



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

ACTA DE EXAMEN DE GRADO

No 00108

Matrícula: 209381861

IMPLEMENTACIÓN DE UNA NUEVA
ESTRATEGIA DE IDENTIFICACIÓN
DEL POTENCIAL P300 PARA
INTERFACES
CEREBRO-COMPUTADORA
UTILIZANDO INFORMACIÓN
ESPECTRAL DEL EEG

En la Ciudad de México, se presentaron a las 12:00 horas del día 29 del mes de agosto del año 2016 en la Unidad Iztapalapa de la Universidad Autónoma Metropolitana, los suscritos miembros del jurado:

DR. DAVID GUTIERREZ RUIZ
M. EN I. OSCAR YAÑEZ SUAREZ
DRA. NORMA CASTAÑEDA VILLA



JORGE AIRY MERCADO GUTIERREZ
ALUMNO

Bajo la Presidencia del primero y con carácter de Secretaria la última, se reunieron para proceder al Examen de Grado cuya denominación aparece al margen, para la obtención del grado de:

MAESTRO EN CIENCIAS (INGENIERIA BIOMEDICA)

DE: JORGE AIRY MERCADO GUTIERREZ

y de acuerdo con el artículo 78 fracción III del Reglamento de Estudios Superiores de la Universidad Autónoma Metropolitana, los miembros del jurado resolvieron:

- APROBAR -

REVISÓ

LIC. JULIO CESAR DE LARA ISASSI
DIRECTOR DE SISTEMAS ESCOLARES

Acto continuo, el presidente del jurado comunicó al interesado el resultado de la evaluación y, en caso aprobatorio, le fue tomada la protesta.

DIRECTOR DE LA DIVISIÓN DE CBI

DR. JOSE GILBERTO CORDOBA HERRERA

PRESIDENTE

DR. DAVID GUTIERREZ RUIZ

VOCAL

M. EN I. OSCAR YAÑEZ SUAREZ

SECRETARIA

DRA. NORMA CASTAÑEDA VILLA



UNIVERSIDAD AUTÓNOMA METROPOLITANA IZTAPALAPA
DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍA

**IMPLEMENTACIÓN DE UNA NUEVA ESTRATEGIA
DE IDENTIFICACIÓN DEL POTENCIAL P300 PARA
INTERFACES CEREBRO-COMPUTADORA
UTILIZANDO INFORMACIÓN ESPECTRAL DEL EEG**

Tesis que presenta:
Jorge Airy Mercado Gutiérrez
Para obtener el grado de:
Maestro en Ciencias en Ingeniería Biomédica

Asesor: **M. en I. Oscar Yáñez Suárez.**

Jurado Calificador:

Presidente: **Dr. David Gutiérrez Ruiz.** CINVESTAV Monterrey

Secretario: **Dra. Norma Castañeda Villa.** UAM Iztapalapa

Vocal: **M. en I. Oscar Yáñez Suárez.** UAM Iztapalapa

Ciudad de México
Agosto de 2016

Dedicada a mis abuelitas, Carmen y Zoila, que fallecieron en los primeros meses de este año. Parece que acordaron acompañarse en el más allá, algo en lo que ambas creían fervientemente. Las dos fueron el pilar de sus familias. Estoy seguro que sus enseñanzas, anécdotas y ejemplo perdurarán por un buen tiempo con nosotros.

A su memoria.

" Items in our world become meaningful when we have willingly torn our flesh in winning them for some resisting force".

- Corey Anton

"The greatest battle is not physical but psychological. The demons telling us to give up when we push ourselves to the limit can never be silenced for good. They must always be answered by the quiet, the steady dignity that simply refuses to give in. Courage. We all suffer. Keep going"

- Graeme Fife

Agradecimientos

Quiero agradecer al CONACYT por la beca para estudios de maestría. Su labor es fundamental para el desarrollo de México y para el de cada uno de nosotros como estudiantes de posgrado.

A mi asesor de tesis, el profesor Óscar Yáñez. Gracias por todo el tiempo e interés dedicados al proyecto; sin su apoyo, confianza y paciencia a lo largo de todo este proceso, no hubiera sido posible concluirlo. A cada momento me iluminó con su claridad mental y capacidad de observación.

A las mujeres de mi vida. Antes que nadie, a mi mamá, Angélica, porque su amor y su confianza en mí han sido siempre mi motor; a mi hermana Jessica, por ser a veces como mi hija adoptiva, cuenta conmigo siempre para alcanzar todo tu potencial; a mis tías Gina y Socorro, mis segundas madres, por su cariño incondicional; a mis primas Yuri y Gema por los fines de semana en casa de la abuela y las vacaciones cuando éramos niños; y a Belén, mi compañera en la vida, gracias por todo lo que hemos vivido, enfrentado y superado juntos en estos 7 años.

Gracias a todas por alentarme y apoyarme incondicionalmente a hacer las cosas que me gustan, y por creer en mí, a veces más que yo mismo.

Un agradecimiento especial a mi hermano Jairo, por apoyarme siempre en los momentos más críticos; no lo hubiera logrado sin tu ayuda. Sabes que tu también puedes contar siempre conmigo.

A Chucho y Ale. Gracias por esta gran amistad de más de 15 años. Ustedes están entre los que se cuentan con los dedos de una mano. Gracias por hacerme sentir siempre como parte de su familia y por ser un ejemplo para mí en todo aspecto. Gracias por escucharme y aconsejarme.

Tlalpan, Ciudad de México.

26 de julio de 2016

Resumen

En este trabajo se propone una metodología para clasificar las épocas de EEG registradas durante la operación de la Interfaz Cerebro-Computadora (BCI) conocida como *Deletreador de Donchin*, cuya operación está basada en el *paradigma de evento raro* y la presencia o no de la onda P300 del Potencial Relacionado a Eventos (ERP).

La propuesta es una alternativa al enfoque clásico de amplitud y latencia, como rasgos para clasificar las épocas de EEG post-estímulo como atendidas y no-atendidas (con/sin P300), y está basada en teorías que relacionan el origen de los ERPs con actividad de las diferentes bandas de frecuencia (espectral) del EEG de fondo.

La estrategia extrae información espectral (tiempo-frecuencia) de épocas de EEG de 1 segundo de duración post-estímulo, mediante la Transformada Wavelet Continua (CWT) con wavelet madre Morlet Compleja, y clasifica las épocas mediante una máquina de soporte vectorial con kernel gaussiano (SVM-RBF).

La metodología consta de 4 etapas: *Preprocesamiento y extracción de rasgos*, *Selección preliminar de rasgos*, *Selección de rasgos e hiperparámetros óptimos* y *Prueba con datos no observados*. El desempeño de clasificación se mide con el parámetro *Área bajo la Curva ROC* (AUROC), una medida general de la potencia de predicción para clasificadores binarios, que toma valores entre 0 y 1.

Con este método se lograron identificar dos tipos de actividad EEG, correspondiente a distintos mecanismos cerebrales que participan en la operación del Deletreador Donchin:

- Un mecanismo predominante en épocas correspondientes a estimulaciones de columnas (procesamiento visual del estímulo), que involucra rasgos en latencias de 100 – 200 ms, bandas de frecuencia α , θ y β y canales parieto-occipitales (Oz, PO8, PO7).
- Un mecanismo predominante en épocas correspondientes a estimulaciones de filas (reconocimiento consciente del estímulo), que presenta rasgos en latencias de 300 – 500 ms, bandas de frecuencia δ y θ , y canales frontales, centrales y parieto-occipitales (Fz, C3, C4, Oz, PO8 y PO7).

En la etapa de *selección de rasgos e hiperparámetros óptimos* (entrenamiento) el máximo desempeño de clasificación promedio (en 10 sujetos) se alcanzó al promediar los rasgos en $N = 10$ épocas de EEG, obteniendo un valor de AUROC = 0.88 ± 0.06 para épocas de columnas y de 0.84 ± 0.05 para épocas de filas. En la etapa de *prueba con datos no observados* el mejor desempeño de clasificación promedio (entre sujetos) se alcanzó también con $N = 10$, con un valor de AUROC = 0.88 ± 0.08 para épocas de columnas y de 0.80 ± 0.08 para épocas de filas.

Para el mejor sujeto en la etapa de *prueba con muestras no-observadas* (LAC) se obtuvo AUROC = 0.97 para épocas de columnas y AUROC = 0.95 para épocas de filas, en ambos casos con $N = 10$.

El método propuesto es útil para clasificar las épocas de EEG registradas durante la operación del Deletreador de Donchin, a partir de los coeficientes CWT y el clasificador SVM-RBF: es capaz de identificar y seleccionar, los rasgos e hiperparámetros que optimizan el desempeño de clasificación (AUROC) para cada sujeto y condición (filas o columnas).

El tipo de mecanismo cerebral predominante, además de la homogeneidad de los subconjuntos de rasgos, parecen ser dos factores que contribuyen directamente al desempeño de clasificación, que es mayor para épocas de columnas que de filas en casi todos los casos. Estas diferencias pueden estar relacionadas con inconvenientes propios del diseño basado en filas-columnas de la matriz del Deletreador de Donchin clásico, y su dependencia de la dirección de la mirada.

Estos resultados motivan a continuar explorando métodos basados en rasgos espectrales para clasificar épocas de EEG registradas en la operación del Deletreador de Donchin, pues éstos parecen ser informativos de la actividad cerebral asociada con los estímulos atendidos, y con la presencia del potencial P300.

Índice General

Índice de Figuras	IV
Índice de Tablas	V
1. Introducción	1
1.1. Hipótesis	5
1.2. Objetivo General	5
1.3. Objetivos Específicos	5
2. Antecedentes	6
2.1. Enfoque del Deletreador de Donchin original	6
2.2. Enfoque espectral	9
2.2.1. Origen de los ERPs	10
2.2.2. Enfoque espectral en las BCI	11
3. Fundamentos	18
3.1. Electroencefalograma (EEG)	18
3.1.1. Actividad oscilatoria en el EEG	20
3.1.2. ERPs	22
3.1.3. Componente P300 del ERP	23
3.1.4. Paradigma de evento raro	25
3.2. Interfaces Cerebro-Computadora	25
3.2.1. Clasificación de las BCI	27
3.2.2. Deletreador de Donchin	29

3.3.	Transformada Wavelet Continua	32
3.3.1.	Escalograma	33
3.3.2.	Wavelet Morlet Compleja	34
3.4.	Selección de rasgos	35
3.4.1.	Métodos de filtrado	38
3.5.	Máquinas de soporte vectorial	42
3.5.1.	Clasificadores Lineales	42
3.5.2.	Kernels	43
3.5.3.	Clasificación de margen amplio	46
3.5.4.	SVM. Clasificador de margen máximo	47
3.5.5.	Hiperparámetros de la SVM	49
3.5.6.	Selección del modelo SVM	50
3.6.	Evaluación del desempeño de clasificadores	51
3.6.1.	Métricas de desempeño	52
3.6.2.	La matriz de confusión	52
3.6.3.	Validación cruzada	55
3.7.	Curvas de operación relativa (ROC)	56
4.	Metodología	60
4.1.	Base de datos	62
4.1.1.	Muestra poblacional	62
4.1.2.	Registro de EEG: Hardware y Software	62
4.1.3.	Sesiones de registro	63
4.2.	Pre-procesamiento	65
4.3.	Extracción de rasgos	66
4.4.	Selección preliminar de rasgos	72
4.5.	Selección de rasgos e hiperparámetros óptimos	80
4.5.1.	Promediación de rasgos entre épocas	80

4.5.2. Conjuntos de entrenamiento y de prueba	80
4.5.3. Evaluación de subconjuntos de rasgos e hiperparámetros	82
4.6. Prueba con datos no-observados	88
5. Resultados	91
5.1. Selección preliminar de rasgos	91
5.2. Selección de rasgos e hiperparámetros óptimos	95
5.3. Prueba con muestras no observadas	95
6. Discusión	99
6.1. Selección preliminar de rasgos	99
6.2. Rasgos, hiperparámetros y desempeño de clasificación	100
6.3. Otras observaciones	104
7. Conclusiones	111
7.1. Conclusiones	111
7.2. Trabajo a Futuro	115
A. Escalogramas p y H por sujeto	119
B. Selección de rasgos e hiperparámetros óptimos por sujeto	141
Bibliografía	152

Índice de Figuras

3.1. Modelo de actualización del contexto del potencial P300.	24
3.2. Diagrama a bloques de una BCI	26
3.3. Matriz de símbolos de un Deletreador de Donchin	30
4.1. Diagrama de la metodología del trabajo	61
4.2. Preprocesamiento y extracción de rasgos	67
4.3. Escalogramas MC_{cwt} para una época atendida de filas.	70
4.4. Escalogramas MC_{cwt} para una época no atendida de filas.	70
4.5. Escalogramas MC_{cwt} para una época atendida de columnas.	71
4.6. Escalogramas MC_{cwt} para una época no atendida de columnas.	71
4.7. Selección preliminar de rasgos	77
4.8. Escalograma-H acumulado del sujeto ASG. Épocas de columnas	78
4.9. Escalograma-H acumulado del sujeto ASG. Épocas de filas	78
4.10. Escalograma-p promedio del sujeto ASG. Épocas de columnas	79
4.11. Escalograma-p promedio del sujeto ASG. Épocas de filas	79
4.12. Selección de rasgos e hiperparámetros óptimos	86
4.13. Selección de rasgos e hiperparámetros óptimos. Sujeto ASG (filas).	87
4.14. Selección de rasgos e hiperparámetros óptimos. Sujeto ASG (columnas)	87
4.15. Prueba con datos no-observados	90
5.1. Gráficas de desempeño por sujeto	97

Índice de Tablas

3.1. Bandas de frecuencia del EEG y estados asociados	21
3.2. Disposición de una matriz de confusión	54
5.1. Resultados de la selección preliminar de rasgos. Filas	93
5.2. Resultados de la selección preliminar de rasgos. Columnas	94
5.3. Resultados: selección de rasgos e hiperparámetros óptimos	96
5.4. Resultados de la prueba con muestras no observadas	98

Capítulo 1

Introducción

Las BCI son sistemas que buscan servir como canales alternativos de comunicación, donde *“los mensajes o comandos que envía el individuo al mundo exterior no pasan por las vías de salida normales del cerebro, los nervios periféricos y los músculos”* [1].

El campo de investigación de las BCI tiene más de 50 años de haber surgido [2], y a lo largo de ese tiempo se ha beneficiado de los continuos avances en el desarrollo de la tecnología para el registro y procesamiento de señales de origen cerebral, y de la investigación en neurofisiología, procesamiento de señales, reconocimiento de patrones, entre otras disciplinas. El impacto de esos avances se refleja en sistemas BCI cada vez más rápidos y robustos, que poco a poco se presentan como verdaderas alternativas de comunicación e interacción para personas en situaciones de incapacidad motora severa o total, pero que se mantienen conscientes [3].

Entre las señales empleadas en la operación de las BCI, predomina el electroencefalograma (EEG) por sus ventajas en cuanto a costo, no-invasividad, facilidad de uso y resolución temporal, frente a otras técnicas como la resonancia magnética funcional (fMRI), el electrocorticograma (ECoG), y los potenciales de acción neuronales.

Dentro de las señales o fenómenos del EEG usados en la operación de las BCI, el componente P300 del ERP tiene un lugar privilegiado. Esto se debe principalmente a que un alto porcentaje de la gente puede usar el P300 de manera confiable y repetible para operar una BCI con muy poco entrenamiento en comparación con otras señales de control [4], como los ritmos sensorial-motores del EEG o los potenciales corticales lentos, en el caso de estos últimos requiriendo incluso meses de entrenamiento continuo [5].

El componente P300 del ERP se considera un potencial cognitivo, y se presenta cuando el sujeto detecta un cambio inesperado entre una serie de estímulos monótonos. Dicho cambio llama la atención del sujeto, obligándolo a utilizar la memoria de trabajo para comparar el estímulo raro o infrecuente, con los estímulos anteriores, frecuentes.

La primera BCI basada en el P300 fue el Deletreador de Donchin [6], que muestra una matriz de símbolos de 6 filas por 6 columnas y está basada en el *paradigma de evento raro*. Esta BCI presenta al sujeto una serie de estímulos en orden aleatorio, que pueden ser de tipo *atendidos*, con poca probabilidad de aparición (la intensificación de la fila o columna que contiene al símbolo objetivo), o bien *no atendidos*, que son más frecuentes (la intensificación de cualquier fila o columna que no contiene al símbolo objetivo) [7].

Después de intensificar varias veces cada fila y columna, se identifican aquellas con mayor probabilidad de ser *atendidas*, a partir de la presencia del P300 en el EEG. Finalmente, el símbolo elegido se determina como la intersección de la fila y columna *atendidas*, se imprime en la pantalla y repitiendo el proceso se pueden escribir palabras o mensajes

completos.

Para una persona totalmente paralizada y sin posibilidad de comunicarse, este sistema puede ayudar a mejorar su interacción con el mundo. Sin embargo, las BCI basadas en el P300 presentan algunos retos para su operación, entre estos:

- Debido a la presencia de ruido de distintos tipos (EEG de fondo, EMG, artefactos oculares, ruido de línea, etc.) el P300 no es fácilmente detectable en una sola época al usar información de amplitud, por lo que generalmente se promedian varias épocas sincronizadas con el estímulo (promediación coherente), para hacer identificable el potencial al aumentar la relación señal a ruido.
- Las variaciones en la latencia y en la forma del pico del potencial P300 en las épocas individuales de EEG, pueden provocar distorsiones temporales y morfológicas en el potencial promedio y generar información espuria, que a su vez puede afectar el desempeño de los algoritmos de clasificación.

Ante estos inconvenientes del método clásico de operación de las BCI basadas en P300, se han propuesto diversos métodos de identificación del potencial P300, utilizando diversos enfoques de preprocesamiento [8,9], selección de rasgos [10–15] y clasificación [8,16], así como variantes y mejoras a la interfaz del Deletreador de Donchin original (matriz de símbolos y comandos) [17–19].

En este trabajo se presenta un método que explora la identificación de las épocas de EEG post-estímulo que contienen el potencial P300, a partir de información espectral de

las señales EEG obtenidas con el Deletreador de Donchin, utilizando análisis tiempo-frecuencia a partir de la Transformada Wavelet Continua (CWT) y Máquinas de soporte vectorial (SVM) para la clasificación.

El documento tiene la siguiente estructura:

En el Capítulo 2 se hace una revisión de algunos trabajos relacionados con este trabajo, sobre todo los métodos de detección del P300 que utilizan información espectral del EEG para clasificar las épocas de EEG en el contexto del Deletreador de Donchin.

En el Capítulo 3 están las definiciones y bases teóricas de las herramientas y métodos empleados en el experimento: la CWT, técnicas de selección de rasgos, SVMs, etc.

En el Capítulo 4 se explica a detalle el procedimiento empleado en cada parte del experimento de la tesis: preprocesamiento y extracción de rasgos, selección preliminar de rasgos, selección de rasgos e hiperparámetros óptimos y prueba con datos no observados.

En el Capítulo 5 se recopilan en gráficas y tablas los principales resultados del trabajo y en el Capítulo 6 se presenta una discusión acerca de las posibles implicaciones y significado de los resultados.

Finalmente, en el Capítulo 7 se presentan las conclusiones del trabajo, así como algunas ideas y propuestas para futuros proyectos a partir de los resultados aquí obtenidos.

En los Anexos A y B se presentan, para cada sujeto, las gráficas con el resultado de los procedimientos *de selección preliminar de rasgos* y *de selección de rasgos e hiperparámetros óptimos*, respectivamente.

1.1. Hipótesis

- Es posible identificar el potencial P300 a partir de información espectral de la señal electroencefalográfica registrada en la operación de la BCI conocida como *Deletreador de Donchin*.

1.2. Objetivo General

- Implementar un método de identificación del potencial P300 para registros de EEG obtenidos de la operación de la BCI conocida como *Deletreador de Donchin*, que utilice rasgos espectrales de la señal EEG.

1.3. Objetivos Específicos

- Determinar los rasgos espectrales de la señal EEG, más adecuados para diferenciar entre épocas atendidas y no-atendidas en el contexto del Deletreador de Donchin.
- Entrenar un clasificador que optimice la clasificación de las épocas de EEG como *atendidas* o *no atendidas* a partir de un conjunto de rasgos seleccionados por sujeto.

Capítulo 2

Antecedentes

En este capítulo se hace una revisión de los trabajos más representativos relacionados con el enfoque de la tesis: el uso de información espectral del EEG para la identificación del potencial P300 en el contexto del Deletreador de Donchin. Primero se describen brevemente los principios de las métricas de identificación del potencial P300 en el Deletreador de Donchin original, mediante información de amplitud y latencia del pico P300. A continuación se revisan las principales ideas y resultados de los trabajos que han explorado las relaciones entre las diferentes bandas del EEG y los ERPs. Finalmente, se presenta una revisión de trabajos representativos que exploran el uso de la información espectral (en magnitud y fase) del EEG para detectar/identificar el potencial P300 en el contexto de las BCI.

2.1. Enfoque del Deletreador de Donchin original

En el capítulo de *Fundamentos* se describirá a detalle la operación de la BCI conocida como el *Deletreador de Donchin*. En el artículo donde se reporta por primera vez ese sistema

[6], los autores utilizan varias métricas para estimar la amplitud del componente P300 del ERP en cada época post-estímulo de EEG analizada: El algoritmo SWDA, la covarianza con respecto a un template, la búsqueda del pico máximo y medición del área, todo dentro de una ventana de análisis de 600 ms post-estímulo. Esas características o *rasgos* están basados en la amplitud de las muestras de la señal de EEG.

Ese enfoque tiene la característica de que, al ser la señal de EEG cruda el resultado de la suma de diversas fuentes generadoras de actividad eléctrica en un volumen conductor, y siendo por su amplitud (μV) muy susceptible a diversas fuentes de interferencia, tiene una baja relación señal-ruido (SNR). Esto hace necesario promediar varias épocas de EEG para aumentar la SNR y así identificar con mayor certeza las épocas o ventanas de EEG *atendidas*, con P300, de las *no-atendidas*, sin P300. En el caso del Deletreador de Donchin original, se utilizaron 30 estimulaciones con un intervalo inter-estímulo de 125 ms, con lo que alcanzaba una velocidad máxima de deletreo de 1 símbolo cada 26 s (muy baja para una aplicación práctica), con una certeza de clasificación del 95 %.

Los algoritmos SWDA y Covarianza, utilizados para estimar la presencia del P300 en el Deletreador de Donchin original, sufren en su desempeño de deletreo ante la variabilidad de la latencia del pico P300, al estar basados en el promedio en un conjunto de datos de entrenamiento. Esto se debe a que las épocas de EEG que presentan el P300 a una latencia anormal, tienen más probabilidad de no ser identificadas como *atendidas* (con P300).

El algoritmo de localización del pico máximo no se ve afectado por la variabilidad

de la latencia, pues se considera una ventana amplia donde puede aparecer (200 – 500 ms aprox.); éste en cambio se ve afectado cuando se usa un intervalo inter-estímulo corto, como 125 ms, pues se puede atribuir el pico P300 de una estimulación anterior a la estimulación actual.

En la práctica, cada uno de esos algoritmos funcionarán mejor para ciertos sujetos, dependiendo de las características de sus señales.

Desde la publicación original del *Deletreador de Donchin* en 1988 [6], diversos autores han encontrado herramientas para mejorar notablemente el desempeño y la velocidad de deletreo, manteniendo el enfoque basado en información temporal y de amplitud de la señal EEG, pero mejorando diversos aspectos del pre-procesamiento, la extracción de información (rasgos) de la señal EEG, y el algoritmo de identificación del potencial P300 a partir de los rasgos.

Sin embargo, parece haber un límite en el incremento de la SNR que se puede alcanzar con información en amplitud, y por lo tanto en el desempeño y velocidad de las BCI basadas en el P300, ante lo cual se han comenzado a buscar nuevos rasgos de la señal EEG, a partir de las teorías que relacionan a los ERPs con información en el dominio de la frecuencia (espectral) de la señal EEG.

Estas nuevas estrategias buscan encontrar información equivalente o complementaria a la que está codificada en las características de latencia y amplitud del potencial P300 en la señal de EEG. A continuación se hace una breve revisión de algunos trabajos represen-

tativos que siguen dicho enfoque espectral para la identificación del potencial P300, con especial énfasis en las aplicaciones a las BCI.

2.2. Enfoque espectral

En los últimos años, el papel de la señal EEG en el procesamiento cerebral ha sido replanteado. El EEG ya no se considera únicamente como actividad de fondo aleatoria que debe eliminarse para poder observar los ERPs mediante promediación, pues se ha encontrado que los cambios en los ritmos del EEG pueden indicar el procesamiento de información en el cerebro [20] o los cambios en el estado del mismo.

Los cambios en el EEG ante la percepción de un estímulo pueden ser de dos tipos: actividad evocada o actividad inducida, dependiendo de si está acoplada en fase o no al estímulo presentado [20]. La actividad evocada se puede extraer del EEG de fondo promediando las muestras de la señal EEG de varias ventanas post-estímulo. La actividad inducida se obtiene al promediar los espectrogramas (gráficas potencia-frecuencia-tiempo) de las muestras de EEG de las ventanas post-estímulo.

Los cambios en los espectrogramas del EEG pueden darse en la forma de una disminución de potencia o amplitud en una banda de frecuencia (Desincronización Relacionada a Evento ó ERD), o como un incremento de amplitud/potencia (Sincronización Relacionada a Evento ó ERS).

Varios estudios han explorado la dinámica espectral del EEG durante la detección de

un evento *atendido* entre una serie de estímulos no atendidos (*paradigma de evento raro*), y generalmente se ha encontrado un ERS en la banda *theta* (θ), con una amplitud pico cerca de los 300 ms post-estímulo [20]. Esta respuesta en la banda θ se ve afectada por las mismas variables que afectan al potencial P300: la probabilidad de presentación del estímulo infrecuente y la dificultad de la tarea.

