	61
4.3.1. Estrategias de combinación de salidas	62
4.3.2. Estrategias de construcción	63
4.4. Comentarios de cierre de capítulo	63
5. Materiales y métodos	65
5.1. Base de datos	65
5.2. Conformación de patrones	68
5.3. Índices de desempeño	71
5.4. Entrenamiento y prueba del clasificador	72
5.5. Experimentos preliminares	72
5.5.1. Elección de wavelet madre para generación de patrones DDWT y WPT	73
5.5.2. Mejoramiento de la relación señal a ruido en el dominio wavelet	76
5.5.3. Selección de medida de discriminación del algoritmo LDB para generación de patrones WPT	80
5.5.4. Algoritmos de selección de bases a partir de un diccionario Wavelet Packets	82
5.5.5. Determinación del punto óptimo de entrenamiento del LDA	83
5.5.6. Evaluación de ensambles de discriminantes lineales	84
5.6. Comentarios de cierre de capítulo	86
6. Experimentos y resultados	88
6.1. Descripción de experimentos y resultados obtenidos	88
6.2. Experimento 1: índices de desempeño de referencia	89

6.3. Experimento 2: generación de características mediante bases ortogonales	91
6.4. Experimento 3: suavizado de patrones DDWT	94
6.5. Experimento 4: selección de características	96
6.5.1. Selección de características mediante eliminación recursiva de características	96
6.5.2. Selección de características mediante algoritmos genéticos	99
6.6. Comentarios de cierre de capítulo	103
7. Discusión	104
7.1. Resultados del Experimento 1	104
7.2. Resultados del Experimento 2	104
7.3. Resultados del Experimento 3	106
7.4. Resultados del Experimento 4	107
7.5. Análisis estadístico	111
7.5.1. Selección de características en patrones temporales	111
7.5.2. Generación de características mediante DDWT y suavizado	112
7.5.3. Generación y selección de características en patrones wavelets	112
7.5.4. Análisis final	113
7.6. Comparación de resultados	113
8. Conclusiones	116
8.1. Contribuciones	117
8.2. Trabajo futuro	118
8.3. Lista de publicaciones derivadas de la tesis	119
8.3.1. Publicaciones en revistas	119
8.3.2. Proceedings de congresos internacionales	121
8.3.3. Proceedings de congresos nacionales	122
Bibliografía	124

Lista de Figuras

1. Esquema de relación de los bloques que conforman una interfaz cerebro computadora. Modificado de [1]	22
1.1. Representación esquemática de la disposición columnar de la corteza auditiva, en la que se muestra como se van agrupando hasta conformar una estructura denominada celdilla.	26
1.2. Señal de electroencefalograma de 1 seg. de duración.	27
1.3. Sistema internacional 10-20 para posicionamiento de electrodos en el cuero cabelludo.	27
1.4. Efectos de la promediación coherente sobre la RSR en registros de potenciales evocados.	29
1.5. Registros de ERP en distintas posiciones del sistema 10-20. La onda P300 se manifiesta mejor en los electrodos centrales F_z , C_z y O_z , observándose un cambio de la polaridad en esta última posición.	30
1.6. Deletreador de Donchin. Es necesario aclarar que las filas y columnas no se iluminan simultáneamente, sino que en el instante $t = t_0$ se ilumina la 3ra. fila y posteriormente en el instante $t = t_0 + \Delta t$ se ilumina la 3ra. columna.	34
1.7. Efecto de la promediación coherente en registros de ERP. Nótese que para el caso de época única es imposible a simple vista reconocer la onda P300.	34
1.8. Arquitectura típica de una interfaz cerebro computadora, a los bloques de extracción de características y clasificación hay que considerar previamente el de instrumentación electrónica.	35

1.9. Propuesta de implementación de las etapas de extracción de características y clasificación de una ICC, donde \mathbf{x} : señal de EEG, Φ : diccionario de átomos o matriz de transformación, \mathbf{a} : señal de coeficientes transformados, \mathbf{a}_L : señal resultante de aplicar selección de características, Clasificador: discriminante lineal, $C_{1,2}$: clases (señales <i>con ERP</i> y <i>sin ERP</i>).	36
2.1. Señal de potencial evocado relacionado con eventos obtenida mediante promediación coherente de 100 épocas y posterior filtrado pasa bajos para suavizar la morfología.	41
2.2. Elementos de la base para la TDF de Fourier en \mathbb{R}^{64} para $k = 1, k = 2, k = 3, k = 4$. Los valores discretos se representan con círculos, y si bien se han unido con líneas punteadas, recuérdese que las muestras de las señales de la base son solamente los valores de los círculos marcados.	42
2.3. Ejemplo de la wavelet de Morlet para distintas escalas y localizaciones. Adaptado de [2].	44
2.4. Esquema del funcionamiento de algoritmo de la DDWT. Los coeficientes de aproximación son filtrados nuevamente y se produce un submuestreo del nivel anterior. Extraído de [2].	46
2.5. Algoritmo de transformada wavelet packet.	47
3.1. Pasos involucrados en la selección de características: generación, evaluación, detención y validación.	54
3.2. Cruza simple de 2 individuos de 8 bits, el <i>intercambio de material genético</i> se produce en la posición 5.	57
3.3. Mutación de un bit en la posición 3 en un individuo de 8 bits.	57
4.1. Representación gráfica de la separación entre clases realizada por un LDA.	60
5.1. Electrodo utilizado para la obtención de los registros de EEG [3].	66
5.2. Registro época única <i>single trial</i> de EEG donde se indica los inicios de estimulación representados por líneas verticales con un tiempo inter estímulo de 125 ms.	67

5.3. Montaje para la adquisición de los registros EEG, en el cual se destaca la gorra con electrodos ubicados en posiciones predefinidas del sistema 10-20 así como también la matriz para estimulación visual que posee la plataforma BCI2000.	68
5.4. Conformación del patrón temporal (640 características) que entra al clasificador, a partir de la concatenación de los vectores individuales (64 características) de cada canal.	69
5.5. Matriz de datos formada por los patrones temporales generada por cada sujeto de la base de datos, cada división en las filas corresponde a una época adquirida en cada uno de los 10 canales. . . .	69
5.6. Conformación del patrón wavelet DDWT o WPT (640 características) que entra al clasificador, a partir de la concatenación de los vectores transformados (64 características) de cada canal.	70
5.7. Descripción esquemática de los experimentos, donde se muestra la etapa de entrenamiento y la de prueba.	73
5.8. Ejemplos de distintas funciones Wavelets que se evaluaron.	74
5.9. Escalogramas obtenidos sobre registros single trial, promedios de 5 y de 10 épocas obtenidos en los canales Cz y Oz.	75
5.10. Descomposición multiresolución hasta el nivel 6 de una señal $x[n]$ muestreada a 64 Hz, en el cual se observan los rangos frecuenciales que abarca cada nivel de descomposición. A_i : i -ésima aproximación, D_i : i -ésimo detalle, i : nivel de descomposición	75
5.11. Base generada mediante Transformada Wavelet Diádica Discreta utilizando como Wavelet madre la Daubechies 4.	76
5.12. Señal de ERP en canal Cz y su respectivo espectro de frecuencias.	77
5.13. Índices de desempeño para cada método propuesto. Ref: patrones temporales de referencia, Smo: patrones DDWT con suavizado, Sur: patrones DDWT con denoising por umbral hard SURE, Uni: patrones DDWT con denoising por umbral hard Universal.	79
5.14. Resultados obtenidos al aplicar sobre todos los sujetos el algoritmo <i>local discriminat basis</i> utilizando las siguientes funciones de costo: simétrica (Si), asimétrica (ASi) y la suma de cuadrados (SuC). . . .	80

5.15. Evolución de la potencia de discriminación de cada coeficiente Wavelet Packet, las gráficas representan el promedio de todos los canales sobre todos los sujetos para seleccionar los valores de p . Tomado de [4]	81
5.16. Resultados obtenidos con el algoritmo <i>local discriminant basis</i> utilizando las diferentes medidas de discriminación y para los distintos valores de $p= 4, 6$ y 8 . LDB_s : LDB con medida simétrica, LDB_{as} : LDB con medida asimétrica y LDB_{sc} : LDB con suma de cuadrados.	82
5.17. Índices de desempeño obtenidos con patrones obtenidos con el algoritmo best orthogonal basis (BOB) y con local discriminat basis utilizando 8 coeficientes (LDB8).	83
5.18. Área debajo la curva ROC vs. cantidad de patrones de entrenamiento. De arriba a abajo utilizando 80, 160 y 320 características en los patrones respectivamente. Las curvas en cada gráfica representan patrones formados utilizando diferente cantidad de promediaciones (1 a 5).	85
5.19. Resultados obtenidos con distintas combinaciones de estrategias de ensambles. <i>Vot S</i> : votación simple, <i>Vot P</i> : votación ponderada, <i>Reg P</i> ; regla del producto, <i>Reg S</i> : regla de la suma, <i>Reg SP</i> : regla de suma ponderada, <i>Reg Ma</i> : regla de máximo, <i>Reg Mi</i> : regla del mínimo, <i>Reg Me</i> : regla de mediana.	86
6.1. Resumen de los experimentos realizados: \mathbf{x} son los patrones temporales, \mathbf{a} son los patrones DDWT o WPT, \mathbf{a}_K son los patrones DDWT con suavizado, \mathbf{a}_L y \mathbf{x}_M son los patrones resultantes de la selección de características, $C_{1,2}$ son las clases posibles.	89
6.2. Índices de desempeño obtenidos en cada uno de los sujetos de la base de datos.	90
6.3. Índices de desempeño de referencia obtenidos sobre registros temporales.	90
6.4. Índices de desempeño obtenidos con patrones DDWT	91
6.5. Evolución de los índices de desempeño en función de la cantidad de características de los patrones WPT	92
6.6. Índices de desempeño obtenidos con patrones WPT utilizando 18 coeficientes por canal	94

6.7. Árboles de descomposición multiresolución resultantes de aplicar smothing eliminando el <i>detalle</i> 1 (a) y los <i>detalles</i> 1 y 2 (b).	95
6.8. Índices de desempeño obtenidos con patrones DDWT sin el <i>detalle</i> 1.	95
6.9. Índices de desempeño obtenidos con patrones DDWT sin los <i>detalles</i> 1 y 2.	96
6.10. Evolución de la tasa de aciertos del algoritmo de <i>eliminación recursiva de características</i> a medida que estas se van eliminando.	97
6.11. Índices de desempeño obtenidos con patrones temporales utilizando <i>eliminación recursiva de características</i> .	98
6.12. Índices de desempeño obtenidos con patrones DDWT sin el Detalle 1 utilizando <i>eliminación recursiva de características</i> .	98
6.13. Índices de desempeño obtenidos con patrones DDWT sin los Detalles 1 y 2 utilizando <i>eliminación recursiva de características</i> .	99
6.14. Generación de patrones de 4 características a partir de un individuo de la población.	100
6.15. Índices de desempeño obtenidos con patrones temporales utilizando <i>algoritmos genéticos</i> .	101
6.16. Índices de desempeño obtenidos con patrones DDWT sin el Detalle 1 utilizando <i>algoritmos genéticos</i> .	102
6.17. Índices de desempeño obtenidos con patrones DDWT sin los Detalles 1 y 2 utilizando <i>algoritmos genéticos</i> .	103
7.1. Comparación de índices de desempeños obtenidos con patrones DDWT y WPT	105
7.2. Comparación de índices de desempeños obtenidos con patrones DDWT, DDWT sin D1 y DDWT sin D1 y D2	106
7.3. Elementos de la base de la DDWT seleccionados por el algoritmo genético, aplicado sobre patrones DDWT sin los Detalles 1 y 2, correspondiente al mejor individuo del sujeto 13 de la base de datos.	107
7.4. Comparación de índices de desempeños obtenidos con <i>algoritmos genéticos</i> aplicado a patrones temporales y DDWT sin los Detalles 1 y 2.	108
7.5. Comparación de índices de desempeños obtenidos con <i>eliminación recursiva de características</i> aplicado a patrones temporales y DDWT sin los Detalles 1 y 2.	108

7.6. Representación de la cantidad de características resultantes en función de la tasa de aciertos. Temp: patrones temporales, Temp ERC: patrones temporales mas ERC, Temp AG: patrones temporales mas AG, DDWT: patrones wavelets, DDWT D1: patrones wavelets sin el detalle 1, DDWT D1yD2: patrones wavelets sin los detalles 1 y 2, DDWT D1 ERC: patrones wavelets sin el detalle 1 mas ERC, DDWT D1 AG: patrones wavelets sin el detalle 1 mas AG, DDWT D1yD2 ERC: patrones wavelets sin los detalles 1 y 2 mas ERC, DDWT D1yD2 AG: patrones wavelets sin los detalles 1 y 2 mas AG, WPT: patrones wavelets packets.	110
---	-----

Lista de Tablas

1.1. Fuentes de actividad eléctrica registrada en el cuero cabelludo . . .	28
6.1. Índices de desempeño de patrones WPT con distintas cantidades de características.	93
7.1. Índices de desempeños obtenidos en los experimentos realizados. .	109
7.2. Comparación de resultados con otros investigadores utilizando la tasa de aciertos como índice de desempeño.	114
7.3. Tasas de aciertos obtenidas con la base del Wadsworth Center. . .	115

Resumen

Las vías naturales de comunicación del cerebro con el exterior eventualmente pueden resultar dañadas de manera irreversible como secuela de accidentes o enfermedades, la consecuencia de esto es la disminución de la capacidad de una persona para comunicarse con el exterior y controlar su entorno.

Actualmente existen dispositivos que posibilitan restituir artificialmente las funciones de las vías de comunicación dañadas y de esta manera permitir a una persona enviar mensajes y comandos a una computadora y desde ésta conectarse o interactuar con el mundo exterior, estos dispositivos se denominan *interfaces cerebro computadoras* (ICC).

Uno de los paradigmas o modelos mas utilizados para implementar una interfaz cerebro computadora es el basado en electroencefalograma, dentro del cual se encuentran las basadas en *potenciales evocados relacionados a eventos* (ERP, del inglés Event Related Potentials).

Este paradigma, también llamado *paradigma P300*, consiste en utilizar una matriz de caracteres alfanuméricos arreglados en filas y columnas que se van iluminando aleatoriamente; de manera sincronizada con el inicio de las iluminaciones se registra en el cuero cabelludo la actividad eléctrica conformada por la señal de electroencefalograma y la de ERP. Esta última se encuentra presente solo si se ilumina la fila o la columna en la que está el caracter al cual se le está prestando atención, por lo que el problema principal a resolver es la detección de la señal de ERP y de esta manera identificar que caracter es al que está prestando atención la persona y así lograr la comunicación.

Esto en si mismo representa un problema de clasificación de 2 clases (*registros de EEG sin ERP y registros de EEG con ERP*), que tiene el inconveniente de que la señal de ERP posee una amplitud mucho menor que la señal de electroencefalograma y en consecuencia una pobre relación señal a ruido.

Habitualmente se utilizan técnicas para mejorar la relación señal a ruido que

necesitan sucesivas estimulaciones y por lo tanto insumen tiempo, lo que representa una desventaja a la hora de utilizar una ICC debido a que es deseable poder transmitir la mayor cantidad de información por unidad de tiempo para permitir una comunicación mas fluida de la persona con el entorno. Por este motivo se trabajará con registros únicos, o *single trial*, lo cual dificulta aún mas el problema de detección del ERP.

El objetivo de esta tesis es el *desarrollo, evaluación y selección de una metodología de descomposición óptima de señales para favorecer la detección de potenciales evocados relacionados con eventos, en registros de single trial, utilizando la menor cantidad de dimensiones de la señal.*

Para esto se propone buscar una representación alternativa de la señal de electroencefalograma de manera de resaltar las características significantes de la señal de ERP y de esta manera mejorar su detección y consecuentemente el desempeño de la ICC. En el contexto de espacios de señales, una señal discreta implícitamente está representada en una base canónica, por lo que mediante un cambio de base apropiado es posible resaltar las características deseables de la señal de ERP.

Mas allá que el cambio de base produzca una mejora en el desempeño de la ICC, es posible que no todas las características aporten a la potencia de discriminación de señales *con ERP* y *sin ERP*. Esto permite pensar en disminuir la dimensión de la señal manteniendo, o inclusive mejorando, el desempeño y a la vez propiciar su implementación en hardware.

Se probaron distintas estrategias combinando representación mediante bases ortogonales derivadas de la *Transformada Wavelet Diádica Discreta* (DDWT) y de la *Transformada Wavelet Packet* (WPT), métodos de mejoramiento de la relación señal a ruido en el dominio Wavelet y estrategias de reducción de dimensionalidad basadas en algoritmos selección de características utilizando algoritmos genéticos y eliminación recursiva.

Se utilizó una base de datos con registros de 18 sujetos saludables, como clasificador un discriminante lineal de Fisher y como índices de desempeño la tasa de aciertos, sensibilidad, especificidad y cantidad de características. Como referencia para comparar el desempeño de las estrategias propuestas se tomaron los índices de desempeño correspondientes a los registros temporales disponibles en la base de datos.

Los resultados obtenidos indican que el mejor desempeño en la detección de

ERP se logra representado la señal en una base ortogonal generada a partir de la WPT utilizando el algoritmo *Local Discriminant Basis* (LDB) logrando una tasa de aciertos del 78,5%, lo que representa una mejora de aproximadamente 12,9% con respecto a los índices de referencia. Con la mayoría de las estrategias se obtuvieron índices de desempeños superiores al 70%, lo cual es adecuado para lograr una comunicación con una ICC, inclusive aquellas que combinaban solo señales temporales y algoritmos de reducción de dimensionalidad, lo cual los hace atractivos para implementación en hardware de una interfaz cerebro computadora.

A futuro se propone evaluar estrategias de selección de canales de registros para disminuir aún mas la dimensionalidad, otros clasificadores y estrategias de combinación mediante ensambles, y sobre todo representaciones de la señal basadas en bases no ortogonales.

Abstract

The natural pathways of communication between the brain and the outside world may eventually be damaged irreversibly in the aftermath of accidents or illness, thereby reducing the ability of a person to communicate and control their environment. *Brain Computer Interfaces* (BCIs) are devices that attempt to artificially restore functions to damaged communication pathways and thus allow a person to send messages and commands to a computer and from there to connect or interact with the outside world.

One of the most used paradigms to implement a BCI is based on recording and analyzing the *electroencephalogram* (EEG), within which are those based on *event-related evoked potentials* (ERP). This paradigm, also called *P300 paradigm*, uses an array of alphanumeric characters arranged in rows and columns that are illuminated randomly, while the EEG is recorded synchronously with the illumination onset. The subject is then asked to pay attention to the row or column in which the next character or command they want to communicate is present. In this context, an ERP is present on the EEG only if the attended row or column is illuminated. Thus, the main problem to be solved is the ERP signal detection in order to identify which character or command the person is trying to communicate. This in itself represents a 2-class classification problem (*EEG recordings without ERP* and *EEG recordings with ERP*), which is challenging because the ERP signal has much smaller amplitude than the EEG it is immersed in, resulting in a poor signal to noise ratio.

The techniques that are used to improve the signal to noise ratio require a large number of repetitions or trials and therefore are time-consuming. This is a disadvantage when using a BCI because it is desirable to transmit as much information per time unit as possible to allow faster and better communication of the person with the environment. For this reason, the present work is mainly based on the analysis of unique records or single trials, which complicate the ERP

problem detection.

The aim of this thesis is *to develop, evaluate and select a methodology for optimal signal decomposition to improve the detection of ERPs in single trial records, using the smallest possible signal dimension*. To this end, an alternative EEG signal representation is proposed in order to highlight the significant features of the ERP signal and thus improve their detection and consequently the performance of the BCI. In the context of signal spaces, a discrete signal implicitly is displayed in a canonical basis, and by changing to an appropriate base is possible to highlight the desirable characteristics of the ERP signal.

Even though changing the base may result in an improvement in the performance of the BCI, it is possible that not all features in the new base contribute to the discriminative power of signals with and without ERP. This implies that reducing the dimension of the base might maintain or even improve the performance of the BCI and simultaneously facilitate their implementation in hardware signal.

In this work, a database with records of 18 healthy subjects was used to test different representation strategies based on orthogonal bases derived of *discrete dyadic wavelet transform* (DDWT) and *wavelet packet transform* (WPT). Furthermore, several methods to improve the signal to noise ratio in the wavelet domain were evaluated and a number of strategies for dimensionality reduction based on feature selection algorithms using genetic algorithms and recursive features elimination were tested as well. Fisher's linear discriminant analysis was chosen as classifier, and sensitivity, specificity, correct hit rate and number of features used were considered as performance indexes. Classification performance rates for unprocessed single trials in the time domain were used as a reference for comparison.

Results showed that the best performance in detecting ERP was achieved by representing the EEG signal on an orthogonal basis from the WPT algorithm using *local discriminant basis* (LDB) achieving a success rate of 78.5%, which was an improvement of about 12.9% compared to the reference indexes. Furthermore, most of the strategies resulted in performance indexes over 70%, which is suitable to achieve communication with the BCI, including those that combined only single trial signals in the time domain and dimensionality reduction algorithms, which makes them attractive for implementation in hardware BCI. Future work intends to evaluate channel selection to further reduce the dimensio-

nality, alternative classifiers and combination strategies through assemblies, and especially representations of the signal based on non-orthogonal bases.

Introducción

Las vías naturales de comunicación del cerebro con el exterior, tales como nervios periféricos y músculos, eventualmente pueden resultar dañadas de manera irreversible como secuela de accidentes, traumas o enfermedades tales como esclerosis lateral amiotrófica, esclerosis múltiple, distrofia muscular, accidentes cerebro vasculares, o tumores entre otras; lo cual disminuye la capacidad de una persona para comunicarse con el exterior y controlar su entorno. En casos extremos, la pérdida del control de toda el sistema muscular es total, incluidos el control ocular; este estado se denomina *síndrome de enclaustramiento* (*locked-in* en inglés) en el cual la persona pierde toda capacidad de comunicación o interacción con el medio a pesar de tener sus capacidades cognitivas inalteradas.

Actualmente existen dispositivos que tienen por objetivo restituir artificialmente las funciones de las vías de comunicación dañadas, y permitir a las personas enviar mensajes y comandos a una computadora y desde ésta conectarse o interactuar con el mundo exterior, estos se denominan *interfaces cerebro computadoras* (ICC).

Para registrar o detectar la actividad cerebral con el objetivo de implementar una ICC se pueden utilizar distintos tipos de tecnología que incluyen la magnetoencefalografía, tomografía por emisión de positrones, imágenes de resonancia magnética funcional, electrocorticografía y la electroencefalografía. De estas, sólo la última se basa en equipos relativamente simples y económicos y permite trabajar sin instalaciones especiales, por este motivo es la técnica que más se utiliza en investigación y desarrollo de nuevos sistemas de comunicación y control no muscular.

Las ICC basadas en señales de electroencefalograma representan un grupo de modelos o paradigmas, dentro de los cuales se encuentra el basado en *potenciales evocados relacionados a eventos* (ERP, del inglés Event Related Potentials) que se utilizará en esta tesis. Independientemente de los distintos paradigmas, la es-

estructura de una ICC se puede representar mediante bloques interconectados tal como se muestra en la Figura 1, entre los cuales se encuentra el de instrumentación o adquisición del electroencefalograma, el de extracción de características y el de clasificación. Se puede mejorar el desempeño de una ICC mejorando el desempeño de cualquiera de estos bloques, en particular en esta tesis se trabajará en el de extracción de características y en menor medida en el de clasificación.

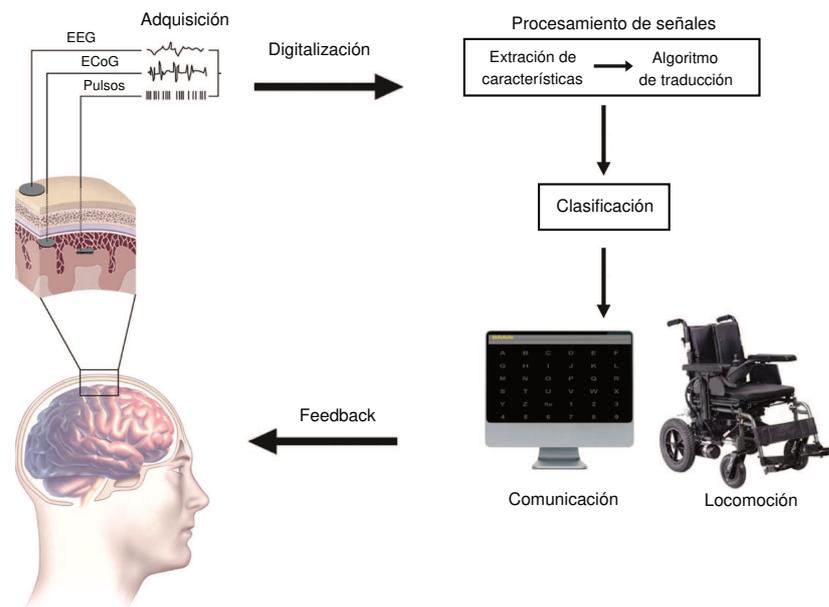


Figura 1: Esquema de relación de los bloques que conforman una interfaz cerebro computadora. Modificado de [1]

El paradigma basado en ERP, también llamado *paradigma P300*, consiste en utilizar una matriz de caracteres alfanuméricos arreglados en filas y columnas, las cuales se van iluminando aleatoriamente. Simultáneamente, y de manera sincronizada con el inicio de las iluminaciones, se registra en el cuero cabelludo la actividad eléctrica conformada por la señal de *electroencefalograma* (EEG) y la de ERP. Esta última se encuentra presente solo si se ilumina la fila o la columna en la que está el carácter al cual se le está prestando atención, por lo que el problema principal a resolver es la detección de la señal de ERP y de esta manera identificar que carácter es al que está prestando atención la persona y así lograr la comunicación. En este contexto, la detección de los ERP se convierte en un problema de clasificación con 2 clases posibles: *registros de EEG sin ERP* y *registros de EEG con ERP*.

La señal de ERP presenta la particularidad de tener una amplitud mucho menor que la señal de EEG, dado que se registran simultáneamente esto provoca que tenga una relación señal a ruido muy baja. Habitualmente se utilizan técnicas para mejorarla, como por ejemplo la promediación coherente, que presentan como ventaja poder tener una mejor estimación de la señal de ERP pero por otro lado requiere de tiempo para su aplicación. Esto representa una desventaja a la hora de utilizar una ICC, ya que es deseable poder transmitir la mayor cantidad de instrucciones o *información* por unidad de tiempo para permitir una comunicación mas fluida de la persona con el entorno. Por este motivo, en esta tesis se trabajará con registros únicos, denominados de una sola época o *single trial* lo cual dificulta el problema de detección mencionado previamente.

En esta tesis se propone buscar una representación alternativa de la señal de EEG registrada, de manera de resaltar las características significativas la señal de ERP y en consecuencia mejorar su detección. En el contexto de espacios de señales, una señal discreta implícitamente está representada en una base canónica conformada por señales delta de Dirac desplazadas. Por lo tanto, mediante un cambio de base apropiado se pueden resaltar las características deseables de la señal de ERP; en este sentido se propone utilizar bases generadas mediante Transformada Wavelet Diádica Discreta y Transformada Wavelets Packets.

A esta altura, se puede considerar a la señal registrada (o su versión transformada) como un patrón de entrada en un sistema de clasificación o de reconocimiento de patrones. Mas allá que una representación en función de otra base produzca una mejora en el desempeño de la ICC, es probable que no todas las características aporten a la potencia de discriminación de señales con ERP y sin ERP. En consecuencia es posible disminuir la dimensión del patrón de entrada manteniendo (o inclusive mejorando) el desempeño de la ICC, lo que propicia la implementación de una ICC en hardware, por lo que se propone utilizar métodos de selección de características para este fin.

Considerando lo mencionado previamente, se propone como objetivo de esta tesis el *Desarrollo, evaluación y selección de una metodología de descomposición óptima de señales para favorecer la detección de potenciales evocados relacionados con eventos, en registros de época única, utilizando la menor cantidad de dimensiones de la señal*. Para esto se propone desarrollar procedimientos y/o algoritmos basados en la teoría de Wavelets que favorezcan la extracción de características significativas de la actividad eléctrica registrada para proceder a la detección de

la señal de ERP de manera robusta y eficiente; así como también algoritmos de selección de características para reducir su dimensionalidad.

Este documento está organizado de la siguiente manera: en el Capítulo 1 se describe que es una ICC, el problema abordado en esta tesis así como también el planteo conceptual de la solución propuesta. En los Capítulos 2, 3 y 4 se presentan los fundamentos teóricos relacionados con los distintos bloques que forman una ICC, es decir, extracción de características mediante representación en bases ortogonales, selección de características y clasificación respectivamente. En el Capítulo 5 se describen los materiales y métodos así como también algunos experimentos preliminares que permiten definir la metodología de los experimentos definitivos que se describen en el Capítulo 6 junto con los resultados obtenidos. Finalmente en los Capítulos 7 y 8 se exponen la discusión de los resultados y las conclusiones a las que se arribó.

Capítulo 1

Descripción del problema

Este capítulo tiene como propósito dar a conocer la motivación para realizar esta tesis y el contexto en el cual se encuadra el desarrollo de la misma. Para esto se exponen algunos conceptos básicos de electrofisiología, se define que es una interfaz cerebro computadora y finalmente se presenta el planteo conceptual de la solución propuesta a problema abordado.

1.1. Conceptos de electrofisiología

1.1.1. Electroencefalograma

En el año 1848 el fisiólogo Du Bois Rédmond fue la primera persona en observar la aparición de actividad eléctrica al paso de un impulso nervioso en nervios periféricos. Luego, en 1875, R. Caton describió el mismo fenómeno a nivel cerebral efectuando registros en animales. Finalmente en 1929 Hans Berger fue el primero en registrar actividad eléctrica cerebral en seres humanos, y fue quien en realidad acuñó el término *electroencefalograma*. Además describió sus características tal como se interpretan actualmente, en especial las variaciones rítmicas asociadas a diferentes estados de atención.

En la actualidad el EEG es una técnica no invasiva de registro de la actividad eléctrica generada en la corteza cerebral, y particularmente el EEG de superficie es la técnica clínica mas empleada en la valoración del funcionamiento cerebral.

La corteza cerebral está organizada en seis láminas horizontales que se orientan de manera paralela a la superficie y se indican con números romanos del I al VI. Sin embargo, las unidades de procesamiento de la información básicas de

la corteza cerebral son columnas verticales definidas fisiológicamente, que se extienden desde la superficie pial hasta la sustancia blanca. No son columnas en el sentido arquitectónico, sino más bien largas barras tridimensionales de $50 \mu m$ de ancho y de largo variable, que a su vez se agrupan para formar una estructura mas grande denominada celdilla tal como se muestra en la Figura 1.1 [5]. El aumento de la superficie de la corteza permite un mayor número de columnas y por lo tanto brinda una mayor capacidad de procesamiento de la información.

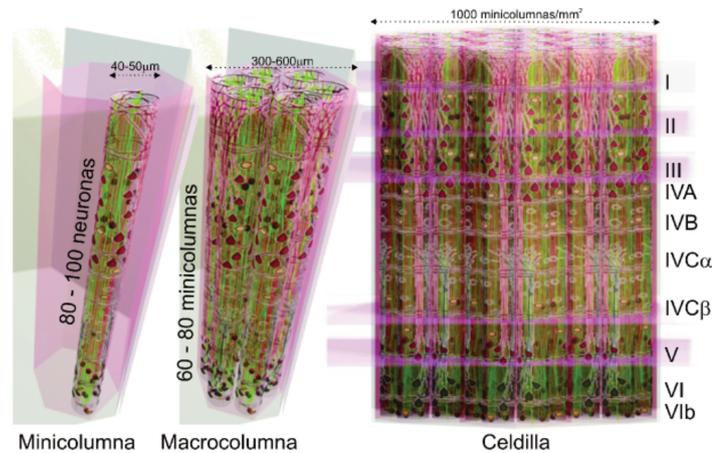


Figura 1.1: Representación esquemática de la disposición columnar de la corteza auditiva, en la que se muestra como se van agrupando hasta conformar una estructura denominada celdilla.

El EEG es el registro de los potenciales post sinápticos de las neuronas ubicadas en la corteza cerebral, que es consecuencia de la organización laminar de esta última y de la activación temporal sincronizadas de grupos o columnas de neuronas [6]. La amplitud del EEG oscila entre los 10 y $200 \mu V$, aunque puede variar en función de diversos parámetros de registro o de las maniobras de activación para favorecer la aparición de una respuesta específica; en la Figura 1.2 se muestra un registro de electroencefalograma de 1 seg de duración.

El espectro de frecuencias del EEG no presenta componentes más allá de los 50 Hz y se definen bandas específicas, que se manifiestan mediante maniobras de activación, que se correlacionan con algunos eventos clínicos [6].

El registro de EEG se realiza mediante electrodos ubicados sobre el cuero cabelludo en posiciones definidas por el sistema internacional 10-20, el cual fue adoptado en el año 1949 por la Federación Internacional de Sociedades de Electroencefalografía y Neurofisiología; cuyo objetivo es estandarizar la ubicación de

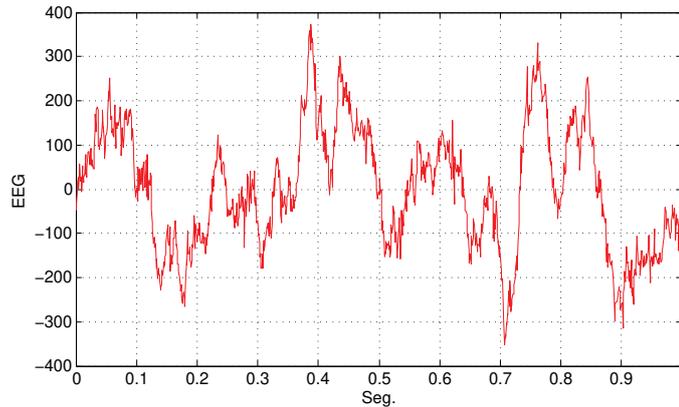


Figura 1.2: Señal de electroencefalograma de 1 seg. de duración.

los electrodos en posiciones normalizadas tal como se muestra en la Figura 1.3. Se observa que los nombres de las posiciones dependen del hemisferio cerebral sobre el cual están posicionados, además los subíndices del lado izquierdo son impares mientras que los del lado derecho son pares y la numeración aumenta a medida que se alejan de la línea media central.