La detección de estímulos esperados (atendidos o infrecuentes) también parece estar asociada con una ERD en la actividad de las bandas *alpha* (α) [21] y *beta* (β) del EEG [20]. La relación entre la banda *gamma* (γ) y el estímulo *objetivo* no es tan clara y difiere entre estudios [22] [23].

Estas referencias sugieren en conjunto una posible correspondencia entre la actividad de las diferentes bandas de frecuencia de EEG y la detección de los estímulos atendidos presentados de acuerdo al *paradigma de evento raro*.

2.2.1. Origen de los ERPs

En la sección anterior se vió que existen relaciones entre las bandas de EEG y el reconocimiento de estímulos objetivo (atendidos). Ahora bien, existen investigaciones que sugieren que los ERPs podrían tener su origen, al menos parcialmente, en el comportamiento de las oscilaciones del EEG. Dos modelos predominan en la literatura acerca del origen de los ERPs [24]:

- El modelo *evocado*, que considera al ERP como actividad eléctrica *acoplada al estímulo*,

provocada por el procesamiento del mismo, que se suma al EEG de fondo. Al promediar varias respuestas al mismo tipo de estímulo, se obtiene únicamente el ERP al eliminar el EEG de fondo.

- El modelo de *reinicio de fase*, el cual afirma que el procesamiento del estímulo reinicia la fase del EEG de fondo, de tal forma que la actividad en ciertas frecuencias se *acopla en fase* al estímulo, dando origen al ERP al promediar varias épocas .

Al primer modelo también se la ha llamado *teoría de modulación en amplitud*, y al segundo *teoría de modulación en fase* de la generación del ERP. Desde los años 1980s había argumentos en apoyo de uno u otro modelo, e incluso algunos consideraban que ambas teorías eran complementarias [25], opinión que sigue vigente, sea por la dificultad de distinguir entre ambas teorías o por una verdadera interrelación entre ellas, como se discute ampliamente en [26].

Recientemente, el debate por la prevalencia del *modelo evocado* o del *reinicio de fase* ha retomado importancia, impulsado por el interés en las BCIs basadas en el componente P300 del ERP.

2.2.2. Enfoque espectral en las BCI

Demiralp y cols. en 1999 [11] presentaron uno de los primeros trabajos donde se utiliza información espectral de la señal EEG para detectar el potencial P300, por medio de la descomposición wavelet de los datos en seis grupos de coeficientes, correspondientes a

las diferentes bandas de frecuencia del EEG, de 0 hasta 125 Hz. El rasgo más asociado con la respuesta P300 fue un coeficiente correspondiente a la banda delta del EEG, a los 310–340 ms post-estímulo. Aunque ese trabajo no está relacionado con las BCI, sienta uno de los primeros precedentes para transferir ese enfoque para su aplicación al Deletreador de Donchin, basado en el P300.

Ya en el contexto de las BCI, uno de los primeros trabajos que sugieren la idea de analizar la información en frecuencia de la señal EEG como rasgos para identificar el potencial P300 es [12], donde se utiliza la CWT y la prueba estadística T de Student de dos muestras, para hacer una búsqueda por templete en el espacio *tiempo-escala* de la CWT, a partir de una wavelet madre del tipo *Mexican-Hat*, que asemeja la forma del potencial P300. Para clasificar utiliza LDA (Análisis Discriminante Lineal) y logra identificar correctamente cualquier símbolo del Deletreador P300 (BCI Competition 2003, Dataset Ib) con sólo 4 secuencias de estimulación.

En 2006, Abootalebi y cols. [27] utilizaron la *Transformada Wavelet Discreta* (DWT) con wavelet madre *B-spline cuadrática* y LDA para evaluar la presencia del potencial P300 en una prueba de detección de mentiras. Con la DWT se descompuso la señal de EEG en 6 conjuntos de coeficientes dentro de las bandas de 0 – 4, 4 – 8, 8 – 16, 16 – 32, 32 – 64 y 64 – 128 Hz, que corresponden a distintas bandas de frecuencia del EEG, y con el análisis discriminante se asignó una combinación lineal de los diferentes coeficientes para discriminar mejor entre grupos (inocente/culpable). Para determinar los coeficientes con

mayores diferencias entre grupos, se utilizó una prueba T de muestras independientes con los coeficientes wavelet de ambos grupos. El enfoque empleado en este trabajo difiere del paradigma de evento raro utilizado para el Deletreador de Donchin, sin embargo es un buen precedente del uso de la información tiempo-frecuencia (DWT) para el análisis de señales EEG con el objetivo de evaluar la presencia del potencial P300 en 2 condiciones distintas.

Costagliola y cols. [13] presentan en 2009 un análisis de varios tipos de wavelets madre y diferentes coeficientes de la DWT para reconocer y clasificar los potenciales P300 en señales de EEG obtenidas con el Deletreador de Donchin. Son tal vez los primeros en reconocer que no es necesario elegir una wavelet madre que asemeje la forma del potencial P300 (como *B-splines cuadrática biortogonal* y *Daubechies 4*), y que es posible obtener información del fenómeno con otros tipos de wavelets madre. Hacen el análisis con varios tipos de wavelets madre, porque consideran que algunas de ellas pueden funcionar mejor para ciertos sujetos, de acuerdo a las características de su potencial P300. Utilizan las señales de la base de datos de Deletreador de Donchin de la *BCI Competition* del 2003. Las wavelets madre que analizan son varias versiones de *Coiflets*, *Daubechies*, *Biortogonales* y *Symlets*, y los coeficientes utilizados están en 6 grupos correspondientes a los rangos de frecuencia de 4 – 7, 8 – 15, 15 – 30, 30 – 60 y 60 – 120 Hz y los residuos en 0.5 – 4 Hz. Los 20 coeficientes ligados a las dos bandas de frecuencia más bajas (δ y θ) son los más relacionados con la presencia del potencial P300.

Al tomar en cuenta 10 canales de los 64 disponibles, la matriz de coeficientes candidatos a ser rasgos para la identificación del P300 es de 20×10 . Para cada uno de esos 200 rasgos se lleva a cabo una *prueba T* y se eligen aquellos rasgos que sobrepasan el umbral del 70 % de la energía total (aquellos que están a lo más un 30 % por debajo del máximo valor del estadístico t de la prueba, obtenido para alguno de los 200 coeficientes). Así, se reduce el número de rasgos de 200 a unos 20 – 30 para cada wavelet madre. Se clasifica mediante *análisis discriminante cuadrático* y para 3 wavelets madre (*Daubechies4*, *Coiflets2* y *Symlets4*) se logra identificar correctamente el 100 % de símbolos deletreados con sólo 4 repeticiones de la secuencia de estimulación.

En 2009 Combaz y cols. [14] reportan un estudio similar a [12], donde para clasificar las señales de EEG extraídas con un Deletreador basado en el sistema Donchin original [6], usan el estadístico t de Student de los coeficientes de la CWT con la wavelet madre *Mexican-Hat*, como en [12]. También usan como rasgos para el clasificador muestras de las señales de EEG submuestreadas de 1 kHz a 40 Hz, y la información obtenida con el método de filtrado espacial conocido como *Patrones Espaciales Comunes* (CSP), generalmente usado en Interfaces Cerebro-Computadora basadas en imaginación motora.

Comparan varios métodos de clasificación lineales y no lineales, y el mejor resultado se obtuvo con una Máquina de Soporte Vectorial de Mínimos Cuadrados (LSSVM). El valor del hiperparámetro γ de la SVM se determinó por validación cruzada de 5 vías mediante un procedimiento de retícula (se compara el desempeño del clasificador con cada uno de

los valores de γ). La clasificación se hace por separado para las épocas asociadas a las estimulaciones de filas y para columnas en el Deletreador P300, con lo que alcanzó una certeza de clasificación promedio (para 4 sujetos) de 94.5 % con 4 promediaciones (series de estimulación para cada símbolo del Deletreador).

Otro antecedente más reciente, donde se aplica el análisis tiempo-frecuencia de la señal EEG para detectar el potencial P300 en el contexto del Deletreador de Donchin es el trabajo de Ming y cols. [28], donde se utiliza la CWT con *Morlet* como wavelet madre, a partir de la cual se obtienen dos parámetros espectrales: la Perturbación Espectral Relacionada a un Evento (ERSP) y la Coherencia entre Trials (ITC). Los autores mencionan que la ERSP refleja la influencia de la estimulación en el espectro de frecuencia de la señal EEG, mientras que la ITC es una medida de la homogeneidad de la fase instantánea entre épocas post-estímulo. El estudio indica que, en el promedio de 20 épocas, el ERSP y el ITC son significativamente mayores para las épocas atendidas que para las épocas no atendidas en el intervalo de 200 – 400 ms post-estímulo, lo que hace pensar que ambas medidas contienen información relacionada con la presencia del componente P300 del ERP.

También en 2010, Wang y cols. [29] presentan un algoritmo de detección del potencial P300 en señales obtenidas con el Deletreador de Donchin, que utiliza la DWT con wavelets madre Haar y Daubechies4, y la Distancia de Fisher, en ventanas de 600 ms post-estímulo, para identificar los rasgos que mejor separan las épocas de EEG atendidas (con P300) de las no atendidas. Encuentran que los coeficientes wavelet que mejor separan las 2 clases

de épocas de EEG están en la banda de frecuencia de 0 – 7.5 Hz, logrando una certeza de clasificación del 85 % utilizando una red neuronal artificial (entrenada con el algoritmo Back Propagation) como método de clasificación.

En resumen, los trabajos citados en esta sección, [29] [28] [12] [14] [13] [11] [27], son un antecedente del enfoque espectral para identificar los estímulos atendidos en el Deletreador de Donchin, al llevar los datos del dominio del tiempo al dominio tiempo-escala mediante la Transformada Wavelet. La mayoría de los trabajos utiliza DWT y la wavelet madre *Daubechies4* en particular; sin embargo, los 2 trabajos citados que utilizan la CWT ([12] [14]) logran buenos resultados, con un desempeño comparable o incluso mejor que la DWT. Esto quiere decir que, de alguna manera, la información que se obtiene del fenómeno por ambos métodos (CWT, DWT) es equivalente, al menos en términos del desempeño de clasificación.

Sin embargo, hay que hacer notar que la wavelet madre utilizada en los trabajos que emplean la CWT (*Mexican-Hat*) es más apropiada para buscar actividad transitoria en la señal EEG con una forma similar a un componente del ERP (como el P300), que para analizar la actividad de las oscilaciones del EEG en el dominio de la frecuencia en términos de las bandas clásicas del EEG que, como ya se mencionó en este capítulo, guardan relaciones funcionales con la detección de estímulos auditivos y visuales [20].

Al elegir una wavelet madre con forma que asemeja un componente del ERP (el P300), como la wavelet *Mexican-Hat* o la wavelet *Spline-B cuadrática* para la DWT, lo que se hace

en realidad es una búsqueda por template (la forma de la wavelet madre), a diferentes escalas (ancho del componente del ERP) y latencias (tiempo desde el instante de estimulación), de los posibles componentes del ERP presentes en las diferentes condiciones en que se registra el EEG (tipo de estímulo). En este caso, el análisis tiempo-escala de la Transformada Wavelet se utiliza para identificar los componentes transitorios del ERP. En caso de que la información de interés sea el comportamiento de las distintas bandas de frecuencia del EEG (como en este trabajo), el análisis debe llevarse a cabo en términos de los cambios en amplitud, latencia o fase de las oscilaciones presentes en el EEG, a partir de la presentación de los estímulos visuales.

Con respecto a ese último caso, Bostanov sugirió en 2004 [12] el uso de las wavelet madre *Morlet* y *Morlet Compleja* en lugar de *Mexican-Hat* para obtener *características mejor localizadas en el dominio de la frecuencia*, refiriéndose a la actividad evocada e inducida en el EEG, que en el caso del Deletreador de Donchin se generan a partir del procesamiento cognitivo de los estímulos visuales presentados al sujeto. Esta sugerencia concreta, así como las diferentes evidencias presentadas de la viabilidad de la CWT para la extracción de rasgos en el contexto del Deletreador de Donchin, sirven como base para la metodología desarrollada y presentada en este trabajo de tesis.

Capítulo 3

Fundamentos

En este capítulo se presenta la información que sustenta el enfoque y procedimientos empleados en la tesis, descritos a detalle en el capítulo de Metodología. Se define y describe el EEG, los ERPs y en particular el componente P300 de éstos, así como su relación con la información en frecuencia (espectral) del EEG. A continuación se aborda el tema de las BCIs y se abunda en el funcionamiento del *Deletreador de Donchin*, utilizado en este trabajo. Finalmente se exponen los conceptos básicos de los procedimientos usados en la parte experimental de la tesis: la Transformada Wavelet Continua, que se aplica en la etapa de extracción de rasgos, algunas definiciones y principios de los métodos de selección de rasgos, y finalmente las SVMs, el método de clasificación empleado en la tesis.

3.1. Electroencefalograma (EEG)

En 1924 Hans Berger llevó a cabo los primeros registros electroencefalográficos(EEG), usando tiras metálicas pegadas al cuero cabelludo como electrodos y un galvanómetro sensible (el galvanómetro de cuerda de Einthoven) como instrumento de registro [30],

logrando así el primer registro electrofisiológico no invasivo en humanos realizado de manera sistemática [31].

El Electroencefalograma(EEG) se define operativamente como:

- *"la diferencia de voltaje entre dos lugares de registro diferentes sobre el cuero cabelludo, graficada con respecto al tiempo",*

y fisiológicamente como:

- *"la suma de la actividad eléctrica generada por los potenciales postsinápticos inhibitorios y excitatorios de las células piramidales en la corteza cerebral" [32].*

Las señales en el EEG se ven modificadas por el tiempo que tardan en llegar al electrodo de registro en el cuero cabelludo por varios factores [32]:

- la impedancia de los tejidos que se encuentran entre la fuente de la señal y el electrodo.
- la orientación de la fuente de la actividad eléctrica con respecto al electrodo.
- las propiedades de conducción de los electrodos y de la interfaz electrodo-cuero cabelludo, las cuales dependen del tamaño y geometría del electrodo, su composición, y la resistencia al flujo de corriente producido por la unión del electrodo y el cuero cabelludo.

A los tejidos que se encuentran entre la fuente de actividad eléctrica y el electrodo, y por los cuales fluye la corriente se les llama en conjunto *"volumen conductor"*.

3.1.1. Actividad oscilatoria en el EEG

En general las oscilaciones existen en todos los sistemas biológicos y físicos al tratar de mantener un equilibrio, y se originan cuando el sistema es controlado por dos procesos opuestos, uno que lo aleja y otro que lo regresa al equilibrio, como sucede en los ritmos del EEG con los potenciales postsinápticos excitatorios e inhibitorios. Las oscilaciones del EEG también pueden servir como factor de combinación en la organización de redes o ensambles neuronales [33].

Las redes neuronales están constituidas por poblaciones de elementos neuronales que forman agregados complejos cuyos elementos, bajo circunstancias apropiadas, tienden a trabajar en sincronía. Una característica general de las poblaciones de neuronas corticales, es que su comportamiento oscilatorio y sincronizado, puede reflejarse en la aparición de oscilaciones a nivel macroscópico, susceptibles de ser detectadas a nivel del cuero cabelludo mediante el EEG [34].

El EEG, aunque a primera vista pueda parecer mero ruido, se compone generalmente de señales con comportamiento rítmico, como el ritmo alfa, que se encuentra como una frecuencia dominante de alrededor de 10 Hz, con una amplitud en el rango de 20 a 200 μV [31]. Este ritmo del EEG fue observado desde los primeros registros de EEG realizados por Berger [30]. Existen otros ritmos del EEG con diferentes rangos de frecuencia de oscilación, que se agrupan en bandas de frecuencia y se asocian con distintos estados cerebrales y atencionales, como se ilustra en la Tabla 3.1:

Tabla 3.1: Bandas de frecuencia del EEG y estados asociados.

<i>Banda del EEG</i>	<i>Rango de Frecuencia</i>	<i>Estado cerebral/mental o acción asociada</i>
Delta (δ)	1-4 Hz	Sueño profundo, coma y anestesia
Theta (θ)	4-8 Hz	Somnolencia, meditación, memoria de corto plazo
Alfa (α)	8-13 Hz	Relajación, ojos cerrados, inactividad visual y cognitiva
Mu (μ)	9-13 Hz	Inactividad sensorial/motora
Beta (frontal) (β)	13-30 Hz	Concentración activa, alerta, ansiedad
Beta (central) (β)	13-30 Hz	Inactividad motora
Gamma (γ)	30-40 Hz	Procesamiento activo de información

Cualquier oscilación se describe por su *frecuencia*, *amplitud* y *fase instantánea*. En el EEG estos parámetros pueden contener información importante [35], relacionada con algún estado o condición de interés particular. Por ejemplo, para la mayoría de las bandas de frecuencia del EEG se han encontrado cambios de amplitud en respuesta a procesos cognitivos. De manera general, la amplitud de las bandas delta (δ), theta(θ) y gamma(γ) aumenta durante esfuerzos cognitivos, y las bandas alfa(α) y beta(β) disminuyen su amplitud durante tareas de procesamiento cognitivo activo [35].

3.1.2. ERPs

El método de los ERPs se introdujo en la neurociencia cognitiva en los años 1960s. Se cree que los ERPs reflejan operaciones psicológicas llevadas a cabo en distintos sistemas cerebrales [33]. El flujo de información en el cerebro se puede evaluar mediante los ERPs. Para provocar un ERP se necesita cumplir la condición de *acoplamiento en el tiempo* (bajo un paradigma experimental) a un cierto evento o tarea motora, cognitiva o sensorial. Esas tareas hacen uso de operaciones psicológicas, como la detección y reconocimiento de estímulos (como en el P300), la actualización de la memoria de trabajo, represión de acciones, etc. Dichas operaciones dan origen a un patrón temporal de activación/inhibición neuronal en ciertas áreas del cerebro. Esta actividad se registra en el EEG como una *deflexión en el potencial*, con localización y latencia bien definidas: un componente del ERP.

La manera más común de analizar las señales de EEG *relacionadas a eventos* es mediante el cálculo de los ERPs. Esto se lleva a cabo al presentar de manera repetida un evento de interés, como un estímulo visual en una pantalla de computadora, y analizar la pequeña fracción de la actividad del EEG que se provoca por este evento. Computacionalmente, el ERP se obtiene al extraer épocas de tiempo del EEG acopladas en tiempo a la presentación del estímulo y calculando el promedio sobre las épocas del EEG [36].

Los componentes clásicos de los ERPs están asociados con la latencia y la dirección de las deflexiones del potencial: P100, N100, N200, P200, P300, N400, donde P indica una deflexión positiva, N una negativa (con respecto a una línea base), y el número indica la

latencia pico dada en ms [33].

3.1.3. Componente P300 del ERP

El componente P300 del ERP se reportó por primera vez en los años 1960s por Sutton y colegas. La presencia del P300, como se conoce de manera abreviada este componente, es un índice de la actividad cerebral que ocurre durante la revisión de la representación mental inducida por un estímulo de entrada. Se produce por una red distribuida de procesos cerebrales asociados con la atención y operaciones de memoria [7].

Después de un procesamiento sensorial inicial, se evalúa la representación del evento previo (un estímulo individual dentro de una serie de estímulos) en la memoria de trabajo mediante un proceso de comparación. Si no se detecta un cambio en el tipo de estímulo, se mantiene el modelo mental del contexto del estímulo, y se generan solamente los potenciales evocados sensoriales N100, P200 y N200. Pero si se encuentra un nuevo estímulo, se produce una actualización de la representación del estímulo, con la consiguiente generación del componente P300. Estas ideas dan forma a la *teoría de actualización del contexto* del P300 [7], esquematizado en la Fig. 3.1.

El componente P300 del ERP se mide mediante la amplitud (en μV , respecto al voltaje de la línea base pre-estímulo) y latencia(en ms, respecto al instante de estimulación) del pico máximo positivo del ERP en la ventana de tiempo de 250 – 500 ms a partir del estímulo. El P300 generalmente se registra con mayor amplitud en los electrodos centrales

del sistema 10 – 20: Fz, Cz y Pz.

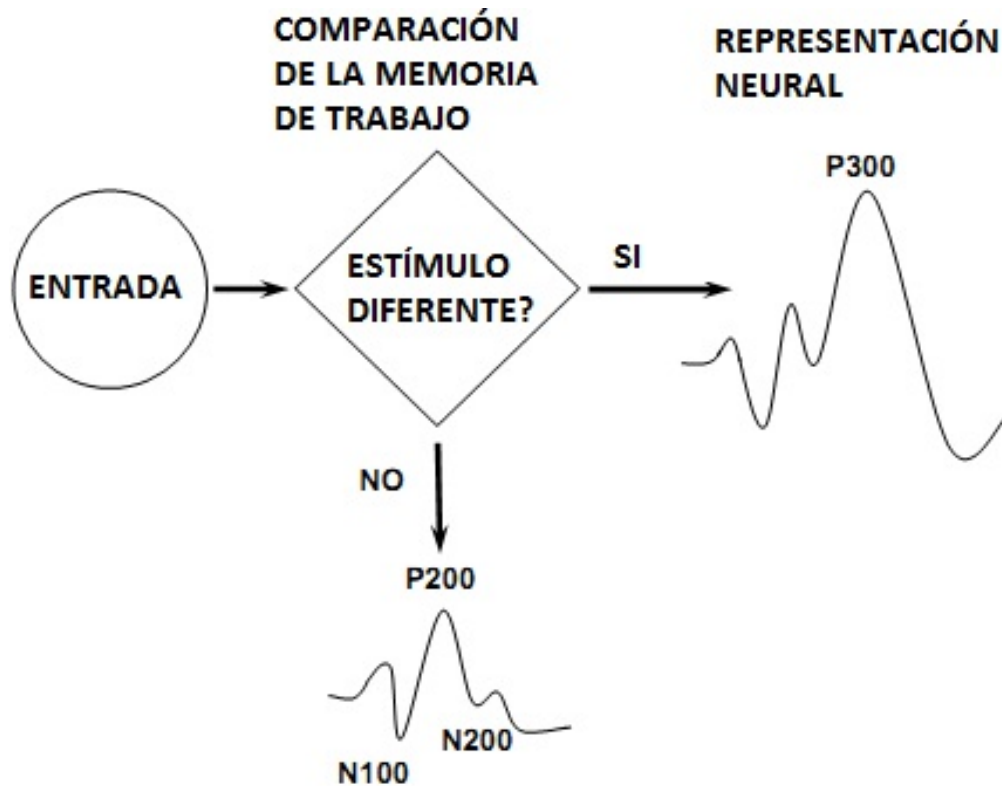


Figura 3.1: Modelo de actualización del contexto del componente P300 del ERP. Ante una serie de estímulos de entrada, un proceso de comparación basado en la atención evalúa si el estímulo es monótono, y genera solo potenciales evocados visuales (N100, P200, N200), o bien, si se trata de un estímulo novedoso o esperado (actualización del contexto), en cuyo caso se genera un potencial P300 (cognitivo) adicional a los componentes visuales. Modificado de [7].

3.1.4. Paradigma de evento raro

Cuando a un sujeto se le presenta una serie de estímulos de 2 categorías distintas, mostrándose una de ellas con menor frecuencia, y se le pide que determine a cual de las 2 categorías pertenece cada estímulo, los eventos infrecuentes o *raros* provocan un ERP con un componente positivo aumentado, a una latencia aproximada de 300 ms, llamado P300. A este tipo de experimento se le llama *paradigma de evento raro* [6].

La amplitud del componente P300 es directamente proporcional a la relevancia de los estímulos que lo provocan, e inversamente proporcional a la probabilidad de dichos estímulos [6]. Además, la aparición del P300 depende de la habilidad del sujeto para discriminar los eventos y asignarlos de acuerdo su categoría .

Por lo tanto, la presencia del componente P300 del ERP indica, mediante una vía alterna de comunicación (no verbal o motora), que el sujeto reconoció el evento *raro*, razón por la cual se considera una señal adecuada para usarse como la base de una BCI.

3.2. Interfaces Cerebro-Computadora

Una BCI es un sistema que traduce las intenciones de un sujeto en comandos para algún dispositivo del mundo externo, empleando información proveniente del sistema nervioso, y transfiriéndola al exterior mediante vías de comunicación alternas a las convencionales (los músculos y nervios periféricos) [1]. En la Fig. 3.2 se muestra un diagrama general de una BCI.

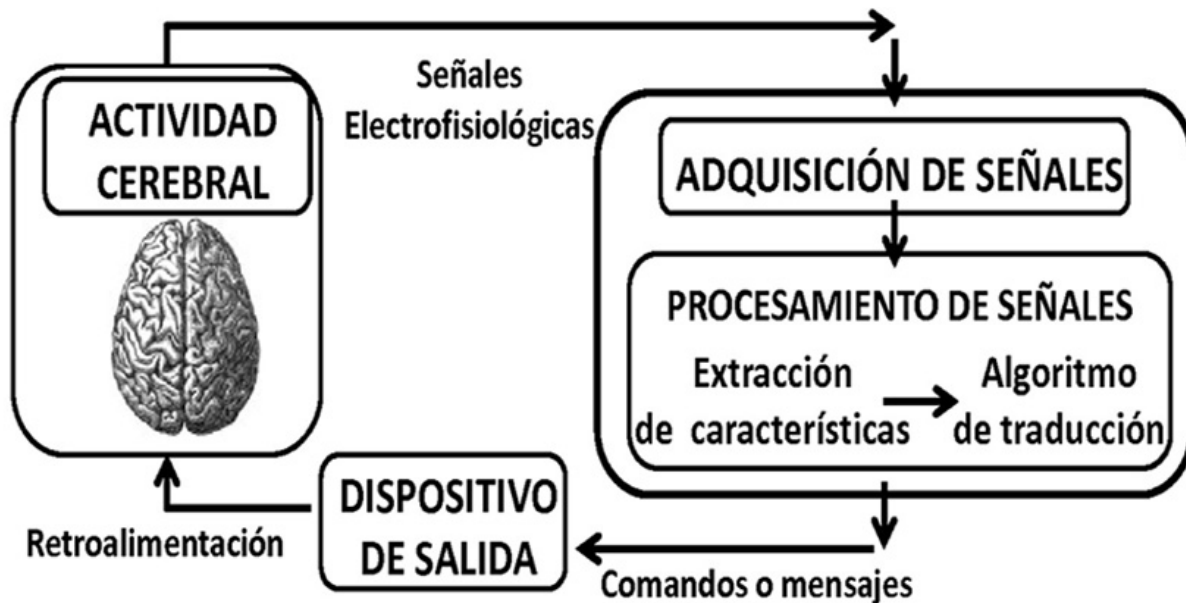


Figura 3.2: Diagrama a bloques de una BCI. Consta de un sistema de adquisición para las señales electrofisiológicas (el EEG), un bloque de procesamiento de señales y un dispositivo de salida, que puede proporcionar retroalimentación al sujeto.

Las BCI buscan proporcionar un medio de interacción y/o comunicación para las personas que padecen incapacidad motora severa, ya sea causada por alguna enfermedad del sistema nervioso (esclerosis lateral amiotrófica, esclerosis múltiple, parálisis cerebral, etc.) o por algún accidente (contusión cerebral o de columna).

Como se aprecia en la figura 3.2, la fuente de información para la operación de una BCI generalmente consiste en señales eléctricas generadas en el cerebro, que el sistema procesa y traduce para interpretar los comandos o mensajes que desea emitir el sujeto. Se necesita además, un dispositivo de salida de información o un actuador, que sirve al mismo tiempo como retroalimentación para el sujeto, un elemento esencial para la operación exitosa de

la BCI.

Se han desarrollado BCI basadas en señales de Electroencefalograma (EEG), Magnetoencefalograma (MEG), Electrocorticograma (ECoG), Resonancia Magnética Funcional (fMRI), Tomografía por Emisión de Positrones (PET) y potenciales de acción de neuronas individuales. Algunas de estas técnicas son muy caras (PET y fMRI) para usarse de manera cotidiana, o bien, necesitan de instalaciones y equipo sofisticado (MEG, PET y fMRI), y además tienen una baja resolución temporal (PET y fMRI) [1]. Las alternativas invasivas (ECoG y potenciales de acción neuronales), con sus desventajas y dificultades obvias, han sido poco utilizadas en humanos, siendo más comunes en investigaciones con animales y en pacientes con electrodos implantados para monitoreo preoperatorio.

El EEG ha demostrado ser la mejor alternativa al tener una buena resolución temporal, facilidad de uso y menor costo de operación, a pesar de su baja resolución espacial, los efectos de la conducción volumétrica y su vulnerabilidad al ruido y artefactos en la señal registrada.

3.2.1. Clasificación de las BCI

Los sistemas BCI pueden ser de dos tipos: *dependientes* o *independientes*. En un sistema BCI *independiente* los nervios periféricos y los músculos del sujeto no tienen un papel importante en su operación, pues no requiere la acción de estos para generar la actividad cerebral (el EEG), en la cual se encuentra codificada la elección o mensaje que el sujeto quiere enviar al mundo exterior. Una BCI dependiente, por el contrario, sí requiere de la

acción consciente de músculos y los nervios periféricos que los activan para generar la señal EEG de interés, aunque esas vías de salida del cerebro no se usen directamente para conducir el mensaje [1].

Wolpaw y colaboradores, en su artículo de revisión clásico sobre BCI [1], las clasifica en cinco categorías, según el tipo de señal que utilizan para su operación:

- Potenciales evocados visuales de estado estacionario (SSVEPs)

- Potenciales Corticales Lentos (SCPs)

- Potenciales provocados P300

- Ritmos sensorial-motores μ y β

- Potenciales de acción

De estos cinco grupos, solo los que están basados en SSVEPs son sistemas BCI *dependientes*. También han surgido recientemente BCI basadas en sensores ópticos (NIRS) [37], y una gran variedad de sistemas BCI *híbridos*, que combinan más de un enfoque para mejorar su desempeño, grado de control o velocidad [38].

Este trabajo de tesis está basado en el sistema BCI conocido como *Deletreador de Donchin*, que opera en base al componente P300 del ERP. A continuación se describen los principios de operación de esta BCI, que es probablemente la más utilizada y estudiada.

3.2.2. Deletreador de Donchin

En 1988 Farwell y Donchin publicaron un artículo [6] donde se presenta un sistema BCI basado en el componente P300 del ERP. Aunque en ese artículo no se encuentra el término BCI (se habla de un *canal biocibernético*), el sistema ahí reportado es la base de la BCI ampliamente conocida como *Deletreador de Donchin* en la actualidad. El sistema busca servir como un canal de comunicación entre la computadora y el usuario, con el objetivo principal de ayudar a las personas con discapacidad motora que les impide comunicarse del todo.

El sistema permite al sujeto usar un *switch* poniendo su atención a uno de entre una serie de *eventos de estimulación*. La discriminación entre el evento en el cual está enfocándose el sujeto y los otros eventos en la serie, codifican la información que el sujeto desea comunicar. Al detectar cuál de los eventos en la serie genera un P300, el algoritmo implementado puede identificar y enviar el mensaje que el sujeto trata de comunicar.

Para llevar esto a cabo, el sistema presenta una matriz con 36 símbolos y comandos simples en la pantalla de un monitor controlado por una computadora (Fig. 3.3). Los *eventos de estimulación* consisten en la *intensificación* (iluminación) de una fila o columna de la matriz. El sujeto intenta comunicar el contenido de una celda de la matriz a la vez. Conforme el sujeto enfoca su atención en una celda o símbolo en particular, la columna y la fila (su intensificación) que contienen a la celda se convierten en eventos *relevantes*.

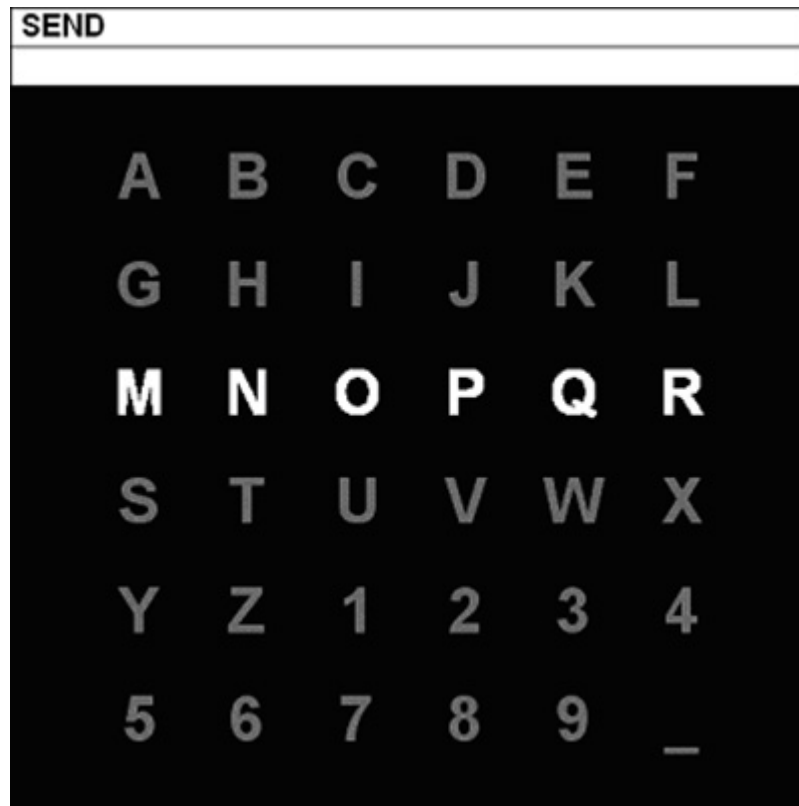


Figura 3.3: Matriz de 6×6 símbolos de un Deletreador de Donchin. La fila que se muestra iluminada se denomina un estímulo atendido solo si el símbolo objetivo (aquel que el sujeto desea deletrear) es alguna de las letras M, N, O, P, Q o R, y como un estímulo no atendido en cualquier otro caso. Las épocas de EEG sincronizadas al instante de presentación de ese estímulo se consideran, bajo la misma lógica, como épocas atendidas o no atendidas. Tomada de [39].

Hay 12 posibles eventos: 6 filas y 6 columnas, de los cuales solo 2 son relevantes a la tarea en cuestión, y se consideran *raros* o *infrecuentes* en el contexto del *paradigma de evento raro*. Por lo tanto, cualquier intensificación que contiene a la celda o símbolo en la cual el sujeto se está enfocando, debería provocar un potencial P300. En ese trabajo [6], mediante varios algoritmos se evalúa la amplitud del P300 posterior a cada estimulación (intensificación), y el símbolo atendido por el sujeto se identifica, después de varias secuencias de intensificación, como la intersección de la fila y la columna que provocan el P300 más grande.

La arquitectura y el principio de operación básico del Deletreador de Donchin arriba descrito, son la base de diversos sistemas BCI [40] [41], y se sigue investigando cómo mejorar la operación del mismo [3], sobre todo en el aspecto que más preocupó desde su publicación en 1988 a los autores: la baja tasa de transferencia de información y velocidad de comunicación del sistema, debido a la necesidad de presentar múltiples series de estimulación al sujeto para obtener, por promediación coherente, un potencial P300 claramente identificable, asociado a los eventos *atendidos*, para así disminuir el número de errores de deletreo. Al final lo que se busca es mejorar el compromiso entre velocidad y precisión del sistema.

3.3. Transformada Wavelet Continua

La Transformada Wavelet es una técnica popular para llevar a cabo análisis tiempo-frecuencia y pertenece a la familia de técnicas de *Análisis Conjunto Tiempo-Frecuencia* (JTFA) que incluyen a la Transformada de Fourier de tiempo corto (STFT), la Transformada Wigner Ville (WVT), y la Transformada Hilbert-Huang (HHT). A diferencia de otras técnicas JTFA de resolución fija, la Transformada Wavelet permite una resolución variable y facilita una mejor resolución temporal para frecuencias altas y una mejor resolución en frecuencia para frecuencias bajas.

Las herramientas clásicas de procesamiento de señales, como la transformada de Fourier, no son adecuadas para analizar señales no estacionarias (como las fisiológicas), porque en su formulación está implícita una suposición de estacionaridad de la señal. La CWT es una generalización de la Transformada de Fourier de tiempo corto (STFT) que permite el análisis de señales no estacionarias a múltiples escalas (la STFT tiene una escala de tiempo fija).

La CWT hace uso de una ventana de análisis (las wavelets) para extraer segmentos de la señal. La ventana se traslada en el tiempo, y además se dilata o contrae dependiendo de la escala de la actividad bajo estudio. Una wavelet dilatada aumenta la sensibilidad de la CWT a eventos con escalas de tiempo grandes, y una wavelet contraída aumenta la sensibilidad a eventos con escalas de tiempo cortas.

La expresión matemática para la transformada Wavelet Continua es:

$$C(s, \tau) = \int \frac{1}{\sqrt{s}} \Psi\left(\frac{t - \tau}{s}\right) x(t) dt \quad (3.1)$$

La correlación entre la señal $x(t)$ y la wavelet $\Psi(t)$ se define como la integral de su producto. La wavelet se traslada por τ y se contrae o dilata por un factor s , antes del cálculo de la correlación con la señal $x(t)$. La CWT mapea $x(t)$ a una función $C(s, \tau)$ que determina la similitud entre $x(t)$ y una wavelet escalada por s a un tiempo τ [42].

La Transformada Wavelet puede considerarse de 2 formas: como un filtrado espectral sobre muchas escalas de tiempo o bien, como un filtro lineal en el tiempo $\Psi[(t - \tau)/s]$ centrado en un tiempo τ con escala s , que se convoluciona con la serie de tiempo $x(t)$.

3.3.1. Escalograma

A la gráfica de la correlación entre la señal y las wavelets escaladas se le llama *escalograma*, el cual permite apreciar como varía en el tiempo la actividad de una señal en un rango de tiempo-escalas. Para valores pequeños de escala s hay una alta resolución temporal y una pobre resolución en frecuencia; para valores grandes de s hay una alta resolución en frecuencia, pero pobre resolución temporal. La CWT permite separar los eventos de corta escala de tiempo (mayor frecuencia) en una porción del plano tiempo-frecuencia y los eventos de gran escala de tiempo (menor frecuencia) en otra región, y analizar simultáneamente la actividad de la señal en múltiples escalas. La CWT es redundante porque varía

el parámetro de escalamiento de la wavelet s de manera continua, con más de un orden de magnitud más wavelets que los componentes originales de la señal. Para extraer información de una descomposición wavelet y eliminar la mayor parte de la redundancia, se pueden considerar solo los mínimos y máximos locales de la CWT.

3.3.2. Wavelet Morlet Compleja

En la CWT el kernel o ventana de análisis es la wavelet $\psi(t)$. Para datos en forma de ondas, una buena elección para la wavelet de análisis es la wavelet Morlet compleja:

$$\psi(t) = e^{-(t/c)^2} e^{i2\pi f_0 t} \quad (3.2)$$

donde t es el tiempo, f_0 es un parámetro de frecuencia, y c es un parámetro de atenuación con unidades de tiempo [43]. Esta ecuación describe una función en el dominio del tiempo, que es el producto de una gaussiana y una exponencial compleja, y cuya frecuencia central es f_0 . La CWT utilizando una wavelet Morlet Compleja tiene una conexión cercana a la Transformada de Fourier.

El parámetro de atenuación c controla la razón a la que se lleva hacia cero la wavelet en el dominio del tiempo, y el espectro en el dominio de la frecuencia. Conforme se incrementa la localización en el tiempo (c grande), la localización en frecuencia disminuye, y viceversa. En esta forma de escribir la wavelet Morlet compleja, se enfatiza el papel central del factor de atenuación c . El valor de este parámetro tiene un efecto de primer orden en cualquier resultado de la CWT.

Cuando se usa para calcular la CWT, el argumento de la wavelet de análisis se traslada y se escala como $\psi(\frac{t-\tau}{s})$. El efecto del escalamiento es doble; en la exponencial real, el valor de la escala multiplica al parámetro de atenuación, ensanchando efectivamente la extensión de la wavelet en el dominio del tiempo. En la exponencial compleja, la escala divide la frecuencia, disminuyéndola precisamente lo suficiente para mantener el número de oscilaciones en el dominio del tiempo. Si dt es la frecuencia de muestreo de la señal y la frecuencia más alta de la CWT es la frecuencia de Nyquist ($dt/2$), y se quiere una expresión completa de la wavelet en cada escala, entonces el valor apropiado de la escala es $c = dt$. Esto permite una localización fina de la CWT en el tiempo, a costa de una pobre resolución en frecuencia. En aplicaciones donde se necesita una mejor localización en frecuencia, es mejor usar un valor de c más grande.

3.4. Selección de rasgos

El aprendizaje maquina es una rama de las ciencias de la computación interesada en reproducir las capacidades de aprendizaje humanas con programas computacionales. Un problema de aprendizaje maquina ocurre cuando una tarea se define como una serie de casos o ejemplos más que por reglas predefinidas [44]. Dado un número de *ejemplos de entrenamiento* (observaciones o muestras) asociados con unos resultados deseados, el proceso de aprendizaje maquina consiste en encontrar la relación entre los patrones y los resultados, usando solamente la información de los *ejemplos de entrenamiento*.

El objetivo en un problema de aprendizaje maquina es predecir el resultado desconocido para nuevos *ejemplos de prueba*, y al desempeño de predicción en los *ejemplos de prueba* se le denomina *capacidad de generalización*. Para llevar esto a cabo, se necesita construir un *modelo predictivo*, que generalmente es una función con parámetros ajustables (una máquina de aprendizaje o clasificador). Los *ejemplos de entrenamiento* se usan para seleccionar un conjunto óptimo de parámetros del modelo predictivo [44].

En un problema de aprendizaje maquina, los *datos* están representados por un número de *rasgos* que pueden ser binarios, categóricos o continuos. Un *rasgo* es sinónimo de *variable de entrada* o *atributo*. La *construcción de rasgos* es uno de los pasos clave en el proceso de análisis de datos, y tiene un efecto mayor en el éxito de cualquier procedimiento estadístico o de aprendizaje maquina posterior [44]. La *selección de rasgos* es parte del proceso de *construcción de rasgos* y se lleva a cabo principalmente para:

- Seleccionar los rasgos relevantes e informativos
- Reducir las necesidades de almacenamiento de datos
- Disminuir tiempo de ejecución del algoritmo
- Reducir el conjunto de rasgos
- Mejorar el desempeño de predicción
- Comprender mejor o visualizar los datos

En general hay tres tipos de métodos de selección de características [44]:

Métodos de filtrado. Se denomina así a los métodos que seleccionan *rasgos* sin optimizar el desempeño de un predictor, y proporcionan un ordenamiento completo de los rasgos mediante algún índice de relevancia. Entre estos índices de relevancia están los coeficientes de correlación, el criterio de Fisher, y algunas pruebas estadísticas (prueba T, prueba-F, χ^2 , etc.)

Métodos wrapper. Involucran al clasificador como parte del proceso de selección de rasgos. Utilizan una máquina de aprendizaje como una caja negra para evaluar subconjuntos de rasgos de acuerdo a su poder predictivo.

Métodos embebidos. Llevan a cabo la selección de características de manera simultánea al proceso de entrenamiento y son específicos al clasificador empleado.

Los métodos *wrapper* y los *embebidos* pueden obtener subconjuntos de rasgos muy distintos ante pequeñas variaciones del conjunto de datos. Los métodos de *filtrado* y los métodos *wrapper* se diferencian principalmente por el criterio de evaluación, pero ambos pueden hacer uso de estrategias de búsqueda para explorar el espacio de posibles combinaciones de rasgos, que es a menudo muy grande para ser explorado de manera exhaustiva.

Existen algunos métodos *híbridos*, en los cuales se usa un filtro para generar una lista de rasgos ordenados; una vez ordenados, se toman subconjuntos anidados de rasgos que se evalúan por medio de una máquina de aprendizaje (clasificador), siguiendo un enfoque *wrapper*.

3.4.1. Métodos de filtrado

Los métodos de filtrado han probado ser bastante efectivos para la selección de rasgos, particularmente cuando el número de éstos es grande y el número de *ejemplos de entrenamiento* es comparativamente pequeño. En esos casos los métodos *wrapper* y *embebidos* pueden ser muy lentos y causar *sobreajuste* [44] (predicción perfecta en *datos de entrenamiento* y baja capacidad de generalización con datos no observados en el entrenamiento).

Los métodos de filtrado de selección de características pueden tener algunas limitaciones, considerando que generalmente se utilizan para ordenar los rasgos por su relevancia individual:

- Los rasgos que no son relevantes de manera individual pueden ser relevantes en el contexto de otros rasgos.
- Los rasgos que son relevantes individualmente pueden no ser útiles por posibles redundancias.

A pesar de estos posibles escenarios, en la práctica se ha visto que los métodos de filtrado simples, al igual que los métodos más complejos y lentos, tienen buenos resultados cuando se combinan con clasificadores como las SVMs [45] [46].

Correlación de Pearson

Los coeficientes de correlación son el enfoque más simple para medir la relevancia de los rasgos. El coeficiente de correlación lineal de Pearson es muy popular en estadística. Es

también un índice de relevancia clásico usado para el ordenamiento de rasgos individuales. Si se denota por x_j al vector de dimensión m que contiene todos los valores del j -ésimo rasgo para todos los ejemplos de entrenamiento, y por y al vector m -dimensional que contiene todos los valores *objetivo* [44], el coeficiente de correlación de Pearson está definido como:

$$C_j = \frac{|\sum_{i=1}^m (x_{i,j} - \bar{x}_j)(y_i - \bar{y})|}{\sqrt{\sum_{i=1}^m (x_{i,j} - \bar{x}_j)^2 \sum_{i=1}^m (y_i - \bar{y})^2}} \quad (3.3)$$

donde la notación de la barra significa el promedio sobre el índice i . Este coeficiente es también el valor absoluto del *coseno* entre los vectores x_i y y_i , después de que han sido centrados (restada su media) [44].

El coeficiente de correlación de Pearson calculado para un rasgo, tratado como una variable aleatoria, y las etiquetas de clase (codificadas con un número entero) se considera una buena medida de la utilidad del rasgo para tareas de *clasificación*. La lista de rasgos ordenados por los valores absolutos decrecientes del coeficiente de correlación, pueden servir como un índice de ordenamiento de acuerdo a la relevancia individual de cada rasgo [44].

Un mayor valor de correlación para un rasgo indica una correlación lineal más alta entre el rasgo y el *objetivo*. Los rasgos con correlación no-lineal a las etiquetas de clase, así como los rasgos sujetos a mucho ruido, resultan en un valor bajo del índice de correlación de Pearson.

Criterio de Fisher

El puntaje de Fisher es una técnica sencilla para medir la discriminación de 2 conjuntos de números reales: Si se tienen unos vectores de entrenamiento \mathbf{x}_k , $k = 1, 2, \dots, m$, y si el número de ejemplos de las dos clases son n_+ y n_- , el puntaje de Fisher para el i -ésimo rasgo es [46]:

$$F(i) = \frac{(\bar{x}_i^{(+)} - \bar{x}_i^{(-)})^2 + (\bar{x}_i^{(-)} - \bar{x}_i^{(-)})^2}{\frac{1}{n_+-1} \sum_{k=1}^{n_+} (x_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n_- -1} \sum_{k=1}^{n_-} (x_{k,i}^{(-)} - \bar{x}_i^{(-)})^2} \quad (3.4)$$

con \bar{x}_i , $\bar{x}_i^{(+)}$, $\bar{x}_i^{(-)}$ el promedio del i -ésimo rasgo de los conjuntos de datos entero, positivo y negativo respectivamente; $x_{k,i}^{(+)}$ es el i -ésimo rasgo de la k -ésima instancia positiva, y $x_{k,i}^{(-)}$ es el i -ésimo rasgo de la k -ésima instancia negativa.

El numerador es un indicador de la discriminación entre los dos conjuntos de datos y el denominador indica la discriminación dentro de cada uno de ellos. Un rasgo tiene más probabilidades de ser discriminativo entre más alto sea su valor del puntaje de Fisher, por lo que se utiliza como un criterio de selección de rasgos.

Prueba T

Para un problema de clasificación de 2 clases, A y B, suponiendo que las distribuciones de los ejemplos de ambas clases son gaussianas, se puede utilizar el estadístico de la prueba T para comparar las medias μ_A y μ_B de las distribuciones gaussianas proyectadas hacia un rasgo en particular. El estadístico de la prueba T es [47]:

$$t = \frac{\mu_A - \mu_B}{s \sqrt{\frac{1}{m_A} + \frac{1}{m_B}}} \quad (3.5)$$

con $s = \frac{(m_A-1)s_A^2 + (m_B-1)s_B^2}{m_A + m_B - 2}$, donde m_A y m_B son el número de datos en cada distribución, y s_A y s_B son la estimación de la desviación estándar de las distribuciones, obtenida de los ejemplos disponibles.

Un rasgo es más informativo si las medias de las dos gaussianas están más separadas. Como índice de ordenamiento se puede usar directamente el valor de t , o bien, el valor p de dos colas de t . Entre más pequeño sea el valor p para un rasgo, mayor será la diferencia entre las medias y el rasgo será más informativo.

Se puede establecer un umbral en el valor p de la prueba por encima del cual los rasgos se consideran estadísticamente significativos. Cuando se prueban muchos rasgos al mismo tiempo se debe hacer una corrección al valor de p .

Cuando las distribuciones de los rasgos no son gaussianas, en lugar de una prueba T se puede utilizar la prueba estadística no-paramétrica *suma de rangos de Wilcoxon-Mann-Whitney*. En ese caso, la hipótesis nula es que las dos clases tienen la misma distribución (arbitraria) de probabilidad proyectadas hacia un rasgo. La hipótesis alternativa es que hay una diferencia en la localización de las distribuciones de las dos clases. Al igual que para la prueba T , los valores p pueden usarse como un índice de ordenamiento de los rasgos. El estadístico de Wilcoxon, W_A , está directamente relacionado al estadístico de Mann-Whitney U_A , el cual equivale al valor de AUROC tomando el valor de un solo rasgo

como el puntaje de predicción de clase [47].

3.5. Máquinas de soporte vectorial

Las SVMs son un método de clasificación moderno que pertenece a la categoría de los métodos de *kernel*. Los métodos de kernel son algoritmos que dependen de los vectores de rasgos solo a través de productos interiores, de modo que el producto se puede reemplazar por una función kernel, que a su vez lleva a cabo un producto interior en un espacio de alta dimensión. Así, usando métodos de clasificadores lineales, se pueden generar fronteras de decisión no-lineales para clasificar los datos [48].

3.5.1. Clasificadores Lineales

Una SVM es un clasificador lineal de dos clases. En un problema de clasificación de dos clases, los datos son objetos asociados a una de dos etiquetas posibles (+1, -1). Un clasificador lineal se define por la siguiente función discriminante [48]:

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (3.6)$$

donde \mathbf{w} es un vector de pesos, \mathbf{x} es el vector de datos de entrada y b es el sesgo.

A un vector de entrada \mathbf{x} se le asigna la clase +1 si $f(x) \geq 0$ y a la clase -1 en otro caso. A la frontera que divide a las regiones clasificadas como positiva (+1) y negativa (-1) se le conoce como *frontera de decisión* del clasificador. La frontera de decisión está definida por

la relación $f(\mathbf{x}) = 0$, que corresponde a un hiperplano $(D - 1)$ -dimensional en el espacio de entrada D -dimensional [49].

El vector \mathbf{w} determina la orientación de la frontera de decisión, y el parámetro de sesgo (b) determina su localización, alejando al hiperplano del origen. Una frontera de decisión se considera lineal si está definida por un hiperplano, porque es lineal respecto a los ejemplos de entrada. A un clasificador con una frontera de decisión lineal se le denomina *clasificador lineal*, y aquél donde la frontera de decisión depende de los datos de manera no-lineal es un *clasificador no-lineal* [48].