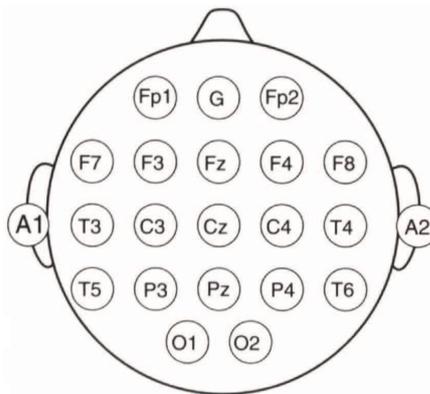


Figura 1.3: Sistema internacional 10-20 para posicionamiento de electrodos en el cuero cabelludo.

1.1.2. Potenciales evocados relacionados con eventos

Los *potenciales evocados* (PE) son señales eléctricas, registradas mediante electrodos en el cuero cabelludo, que se producen como respuesta de los sistemas sensoriales a la aplicación de un estímulo adecuado. Este puede ser acústico, visual o eléctrico y producirá la respuesta del sistema sensorial correspondiente y en función de esto será la modalidad del potencial evocado que se produzca [7].

La actividad eléctrica registrada simultáneamente a la aplicación del estímulo se denomina *época* y está conformada por la señal de interés (PE) y actividad eléctrica no correlacionada con el estímulo denominado *ruido*. Este último tiene distintas componentes, las cuales se resumen en la Tabla [1.1](#).

Tabla 1.1: Fuentes de actividad eléctrica registrada en el cuero cabelludo

Fisiológicas	No fisiológicas
PE (señal de interés)	Potenciales electromagnéticos
EEG	señales de radio
EMG	tensión de línea (50 Hz)
ECG	artefacto de estímulo
EOG	Ruido de instrumentación

El registro de la señal de PE presenta el inconveniente de tener una *relación señal a ruido* (RSR) muy baja; dado que el ruido tiene una amplitud aproximadamente 20 veces mayor que el PE. Se podría pensar que este inconveniente puede ser resuelto con un proceso de filtrado tradicional lineal, sin embargo, el hecho que los espectros de frecuencias del ruido y del PE están superpuestos hacen esto imposible [\[8\]](#).

La técnica mas usada para separar ambas señales es la *promediación coherente*, la cual consiste en registrar la actividad eléctrica en el cuero cabelludo en forma sincronizada con el estímulo aplicado. Para llevar adelante este proceso se asumen 3 premisas básicas como válidas:

- a) la señal de PE y el ruido se suman linealmente.
- b) la señal de PE es estable.
- c) el ruido se supone del tipo blanco, estacionario y con valor medio cero.

Considerando esto, durante el proceso de promediación la señal de PE permanece constante mientras que el ruido va disminuyendo su varianza dado que su morfología es diferente en cada época [\[8\]](#).

La promediación mejora la RSR inicial de acuerdo a la siguiente ecuación:

$$RSR_p = RSR_i \cdot N^{\frac{1}{2}} \quad (1.1)$$

donde RSR_p es la relación señal a ruido después de la promediación, RSR_i es la relación señal a ruido inicial y N es la cantidad de épocas promediadas. En la Figura [1.4](#) se muestra como mejora la RSR de un registro de potenciales evocados mediante promediación coherente, en la primera fila se presenta el caso

de un registro de época única o *single trial* que consiste en la actividad registrada aplicando un solo estímulo. En las filas de abajo se presentan los caso de promediaciones de 4 y 16 épocas respectivamente, se muestra claramente que la amplitud de la señal de PE se mantiene constante mientras que el ruido disminuye su varianza en forma proporcional a la cantidad de épocas promediadas.

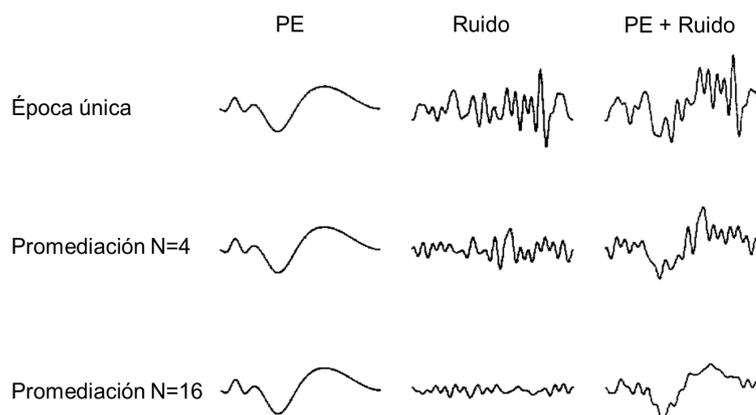


Figura 1.4: Efectos de la promediación coherente sobre la RSR en registros de potenciales evocados.

Mas allá de la modalidad (visuales, auditivos o somatosensoriales) los potenciales evocados se clasifican en función del tiempo de aparición después de aplicado un estímulo; aquellos que aparecen dentro de los primeros 10 ms se denominan de *latencia corta*, los que aparecen entre los 10 ms y 60 ms se denominan de *latencia media* y los que están mas allá de los 60 ms se denominan de *latencia larga*. Estos últimos también se denominan *potenciales evocados relacionados a eventos* (ERP, del inglés event related potentials), dado que la aparición de las ondas que lo conforman depende fuertemente del estado de atención de la persona [7].

En la Figura 1.5 se muestra la morfología de ERP, se puede observar la variabilidad entre registros según la posición en la cual se registre. La onda mas importante de los ERP es la denominada P300, generalmente aparece alrededor de los 300 *mseg* luego de aplicado el estímulo habiendo realizado alguna actividad cognitiva. Esta es de origen endógeno y se caracteriza por tener polaridad positiva en la mayoría de las posiciones de registro, esta situación se muestra claramente en la figura donde la polaridad es negativa solo en la posición Oz. La amplitud de la onda P300 puede llegar a los 5 μV , siendo la misma inversamente proporcional a la frecuencia de aplicación del estímulo. Las características principales de esta

onda son la amplitud, latencia y duración; los cuales son influenciados por diversas causas tales como la frecuencia de estimulación, el estado de atención de la persona, entre las mas importantes [9].

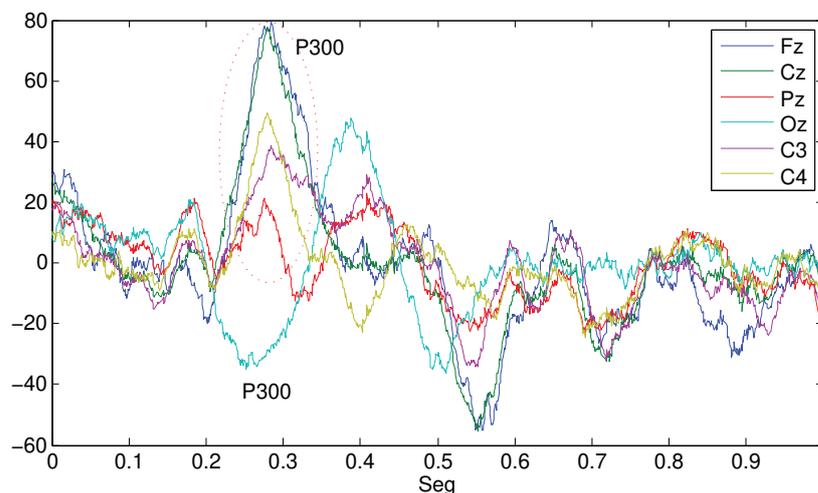


Figura 1.5: Registros de ERP en distintas posiciones del sistema 10-20. La onda P300 se manifiesta mejor en los electrodos centrales F_z , C_z y O_z , observándose un cambio de la polaridad en esta última posición.

1.2. Interfaces cerebro computadoras

En la actualidad, la definición internacionalmente aceptada de ICC es la siguiente: "Una interfaz cerebro computadora es un sistema de comunicación que no depende de las vías de salida normales del cerebro, como nervios periféricos y músculos" [10]. Al tratarse de una tecnología reciente y compleja son evidentes los avances realizados y el aumento en el número de grupos de investigación dedicados a este tema durante las últimas décadas. Sin embargo, hasta hoy persisten algunas limitaciones que se intentan superar, la cantidad de información transmitida por unidad de tiempo es una de ellas.

Independientemente del paradigma de ICC, hay dos instancias claramente definidas en el funcionamiento de la misma: una es la etapa de entrenamiento, el cual es un proceso *fuera de línea* en el que se le *enseña* a la ICC a resolver el problema; y la otra es la etapa de funcionamiento propiamente dicha en la cual se requiere la interpretación de las señales en tiempo real.

En la actualidad hay varias taxonomías o tipos de ICCs, según como se inducen las señales cerebrales pueden ser *dependientes* o *independientes*, según la naturaleza de las señales cerebrales pueden ser *endógenas* o *exógenas*, según el modo de procesamiento de los datos pueden ser *síncrona* o *asíncrona* y según la modalidad de adquisición de las señales cerebrales pueden ser *invasivas* o *no invasivas* [11, 12].

En el caso de las ICC invasivas se utilizan registros de electrocorticograma utilizando electrodos implantables, la colocación de estos requiere una intervención quirúrgica y además poseen vida útil limitada. Entre las ventajas que ofrecen este tipo de ICCs es que mejoran la relación señal a ruido obteniéndose señales del orden de los milivolts [13, 14].

Entre las ICC no invasivas se encuentran las basadas en EEG, estas funcionalmente se conforman por una etapa de *codificación* de la acción o comando de control, una de *decodificación* del comando mediante algoritmos adecuados y una de *realimentación* en tiempo real del resultado de la acción tomada [15]. A continuación se describen cada una de estas.

- a) *Codificación*: en el contexto de una ICC, la idea de *control voluntario* a través de señales de EEG es posible. Por ejemplo, si un individuo realiza un movimiento voluntario se generan en el EEG los mismos patrones que si el individuo sólo piensa en realizar ese movimiento [16]; sin embargo, estas señales no son útiles porque la información presente no es suficiente para realizar la decodificación. Otra alternativa es intentar asignar un estado mental a una determinada tarea, en este caso el usuario de la ICC debe realizar la tarea mental correspondiente a *codificar* el comando de control deseado, ya sea por regulación del EEG (por ejemplo en una ICC basada en ritmos μ o β) o por variación voluntaria de la atención (como en el caso de una ICC basada en ERP). La codificación es un proceso que depende del entrenamiento o atención del usuario, según corresponda, los usuarios más entrenados generan señales más fuertes lo cual ayuda al mejor desempeño del sistema.
- b) *Decodificación*: para *traducir* las señales de EEG en comandos de control es necesario la *adquisición* y su posterior *procesamiento digital* con el fin de mejorar la RSR de las señales registradas y *extraer características* significativas para la etapa de *clasificación*.

En general, la etapa de clasificación de una ICC se implementa mediante un

clasificador adecuado y optimizado (entrenado) para que realice la separación de señales (patrones) de la mejor manera posible en las clases correspondientes. Este clasificador debe ser entrenado con un grupo de patrones y luego evaluado con otros que no se utilizaron en el entrenamiento. Varios clasificadores han sido estudiados para esta aplicación [17], los más destacados son el *discriminante lineal* (ver sección 4.2) y las *máquinas de soporte vectorial*. Algunos autores sugieren que como el cerebro es un sistema adaptativo y dinámico, por esta razón un clasificador lineal con baja complejidad es probable que tenga mejor capacidad de generalización y sea más estable que los no lineales como por ejemplo una red neuronal multicapa [18]. Al emplear clasificadores de este tipo se está suponiendo que las clases son linealmente separables en algún espacio *n-dimensional* de representación.

- c) *Realimentación*: puede pensarse que existen dos enlaces entre el usuario y el dispositivo externo a controlar: uno está determinado por el bloque mencionado anteriormente y es un vínculo físico, y el otro lo constituye la realimentación, que es percibida por una vía sensorial del usuario, ya sea visual, auditiva o táctil. Esta realimentación ayuda a ajustar la actividad cerebral del usuario para mejorar la salida de la ICC, como sucede en la actividad voluntaria realizada naturalmente.

Se pueden clasificar las ICC basadas en EEG en base a las causas que generan los patrones de actividad eléctrica neuronal. Existe un tipo de ICC que no requiere estimulación externa (asíncrona) para generar los patrones eléctricos, tal es el caso de las ICC basadas en modulación voluntaria por parte de la persona de la energía en la banda α o la imaginación motora basado en bandas μ y β , estas son ejemplos de las ICC endógenas. Las ICC exógenas son aquellos que necesitan de estimulación externa para producir una serie de respuestas fisiológicas que pueden ser moduladas de forma voluntaria por la persona mediante tareas cognitivas. Un ejemplo es la ICC basada en ERP [19], en el cual la persona puede modular voluntariamente la aparición de una onda de la señal de ERP denominada P300 [12], este modelo es de particular interés ya que es el utilizado en esta tesis y se describe en detalle a continuación.

1.2.1. Interfaces cerebro computadoras basadas en ERP

Anteriormente se describió el proceso de promediación coherente como técnica habitual para estimar la señal de PE, sin embargo hay situaciones en las que no se necesita estimar la morfología de la señal y solo basta detectar la señal de PE en la actividad eléctrica registrada; este es el caso de las ICC basadas en ERP [20]. En este caso es deseable realizar la detección de la señal de ERP utilizando una cantidad reducida de épocas, de ser posible a partir de una época única (*single trial* de ahora en adelante), lo que actualmente es uno de los focos principales en investigación de potenciales evocados; ya que esto incide directamente en la tasa de transferencia de información que se logra entre la persona y la computadora. Cuanto menos épocas se utilicen se podrán tomar decisiones más rápidas y en consecuencia enviar mas comandos por unidad de tiempo. En este sentido diversas técnicas modernas de procesamiento de señales han sido evaluadas, incluyendo limpieza de ruido con wavelets [21, 22], filtrado Kalman [23], análisis de componentes independientes [24], descomposición modal empírica [25] y redes neuronales [26] entre otras.

La primera ICC basada en ERP fue el denominado deletreador de Donchin [20], este utilizaba el *paradigma oddball* que consiste en la presentación aleatoria de un evento o estímulo *frecuente* y otro *infrecuente*; cuando aparece un estímulo infrecuente se genera una onda P300. En este contexto un estímulo consiste en la iluminación aleatoria de las filas o columnas de una matriz en la que se representan las letras del abecedario. En la Figura 1.6 se muestra la matriz de 6x6 caracteres alfanuméricos utilizada por este tipo de ICC, cuenta con 36 caracteres representando las 26 letras del abecedario y comandos específicos de una sola palabra arreglados en filas y columnas. El objetivo es que el usuario *comunique* la selección de un caracter, para esto debe prestar atención al mismo, cuando el caracter elegido es iluminado (la fila o la columna en la que se encuentra), y suponiendo que el usuario de la ICC se encuentra concentrado adecuadamente, se genera un ERP. Existen doce eventos posibles en un barrido completo de la matriz, la intensificación luminosa de las 6 columnas y las 6 filas, de los cuales diez son frecuentes y dos infrecuentes, estos últimos corresponden a la fila y la columna del caracter escogido. En cada estímulo se registran simultáneamente señales de EEG en distintas posiciones del cuero cabelludo, así se obtiene un conjunto de registros formados por dos épocas con ERP y diez épocas sin ERP. El objetivo de la ICC es identificar cual fue la fila y la columna que generaron ERP

correspondientes a los estímulos infrecuentes y así determinar el caracter elegido por la persona.

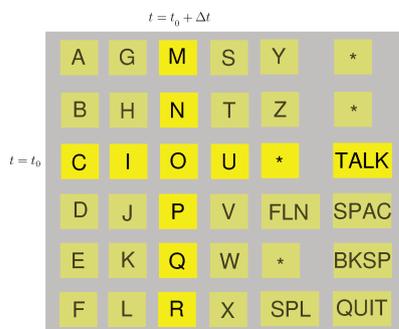


Figura 1.6: Deletreador de Donchin. Es necesario aclarar que las filas y columnas no se iluminan simultáneamente, sino que en el instante $t = t_0$ se ilumina la 3ra. fila y posteriormente en el instante $t = t_0 + \Delta t$ se ilumina la 3ra. columna.

Si bien el procedimiento descrito para determinar el caracter seleccionado mediante el deletreador es sencillo, hay que recordar el inconveniente que presentan las señales de ERP debido a la pobre RSR de la señal registrada. En la Figura 1.7 se muestra claramente la dificultad del problema abordado en esta tesis, dado que para el caso de promedios de 50 épocas se puede observar nítidamente la onda P300 mientras que va desapareciendo a medida que disminuyen la cantidad de épocas promediadas hasta hacerse irreconocible para el caso de *single trial*.

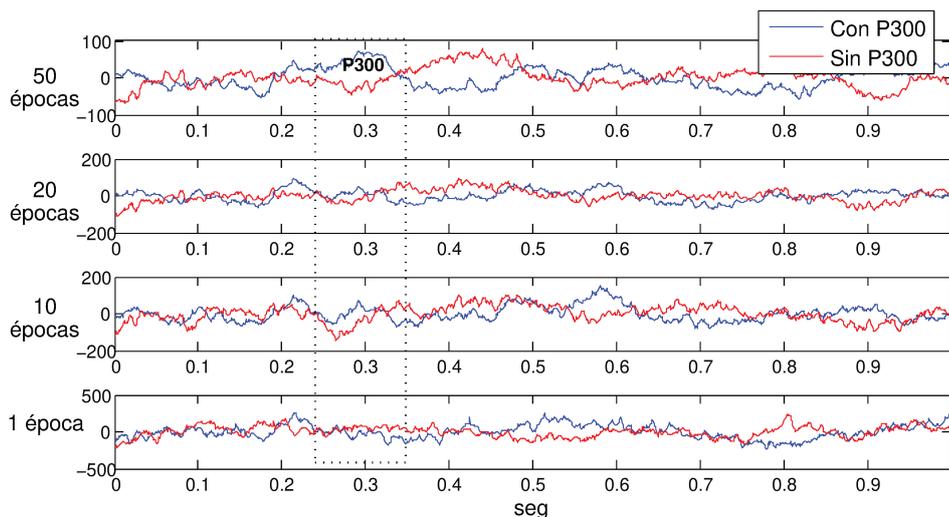


Figura 1.7: Efecto de la promediación coherente en registros de ERP. Nótese que para el caso de época única es imposible a simple vista reconocer la onda P300.

1.3. Planteo conceptual de la solución propuesta

Recordando que el objetivo planteado en esta tesis es *desarrollar, evaluar y seleccionar una metodología de descomposición óptima de señales para favorecer la detección de potenciales evocados relacionados con eventos en interfaces cerebro computadoras*, se propone abordar el problema en el contexto de área de *decodificación* de una ICC descrita anteriormente.

La arquitectura de una ICC está conformada por una serie de bloques entre los que se encuentra el de instrumentación electrónica para el registro de las señales, el de extracción de características y el de clasificación. En la Figura 1.8 se muestran estos 2 últimos que son los de interés en esta tesis y que representan un sistema de reconocimiento de patrones.

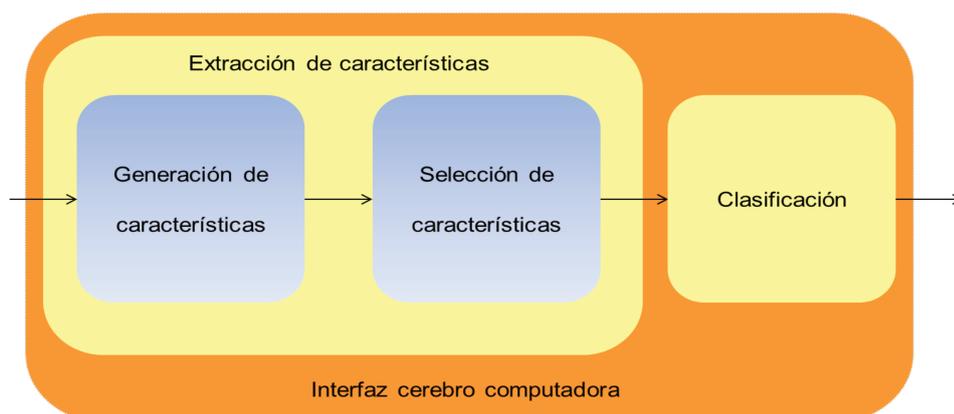


Figura 1.8: Arquitectura típica de una interfaz cerebro computadora, a los bloques de extracción de características y clasificación hay que considerar previamente el de instrumentación electrónica.

Considerando el objetivo planteado, en la Figura 1.9 se muestra la propuesta de implementación de las etapas de extracción de características y clasificación de una interfaz cerebro computadora. La etapa de generación de características se realizará mediante transformaciones de la señal registrada utilizando bases ortogonales. Además, se evaluarán algoritmos de selección de características dado que es de interés realizar la implementación futura en *tiempo real* de la ICC mediante microprocesadores, FPGAs u otra alternativa disponible; por lo tanto es relevante tratar de disminuir la cantidad de dimensiones de la señal o características. Si bien cada una de las etapas se desarrollarán en profundidad a lo largo de los

Capítulos 2, 3 y 4; a continuación se hará una breve descripción de cada una de ellas.

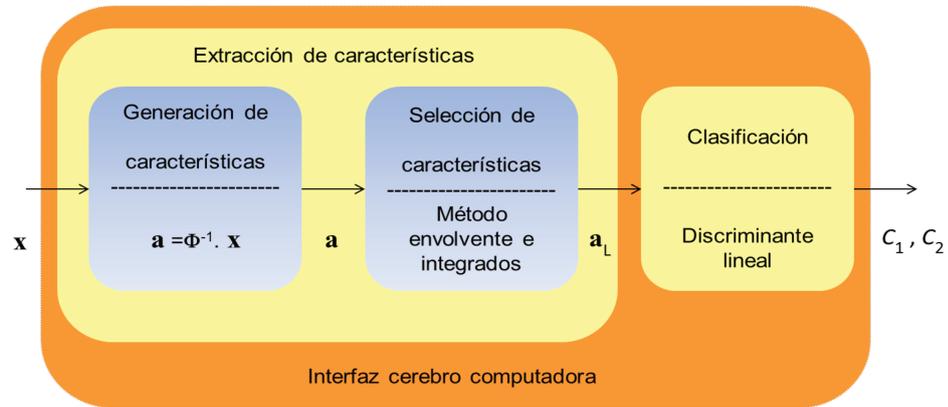


Figura 1.9: Propuesta de implementación de las etapas de extracción de características y clasificación de una ICC, donde \mathbf{x} : señal de EEG, Φ : diccionario de átomos o matriz de transformación, \mathbf{a} : señal de coeficientes transformados, \mathbf{a}_L : señal resultante de aplicar selección de características, Clasificador: discriminante lineal, $C_{1,2}$: clases (señales *con ERP* y *sin ERP*).

1.3.1. Generación de características

En el contexto de esta tesis la frase *representación óptima* hace referencia a tres propiedades deseables para la representación de la señal propuesta:

- que permita separar lo mejor posible las dos clases involucradas en el problema planteado (*registros con ERP* y *registros sin ERP*).
- que codifique características significativas de la señal de interés.
- que la cantidad de coeficientes con valores diferentes de cero sea la menor posible.

La cuestión acerca de la búsqueda de una base óptima y su correspondiente representación ha sido abordada por diferentes autores, sin embargo la mayoría de los métodos desarrollados no tiene en cuenta la información asociada a las clases para un problema de clasificación de señales, dando lugar a representaciones subóptimas para este tipo de aplicaciones [2]. A continuación se formaliza el planteo del problema acerca de cómo encontrar una representación de una señal, comenzando con el caso más general de representación de una señal mediante un diccionario de señales.

Sea $\mathbf{x} \in \mathbb{R}^N$ una señal de dimensión N a la cual se la quiere representar mediante una combinación lineal de señales que conforman una base Φ , de tamaño $N \times M$, y un conjunto de coeficientes $\mathbf{a} \in \mathbb{R}^M$. Entonces, la expresión que describe a la señal es la siguiente:

$$\mathbf{x} = \sum_{\gamma \in \Gamma} \vec{\phi}_{\gamma} a_{\gamma} + \varepsilon = \Phi \mathbf{a} + \varepsilon, \quad (1.2)$$

donde $\varepsilon \in \mathbb{R}^N$ constituye el término de ruido aditivo. La base Φ consiste en una colección de formas de onda o funciones parametrizadas $(\vec{\phi}_{\gamma})_{\gamma \in \Gamma}$, donde cada forma de onda $\vec{\phi}_{\gamma}$ constituye un *átomo*. Aunque la apariencia de la Ecuación (1.2) resulta sencilla, el principal problema consiste en que para el caso más general Φ , \mathbf{a} y ε son desconocidos, existiendo infinitas soluciones. Aún en el caso sin ruido ($\varepsilon = \mathbf{0}$) y conociendo Φ de antemano, si los átomos son más que la cantidad de muestras de \mathbf{x} o si no forman una base, esto produce representaciones no únicas de la señal. Por lo tanto se debe encontrar un criterio que permita seleccionar alguna de ellas; en este caso, y a pesar de que la ecuación de síntesis es lineal, los coeficientes que se eligen para formar parte de la solución resultan en general de una función no lineal de los datos \mathbf{x} .

Para el caso completo ($N = M$) y sin ruido, la relación entre los datos y los coeficientes resulta lineal y está dada por Φ^{-1} ; en el caso de transformaciones unitarias, como por ejemplo la *Transformada Discreta de Fourier*, esta inversión se simplifica debido a que $\Phi^{-1} = \Phi^T$.

En base a lo mencionado hasta aquí, un criterio de interés para seleccionar una representación, de entre todas las posibles, consiste en que ésta sea lo más rala posible; esto significa que se espera que unos pocos coeficientes a_{γ} en la Ecuación (1.2), sean diferentes de cero [2].

Definición 1 Sea $\mathcal{C}(\mathbf{a}, \Phi|\mathbf{x})$ una función criterio con valores escalares que da cuenta de la capacidad de aproximación, de la dispersión de los coeficientes \mathbf{a} y de la discriminación entre clases C_i . Entonces, a partir de la Ecuación (1.2), es posible definir el problema de la representación de \mathbf{x} con respecto a $\mathcal{C}(\mathbf{a}, \Phi|\mathbf{x})$ como:

$$\left[\hat{\mathbf{a}}, \hat{\Phi} \right] = \arg \max_{\mathbf{a}, \Phi} \mathcal{C}(\mathbf{a}, \Phi|\mathbf{x}) \quad \text{sujeto a} \quad \Phi \mathbf{a} + \varepsilon = \mathbf{x}. \quad (1.3)$$

El problema planteado en la Definición 1 en lenguaje cotidiano podría plan-

tearse de la siguiente forma: *Dado un diccionario ¿cómo encontrar la menor cantidad de coeficientes que representen mejor a la señal original, destacando características que favorezcan la clasificación de las señales?*

A este problema se lo denomina de *inferencia* [27], y es posible también incluir restricciones adicionales a fin de disminuir la cantidad de soluciones posibles del problema, que resulten a su vez útiles para en la aplicación final. Esto implica concebir un método que permita encontrar una solución adecuada considerando la formulación original y a las nuevas restricciones planteadas.

Formalmente el problema de inferencia, siempre a partir de la Ecuación (1.2), se puede expresar como una maximización (o minimización según corresponda) de una función objetivo \mathcal{I} que incluya el o los criterios considerados:

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a}} \{ \mathcal{I}(\mathbf{a} | \mathbf{x}, \Phi) \}, \quad (1.4)$$

donde \mathbf{x} y Φ están dados de antemano.

Existen diferentes enfoques para plantear la función \mathcal{I} de manera de obtener una solución a la Ecuación (1.4). Desde el punto de vista determinístico las restricciones aparecen en forma de la minimización de ciertas medidas, distancias o funciones de costo y la solución en forma de un problema de optimización o mediante teoría de regularización [2]. Con el enfoque estadístico las restricciones aparecen sobre el tipo de funciones de densidad de probabilidad de los coeficientes y el ruido, y la solución se obtiene nuevamente como un problema de optimización, maximizando alguna *verosimilitud* o *probabilidad posterior*. Para un análisis más detallado de las conexiones entre el enfoque probabilista y el determinista puede consultarse [28].

Las mismas medidas empleadas en la búsqueda de una solución permiten también valorar la eficacia de la representación lograda, es decir la *bondad* de la codificación de los datos \mathbf{x} en términos de los coeficientes \mathbf{a} , mediante la Ecuación (1.2). Para esta tesis Φ se restringirá al caso de colecciones de funciones que, o bien constituyan una base ortogonal, o correspondan a una familia de bases ortogonales, particularmente bases basadas en Wavelets.

1.3.2. Selección de características

Si bien la generación de características podría verse como una etapa de procesamiento de la señal, y en ocasiones es suficiente para resolver determinadas

aplicaciones, hay otras en las cuales se hace necesario seleccionar un subconjunto relevante de las mismas [29]. Tal es el caso del problema abordado en esta tesis, en la cual mas allá del desempeño obtenido por la estrategia propuesta es deseable que la cantidad de información (características relevantes) sea la mínima posible.

Los métodos de selección de características se pueden clasificar de la siguiente manera: *filtros* (filters), *envolventes* (wrappers) e *integrados* (embedded) [29], los cuales serán descritos en detalle en el Capítulo 3.

1.3.3. Clasificación

En una ICC el clasificador es el encargado de *detectar* la presencia de potenciales evocados relacionados a eventos (P300) y así poder tomar una decisión sobre la intención de la persona de ejecutar una acción en función del caracter seleccionado. Los clasificadores pueden agruparse según su modo de entrenamiento en *supervisados* o *no supervisados*, según su capacidad de discriminación en *lineales* y *no lineales*, según el paradigma de implementación en *clásicos* o *bioinspirados*, entre otras formas.

El clasificador propuesto en esta tesis se enmarca en el contexto del *Análisis Discriminante*, se lo puede enmarcar dentro de los clasificadores de aprendizaje supervisado y cuyo objetivo es encontrar reglas de asignación de patrones a una de las clases C_i presentes en el problema a resolver. Si bien es posible utilizar distintos tipos de funciones discriminantes, aquí se utilizó un función discriminante de Fisher [30], la cual se describirá en el Capítulo 4.

1.4. Comentarios de cierre de capítulo

En este capítulo se describió una interfaz cerebro computadora, su arquitectura y se expuso la dificultad que representa detectar la señal de ERP en registros *single trial* en el contexto de una ICC basada en ERP. Además, y como punto a destacar, se presentó el planteo conceptual de la solución propuesta para abordar el problema mencionado, la cual básicamente consiste en implementar la etapa de extracción de características de una ICC mediante transformaciones ortogonales combinada con técnicas de selección de características, y la etapa de clasificación mediante un discriminante lineal. Cada una de estas estrategias se abordarán en detalle en los Capítulos 2, 3 y 4.

Capítulo 2

Generación de características mediante bases ortogonales

En este capítulo se presentan los conceptos teóricos relacionados con la Transformada Wavelet Diádica Discreta y la Transformada Wavelet Packet, a partir de las cuales se generarán características mediante proyección de la señal registrada en las bases ortogonales obtenidas para cada transformación.

2.1. Conceptos generales

Las características importantes de una señal, tales como bordes, espigas, transitorios u oscilaciones; en general son caracterizadas por información en el dominio del tiempo, en el de la frecuencia o en ambos como en el caso de los ERP.

En el ámbito de las ICC basadas en ERP, para determinar la respuesta de una persona es necesario detectar la presencia de la señal de ERP, en particular la de la onda P300 que es la característica mas importante de esta señal. En la Figura [2.1](#) se muestra la morfología típica de la señal de ERP donde se puede observar la onda P300, nótese que es un evento de muy corta duración (menos de 200 ms).

Desde el enfoque de espacios de señales, se puede considerar a una señal discreta $\mathbf{x} = x(n)$, con $1 \leq n \leq N$, como un elemento perteneciente a \mathbb{R}^N , la cual implícitamente está representada en la base *canónica* mediante una combinación lineal de señales deltas de Dirac desplazadas de acuerdo a la Ecuación [2.1](#):

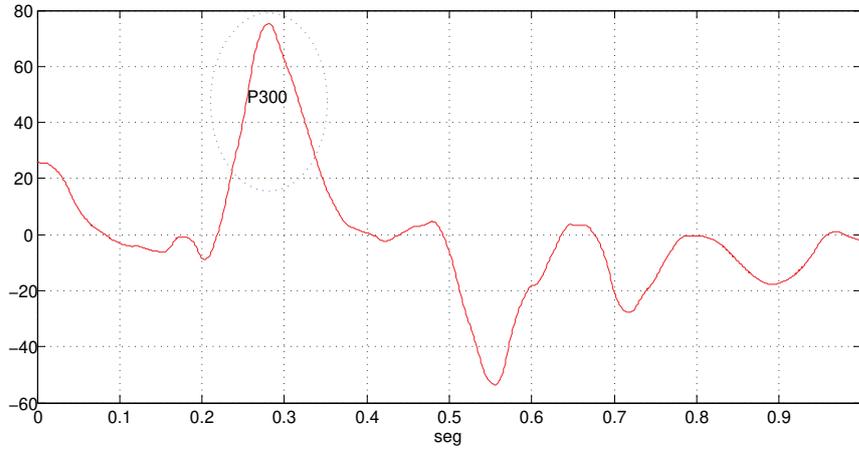


Figura 2.1: Señal de potencial evocado relacionado con eventos obtenida mediante promediación coherente de 100 épocas y posterior filtrado pasa bajos para suavizar la morfología.

$$x(n) = \sum_{k=1}^N x(k) \cdot \delta(n - k) \quad (2.1)$$

Recordando que $\mathbf{x} = \Phi \mathbf{a} + \boldsymbol{\varepsilon}$ y asumiendo que $\boldsymbol{\varepsilon} = \mathbf{0}$, es posible representar la señal \mathbf{x} en otra base que permita resaltar determinadas características de esta, puntualmente de la onda P300 que se encuentra inmersa en el ruido conformado básicamente por el EEG. La operación de cambio de base se realiza mediante una matriz de cambio de base o transformación representada por Φ^{-1} , mediante la Ecuación [2.2](#):

$$\mathbf{a} = \Phi^{-1} \cdot \mathbf{x} \quad (2.2)$$

y $a_i = \frac{\langle \mathbf{x}, \phi_i \rangle}{\|\phi_i\|^2}$ son las proyecciones de \mathbf{x} en la nueva base formada por las columnas ϕ_i de Φ . Los coeficientes que conforman \mathbf{a} se obtienen mediante el producto interno entre \mathbf{x} y los elementos ϕ_i de la nueva base; en consecuencia indican que tanto hay de cada elemento de la base en la señal \mathbf{x} [\[31\]](#).

A continuación se describirá a manera introductoria la base generada mediante la Transformada Discreta de Fourier y posteriormente las bases ortogonales generadas por las Transformada Wavelet Diádica Discreta y por la Transformada Wavelet Packets, las cuales serán utilizadas en esta tesis.