3.5.2. Kernels

Se puede usar la formulación de los clasificadores lineales para generar fronteras de decisión no-lineales, mapeando los datos del espacio de entrada X a un espacio de características F usando una función no-lineal ϕ , donde la función discriminante es:

$$f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \quad (3.7)$$

Aplicar un mapeo con una función no-lineal ϕ a un conjunto de datos, resulta en un aumento (cuadrático en el mejor de los casos) en las dimensiones del espacio de características F , así como en el espacio de memoria y el tiempo de cálculo de la función discriminante del clasificador. Esto se complica aún más si el espacio de características F es de dimensiones elevadas, haciendo impráctico el enfoque del mapeo no-lineal. Los métodos

de kernel resuelven este problema evitando mapear explícitamente los datos a un espacio de características de una alta dimensión (no calculan la función ϕ directamente). El vector de pesos \mathbf{w} se puede expresar como una combinación lineal de los ejemplos de entrenamiento [50], $\mathbf{w} = \sum_{i=1}^n \alpha_i \mathbf{x}_i$, de modo que:

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i \mathbf{x}_i^T \mathbf{x} + b \quad (3.8)$$

En el espacio de características esta expresión se convierte en:

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i \phi(\mathbf{x}_i)^T \phi(\mathbf{x}) + b \quad (3.9)$$

Esta formulación de la función discriminante en términos de las variables α_i se denomina *representación dual* de la frontera de decisión. Como ya se mencionó antes, si el espacio de características F es de una dimensión alta, el truco es impráctico a menos que se pueda calcular de manera eficiente la función kernel $\kappa(\mathbf{x}, \mathbf{x}')$, definida como:

$$\kappa(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}') \quad (3.10)$$

En términos de la función kernel, la función discriminante es:

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i \kappa(\mathbf{x}, \mathbf{x}_i) + b \quad (3.11)$$

Al sustituir en las dos ecuaciones anteriores la expresión de ciertas funciones de mapeo no-lineal ϕ , se puede mostrar que el kernel puede calcularse sin llevar a cabo explíci-

tamente el mapeo ϕ . El kernel queda expresado en términos de productos punto de los datos de entrada \mathbf{x} . Al uso de esta propiedad para convertir un clasificador lineal en uno no-lineal se le denomina *el truco del kernel*. Los kernels más utilizados son el polinomial y el Gaussiano.

El kernel polinomial de grado d está definido por:

$$\kappa(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + 1)^d \quad (3.12)$$

El espacio de características para este kernel son todos los monomios con grado $\leq d$. Con $d = 1$ se tiene el kernel lineal (generalmente se omite el término constante 1).

El kernel Gaussiano se define como:

$$\kappa(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2) \quad (3.13)$$

Donde $\gamma > 0$ es un parámetro que controla el ancho de la Gaussiana y $\|\mathbf{x}\|$ es la norma de \mathbf{x} y está dada por $\sqrt{\mathbf{x}^T \mathbf{x}}$. El parámetro γ controla la flexibilidad del clasificador de manera similar al parámetro d del kernel polinomial.

Para que el truco del kernel sea útil en la práctica, el algoritmo de entrenamiento del clasificador también debe ser *kernelizable*. Las SVMs y algunos otros algoritmos cumplen con ese requisito [51].

3.5.3. Clasificación de margen amplio

Un conjunto de datos es *linealmente separable* si existe una frontera de decisión lineal que separe los ejemplos positivos de los negativos (las dos clases). Para un hiperplano dado, se denota por \mathbf{x}_+ (\mathbf{x}_-) al punto más cercano al hiperplano entre los ejemplos positivos (negativos). El margen se define como la menor distancia entre la frontera de decisión y cualquiera de los ejemplos [49]. Geométricamente, el margen de un hiperplano definido por un vector de pesos \mathbf{w} con respecto a un conjunto de datos D es:

$$m_D(\mathbf{w}) = \frac{1}{2} \hat{\mathbf{w}}^T (\mathbf{x}_+ - \mathbf{x}_-) \quad (3.14)$$

siendo $\hat{\mathbf{w}}$ un vector unitario en la dirección de \mathbf{w} y asumiendo que \mathbf{x}_+ y \mathbf{x}_- son equidistantes a la frontera de decisión:

$$f(\mathbf{x}_+) = \mathbf{w}^T \mathbf{x}_+ + b = a \quad (3.15)$$

$$f(\mathbf{x}_-) = \mathbf{w}^T \mathbf{x}_- + b = -a \quad (3.16)$$

para un constante $a > 0$.

Para hacer relevante el margen geométrico, se fija el valor de la función discriminante a los puntos más cercanos al hiperplano, y se hace $a = 1$ en las dos ecuaciones anteriores, de modo que al sumarlas y dividir las por $\|\mathbf{w}\|$, el margen queda expresado como:

$$m_D(\mathbf{w}) = \frac{1}{2} \hat{\mathbf{w}}^T (\mathbf{x}_+ - \mathbf{x}_-) = \frac{1}{\|\mathbf{w}\|} \quad (3.17)$$

3.5.4. SVM. Clasificador de margen máximo

El clasificador de margen máximo es la función discriminante que maximiza el margen geométrico $\frac{1}{\|\mathbf{w}\|}$, equivalente a minimizar $\|\mathbf{w}\|^2$, y que lleva al siguiente problema de optimización con restricciones:

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad (3.18)$$

$$\text{sujeto a: } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, i = 1, \dots, n$$

Si los datos son linealmente separables, la ecuación anterior asegura que el clasificador de margen máximo clasifique cada muestra correctamente. Permitiendo que el clasificador se equivoque con algunos puntos puede ayudar a mejorar el desempeño de clasificación, tanto si los datos son o no linealmente separables. Para permitir estos errores se reemplazan las restricciones de desigualdad de la ecuación anterior por:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i \quad (3.19)$$

Donde ξ_i son *variables de relajación* que permiten que un ejemplo esté sobre el margen o que sea mal-clasificado. Al objetivo de maximizar el margen (minimización de $\|\mathbf{w}\|$) se

agrega un término $C \sum_i \xi_i$, de modo que se penalice tanto la mal-clasificación como los errores de margen. Así, se tiene el siguiente problema de optimización:

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1} \xi_i \quad (3.20)$$

$$\text{sujeto a: } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0$$

La constante $C > 0$ establece la importancia relativa de maximizar el margen y minimizar la cantidad de relajación. A esta formulación se le llama SVM de *margen suave* [52]. Usando el método de multiplicadores de Lagrange, se obtiene la *formulación dual*, que se expresa en términos de las variables α_i :

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j \quad (3.21)$$

$$\text{sujeto a: } \sum_{i=1}^n y_i \alpha_i = 0, 0 \leq \alpha_i \leq C$$

La formulación dual lleva a una expansión del vector de pesos en términos de los ejemplos de entrada:

$$\mathbf{w} = \sum_{i=1}^n y_i \alpha_i \mathbf{x}_i \quad (3.22)$$

Los ejemplos para los cuales $\alpha_i > 0$ son los puntos que están sobre el margen, o dentro del margen en una SVM de margen suave. A estos ejemplos se les llama *vectores de soporte*.

La formulación dual de la SVM depende de los datos solo a través de productos punto, y puede entonces reemplazarse con una función kernel no-lineal, que separe con margen amplio los datos en el espacio de características del kernel.

3.5.5. Hiperparámetros de la SVM

Al entrenar una SVM se busca el hiperplano de margen amplio, es decir, se establecen los valores de los parámetros α_i y b . Además, una SVM con kernel no-lineal cuenta con dos *hiperparámetros*: la constante de margen suave C , y el parámetro de la función kernel (el ancho del kernel gaussiano o el grado de un kernel polinomial).

Para un valor grande del hiperparámetro C , se asigna una penalización grande a los errores de clasificación y los errores del margen, de modo que la orientación del hiperplano se ve afectada por los puntos más cercanos a éste, provocando que el hiperplano esté cerca de muchos otros puntos de datos. Al disminuir el valor de C , los puntos más cercanos al hiperplano se convierten en errores de margen (se ignoran y no determinan la orientación del hiperplano). Así se genera un margen más amplio para los demás datos. El parámetro del kernel controla la flexibilidad del clasificador resultante.

El kernel Gaussiano se define como:

$$\kappa(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2) \quad (3.23)$$

Esta expresión es prácticamente cero si la distancia entre \mathbf{x} y \mathbf{x}' es mucho mayor que $1/\sqrt{\gamma}$.

Para una γ pequeña, un punto de datos dado, x , tiene un valor de kernel distinto de 0 relativo a cualquier otro ejemplo en el conjunto de vectores de soporte. Por lo tanto el conjunto entero de vectores de soporte afecta al valor de la función discriminante en x , lo que lleva a una frontera de decisión suave. Si γ se incrementa, aumenta la *localidad* de la expansión de los vectores de soporte, llevando a una mayor curvatura de la frontera de decisión. Cuando γ es grande, el valor de la función discriminante es prácticamente constante fuera de la vecindad de la región donde los datos se concentran. En tal caso se dice que el clasificador está *sobre-ajustado* a los datos (de entrenamiento).

La flexibilidad de los kernels Gaussiano y polinomial puede causar *sobre-ajuste* en conjuntos de datos multi-dimensionales que cuentan con pocos ejemplos.

3.5.6. Selección del modelo SVM

Los hiperparámetros determinan la frontera de decisión, y al mismo tiempo la certeza de clasificación de la SVM. Para los kernels polinomial y gaussiano, se busca, además del valor del hiperparámetro C , el valor del hiperparámetro específico del kernel, que utilizados para entrenar un modelo SVM optimicen una función de costo, o bien un criterio de desempeño del clasificador en un conjunto de validación. Es más fácil optimizar una SVM con kernel lineal porque el único parámetro que afecta el desempeño es la constante de margen suave (C).

El espacio de búsqueda para las SVMs con kernel gaussiano tiene entonces 2 dimensiones: C y γ . La manera más simple de encontrar los valores óptimos de estos hiperparáme-

tros es evaluando la certeza de clasificación para cada combinación de un conjunto de valores de C y γ . A esto se le conoce como búsqueda por retícula (o cuadrícula). La principal desventaja de este método es que, si la retícula de hiperparámetros es muy densa, el tiempo de cómputo puede llegar a ser muy elevado, haciendo impráctica la búsqueda.

3.6. Evaluación del desempeño de clasificadores

La evaluación del desempeño de un sistema es la última etapa del procedimiento de diseño de un sistema de clasificación. Suponiendo que se ha diseñado un clasificador óptimo, basado en un conjunto seleccionado de vectores de rasgos de entrenamiento, ahora se debe evaluar su desempeño con respecto a la probabilidad de error de clasificación asociado con el sistema diseñado. Una vez que se considera satisfactorio el error estimado, se lleva a cabo una evaluación completa del desempeño del sistema en el ambiente real para el cual el sistema fue diseñado [53]. La evaluación del desempeño de un sistema determinará si el sistema diseñado cumple con los requerimientos impuestos por la aplicación específica y el uso deseado del sistema. Aún más, la probabilidad de mala clasificación también puede usarse como un índice de desempeño en la etapa de selección de características, para elegir los mejores rasgos asociados con un clasificador específico.

3.6.1. Métricas de desempeño

Dado un modelo y un conjunto de datos etiquetado de manera apropiada, tal vez la métrica más intuitiva para el desempeño del modelo es el *error* del modelo. El *error* se define como el número de errores que el modelo comete en el conjunto de datos dividido por el número total de observaciones en el mismo [54]:

$$error = \frac{\text{No. de errores}}{\text{No. total de observaciones}} \quad (3.24)$$

También se puede caracterizar el desempeño de un modelo en términos de su *certeza*:

$$certeza = \frac{\text{No. de predicciones correctas}}{\text{No. total de observaciones}} \quad (3.25)$$

El *error* y la *certeza* de un modelo están relacionados por:

$$certeza = 1 - err \quad (3.26)$$

3.6.2. La matriz de confusión

Cuando se trata con un problema de clasificación binario, hay 4 posibles resultados cuando un modelo se aplica a una observación. Sea $(\bar{x}, y) \in \mathbb{R}^n \rightarrow \{+1, -1\}$ una observación y sea $\hat{f} : \mathbb{R}^n \rightarrow \{+1, -1\}$ un modelo. Entonces se tienen las siguientes cuatro posibilidades cuando el modelo se aplica a la observación [54]:

$$\hat{f}(\bar{x}) = \begin{cases} +1 & \text{si } y = +1, \text{ llamado } \textit{verdadero positivo} \\ -1 & \text{si } y = +1, \text{ llamado } \textit{falso negativo} \\ +1 & \text{si } y = -1, \text{ llamado } \textit{falso positivo} \\ -1 & \text{si } y = -1, \text{ llamado } \textit{verdadero negativo} \end{cases} \quad (3.27)$$

Si la salida del modelo $\hat{f}(\bar{x})$ coincide con la etiqueta y de la observación, tenemos un resultado verdadero positivo o verdadero negativo. Si la salida del modelo no coincide con la etiqueta observada, tenemos un resultado falso positivo o un falso negativo, que en ambos es un resultado erróneo.

En muchos casos es importante distinguir estos dos resultados erróneos cuando se evalúa el desempeño de un modelo. Una representación del desempeño del modelo llamado *matriz de confusión* distingue entre los dos tipos de errores y es por lo tanto la herramienta de elección cuando se analiza el desempeño de un modelo, cuando uno u otro tipo de error puede tener implicaciones serias.

Una matriz de confusión para un problema de clasificación binario es una tabla de 2×2 que despliega las etiquetas observadas contra las etiquetas predichas para un conjunto de datos. Una manera de visualizar la matriz de confusión es considerar que aplicar un modelo \hat{f} a una observación (\bar{x}, y) nos dará dos etiquetas. La primera etiqueta, y , se debe a la observación, y la segunda etiqueta, $\hat{y} = \hat{f}(\bar{x})$, se debe a la predicción del modelo. Esto es, para cada observación, se obtiene un par de etiquetas (y, \hat{y}) . Este par de etiquetas especifican las coordenadas de cada observación dentro de la matriz de confusión: la primera

etiqueta especifica la fila de la matriz y la segunda etiqueta especifica la columna de la matriz. Por lo tanto, una observación con el par de etiquetas (y, \hat{y}) será mapeada a una matriz de confusión como la de la Tabla 3.2 de la siguiente manera:

$$(+1, +1) \rightarrow \text{VP}$$

$$(-1, +1) \rightarrow \text{FP}$$

$$(+1, -1) \rightarrow \text{FN}$$

$$(-1, -1) \rightarrow \text{VN}$$

Para un modelo que no comete errores, todas las predicciones serán mapeadas en los campos superior izquierdo e inferior derecho de la matriz de confusión. Para modelos que cometen errores, los errores serán mapeados de acuerdo al tipo de error que comete el modelo.

Tabla 3.2: Disposición de una matriz de confusión.

		Predicción (\hat{y})	
		+1	-1
Observación (y)	+1	Verdadero positivo (VP)	Falso Negativo (FN)
	-1	Falso Positivo (FP)	Verdadero Negativo (VN)

Dada una matriz de confusión de un modelo como la de la Tabla 3.2, podemos calcular el error del modelo o la certeza del mismo directamente de la matriz de confusión como:

$$error = (FP + FN)/(VP + VN + FP + FN) \quad (3.28)$$

$$certeza = (VP + VN)/(VP + VN + FP + FN) \quad (3.29)$$

respectivamente.

Además de estas métricas, hay otras dos que son comúnmente usadas para caracterizar el desempeño del modelo: la *sensibilidad* y la *especificidad*. La *sensibilidad* de un modelo se define como el número de *predicciones verdaderas positivas* divididas por la suma de todas las *observaciones positivas*:

$$\text{Sensibilidad} = VP/(VP + FN) \quad (3.30)$$

La *especificidad* de un modelo se define como el número de *predicciones verdaderas negativas* divididas por la suma de todas las *observaciones negativas*:

$$\text{Especificidad} = VN/(VN + FP) \quad (3.31)$$

3.6.3. Validación cruzada

En muchas aplicaciones los datos para entrenamiento y validación son limitados y para construir buenos modelos, es deseable usar tantos datos como sea posible para el entrenamiento. En esos casos el conjunto de validación será pequeño y ofrecerá un estimado

ruidoso del desempeño predictivo del clasificador [49].

Una solución a tal problema es usar la técnica conocida como *validación cruzada de k -vías*, la cual permite usar una porción $(k - 1)/k$ de los datos disponibles para el entrenamiento, mientras se hace uso de todos los datos para evaluar el desempeño.

La técnica de *validación cruzada de k -vías* involucra tomar todos los datos disponibles y repartirlos en k grupos, generalmente del mismo tamaño. Después se toman $k - 1$ grupos para entrenar una serie de modelos que son evaluados en el grupo restante. Este procedimiento se repite para todas las k opciones posibles del grupo apartado (en el que se evalúa el desempeño), y al final se promedian los puntajes de desempeño de las k corridas (vías) para estimar el desempeño.

En caso de que los datos sean muy escasos, tomando $k = N$, con N el número total de datos, se tiene el caso de la técnica conocida como *leave-one out*.

3.7. Curvas de operación relativa (ROC)

Una gráfica ROC (del acrónimo en inglés *receiver operating characteristic*) es una técnica útil para visualizar, organizar y seleccionar clasificadores en base a su desempeño. El Análisis ROC se ha extendido de la teoría de detección de señales a la visualización y análisis del desempeño de sistemas de diagnóstico. Se usan de manera común para evaluar la calidad de las decisiones médicas, y cada vez se aplican más para el aprendizaje maquina y la minería de datos, debido en parte al hecho de que la *certeza* de clasificación

simple es a menudo una métrica pobre para la medición del desempeño [55].

Las gráficas ROC tienen propiedades que las hacen especialmente útiles en dominios con distribuciones de clases sesgadas y costos de error de clasificación no-balanceados. Las gráficas ROC son ilustraciones bi-dimensionales en las cuales la razón de verdaderos positivos (VP) se grafica en el eje Y, y la razón de falsos positivos (FP) se grafica en el eje X. De manera informal, un punto en el espacio ROC es mejor que otro si está ubicado al *nor-oeste* del primero (la razón VP es mayor, la razón FP es menor, o ambos)

Curvas en el espacio ROC

Muchos clasificadores se diseñan para producir como salida solamente una decisión de clase para cada instancia, en cuyo caso se conoce como un clasificador discreto. Un clasificador discreto que se aplica a un conjunto de prueba, entrega un solo par (FP,VP), y una sola matriz de confusión, que a su vez corresponde a un solo punto en el espacio ROC.

Algunos clasificadores como el *clasificador bayesiano ingenuo* o las *redes neuronales* o las SVM, entregan de manera natural una probabilidad o *puntaje* a cada instancia; un valor numérico que representa el grado al cual la instancia es un miembro de una clase. Estos valores pueden ser probabilidades, que se adhieren a teoremas estándar de probabilidad, o pueden ser puntajes generales, no calibrados, en cuyo caso la única propiedad que importa es que un mayor puntaje indique una mayor probabilidad de pertenencia a una clase, a pesar del hecho de que la salida no sea estrictamente una probabilidad [55].

Se recorre un umbral, de 0 a 1, sobre la probabilidad posterior (o puntaje) entregada por el clasificador, asignando la clase +1 si la probabilidad está por encima del umbral y la clase -1 en otro caso. Cada valor de umbral produce un punto diferente en el espacio ROC, y con el barrido se traza una curva a través del espacio ROC: una curva ROC.

Las curvas ROC tienen una propiedad atractiva: son insensibles a cambios en la distribución de clases de los datos. Si la proporción de instancias positivas y negativas cambia en un conjunto de prueba, las curvas ROC no cambiarán. A diferencia de otras métricas de desempeño comúnmente utilizadas (certeza, precisión, F-score, etc.) que usan valores de ambas columnas de la matriz de confusión, las curvas ROC están basadas en la razón de VP y FP, así que no dependen de las prevalencias de clase [55].

Área bajo la curva ROC (AUROC)

Una curva ROC es una ilustración bi-dimensional del desempeño de un clasificador. Para comparar clasificadores puede ser deseable reducir el desempeño ROC a un solo valor escalar que represente el desempeño esperado. Un método común es calcular el *área bajo la curva ROC*, abreviado AUROC [55].

Dado que el AUROC es una porción del área del cuadrado unitario, su valor estará entre 0 y 1. Sin embargo, dado que una predicción aleatoria produce la línea diagonal entre (0, 0) y (1, 1), que tiene un área de 0.5, ningún clasificador realista debería tener un AUROC menor a 0.5.

El AUROC tiene una propiedad estadística importante: el AUROC de un clasificador

es equivalente a la probabilidad de que el clasificador *ordene* una instancia positiva elegida al azar, en una posición mayor que una instancia negativa elegida al azar, lo que es equivalente a la prueba de rangos de Wilcoxon. El AUROC se usa a menudo como una medida general del poder de predicción de un clasificador [55].

Otra interpretación del AUROC es: *el valor promedio de la sensibilidad para todos los posibles valores de especificidad* [56]. Este índice es especialmente útil en un estudio comparativo de 2 pruebas diagnósticas o clasificadores. Si se van a comparar dos sistemas, es deseable comparar la curva ROC completa más que un punto en particular. El AUROC máxima (1.0) significa que el clasificador es perfecto en la diferenciación entre las clases; esto es, cuando la distribución de los resultados de prueba para las 2 clases no se traslapan.

Capítulo 4

Metodología

En este capítulo se detalla el procedimiento seguido en este trabajo de tesis. Primero se describe la base de datos de registros de EEG utilizada, después el pre-procesamiento y la extracción de rasgos en el dominio tiempo-frecuencia. A continuación se expone el procedimiento de selección preliminar de rasgos, y el promediado de los rasgos entre N épocas de EEG. La siguiente etapa abordada es la generación de conjuntos de datos de prueba y de entrenamiento, la selección de los rasgos e hiperparámetros óptimos y finalmente la prueba con datos no-observados.

En la Fig. 4.1 se presenta un diagrama a bloques simplificado que ilustra de manera general la metodología propuesta en este trabajo. A lo largo del capítulo, en las Figs. 4.2, 4.7, 4.12 y 4.15, se detallan los procedimientos que corresponden a los bloques de *Preprocesamiento y Extracción de rasgos*, *Selección preliminar de rasgos*, *Selección de rasgos e hiperparámetros óptimos* y *Prueba con datos no-observados*, de la Fig. 4.1.

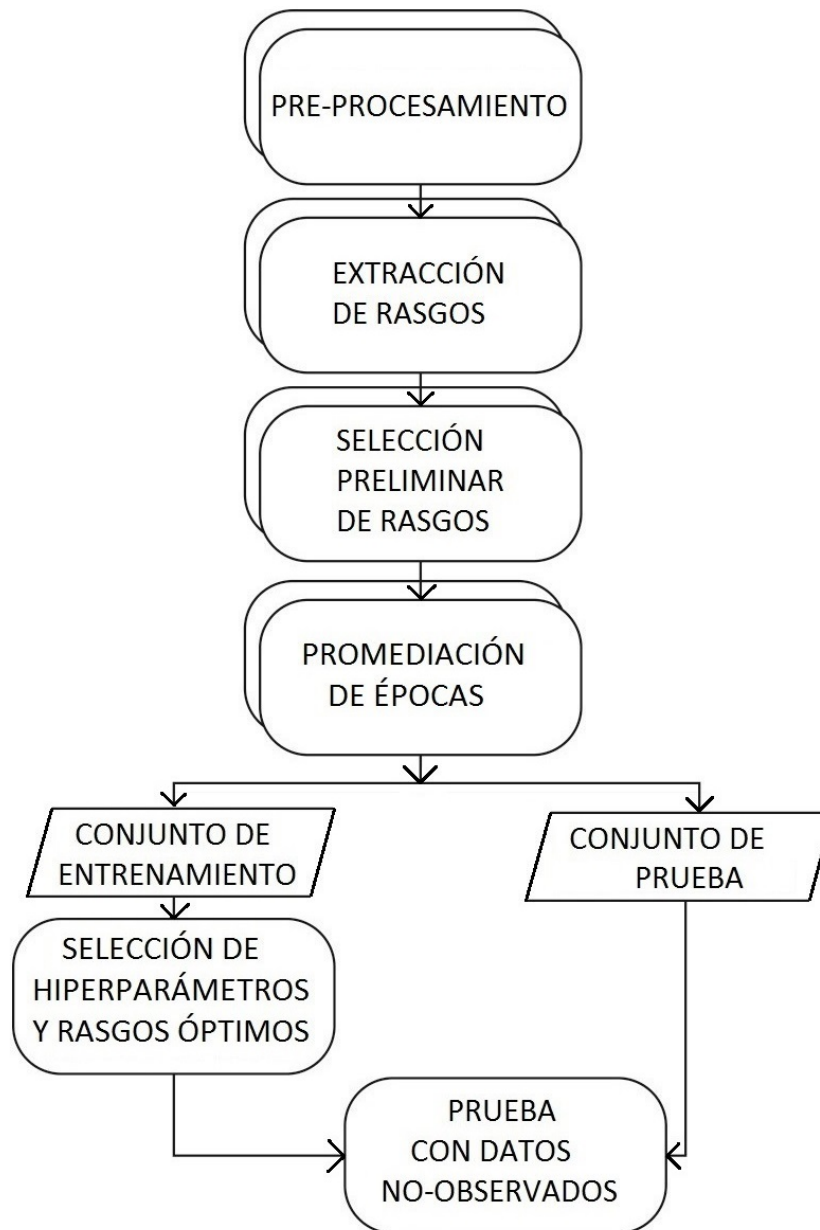


Figura 4.1: Diagrama a bloques de la metodología. En el PRE-PROCESAMIENTO se filtran pasa-banda (0.5 – 30 Hz) los registros de EEG y se extraen todas las ventanas (épocas) de 1 seg. de muestras post-estímulo: atendidas y no-atendidas, de filas o de columnas. A partir de la etapa EXTRACCIÓN DE RASGOS, el procedimiento se realiza por separado para épocas de filas y de columnas. Después del bloque PROMEDIACIÓN DE ÉPOCAS, el proceso se repite para varios valores de N : el número de épocas en las que se promedian los rasgos del EEG.