2.2. Transformada Discreta de Fourier

La *Transformada Discreta de Fourier* (TDF) se utiliza habitualmente en el ámbito del procesamiento digital de señales, conceptualmente consiste en la representación de una señal discreta en una base formada por exponenciales complejas discretas. En este caso la Ecuación 2.2 se define de la siguiente manera:

$$\mathbf{X} = \mathbf{\Phi}^{-1} \cdot \mathbf{x} \quad (2.3)$$

donde $\mathbf{\Phi} = \{\phi[k, n]\}$, las columnas $\phi[k, n]$ conforman la base de la transformada discreta de Fourier y están dadas por la Ecuación 2.4:

$$\phi[k, n] = e^{j\frac{2\pi kn}{N}} \quad (2.4)$$

con $0 \leq n \leq N$ y $\mathbf{\Phi}^{-1}$ se define como la conjugada de su transpuesta de manera que $\mathbf{\Phi}^{-1} = \mathbf{\Phi}^{*T}$ lo que define a la TDF como una transformación unitaria.

En la Figura 2.2 se muestran 4 elementos de la base de TDF para señales en \mathbb{R}^{64} dados por $\phi[k, n]$ donde $k=1..4$ y $N=64$.

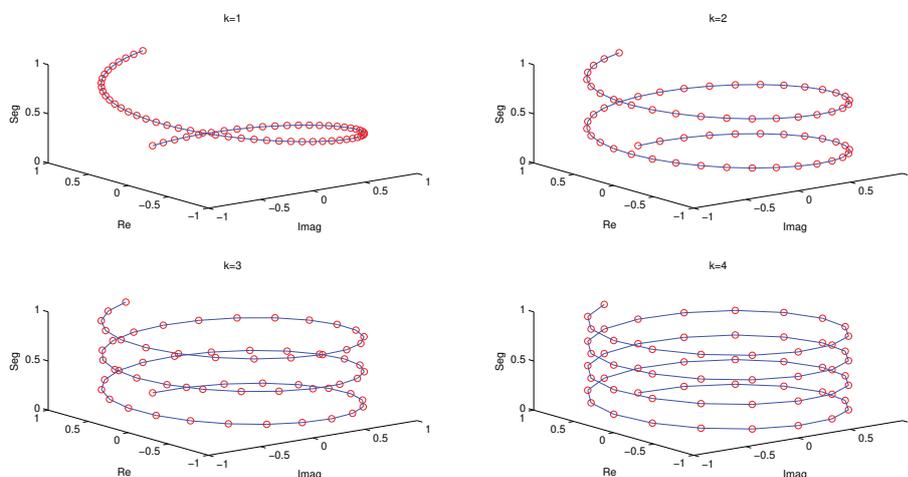


Figura 2.2: Elementos de la base para la TDF de Fourier en \mathbb{R}^{64} para $k = 1, k = 2, k = 3, k = 4$. Los valores discretos se representan con círculos, y si bien se han unido con líneas punteadas, recuérdese que las muestras de las señales de la base son solamente los valores de los círculos marcados.

Si bien la TDF es ampliamente utilizada, su aplicación en el problema abordado en esta tesis no resulta adecuada, ya que es muy difícil detectar un evento transitorio de corta duración (onda P300) utilizando una base cuyos elementos

morfológicamente no son similares a esta y además no tienen una buena localización temporal. En el desarrollo de esta tesis se realizaron experimentos para evaluar el desempeño de la TDF, los cuales como era de esperar generaron resultados que no fueron satisfactorios por lo que no se la considerará como una alternativa válida.

2.3. Transformada Wavelet Diádica Discreta

Se define como *wavelet* a una función que debe cumplir las siguientes condiciones:

- a) debe tener valor medio *cero* y derivadas continuas hasta cierto orden
- b) debe tener energía acotada en un rango de tiempo
- c) no poseer componentes frecuenciales nulos.

Estas brindan información del dominio temporal y frecuencial simultáneamente, y es por esto que son particularmente útiles para el análisis de señales transitorias, aperiódicas o no estacionarias. Además existe una gran cantidad de funciones o familias wavelet con diferentes propiedades, lo que permite la elección de una *wavelet madre* en función de la señal a analizar.

La *Transformada Wavelet Continua* (CWT, del inglés Continuous Wavelet Transform) es una herramienta de análisis tiempo-frecuencia que, a diferencia de la Transformada de Fourier, permite una buena localización en el tiempo en señales de alta frecuencia.

Sea $x(t)$ una señal $\in L^2(\mathbb{R})$ y $\psi(t)$ una función wavelet madre también $\in L^2(\mathbb{R})$, entonces se define la CWT de $x(t)$ como:

$$c_{a,b} = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{|a|}} \psi^* \left(\frac{t-b}{a} \right) dt. \quad (2.5)$$

donde cada elemento del diccionario es una versión escalada de la wavelet madre mediante un factor de escala a , dado la Ecuación [2.6](#), y trasladada mediante un factor b .

$$\psi_a(t) = \frac{1}{\sqrt{|a|}} \psi \left(\frac{t}{a} \right), \quad (2.6)$$

En la Ecuación [2.6](#) se observa que factores de escala grandes generan versiones comprimidas de $\psi(t)$ y factores de escala chicos generan versiones dilatadas de $\psi(t)$ [\[32\]](#), esto se muestra en la Figura [2.3](#). De ahora en adelante se utilizará el

término *escala* en lugar de factor de escala.

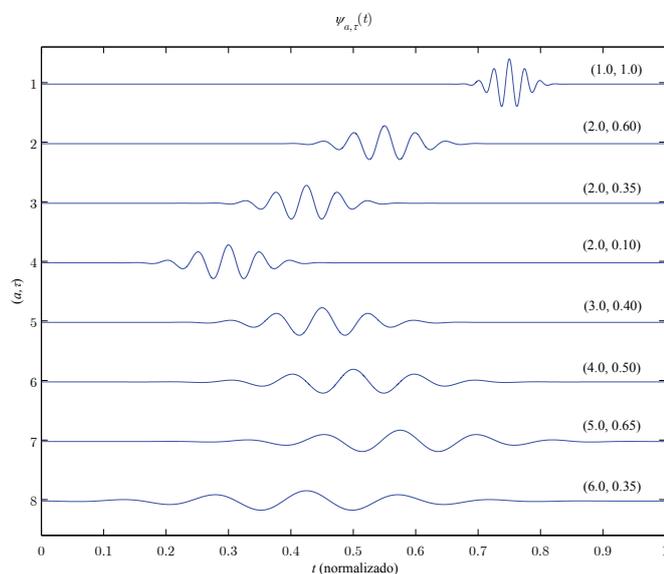


Figura 2.3: Ejemplo de la wavelet de Morlet para distintas escalas y localizaciones. Adaptado de [2]

En la Ecuación 2.5, $c_{a,b}$ denota los coeficientes wavelets que son el resultado de comparar la señal $x(t)$ con cada una de las versiones de $\psi(t)$ escaladas y trasladadas en el tiempo. El valor de $c_{a,b}$ indica el parecido de $x(t)$ con cada elemento del diccionario, por lo tanto si se define una wavelet madre con una morfología similar al P300 se podría caracterizar esta con muy pocos coeficientes. Esta aproximación permite que cualquier señal pueda ser representada como una descomposición en wavelets, con lo cual $\psi_{a,b}(t)_{a,b \in \mathbb{Z}}$ podría conformar una base ortogonal. Este no es siempre el caso, ya que el conjunto es redundante [2].

Una manera de pensar la CWT es que los escalamientos y traslaciones de la wavelet madre generan un *diccionario de átomos tiempo-frecuencia*, lo cual permite disponer de una cantidad muy grande de elementos para formar una base, utilizando un criterio adecuado es posible definir una que sea ortogonal para representar la señal de ERP.

La justificación de la utilización de wavelets en esta tesis es que, como se mencionó previamente, estas presentan un soporte compacto que permite capturar características locales de la onda P300 en el plano tiempo-frecuencia; que a su vez resulten discriminativas de una manera computacionalmente eficiente [2]. Entre los distintos criterios más usuales que podrían resultar útiles en esta aplicación se pueden mencionar el de la entropía de los coeficientes wavelets [33],

el de máxima discriminación entre clases [34] y el de mejor dispersión [35].

En los casos de aplicación de la CWT, su implementación es costosa desde el punto de vista computacional, mediante la discretización adecuada de los parámetros a y b se obtiene una implementación eficiente y en algunos casos una representación mediante una base *ortonormal*, por este motivo se propone utilizar la transformada Wavelet discreta introducida por Mallat [32] que se describe a continuación.

La *Transformada Wavelet Diádica Discreta* (DDWT, del inglés Diadyc Discrete Wavelet Transform) genera una base ortonormal, la cual se obtiene como consecuencia de una discretización diádica de los parámetros de escala a y de traslación b de la Ecuación 2.5.

Una Wavelet diádica es una función $\psi(t) \in L^2(\mathbb{R})$ que tiene la propiedad de que la familia de funciones $\psi_{k,n}(t) = 2^{k/2}\psi(2^k t - n)$ para $k, n \in \mathbb{Z}$ conforma una base ortonormal en $L^2(\mathbb{R})$. Esto implica la existencia de una *sola* función cuyas traslaciones y dilataciones apropiadas forman una base ortonormal [2].

La DDWT permite realizar un análisis multiresolución, en el que cada escala pertenece a una determinada resolución y la señal queda representada en un conjunto de subespacios de $L^2(\mathbb{R})$.

El *análisis multiresolución* (MRA, del inglés Multi Resolution Analysis) es una secuencia $(V_k)_{k \in \mathbb{Z}}$ de subespacios de $L^2(\mathbb{R})$ tales que:

- $V_k \subset V_{k+1}$ para cualquier $k \in \mathbb{Z}$,
- $\overline{\bigcup_{k \in \mathbb{Z}} V_k}$ es denso en $L^2(\mathbb{R})$ y $\bigcap_{k \in \mathbb{Z}} V_k = \{0\}$,
- $x(t) \in V_0 \Leftrightarrow x(2^{-k}t) \in V_k$ para cualquier $k \in \mathbb{Z}$,
- \exists una función $\phi(t) \in V_0$, llamada la *función de escala*, tal que la familia $\{\phi(t - n)\}_{n \in \mathbb{Z}}$ es una base ortonormal para V_0 .

Considerando la definición de MRA es posible aproximar cualquier función o señal en $L^2(\mathbb{R})$ por una función en alguno de los V_k , se dice que ésto constituye una *aproximación* a la resolución o escala k .

Dado que $\{2^{k/2}\phi(2^k t - n)\}_{n \in \mathbb{Z}}$ es una base ortonormal para V_k , ésto da la posibilidad de analizar una señal con “ventanas” de diferentes tamaños, a diferencia del análisis basado en la TDF. Las condiciones sobre los subespacios V_k implican que por estar en el espacio de Hilbert $L^2(\mathbb{R})$ existen subespacios $(W_k)_{n \in \mathbb{Z}}$

tal que cada V_{k+1} es la suma directa de V_k con W_k . Las W_k tienen la interpretación de que representan la información de *detalle* que se requiere cuando se pasa de una aproximación a la resolución k , a una aproximación a la resolución $k + 1$. En este caso $V_{k+1} = V_k \oplus W_k$, y continuando el proceso se obtiene $V_{k+1} = V_l \oplus W_l \oplus W_{l+1} \oplus \dots \oplus W_{k-1} \oplus W_k$. Entonces una aproximación a una resolución más alta puede ser representada en términos de una resolución más baja con los detalles adicionales, además $L^2(\mathbb{R})$ es la suma directa de los W_k [32].

En otras palabras, las frecuencias más altas se analizan utilizando ventanas estrechas para obtener buena resolución temporal, mientras que las bajas frecuencias se analizan utilizando ventanas anchas para obtener una buena resolución frecuencial. A diferencia del caso de la TDF en la cual se tiene una resolución frecuencial fija, en el MRA se analiza la señal de manera tal que cada componente de frecuencia se estudia con una resolución diferente.

En su implementación, la familia de Wavelets puede ser vista como un banco de filtros en octavas. Un filtro pasa-altos con respuesta al impulso $h[n]$, el cual genera los coeficientes de detalles y un filtro pasa-bajos con respuesta al impulso $g[n]$, que genera los coeficientes de aproximación. Estos filtros, llamados *filtros espejos de cuadratura*, se aplican de forma iterativa sobre cada aproximación generada en cada nivel de descomposición [36].

En cada nivel se realiza un submuestreo del nivel anterior, este proceso se muestra en la Figura 2.4, en la cual se tiene una señal $x(t)$ muestreada con una frecuencia de muestreo f_m , obteniéndose una secuencia $x[n]$, con N igual al número total de muestras tomadas.

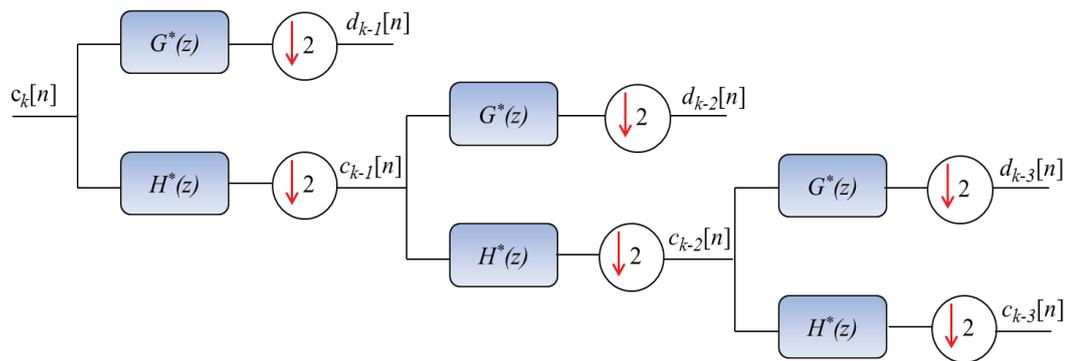


Figura 2.4: Esquema del funcionamiento de algoritmo de la DDWT. Los coeficientes de aproximación son filtrados nuevamente y se produce un submuestreo del nivel anterior. Extraído de [2].

Este análisis resulta en la descomposición de una señal $x[n]$ en sus diferen-

tes componentes de escalas (o frecuencias), donde en cada nivel se produce una reducción de la cantidad de muestras a la mitad de las del nivel anterior [32].

2.4. Transformada Wavelet Packet

La *Transformada Paquete de Onditas* (WPT, del inglés Wavelet Packet Transform) es una versión generalizada de la DDWT, la cual puede pensarse como la aplicación de los filtros $h[n]$ y $g[n]$ tanto a las *aproximaciones* como a los *detalles*, tal como se muestra en la Figura 2.5.

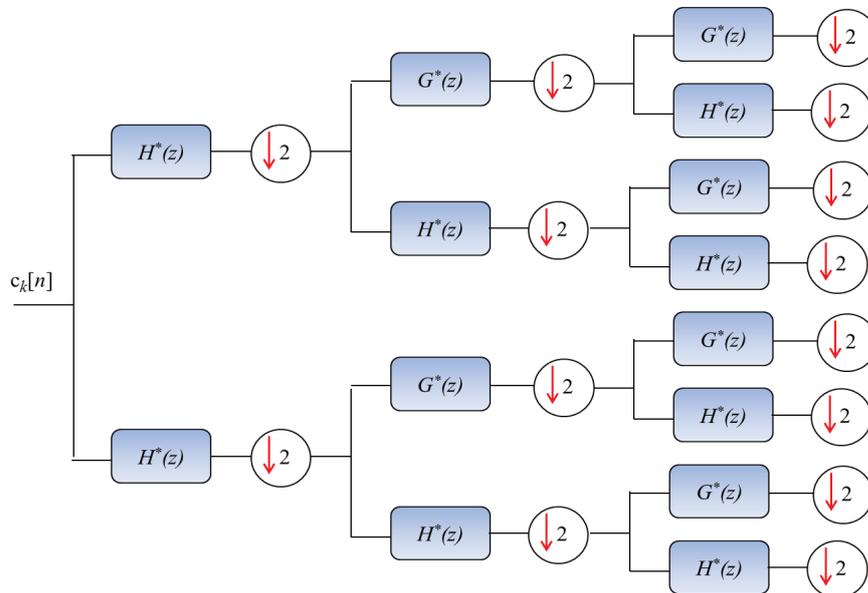


Figura 2.5: Esquema del funcionamiento de algoritmo de la WPT. Los coeficientes de aproximación y de detalle se filtran en cada nivel generado. Extraído de [2].

Tanto en el caso de la Figura 2.4 como en el de la Figura 2.5 se realiza la descomposición de la señal $x[n]$ hasta un nivel $J=3$, se observa que la WPT presenta mayor cantidad de nodos y por ende descompone en mayores bandas frecuenciales a la señal.

En general, cuando se realiza la descomposición hasta un cierto nivel J se obtiene un diccionario de funciones con el que se pueden formar $2^{2(J-1)}$ bases ortogonales para conformar la matriz de transformación Φ [34]; por lo que se tiene a disposición una cantidad finita de posibles bases ortogonales para representar a la señal de interés en función de cómo se vaya recorriendo el árbol. La elección de la base adecuada se realiza normalmente utilizando métodos que tienen co-

mo objetivo retener la mayor cantidad de información en la menor cantidad de coeficientes [37].

Estos métodos se basan en el paradigma denominado *best-basis*, el cual puede describirse mediante una secuencia de pasos:

1. Seleccionar la *mejor base* para un problema particular a partir de un diccionario de posibles bases.
2. Ordenar los elementos de la base (vectores base) según la *importancia* para el problema analizado y descartar el resto.
3. Utilizar los vectores bases *sobrevivientes* para resolver el problema.

Dentro de los posibles algoritmos para seleccionar la mejor base se pueden mencionar el Best Orthogonal Basis desarrollado por Coifman *et al* y el Local Discriminant Basis desarrollado por Saito *et al*, los cuales se describen a continuación.

2.4.1. Algoritmo Best Orthogonal Basis

El algoritmo *Best Orthogonal Basis* (BOB) es uno de los más utilizados, permite seleccionar una base ortogonal a partir de un diccionario de funciones wavelet obtenido mediante WPT, minimizando una función de costo determinada. La principal aplicación de este algoritmo es la compresión eficiente de señales o imágenes, es decir, cuando se busca una representación en unos pocos componentes de gran importancia sin degradar significativamente a la señal o imagen analizada [38].

La función de costo en este algoritmo es la entropía de Shannon, esta mide la concentración de energía de una señal y da idea del grado de aleatoriedad presente en la misma. Considerando una secuencia $\mathbf{p} = \{p_i\}$, donde $\sum_i p_i = 1$, entonces se define esta entropía de la siguiente manera:

$$H(\mathbf{p}) = -\sum_i p_i \log p_i, \quad (2.7)$$

con la convención de que $0 \cdot \log 0 = 0$.

Considerando una señal \mathbf{x} y haciendo $p_i = \frac{|x_i|^r}{\|\mathbf{x}\|_r^r}$, donde $\|\cdot\|_r$ es la norma l^r y $1 \leq r \leq \infty$, la entropía de dicha señal se define como:

$$H_r(\mathbf{x}) = - \sum_i \frac{|x_i|^r}{\|\mathbf{x}\|_r^r} \log \frac{|x_i|^r}{\|\mathbf{x}\|_r^r} \quad (2.8)$$

en general $r = 1$ o $r = 2$ [39]. Esta función de costo mide la bondad de cada nodo rama del árbol binario, en cada subespacio, para la compresión de la señal. El siguiente algoritmo esencialmente “poda” este árbol binario mediante la comparación de la eficiencia de cada nodo padre en función de sus nodos hijos [38].

Para una señal \mathbf{x} , el algoritmo BOB puede describirse en los siguientes pasos:

1. Elegir un diccionario de bases ortogonales, en el caso de la WPT comprende seleccionar la wavelet madre y especificar un máximo nivel de descomposición J .
2. Calcular la entropía presente en cada nodo de árbol binario.
3. Podar el árbol binario mediante la eliminación de aquellos nodos hijos si la suma de su entropía es menor o igual a la de los nodos padres.

Si bien inicialmente se consideró este algoritmo como alternativa para obtención de una base ortogonal y se hicieron pruebas preliminares que se presentan en la Sección 5.5.4, los resultados obtenidos no fueron satisfactorios por lo que no se lo considerará como una alternativa para generar una base.

2.4.2. Algoritmo Local Discriminant Basis

A diferencia de BOB, el algoritmo *Local Discriminant Basis* (LDB) fue desarrollado para aplicarse a problemas de clasificación [34], permite seleccionar una base ortogonal a partir de un diccionario WPT con el objetivo de maximizar la separación entre las clases posibles de un problema, lo cual lo hace adecuado para resolver el problema planteado en esta tesis. El algoritmo LDB propone que para problemas de clasificación es más eficiente evaluar el poder de discriminación entre las clases en cada subespacio del árbol binario (cada nodo rama), a partir de alguna función de costo aditiva, antes que evaluar la eficiencia en la representación de la señal como lo hace el algoritmo BOB. Para determinar el poder de discriminación se utilizan distintas alternativas, todas ellas básicamente miden *distancias* a lo largo de las clases involucradas en el problema en cuestión [39].

Sean $\mathbf{p} = (p_i)_{i=1}^n$, $\mathbf{q} = (q_i)_{i=1}^n$ dos secuencias no negativas con $\sum p_i = \sum q_i = 1$, que pueden pensarse como la distribución de energía normalizada de señales pertenecientes a la *clase 1* y a la *clase 2* de un problema de clasificación. La potencia de discriminación está dada por una función de discriminación de información $D(\mathbf{p}, \mathbf{q})$ entre estas dos secuencias; a continuación se describen 3 alternativas para medir la potencia de discriminación:

- a) Asimétrica: también denominada *distancia Kullback-Leibler o Divergencia-I*, esta dada por la siguiente ecuación:

$$I(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^n p_i \log \frac{p_i}{q_i} \quad (2.9)$$

donde $\log(0) = -\infty$, $\log(x/0) = +\infty$ para $x > 0$. Esta no es estrictamente una distancia ya que no es simétrica y no satisface la igualdad del triángulo.

- b) Divergencia-J: es una alternativa de distancia simétrica entre \mathbf{p} y \mathbf{q} , la cual se define de la siguiente manera:

$$J(\mathbf{p}, \mathbf{q}) = I(\mathbf{p}, \mathbf{q}) + I(\mathbf{q}, \mathbf{p}) \quad (2.10)$$

- c) Suma de cuadrados: esta medida se define de la siguiente manera:

$$W(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|^2 = \sum_{i=1}^n (p_i - q_i)^2 \quad (2.11)$$

En resumen, a partir de la WPT se dispone de un diccionario constituido por una familia de posibles bases ortogonales. El algoritmo LDB, mediante alguna de las medidas de discriminación descritas, permite obtener una base ortogonal que maximice la separación entre las clases.

Con el fin de utilizar plenamente la localización tiempo-frecuencia de las características del diccionario de bases, se calcula un mapa de energía tiempo-frecuencia Γ_c para cada clase para ser suministrado a D . Este mapa de energía Γ_c se calcula mediante la acumulación de los cuadrados de los coeficientes de expansión de las señales en cada posición del árbol binario seguido por la normalización de la energía total de las señales que pertenecen a la clase C_i . La normalización puede ser importante, especialmente si hay diferencias significativas en el número de muestras entre las clases [39].

La implementación del algoritmo LDB se realiza mediante una secuencia de 6 pasos [40]:

1. Elegir un diccionario de bases ortogonales, en el caso de la WPT comprende seleccionar la wavelet madre y especificar un máximo nivel de descomposición J . Además se debe seleccionar la medida de discriminación D .
2. Construir el mapa de energía Γ para cada clase.
3. Calcular la medida de discriminación D a lo largo de los mapas de energía en cada nodo rama.
4. Podar el árbol binario: eliminar aquellos nodos hijos si la suma de su medida de discriminación es menor o igual a la de los nodos padres.
5. Ordenar los vectores bases de mayor a menor según su poder de discriminación
6. Utilizar los p vectores bases más discriminativos para generar los patrones para la clasificación mediante la expansión de la señal en esta base.

A partir del paso 4 se dispone de una base ortogonal, pudiéndose utilizar todos los coeficientes de expansión de la señal en esta base como características para la clasificación, los pasos 5 y 6 son necesarios si se desea reducir la dimensionalidad del problema.

2.5. Comentarios de cierre de capítulo

En este capítulo se presentaron conceptos relacionados con transformaciones unitarias como herramienta para obtener nuevas bases ortogonales en las cuales proyectar una señal para destacar rasgos de interés de esta, en particular se describieron la Transformada Wavelet Diádica Discreta y la Transformada Wavelet Packets. En la DDWT se mostró como se genera la base ortogonal a través de la descomposición multiresolución a partir de un diccionario Wavelet, mientras que en el caso de la WPT se describieron los algoritmos BOB y LDB para este fin.

Capítulo 3

Selección de características

En este capítulo se describe la motivación para utilizar métodos de selección de características, los conceptos generales y descripción de los métodos implementados.

3.1. Conceptos generales

El proceso de selección de características permite reducir la dimensión de un conjunto de datos o *patrón*, cuyo objetivo es remover aquellas que resultan irrelevantes o aquellas que resultan redundantes para modelar un determinado comportamiento o resolver un problema. El número de características influye en la complejidad del modelo, al reducirlo permite disminuir la cantidad de tiempo de cómputo, construir modelos más generales y simples de comprender, pero por sobre toda las cosas reducir la dimensionalidad del problema favoreciendo su resolución [41].

En la Figura 3.1 se presenta un esquema con los pasos involucrados en el proceso de selección de características, los que se describen a continuación:

- a) Procedimiento de generación: es un proceso que tiene como objetivo generar subconjuntos de características que serán ponderados por una función de evaluación. Este primer paso puede iniciarse con un conjunto vacío, con el conjunto total de características o con un conjunto de características aleatorio que sea subconjunto del conjunto completo. En los primeros dos casos, la generación de subconjuntos se realiza agregando o quitando respectivamente características en forma iterativa y determinista, mientras que en el tercer caso

pueden tanto agregarse como quitarse características y las mismas se pueden seleccionar de manera aleatoria.

- b) Función de evaluación: esta determina la calidad del subconjunto de características seleccionado por el proceso de generación, lo cual permite realizar la comparación de dicho subconjunto de características con otros subconjuntos a fin de determinar cuál de ellos es mejor para una determinada aplicación.
- c) Criterio de detención: dado que el proceso de selección de características (generación, evaluación y comparación) es un proceso iterativo, es necesario un criterio de detención del mismo. Un criterio de detención basado en el procedimiento de generación puede contemplar, por ejemplo, si se alcanzó un número de características seleccionadas predefinido o si se alcanzó cierto número de generaciones. Por otra parte, un criterio de detención basado en la función de evaluación puede contemplar, por ejemplo, si luego de cierto número de iteraciones no se logró una mejora en la calidad de los subsecuentes subconjuntos de características o si se alcanzó algún subconjunto que sea óptimo de acuerdo con alguna de las funciones de evaluación. Una vez que el criterio de detención se cumple, el proceso entrega como resultado uno o varios subconjuntos candidatos.
- d) Procedimiento de validación: no se considera que forma parte del proceso de selección de características en sí mismo, sino que es un paso posterior que permite validar el método. En este último paso se realizan distintos análisis, pruebas y comparaciones con resultados previos o con resultados obtenidos por otros métodos de selección de características.

Los métodos de selección de características se pueden encuadrar dentro de alguno de los siguientes modelos: filtros (filters), envolventes (wrappers) e integrados (embedded) [29].

Los métodos filtros realizan la selección de un subconjunto de las características originales aplicando algún criterio que resulte independiente del desempeño del clasificador, es decir que no incorporan el aprendizaje en la selección de las características. Los componentes básicos de este tipo de algoritmos son un método de búsqueda de características y un criterio de selección [29], algunos ejemplos utilizados en el caso de señales temporales son:

- a) submuestreo o diezmado: mediante este procedimiento se genera una nueva

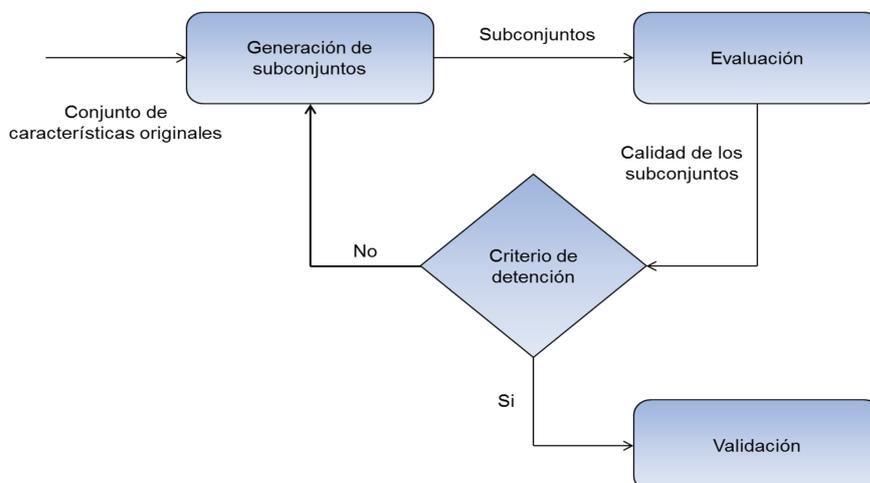


Figura 3.1: Pasos involucrados en la selección de características: generación, evaluación, detención y validación.

señal muestreada a un frecuencia menor, esto fue realizado sobre los registros de la base de datos y se dará mas detalle en la Sección [5.1](#).

- b) transformaciones: en este caso se genera un conjunto de características a partir de las originales, tal es el caso de la Transformada Discreta de Fourier, de la Transformada Wavelet Diádica Discreta o de la Transformada Wavelets Packets.

Los métodos envolventes usan una técnica de aprendizaje tipo caja negra, representando un clasificador, y utilizan el desempeño de estos para realizar la elección del subconjunto de características [\[42\]](#). Estos a su vez pueden dividirse internamente en *búsqueda de características* y *evaluación del subconjunto de características*. El primer componente es responsable de realizar la búsqueda combinatoria entre los posibles subconjuntos de características, mientras que el segundo componente evalúa la eficacia del subconjunto seleccionado y en base a esto guía la búsqueda de características [\[29\]](#). Una implementación de este tipo de métodos se puede realizar utilizando *algoritmos genéticos* para realizar la búsqueda de características combinado con un clasificador para evaluar los distintos subconjuntos generados; esta alternativa es la que se implementó en esta tesis y se describirá en la Sección [6.5](#).

Los métodos integrados difieren de los demás métodos de selección de características en la forma de interacción entre la selección de las características y el

aprendizaje [42], en este caso las partes constitutivas del método no pueden ser separadas. Los algoritmos integrados, realizan la selección de las características durante el entrenamiento del clasificador y generalmente son específicos para cada tipo de máquina de aprendizaje. Estos se pueden clasificar de la siguiente manera:

- a) *selección hacia adelante*: comienzan con una o pocas características y en forma iterativa agregan características hasta que se alcanza el criterio de parada.
- b) *eliminación hacia atrás*: comienzan incluyendo todas las características y en forma iterativa remueven una o varias características
- c) *anidados*: durante las iteraciones se pueden agregar así como remover características [29, 42].

En esta tesis se utilizó el algoritmo denominado *eliminación recursiva de características*, el cual se enmarca dentro de los métodos integrados de eliminación hacia atrás y su implementación se describirá en la Sección 6.5.

A continuación se presentan los fundamentos teóricos de los algoritmos genéticos y del método de eliminación recursiva de características.

3.2. Algoritmos genéticos

Los *algoritmos genéticos* (AGs) pueden ser usados en el esquema envolvente junto a distintos tipos de clasificadores; estos son una herramienta de optimización en los que hay que considerar cuatro aspectos en su utilización: codificación, población, operadores y aptitud (o *fitness*) [43].

El paradigma de implementación de un AG se muestra a continuación:

1. Inicialización de población
2. Evaluación de la población
3. Mientras $MejorFitness \leq FitnessDeseado$ Hacer
 - Selección*
 - Cruzas o recombinaciones*
 - Mutaciones*
 - Evaluación de la población*
4. Fin mientras

1. Inicialización de población: los AGs manipulan una población de posibles soluciones de un problema codificadas en cadenas binarias, estas se denominan *individuos* y representan el material genético de una *población de individuos* [31].

2. Evaluación de la población: cada potencial solución se asocia con un valor de *fitness* o *aptitud* que mide que tan buena es comparada con otras soluciones (individuos) de la población. De esta manera el fitness simula el papel que juega el ambiente en la evolución natural darwiniana, esta debe ser monótonamente creciente con la bondad de la solución que un individuo representa. Si esta dependiera de varios factores, de la forma en que estos son pesados en la función de fitness dependerá la optimización que realice el AG [31].

3. Operadores de variación: utilizando operadores artificiales de *selección*, *cruzas* y *mutación* se genera una nueva población en la cual se buscan los mejores individuos que a su vez implican mejores soluciones a través de la simulación del proceso evolutivo natural. Estos operadores se describen a continuación:

- a) Operador de selección: permite elegir los individuos que oficiarán como progenitores o padres de una nueva población. Hay varias alternativas que se utilizan para este fin, los más usados son el método de la ruleta, el método de ventanas, el método de competencia y el elitismo [31]. En esta tesis se utilizó el de competencias y será descrito en el Capítulo 4.
- b) Operador de cruce o recombinación: permite el intercambio de parte del material genético entre dos individuos progenitores elegidos al azar. Puede ser de un punto (cruza simple) o de dos puntos, en el caso de la cruce simple se elige al azar un punto de intercambio en el individuo y a partir de ese punto se intercambian todos los bits hasta el final de los dos individuos seleccionados; en el caso de la cruce de dos puntos se eligen al azar dos puntos y se intercambian los bits definidos por estos. En la Figura 3.2 se esquematiza el procedimiento de cruce simple de 2 individuos de 8 bits de longitud.
- c) Operador de mutación: produce el cambio de estado de un bit seleccionado al azar, y a diferencia de la cruce se dan con una probabilidad baja, si esta es demasiado baja la exploración es mínima y las generaciones sucesivas van eliminando la variedad de la población y una vez que la población converge en un genotipo toda otra exploración cesa. Por otro lado, si la probabilidad es alta a lo largo de las generaciones el proceso de búsqueda se degenera convirtiéndose en una búsqueda aleatoria sin explotación de la información adquirida en las generaciones anteriores. En la Figura 3.3 se esquematiza el procedimiento de mutación en un individuo de 8 bits de longitud.

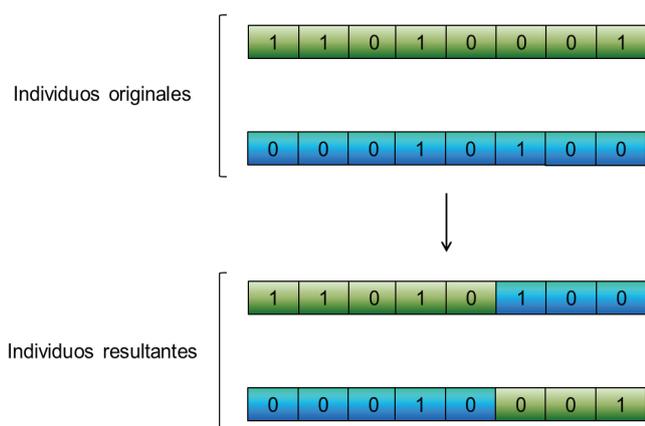


Figura 3.2: Cruza simple de 2 individuos de 8 bits, el *intercambio de material genético* se produce en la posición 5.

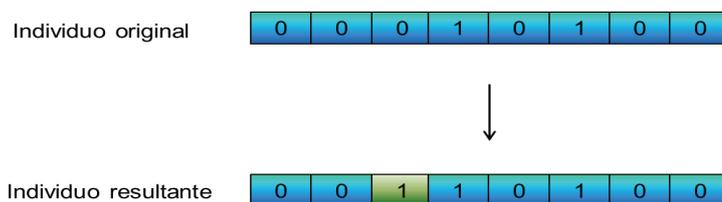


Figura 3.3: Mutación de un bit en la posición 3 en un individuo de 8 bits.

3.3. Eliminación recursiva de características

El algoritmo de *eliminación recursiva de características* (ERC) es un proceso simple y relativamente eficiente, la idea es utilizar un clasificador y utilizar información del mismo para ordenar las características de acuerdo a su importancia. Una vez ordenadas, se elimina el subconjunto de menor importancia (en algunos casos se elimina una variable a la vez, en otros un número mayor) y se repite el proceso. La iteración, que agrega una carga computacional, es necesaria para poder enfrentar el problema de la correlación entre características. Cuando dos características determinantes tienen una alta correlación, el clasificador tiende a repartir la importancia entre ambas, recién cuando el algoritmo remueve una de ellas la otra toma toda la importancia y pasa a ocupar lugares más altos en el ranking [29]. Este algoritmo se puede combinar con cualquier clasificador que permita extraer información sobre la influencia de cada variable en la asignación de clase.

El algoritmo es un proceso iterativo que se puede sintetizar en tres pasos que se muestran a continuación:

- Entrenamiento del clasificador
- Cálculo del criterio para establecer un ranking entre las características
- Eliminación de las características que se encuentran en la parte inferior del ranking

Una de las ventajas de este algoritmo es que se pueden eliminar más de una característica por iteración y el algoritmo termina cuando el criterio de detención se ha alcanzado o cuando no hay más características que evaluar [29]. El criterio de eliminación utilizado es el valor absoluto de la amplitud de los pesos, es decir, el valor de los coeficientes que conforman la combinación lineal que posibilita la clasificación.

3.4. Comentarios de cierre de capítulo

En este capítulo se expuso la importancia de reducir la dimensionalidad de los patrones, los conceptos básicos relacionados con el proceso de selección de características y las distintas estrategias para llevarla a cabo. En particular se describieron las estrategias que se implementan en esta tesis, una del tipo wrapper basada en algoritmos genéticos y otra del tipo integrada denominada eliminación recursiva de características. Además de estas, implícitamente también se utiliza una estrategia del tipo filtros, ya que tanto el submuestreo de las señales registradas (que se detallará en el Capítulo 5) como las transformaciones o cambio de bases descriptas en el Capítulo 2 son ejemplos de este paradigma.

Capítulo 4

Clasificación

En este capítulo se describe las características de la etapa de clasificación, la cual está conformada por un clasificador lineal del tipo discriminante lineal. Complementariamente, se describe un ensamble de clasificadores como estrategia alternativa para mejorar el desempeño de un clasificador único.

4.1. Conceptos generales

Como se ha mencionado previamente, el rendimiento de una ICC depende fuertemente de las estrategias de extracción de características que codifican las intenciones de un usuario y del clasificador que se utilice [44]. En este contexto se han utilizado distintos tipos de clasificadores, como máquinas de soporte vectorial [45], [46], redes neuronales [47], entre otros. Mas allá de esto, en [48] se demuestra que algoritmos básicos de clasificación, como por ejemplo discriminantes lineales, pueden convertirse en una potente herramienta para la clasificación de ERP ya que son simple de implementar, de bajo costo computacional $O(n^3)$, fácil de aplicar y genera resultados similares a los métodos de clasificación más complejos [17]. Debido a que el objetivo de esta tesis es evaluar estrategias de extracción de características y no de clasificación, y que de la literatura revisada se desprende que el discriminante lineal es uno de los clasificadores más utilizados en el área de las ICC con buenos resultados [17, 49], por estos motivos se propone utilizarlo y se lo describe a continuación.

4.2. Discriminante lineal

Como se mencionó en el Capítulo 1, en esta tesis se utilizó como clasificador un *discriminante lineal* (LDA, del inglés Linear Discriminant Analysis), el cual es un clasificador del tipo binario que puede ser caracterizado a partir de una proyección de un vector \mathbf{w} y un término de bias o umbral b que hace referencia al hiperplano de separación definido mediante $\mathbf{w}^T \mathbf{x} + b = 0$. La función de clasificación asigna a una nueva entrada $\mathbf{x} \in \mathbb{R}^N$ la etiqueta de una de las dos clases posibles C_i según sea el signo de la función $\mathbf{w}^T \mathbf{x} + b$.

El vector de proyección se define como:

$$\mathbf{w} = \hat{\Sigma}_c^{-1}(\hat{\mu}_2 - \hat{\mu}_1) \quad (4.1)$$

donde μ_i es la media estimada de cada clase C_i y $\hat{\Sigma}_c = \frac{1}{2}(\hat{\Sigma}_1 + \hat{\Sigma}_2)$ es la matriz común de covarianza estimada, es decir, el promedio de las matrices empíricas de covarianza de cada clase [48].

En la Figura 4.1 se muestra un ejemplo gráfico de la separación lineal que este clasificador genera sobre dos clases hipotéticas C_1 y C_2 , se clasificarán como pertenecientes a C_1 a aquellas entradas que cumplan con $\mathbf{w}^T \mathbf{x} + b > 0$ mientras que el resto sería clasificada como pertenecientes a C_2 .

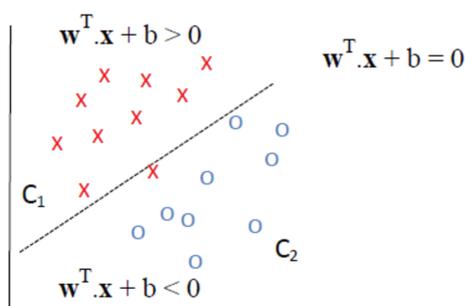


Figura 4.1: Representación gráfica de la separación entre clases realizada por un LDA.

Dado $\mathbf{x} = \Phi \mathbf{a}$, entonces \mathbf{a} son las características extraídas de la señal \mathbf{x} para proceder a su posterior clasificación, las cuales a su vez pueden ser el resultado de aplicar un algoritmo de selección de características; las que deberían ser lo más *discriminativas posibles* para facilitar la clasificación de las señales.

Suponiendo que se tiene un conjunto de K señales en una matriz $\mathbf{X} = [\vec{x}_1, \vec{x}_2, \dots, \vec{x}_K]$ con su correspondiente representación $\mathbf{A} = [\vec{a}_1, \vec{a}_2, \dots, \vec{a}_K]$ en

términos de un diccionario sobrecompleto Φ , del cual K_i señales corresponden a la clase C_i , para $1 \leq i \leq C$. La media $\vec{\mu}_i$ y la varianza $\vec{\sigma}_i^2$ para la clase C_i en el espacio de las características se calcula como:

$$\vec{\mu}_i = \frac{1}{K_i} \sum_{\vec{a}_k \in C_i} \vec{a}_k \quad (4.2)$$

$$\vec{\sigma}_i^2 = \frac{1}{K_i} \sum_{\vec{a}_k \in C_i} \|\vec{a}_k - \vec{\mu}_i\|_2^2 \quad (4.3)$$

Como se mencionó en la Sección [1.3.3](#), si bien hay distintas alternativas para definir la función discriminante aquí se utilizó la función de Fisher. Su utilización aquí es pertinente, dado que esta función o criterio esta motivada por la idea intuitiva de que la potencia de discriminación se maximiza cuando la distribución espacial de las muestras de clases distintas se encuentran tan distantes como sea posible y la distribución espacial de las muestras dentro de la misma clase se encuentran tan cerca como sea posible [\[41\]](#).

Considerando que el valor medio de todas las señales se define como: $\vec{\mu} = \frac{1}{K_i} \sum_{k=1}^K \vec{a}_k$ entonces el criterio de discriminación de Fisher se define de la siguiente manera:

$$F(\vec{\Phi}, \vec{A} | \vec{X}) = \frac{SB}{SW} = \frac{\left\| \sum_{i=1}^C K_i (\vec{\mu}_i - \vec{\mu})(\vec{\mu}_i - \vec{\mu})^T \right\|_2^2}{\sum_{i=1}^C \vec{\sigma}_i^2} \quad (4.4)$$

donde SB puede ser interpretada como la distancia interclase y SW como la dispersión intraclase.

4.3. Ensamble de clasificadores

En la sección previa se mencionaron los motivos por el cual se seleccionó un LDA como clasificador, sin embargo a la hora de evaluar su desempeño hay que considerar que este puede ser influido por la información propia del problema presente en los patrones, que sería lo deseable; o por aspectos relacionados por ejemplo al entrenamiento que son propios de cada tipo del clasificador [\[44\]](#). Con respecto a este último punto, en la Sección [5.5.5](#) se analiza como influye en un LDA la relación entre la cantidad de patrones de entrenamiento y la dimensión de estos para establecer un punto de trabajo en el que el desempeño solo dependa

de la información de los patrones.

Una vez establecido el punto de trabajo adecuado del LDA se propuso evaluar la hipótesis de mejorar el desempeño de este utilizando un conjunto clasificadores en paralelo, denominado *ensamble de clasificadores*, los cuales se pueden relacionar de distintas maneras así como también combinar las respectivas salidas.

Se pueden mencionar varias razones para la utilización de un ensamble de clasificadores, entre las cuales se destacan razones de estadística, grandes volúmenes de datos, muy pocos datos disponibles, utilización de estrategias del tipo *divide y vencerás*, entre otras [50].

Los ensambles están formados por dos componentes: a) en primer lugar se requiere una estrategia para construir un ensamble tan diverso como sea posible, para ello se utilizan estrategias de combinación de clasificadores tales como fusión y selección de clasificadores, y b) es necesaria una estrategia para combinar las salidas de los clasificadores individuales que conforman el ensamble de tal manera que las decisiones correctas sean amplificadas, y las incorrectas tiendan a ser anuladas [50].

A continuación se describen las estrategias más usuales para combinar clasificadores individuales y para obtener la salida de un ensamble.

4.3.1. Estrategias de combinación de salidas

- a) Votación simple: se realiza un conteo de la cantidad de votos recibidos para cada clase por parte de cada clasificador, y la salida del ensamble es la clase que fue más votada.
- b) Votación ponderada: cada clasificador recibe un determinado peso en la votación final en función de su desempeño durante el proceso de entrenamiento.
- c) Regla del mínimo/máximo/mediana: como su nombre lo indica, son funciones simplemente para tomar la mínima, máxima o mediana entre las salidas continuas de los clasificadores individuales.
- d) Regla del producto: se multiplican los valores continuos de salidas de cada clasificador, esta regla es muy sensible a los clasificadores más pesimistas
- e) Regla de la media: el apoyo a la clase C_j se obtiene como la media de todas las salidas continuas de los clasificadores. Ésta es equivalente a la regla de la suma por un factor de normalización $1/N$, siendo N la cantidad de clasificadores.

4.3.2. Estrategias de construcción

- a) Bagging: es la abreviatura de *bootstrap aggregating* y es uno de los primeros algoritmos basados en ensamble. Este método obtiene la diversidad mediante el uso de réplicas bootstrap de los datos de entrenamiento, o sea, para cada subconjunto de datos de entrenamiento se utiliza un clasificador diferente del mismo tipo. Cada clasificador se combina a través de una votación de mayoría de sus decisiones, en este caso la clase elegida por la mayoría de los clasificadores es la decisión de conjunto.
- b) Boosting: en esta estrategia se crean 3 clasificadores débiles: un primer clasificador (C1) se entrena con un subconjunto aleatorio de los datos de entrenamiento disponibles. El subconjunto de datos de entrenamiento para el segundo clasificador (C2) es elegido como el subconjunto más informativo, dado C1. Es decir, C2 está entrenado con una base de datos de entrenamiento donde sólo la mitad de los patrones se han clasificado correctamente por C1. El tercer clasificador (C3) se entrena con los patrones en los que no están de acuerdo C1 y C2. La salida del ensamble se obtiene mediante una mayoría de votación [50].
- c) K-means: esta es una estrategia de selección alternativa de los grupos de entrenamiento, donde se separa el conjunto de patrones de entrenamiento en un número determinado subconjuntos, procurando tener una distribución particular para cada clasificador. De esta manera cada clasificador se especializa en un determinada área local del espacio total de características [50] [51].

Se realizó el análisis preliminar de desempeño de distintas configuraciones de ensambles de LDA y se los contrastó con el obtenido por un LDA de manera individual, las mejoras obtenidas no justifican su utilización en el marco de los experimentos definitivos de esta tesis, lo cual se detalla en la Sección 5.5.6.

4.4. Comentarios de cierre de capítulo

En este capítulo se describió el discriminante lineal de Fisher, que si bien es un clasificador de fácil implementación y de baja complejidad, es uno de los más utilizados en el ámbito de las ICC con buenos desempeños. Dado que en la tesis el objetivo es evaluar diversas estrategias de representación de señales combinadas

con técnicas de reducción de dimensionalidad, no se evaluaron otros clasificadores como alternativas a este.

Capítulo 5

Materiales y métodos

En este capítulo se describen los materiales y métodos utilizados para llevar adelante los experimentos que se describirán en el Capítulo 6 así como también algunos resultados preliminares, que sumado a lo anterior permitirán definir la metodología de los experimentos finales.

5.1. Base de datos

En el área de interfaces cerebro computadoras, hay grupos que generan sus propios registros de electroencefalograma y/o potenciales evocados así como también hay bases de datos disponibles para poder contrastar resultados. Algunas de las más conocidas son la generada por el BCI Laboratory of the Wadsworth Center, New York State Department of Health (EEUU) [52] y la generada por el Laboratorio de Neuroimagenología (LINI) del Departamento de Ingeniería Eléctrica de la Universidad Autónoma Metropolitana (México) [53] [54]. Ambas contienen registros de ERP obtenidos con el software BCI2000 [55] a partir de su aplicación P300 Speller basada en el deletreador propuesto por Donchin y Farwell [56]. EL software BCI2000 es una de las plataformas disponibles que permiten diseñar, evaluar y usar interfaces cerebro computadoras, al igual que otras como por ejemplo OpenVibe [57]. Sin embargo, una diferencia importante entre ambas bases de datos es la cantidad de sujetos disponible en cada una, mientras que la base de datos del Wadsworth Center contiene registros de solo 2 sujetos la del LINI contiene de 18 sujetos. Por este motivo, para poder realizar un análisis estadístico de los resultados que permita generalizar mejor se optó por utilizar esta última para el desarrollo de la tesis, y en consecuencia se describe a continuación.

Para la adquisición de los registros se utilizó un amplificador de 16 canales marca *g.tec*® modelo gUSBamp, con el cual se registraron simultáneamente 10 canales de EEG en las posiciones del sistema 10-20 indicadas en color verde en la Figura 5.1, el electrodo de referencia se ubicó en la posición A2 (lóbulo de la oreja derecha) y el de tierra en M2 (apófisis mastoidea del mismo lado). La base de datos está conformada por registros de 18 personas sanas, de las que se obtuvieron registros muestreados a una frecuencia de muestreo de 256 Hz y se aplicó un filtro Chebyshev pasa banda en 0,1-60 Hz, de orden 8 y un filtro notch a 60 Hz durante la adquisición.

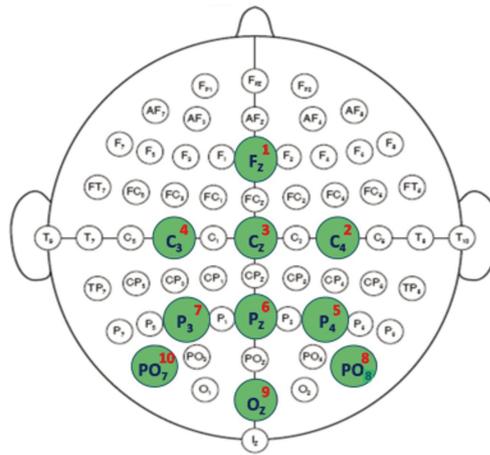


Figura 5.1: Electrodo utilizados para la obtención de los registros de EEG [3].

La adquisición de los registros se realizó utilizando el paradigma descrito en la Sección 1.2.1, en la Figura 5.2 se muestra un registro single trial de EEG filtrado correspondiente al canal Fz de uno de los sujetos de la base de datos. Las líneas verticales indican el comienzo de estimulación para un bloque completo, las líneas claras indican aquellas épocas en donde no se producirán ERP mientras que las oscuras indican las que si producirán ERP. El tiempo inter-estímulo es de 125 ms.

Los registros se llevaron a cabo en cuatro fases, una fase de *deletreo dirigido*, donde se le indicó a la persona las palabras a deletrear (CALOR, CARINO, SUSHI) de 15 secuencias cada una. Una segunda fase de *deletreo dirigido con matriz de clasificación*, se realizó una única corrida de deletreo dirigido con 15 secuencias para la palabra SUSHI. Las tercera y cuarta fase fueron de *deletreo libre* donde el sujeto elegía que palabra a deletrear. Esta última se realizó con menos secuencias de intensificación (iluminación de fila o columna), dependiendo

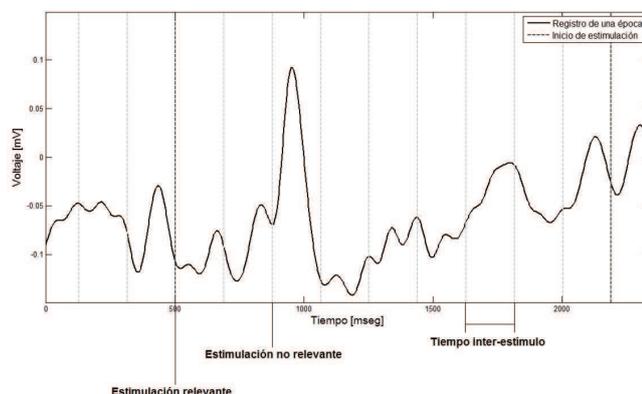


Figura 5.2: Registro época única *single trial* de EEG donde se indica los inicios de estimulación representados por líneas verticales con un tiempo inter estímulo de 125 ms.

del nivel de cansancio, disponibilidad de tiempo y desempeño del sujeto a lo largo de la prueba [3]. De todas estas fases, en esta tesis sólo se utilizaron las dos primeras que corresponden a las fases de deletreo dirigido, ya que las mismas poseen las etiquetas correspondientes a épocas con ERP (de ahora en adelante se denominarán *con P300*) o épocas sin respuesta. Teniendo en cuenta que las sesiones se realizaron con una matriz de 6x6 donde el caracter deseado se intensifica dos veces por cada una de las 15 secuencias, y teniendo un total de 21 caracteres a deletrear (CALOR, CARINO, SUSHI, SUSHI), para las dos primeras sesiones se dispone de 3780 registros *single trial* por canal, donde 630 contienen P300 y el resto son registros sin respuesta o de línea base. La forma en la que fueron adquiridos los datos se muestra en la Figura 5.3, donde puede observarse la colocación de la gorra de electrodos mientras es estimulado por la intensificación de las filas o columnas de la pantalla.

En una primera etapa de acondicionamiento se filtraron todos los registros de la base de datos con un filtro FIR pasa bajos de orden 30, la frecuencia de corte del filtro se fijó en 15 Hz [58] y se aplicó a cada época de EEG de 1 seg. de duración. Finalmente se normalizaron los registros de cada canal de manera global utilizando el valor medio y la desviación estándar del total de las épocas adquiridas en cada estimulación.

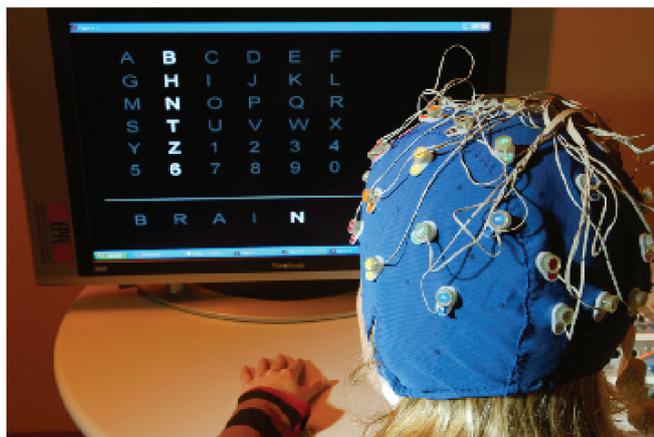


Figura 5.3: Montaje para la adquisición de los registros EEG, en el cual se destaca la gorra con electrodos ubicados en posiciones predefinidas del sistema 10-20 así como también la matriz para estimulación visual que posee la plataforma BCI2000.

5.2. Conformación de patrones

Retomando la idea expresada en la Sección [1.3](#) de pensar una ICC como un sistema de reconocimiento de patrones, de ahora en adelante se hará énfasis en este enfoque y en consecuencia a las señales de entrada al clasificador se las denominará *patrones*.

De acuerdo a lo descrito en la Sección [5.1](#), para cada intensificación se adquiere 1 seg. de señal de EEG en 10 canales simultáneos utilizando una frecuencia de muestreo de 256 Hz, los que posteriormente se submuestrearon a 64 Hz para reducir la dimensionalidad de la señal registrada.

En consecuencia, se dispone de 10 señales discretas (o vectores) de 64 muestras cada una, de ahora en adelante a las muestras se las denominarán *características*. Sin embargo, no entran al clasificador cada uno de estos vectores por separado, sino que se concatenan uno a continuación del otro para armar un solo patrón de entrada. Este proceso se describe en la Figura [5.4](#), y a los patrones conformados de esta manera se los denominará *patrones temporales*.

Aplicando este proceso sobre el total registros disponibles por cada sujeto se obtiene una matriz de 3780 filas por 640 columnas, las cuales se ordenaron según el tipo de respuesta: las primeras 630 filas poseen P300 y las siguientes 3150 no poseen P300, la Figura [5.5](#) esquematiza la estructura de *datos temporales* generada.

Entonces, a partir de la estructura de datos temporales se generó un nuevo

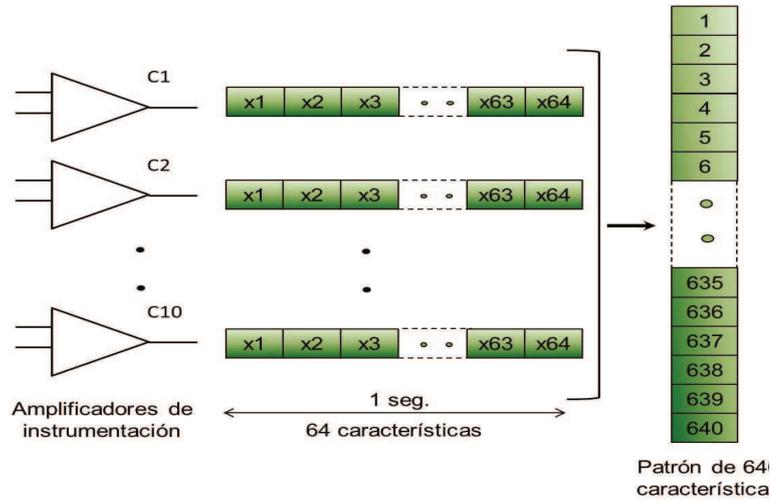


Figura 5.4: Conformación del patrón temporal (64 características) que entra al clasificador, a partir de la concatenación de los vectores individuales (64 características) de cada canal.

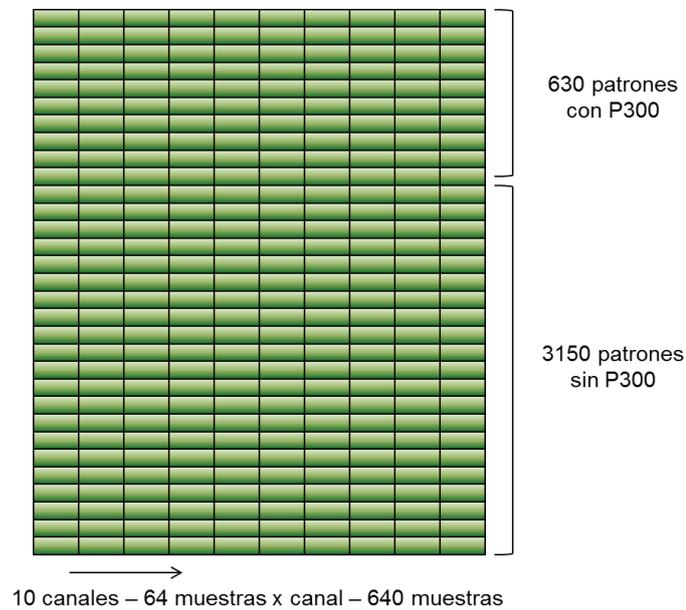


Figura 5.5: Matriz de datos formada por los patrones temporales generada por cada sujeto de la base de datos, cada división en las filas corresponde a una época adquirida en cada uno de los 10 canales.

conjunto de patrones mediante una representación alternativa con el objetivo de facilitar la detección de la información de interés, en particular la existencia de la onda P300; para lo cual se utilizaron la Transformada Wavelet Diádica Discreta y la Transformada Wavelet Packets descritas en el Capítulo 2. Este proceso se realizó sobre cada uno de los 10 canales de forma separada a través de la Ecuación 5.1:

$$\mathbf{a} = \Phi^{-1} \mathbf{x}, \quad (5.1)$$

donde \mathbf{x} representa el vector de 64 características temporales generado por cada canal, Φ es la matriz de cambio de base cuyas columnas contienen los átomos obtenidos por aplicación de DDWT o WPT y \mathbf{a} es el respectivo vector de coeficientes obtenidos. De manera análoga a la estructura de los patrones temporales, se concatenan los vectores de coeficientes de cada canal para generar un patrón de 640 características, en este caso de coeficientes Wavelets, este proceso se muestra en la Figura 5.6. A los patrones generados de esta manera se los denominará de ahora en adelante *patrones DDWT* (si Φ corresponde a la base DDWT) y *patrones WPT* (si Φ corresponde a la base wavelet packet seleccionada por alguno de los algoritmos descritos en la Sección 2.4). En ambos casos se genera una matriz de patrones similar a la mostrada en la Figura 5.5.

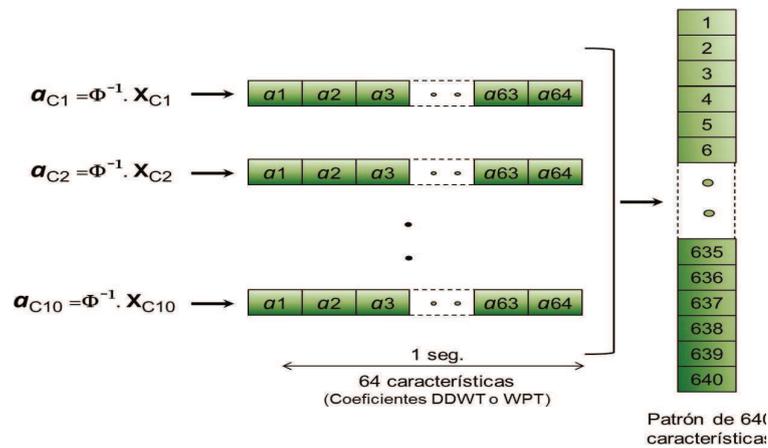


Figura 5.6: Conformación del patrón wavelet DDWT o WPT (640 características) que entra al clasificador, a partir de la concatenación de los vectores transformados (64 características) de cada canal.

En todos los experimentos realizados en esta tesis se utilizó el software Matlab v7.2 (MathWorks Inc.), en particular se utilizó la librería PRTools para la implementación del discriminante lineal [59], la librería Wavelab850 para la descompo-

sición WPT [60] y la librería MLDB7 para la implementación del algoritmo LDB [34].

5.3. Índices de desempeño

Dado que en la tesis se aborda un problema de clasificación, entonces resulta natural definir indicadores que evalúen que tan buenos son los métodos implementados para clasificar, y consecuentemente detectar, correctamente las clases involucradas en el problema en cuestión. Sin embargo, dado que también es importante la potencial implementación hardware de los mismos, se definió un criterio secundario que indique la cantidad de información necesaria para lograr el desempeño.

Los índices de desempeño que se utilizaron son:

- a) Tasa de aciertos: es una métrica que indica el promedio de todas los casos bien clasificados y esta dado por la siguiente Ecuación.

$$Tasa\ de\ acierto = \frac{patrones\ bien\ clasificados}{total\ de\ patrones}$$

- b) Sensibilidad: es una métrica que indica la capacidad del clasificador para reconocer como *casos positivos* los registros que poseen P300, en consecuencia es la capacidad para detectar los ERP.

$$Sensibilidad = \frac{vp}{vp + fn}$$

donde vp representa los verdaderos positivo (registros con P300) y fn los falsos negativo.

- c) Especificidad: es una métrica que indica la capacidad del clasificador para reconocer como *casos negativos* los registros que no poseen P300, es decir la proporción de épocas sin P300 correctamente identificadas.

$$Especificidad = \frac{vn}{vn + fp}$$

donde vn representa los verdaderos negativo (registros sin P300) y fp los falsos positivo.

- d) Cantidad de características: en caso de obtener desempeños similares basados en los índices descriptos previamente se utilizará la cantidad de muestras de los patrones resultantes de cada método.

5.4. Entrenamiento y prueba del clasificador

Considerando la estructura de datos de la Figura 5.5 se realizó una partición de los patrones (temporales y Wavelets) de cada uno de los 18 sujetos en la base de datos, de las 3780 épocas disponibles se separaron 2880 épocas para entrenamiento (480 con P300 y 2400 sin P300) y 900 épocas para el testeo (150 con P300 y 750 sin P300). Esta partición se utilizó para obtener algunos resultados preliminares que se presentarán a continuación así como también para todos los experimentos que se describirán en el Capítulo 6, este esquema permite comparar los resultados obtenidos con cada una de las estrategias planteadas.

Todos los experimentos contemplaron una etapa de entrenamiento y una de prueba o testeo, las cuales se esquematizan en la Figura 5.7 y se describen a continuación:

- a) Entrenamiento del LDA: en esta etapa se utilizaron los patrones de la partición de entrenamiento descrita anteriormente, estos pueden ser temporales, DDWT o WPT. En todos los casos, excepto para los patrones WPT, estos pueden ser resultado de aplicar alguno de los métodos de selección de características.
- b) Prueba: una vez entrenado el LDA se procede a la prueba del mismo utilizando los patrones de testeo, en esta etapa se obtienen los índices de desempeño (sensibilidad, especificidad, tasa de aciertos) que permiten comparar los desempeños entre todas las estrategias implementadas, así como también la cantidad de características resultantes.

5.5. Experimentos preliminares

En esta sección se presentan una serie de resultados parciales, o preliminares, que permiten definir la metodología para llevar adelante la propuesta de experimentos de esta tesis.

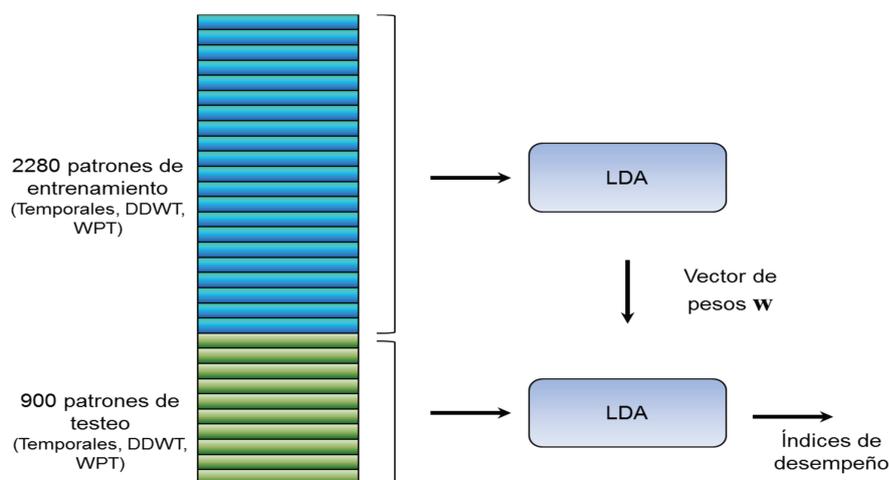


Figura 5.7: Descripción esquemática de los experimentos, donde se muestra la etapa de entrenamiento y la de prueba.

5.5.1. Elección de wavelet madre para generación de patrones DDWT y WPT

En la Sección [5.2](#) se describió como generar los patrones DDWT y WPT a través de una transformación Φ sin hacer referencia a una Wavelet madre en particular, por lo que aquí se describirá como se realizó esta elección para conformar los patrones que se utilizarán posteriormente.

La matriz Φ define el tipo de patrones a generar, ya sean DDWT o WPT, para lo cual es necesario seleccionar una Wavelet madre. En ambas transformaciones se utilizó la Wavelet *Daubechies 4* (*db4* en adelante), esta decisión inicialmente se basó en el análisis de publicaciones en las cuales se sustenta su elección en base a la similitud morfológica entre la onda P300 y la Wavelet *db4* [\[61–63\]](#).

Si bien esto podría considerarse suficiente para justificar esta elección, y teniendo en cuenta que en el problema planteado la morfología de la onda P300 queda enmascarada por el ruido en registros de época única; se evaluaron wavelets de las familias *Daubechies* (1 a 20), *Symlets* (2 a 20), *Coiflets* (1 a 5), *Meyer* y *Biortogonal* (1.1 a 6.8), lo cual permitió generar resultados propios para fortalecer la decisión tomada. En la Figura [5.8](#) se muestran algunos ejemplos de las Wavelets evaluadas.