4.1. Base de datos

Se utilizó una base de datos de registros de EEG, generada con anterioridad [57] en el Laboratorio de Investigación en Neuroimagenología (LINI) de la Universidad Autónoma Metropolitana-Iztapalapa (UAM-I), la cual tiene las siguientes características:

4.1.1. Muestra poblacional

Consiste en 25 sujetos, 9 mujeres y 16 hombres, estudiantes de la UAM-Iztapalapa de 21 a 25 años de edad. Se recopiló información sobre algunas condiciones previas a los registros de EEG: horas de sueño la noche anterior, medicamentos ingeridos, y en caso de ser fumador, tiempo del último cigarro fumado previo al registro.

La metodología propuesta en este trabajo se prueba únicamente en 10 sujetos de la base de datos con los siguientes identificadores: DLP, DMA, ASG, CLL, LAG, LGP, LAC, IZH, GCE y PGA.

4.1.2. Registro de EEG: Hardware y Software

Se utilizó un amplificador de biopotenciales de 16 canales, modelo *g.USBamp* de la empresa *g.tec*, con el que se registró el EEG en 10 posiciones del *Sistema Internacional 10-20* (Fz, C4, Cz, C3, P4, Pz, P3, PO8, Oz y P07) durante la operación de la aplicación *Deletreador P300* de la plataforma experimental para Interfaces Cerebro-Computadora BCI2000 [40], basada en el *Deletreador de Donchin* original [6]. El electrodo de referencia se colocó en el

lóbulo de la oreja derecha y el electrodo de tierra en el mastoides del mismo lado.

Las siguientes opciones del amplificador se establecieron desde BCI2000:

- Filtro Pasabanda: Chebyshev, banda de paso: 0.1-60 Hz, orden 8.
- Filtro Notch: Chebyshev, banda de rechazo: 58-62 Hz, orden 4.
- Frecuencia de muestreo: 256 mps (muestras por segundo).

Las opciones de la aplicación *Deletreador P300* en BCI2000 fueron las siguientes:

- Duración del estímulo: 62.5 ms.
- Duración de la pausa inter-estímulo: 125 ms.
- Tamaño de la matriz de caracteres: 6 filas × 6 columnas (36 símbolos: 26 letras del alfabeto, los dígitos del 0 al 9 y el caracter del guion bajo).

4.1.3. Sesiones de registro

Los sujetos se sometieron a 4 sesiones de registro organizadas de la siguiente manera:

- Sesión 1. Deletreo Dirigido.

Número de secuencias de intensificación por símbolo: 15.

Registro 1. Palabra objetivo: CALOR.

Registro 2: Palabra objetivo: CARINO.

Registro 3: Palabra objetivo: SUSHI.

- Sesión 2. Deletreo Dirigido con matriz de clasificación.

Número de secuencias de intensificación por símbolo: 15.

Registro 1: Palabra objetivo: SUSHI.

- Sesión 3. Deletreo Libre con matriz de clasificación

Número de secuencias de intensificación por símbolo: 15.

De 1 a 4 registros. Palabras objetivo elegidas por el sujeto.

- Sesión 4. Deletreo Libre con menos secuencias de intensificación.

Número de secuencias de intensificación: variable, de 1 a 15.

De 1 a 10 registros. Palabras objetivo elegidas por el sujeto.

Algunos sujetos no llegaron a hacer una cuarta sesión de registro.

En los registros de deletreo dirigido (sesiones 1 y 2), las palabras objetivo están predefinidas y se indican al sujeto uno por uno los símbolos que las conforman, realizando 15 secuencias de estimulación por símbolo. Una secuencia de estimulación consiste en la intensificación (los símbolos contenidos en una fila o columna se iluminan en color blanco) aleatoria de cada una de las 6 filas y las 6 columnas de la matriz de símbolos.

En los registros de las sesiones de deletreo libre (3 y 4) las palabras objetivo son elegidas de manera libre por el sujeto y el número de secuencias de estimulación por símbolo en cada registro varía entre 1 y 15, también a elección del sujeto.

4.2. Pre-procesamiento

Para cada uno de los 10 sujetos considerados, cada uno de los 4 registros de EEG (de deletreo digirido) se puede expresar como $x_{i,ch}(n)$, donde $i \in \{1, 2, 3, 4\}$ representa el número de registro, $ch \in \{1, 2, \dots, 10\}$ el número de canal, $n = 1, 2, \dots, N$ son los instantes de tiempo de muestreo de la señal EEG, y N es el número total de muestras del registro i , que depende del número de símbolos deletreados (5 ó 6). Con la señal de EEG cruda de cada uno de los 10 canales, de cada registro $x_{i,ch}(n)$, se hace lo siguiente (Fig. 4.2):

- Se filtra pasa-banda (0.5 – 30 Hz, FIR, orden 128) la señal de EEG de cada canal ch de los registros $x_{i,ch}(n)$ (Fig. 4.2-A).
- Una época de EEG se expresa como $x_{i,ch}^{k,y}(n)$, con $n = \{1, 2, \dots, 257\}$, y se define como una ventana de 257 muestras (posteriores al instante de estimulación) del registro i y clase $k \in \{a, u\}$, donde $k = a$ indica una época de la clase atendida y $k = u$ una época de la clase no-atendida.

El superíndice $k = a$, corresponde a la época sincronizada con la intensificación de una fila o columna de la matriz del Deletreador P300, que contiene al símbolo objetivo (evento infrecuente), y $k = u$ corresponde a una época de EEG asociada a una intensificación que no incluye un símbolo objetivo (evento frecuente). El superíndice y indica el tipo de época: filas ($y = f$) o columnas ($y = c$).

- En cada uno de los 4 registros de EEG (i) de cada sujeto, se extraen todas las épocas

disponibles y se dividen en 4 grupos (Fig. 4.2-B):

épocas atendidas de filas: $x_{i,ch}^{a,f}(n)$

épocas atendidas de columnas: $x_{i,ch}^{a,c}(n)$

épocas no-atendidas de filas: $x_{i,ch}^{u,f}(n)$

épocas no-atendidas de columnas: $x_{i,ch}^{u,c}(n)$

4.3. Extracción de rasgos

La información espectral de la señal EEG se extrae calculando la CWT a cada época de EEG $x_{i,ch}^{k,y}(n)$, cuyos coeficientes están dados por:

$$C_{cwt}(x_{i,ch}^{k,y}(n), s, m) = \frac{1}{\sqrt{s}} \sum_{n=0}^{M-1} x_{i,ch}^{k,y}(n) \psi^*\left(\frac{n-m}{s}\right) \quad (4.1)$$

donde M es la duración de la señal (en número muestras), $m \in \mathbb{Z}^+$ y $s \in \mathbb{R}^+$ son respectivamente, el número de muestras que se recorre y la escala a la cual se dilata o comprime la wavelet madre ψ [58].

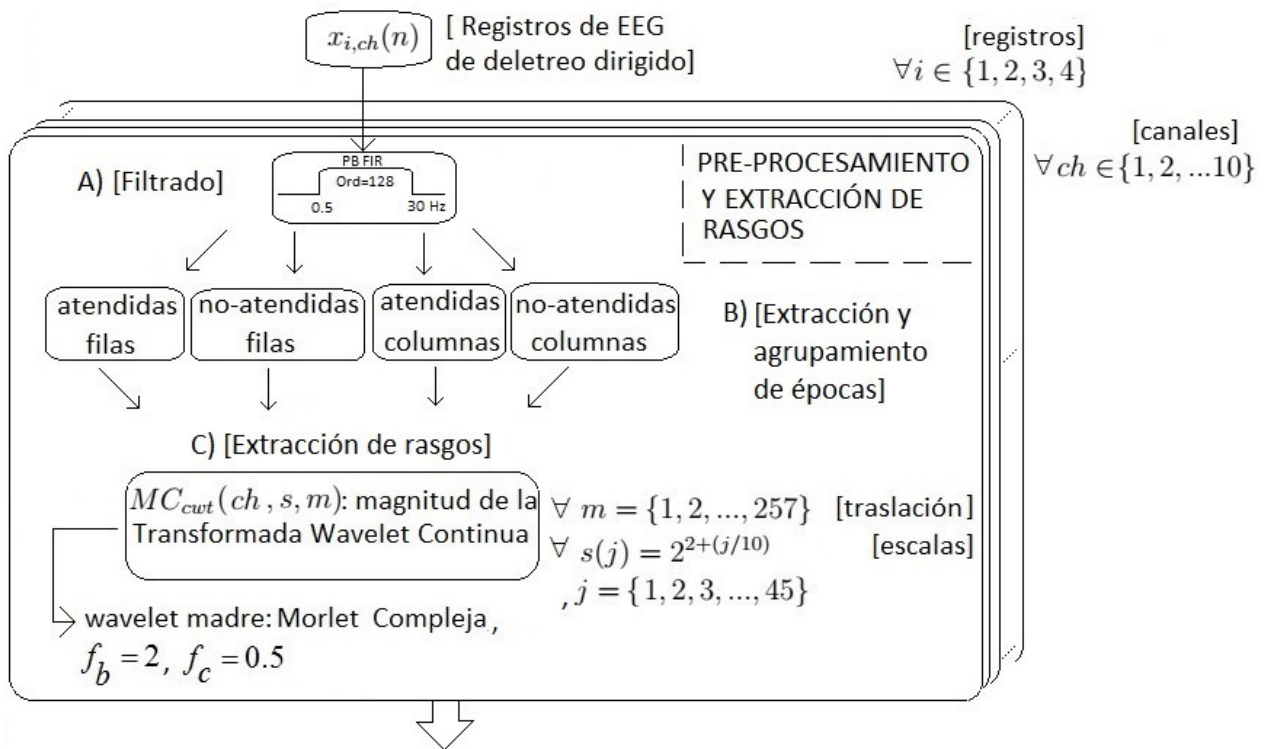


Figura 4.2: Preprocesamiento y extracción de rasgos. (A) Mediante un filtro pasabanda FIR de orden 128, se filtra de 0.5 a 30 Hz la señal cruda de EEG de cada canal (ch), en cada registro de delecteo dirigido (i). (B) De las señales filtradas se extraen todas las épocas de EEG de 1 seg. de duración post-estímulo, y se reparten en 4 grupos: épocas atendidas de filas o de columnas, y épocas no-atendidas de filas o de columnas. (C) Se calcula la Magnitud de la Transformada Wavelet Continua (MC_{cwt}) de cada época, con wavelet madre Morlet Compleja y valores de traslación en el tiempo (m) y escalas (s) en el rango temporal y espectral de interés.

Para calcular la CWT se hace lo siguiente (Fig. 4.2-C):

- Se elige una wavelet madre $\psi(t)$ Morlet Compleja (sección 3.3.2) con parámetros de ancho de banda $f_b = 2$ y frecuencia central $f_c = 0.5$, elegidos de manera heurística a partir de señales simuladas de prueba y épocas reales de EEG. Esta wavelet es apta para analizar la actividad oscilatoria de la señal EEG en el dominio tiempo-frecuencia [12].
- Las escalas utilizadas en la Ec. 4.1 están determinadas por:

$$s(j) = 2^{2+(j/10)}, j = \{1, 2, 3, \dots, 45\} \quad (4.2)$$

La siguiente expresión asocia una wavelet madre ψ con parámetro de frecuencia central f_c , a una señal puramente periódica de frecuencia F_s :

$$F_s = \frac{f_c}{s \times \Delta} \quad (4.3)$$

, donde F_s está en Hz, s es una escala en particular y Δ es el periodo de muestreo ($\frac{1}{256}s$). En este caso, las escalas dadas por la Ec. 4.2 corresponden a pseudo-frecuencias entre 1.41 ($s = 45$) y 30 Hz ($s = 1$), que incluye a las bandas δ , θ , α y β del EEG, de interés en este trabajo.

- Para hacer compatible $\psi(t)$ con la Ec. 4.1, después de ser escalada por cada valor del factor s (Ec. 4.2) se toman $M = 257$ valores de ésta (mediante interpolación), correspondientes a las muestras de las épocas $x_{i,ch}^{k,y}(n)$. Esto resulta, para cada valor

de s , en una versión discreta de la función wavelet madre continua, $\psi(n)$, que es la que finalmente se usa en la Ec. 4.1 para calcular los coeficientes de la CWT.

- En la expresión 4.1 la wavelet madre $\psi(n)$ se recorre a los intervalos de tiempo $m = \{1, 2, \dots, 257\}$, correspondientes a las muestras de las épocas de EEG $x_{i,ch}^{k,y}(n)$.
- A continuación, a cada época $x_{i,ch}^{k,y}(n)$ se calcula su Magnitud Wavelet, como la magnitud de los coeficientes wavelet complejos, obtenidos del cálculo de la CWT en cada punto escala-tiempo-canal (s, m, ch) (Fig.4.2-C):

$$MC_{cwt}(x_{i,ch}^{k,y}(n), s, m) = \left| C_{cwt}(x_{i,ch}^{k,y}(n), s, m) \right| \quad (4.4)$$

Cada punto (s, m, ch) del escalograma de Magnitud Wavelet (Ec. 4.4), se considera un *rasgo* de la señal EEG. Cada rasgo contiene información de la actividad oscilatoria en el dominio tiempo-frecuencia. En las Figs. 4.3, 4.4, 4.5 y 4.6, se muestran ejemplos del escalograma de magnitud wavelet $MC_{cwt}(ch, s, m)$ correspondientes a épocas atendidas y no atendidas, de filas y de columnas, calculados en cada caso a partir de una época individual del registro EEG (sujeto ASG).

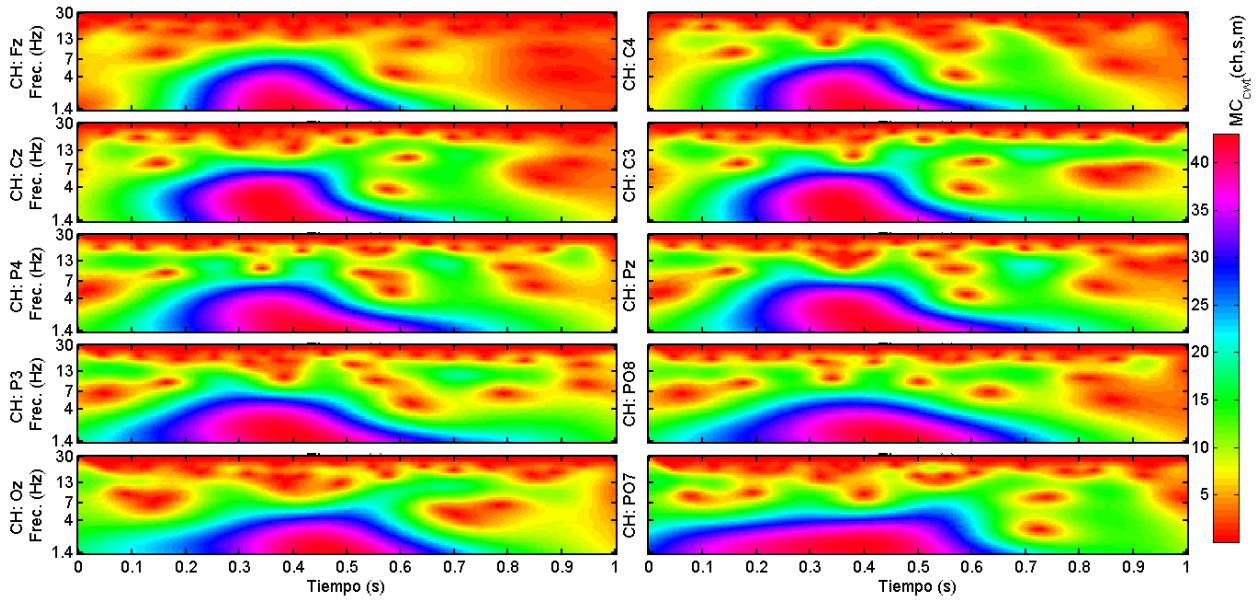


Figura 4.3: Escalogramas del valor $MC_{cwt}(ch, s, m)$ por canal, en una época atendida de filas.

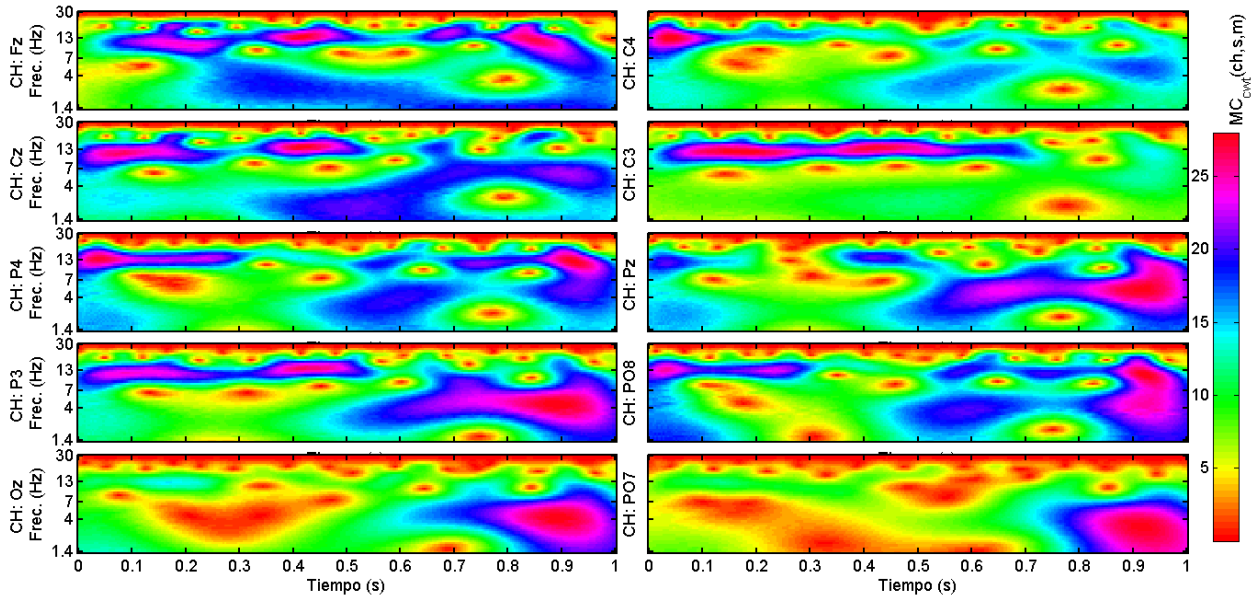


Figura 4.4: Escalogramas del valor $MC_{cwt}(ch, s, m)$ por canal, en una época no atendida de filas.

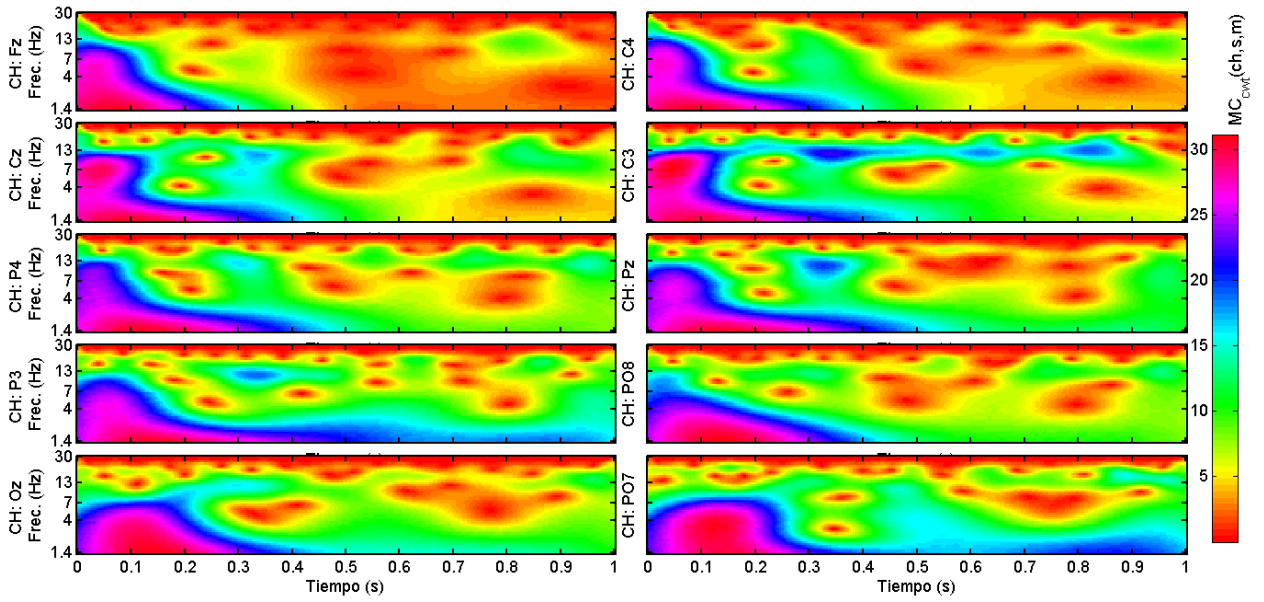


Figura 4.5: Escalogramas del valor $MC_{cwt}(ch, s, m)$ por canal, en una época atendida de columnas.

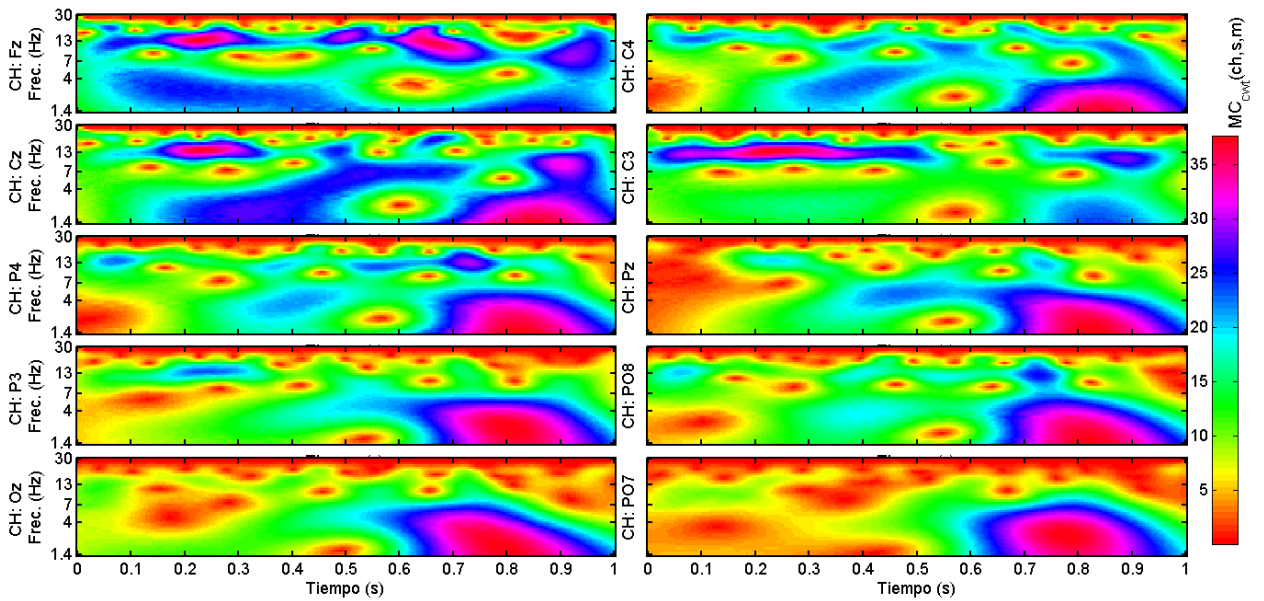


Figura 4.6: Escalogramas del valor $MC_{cwt}(ch, s, m)$ por canal, en una época no atendida de columnas.

4.4. Selección preliminar de rasgos

- El primer paso para seleccionar un conjunto de rasgos óptimo, es obtener una métrica de la utilidad de cada *rasgo* de manera individual, para diferenciar entre las épocas atendidas ($k = a$) y no-atendidas ($k = u$). Para cada rasgo $MC_{cwt}(x_{i,ch}^{k,y}(n), s, m)$, se lleva a cabo una *prueba T de 2 muestras* (no pareada) entre las épocas atendidas y no-atendidas de cada registro de EEG (filas y columnas por separado), con un nivel de significancia estadística del 5% ($\alpha = 0.05$) (Fig. 4.7-A). La prueba se expresa como:

$$T2(MC_{cwt}(x_{i,ch}^{a,y}(n), s, m), MC_{cwt}(x_{i,ch}^{u,y}(n), s, m)) \quad (4.5)$$

donde i , ch e y indican respectivamente el número de registro, el número de canal y el tipo de épocas (filas o columnas, $y = f$ o $y = c$) para los cuales se realiza la prueba estadística.

En total, para cada uno de los 4 registros se llevan a cabo 115650 pruebas estadísticas (45 escalas \times 257 muestras \times 10 canales), para cada valor de y .

- El valor-p que resulta de la prueba T se utiliza para determinar el nivel de relevancia de cada rasgo, en referencia a su capacidad para separar las 2 clases de épocas, usando únicamente los valores de ese rasgo. El umbral del valor-p para considerar a un rasgo como relevante es $p < 0.05$.
- En cada prueba estadística hay dos hipótesis posibles:

- La hipótesis nula, H_0 : afirma que no existe diferencia entre las medias de las dos clases a un nivel $\alpha = 0.05$ para ese rasgo en particular.
- La hipótesis alternativa, H_1 : implica que existe una diferencia en los valores de las medias de las dos clases (se rechaza la hipótesis nula H_0).

Si ahora se define H como la variable que representa el resultado de la prueba estadística, $H = 0$ indica que se acepta la hipótesis nula, y $H = 1$ indica que se rechaza la hipótesis nula y se acepta la alternativa. $H = 1$, es equivalente a cumplir la condición $p < 0.05$, el umbral para considerar a un rasgo como relevante.