Para simplificar la tarea solo se trabajó con registros de un solo canal en lugar de los 10 disponibles, puntualmente con los registros del canal Oz, dado que es uno en los cuales mejor se manifiesta la onda P300; tal como se refleja en los

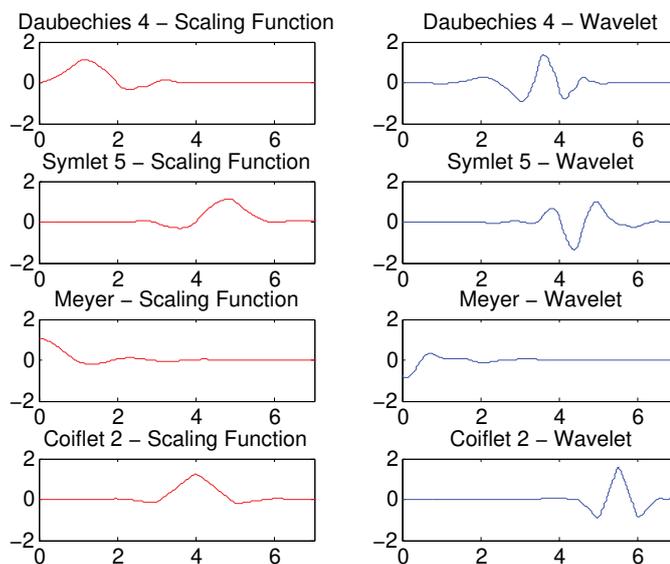


Figura 5.8: Ejemplos de distintas funciones Wavelets que se evaluaron.

escalogramas de los patrones temporales con P300 y sin P300 de los canales Cz y Oz mostrados en la Figura 5.9. Las columnas 1 y 2 corresponden a registros con P300 y sin P300 respectivamente en Cz, las columnas 3 y 4 corresponden a registros con P300 y sin P300 respectivamente en Oz.

Para resaltar al menos un poco la morfología de la onda P300 se utilizaron patrones generados por promediación coherente de 2 épocas, se utilizó como clasificador un LDA y la tasa de aciertos como índice de desempeño. En concordancia con la bibliografía, los resultados obtenidos indicaron que las Wavelets de la familia Daubechies fueron las que produjeron los mejores resultados; más detalles se pueden encontrar en [64, 65].

Por lo tanto, para generar los patrones DDWT, y recordando el esquema mostrado en la Figura 5.6, en cada canal se aplicó el algoritmo de la DDWT hasta el nivel de descomposición 6 utilizando la Wavelet *db4*; generando el árbol de descomposición mostrado en la Figura 5.10. En consecuencia se generó un vector de 64 características formado por los coeficientes wavelets de los *Detalles* 1 a 6 y de la *Aproximación* 6.

En la Figura 5.11 se muestra los elementos de la base generada a partir de la DDWT utilizando la wavelet *db 4*, en la cual se grafican los elementos de la *Aproximación* 6 (azul), *Detalle* 6 (rojo), *Detalle* 5 (verde), *Detalle* 4 (violeta), *Detalle* 3 (cyan), *Detalle* 2 (negro) y *Detalle* 1 (marrón).

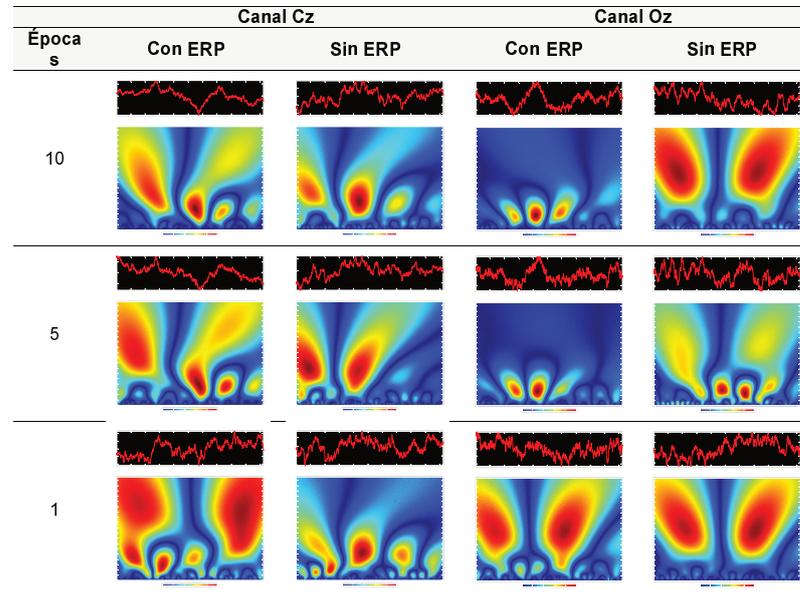


Figura 5.9: Escalogramas obtenidos sobre registros single trial, promedios de 5 y de 10 épocas obtenidos en los canales Cz y Oz.

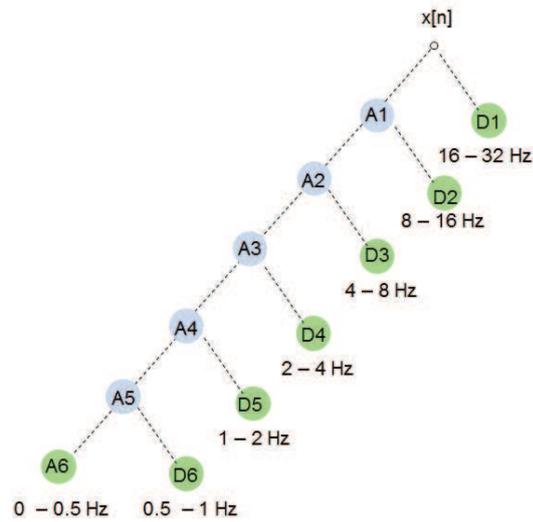


Figura 5.10: Descomposición multiresolución hasta el nivel 6 de una señal $x[n]$ muestreada a 64 Hz, en el cual se observan los rangos frecuenciales que abarca cada nivel de descomposición. A_i : i -ésima aproximación, D_i : i -ésimo detalle, i : nivel de descomposición

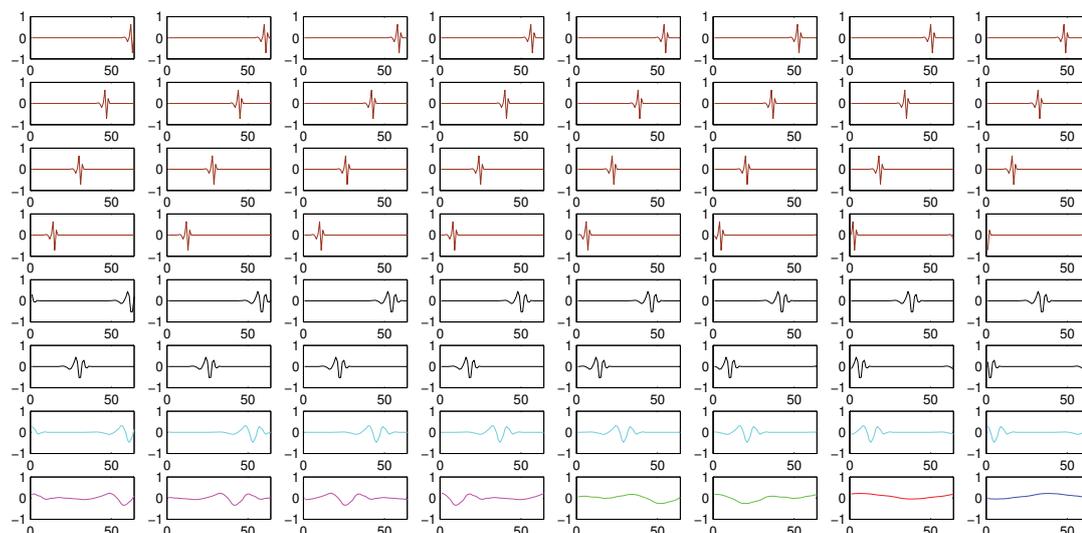


Figura 5.11: Base generada mediante Transformada Wavelet Diádica Discreta utilizando como Wavelet madre la Daubechies 4.

5.5.2. Mejoramiento de la relación señal a ruido en el dominio wavelet

En la Sección [1.1.2](#) se describió las dificultades presentes en el registro de la señal de ERP debido a la pobre RSR, por lo cual es deseable mejorarla y de esa manera propiciar un mejoramiento en el desempeño de la ICC. A partir de los patrones DDWT se pueden realizar acciones de *filtrado* de la señal con este propósito, las que se denominan *smoothing* y *denoising*, permitiendo generar un nuevo conjunto de patrones.

El smoothing, también denominado *suavizado*, es el proceso por el cual se *eliminan* los coeficientes DDWT correspondientes a las altas frecuencias (*detalles*) y se preservan los coeficientes DDWT asociados a las bajas frecuencias presentes en una señal (*aproximación*). Esta estrategia para mejorar la relación señal a ruido de una señal se basa en el conocimiento *a priori* del espectro de frecuencias del P300, en la Figura [5.12](#) se muestra que toda la energía de la señal se encuentra por debajo de los 15 Hz aproximadamente, por lo que los coeficientes Wavelets asociados a frecuencias superiores se consideran asociados al ruido y se eliminan [\[66\]](#).

El denoising es el proceso por el cual se *modifican* los coeficientes DDWT independientemente de su contenido frecuencial, hay varios métodos para aplicarlo y se basan en el cálculo de un umbral λ . Actualmente existe un gran número de

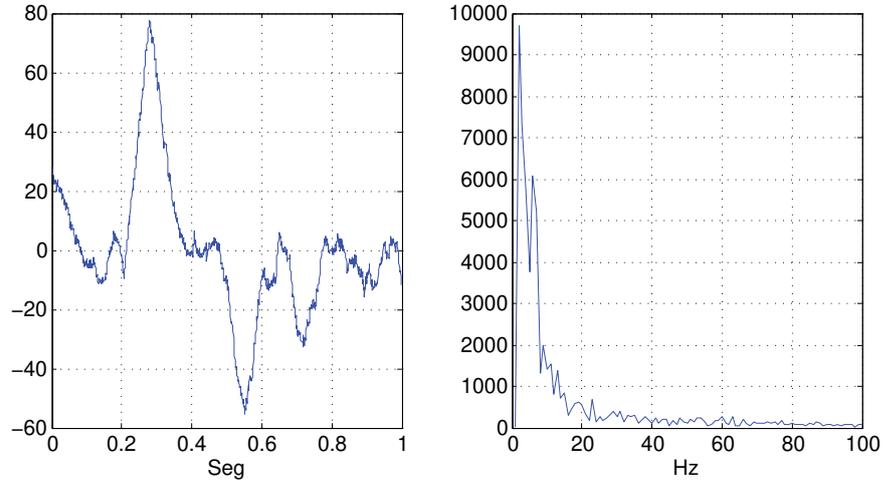


Figura 5.12: Señal de ERP en canal Cz y su respectivo espectro de frecuencias.

métodos para calcular el valor de λ , además se puede calcular un umbral para cada nivel de descomposición o sino calcular un umbral global para todos. Una vez obtenido λ se compara su valor con cada coeficiente wavelet de la descomposición multiresolución y se realiza una operación no lineal [67]. Las ecuaciones 5.2 y 5.3 corresponden a las operaciones que se realizan sobre cada coeficiente $a_{k,b}$ luego de ser comparado con el umbral, se denominan *hard* y *soft* respectivamente.

$$a_{k,b} = \begin{cases} 0 & \text{si } |a_{k,b}| < \lambda \\ a_{k,b} & \text{si } |a_{k,b}| \geq \lambda \end{cases} \quad (5.2)$$

$$a_{k,b} = \begin{cases} 0 & \text{si } |a_{k,b}| < \lambda \\ \text{sig}(ak, b)(|a_{k,b}| - \lambda) & \text{si } |a_{k,b}| \geq \lambda \end{cases} \quad (5.3)$$

El umbral puede ser *global* o *por nivel*, esto significa que puede calcularse un solo valor de λ para todos los coeficientes de la señal, o se pueden calcular λ_j umbrales, es decir uno para cada nivel de descomposición. A diferencia del suavizado, el denoising modifica los coeficientes Wavelet con el objetivo de eliminar las componentes de ruido presentes en ellos, la operación dará nuevos coeficientes Wavelet $z_{k,b}$, obtenidos de la siguiente manera:

$$z_{k,b} = \mathfrak{D}(a_{k,b}; \lambda), \quad (5.4)$$

en donde la operación de denoising se denota como $\mathfrak{D}(\cdot; \lambda)$, los nuevos coeficientes Wavelet obtenidos de aplicar el denoising se expresan como $z_{k,b}$, y son derivados de los coeficientes originales $a_{k,b}$.

Para calcular el umbral λ algunos métodos tienen en cuenta solamente la cantidad de muestras presentes en la señal; en cambio existen otros métodos que son adaptativos y tienen en cuenta el valor de los datos presentes en la señal para el cálculo. Independientemente del método que se elija, la determinación del umbral siempre requiere de una *estimación de ruido* presente en la señal; en [67] se propuso un método de *estimación robusta de ruido* en el dominio Wavelet, dada por la siguiente ecuación:

$$\theta = \frac{\text{mediana}(|a_{k,b}|)}{0,6745}, \quad (5.5)$$

donde los $a_{k,b}$ son los coeficientes pertenecientes a los niveles donde se supone que hay mas ruido. Una vez que se ha calculado θ el umbral λ que se aplicará sobre cada coeficiente será:

$$\lambda = \theta\lambda^*, \quad (5.6)$$

donde λ^* representa el umbral que se obtiene con el método elegido.

Los métodos para calcular λ^* se describen a continuación:

- a) Umbral Universal: este método sólo tiene en cuenta el número de coeficientes, se calcula según la siguiente ecuación:

$$\lambda^* = \sqrt{2\log(N)}, \quad (5.7)$$

siendo N el número de muestras de la señal si se trata de un umbral global, o el número de coeficientes del nivel, si se trata de un umbral por nivel. Como puede verse el umbral aumenta si la señal es más larga, esto es porque estos métodos suponen que la señal es coherente. Esto significa que la mayor parte de la energía puede representarse en unos pocos coeficientes Wavelet, entonces supone que a mayor número de muestras mayor ruido.

- b) Umbral SURE: el nombre proviene del inglés Stein Unbiased Risk Estimation, este método utiliza la información presente en la señal para calcular el valor del umbral, es un método de los llamados adaptativos, emplea una ecuación

de riesgo y toma la raíz cuadrada de valor que la minimiza. En general, los valores de umbrales obtenidos con este método son menores a los obtenidos con el método Universal.

Tanto el denoising como el suavizado son estrategias muy usadas en el ámbito del procesamiento digital de señales, sin embargo en experiencias realizadas durante el desarrollo de esta tesis se obtuvieron mejores resultados con este último [68]. En la Figura 5.13 se presentan los índices de desempeño, obtenidos sobre los 10 sujetos de la base de datos que presentan registros mas limpios, en la cual se muestra los resultados obtenidos mediante suavizado eliminando los coeficientes de los *detalles 1 y 2*, y denoising aplicando umbrales hard calculados por el método universal y por el método SURE.

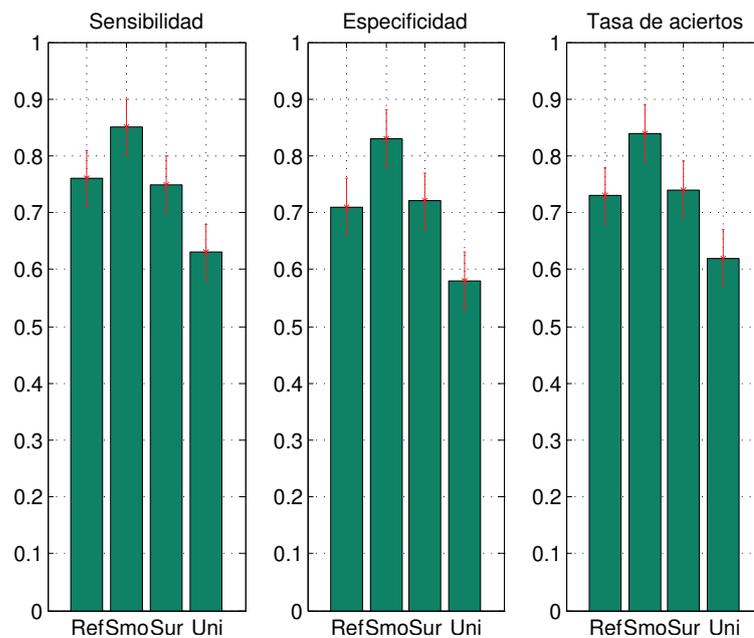


Figura 5.13: Índices de desempeño para cada método propuesto. Ref: patrones temporales de referencia, Smo: patrones DDWT con suavizado, Sur: patrones DDWT con denoising por umbral hard SURE, Uni: patrones DDWT con denoising por umbral hard Universal.

Una posible causa por la cual el suavizado tiene un mejor desempeño es que la estrategia de denoising asume como hipótesis que la energía de la señal de interés es superior a la del ruido, y por ese motivo se modifican los coeficientes cuyos valores están por debajo del umbral. En consecuencia, se utilizará la estrategia de suavizado para mejorar la RSR en los experimentos propuestos mas adelante.

5.5.3. Selección de medida de discriminación del algoritmo LDB para generación de patrones WPT

Para el caso de los patrones WPT generados por el algoritmo LDB, además de elegir una Wavelet madre también es necesario seleccionar cual es la medida de discriminación para obtener el mejor desempeño en el problema en cuestión.

El algoritmo LDB devuelve los vectores ϕ_i que conforman la base ortogonal (columnas de Φ), los coeficientes LDB asociados a cada vector base y sus respectivos valores de *scores*, un valor de score grande indica una mejor potencia de discriminación.

Para generar la base el algoritmo utiliza una medida de discriminación, por lo cual fue necesario evaluar el desempeño de cada una de las medidas descriptas en la Sección 2.4 en el contexto del problema abordado en esta tesis. Para esto se utilizó la base de datos descripta, se generó un conjunto de patrones WPT por cada medida de discriminación y se utilizó como índice de desempeño la tasa de aciertos de un discriminante lineal de Fisher. Según se muestra en la Figura 5.14 los resultados obtenidos no indicaron una superioridad de una medida por sobre otra, lo cual no permitió tomar la decisión sobre cual elegir.

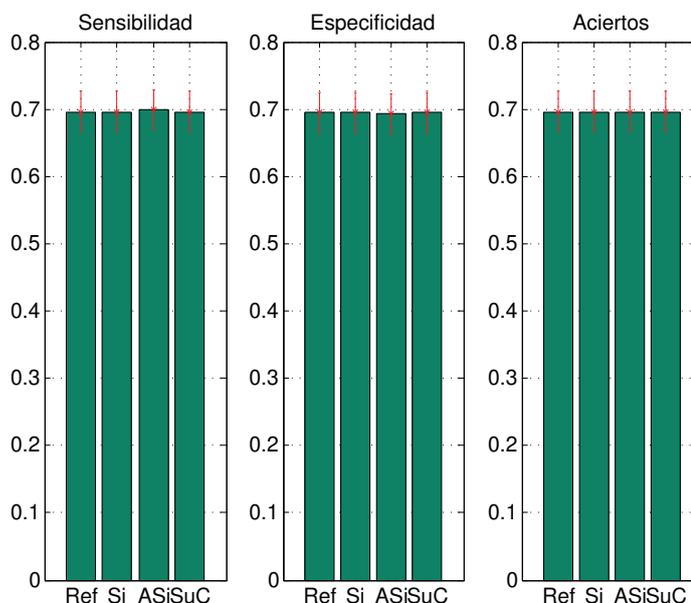


Figura 5.14: Resultados obtenidos al aplicar sobre todos los sujetos el algoritmo *local discriminat basis* utilizando las siguientes funciones de costo: simétrica (Si), asimétrica (ASi) y la suma de cuadrados (SuC).

Considerando esta situación, se evaluó el desempeño de las distintas medidas en función de la cantidad de características p de los patrones WPT, para lo cual previamente se analizó la evolución de la potencia de discriminación de los patrones generados en función de la cantidad de característica por canal sobre todos los sujetos de la base de datos. Esto se muestra en la Figura 5.15 donde en el eje de las abscisas se muestra la cantidad de características por canal y en el eje de las ordenadas el valor de discriminación en escala logarítmica.

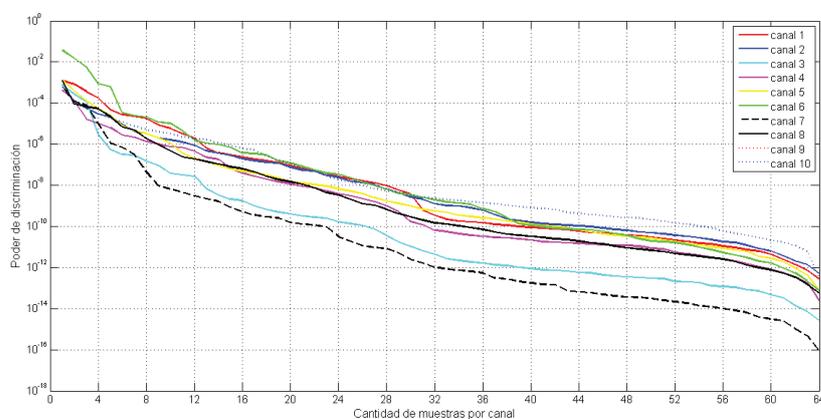


Figura 5.15: Evolución de la potencia de discriminación de cada coeficiente Wavelet Packet, las gráficas representan el promedio de todos los canales sobre todos los sujetos para seleccionar los valores de p . Tomado de [4]

Del análisis de la gráfica se desprende que los vectores más discriminativos se encuentran entre los 20 primeros vectores bases, sin embargo de acuerdo a la Figura 5.14 no hay diferencia en el poder de discriminación cuando se utilizan los 64 coeficientes por canal. Por lo tanto se analizó la tendencia en el desempeño de cada medida de discriminación en función de la cantidad de características, a manera de ejemplo en la Figura 5.16 se muestran los resultados obtenidos para los valores de p : 4, 6 y 8. En la misma se observa que a medida que aumenta el valor de p aumenta el poder de discriminación de todas las medidas, siendo superior en todos los casos la *suma de cuadrados*, por lo que se decidió utilizarla para generar los patrones WPT.

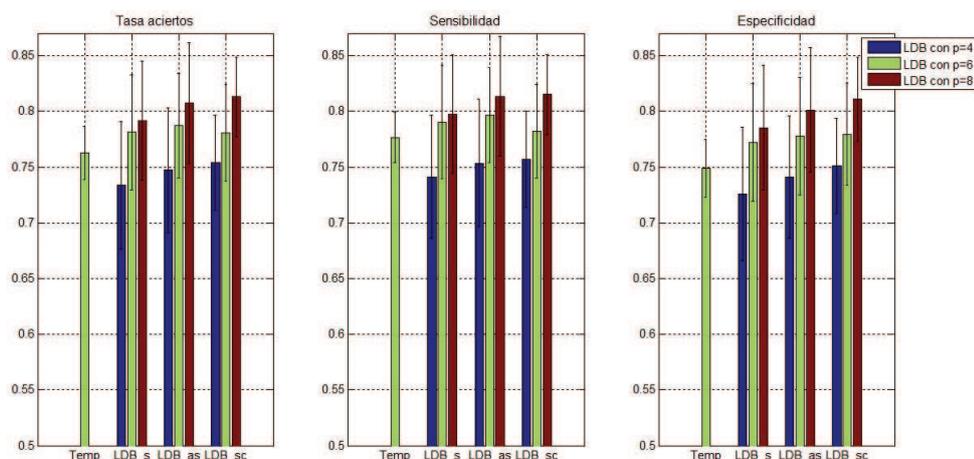


Figura 5.16: Resultados obtenidos con el algoritmo *local discriminant basis* utilizando las diferentes medidas de discriminación y para los distintos valores de $p=4, 6$ y 8 . LDB_s : LDB con medida simétrica, LDB_{as} : LDB con medida asimétrica y LDB_{sc} : LDB con suma de cuadrados.

5.5.4. Algoritmos de selección de bases a partir de un diccionario Wavelet Packets

En la Sección 2.4 se describieron los algoritmos BOB y LDB para generar los patrones WPT, considerando que en la sección anterior se definió la medida de discriminación del algoritmo LDB para generar una base ortogonal, entonces se está en condiciones de comparar el desempeño de los patrones generados por este algoritmo y los generados por el algoritmo BOB.

Para esto se generaron dos conjuntos de patrones, con una estructura similar a la mostrada en la Figura 5.5, a partir de aplicar cada uno de los algoritmos mencionados sobre el diccionario Wavelet Packets. La metodología utilizada para realizar la comparación es la misma de la sección anterior, es decir, un discriminante lineal de Fisher como clasificador y la tasa de aciertos, sensibilidad y especificidad como índices de desempeño.

En la Figura 5.17 se muestran los índices de desempeños obtenidos por los patrones generados por BOB y por los patrones generados por LDB utilizando solo 8 coeficientes. Es evidente que los patrones generados mediante BOB no producen mejoras con respecto a los patrones de referencia a diferencia de los patrones generados mediante LDB, por lo cual estos fueron descartados como una alternativa a evaluar en los experimentos del Capítulo 6.

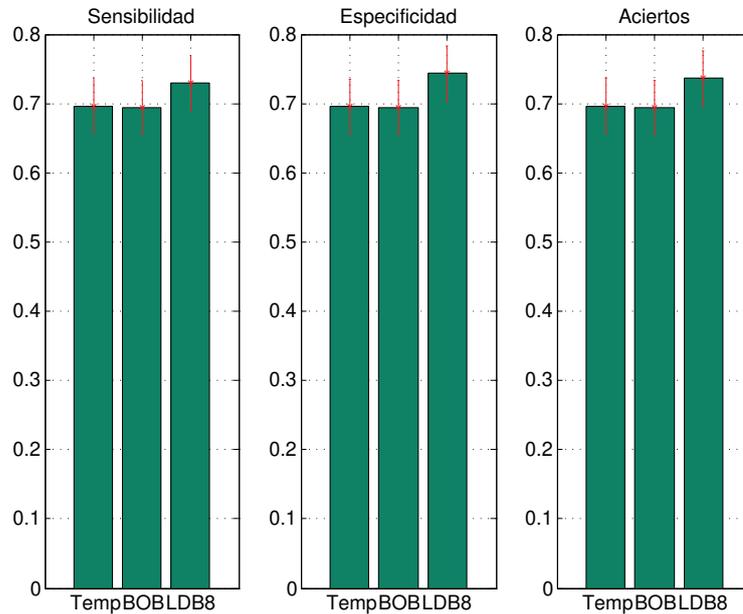


Figura 5.17: Índices de desempeño obtenidos con patrones obtenidos con el algoritmo best orthogonal basis (BOB) y con local discriminat basis utilizando 8 coeficientes (LDB8).

5.5.5. Determinación del punto óptimo de entrenamiento del LDA

El LDA es uno de los clasificadores más utilizados en aplicaciones en ICC [69, 70], sin embargo no se ha analizado el efecto de la variación en la cantidad de patrones de entrenamiento utilizando señales de EEG para poder determinar la cantidad necesaria para llegar a la meseta de rendimiento. Esto es el punto en que no hay ganancia significativa del rendimiento al agregar una mayor cantidad de patrones de entrenamiento, el cual se denominará *punto óptimo*, el poder determinarlo permite asumir que los resultados obtenidos al utilizar el LDA solo dependen de la información relativa al problema contenida en los patrones.

Para esto se utilizaron 1200 patrones temporales (600 de cada clase) y se aplicaron 3 tazas de submuestreo diferentes obteniéndose patrones con 80, 160 y 320 características (8, 16 y 32 características por canal respectivamente, 10 canales). Se utilizaron patrones *single trial* así como también generados por promediación de 2 a 5 épocas, esto permitió corroborar el efecto de la promediación coherente en la mejora de la RSR.

En la Figura [5.18](#) se muestra, desde arriba hacia abajo, la evolución del *área bajo la curva ROC* (AUC, del inglés Area Under the Curve) cuando se utilizan los patrones de 80, 160 y 320 características. En la misma puede observarse que cuanto mas grande es la cantidad de características que conforman el patrón el LDA tarda mas en llegar a la meseta de rendimiento, de hecho solo se alcanzó completamente para el caso de los patrones single trial con 80 características cerca de los 600 patrones de entrenamiento. En el caso de los patrones single trial con 160 características se observa una tendencia a la estabilización del *área bajo la curva* hacia los 1200 patrones de entrenamiento, y algo similar pasa en el caso de los patrones single trial de 320 características.

Además, puede observarse que el desempeño para el caso de patrones promediados aumenta en función de la cantidad de épocas promediadas; esto es lógico debido a que cuantas mas épocas se promedian mas se mejora la RSR y en consecuencia se facilita la detección de la señal de ERP. Sin embargo, en ninguno de los casos se alcanza la meseta de rendimiento como en el caso de los patrones single trial; para lograrlo es necesario disponer de mas patrones de entrenamiento.

Por lo tanto, asumiendo un comportamiento lineal de la evolución se puede suponer que “*para llegar al punto óptimo de trabajo del LDA se necesitan aproximadamente entre 6 y 8 patrones de entrenamiento por cada característica que los conforman*”. Es importante tener en cuenta este resultado al utilizar el LDA como herramienta de evaluación de desempeño, ya que alguna de las estrategias utilizadas en la extracción de características suelen modificar la cantidad de estas y en consecuencia afectar el rendimiento de la ICC sin que cambie la capacidad de discriminación intrínseca de dichas características. Más detalles de estos resultados fueron publicados en [\[71\]](#), [\[72\]](#).

5.5.6. Evaluación de ensambles de discriminantes lineales

Bajo la hipótesis de que un ensamble de LDAs podría mejorar el desempeño de un LDA único en la detección de ERP, se comparó el desempeño de este con los de 3 configuraciones distintas de ensamble formados por 3 LDA.

Para la realización del experimento se utilizó la base de datos del LINI, de la cual se seleccionaron 10 sujetos cuyos registros presentaban menor cantidad de artefactos con el fin de evitar que estas variables influyan en los resultados.

Como índice de desempeño se utilizó el *área debajo de la curva ROC*, dado que esta curva representa la tasa de verdaderos positivos frente a la tasa de falsos

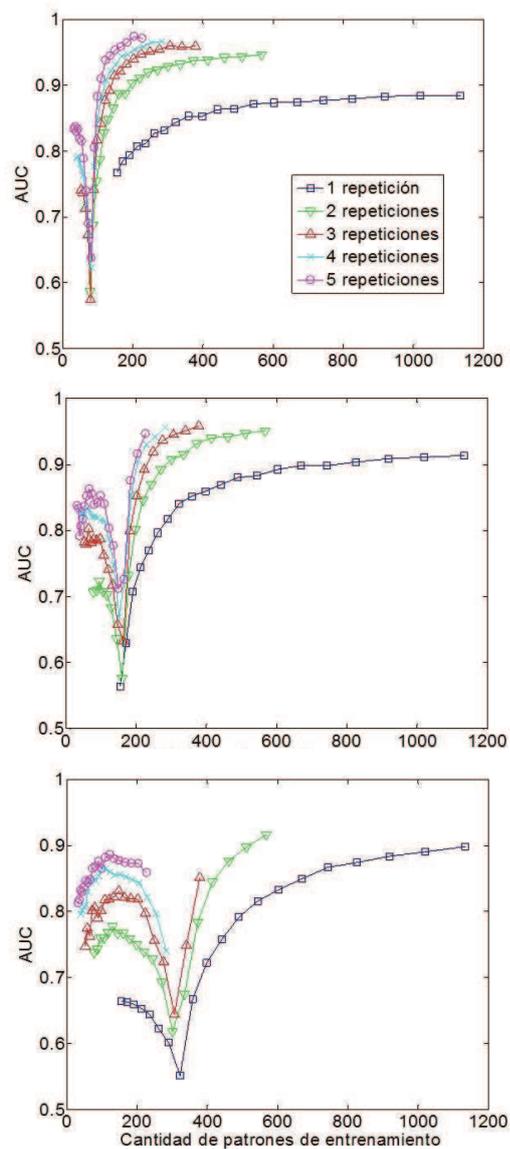


Figura 5.18: Área debajo la curva ROC vs. cantidad de patrones de entrenamiento. De arriba a abajo utilizando 80, 160 y 320 características en los patrones respectivamente. Las curvas en cada gráfica representan patrones formados utilizando diferente cantidad de promediaciones (1 a 5).

positivos, entonces cuanto mayor es el área debajo la curva mejor es el desempeño del clasificador.

Las estrategias de construcción de ensambles que se evaluaron fueron boosting, bagging y k-means; simultáneamente se evaluaron como estrategias para obtener la salida la votación simple, votación ponderada y las siguientes regla: producto, suma, suma ponderada, máximo, mínimo y mediana.

Los resultados obtenidos se muestran en la Figura 5.19 [73], se advierte que excepto el bagging ninguna de las estrategias de ensambles supera el desempeño logrado por el LDA único, por lo que se utilizará esta estrategia en los experimentos del Capítulo 6. Esto puede deberse a la falta de diversidad en los clasificadores incluidos en el conjunto y a las características de la distribución estadística de los patrones utilizados.

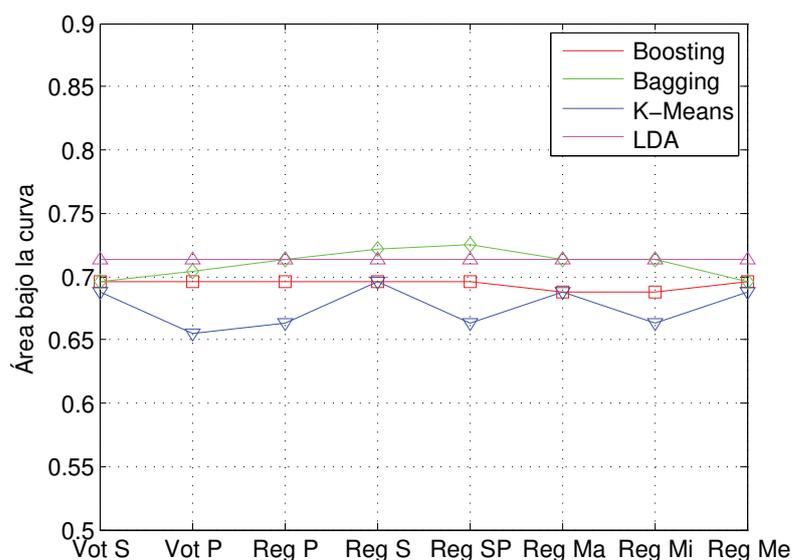


Figura 5.19: Resultados obtenidos con distintas combinaciones de estrategias de ensambles. Vot S: votación simple, Vot P: votación ponderada, Reg P; regla del producto, Reg S: regla de la suma, Reg SP: regla de suma ponderada, Reg Ma: regla de máximo, Reg Mi: regla del mínimo, Reg Me: regla de mediana.

5.6. Comentarios de cierre de capítulo

En este capítulo se describió la base de datos que se utilizó en la tesis así como también el proceso para generar los distintos tipos de patrones a partir de los temporales originales.

Además, se presentaron resultados de experimentos preliminares propuestos para definir claramente los experimentos del Capítulo 6. En primer lugar se justificó la elección de la Wavelet Daubechies 4 como Wavelet madre para generar tanto los patrones DDWT como los WPT, posteriormente se explicitó el motivo por el cual se utilizará el método de suavizado en lugar del método de denoising para mejorar la relación señal a ruido con el objetivo de mejorar la detección del P300. Luego se presentaron los resultados de dos experiencias que están relacionadas entre sí, por un lado la elección de la *suma de cuadrados* como medida de dispersión para generar los patrones WPT y por otro la comparación de los desempeños de patrones WPT generados mediante los algoritmos LDB y BOB, dado que estos últimos no produjeron mejoras con respecto a los de referencia no se considerarán en esta tesis. Relacionado con la etapa de clasificación, se determinó el punto de trabajo óptimo del discriminante de Fisher, puntualmente se obtuvo la relación entre la cantidad de patrones de entrenamiento y las dimensiones de estos, de manera que el desempeño dependa sólo de la información del problema presente en los patrones de entrenamiento. También se evaluó el desempeño de distintas estrategias de ensambles de clasificadores no obteniéndose resultados satisfactorios, en consecuencia se decidió utilizar solo un clasificador.

Capítulo 6

Experimentos y resultados

En este capítulo se describen los experimentos definitivos planteados para cumplir el objetivo propuesto en la tesis y se presentan los resultados obtenidos en cada uno de estos.

6.1. Descripción de experimentos y resultados obtenidos

En la Figura [6.1](#) se muestra un esquema que resume los experimentos realizados en el marco de esta tesis, aquí la señal \mathbf{x} representa los registros de EEG de N muestras o características y \mathbf{a} es el vector de coeficientes generados al aplicar la transformación Φ mediante la Ecuación $\mathbf{a} = \Phi^{-1} \cdot \mathbf{x}$.

La relación entre patrones involucrados en cada experimento están indicados con distintos colores: a) en azul se indican los patrones temporales, b) en rojo se indica la generación de patrones DDWT y WPT mediante la transformación Φ , c) en verde se indica la generación de patrones mediante suavizado de patrones DDWT, cuyo vector resultante es \mathbf{a}_K donde K es la dimensión del patrón y $K \leq N$ y d) en color amarillo se indica la generación de patrones mediante selección de características, ya sean aplicados a patrones temporales o DDWT suavizados. En el caso temporal el patrón resultante es \mathbf{x}_M , donde M es la dimensión del patrón y $M \leq N$, en el caso DDWT suavizado el patrón resultante es \mathbf{a}_L , donde L es la dimensión del patrón y $L \leq K \leq N$.

A continuación se describe en detalle la metodología empleada en cada experimento y los resultados que se obtuvieron en cada uno de ellos.

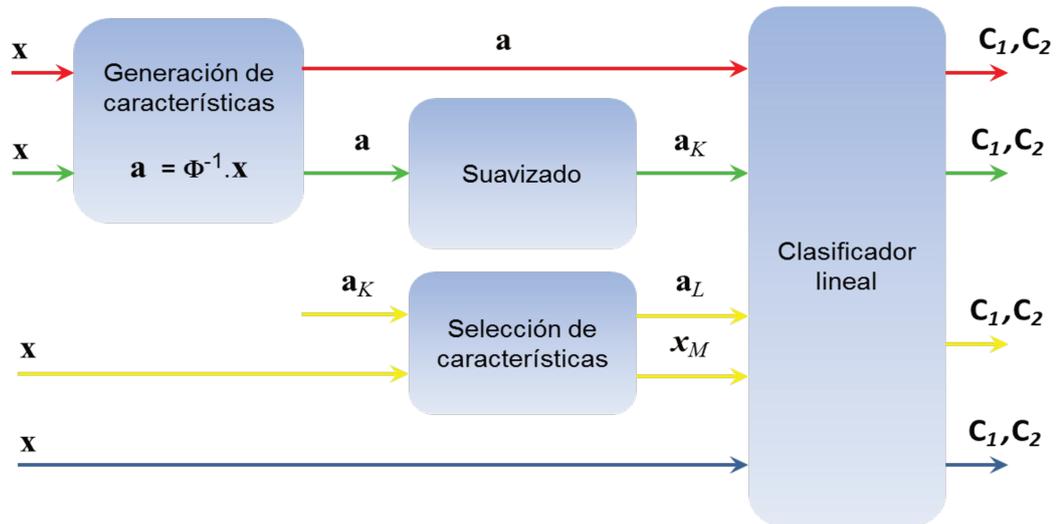


Figura 6.1: Resumen de los experimentos realizados: \mathbf{x} son los patrones temporales, \mathbf{a} son los patrones DDWT o WPT, \mathbf{a}_K son los patrones DDWT con suavizado, \mathbf{a}_L y \mathbf{x}_M son los patrones resultantes de la selección de características, $C_{1,2}$ son las clases posibles.

6.2. Experimento 1: índices de desempeño de referencia

En esta sección se determinaron los valores de los índices de desempeño correspondientes a los patrones temporales sin ningún tipo de modificación o procesamiento, en adelante estos se considerarán como *los índices de referencia* para comparar los desempeños del resto de las estrategias propuestas en la tesis.

En la Figura [6.2](#) se muestran los índices de desempeño obtenidos para cada uno de los sujetos de la base de datos. Se destaca la variabilidad que hay entre cada uno, lo cual depende de factores tales como el estado de atención del sujeto en el momento de los registros, las condiciones de instrumentación (por ejemplo impedancia de contacto de electrodos) entre otros.

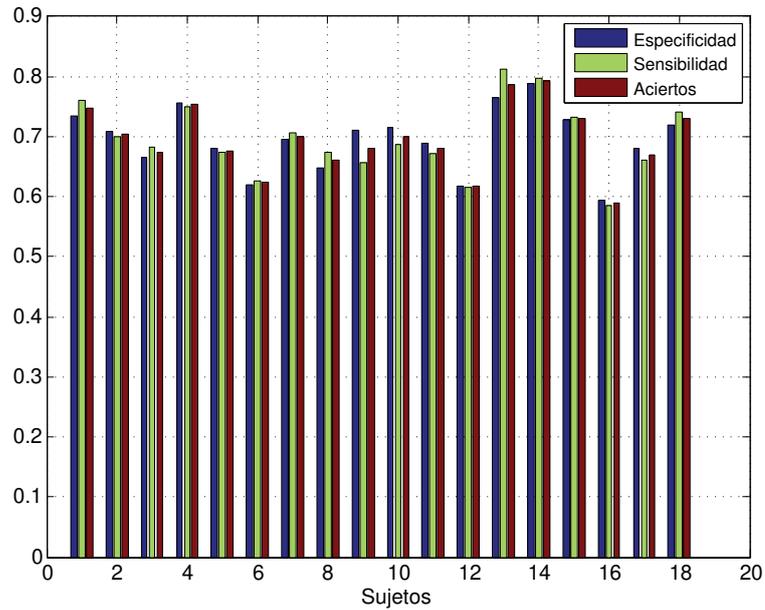


Figura 6.2: Índices de desempeño obtenidos en cada uno de los sujetos de la base de datos.

Los gráficos de caja (boxplots) de los índices de desempeño obtenidos sobre todos los sujetos de la base de datos se muestran en la Figura 6.3. La tasa de aciertos promedio fue de $0,695 \pm 0,05$, la sensibilidad promedio de $0,696 \pm 0,05$ y la especificidad promedio $0,695 \pm 0,05$. La longitud del patrón es de 640 características.

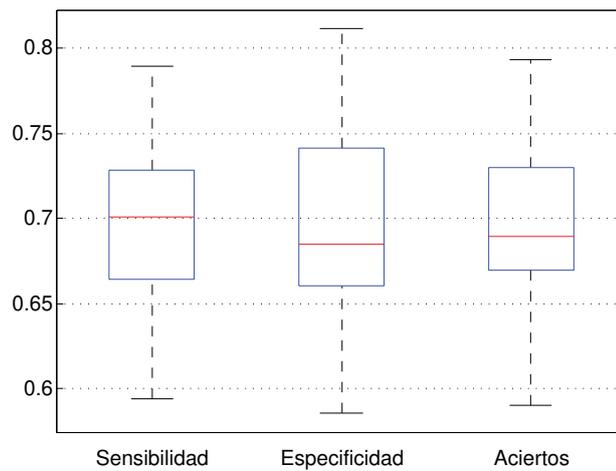


Figura 6.3: Índices de desempeño de referencia obtenidos sobre registros temporales.

6.3. Experimento 2: generación de características mediante bases ortogonales

En esta sección se obtuvieron los índices de desempeño utilizando patrones generados mediante DDWT y WPT de acuerdo a lo descrito en la Sección 5.2.

En el caso de los patrones DDWT se utilizó la Wavelet *db4* y se realizó la descomposición multiresolución hasta el nivel 6. Por cada uno de los canales se conformó el patrón con los coeficientes correspondientes al detalle 1 (32 coeficientes), detalle 2 (16 coeficientes), detalle 3 (8 coeficientes), detalle 4 (4 coeficientes), detalle 5 (2 coeficientes) y al detalle 6 y aproximación 6 (1 coeficiente cada uno), lo cual totaliza 64 coeficientes por canal y en consecuencia un patrón de 640 características.

En la Figura 6.4 se muestran los boxplots de los índices de desempeño. La tasa de aciertos promedio fue de $0,695 \pm 0,05$, la sensibilidad promedio de $0,695 \pm 0,05$ y la especificidad promedio $0,695 \pm 0,05$.

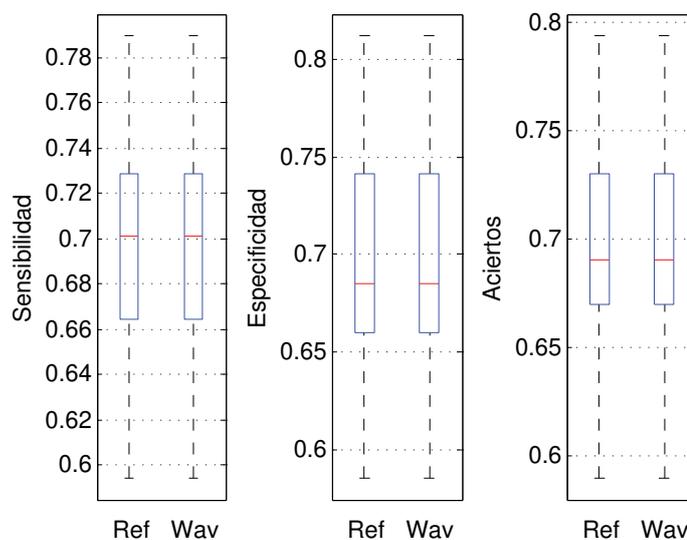


Figura 6.4: Índices de desempeño obtenidos con patrones DDWT

Los patrones WPT se generaron a partir de la descomposición Wavelet Packet utilizando la Wavelet *db4*, los coeficientes se obtuvieron mediante el algoritmo LDB utilizando como medida de discriminación la *suma de cuadrados*.

Dado que el algoritmo LDB devuelve los coeficientes asociados a los elementos de la base ordenados según su potencia de discriminación, fue posible generar patrones de diferentes longitudes e ir evaluando el desempeño de estos para de-

terminar si es necesario utilizar los 64 coeficientes disponibles por canal o si es posible con una cantidad menor obtener un desempeño igual o mejor, y de esta manera disminuir el tamaño del patrón final. Para verificar esto se conformaron patrones con longitudes que fueron variando de 2 en dos, desde 2 hasta 64, y se evaluó el desempeño para cada uno de estos conjuntos, en la Figura 6.5 se muestran la evolución de los índices de desempeño en función de la cantidad de coeficientes WPT por canal [4].

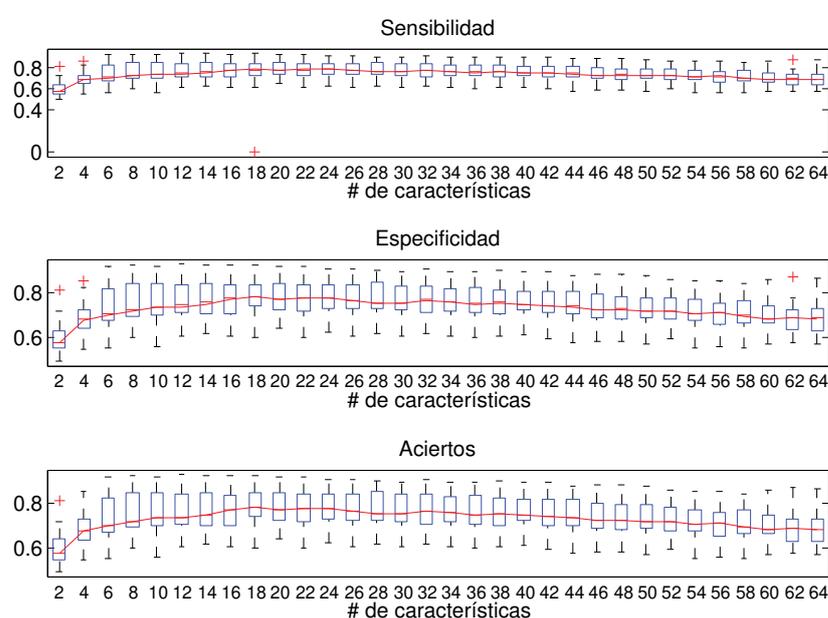


Figura 6.5: Evolución de los índices de desempeño en función de la cantidad de características de los patrones WPT

En la Tabla 6.1 se presentan los índices de desempeños obtenidos sobre todos los sujetos utilizando patrones WPT con diferentes cantidades de características, en negrita se remarcan los valores máximos correspondiente a 18 coeficientes; por lo cual se conformó el patrón WPT mediante concatenación de los 18 primeros coeficientes WPT de cada canal generando un patrón final de 180 características.

En la Figura 6.6 se muestran los boxplots de los índices de desempeño para los patrones generados con 18 coeficientes por canal. La tasa de aciertos promedio fue de $0,785 \pm 0,08$, la sensibilidad promedio de $0,803 \pm 0,08$ y la especificidad promedio $0,771 \pm 0,08$.

Tabla 6.1: Índices de desempeño de patrones WPT con distintas cantidades de características.

Características	Sensibilidad	Especificidad	Tasa de aciertos
2	0,60	0,60	0,60
4	0,69	0,69	0,69
6	0,72	0,72	0,72
8	0,76	0,76	0,76
10	0,76	0,76	0,76
12	0,77	0,77	0,77
14	0,77	0,77	0,77
16	0,78	0,78	0,78
18	0,79	0,79	0,79
20	0,78	0,78	0,78
22	0,78	0,78	0,78
24	0,78	0,78	0,78
26	0,78	0,78	0,78
28	0,78	0,78	0,78
30	0,77	0,77	0,77
32	0,77	0,77	0,77
34	0,76	0,76	0,76
36	0,76	0,76	0,76
38	0,76	0,76	0,76
40	0,75	0,75	0,75
42	0,75	0,75	0,75
44	0,74	0,74	0,74
46	0,73	0,73	0,73
48	0,73	0,73	0,73
50	0,72	0,72	0,72
52	0,72	0,72	0,72
54	0,71	0,71	0,71
56	0,71	0,71	0,71
58	0,71	0,71	0,71
60	0,70	0,70	0,70
62	0,69	0,69	0,69
64	0,69	0,69	0,69

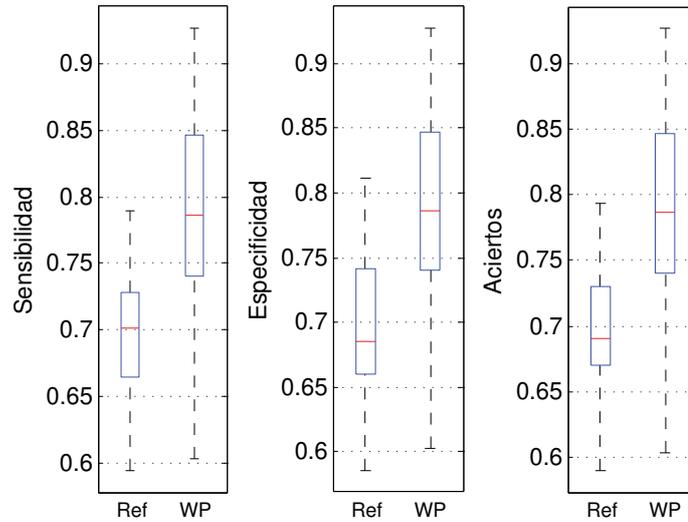


Figura 6.6: Índices de desempeño obtenidos con patrones WPT utilizando 18 coeficientes por canal

6.4. Experimento 3: suavizado de patrones DDWT

Dado que en la descomposición multiresolución se conoce el rango de frecuencias correspondiente a cada uno de los *detalles* y a la *aproximación* resultante, se procedió a eliminar los coeficientes de los detalles 1 y 2 tal como se muestra en la Figura 6.7. En el caso en que se eliminó el *detalle* 1 (rango de frecuencia de 16 - 32 Hz) se generaron patrones de 320 coeficientes (características) a los que se los denominan *patrones DDWT sin D1*; y en el caso en que se eliminó también el *detalle* 2 (rango de frecuencia de 8 - 16 Hz) los patrones generados son de 160 coeficientes y se los denomina *patrones DDWT sin D1 y D2*.

El proceso de suavizado puede pensarse como la proyección de la señal \vec{x} en una base ortogonal subcompleta, de manera tal que si se considera cada canal de forma individual se pasa la representación en un espacio \mathbb{R}^{64} a otro definido en \mathbb{R}^{32} o \mathbb{R}^{16} .

En la Figura 6.8 se muestran los boxplots de los índices de desempeño para los patrones DDWT sin D1. La tasa de aciertos promedio fue de $0,768 \pm 0,06$, la sensibilidad promedio de $0,770 \pm 0,06$ y la especificidad promedio $0,767 \pm 0,06$. La longitud del patrón resultante es de 320 características.

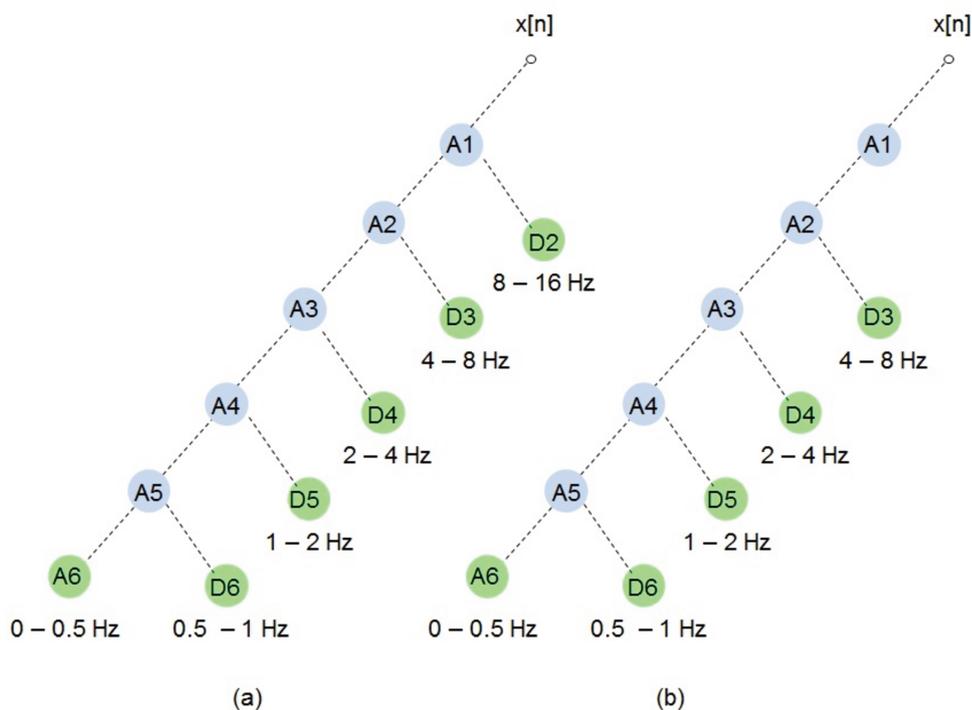


Figura 6.7: Árboles de descomposición multiresolución resultantes de aplicar smothing eliminando el *detalle* 1 (a) y los *detalles* 1 y 2 (b).

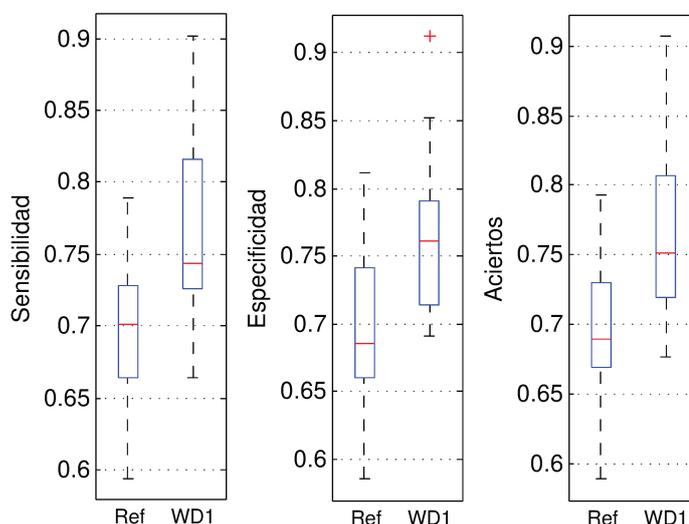


Figura 6.8: Índices de desempeño obtenidos con patrones DDWT sin el *detalle* 1.

En la Figura [6.9](#) se muestran los boxplots de los índices de desempeño para los patrones DDWT sin D1 y D2. La tasa de aciertos promedio fue de $0,747 \pm 0,05$, la sensibilidad promedio de $0,758 \pm 0,06$ y la especificidad promedio $0,739 \pm 0,05$.

La longitud del patrón resultante es de 160 características.

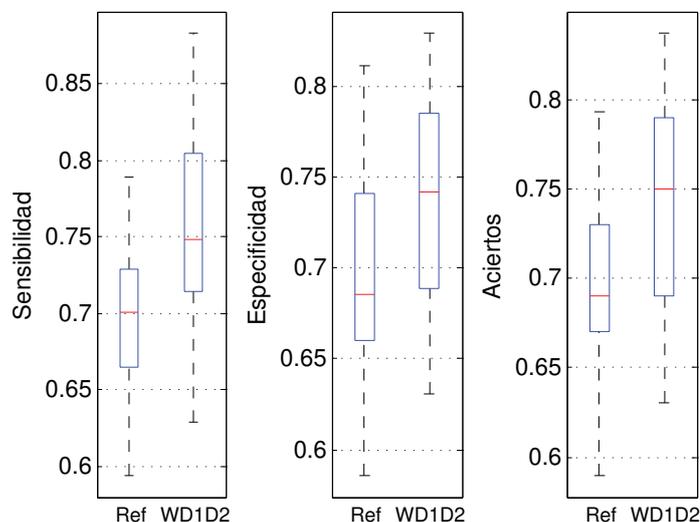


Figura 6.9: Índices de desempeño obtenidos con patrones DDWT sin los *detalles 1 y 2*.

6.5. Experimento 4: selección de características

En esta sección se evalúa el desempeño de las estrategias de eliminación de características descritas en el Capítulo 3: la *eliminación recursiva de características* (ERC) y los *algoritmos genéticos* (AGs).

6.5.1. Selección de características mediante eliminación recursiva de características

En la ERC se inicia el proceso con patrones de 640 características (temporales y DDWT), 320 características (DDWT sin D1) o 160 características (DDWT sin D1 y D2); en todos los casos la implementación del método es la siguiente:

1. Entrenamiento del LDA
2. Cálculo y almacenamiento de la tasa de acierto
3. Determinación del peso de conexión con menor valor absoluto
4. Eliminación del patrón de la característica asociada a ese peso

Este proceso se repite hasta que el patrón queda con una sola característica, finalmente se determina cual es el máximo valor de las tasas de aciertos y a que cantidad de características corresponde; este último es el que se asigna al sujeto en cuestión. A manera de ejemplo, en la Figura 6.10 se muestra la evolución de la tasa de aciertos a medida que se van eliminando características de patrones DDWT sin D1 y D2 (160 características). En este caso el máximo valor de tasa de acierto se da cuando se eliminan 91 características, por lo que la longitud del patrón resultante es de 69 características.

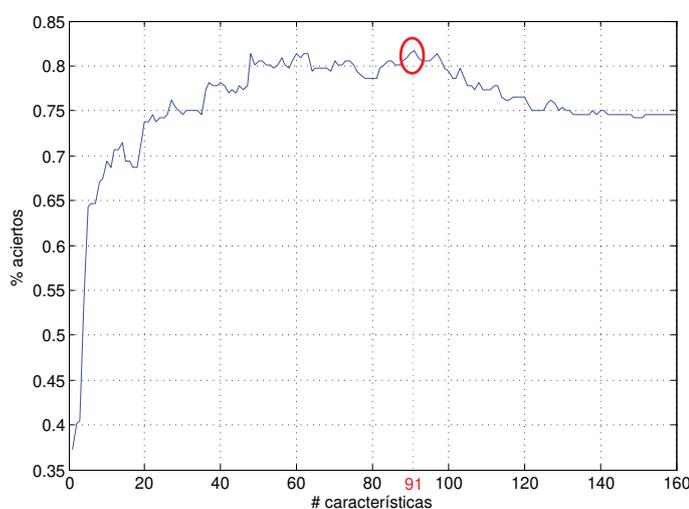


Figura 6.10: Evolución de la tasa de aciertos del algoritmo de *eliminación recursiva de características* a medida que estas se van eliminando.

- Eliminación recursiva de características aplicado a patrones temporales
En la Figura 6.11 se muestran los boxplots de los índices de desempeño obtenidos mediante aplicación sobre los patrones temporales del método integrado de selección de características; la tasa de aciertos promedio fue de $0,777 \pm 0,07$, la sensibilidad promedio de $0,768 \pm 0,08$ y la especificidad promedio $0,786 \pm 0,08$. En este caso la longitud promedio del patrón resultante es de 120 características.
- Eliminación recursiva de características aplicado a patrones DDWT sin D1
En la Figura 6.12 se muestran los boxplots de los índices de desempeño utilizados. La tasa de aciertos promedio fue de $0,778 \pm 0,06$, la sensibilidad promedio de $0,766 \pm 0,07$ y la especificidad promedio $0,789 \pm 0,07$. La longitud del patrón resultante es de 111 características.

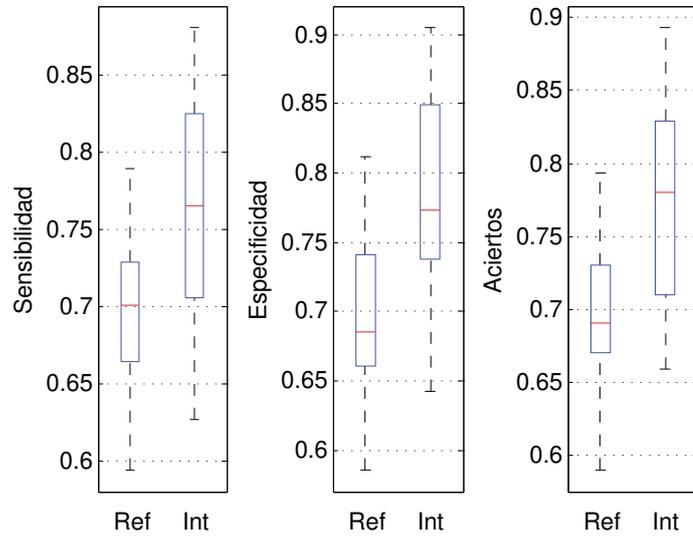


Figura 6.11: Índices de desempeño obtenidos con patrones temporales utilizando *eliminación recursiva de características*.

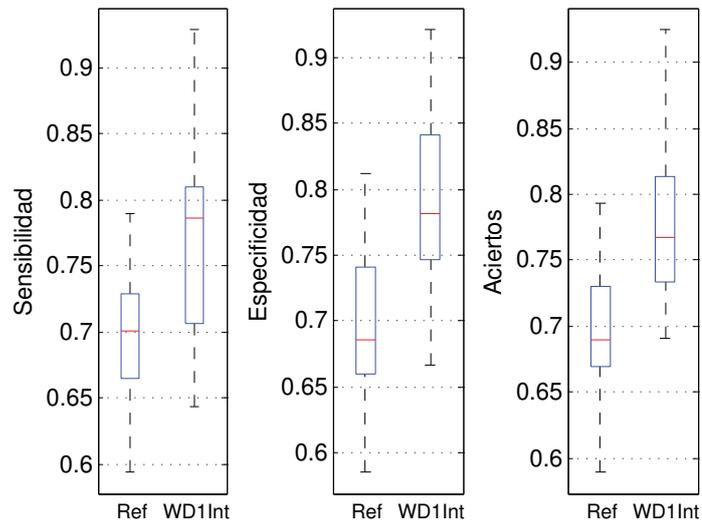


Figura 6.12: Índices de desempeño obtenidos con patrones DDWT sin el Detalle 1 utilizando *eliminación recursiva de características*.

- Eliminación recursiva de características aplicado a patrones DDWT sin D1 y D2

En la Figura 6.13 se muestran los boxplots de los índices de desempeño utilizados. La tasa de aciertos promedio fue de $0,738 \pm 0,06$, la sensibilidad promedio de $0,727 \pm 0,07$ y la especificidad promedio $0,749 \pm 0,08$. La longitud del patrón resultante es de 63 características.

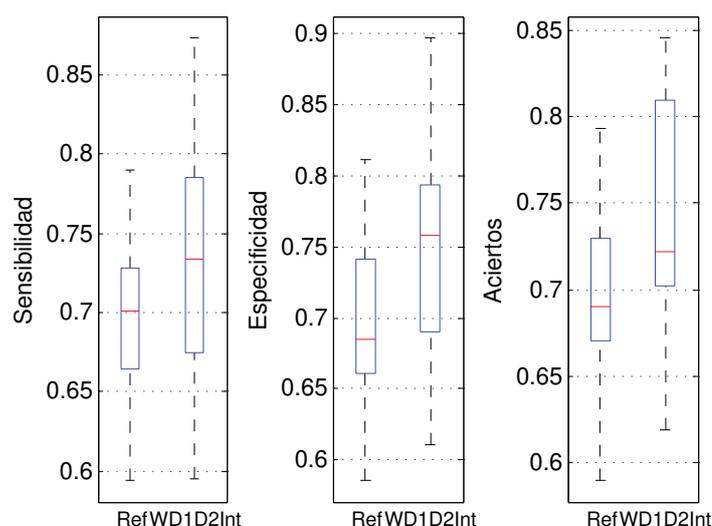


Figura 6.13: Índices de desempeño obtenidos con patrones DDWT sin los Detalles 1 y 2 utilizando *eliminación recursiva de características*.

6.5.2. Selección de características mediante algoritmos genéticos

Las selección de características mediante AG se hizo de la siguiente manera:

1. Inicialización de población: se inicializó aleatoriamente una población de 100 individuos, cuya cantidad de bits (dimensión) se eligió de modo que cada uno de ellos este correlacionado con una característica del patrón de entrada correspondiente a un canal (64, 32 o 16). El valor del bit indica la presencia o ausencia de la característica correspondiente, un 1 significa que se mantiene la característica y un 0 que se la elimina.
2. Evaluación de la población: para realizar esta acción por cada individuo se conforma un nuevo grupo de patrones, de entrenamiento y de prueba,

en función de la distribución de unos y ceros presentes. Por ejemplo, en la Figura 6.14 se muestra un individuo de 16 bits en el cual solo en 4 posiciones hay unos, en consecuencia se generará un nuevo conjunto de patrones con solo 4 características.

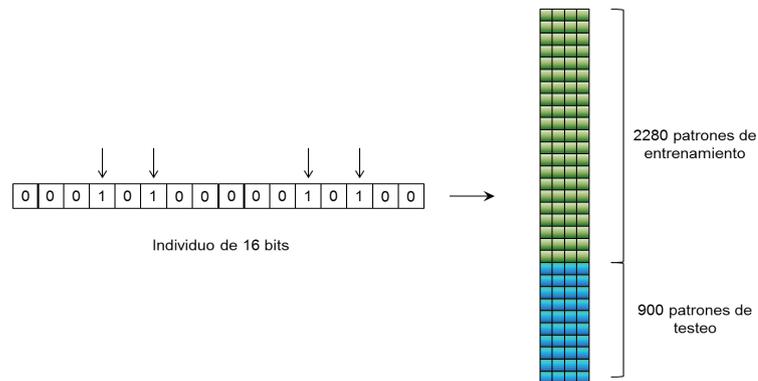


Figura 6.14: Generación de patrones de 4 características a partir de un individuo de la población.

Con este nuevo conjunto se entrena el LDA y se obtiene la tasa de aciertos, la función de fitness toma en cuenta esta última así como también la cantidad de bits activos del individuo y se define mediante la Ecuación 6.1:

$$fitness = w_a * aciertos + \frac{w_c}{\sum c_i} \quad (6.1)$$

asignando un peso mayor a la componente correspondiente a la tasa de aciertos, $w_a = 0,80$, con respecto al componente de la ecuación que representa a la cantidad de características activas, $w_c = 0,20$. Esta ecuación pondera al porcentaje de clasificación, pero también permite detectar como mejor solución aquella que con similar desempeño tiene un número menor de características; fue utilizada por Zhou *et al* para otra aplicación con buen resultado [74].

3. Mientras Mejor $Fitness \leq Fitness Deseado$ Hacer

- Selección: se utilizó el método de competencias para hacer la selección de progenitores; este consiste en seleccionar aleatoriamente 2 individuos de la población y se compara sus respectivos fitness, el que posea el valor mayor queda seleccionado. Este proceso se repite 20 veces de

manera tal de seleccionar 20 individuos de la población como progenitores de la nueva generación [31].