- El mapa que se genera a partir del valor-H de todas las pruebas estadísticas realizadas para un registro i se llama *escalograma-H*, $EH_i(s, m, ch)$ (Fig. 4.7-A). El gráfico del *escalograma-H* permite visualizar las combinaciones (s, m, ch) para las cuales los valores de $MC_{cut}(x_{i,ch}^{k,y}(n), s, m)$ presentan diferencias significativas entre clases (los *rasgos relevantes*)
- Para identificar los rasgos más relevantes para identificar las clases, en cada punto (s, m, ch) del escalograma-H se calcula:

$$EHA(s, m, ch) = \sum_{i=1}^4 EH_i(s, m, ch) \quad (4.6)$$

, donde $EHA(s, m, ch)$ es el escalograma-H acumulado de los 4 registros de EEG (Fig. 4.7-B.1), que al graficarse muestra las zonas del espacio tiempo-frecuencia-canal

que concentran los rasgos que son relevantes ($H = 1$) en un mayor número de registros. En las Figs. 4.8 y 4.9 se muestran ejemplos de gráficos de EHA para épocas de columnas y de filas de un sujeto de la base de datos (ASG). A partir de éstas, se pueden apreciar con facilidad zonas del mapa tiempo-frecuencia-canal, en donde los rasgos $EHA = 4$ (zonas color café), $EHA = 3$: (naranja), $EHA = 2$: (verde), $EHA = 1$: (azul cielo), o $EHA = 0$ (azul fuerte). La idea es utilizar en las siguientes etapas aquellos rasgos que tienen los mayores valores de EHA , que en teoría son los más robustos y representativos de la actividad de interés.

- Para determinar los mejores rasgos para un sujeto (*rasgos candidatos*), se hace una corrida de las pruebas estadísticas para todos los rasgos de cada registro i , y se verifica que al menos 100 puntos de $EHA(s, m, ch)$ tengan valor de $k = 4$ (Fig. 4.7-B.2), lo que implica que en cada registro, esos rasgos obtengan $H = 1$ en la prueba estadística ($p < 0.05$).

De no obtenerse el mínimo de 100 rasgos que cumplan la condición $k = 4$, se repite el procedimiento, disminuyendo k de uno en uno, hasta conseguir el mínimo de 100 rasgos relevantes, o hasta que $k = 0$, en cuyo caso se detiene el algoritmo, por no encontrar suficientes rasgos con diferencias entre clases.

- De manera similar al valor-H, el *escalograma-p* es la gráfica del valor-p derivado de la prueba estadística para cada rasgo $MC_{cwt}(x_{i,ch}^{k,y}(n), s, m)$ del registro i , y se denota como $EP_i(s, m, ch)$ (Fig. 4.7-A).

Si a partir de EP_i se calcula:

$$EPP(s, m, ch) = \frac{1}{4} \sum_{i=1}^4 EP_i(s, m, ch) \quad (4.7)$$

, se obtiene $EPP(s, m, ch)$, el *escalograma-p promedio* (Fig. 4.7-C), que sirve para localizar, de entre los llamados *rasgos relevantes*, aquellos con mayor separación entre clases, considerando los 4 registros de EEG.

En las Figs. 4.10 y 4.11 se muestra un ejemplo gráfico del escalograma EPP para épocas de columnas y de filas del sujeto ASG . En el Anexo A se muestran los gráficos de los escalogramas EHA y EPP de todos los sujetos cuyos registros se utilizaron en este trabajo. Los círculos negros, que también están presentes en las Figs. 4.8 y 4.9, indican la localización de los rasgos con menor valor de EPP en cada canal, y el cuadrado blanco con contorno negro indica el rasgo con menor valor de EPP entre todos los canales. Estos puntos son de utilidad para darse una idea de las zonas en que se concentran los rasgos relevantes y como varían entre canales.

- Los rasgos seleccionados en el paso anterior por valor de EHA , se ordenan ascendentemente por valor de EPP y se eligen únicamente los primeros 100 (los que minimizan EPP). Estos forman un conjunto de rasgos potencialmente útiles para predecir la clase de las épocas de EEG: los *rasgos candidatos* o RC_r (Fig. 4.7-D), donde r indica el número de rasgos que contiene (aquí $r = 100$). En los escalogramas EPP

de las Figs. 4.10 y 4.11, los rasgos candidatos se concentran en las zonas de color naranja intenso (rodeadas de amarillo), aquellos con menor valor EPP .

- Para cada uno de los 4 grupos descritos en la Fig. 4.2-B, se conjuntan las épocas de los 4 registros de EEG, pero tomando solo los rasgos $\in RC_r$. Esos rasgos generalmente son distintos en épocas de filas y de columnas, como el ejemplo de las Figs. 4.10 y 4.11.

Finalmente, en cada uno de los 4 grupos de épocas se eligen aleatoriamente únicamente 315 de ellas, para tener clases balanceadas en las secciones 4.5 y 4.6.

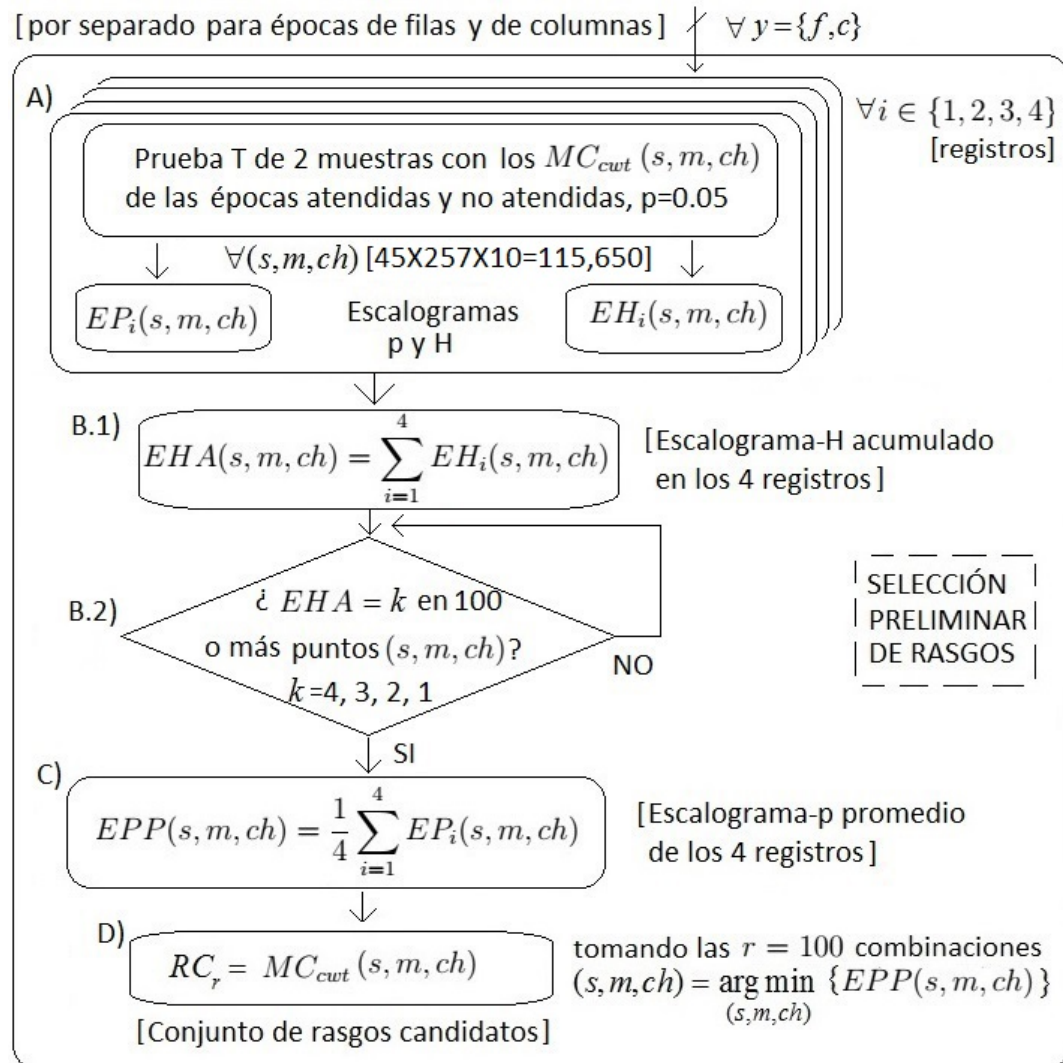


Figura 4.7: Selección preliminar de rasgos. (A.1) Con todas las épocas atendidas y no atendidas de cada registro (i), se calcula una prueba T de dos muestras para cada rasgo $MC_{cwt}(s, m, ch)$. (A.2) De las pruebas estadísticas se obtienen escalogramas de valores p (EP_i) y H (EH_i). (B.1) El Escalograma-H acumulado (EHA) es la suma del valor de EH_i en los 4 registros. (B.2) Se verifica que al menos 100 puntos de EHA tengan valor de $k = 4, 3, 2$ ó 1 (número de registros con diferencias significativas), y en aquellos que cumplen la condición, (C) se promedia el valor de EP en los 4 registros (EPP). (D) El Conjunto de Rasgos Candidatos, RC_R , contiene el valor de MC_{cwt} de los 100 puntos (s, m, ch) que minimizan a EPP .

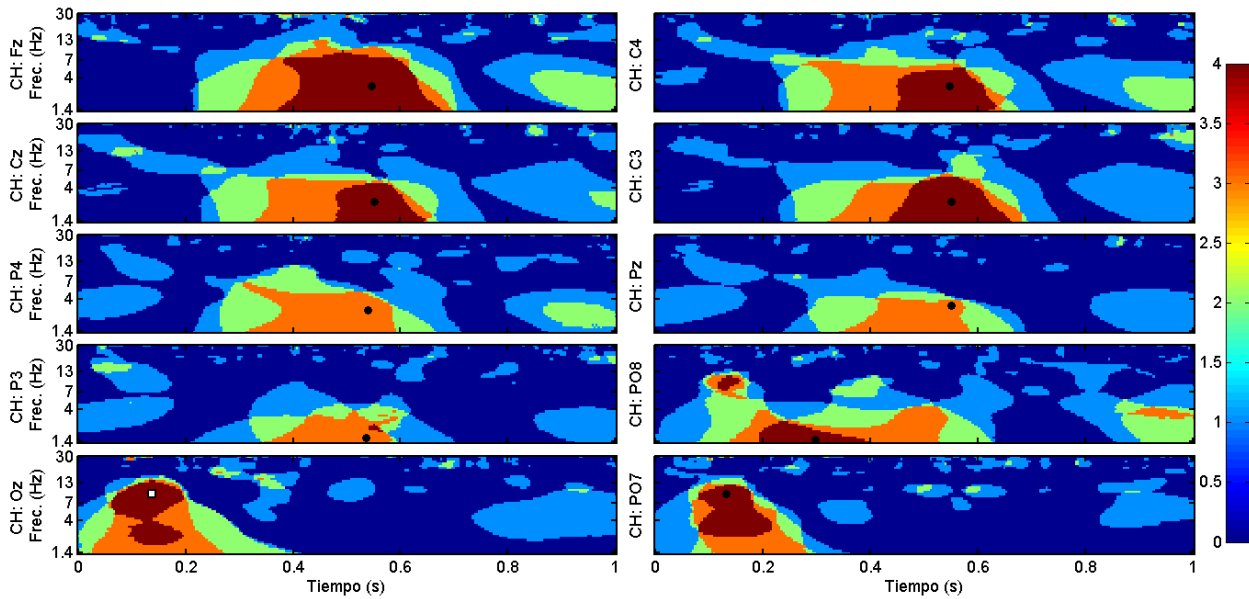


Figura 4.8: Escalograma-H acumulado en los 10 canales de registro del sujeto ASG (columnas).

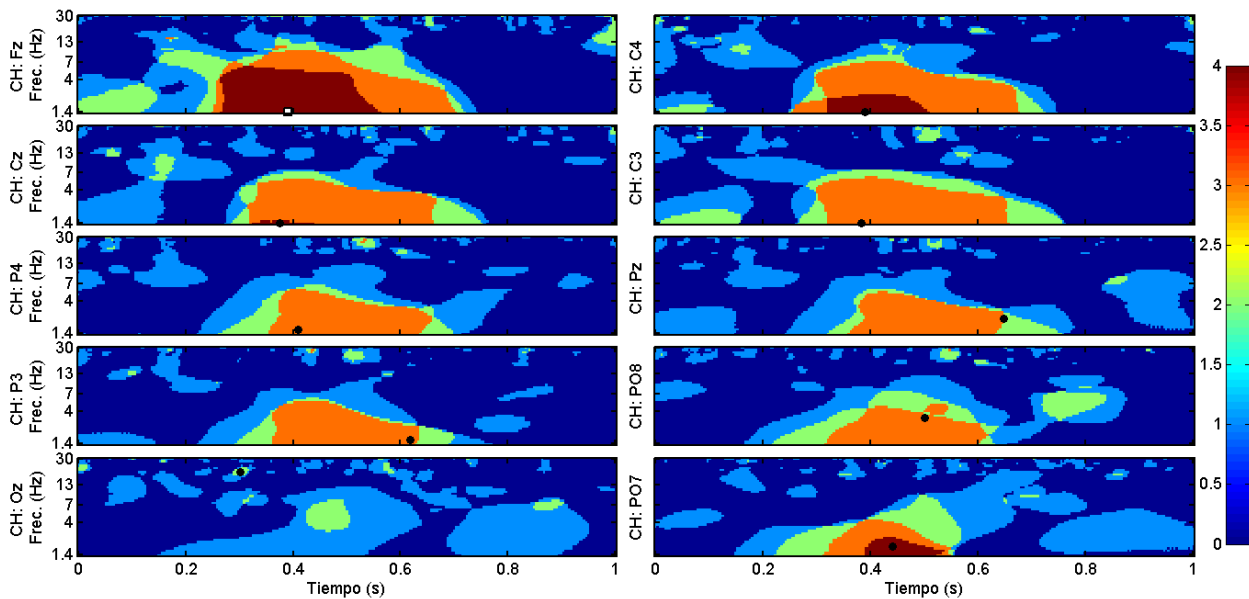


Figura 4.9: Escalograma-H acumulado en los 10 canales de registro del sujeto ASG (filas).

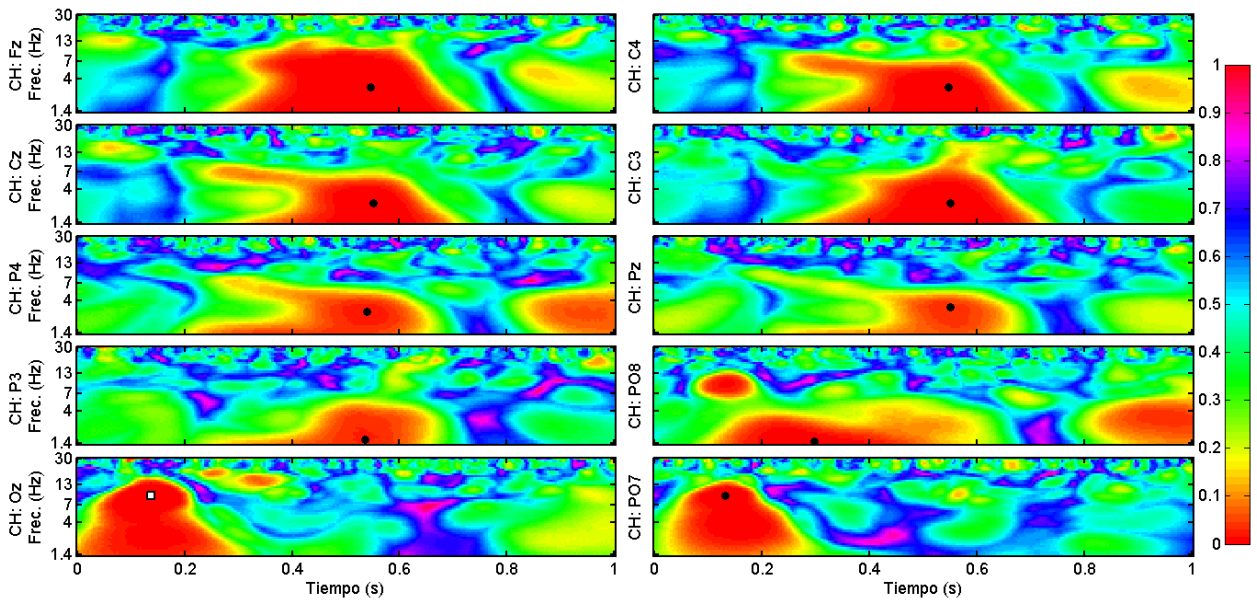


Figura 4.10: Escalograma-p promedio entre los 4 registros, en los 10 canales de registro del sujeto ASG (columnas).

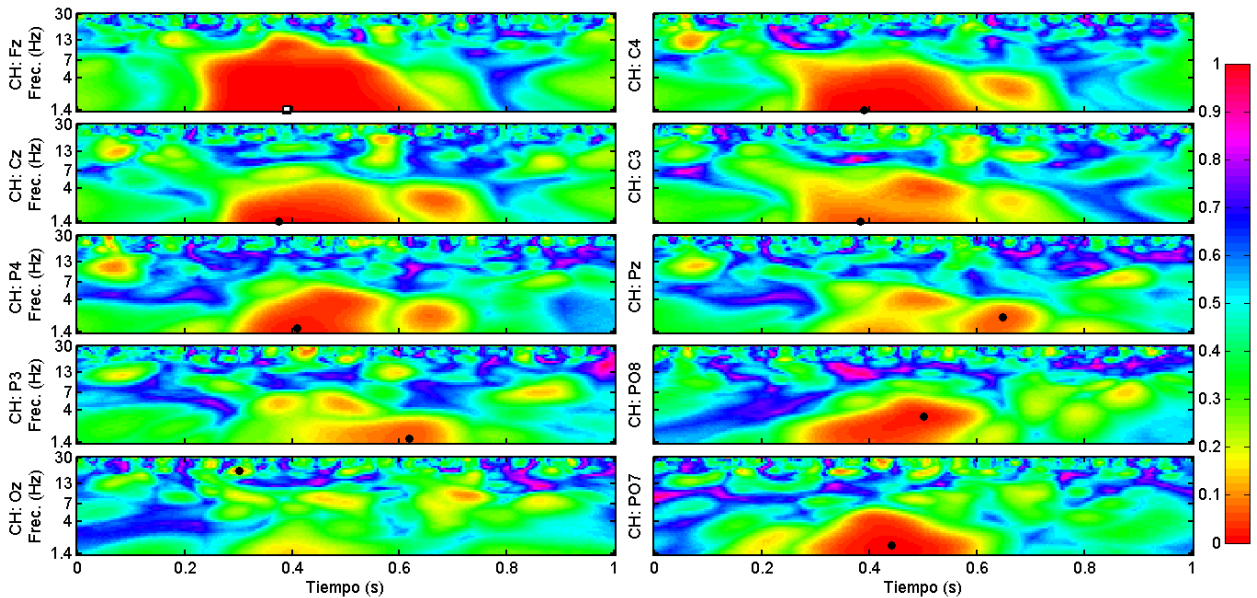


Figura 4.11: Escalograma-p acumulado en los 10 canales de registro del sujeto ASG (filas).

4.5. Selección de rasgos e hiperparámetros óptimos

4.5.1. Promediación de rasgos entre épocas

- Se busca conocer el efecto de la promediación de los valores de los rasgos candidatos RC_r entre N épocas de EEG en el desempeño de clasificación. Los promedios de los rasgos se calculan de la siguiente manera (Fig. 4.12-A.1):

$$RCp_r(N) = \frac{1}{N} \sum_{n=1}^N RC_r, N = \{1, 2, 3, 5, 8, 10\} \quad (4.8)$$

donde N es el número de épocas a promediar, RC_r es el conjunto de r rasgos candidatos ($r = 100$ a menos que se indique otra cosa, como en la sección 4.5.3), y $RCp_r(N)$ es el conjunto de los promedios de cada rasgo candidato $\in RC_r$, calculados en cada grupo de N épocas que se pueden formar, sin repetición, con las 315 disponibles. Así, se generan 6 conjuntos de promedios de rasgos $RCp_r(N)$ con el sig. número de elementos en cada uno:

$$315/N = \{315, 157, 105, 63, 39, 31\}, N = \{1, 2, 3, 5, 8, 10\} \quad (4.9)$$

, que sustituyen a los 315 valores de los rasgos candidatos originales, RC_r .

4.5.2. Conjuntos de entrenamiento y de prueba

En un problema de clasificación supervisada es preciso separar los datos disponibles en dos conjuntos: uno de mayor tamaño, llamado *conjunto de entrenamiento* que se usa para

seleccionar un subconjunto de rasgos y modelo de clasificación *óptimos*, y el otro conjunto, más pequeño, para evaluar el desempeño del clasificador y rasgos seleccionados, con datos no-observados en las etapas previas, llamado *conjunto de prueba*. A continuación se describe el procedimiento para generar estos conjuntos de datos.

Para cada valor de N , se genera un conjunto de entrenamiento ($CEN_r(N)$) y un conjunto de prueba ($CPR_r(N)$) a partir del conjunto correspondiente de promedios de rasgos ($RCp_r(N)$). Estos conjuntos se forman de la siguiente manera (Fig. 4.12-A.2):

- *Conjunto de entrenamiento*. $CEN_r(N)$ se conforma con el 70% de los elementos en $RCp_r(N)$, elegidos aleatoriamente (promedios sobre N épocas de los $r = 100$ rasgos candidatos). Así $CEN_r(N)$ contiene el siguiente número de elementos :

$$0.7(315/N) = \{220, 110, 73, 44, 27, 22\}, N = \{1, 2, 3, 5, 8, 10\} \quad (4.10)$$

, donde $CEN_r(N) \subset RCp_r(N)$, y $r = 100$ es el número total de rasgos candidatos (sección 4.4). El conjunto $CEN_r(N)$ se usa:

- En la etapa de *selección del subconjunto de rasgos e hiperparámetros óptimos*, se toman particiones de los elementos de $CEN_r(N)$ para entrenar y probar los clasificadores en las vías de las validaciones cruzadas (Fig. 4.12-B).
- Con todos sus elementos, para el entrenamiento del *clasificador óptimo* en la etapa de *Prueba con datos no-observados* (Fig. 4.15-A.1).

- *Conjunto de Prueba.* $CPR_r(N)$ consiste en el 30% de los elementos de $RCp_r(N)$ no incluidos en el conjunto $CEN_r(N)$ (Ec. 4.10). Así, se tienen 6 conjuntos de prueba $CPR_r(N)$ con el siguiente número de elementos:

$$0.3(315/N) = \{94, 47, 31, 19, 11, 9\}, N = \{1, 2, 3, 5, 8, 10\} \quad (4.11)$$

, donde $CRP_r(N) \subset RCp_r$, y $r = 100$ es el número total de rasgos candidatos (RC_r).

El conjunto $CPR_r(N)$ se utiliza para evaluar el desempeño del clasificador óptimo en la etapa de *Prueba con datos no-observados*.

4.5.3. Evaluación de subconjuntos de rasgos e hiperparámetros

Una vez conformados los conjuntos $CEN_r(N)$ y $CPR_r(N)$ para cada valor de N , se evalúan 4 subconjuntos de los $r = 100$ rasgos en RCp_r , para encontrar el que maximiza la correcta clasificación de los elementos en $CEN_r(N)$ (promedios en N épocas de los rasgos en RCp_r) como pertenecientes a la clases *atendida* o *no-atendida*.

Para la clasificación se utilizan Máquinas de Soporte Vectorial con kernel de función de base radial (SVM-RBF), entrenadas y probadas con funciones de la librería para SVMs *LIBSVM* [59] versión 3.20.

Para encontrar la combinación del subconjunto de r rasgos de RCp_r y los hiperparámetros *óptimos* de la SVM-RBF, para cada valor de N se lleva a cabo el siguiente procedimiento (Fig. 4.12-B):

- Para cada valor de $r \in \{1, 20, 50, 100\}$, se busca la combinación óptima de los hiper-

parámetros de costo (C) y ancho de banda del kernel (γ) del clasificador SVM-RBF, entrenado y evaluado con el subconjunto de r rasgos de los elementos en $CEN_r(N)$.

- Para cada valor de r y el $CEN_r(N)$ correspondiente, se hace lo siguiente:
 - Se barren los hiperparámetros, con valores de $C(a) = 10^a$, $a = \{-2, -1, 0, 1, 1.5\}$, que dan $C(a) = \{0.01, 0.1, 1, 10, 31.62\}$ y valores de $\gamma(b) = 10^b$, $b = \{-4.3, -4, -3.3, -3, -2\}$, que dan $\gamma(b) = \{0.05, 1, 5, 10, 100\} \times 10^{-4}$. Esto da como resultado 25 combinaciones de $(C(a), \gamma(b))$.
 - Para cada combinación de $(C(a), \gamma(b))$ se llevan a cabo tres corridas de validación cruzada de 4 vías con el conjunto $CEN_r(N)$. En cada corrida se procede como sigue:
 - Con la función de entrenamiento de LIBSVM, se genera un clasificador SVM con kernel RBF e hiperparámetros $(C(a), \gamma(b))$, entrenado con 3/4 partes de las épocas de $CEN_r(N)$.
 - Se evalúa el clasificador SVM($C(a), \gamma(b)$), usando la función de predicción (clasificación) de LIBSVM con la porción restante (1/4 parte) de las épocas en $CEN_r(N)$. El clasificador entrega su predicciones de etiquetas de clase (atendida/no-atendida) y las probabilidades a posteriori de pertenencia a cada clase (0.0 – 1.0).
 - A partir de las probabilidades a posteriori obtenidas, se calcula el índice de

desempeño AUROC, como se describe en la sección 3.7. El desempeño de clasificación en cada validación cruzada está dado por:

$$\text{AUROC}_{v,c_r}(C(a), \gamma(b)) = \frac{1}{4} \sum_{v=1}^4 \text{AUROC}_{r,v}(C(a), \gamma(b)) \quad (4.12)$$

, donde v es el número de vía de la validación cruzada, r es el tamaño del subconjunto de rasgos bajo evaluación, y $\text{AUROC}_{r,v}(C(a), \gamma(b))$ es el índice de desempeño del clasificador en cada vía v de la validación cruzada.

- Para cada combinación (r, C, γ) , el índice de desempeño de clasificación promedio de las 3 corridas de validación cruzada se calcula como:

$$\text{AUROC}_{\text{prom}_r}(C, \gamma) = \frac{1}{3} \sum_{c=1}^3 (\text{AUROC}_{v,c_r}(C, \gamma)) \quad (4.13)$$

, donde la c de la sumatoria es el índice de las corridas de validación cruzada.

- El subconjunto de rasgos e hiperparámetros óptimos son la combinación (r, C, γ) que maximiza el valor de $\text{AUROC}_{\text{prom}_r}(C, \gamma)$, como se muestra a continuación:

$$(r_{opt}, C_{opt}, \gamma_{opt}) = \arg \max_{(r,C,\gamma)} \{\text{AUROC}_{\text{prom}_r}(C, \gamma)\} \quad (4.14)$$

Estos rasgos e hiperparámetros óptimos se utilizarán en la siguiente etapa (sección 4.6), para entrenar un clasificador óptimo y evaluar su desempeño con datos no-observados hasta este punto.

En las Figs. 4.13 y 4.14 se muestra un ejemplo gráfico del resultado del procedimiento de *selección de rasgos e hiperparámetros óptimos* que se acaba de describir, para épocas de filas

y columnas del sujeto ASG. Sólo hay una combinación C_{opt} , γ_{opt} y r_{opt} , que maximiza el valor de $AUROC_{prom,r}(C, \gamma)$ para cada valor de N (columnas), indicada por la localización de los pequeños círculos de color negro. Este gráfico nos permite ver de manera rápida el comportamiento del parámetro $AUROC_{prom}$, conforme varía N y las otras variables, y para identificar algún comportamiento peculiar, como en la Fig. 4.14, donde $C_{opt} = 100$ para todo valor de N . En el Anexo B se muestran los gráficos equivalentes de los 10 sujetos usados en este trabajo.

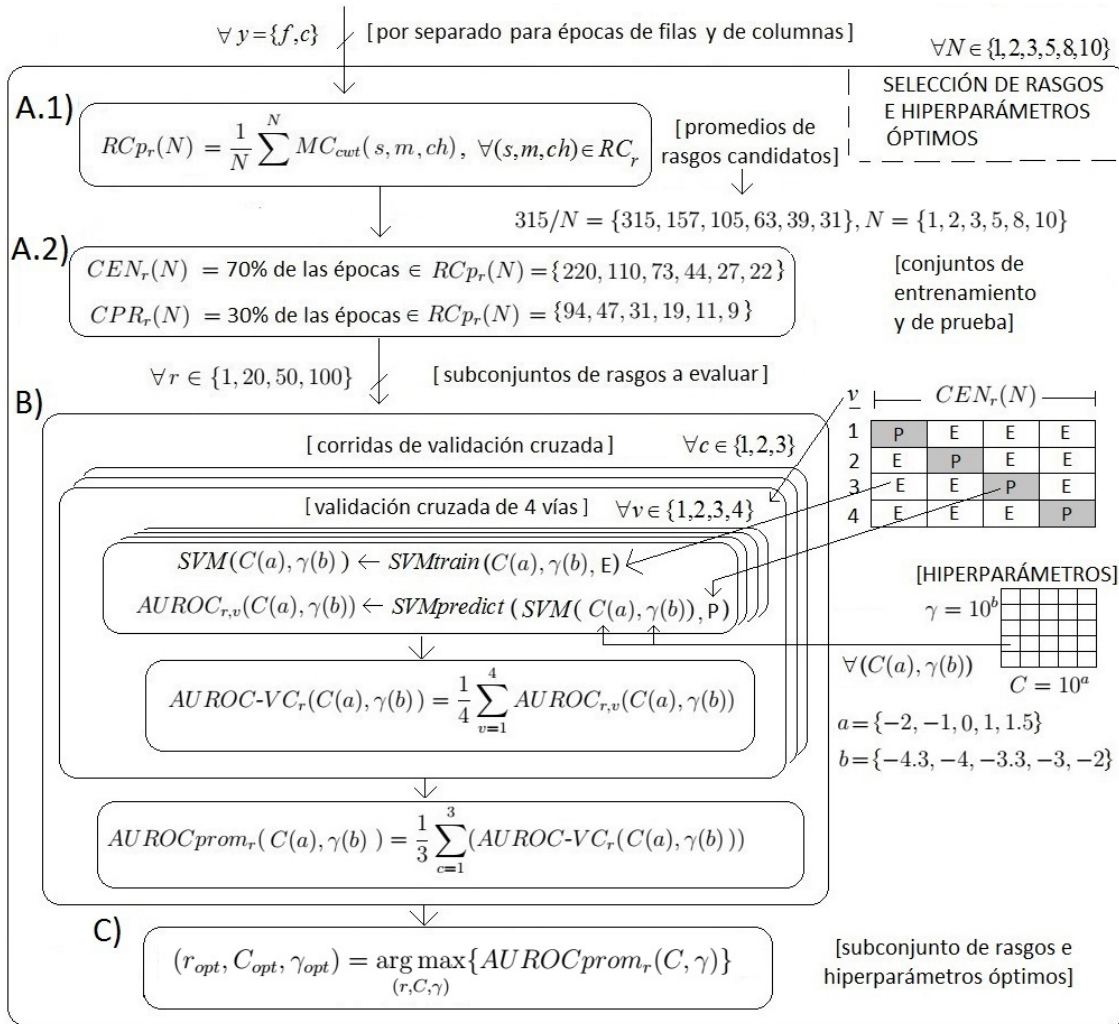


Figura 4.12: Selección de rasgos e hiperparámetros óptimos. Para cada N y con épocas de filas y de columnas por separado, (A.1) los rasgos en $RCp_r(N)$ se promedian en grupos de N épocas ($RCp_r(N)$). (A.2) Los conjuntos de entrenamiento y prueba ($CEN_r(N)$, $CPR_r(N)$) se integran con el 70% y el 30% de elementos en $RCp_r(N)$. (B) Para cada subconjunto de $r = 1, 10, 50, 100$ mejores rasgos en $RCp_r(N)$, se obtiene $AUROC_{prom_r}(C, \gamma)$, el parámetro de desempeño de clasificación promedio en 3 corridas de validación cruzada de 4 vías. El clasificador es una SVM-RBF con hiperparámetros (C, γ) , entrenado y evaluado con particiones de $CEN_r(N)$. (C) El subconjunto de rasgos y los hiperparámetros óptimos $(r_{opt}, C_{opt}, \gamma_{opt})$, son aquellos que maximizan el valor de $AUROC_{prom_r}(C, \gamma)$.

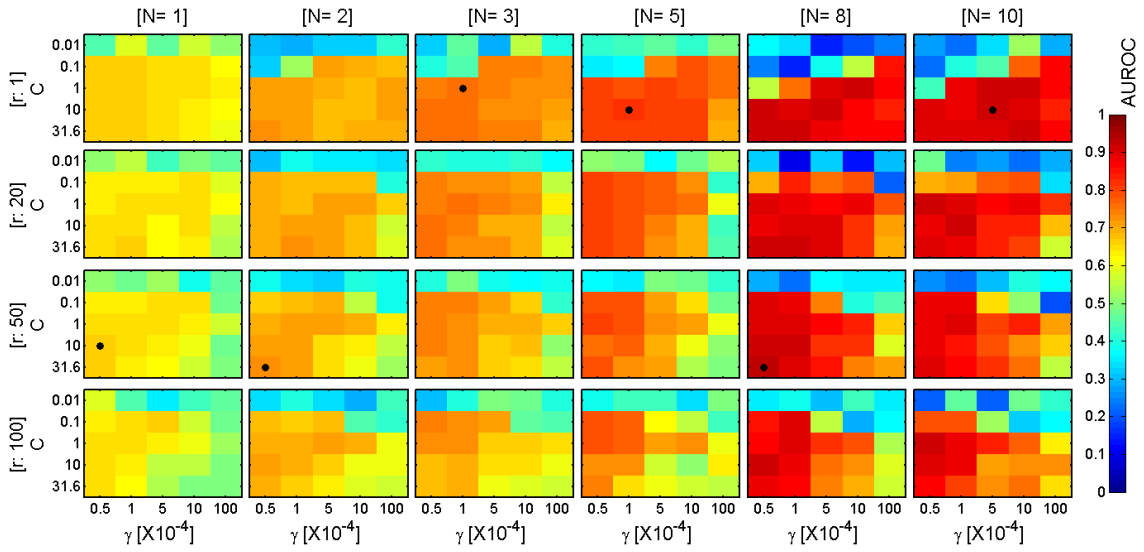


Figura 4.13: Resultado de Selección de rasgos e hiperparámetros óptimos. Sujeto ASG (filas).

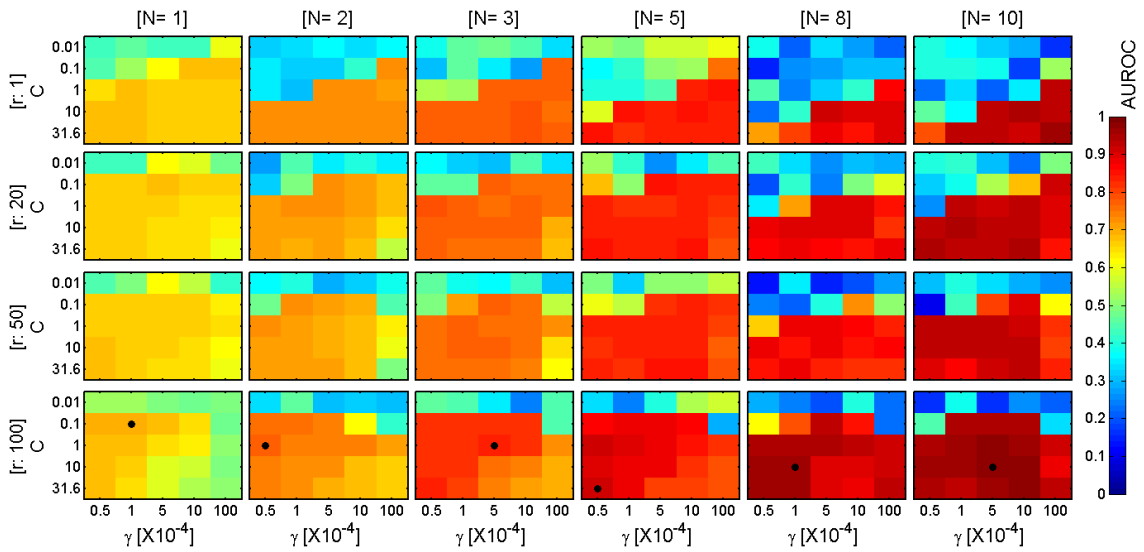


Figura 4.14: Resultado de Selección de rasgos e hiperparámetros óptimos. Sujeto ASG (columnas).

4.6. Prueba con datos no-observados

Una vez obtenidos los rasgos e hiperparámetros óptimos para cada N , ahora se busca evaluar la capacidad de estos para clasificar datos no-observados previamente. Los conjuntos de datos $CPR_r(N)$ fueron apartados (sección 4.5.2) justamente para ser utilizados para ese fin, de modo que no fueran conocidos a lo largo de todo el proceso de entrenamiento y selección de rasgos e hiperparámetros (sección 4.5). El objetivo es obtener un índice fidedigno de la capacidad de generalización de un clasificador entrenado con esos parámetros y rasgos seleccionados como *óptimos*.

Esta evaluación se lleva a cabo para cada valor de $N \in \{1, 2, 3, 5, 8, 10\}$, y con épocas de filas y columnas por separado, de la siguiente manera (Fig. 4.15):

- Se genera el clasificador óptimo $SVM_{opt}(N)$ mediante la función de entrenamiento de LIBSVM (*svmtrain*), con $(r_{opt}, C_{opt}, \gamma_{opt})$ y todos los elementos del conjunto de datos no-observado $CEN_{r_{opt}}$ como argumentos de entrada (Fig. 4.15-A.1):

$$SVM_{opt}(N) \leftarrow svmtrain(r_{opt}, C_{opt}, \gamma_{opt}, CEN_{r_{opt}}(N)) \quad (4.15)$$

- Se prueba el desempeño del clasificador óptimo, usando como entradas a la función de predicción de LIBSVM (*svmpredict*) los elementos de $CPR_{r_{opt}}(N)$ y el modelo de clasificación $SVM_{opt}(N)$. A partir de las probabilidades a posteriori entregadas por la función *svmpredict*, se obtiene el índice de desempeño AUROCnob (Fig. 4.15-A.2):

$$AUROCnob(N) \leftarrow svmpredict(SVM_{opt}, CPR_{r_{opt}}(N)) \quad (4.16)$$

A partir de los valores de $AUROC_{nob}(N)$ para cada valor de N , se genera una curva AUROC VS N por sujeto (Figs. 4.15-B y 5.1), que puede resultar útil para identificar un valor adecuado de N para alcanzar cierto nivel de desempeño (un umbral de AUROC de 0.9 por ejemplo), o una tasa de transferencia de información adecuada (por ejemplo, 4 símbolos por minuto), considerando el número N de épocas a promediar.

Así, al elegir un valor de N con buen compromiso desempeño/velocidad de deletreo (AUROC/símbolos deletreados por minuto), y el clasificador $SVM_{opt}(N)$ correspondiente, éstos pueden utilizarse en experimentos posteriores con el Deletreador de Donchin para ese sujeto en particular. Además, si las características espectrales importantes en esos nuevos registros no cambian demasiado respecto a las aquí seleccionadas (r_{opt}), el desempeño de clasificación se mantendrá en valores cercanos a los de las curvas de AUROC VS N mostradas en la Fig. 5.1.

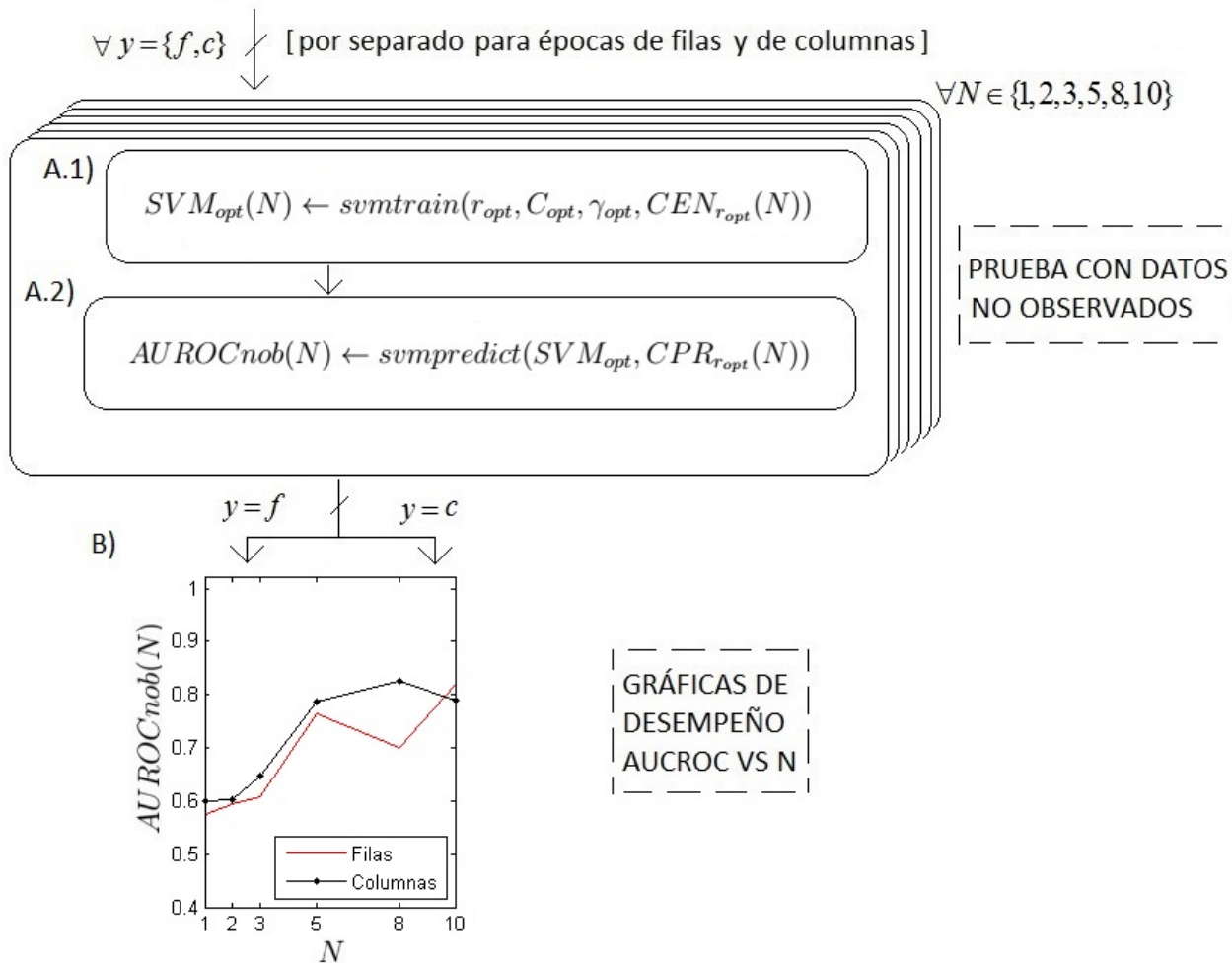


Figura 4.15: Prueba con datos no-observados. El procedimiento se lleva a cabo para cada valor de N , y para épocas de filas y de columnas por separado. (A.1) Se entrena el clasificador SVM_{opt} con las épocas de EEG del conjunto de entrenamiento $CEN_{r_{opt}}(N)$ y la combinación $(r_{opt}, C_{opt}, \gamma_{opt})$. (A.2) Con SVM_{opt} se clasifican los rasgos r_{opt} de las épocas no-observadas del conjunto de prueba $(CPR_{r_{opt}}(N))$, obteniendo el parámetro de desempeño $AUROC_{nob}(N)$. (B) El desempeño del clasificador óptimo (SVM_{opt}) para cada valor de N , se resume en una gráfica de $AUROC_{nob}(N)$ vs N , para épocas de filas (rojo) y de columnas (negro).

Capítulo 5

Resultados

En este trabajo se busca averiguar si es posible utilizar información en el dominio de la frecuencia (coeficientes de la CWT) para clasificar épocas de EEG, como *atendidas* o *no-atendidas* en el contexto de la BCI conocida como Deletreador de Donchin, como sugieren la hipótesis (sección 1.1) y el objetivo general del trabajo (sección 1.2). En este capítulo se presenta evidencia que apunta a esa hipótesis, en base a los resultados de las etapas *selección preliminar de rasgos* (sección 5.1) y *selección de rasgos e hiperparámetros óptimos* (sección 5.2), así como el desempeño de los clasificadores óptimos con datos no observados (sección 5.3).

5.1. Selección preliminar de rasgos

En el capítulo de Metodología se presentaron ejemplos para el sujeto ASG de los *escalogramas-H* y *escalogramas-p* (Figs. 4.8, 4.9, 4.10 y 4.11), que resultan del procedimiento de *selección preliminar de rasgos* realizado en los más de 100,000 puntos del mapa tiempo-frecuencia-canal, para épocas de filas y de columnas. En el Anexo A se recopilan los gráfi-

cos correspondientes a los 10 sujetos con los que se probó la metodología.

En la tablas 5.1 y 5.2 se resume el resultado de la *selección preliminar de rasgos* para cada sujeto, indicando además los canales, bandas de frecuencia y latencias en las cuales se concentran los 4 subconjuntos de rasgos candidatos ($r = 1, 20, 50, 100$), que se evalúan posteriormente en la *selección de rasgos e hiperparámetros óptimos*.

Esas tablas resultan útiles para observar similitudes y diferencias entre los rasgos importantes para los diferentes sujetos, y entre condiciones (filas y columnas).

Es pertinente aclarar aquí, que las bandas de frecuencia del EEG, como se usan en las tablas 5.2 y 5.1 son: δ (*delta*: 1.4 – 4 Hz), θ (*theta*: 4 – 7 Hz), α (*alfa*: 7 – 13 Hz) y β (*beta*: 13–30 Hz). También es preciso indicar que el término *latencia*, se refiere al tiempo posterior al instante de estimulación (en s o ms), en el que se presentan los rasgos candidatos.

Tabla 5.1: Resultados de la selección preliminar de rasgos para épocas de filas: se indican los canales (ch), bandas de frecuencia (bf) y latencias (t(s)) donde se concentran los subconjuntos de r rasgos relevantes seleccionados de RC_r (Sección 4.4, Fig. 4.7).

Sujeto	Filas			
	$r = 1$	20	50	100
DLP: ch	Fz	Fz	Fz/PO8	Fz/PO8
bf	θ	θ	θ/β	$\theta-\alpha/\beta$
t(s)	0.50	0.48-0.52	0.47-0.53/0.78-0.79	0.46-0.54/0.33,0.79
DMA: ch	PO8	PO8	PO8	PO8/C4/Oz/P4
bf	δ	δ	δ	$\delta/\delta/\delta/\delta$
t(s)	0.37	0.36-0.39	0.35-0.4	0.34-0.4/0.4/0.44/0.4
ASG: ch	Fz	Fz	Fz	Fz
bf	δ	δ	δ	δ
t(s)	0.39	0.36-0.43	0.35-0.45	0.33-0.46
CLL: ch	Oz	Oz	Oz	Oz/PO8
bf	α	α	α	α/α
t(s)	0.16	0.14-0.18	0.14-0.19	0.12-0.20/0.12-0.145
LAG: ch	Oz	Oz	Oz	Oz
bf	δ	δ	$\delta-\theta$	$\delta-\theta$
t(s)	0.18	0.17-0.20	0.17-0.21	0.15-0.21
LGP: ch	PO7	PO7/Cz	PO7/Cz/C3/P3	PO7/Cz/C3/P3
bf	α	α/α	$\alpha/\alpha/\alpha/\beta$	$\alpha/\alpha-\beta/\alpha-\beta/\beta$
t(s)	0.16	0.16-0.18/0.1	0.15-0.18/0.13/0.1/0.8	0.14-0.19/0.1-0.13/0.1/0.8
LAC: ch	C4	C4	C4/Cz	C4/Cz
bf	δ	δ	δ/δ	δ/δ
t(s)	0.42	0.4-0.43	0.38-0.44/0.4-0.41	0.38-0.44/0.39-0.43
IZH: ch	C4	C4	C4	C4/C3
bf	δ	δ	$\delta-\theta$	$\delta-\theta/\delta-\theta$
t(s)	0.42	0.41-0.43	0.4-0.44	0.39-0.45/0.43-0.45
GCE: ch	C3	C3/Cz	C3/Cz	C3/Cz
bf	δ	δ/δ	δ/δ	δ/δ
t(s)	0.42	0.41-0.44/0.41-0.44	0.4-0.46/0.41-0.46	0.39-0.48/0.41-0.47
PGA: ch	Fz	Fz	Fz	Fz
bf	δ	δ	δ	δ
t(s)	0.44	0.43-0.46	0.41-0.47	0.4-0.48

Tabla 5.2: Resultados de la selección preliminar de rasgos para épocas de columnas: se indican los canales (ch), bandas de frecuencia (bf) y latencias (t(s)) donde se concentran los subconjuntos de r rasgos relevantes seleccionados de RC_r (Sección 4.4, Fig. 4.7).

Sujeto	Columnas			
	$r = 1$	20	50	100
DLP: ch	Fz	Fz	Fz	Fz
bf	δ	δ	δ	δ
t(s)	0.44	0.43-0.45	0.43-0.46	0.42-0.47
DMA: ch	PO8	PO8	PO8/Fz/PO7	PO8/Fz/PO7
bf	β	β	$\beta/\beta/\delta/$	$\alpha-\beta/\beta/\delta/$
t(s)	0.35	0.33-0.36	0.33-0.37/0.92/0.33	0.33-0.37/0.92/0.3-0.36
ASG: ch	Oz	Oz	Oz	Oz/Fz
bf	α	α	α	$\theta-\alpha/\delta$
t(s)	0.13	0.12-0.14	0.11-0.15	0.1-0.16/0.53-0.55
CLL: ch	Oz	Oz	Oz/PO8	Oz/PO8/P3
bf	θ	θ	θ/β	$\theta/\beta/\beta$
t(s)	0.15	0.13-0.17	0.12-0.18/0.3	0.11-0.19/0.3-0.31/0.24-0.25
LAG: ch	PO8	PO8	PO8	PO8
bf	α	$\alpha-\beta$	$\alpha-\beta$	$\alpha-\beta$
t(s)	0.36	0.31-0.37	0.3-0.37	0.30-0.38
LGP: ch	PO7	PO7	PO7	PO7
bf	β	$\alpha-\beta$	$\alpha-\beta$	$\alpha-\beta$
t(s)	0.13	0.12-0.15	0.11-0.15	0.1-0.16
LAC: ch	PO7	PO7	PO7	PO7
bf	θ	θ	θ	$\delta-\theta$
t(s)	0.17	0.15-0.18	0.14-0.18	0.13-0.19
IZH: ch	Fz	Fz	Fz	Fz
bf	θ	$\delta-\theta$	$\delta-\alpha$	$\delta-\alpha$
t(s)	0.39	0.34-0.39	0.34-0.4	0.33-0.4
GCE: ch	PO8	PO8	PO8	PO8
bf	θ	θ	$\delta-\theta$	$\delta-\theta$
t(s)	0.14	0.1-0.17	0.1-0.19	0.1-0.2
PGA: ch	PO8	PO8	PO8/C4	PO8/C4
bf	α	α	α/α	$\theta-\alpha/\alpha$
t(s)	0.15	0.14-0.17	0.13-0.17/0.34-0.35	0.12-0.18/0.33-0.36

5.2. Selección de rasgos e hiperparámetros óptimos

En la tabla 5.3 se concentran los resultados del procedimiento de selección de rasgos e hiperparámetros óptimos para cada sujeto (r_{opt} , C_{opt} , y γ_{opt}). En las 3 filas inferiores de la tabla, se presenta la moda de éstos valores óptimos, para cada valor de N , calculada entre todos los sujetos. En negritas se resaltan los casos en los que el valor de un parámetro fue el mismo en al menos 5 de los 6 valores de N . En el capítulo 6 (Discusión), se hacen algunas observaciones respecto a los resultados presentados en las tablas de este capítulo.