- Cruzas: se utilizaron cruza simples con una probabilidad del 0,95 [75].
- Mutaciones: se utilizó mutaciones de un solo bit por individuo con una probabilidad de 0,05 [75].
- Evaluación de la población

4. Fin mientras

- Algoritmo genéticos aplicado a patrones temporales

En la Figura 6.15 se muestran los boxplots de los índices de desempeño obtenidos mediante aplicación sobre los patrones temporales del método envolvente de selección de características; la tasa de aciertos promedio fue de $0,770 \pm 0,05$, la sensibilidad promedio de $0,766 \pm 0,07$ y la especificidad promedio $0,774 \pm 0,07$. En este caso la longitud promedio del patrón resultante es de 131 características.

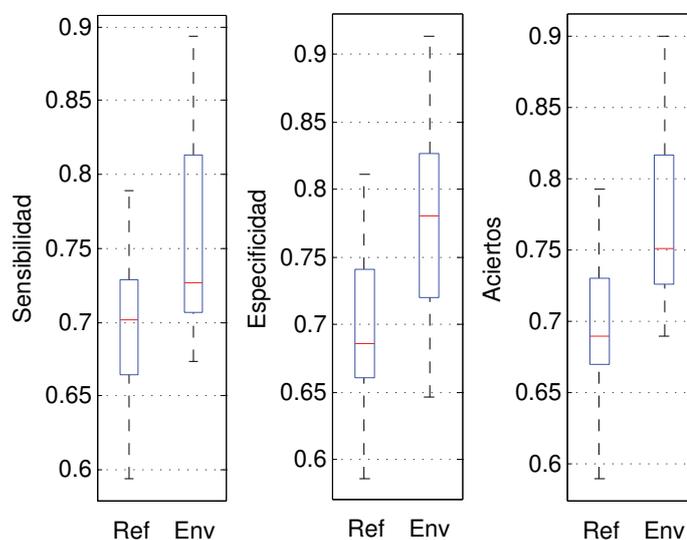


Figura 6.15: Índices de desempeño obtenidos con patrones temporales utilizando *algoritmos genéticos*.

- Algoritmos genético aplicado a patrones DDWT sin D1

En la Figura 6.16 se muestran los boxplots de los índices de desempeño utilizados. La tasa de aciertos promedio fue de $0,781 \pm 0,06$, la sensibilidad

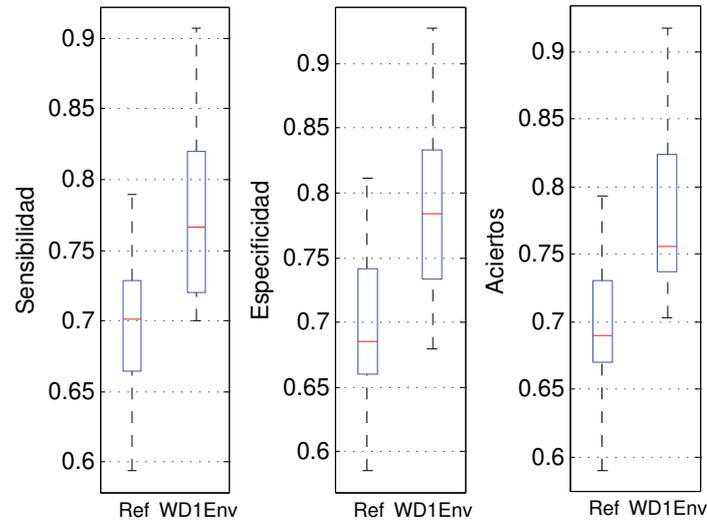


Figura 6.16: Índices de desempeño obtenidos con patrones DDWT sin el Detalle 1 utilizando *algoritmos genéticos*.

promedio de $0,777 \pm 0,06$ y la especificidad promedio $0,786 \pm 0,07$. La longitud del patrón resultante es de 111 características.

- Algoritmo genético aplicado a patrones DDWT sin D1 y D2

En la Figura [6.17](#) se muestran los boxplots de los índices de desempeño utilizados. La tasa de aciertos promedio fue de $0,748 \pm 0,06$, la sensibilidad promedio de $0,732 \pm 0,07$ y la especificidad promedio $0,763 \pm 0,07$. La longitud del patrón resultante es de 140 características.

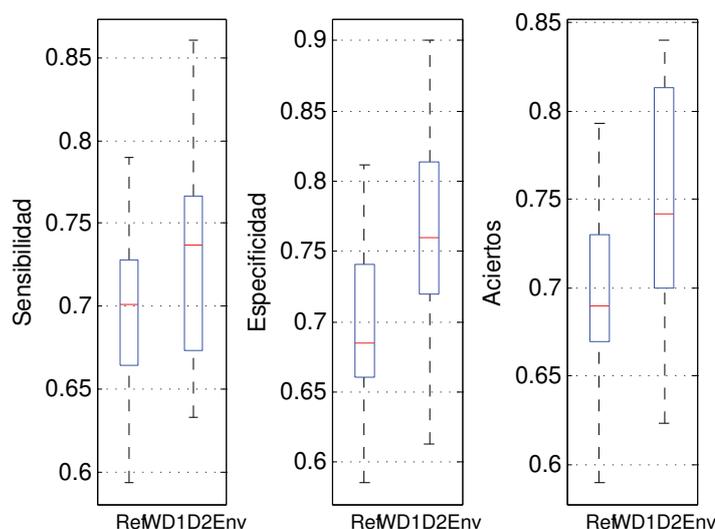


Figura 6.17: Índices de desempeño obtenidos con patrones DDWT sin los Detalles 1 y 2 utilizando *algoritmos genéticos*.

6.6. Comentarios de cierre de capítulo

En este capítulo se describieron los experimentos propuestos para cumplir el objetivo de la tesis y los resultados que se obtuvieron en cada uno de ellos. En el experimento 1 se obtuvieron los índices de desempeño correspondientes a los patrones temporales que se tomaron como referencia para el resto de los experimentos. Posteriormente, en el experimento 2 se obtuvieron los índices de desempeños los patrones generados mediante DDWT y WPT, obteniéndose muy buenos resultados con estos últimos. En el experimento 3 se obtuvieron los índices de desempeño de los patrones generados mediante suavizado de los patrones DDWT, puntualmente de los patrones DDWT sin el detalle 1 y patrones DDWT sin los detalles 1 y 2. Finalmente, en el experimento 4 se obtuvieron los índices de desempeño de las estrategias de selección de características aplicadas a los patrones temporales y a los patrones generados en el experimento 3.

Capítulo 7

Discusión

En este capítulo se discuten los resultados obtenidos en cada uno de los experimentos, se presenta el análisis de significancia estadística de estos y se los compara con resultados obtenidos por otros investigadores que trabajaron con esta base de datos.

7.1. Resultados del Experimento 1

Se determinaron los valores de referencia de los índices de desempeño que son utilizados en la tesis para contrastar los resultados obtenidos en las distintas estrategias propuestas, a saber:

- Tasa de aciertos promedio: $0,695 \pm 0,05$.
- Sensibilidad promedio: $0,696 \pm 0,05$.
- Especificidad promedio $0,695 \pm 0,05$.

Además de estos índices, en todos los experimentos se consideró la longitud final de los patrones resultantes para evaluar el desempeño global de las distintas estrategias. En este caso la longitud es 640 características, dado que el patrón se conformó concatenando las épocas de 1 seg. de duración, adquiridas a 256 Hz y posteriormente decimadas a 64 Hz, en 10 canales de registro.

7.2. Resultados del Experimento 2

Se evaluaron los patrones generados mediante un cambio de base de los patrones temporales originales, puntualmente se utilizaron bases ortogonales generadas mediante descomposición multiresolución a partir de la Transformada Wavelet

Diádica Discreta (patrones DDWT) y a partir de la Transformada Wavelet Packet mediante el algoritmo *local discriminant basis* utilizando como medida de discriminación la *suma de cuadrados* (patrones WPT). En ambos casos se utilizó como wavelet madre la *db4* y se realizó la descomposición hasta el nivel 6.

En la Figura 7.1 se muestra un gráfico de cajas que permite comparar los índices de desempeño para ambos tipos de patrones con los de referencia. En el caso de los patrones DDWT los índices de desempeño obtenidos fueron muy similares a estos últimos; por lo que este cambio de base no representa una alternativa que permita mejorar el desempeño de la ICC. En el caso de los WPT, previamente se

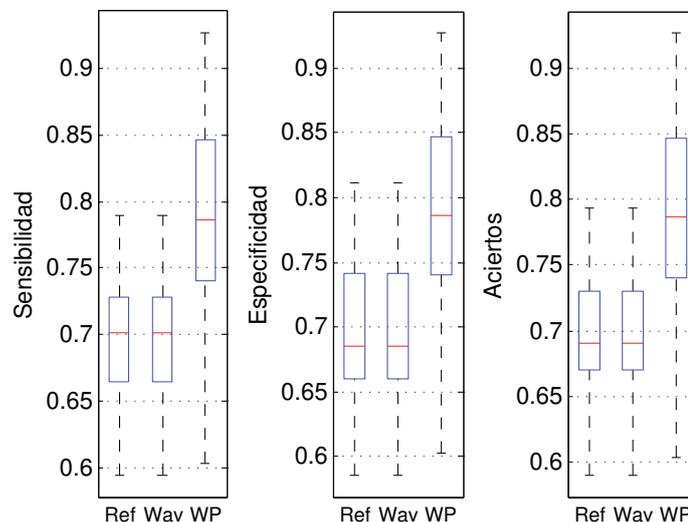


Figura 7.1: Comparación de índices de desempeños obtenidos con patrones DDWT y WPT

realizó un análisis de la conformación de los patrones, ya que el algoritmo LDB devuelve los vectores bases (y sus coeficientes asociados) ordenados en función de su potencia de discriminación. En la Tabla 6.1 se muestran los valores obtenidos para distintas conformaciones de patrones WPT notándose que para el caso de 18 coeficientes se obtienen los mayores índices de desempeño; y por este motivo se tomó la decisión de conformar los patrones con esta cantidad de características. Los índices de desempeño obtenidos con estos patrones fueron superiores a los de referencia, lográndose una mejora de aproximadamente 12,9% sobre los índices de referencia. Además se disminuyó la dimensión de los patrones pasando de 64 a 18 características por canal, lo cual representa un 28% de la dimensión original.

7.3. Resultados del Experimento 3

Se utilizó el proceso de *suavizado* para mejorar la relación señal a ruido y así mejorar la detección de la onda P300. Para cada canal puede pensarse que esto representa la proyección en una base ortogonal subcompleta definida en \mathbb{R}^{32} (eliminando el *detalle 1*) o \mathbb{R}^{16} (eliminando los *detalles 1 y 2*). En la Figura 7.2 se puede comparar los índices de desempeño de los patrones a los cuales se les aplicó suavizado sobre los patrones DDWT.

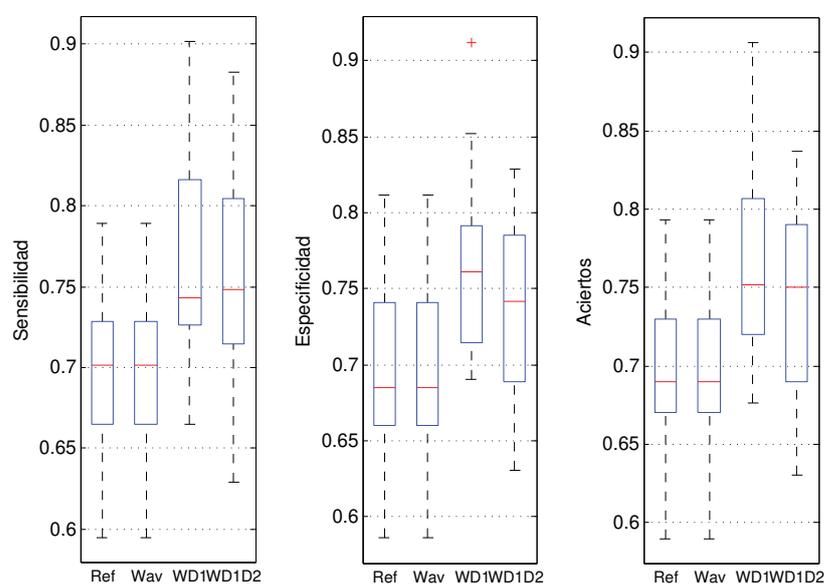


Figura 7.2: Comparación de índices de desempeños obtenidos con patrones DDWT, DDWT sin D1 y DDWT sin D1 y D2

Desde el enfoque de espacios de señales se puede pensar que los elementos de la base original que se eliminan, pertenecientes a los *detalles 1 y 2*, son los que contienen mayormente proyecciones del ruido. Mientras que los que quedan, pertenecientes a los *detalles 3, 4, 5 y 6* y a la *aproximación 6*, son los que contienen mayormente las proyecciones de la señal de ERP; en particular la onda P300.

Los índices de desempeño de los patrones con suavizado fueron superiores a los de referencia, en el caso de los DDWT sin D1 se obtuvo una mejora del 11% y en el caso de los DDWT sin D1 y D2 se obtuvo una mejora del 10,7%. Además, la dimensión de los patrones pasó a 32 características por canal en el caso de los patrones DDWT sin D1 y a 16 características por canal en el caso de los patrones DDWT sin D1 y D2; lo cual representa el 50% y 25% de la dimensión original.

7.4. Resultados del Experimento 4

Se utilizaron métodos de selección de características para reducir la dimensión de los patrones tratando de mantener o mejorar los desempeños obtenidos. Los métodos utilizados fueron eliminación recursiva de características y algoritmos genéticos. Estos se aplicaron sobre patrones temporales y wavelets, no así sobre patrones wavelets packets, dado que el algoritmo LDB realiza una *especie de selección de características* al devolver ordenados los coeficientes en función de la potencia de discriminación.

En la Figura 7.3 se muestra a manera de ejemplo los elementos de la base seleccionada mediante la aplicación de AGs sobre patrones wavelets sin los coeficientes de los Detalles 1 y 2; se observa que la morfología de los elementos de la base seleccionados son similares a la onda P300 en particular el elemento correspondiente a la ubicación 4 del individuo.

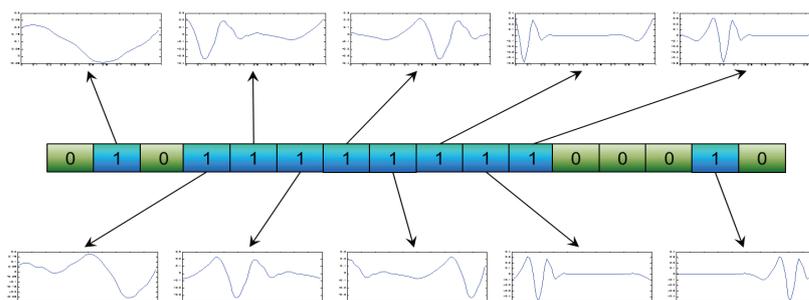


Figura 7.3: Elementos de la base de la DDWT seleccionados por el algoritmo genético, aplicado sobre patrones DDWT sin los Detalles 1 y 2, correspondiente al mejor individuo del sujeto 13 de la base de datos.

En la Figura 7.4 se muestran los índices de desempeño de los algoritmos genéticos aplicados a patrones temporales y a patrones wavelets a los que se les aplicó suavizado (eliminación de detalles 1 y 2).

En la Figura 7.5 se muestran los índices de desempeño del método de eliminación de características aplicados a patrones temporales y a patrones wavelets a los que se les aplicó suavizado (eliminación de detalles 1 y 2).

Tanto los algoritmos genéticos, como la eliminación recursiva de características, obtuvieron índices de desempeño superiores a los de referencia. Con los AGs se obtuvieron mejoras del 10,7% cuando se los aplicó sobre patrones temporales, del 12,4% cuando se los aplicó sobre patrones DDWT sin D1 y del 8% para los patrones DDWT sin D1 y D2. Con el método ERC se obtuvieron mejoras del

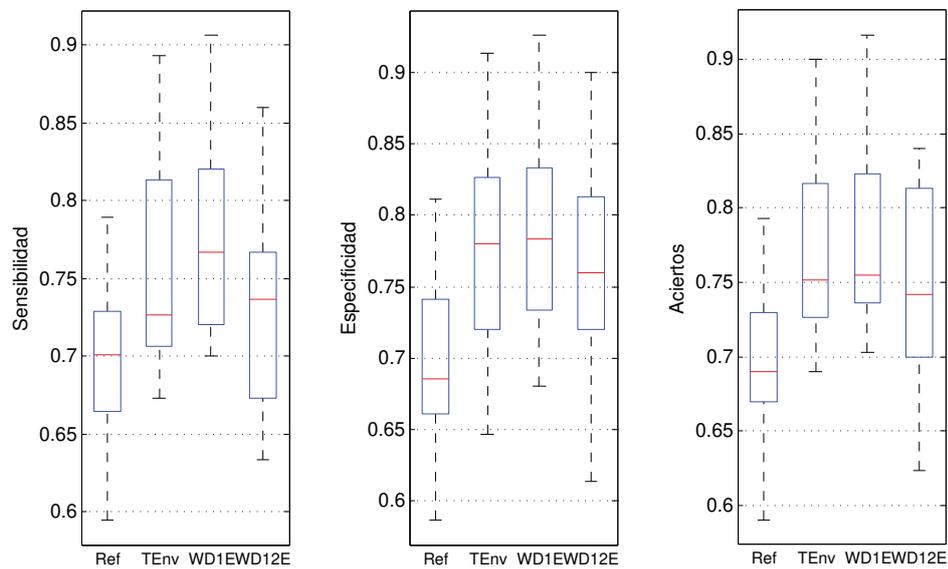


Figura 7.4: Comparación de índices de desempeños obtenidos con *algoritmos genéticos* aplicado a patrones temporales y DDWT sin los Detalles 1 y 2.

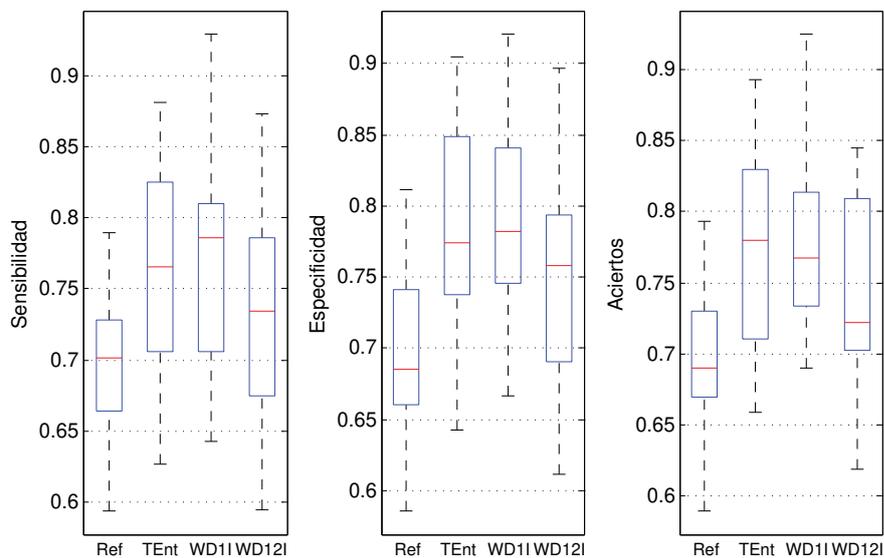


Figura 7.5: Comparación de índices de desempeños obtenidos con *eliminación recursiva de características* aplicado a patrones temporales y DDWT sin los Detalles 1 y 2.

11,8% cuando se los aplicó sobre patrones temporales, del 12% cuando se los aplicó sobre patrones DDWT sin D1 y del 6% para los patrones DDWT sin D1 y D2.

En lo que respecta a la disminución de la dimensión de los patrones con los AGs se obtuvieron 131 características para patrones temporales, 163 para patrones DDWT sin D1 y 140 para patrones DDWT sin D1 y D2; esto representa respectivamente el 20%, el 25% y el 21,8% de la dimensión original. Con el método de ERC se obtuvieron 120 características para patrones temporales, 111 para patrones DDWT sin D1 y 63 para patrones DDWT sin D1 y D2; esto representa respectivamente el 18,7%, el 17,3% y el 9,8% de la dimensión original.

Los índices de desempeño obtenidos por cada uno de los métodos implementados se encuentran resumidos en la Tabla 7.1, estos se encuentran agrupados según los experimentos descritos en el Capítulo 6.

Tabla 7.1: Índices de desempeños obtenidos en los experimentos realizados.

Método	Sensibilidad	Especificidad	Tasa de aciertos	Características
Temporales	$0,696 \pm 0,05$	$0,695 \pm 0,06$	$0,695 \pm 0,05$	640
DDWT	$0,695 \pm 0,05$	$0,696 \pm 0,06$	$0,695 \pm 0,05$	640
WPT + LDB	$0,785 \pm 0,08$	$0,785 \pm 0,08$	$0,785 \pm 0,08$	180
DDWT sin D1	$0,770 \pm 0,06$	$0,767 \pm 0,06$	$0,768 \pm 0,06$	320
DDWT sin D1 y D2	$0,758 \pm 0,06$	$0,739 \pm 0,05$	$0,747 \pm 0,05$	160
Temporales + AGs	$0,766 \pm 0,07$	$0,774 \pm 0,07$	$0,770 \pm 0,05$	131 ± 29
DDWT sin D1 + AGs	$0,777 \pm 0,06$	$0,786 \pm 0,07$	$0,781 \pm 0,06$	163 ± 41
DDWT sin D1 y D2 + AGs	$0,732 \pm 0,07$	$0,763 \pm 0,07$	$0,748 \pm 0,06$	140 ± 16
Temporales + ERC	$0,768 \pm 0,08$	$0,786 \pm 0,08$	$0,777 \pm 0,07$	120 ± 74
DDWT sin D1 + ERC	$0,766 \pm 0,07$	$0,789 \pm 0,07$	$0,778 \pm 0,06$	111 ± 58
DDWT sin D1 y D2 + ERC	$0,727 \pm 0,07$	$0,749 \pm 0,08$	$0,738 \pm 0,06$	63 ± 34

Al analizar la tabla se observa que la mayoría de los resultados obtenidos son prometedores, ya sea si se considera la tasa de aciertos, la sensibilidad o la especificidad; inclusive los valores de estos dos últimos índices son parejos para todos los métodos. La única excepción es el resultado correspondiente a los patrones DDWT, los cuales no generan mejoras con respecto a los patrones temporales de referencia, por lo tanto no serán considerados en los análisis de ahora en adelante. El resto de los métodos la tasa de aciertos está por encima de 0,73, lográndose el mejor desempeño utilizando los patrones WPT y el algoritmo LDB.

El análisis anterior se hizo en base a la tasa de aciertos, sin embargo cuando se tiene en cuenta la dimensionalidad de los patrones resultantes de cada estra-

tegia planteada en los experimentos, se observa que los métodos de selección de características aplicados a patrones temporales tienen desempeños similares a los de los métodos mencionados anteriormente pero con una reducción considerable de la dimensión de los patrones. En la Figura 7.6 se muestra claramente esta situación, en la cual se representa la cantidad de características de los patrones resultante de cada una de las estrategias en función de la tasa de aciertos.

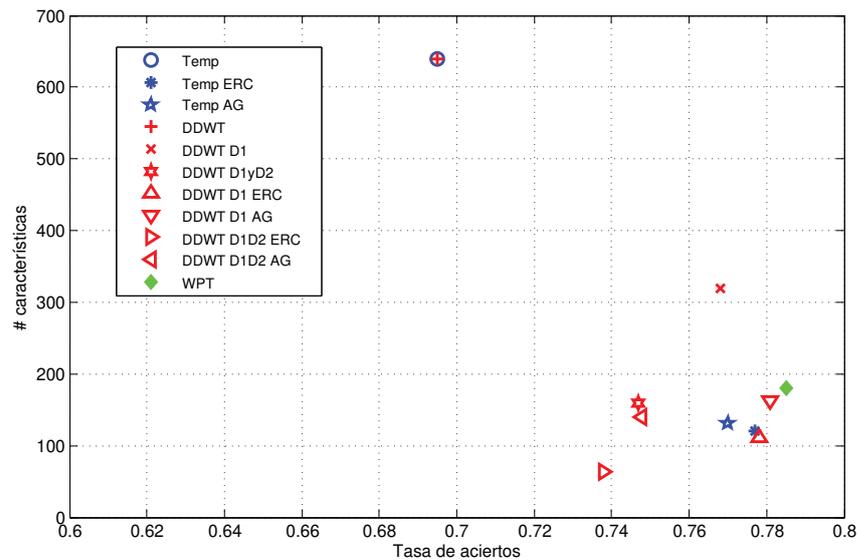


Figura 7.6: Representación de la cantidad de características resultantes en función de la tasa de aciertos. Temp: patrones temporales, Temp ERC: patrones temporales mas ERC, Temp AG: patrones temporales mas AG, DDWT: patrones wavelets, DDWT D1: patrones wavelets sin el detalle 1, DDWT D1yD2: patrones wavelets sin los detalles 1 y 2, DDWT D1 ERC: patrones wavelets sin el detalle 1 mas ERC, DDWT D1 AG: patrones wavelets sin el detalle 1 mas AG, DDWT D1yD2 ERC: patrones wavelets sin los detalles 1 y 2 mas ERC, DDWT D1yD2 AG: patrones wavelets sin los detalles 1 y 2 mas AG, WPT: patrones wavelets packets.

Esto indicaría que los cambios de base producidos tanto por la Transformada Wavelet Packets, como por la Transformada Wavelet Diádica Discreta, generan características (proyecciones de la señal registrada en la nueva base) que permiten discriminar mejor los patrones correspondiente a una clase u otra. Aunque en el caso de la DDWT es necesario mejorar la relación señal a ruido (*suavizado*) para obtener una mejora estadísticamente significativa en el desempeño con respecto al obtenido con los patrones temporales.

Mas allá de esto, la utilización de algoritmos de selección de características mejora notablemente el desempeño de los métodos mencionados anteriormente,

lo cual puede justificarse en la reducción de la dimensionalidad de los datos producidos por estos.

7.5. Análisis estadístico

Para determinar si estadísticamente hay diferencia significativa al utilizar los patrones WPT sobre el resto de las estrategias, se evaluó la normalidad de las distribuciones de los índices de desempeño mediante la prueba de Shapiro-Wilk. En la mayoría de los casos las distribuciones fallaron la prueba, por lo que para evaluar las diferencias entre métodos para cada variable se utilizó la prueba de Friedman, que es el equivalente no paramétrico al análisis de varianza para medidas repetidas. En el caso de que la prueba de Friedman detectase una diferencia significativa entre grupos, se utilizó la prueba de Tukey para las comparaciones *post hoc* entre pares de métodos, también para cada índice de desempeño. El nivel de significancia estadística fue fijado a 0,05 y se utilizó el software SigmaPlot v12 [76].

Pensando en aumentar el poder estadísticos de la prueba, y evitando relizar comparaciones innecesarias, se propone agrupar los métodos de manera tal de determinar cual/cuales son los de mejor desempeño en cada grupo para posteriormente hacer una comparación final. En consecuencia los métodos se agruparon de la siguiente manera:

7.5.1. Selección de características en patrones temporales

Tipos de patrones: Temporales + ERC, y Temporales + AGs.

Se encontraron diferencias significativas entre los índices de desempeño correspondientes a los métodos propuestos y los de referencia (Friedman, $p < 0,001$ para todas las variables), ambos entregan tasas de sensibilidad, especificidad y aciertos significativamente superiores al método de referencia con un menor número de características (Tukey, $p < 0,05$). Sin embargo no hay diferencias significativas entre ellos (Tukey, $p > 0,05$), por lo cual se considerarán ambos para la comparación final.

7.5.2. Generación de características mediante DDWT y suavizado

Tipos de patrones: DDWT, DDWT sin D1 y DDWT sin D1 y D2.

Se encontraron diferencias significativas entre los índices de desempeño correspondientes a los métodos propuestos y los de referencia (Friedman, $p < 0,001$ para todas las variables). Los métodos DDWT sin D1 y DDWT sin D1 y D2 presentan sensibilidad significativamente superiores al de referencia (Tukey, $p < 0,05$), sin embargo no hay diferencias significativas entre ellos (Tukey, $p > 0,05$). El método DDWT sin D1 entrega tasas de especificidad significativamente superiores al de referencia y DDWT sin D1 y D2 (Tukey, $p < 0,05$). Los métodos DDWT sin D1 y DDWT sin D1 y D2 entregan tasas de aciertos significativamente superiores al método de referencia (Tukey, $p < 0,05$), sin embargo no hay diferencias significativas entre ellos (Tukey, $p > 0,05$). En resumen, podría decirse que el método DDWT sin D1 parece ser superior debido a que la sensibilidad y tasa de aciertos son comparables a DDWT sin D1 y D2, sin embargo tiene mejor especificidad que este último.

7.5.3. Generación y selección de características en patrones wavelets

Tipos de patrones: DDWT sin D1 + algoritmos genéticos (DDWT sin D1+AGs), DDWT sin D1 + eliminación recursiva de características (DDWT sin D1+ERC), DDWT sin D1 y D2 + algoritmos genéticos (DDWT sin D1 y D2+AGs), DDWT sin D1 y D2 + eliminación recursiva de características (DDWT sin D1 y D2+ERC).

Se tomó como criterio principal para determinar la significancia estadística a los índices de desempeño evaluados en los grupos anteriores. Sin embargo, en caso que haya dos métodos con igual desempeño se consideró la cantidad de características como un segundo criterio a utilizar. Para todos los métodos se encontraron diferencias en la sensibilidad (Friedman, $p < 0,001$), especificidad (Friedman, $p = 0,002$), tasa de aciertos (Friedman, $p < 0,001$) y cantidad de características (Friedman, $p < 0,001$) con respecto a los de referencia. En base al criterio principal, se descartan DDWT sin D1 y D2+ERC y DDWT sin D1 y D2+AG por tener significativamente menor sensibilidad que el resto (Tukey, $p < 0,05$). También se descarta DDWT sin D1 y D2 por tener menor especificidad que DDWT sin D1+AG y DDWT sin D1+ERC y menor tasa de aciertos que

DDWT sin D1+ERC (Tukey, $p < 0,05$ en todos los casos). Como resultado de aplicar el criterio principal, se seleccionan DDWT sin D1+AG, DDWT sin D1+ERC. En base al criterio secundario, y considerando además los patrones DDWT sin D1, debería seleccionarse sólo DDWT sin D1+ERC, ya que a tasas comparables de sensibilidad, especificidad y aciertos, utiliza significativamente menos características que DDWT sin D1+AG y DDWT sin D1 (Tukey, $p < 0,05$).

7.5.4. Análisis final

Tipos de patrones: en base a los resultados obtenidos en los ítems anteriores se consideraron los patrones Temporales+ERC, Temporales+AGs, DDWT sin D1+ERC y WPT.

No se encontraron diferencias significativas entre los métodos mencionados en base al criterio principal; ya sea sensibilidad (Friedman, $p = 0,615$), especificidad (Friedman, $p = 0,376$) o tasa de aciertos (Friedman, $p = 0,457$). En base al criterio secundario, existen diferencias significativas entre los métodos mencionados en términos de cantidad de características (Friedman, $p < 0,001$). En particular, Temporales+ERC, Temporales+AG y DDWT sin D1+ERC utilizan significativamente menos características que WPT (Tukey, $p < 0,05$), sin ninguna otra diferencia significativa.

7.6. Comparación de resultados

Los resultados obtenidos fueron contrastados con los obtenidos por investigadores que utilizaron la base de datos de la Universidad Autónoma Metropolitana. Dado que en lo que respecta a la tasa de aciertos no se pudo establecer diferencias significativas entre los grupos que mejor desempeño obtuvieron, lo cual implica que se podría utilizar cualquiera de estos, se tomarán los índices de desempeño de los patrones WPT para la comparación.

Lindig-León *et al* en [77] desarrolló un método que tiene como objetivo reducir el compromiso existente entre la tasa de deletreo y la precisión, esto es una características de cualquier ICC basada en P300 y se debe a que en este caso se utilizan patrones generados mediante promediación coherente. Para esto utilizó un enfoque bayesiano para estimar la probabilidad *a posteriori* asociada con la clasificación de cada objetivo utilizando un discriminante lineal como clasificador.

La tasa de aciertos obtenida en este caso es de 0,75 utilizando patrones generados mediante promediación coherente de 2 épocas .

Saavedra *et al* en [78], con el objetivo de mejorar la relación señal a ruido en registros de ERP en una ICC, propuso un nuevo enfoque de denoising en el dominio wavelet teniendo en cuenta la correlación entre los canales registrados. Esto se hace combinando la información de fase y amplitud en el dominio wavelet para seleccionar automáticamente una ventana temporal en cada canal de registro que maximice la separabilidad de clases (registros *con P300* y *sin P300*). Los resultados reportados en este trabajo indican una tasa de aciertos de 0,56 para registros de época única.

Kindermans *et al* en [79] propuso una ICC capaz de realizar delecteo sin ningún tipo de entrenamiento, para lo cual utiliza información inter-sujetos así como también modelos del lenguaje utilizando modelos ocultos de Markov; el clasificador utilizado es del tipo no supervisado. Con esta propuesta se logran tasas de acierto del 0,89 para patrones generados con promedios de 3 épocas.

En la Tabla 7.2 se sumaría los resultados descriptos, si bien Kindermans *et al* obtiene una mayor tasa de aciertos esto se debe a que los patrones se conforman con promedios de 3 épocas; lo cual representa una ventaja muy grande comparado con patrones conformados por una sola época. Para los otros casos los resultados obtenidos en esta tesis son superiores.

Tabla 7.2: Comparación de resultados con otros investigadores utilizando la tasa de aciertos como índice de desempeño.

Autores	Tasa de aciertos	Cantidad de épocas
Acevedo <i>et al</i>	0,78	Epoca única
Lindig <i>et al</i>	0,75	2 épocas
Saavedra <i>et al</i>	0,56	Epoca única
Kindermans <i>et al</i>	0,89	3 épocas

Además, para corroborar el desempeño de los métodos propuestos en esta tesis, se evaluaron algunos de ellos utilizando la base de datos del Wadsworth Center, la cual fue obtenida mediante un procedimiento similar al descrito en la Sección 5.1 con la diferencia que solo participaron 2 sujetos y se adquirieron 64 canales en posiciones del sistema 10-20 [52].

Para este experimento no se utilizó la totalidad de canales sino aquellos en las posiciones Fz, C4, Cz, C3, P3, P4, Pz, PO8, Oz, y PO7; y se sub-muestrearon las

señales con una frecuencia de muestreo de 64 Hz. Además, se probaron métodos representativos de cada experimento planteado en la tesis y con tasas de acierto superiores al 0,75 con la base de datos de la Universidad Autónoma Metropolitana, a saber: patrones WPT (experimento 2), DDWT con soothing (experimento 3) y patrones temporales con selección de características mediante AGs (experimento 4). Los resultados obtenidos se muestran en la Tabla 7.3, donde se corrobora que la estrategia que obtiene el mejor desempeño es la combinación de patrones WPT con el algoritmo LDB.

Tabla 7.3: Tasas de aciertos obtenidas con la base del Wadsworth Center.

Método	Tasa de aciertos	Características
Temporales	$0,67 \pm 0,04$	640
WPT + LDB	$0,70 \pm 0,03$	180
DDWT sin D1	$0,64 \pm 0,02$	320
Temporales + AGs	$0,68 \pm 0,001$	275 ± 7

Capítulo 8

Conclusiones

Hay situaciones que producen una disminución en la capacidad de comunicación de una persona con su entorno, por ejemplo traumatismos debido a accidentes, o enfermedades como la esclerosis lateral amiotrófica o esclerosis múltiple entre otras. En este contexto el caso extremo es cuando se pierde el control total el sistema muscular, quedando la persona cognitivamente en buen estado pero sin poder mover su cuerpo, este estado se denomina síndrome de enclaustramiento.

En la actualidad existen dispositivos que permiten restituir artificialmente, y de manera parcial, las funciones de las vías de comunicación dañadas, otorgando a una persona en estado de enclaustramiento la posibilidad de una interacción reducida enviando mensajes y comandos a una computadora y desde ésta conectarse o interactuar con el mundo exterior, estos dispositivos se denominan *interfaces cerebro computadoras* (ICC).

Se pueden implementar ICC utilizando diversos paradigmas, el más popular dentro de la comunidad científica es el basado en electroencefalograma, y dentro de este modelo el denominado paradigma P300. Básicamente este consta de una matriz con caracteres alfanuméricos en la que se van iluminando sus filas y columnas aleatoriamente, cuando se ilumina la fila o la columna en la cual está el carácter al que la persona está prestando atención se genera una señal de potencial evocado relacionado con evento cuya principal característica es la onda P300. En consecuencia, bastaría con identificar cual es la fila y la columna en la cual se generó esta onda y así saber cual es el carácter que la persona quiere seleccionar.

La solución no es sencilla debido a que la onda P300 se encuentra inmersa en ruido de fondo conformado principalmente por electroencefalograma con una relación señal a ruido muy pobre. Si bien hay técnicas que permiten mejorarla,

y en consecuencia facilitar la detección de la onda P300, estas insumen tiempo de procesamiento; lo cual no es deseable ya que en este ámbito se busca lograr la mayor tasa de transferencia de instrucciones por unidad de tiempo.

Por este motivo, en los últimos años ha aumentado el interés por investigar el problema de detección de la onda P300 en registros de electroencefalograma de época única o *single trial*. Esto permitiría mejorar la tasa de transferencia por lo comentado en el párrafo anterior, pero por otro lado representa un problema de difícil solución.

8.1. Contribuciones

En esta tesis se abordó el problema de detección de la onda P300 en registros *single trial* en el marco de una interfaz cerebro computadora, para lo cual se evaluaron distintas estrategias combinando representaciones en bases ortogonales derivadas de la Transformada Wavelet Diádica Discreta y de la Transformada Wavelet Packet, suavizado de la señal en el dominio Wavelet y estrategias de reducción de dimensionalidad utilizando algoritmos genéticos y eliminación recursiva de características.

Se evaluaron representaciones de la señal de electroencefalograma obtenidas en bases ortogonales generadas por la Transformada Wavelet Diádica Discreta y por la Transformada Wavelet Packets, en ambos casos utilizando como Wavelet madre la Daubechies 4. Este enfoque desde el punto de vista de espacios de señales es relativamente novedoso en el ámbito de la investigación en ICC, y permite a futuro profundizar más sobre representaciones de señales en bases no ortogonales.

La utilización de estrategias de mejoramiento de la relación señal a ruido en el dominio Wavelet, en este caso particular el suavizado o smoothing, probó ser una alternativa válida para mejorar la detección de la onda P300, ya sea porque la resalta del ruido en la que está inmersa así como también porque implícitamente provoca una reducción de la dimensionalidad de los patrones resultantes.

Relacionado con este último concepto, y considerando además la posibilidad de futuras implementaciones de una ICC en alguna plataforma hardware se estudiaron métodos de reducción de dimensionalidad de las señales mediante algoritmos genéticos y eliminación recursiva de características, las cuales permitieron mejorar aún más los desempeños obtenidos en los casos descritos anteriormente, excepto para el caso de los patrones WPT.

En términos generales, se puede decir que se obtuvieron índices de desempeño superiores a 0,70 en la mayoría de la estrategias propuestas, y superiores a 0,75 en mas de la mitad; lo cual es un logro importante considerando que con estos valores es posible establecer la comunicación de una persona mediante una ICC y eventualmente controlar un dispositivo [80, 81].

Los mejores desempeños se obtuvieron con los patrones WPT, con los que se logró una tasa de aciertos de 0,785, representando una mejora del 12,9% con respecto a los índices de referencia. Además, las siguientes estrategias también produjeron buenos resultados: a) patrones Temporales combinados con ERC, b) patrones Temporales combinados con AGs y c) patrones DDWT sin el detalle 1 combinado con ERC; obteniéndose tasas de aciertos de 0,777 (mejora del 11,7%), 0,770 (mejora del 10,8%) y 0,778 (mejora del 11,9%) respectivamente con respecto a los índices de referencia. Si bien no se pudieron establecer diferencias estadísticamente significativas entre ellas, se puede concluir que cualquiera es una buena opción.

La situación es diferente si se piensa implementar una interfaz cerebro computadora en una plataforma hardware, en esta situación las estrategias que utilizan patrones temporales y selección de características presentan ventajas por sobre las estrategias basadas en patrones DDWT y WPT, ya que son mas simples y generan patrones con dimensionalidad menor. A manera de ejemplo, esto se observa claramente si se analizan las dimensiones resultantes de las estrategias a) y b) mencionadas en el párrafo anterior, ya que fueron el 18% y el 20% de la dimensión de los patrones de referencia respectivamente.

8.2. Trabajo futuro

En esta sección se presentan algunas líneas de trabajo futuro para mejorar los desempeños logrados en esta tesis.

- a) Selección de canales de registro: en esta tesis se utilizaron la totalidad de canales de registros disponibles en la base de datos de la Universidad Autónoma Metropolitana, sin embargo disminuir la cantidad de canales manteniendo o mejorando el desempeño resultaría conveniente, ya que esto simplificaría el setup de instrumentación del sujeto y por sobre todas las cosas reduciría aún mas la dimensionalidad de los patrones y propiciaría la implementación hardware de una ICC. Algunas alternativas a evaluar son el algoritmo CSP (del

- inglés *Common Spatial Pattern*) [82] y algoritmos genéticos [83], entre otros.
- b) Ensamble de clasificadores: la etapa de clasificación de la ICC se conformó con un clasificador lineal al cual le ingresa la información de los 10 canales concatenados, ya sean temporales o wavelets. Tal como se motró, durante el desarrollo de la tesis, se realizó una primera experiencia evaluando 3 métodos de construcción de ensambles de clasificadores lineales así como también diferentes métodos de toma de decisión. Si bien los resultados obtenidos no son satisfactorios, orientan la búsqueda hacia el estudio de otros tipos de clasificadores con características de mayor diversidad en su respuesta para ser utilizados en ensambles, como máquinas de soporte vectorial, Perceptrones multicapa, otros.
 - c) Representación a través de bases no ortogonales: en el contexto de la tesis la matriz de transformación o cambio de base Φ corresponde a una base ortogonal completa, sin embargo resulta de interés evaluar alternativas que no necesariamente cumplan con esta condición. En este sentido se podría considerar la Descomposición Modal Empírica, Análisis de Componentes Independientes y bases no ortogonales obtenidas a partir de diccionarios de átomos tiempo-frecuencia sobrecompletos mediante algoritmos como Basis Pursuits o Matching Pursuits, entre otros.
 - d) Implementación en hardware: al momento de presentación de esta tesis ya se están realizando algunas experiencias de implementación de los bloques de extracción de características y clasificación de una interfaz cerebro computadora en la plataforma EDU-CIAA-NXP basada en un microcontrolador ARM CORTEX M4 de 32 bits [84]. El objetivo es evaluar el desempeño de esta plataforma en implementaciones de distintas alternativas de selección de características, clasificadores lineales, tipos de datos, entre otras.

8.3. Lista de publicaciones derivadas de la tesis

8.3.1. Publicaciones en revistas

1. Carolina Posada Zuluaga, Rubén Acevedo, “Estimación de potenciales evocados auditivos del tronco cerebral mediante descomposición modal empírica”, *Revista Ingeniería Biomédica (EIA, Colombia)*, Vol. 2, No. 3, pp. 27-32,

Junio 2008, ISSN 1909–9762.

2. J. Gebhart, G. Gentiletti, A. Bonardi, M. Richard, R. Acevedo, “Comando en tiempo real de una SRAI simulada mediante interfaz cerebro computadora basada en P300”, *Revista Argentina de Bioingeniería (SABI)*, Vol. 15 No. 1, pp. 17-22, Junio 2009, ISSN: 0329-5257.
3. G.G. Gentiletti, J.G. Gebhart, R.C. Acevedo, O. Yáñez-Suárez, V. Medina-Bañuelos, “Command of a simulated wheelchair on a virtual environment using a brain-computer interface”, *Ingénierie et Recherche Biomédicale*, Elsevier Publisher, Vol. 30, No. 5, pp. 218-225, Noviembre de 2009.
4. I. Gareis, G. Gentiletti, R. Acevedo, L. Rufiner, “Feature Extraction on Brain Computer Interfaces using Discrete Dyadic Wavelet Transform: Preliminary Results”, *Journal of Physics: Conference Series* 313, 2011. ([doi: 10.1088/1742-6596/313/1/012011](https://doi.org/10.1088/1742-6596/313/1/012011)).
5. Cristian Arjona, José Pentácolo, Iván Gareis, Yanina Atum, Gerardo Gentiletti, Rubén Acevedo and Leonardo Rufiner, “Evaluation of LDA Ensembles Classifiers for Brain Computer Interface”, *Journal of Physics: Conference Series* 332, 2011. ([doi: 10.1088/1742-6596/332/1/012025](https://doi.org/10.1088/1742-6596/332/1/012025)).
6. Iván Gareis, Yanina Atum, Gerardo Gentiletti, Rubén Acevedo, Verónica Medina Bañuelos and Leonardo Rufiner, “On the use of LDA performance as a metric of feature extraction methods for a P300 BCI classification task”, *Journal of Physics: Conference Series* 332, 2011. ([doi: 10.1088/1742-6596/332/1/012021](https://doi.org/10.1088/1742-6596/332/1/012021)).
7. Larrazábal A., Acevedo R., Rufiner L., “Mejoramiento de la relación señal a ruido en potenciales evocados mediante métodos de regularización”, *Revista Argentina de Bioingeniería (SABI)*, Vol. 18 No. 1, Junio 2012, ISSN: 0329-5257.
8. Victoria Peterson, Yanina Atum, Florencia Jauregui, Iván Gareis, Rubén Acevedo, Leonardo Rufiner, “Detección de potenciales evocados relacionados a eventos en interfaces cerebro-computadora mediante transformada wavelet”, *Revista Ingeniería Biomédica (EIA, Colombia)*, Vol. 7, No. 4, pp. 50-58, Diciembre 2013, ISSN 1909–9762.

9. Yanina Atum, Marianela Pacheco, Rubén Acevedo “Evaluation of Feature Selection Methods for P300 Brain Computer Interfaces”, Special Issue on Event Related Potentials Brain Computers Interfaces, Computers, ISSN 2073-431X (http://www.mdpi.com/journal/computers/special_issues/erp_bci). *En evaluación.*
10. R. Acevedo, Y. Atum, V. Peterson, I. Gareis, J. Biurrun Manresa, V. Medina Bañuelos, H. Rufiner, “Feature extraction using orthogonal basis in P300 brain computers interfaces”, Medical & Biological Engineering & Computing, ISSN 1741-0444. *En preparación.*

8.3.2. Proceedings de congresos internacionales

1. G.G. Gentiletti Faenze, J.G. Gebhart, M. Richard, E. García Cossio, R.C. Acevedo, V. Medina Bañuelos, “Robotics Simulation Platform for BCI Application: A Wheelchair Driving Example Using P300 Paradigm”, Proceedings of the 4th International Brain-Computer Interface Workshop and Training Course, Graz (Austria), 2008.
2. Yanina Atum, Gerardo G. Gentiletti, Rubén Acevedo, “Detección de Potenciales Evocados en Interfaces Cerebro Computadora mediante Técnicas de Aprendizaje Maquinal”, II Jornadas Chilenas de Ingeniería Biomédica, Valparaíso (Chile), Octubre de 2009.
3. Yanina Atum, Iván Gareis, Gerardo Gentiletti, Rubén Acevedo, Leonardo Rufiner “Genetic Feature Selection to Optimally Detect P300 in Brain Computer Interfaces”, Proceedings of the 32nd Annual International Conference of the IEEE EMBS Buenos Aires, Argentina, August 31 - September 4, 2010, pp. 3289-3292.
4. Iván E. Gareis, Rubén C. Acevedo, Yanina V. Atum, Gerardo G. Gentiletti, Verónica Medina Bañuelos, Hugo L. Rufiner, “Determination of an optimal training strategy for a BCI classification task with LDA”, Proceedings of the 5th International IEEE EMBS Neural Engineering Conference, Cancún (México), 27 de Abril al 01 de Mayo, 2011, pp 25.
5. I. E. Gareis, R. C. Acevedo, Y. V. Atum, G. G. Gentiletti, V. Medina Bañuelos, H. L. Rufiner, “Efecto de la cantidad y dimensión de los patro-

nes en una interfaz cerebro computadora basada en discriminante lineal de Fisher”, Proceedings of V Latin American Congress on Biomedical Engineering, págs. 1186–1189. Springer International Publishing, 2012. ISBN 978-3-319-13117-7.

6. V. Peterson, Y. Atum, F. Jauregui, I. Gareis, L. Rufiner, R. Acevedo, “Extracción de características en interfaces cerebro computadoras basadas en transformaciones ortogonales: resultados preliminares”, Memorias de las III Jornadas Chilenas de Ingeniería Biomédica, Valparaíso (Chile), Octubre de 2012.
7. Y. Atum, J. Biurrun Manresa, L. Rufiner and R. Acevedo. Genetic Feature Selection for a P300 Brain Computer Interface, Proceedings of VI Latin American Congress on Biomedical Engineering, págs. 603–606. Springer International Publishing, 2015. ISBN 978-3-319-13117-7.
8. V. Peterson, R. Acevedo, H. Rufiner, y R. Spies. Local Discriminant Wavelet Packet Basis for Signal Classification in Brain Computer Interface, Proceedings of VI Latin American Congress on Biomedical Engineering, págs. 584–587. Springer International Publishing, 2015. ISBN 978-3-319-13117-7.

8.3.3. Proceedings de congresos nacionales

1. A. Quiroga, R. Acevedo, H. Rufiner, “Extracción de Características en Interfaces Cerebro Computadoras mediante Distribuciones Tiempo-Frecuencia: Resultados preliminares”, Memorias del XIXI Congreso Argentino de Bioingeniería (SABI 2013), Tucumán, 4 al 6 de septiembre, 2013.
2. V. Peterson, Y. Atum, I. Gareis, L. Rufiner, R. Acevedo, “Extracción de características en interfaces cerebro computadoras mediante diccionarios wavelets packets discriminativos”, Memorias del XIXI Congreso Argentino de Bioingeniería (SABI 2013), Tucumán, 4 al 6 de septiembre, 2013.
3. A. Larrazábal, L. Rufiner, R. Acevedo, “Comparación de métodos de regularización para estimación de potenciales evocados”, Memorias del XIXI Congreso Argentino de Bioingeniería (SABI 2013), Tucumán, 4 al 6 de septiembre, 2013.

4. F. Jauregui, V. Peterson, R. Acevedo, “Denoising en el dominio wavelet para detección de P300 en interfaz cerebro computadora”, Memorias del XIXI Congreso Argentino de Bioingeniería (SABI 2013), Tucumán, 4 al 6 de septiembre, 2013.

Bibliografía

- [1] J. Mak and J. Wolpaw. Clinical applications of brain-computer interfaces: Current state and future prospects. *IEEE reviews in biomedical engineering*, 2:187–199, January 2009. ISSN 1941-1189. doi: 10.1109/RBME.2009.2035356. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2862632&tool=pmcentrez&rendertype=abstract>. [8](#), [22](#)
- [2] L. Rufiner. *Análisis y modelado digital de la voz. Técnicas recientes y aplicaciones*. Ediciones UNL, Colección Ciencia y Técnica, 1a. ed. edition, 2006. ISBN 9789876570909. [9](#), [36](#), [37](#), [38](#), [44](#), [45](#), [46](#), [47](#)
- [3] Laboratorio de neuroimagenología. *Deletreador p300 de Donchin*, 2016. URL <http://akimpech.izt.uam.mx/dokuwiki/doku.php>. [9](#), [66](#), [67](#)
- [4] V. Peterson, R. Acevedo, L. Rufiner, and R. Spies. *Local Discriminant Wavelet Packet Basis for Signal Classification in Brain Computer Interface*, pages 584–587. Springer International Publishing, 2015. ISBN 978-3-319-13117-7. doi: 10.1007/978-3-319-13117-7_149. URL http://dx.doi.org/10.1007/978-3-319-13117-7_149. [11](#), [81](#), [92](#)
- [5] E. Kandel, J. Schwartz, and T. Jessell. *Principios de Neurociencia*. McGraw-Hill Interamericana, 2001. [26](#)
- [6] D. Daly and T. Pedley. *Current Practice of Clinical Electroencephalography 2 Ed*. Lippincott-Raven Pub., 1997. [26](#)
- [7] K. Chiappa. *Evoked Potentials in Clinical Medicine 3rd Ed*. Lippincott-Raven, 1998. [27](#), [29](#)
- [8] R. Acevedo, C. Bonell, G. Gentiletti, J. Aldonate, C. Mercuri, S. Escobar, and J. Biurrun Manresa. Potenciales evocados auditivos de tronco cerebral:

- Aspectos tecnológicos del registro. *Revista Argentina de Bioingeniería*, 12 (1):10–17, 2006. [28](#)
- [9] J. Kropotov. *Quantitative EEG, Event-Related Potentials and Neurotherapy*. Academic Press, 2008. [30](#)
- [10] J. Wolpaw, N. Birbaumer, W. Heetderks, D. McFarland, P. Peckham, G. Schalk, E. Donchin, L. Quatrano, C. Robinson, and T. Vaughan. Brain-computer interface technology: a review of the first international meeting. *IEEE transactions on rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, 8(2):164–73, June 2000. ISSN 1063-6528. [30](#)
- [11] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, and T. Vaughan. Braincomputer interfaces for communication and control. *Clinical Neurophysiology*, 113:767–791, 2002. [31](#)
- [12] G. Gentiletti, C. Taberning, and R. Acevedo. Interfaces cerebro computado-ra: Definición, tipos y estado actual. In *IFMBE Proceedings Vol.18*, 2007. [31](#), [32](#)
- [13] M. Rizk, C. Bossetti, T. Jochum, S. Callender, M. Nicoletis, D. Turner, and P. Wolf. A fully implantable 96-channel neural data acquisition system. *J. Neural Eng. Apr*, 6(2), 2009. [31](#)
- [14] J. O’Doherty, M. Lebedev, T. Hanson, N. Fitzsimmons, and M. Nicoletis. A brain-machine interface instructed by direct intracortical microstimulation. *Front. Integr. Neurosci.*, (3):1–10, 2009. [31](#)
- [15] B. Hong, W. Yijun, G. Xiaorong, and G. Shang kai. Quantitative eeg-based brain-computer interface. In Shanbao Tong and Nitish V. Thakor, editors, *Quantitative EEG Analysis Methods and Clinical Applications*, chapter 8. 2009. ISBN 9781596932043. [31](#)
- [16] L. Hochberg, M. Serruya, G. Friehs, J. Mukand, M. Saleh, A. Caplan, A. Branner, D. Chen, R. Penn, and J. Donoghue. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442(7099): 164–171, July 2006. ISSN 1476-4687. doi: 10.1038/nature04970. URL <http://www.ncbi.nlm.nih.gov/pubmed/16838014>. [31](#)

- [17] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for eeg-based brain-computer interfaces. *Journal of neural engineering*, 4(2), June 2007. ISSN 1741-2560. doi: 10.1088/1741-2560/4/2/R01. URL <http://www.ncbi.nlm.nih.gov/pubmed/17409472>. [32](#), [59](#)
- [18] K. Muller, C. Anderson, and G. Birch. Linear and nonlinear methods for brain-computer interfaces. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 11(2):165–169, 2003. [32](#)
- [19] M. López Gordo. *Interfaz BCI de altas prestaciones basada en la detección y procesamiento de la actividad cerebral*. Tesis doctoral, Universidad de Granada, 2009. [32](#)
- [20] L. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical neurophysiology*, 70(6):510–523, December 1988. ISSN 0013-4694. URL <http://www.ncbi.nlm.nih.gov/pubmed/2461285>. [33](#)
- [21] R. Quian Quiroga and H. Garcia. Single-trial event-related potentials with wavelet denoising. *Clinical Neurophysiology*, 113:376–390, 2003. [33](#)
- [22] A. Herrera, J. Biurrun Manresa, and R. Acevedo. Promediación coherente mejorada mediante transformada wavelet de potenciales evocados auditivos de tronco cerebral. In *IV Latin American Congress of Biomedical Engineering*, 2007. [33](#)
- [23] S. Georgiadis, P. Ranta-aho, M. Tarvainen, and A. Pasi. Single-trial dynamical estimation of event-related potentials: A kalman filter-based approach. *IEEE Trans. on Biomedical Engineering*, 53(8), 2005. [33](#)
- [24] D. Iyer and G. Zouridakis. Single-trial evoked potential estimation: Comparison between independent component analysis and wavelet denoising. *Clinical Neurophysiology*, 118:495–504, 2007. [33](#)
- [25] C. Posada Zuluaga and R. Acevedo. Estimación de potenciales evocados auditivos del tronco cerebral mediante descomposición modal empírica. *Revista Ingeniería Biomédica*, 2(3):27–32, 2008. [33](#)

- [26] Q. Wei, S. Kenneth, H. Francis, F. Lam, W. Paul, and P. Roger. Adaptive filtering of evoked potentials with radial-basis-function neural network prefilter. *IEEE Trans. on Biomedical Engineering*, 49:225–232, 2002. [33](#)
- [27] M. Lewicki and T. Sejnowski. Learning overcomplete representations. *Neural Computation*, 12:337–365, 2000. [38](#)
- [28] K. Kreutz-Delgado and B. Rao. Sparse basis selection, ica, and majorization: Towards a unified perspective. In *IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING ICASSP99*, 1999. [38](#)
- [29] I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh. *Feature Extraction, Foundations and Applications. Series Studies in Fuzziness and Soft Computing*. Springer Verlag, 2006. [39](#), [53](#), [54](#), [55](#), [57](#), [58](#)
- [30] R. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 2002. [39](#)
- [31] D. Milone, L. Rufiner, R. Acevedo, L. Di Persia, and H. Torres. *Introducción a las Señales y a los Sistemas Discretos*. EDUNER, 2006. [41](#), [55](#), [56](#), [101](#)
- [32] S. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way, 3^o ed.* Academic Press, 2008. [43](#), [45](#), [46](#), [47](#)
- [33] R. Coifman and M. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713–718, 1992. [44](#)
- [34] N. Saito. *Local Feature Extraction and Its Applications Using a Library of Basis*, December 1998. [45](#), [47](#), [49](#), [71](#)
- [35] N. Saito, B. Larson, and B. Benichou. Sparsity vs. statistical independence from a best-basis viewpoint. *Wavelet Applications in Signal and Image Processing VIII, Proc. SPIE 4119*, pages 474–486, 2000. [45](#)
- [36] G. Strang and T. Nguyen. *Wavelets and filter banks*. Wellesley-Cambridge Press, 1996. [46](#)
- [37] V. Samar. Wavelet analysis of neuroelectric waveforms: A conceptual tutorial. *Brain and Language*, 66:7–60, 1999. [48](#)

- [38] R. Coifman and M. Wickerhauser. Entropy-based algorithms for best basis selection. *Transactions on Information Theory*, 38:713–718, 1992. [48](#), [49](#)
- [39] N. Saito and R. Coifman. Local discriminant bases and their applications. *Journal of Mathematical Imaging and Vision*, 5:337–358, 1995. [49](#), [50](#)
- [40] N. Saito and R. Coifman. On local discriminant bases for classification and regression. *Acoustics, Speech, and Signal Processing*, 3:1529–1532, 1995. [51](#)
- [41] J. Orallo, Ma. José Ramírez Quintana, and C. Ferri Ramírez. *Introducción a la minería de datos*. PEARSON Prentice Hall, 2004. [52](#), [61](#)
- [42] A. Webb and A. Copsey. *Statistical pattern recognition 3ed*. Wiley Chichester, 2011. [54](#), [55](#)
- [43] M. Mitchell. *An introduction to genetic algorithms 5ed*. MIT Press Cambridge, 1999. [55](#)
- [44] A. Bashashati, M. Fatourehchi, R. Ward, and G. Birch. A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural Engineering*, (4):32–57, 2007. [59](#), [61](#)
- [45] R. Swarnkar, A. Keskar, P. Prasad, and N. Shivprakash. A new approach to detect p300 in a single trial based on pca and svm classifier. In *2016 IEEE Region 10 Symposium (TENSYMP)*, pages 355–360, May 2016. doi: 10.1109/TENCONSpring.2016.7519432. [59](#)
- [46] S. Tayeb, A. Mahmoudi, F. Regragui, and M. Himmi. Efficient detection of p300 using kernel pca and support vector machine. In *Complex Systems (WCCS), 2014 Second World Conference on*, pages 17–22, Nov 2014. doi: 10.1109/ICoCS.2014.7060953. [59](#)
- [47] S. Xie, Y. Wu, Y. Zhang, J. Zhang, and C. Liu. Single channel single trial p300 detection using extreme learning machine: Compared with bpnn and svm. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 544–549, July 2014. doi: 10.1109/IJCNN.2014.6889400. [59](#)
- [48] B. Blankertz, S. Lemm, M. Treder, S. Hauf, and K. Müller. Single-trial analysis and classification of erp component- a tutorial. *Neuroimage*, 56: 814–825, 2011. [59](#), [60](#)

- [49] D. Krusienski, E. Sellers, F. Cabestaing, S. Bayouth, D. McFarland, T. Vaughan, and J. Wolpaw. A comparison of classification techniques for the p300 speller. *Journal of Neural Engineering*, 3(1):299–305, 2007. [59](#)
- [50] R. Polikar. Ensemble based systems in decision making. *IEEE Circuits and System Magazine*, (6):21–40, 2006. [62](#), [63](#)
- [51] R. Duda, P. Hart, and D. Stork. *Pattern Classification 2 Ed.* Wiley-Interscience, 2000. [63](#)
- [52] New York State Department of Health. Bci laboratory of the wadsworth center, Junio 2006. URL <http://www.bbci.de/competition/iii/>. [65](#), [114](#)
- [53] UAM Universidad Autónoma Metropolitana. Laboratorio de neuroimagenología, Julio 2013. URL <http://akimpech.izt.uam.mx/p300db/doku.php>. [65](#)
- [54] C. Ledesma-Ramirez, E. Bojorges-Valdez, O. Yañez-Suarez, C. Saavedra, L. Bougrain, and G. Gentiletti. An open-access p300 speller database. In *Fourth international BCI meeting, Monterrey, USA, California*, 2010. [65](#)
- [55] G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw. Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Trans. on Biomedical Engineering*, 51:1034–1043, 2004. [65](#)
- [56] L. Farwell and E. Donchin. Talking off the top of your head: toward a metal prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical neurophysiology*, 70:510–523, 1988. [65](#)
- [57] Y. Renard, F. Lotte, G. Gibert, M. Congedo, E. Maby, V. Delannoy, O. Bertrand, and A. Lécuyer. Openvibe: An open-source software platform to design, test, and use brain computer interfaces in real and virtual environments. *Presence*, 19(1):35–53, Feb 2010. ISSN 1054-7460. doi: 10.1162/pres.19.1.35. [65](#)
- [58] L. Bougrain, C. Saavedra, and R. Ranta. Finally, what is the best filter for p300 detection. *Tools for Brain-Computer Interaction (TOBI) Workshop III*, 1, 2012. [67](#)

- [59] R. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. de Ridder, and D. Tax. *PRTools4 - A Matlab Toolbox for Pattern Recognition*, 2004. [70](#)
- [60] J. Buckheit, S. Chen, D. Donoho, I. Johnstone, and J. Scargle. *WaveLab 850*, 2005. [71](#)
- [61] C. Arboleda, E. García Cossioa, A. Posada, and R. Torres. Diseño y construcción de un prototipo de interfaz cerebro-computador para facilitar la comunicación de personas con discapacidad motora. *Revista EIA*, (11):105–115, 2009. [73](#)
- [62] J. Ramirez-Cortes, V. Alarcon-Aquino, G. Rosas-Cholula, P. Gomez-Gil, and J. Escamilla-Ambrosio. Anfis-based p300 rhythm detection using wavelet feature extraction on blind source separated eeg signals. In *Intelligent Automation and Systems Engineering*, pages 353–365. Springer, 2011.
- [63] S. Costagliola, B. Dal Seno, and M. Matteucci. Recognition and classification of p300s in eeg signals by means of feature extraction using wavelet decomposition. In *Proceedings of the 2009 international joint conference on Neural Networks*. IEEE, 2009. [73](#)
- [64] R. Acevedo, G. Gentiletti, V. Medina Bañuelos, and L. Rufiner. Extracción de características en icc mediante métodos basados en diccionarios óptimos: Resultados preliminares. In *II Jornadas Argentinas sobre Interfaces Cerebro Computadora*, 2009. [74](#)
- [65] I. Gareis, G. Gentiletti, R. Acevedo, and L. Rufiner. Extracción de características en interfaces cerebro computadoras mediante transformada wavelet discreta. In *XVII Congreso Argentino de Bioingeniería*, 2009. [74](#)
- [66] T. Picton. The p300 wave of the human event-related potential. *Journal of clinical neurophysiology*, 9(4):456–479, 1992. [76](#)
- [67] D. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, May 1995. ISSN 00189448. doi: 10.1109/18.382009. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=382009>. [77](#), [78](#)

- [68] F. Jauregui, V. Peterson, and R. Acevedo. Denoising en el dominio wavelet para detección de p300 en interfaz cerebro computadora. In *XIX Congreso Argentino de Bioingeniería*, 2013. [79](#)
- [69] S. Raudys and R. Duin. Expected classification error of the fisher linear classifier with pseudo-inverse covariance matrix. *Pattern Recognition Letters archive*, 19, 2008. [83](#)
- [70] G. McLachlan. *The clinical neurophysiology primer*. John Wiley & Sons, Inc., 1992. [83](#)
- [71] I. Gareis, R. Acevedo, Y. Atum, V. Medina Bañuelos, L. Rufiner, and G. Gentiletti. Efecto de la cantidad y dimensión de los patrones en una interfaz cerebro computadora basada en discriminante lineal de fisher. In *V Latin American Congress of Biomedical Engineering*, 2011. [84](#)
- [72] I. Gareis, R. Acevedo, Y. Atum, G. Gentiletti, V. Medina Bañuelos, and L. Rufiner. Determination of an optimal training strategy for a bci classification task with lda. In *5th International IEEE EMBS Conference on Neural Engineering*, 2011. [84](#)
- [73] C. Arjona, J. Pentácolo, I. Gareis, Y. Atum, G. Gentiletti, R. Acevedo, and L. Rufiner. Evaluation of lda ensembles classifiers for brain computer interface. *Journal of Physics: Conference Series*, (332), 2011. [86](#)
- [74] L. Zhuo, J. Zheng, F. Wang, X. Li, B. Ai, and J. Qian. A genetic algorithm based wrapper feature selection method for classification of hyperspectral images using support vector machine. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37:397–402, 2008. [100](#)
- [75] D. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Publishing, 1989. [101](#)
- [76] Systat Software Inc. <http://www.sigmaplot.com>, 2016. [111](#)
- [77] C. Lindig León and O. Yáñez Suárez. Optimized detection of the infrequent response in p300-based brain-computer interfaces. *Revista Mexicana de Ingeniería Biomédica*, 34(1):53–70, 2013. [113](#)

- [78] C. Saavedra and L. Bougrain. Wavelet-based semblance for p300 single-trial detection. In *International Conference on Bio-Inspired Systems and Signal Processing BIOSIGNAL 2013*, 2013. [114](#)
- [79] P. Kindermans, H. Verschore, D. Verstraeten, and B. Schrauwen. A p300 bci for the masses: Prior information enables instant unsupervised spelling. *Advances in neural information processing systems*, (25):719–727, 2012. [114](#)
- [80] A. Kubler, V. Mushahwar, L. Hochberg, and J. Donoghue. Bci meeting 2005-workshop on clinical issues and applications. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):131–134, June 2006. ISSN 1534-4320. doi: 10.1109/TNSRE.2006.875585. [118](#)
- [81] K. Li, V. Narayan Raju, R. Sankar, Y. Arbel, and E. Donchin. *Advances and Challenges in Signal Analysis for Single Trial P300-BCI*, pages 87–94. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-21852-1. doi: 10.1007/978-3-642-21852-1_11. URL http://dx.doi.org/10.1007/978-3-642-21852-1_11. [118](#)
- [82] J. Lv and M Liu. Common spatial pattern and particle swarm optimization for channel selection in bci. In *The 3rd International Conference on Innovative Computing Information and Control (ICICIC'08)*, 2008. [119](#)
- [83] K. Aswineshadri and V. Thulasi Bai. Feature selection in brain computer interface using genetics method. In *IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, 2015. [119](#)
- [84] Computador Industrial Abierta Argentina. <http://goo.gl/aA5IwY>, 2016. [119](#)

UNIVERSIDAD NACIONAL DEL LITORAL



Representación Óptima de Señales de Potenciales Evocados Relacionados a Eventos para la Implementación en Interfaces Cerebro Computadoras

Rubén Carlos Acevedo

FICH

FACULTAD DE INGENIERIA Y CIENCIAS HIDRICAS

sinc(i)

INSTITUTO DE INVESTIGACIÓN EN SEÑALES, SISTEMAS E INTELIGENCIA
ARTIFICIAL

Tesis de Doctorado **2017**