5.3. Prueba con muestras no observadas

Como se explicó en la sección 4.6, para cada sujeto se obtienen 12 clasificadores $SVM_{opt}(N)$, uno para cada combinación de N y tipo de épocas (filas ó columnas) (Fig. 4.15). Cada clasificador se entrena con la combinación de hiperparámetros (C_{opt} , γ_{opt}) y el subconjunto de rasgos ($RCp_{r_{opt}}$) que maximizan el valor del parámetro $AUROC_{prom_r}(C, \gamma)$ en 3 corridas de validación cruzada de 4 vías (Fig. 4.12).

En la Fig. 5.1 se muestran las gráficas de desempeño de los clasificadores $SVM_{opt}(N)$ (Fig. 4.15), y de los clasificadores $SVM(C, \gamma)$ con rasgos e hiperparámetros óptimos (Fig. 4.12). Las líneas rojas corresponden a épocas de filas y las negras de columnas. Las líneas sólidas representan el valor de $AUROC_{prom_r}(C_{opt}, \gamma_{opt}) \pm$ la desviación estándar en las 3 corridas de validación cruzada (Fig. 4.12). Las líneas punteadas corresponden al valor de $AUROC_{nob}(N)$ obtenido al evaluar los $SVM_{opt}(N)$ con el conjunto $CPR_{r_{opt}}(N)$ (Fig. 4.15).

Tabla 5.3: Resultados de selección de rasgos e hiperparámetros óptimos: se muestra el valor de r_{opt} , C_{opt} y γ_{opt} para cada sujeto y valor de N , así como la moda de cada parámetro entre los 10 sujetos. Los casos con el mismo valor del hiperparámetro o valor de r para 5 ó 6 valores de N se muestran en negritas, pues sugieren homogeneidad y consistencia de los datos.

Sujeto	N=	Filas						Columnas					
		1	2	3	5	8	10	1	2	3	5	8	10
DLP	r_{opt}	1	50	100	100	50	100	1	1	1	20	1	1
	C_{opt}	10	1	10	31	10	31	0.1	10	10	1	31	0.1
	$\gamma_{opt}[\times 10^{-4}]$	1	0.5	0.5	0.5	100	0.5	5	0.5	100	1	0.5	100
DMA	r_{opt}	50	100	100	100	50	100	100	100	100	100	100	50
	C_{opt}	31	1	10	10	10	31	1	31	1	1	10	31
	$\gamma_{opt}[\times 10^{-4}]$	5	0.5	1	1	5	0.5	1	100	1	100	1	5
ASG	r_{opt}	50	50	1	1	50	1	100	100	100	100	100	100
	C_{opt}	10	31	1	10	31	10	0.1	1	1	31	10	10
	$\gamma_{opt}[\times 10^{-4}]$	0.5	0.5	1	1	0.5	5	1	0.5	5	0.5	1	5
CLL	r_{opt}	50	1	50	20	50	50	1	50	20	1	1	20
	C_{opt}	31	10	1	31	1	31	1	10	10	31	31	1
	$\gamma_{opt}[\times 10^{-4}]$	100	10	0.5	0.5	5	100	5	1	0.5	1	5	10
LAG	r_{opt}	1	1	1	1	1	1	20	20	50	50	100	50
	C_{opt}	31	10	10	1	1	10	10	10	10	10	31	10
	$\gamma_{opt}[\times 10^{-4}]$	0.5	1	10	100	100	5	1	5	1	1	1	0.5
LGP	r_{opt}	20	20	50	50	20	100	50	50	1	100	20	1
	C_{opt}	31	10	10	1	31	10	10	1	10	0.1	10	31
	$\gamma_{opt}[\times 10^{-4}]$	100	5	10	5	5	0.5	1	1	10	5	10	5
LAC	r_{opt}	1	1	1	20	20	20	100	100	100	100	20	1
	C_{opt}	31	1	31	10	10	10	31	31	31	10	1	31
	$\gamma_{opt}[\times 10^{-4}]$	0.5	10	1	5	1	1	0.5	1	1	0.5	1	5
IZH	r_{opt}	50	1	1	20	1	20	100	100	20	20	20	20
	C_{opt}	0.1	31	10	1	31	10	31	1	10	31	31	1
	$\gamma_{opt}[\times 10^{-4}]$	100	0.5	5	1	5	5	100	10	100	0.5	1	1
GCE	r_{opt}	1	1	1	20	50	1	50	100	100	100	100	100
	C_{opt}	0.1	10	10	10	1	31	1	10	10	0.1	31	10
	$\gamma_{opt}[\times 10^{-4}]$	100	1	5	1	100	5	0.5	0.5	1	5	10	0.5
PGA	r_{opt}	50	1	1	1	20	1	100	100	50	100	100	100
	C_{opt}	10	10	31	10	10	10	1	1	1	1	1	10
	$\gamma_{opt}[\times 10^{-4}]$	100	1	1	0.5	1	10	0.5	0.5	5	1	5	1
Moda	r_{opt}	50	1	1	20	50	1	100	100	100	100	100	100
Moda	C_{opt}	31	10	10	10	10	10	1	1	10	1	31	10
Moda	$\gamma_{opt}[\times 10^{-4}]$	100	0.5	1	1	5	5	1	0.5	1	1	1	5

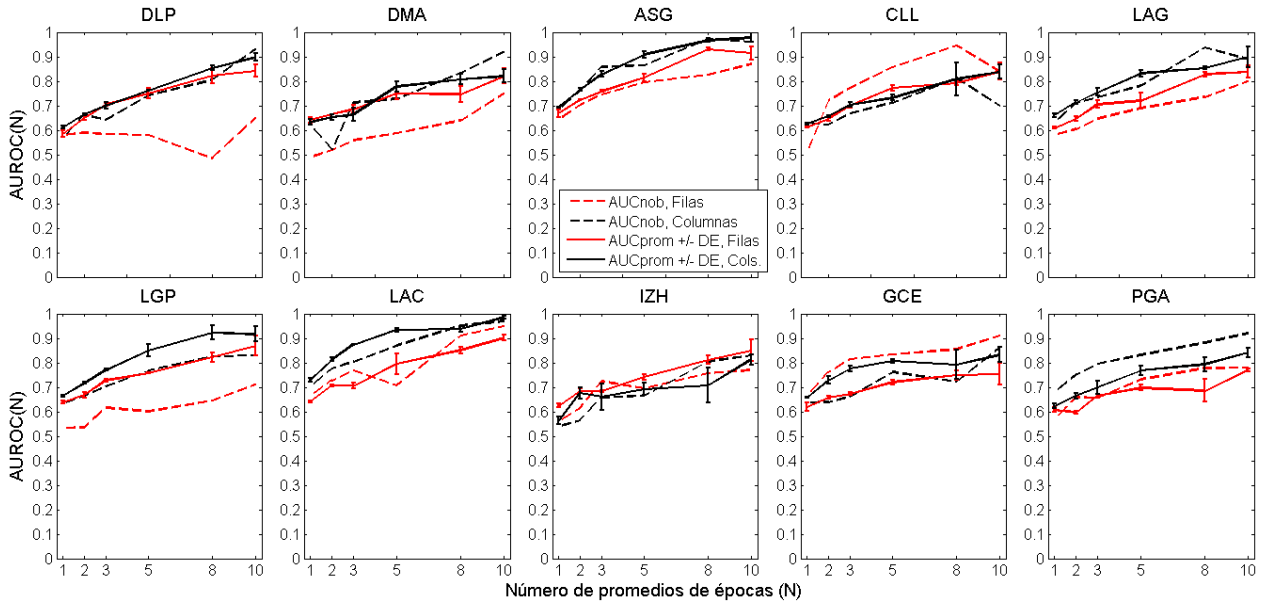


Figura 5.1: Gráficas de desempeño por sujeto. Las líneas rojas son para épocas de filas y las negras de columnas. Las líneas punteadas corresponden al $AUROCprom_r(C, \gamma) \pm$ desviación estándar (DE), obtenidas con la combinación de rasgos e hiperparámetros óptimos ($r_{opt}, C_{opt}, \gamma_{opt}$). Las líneas sólidas corresponden al $AUROCnob(N)$ obtenido al probar los clasificadores $SVM_{opt}(N)$ con datos no observados (NO).

En la Tabla 5.4 se muestran los valores numéricos de $AUROCnob(N)$ de la Fig. 5.1. Para cada N se calcula el promedio de $AUROCnob(N)$ y la desviación estándar entre los 10 sujetos. Se resaltan en negritas los valores máximos de $AUROCnob(N)$ para cada sujeto, y también los identificadores de los dos sujetos con mejor desempeño general para épocas de columnas y de filas: ASG (0.97 y 0.87) y LAC (0.97 y 0.95).

Tabla 5.4: Resultados de la prueba con muestras no observadas. Para cada sujeto y valor de N se muestra el valor $AUROC_{nob}$ obtenido con el clasificador $SVM_{opt}(N)$ (Sección 4.6, Fig. 4.15). Para fines de comparación, las 4 últimas filas muestran el promedio y desviación estándar del valor de $AUROC_{nob}$, y del valor de $AUROC_{prom}$, donde P-SRHO: Promedio del máximo $AUROC_{prom}$, y DE-SRHO: Desviación estándar del máximo $AUROC_{prom}$, en la etapa de selección de rasgos e hiperparámetros óptimos (SRHO). En negritas se muestran los máximos valores de $AUROC$ para cada sujeto y los promedios, además de los identificadores de los dos sujetos con mejor desempeño general en filas y columnas.

Sujeto	Columnas						Filas					
	N=1	2	3	5	8	10	N=1	2	3	5	8	10
DLP	0.57	0.66	0.64	0.74	0.81	0.93	0.58	0.59	0.59	0.58	0.48	0.65
DMA	0.62	0.52	0.71	0.73	0.83	0.92	0.49	0.52	0.56	0.59	0.64	0.75
ASG	0.68	0.76	0.86	0.86	0.97	0.96	0.64	0.70	0.75	0.80	0.83	0.87
CLL	0.63	0.62	0.67	0.71	0.81	0.70	0.51	0.72	0.77	0.86	0.94	0.84
LAG	0.63	0.71	0.73	0.78	0.94	0.89	0.58	0.61	0.65	0.69	0.74	0.80
LGP	0.63	0.66	0.70	0.77	0.82	0.83	0.53	0.53	0.62	0.60	0.65	0.71
LAC	0.70	0.77	0.80	0.87	0.95	0.97	0.66	0.73	0.77	0.71	0.91	0.95
IZH	0.54	0.56	0.66	0.66	0.81	0.83	0.56	0.62	0.72	0.70	0.76	0.77
GCE	0.64	0.64	0.66	0.76	0.72	0.86	0.67	0.77	0.81	0.84	0.85	0.91
PGA	0.68	0.75	0.80	0.83	0.88	0.92	0.57	0.66	0.66	0.73	0.77	0.78
Promedio	0.63	0.67	0.72	0.77	0.85	0.88	0.58	0.64	0.69	0.71	0.76	0.80
Desv. Est.	0.05	0.08	0.07	0.06	0.07	0.08	0.06	0.08	0.08	0.09	0.13	0.08
P-SRHO	0.65	0.70	0.74	0.81	0.84	0.88	0.62	0.66	0.70	0.75	0.80	0.84
DE-SRHO	0.04	0.05	0.07	0.07	0.08	0.06	0.02	0.03	0.03	0.03	0.06	0.05

Capítulo 6

Discusión

En este capítulo se presentan observaciones sobre los resultados de las diferentes etapas de la metodología, así como algunas ideas que parecen explicar parte del desempeño de clasificación en los diferentes sujetos y condiciones (épocas de filas o de columnas). En especial, se proponen relaciones entre la composición de los subconjuntos de rasgos óptimos y el desempeño de clasificación, tanto en el entrenamiento (selección de rasgos e hiperparámetros óptimos), como en la prueba con datos no-observados. El término composición se refiere a los rangos de latencia y de frecuencia, y los canales de registro, en los cuales se concentran los subconjuntos de rasgos candidatos para los diferentes sujetos.

6.1. Selección preliminar de rasgos

De la tabla 5.2, se puede observar que para la mayoría de los sujetos, los rasgos candidatos para épocas de columnas se concentran en canales occipitales/parieto-occipitales (Oz, PO8 y PO7), latencias de 100 – 200 ms, y bandas de frecuencia θ y α y en menor medida β (LGP). Las excepciones son los sujetos DLP e IZH: canal Fz, bandas δ y θ , latencias

de 330 – 470 ms, y el sujeto DMA: banda de frecuencia β , y latencias a 330 – 370 ms.

Para épocas de filas, (Tabla 5.1) vemos que los rasgos candidatos se reparten en varios canales (Fz, C3, C4, Oz, PO8, PO7), en frecuencias de la banda δ y también α (CLL, LGP), y a latencias de 300 – 500 ms. La excepciones son el sujeto DLP: bandas θ y β , y latencias alrededor de 500 y 780 ms, y tres sujetos que tienen sus rasgos candidatos a latencias de 100 – 200 ms (CLL, LAG, LGP), aunque coinciden con otros sujetos en las bandas δ y α (comunes en épocas de filas).

De las mismas tablas, vemos que para épocas de columnas, los rasgos son más homogéneos entre los diferentes subconjuntos de rasgos candidatos, que para épocas de filas. Esto quiere decir que para columnas, desde $r = 1$ hasta $r = 100$, los rasgos están más concentrados en uno o unos cuantos canales, y en un rango estrecho de frecuencia y latencia, mientras que para filas, los rasgos están repartidos, sobre todo, entre más canales y bandas de frecuencias distintas (son menos homogéneos en sus características).

6.2. Rasgos, hiperparámetros y desempeño de clasificación

Es de notarse que los sujetos con desempeño notablemente menor para épocas de filas en muestras no-observadas en comparación con el entrenamiento (DLP, DMA, LGP) (Fig. 5.1), tienen valores de $r_{opt} = 50$ y 100 para la mayoría de valores de N , sobre todo en $N = 10$ (5.3), donde se tiene el mayor valor de AUROC_{nob} para los tres sujetos (Tabla 5.4). Esto también puede estar relacionado con el hecho de que, para estos sujetos se agregan

rasgos en los subconjuntos de $r = 50$ y 100 , que están en canales, latencias y bandas de frecuencia que no están presentes en los subconjuntos de $r = 1$ y 20 .

Ante la presencia de esos nuevos rasgos diferentes para $r = \{50, 100\}$ (Tabla 5.4), el procedimiento de selección de rasgos e hiperparámetros óptimos determina que es importante tomarlos en cuenta ($r_{opt} = \{50, 100\}$, Tabla 5.3), pero el clasificador óptimo $SV M_{opt}$ entrenado con éstos rasgos no logra clasificar las muestras no observadas con el mismo desempeño que durante el entrenamiento (Fig. 5.1, Tabla 5.4). Esto sugiere un posible sobre-entrenamiento del clasificador, probablemente a causa de la baja homogeneidad en los rasgos del subconjunto óptimo, que se mencionó en la sección anterior (6.1).

El hecho de que, en los sujetos con buen desempeño de clasificación (en entrenamiento y prueba), para épocas de filas predomine $r_{opt} = 1$ (Tabla 5.3), también podría estar relacionado con la alta no-homogeneidad de los rasgos en los subconjuntos de $r = 50$ y $r = 100$ (Fig. 5.1).

La tendencia (la moda) en épocas de columnas del valor $r_{opt} = 100 \forall N$ (Fig. 5.3), podría también estar relacionada con el mejor desempeño de clasificación general en comparación con las épocas de filas, para las cuales la moda de $r_{opt} \neq 100 \forall N$. De hecho para épocas de filas $r_{opt} = 1$ para $N = \{2, 3, 10\}$ (Tabla 5.3). Es decir, el comportamiento de r_{opt} es prácticamente opuesto entre épocas de filas y de columnas.

Para algunos sujetos (PGA, GCE y ASG en épocas de filas, y LAC, CLL en épocas de columnas), $r_{opt} = 1$ (Tabla 5.3) para el valor de N con el mayor desempeño de clasificación

en muestras no-observadas (Tabla 5.4), aún cuando para otros valores de N , $r_{opt} > 1$ (Tabla 5.3). Este comportamiento es notable, pues significa que un solo rasgo individual (un punto del mapa tiempo-frecuencia canal), clasifica mejor las épocas de EEG que $r = 20, 50$ o 100 rasgos juntos.

Ese comportamiento difícilmente puede ser atribuible por completo al azar, toda vez que el desempeño alcanzado en la fase de entrenamiento, donde se selecciona r_{opt} , se verifica y se mantiene a un nivel similar en la prueba con datos no observados en todos los casos ($AUROC_{nob} \approx AUROC_{prom}$).

El caso del sujeto LAG es notable, pues para épocas de filas, del procedimiento de selección de rasgos óptimos se obtiene que $r_{opt} = 1 \forall N$ (Tabla 5.3). Es decir, para cualquier valor de N se tuvo un mayor desempeño de clasificación con el mejor rasgo individual que con los mejores 20, 50 o 100 rasgos juntos. El mejor rasgo individual para épocas de filas, del subconjunto $r_{opt} = 1$, está en el canal Oz, banda δ y latencia de 0.18 s (Tabla 5.1).

Para los otros 3 subconjuntos de rasgos del sujeto LAG ($r=\{20,50,100\}$), los rasgos son bastante similares al de $r_{opt} = 1$ tanto para épocas de filas como de columnas, manteniéndose en los mismos canales y rangos de latencia, y solo para épocas de columnas presenta actividad en la banda β , adyacente a la banda α donde está el rasgo único para $r_{opt} = 1$ (Tablas 5.2 y 5.1).

Respecto a los valores del hiperparámetro C de la SVM-RBF, de la Tabla 5.3 se puede apreciar que para épocas de filas la moda entre sujetos es $C_{opt} = 10 \forall N \neq 1$, y para épocas

de columnas la moda de $C_{opt} = 1$ para $N = \{1, 2, 5\}$, aunque para $N = \{5, 10\}$ la moda es $C_{opt} = 10$ (como en las filas).

En cuanto a los valores de γ , el otro hiperparámetro de la SVM-RBF, encontramos que para épocas de columnas, la moda de $\gamma_{opt} = 1 \times 10^{-4}$ para $N = \{1, 3, 5, 8\}$, mientras que para épocas de filas, la moda de $\gamma_{opt} = 1 \times 10^{-4}$ sólo para $N = \{3, 5\}$. Por otra parte, la moda de $\gamma_{opt} = 5 \times 10^{-4}$ para $N = 10$, en épocas de filas y de columnas (Tabla 5.3).

Finalmente podemos observar, que para $N = 10$, donde se tienen los mayores desempeños de clasificación (Tabla 5.4), la moda de $C_{opt} = 10$ y la moda de $\gamma_{opt} = 5 \times 10^{-4}$, tanto para épocas de filas como de columnas (Tabla 5.3). En cambio, en el caso del subconjunto óptimo de rasgos, la moda de $r_{opt} = 1$ para épocas de filas (para $N = \{2, 3, 10\}$), pero para épocas de columnas es $r_{opt} = 100$ ($\forall N$).

Lo anterior permite sugerir que el mejor desempeño de clasificación general para épocas de columnas que de filas, puede depender de la composición de los subconjuntos de rasgos y en particular con el subconjunto elegido como óptimo, más que de los valores de los hiperparámetros óptimos. Esto parece estar en línea o apoyar la idea, ya mencionada en ésta y en la sección anterior, de que la homogeneidad de los subconjuntos de rasgos parecen ser un factor que influye de manera importante en el desempeño de clasificación.

6.3. Otras observaciones

De manera general el desempeño de clasificación fue mayor para épocas de filas que de columnas, como se observa en la Fig. 5.1 y la Tabla 5.4. En la primera vemos que para 6 sujetos (ASG, LAG, LGP, LAC, GCE, PGA), las gráficas correspondientes a la etapa de entrenamiento (líneas sólidas) están totalmente separadas, y no se traslapan sus barras de desviación estándar, lo cual denota una clara diferencia general en sus valores de AUROCprom. De la Figura 5.1 se observa, que los casos de los tres sujetos (DLP, DMA, LGP) que tuvieron un desempeño de clasificación menor con muestras no-observadas respecto al entrenamiento (baja capacidad de generalización), se dieron en épocas de filas.

También de la Tabla 5.4, se observa que los dos sujetos con mejor desempeño de clasificación general en muestras no observadas son LAC (0.97 columnas, 0.95 filas, en $N = 10$ en ambos casos) y ASG (0.97 en columnas, $N = 8$, y 0.87 en filas, $N = 10$), cuyos identificadores se resaltan en negritas.

En general, se observan las siguientes diferencias en los rasgos que conforman los subconjuntos de rasgos candidatos, entre épocas de filas y columnas (Tablas 5.1, 5.2 y 5.3):

- La prevalencia de 100–200 ms en las latencias para épocas de columnas y de 300–500 ms para épocas de filas, da cuenta de una diferencia esencial en la información que el presente método determina como relevante para diferenciar entre épocas atendidas y no atendidas.

- Así mismo, la prevalencia de rasgos en frecuencias de la banda δ para épocas de filas, y en las bandas α y θ para épocas de columnas, sugiere que los rasgos elegidos representan mecanismos distintos en épocas de filas y de columnas.
- A los dos puntos anteriores se puede agregar, respecto a los canales, que en épocas de columnas son PO8, PO7, Oz y Fz los de mayor presencia en los rasgos candidatos, mientras que para épocas de filas son los canales Fz, Oz, C4, C3 y Cz.

Estas consistentes diferencias en los rasgos espectrales entre filas y columnas, apuntan a la idea de que los mecanismos cerebrales más informativos para identificar la presencia del potencial P300 (o bien para diferenciar entre épocas atendidas y no atendidas), son distintos entre épocas de filas y de columnas, a nombrar:

- Actividad en electrodos parieto-occipitales, en las bandas θ y α , a latencia de 100 – 200 ms para épocas de columnas. Este mecanismo probablemente esté relacionado con el procesamiento o integración de los estímulos visuales, dadas las características de los rasgos.
- Actividad en electrodos frontales, occipitales y centrales, en la banda δ , a latencia de 300 – 500 ms para épocas de filas. Las características de este mecanismo son similares a las del potencial P300 en el paradigma de evento raro [7], que involucra el reconocimiento consciente (cognitivo) de los estímulos esperados, y que supone la presencia del componente P300 en los ERPs.

Esto no quiere decir que ambos mecanismos no puedan estar presentes en los dos tipos de épocas, pero al menos si se puede sugerir, a partir de la evidencia de los resultados y los argumentos de este capítulo, que en cada caso predomina un tipo de actividad cerebral distinto.

Es posible que, aumentando el número de rasgos candidatos a considerar ($r \gg 100$) y combinando intencionalmente rasgos relativos a ambos mecanismos, pueda mejorar el desempeño de clasificación, pues se estarían tomando en cuenta dos mecanismos que quizás son complementarios durante la operación del Deletreador de Donchin.

El mecanismo o actividad predominante en columnas probablemente esté relacionada con la presencia de los componentes N100 o P200 del ERP [33], y todo parece indicar que el mecanismo relevante para las filas presenta una dinámica similar a la del potencial P300 clásico [7].

Probablemente ambos mecanismos, el cognitivo (P300) y el sensorial (N100, P200), colaboren en la identificación consciente de los estímulos atendidos [7]. Lo interesante es que para épocas de columnas, donde se tuvieron mayores desempeños de clasificación (5.4), predomine el mecanismo sensorial (electrodos parieto-occipitales, bandas θ y α), y que para épocas de filas, predomine el mecanismo cognitivo (electrodos fronto-centrales, banda δ). Esto sugiere que el diseño clásico de la matriz de símbolos en el Deletreador de Donchin es un factor en el nivel de desempeño, que es importante considerar, pues tiene una incidencia en la dificultad de la tarea para la identificación de las épocas atendidas de

filas, que afecta negativamente el desempeño de clasificación.

En la literatura se encontró que efectivamente, el diseño del Deletreador Donchin, que se conoce como Paradigma Fila Columna (RCP del inglés *Row-Column Paradigm*) presenta algunos inconvenientes [60]: *efectos de aglomeración*, que se dan cuando el objetivo está rodeado de objetos similares, que se convierten en distractores, y *errores de adyacencia*, que ocurren en localizaciones cercanas a los símbolos objetivo, que pueden generar potenciales P300 en la estimulación de algunas filas o columnas diferentes a las que contienen al objetivo.

Además, se ha encontrado que el diseño tradicional de la matriz del Deletreador de Donchin depende, además del potencial P300, de los potenciales evocados visuales, en los cuales está involucrada la dirección de la mirada [61].

Es muy probable que estos efectos no deseables del diseño de la matriz de símbolos, jueguen un papel importante en las diferencias de desempeño de clasificación encontradas entre épocas de filas y columnas en este trabajo, pues parece que tienen un efecto en los rasgos espectrales, como se ha descrito a detalle a lo largo de éste y el capítulo anterior.

Ante estos inconvenientes, se han propuesto en la literatura otras alternativas en el diseño de la interfaz del Deletreador que buscan evitar los inconvenientes del diseño clásico de filas y columnas (RCP) [62]: Paradigmas de carácter único (SCP), basados en regiones (RBP), de tablero de ajedrez (CBP), y con estímulos móviles y alternativos, con los cuales se han logrado mejoras en el desempeño de clasificación respecto al Paradigma

clásico (RCP).

En el capítulo de Antecedentes (2), se mencionaron algunos estudios sobre la dinámica espectral durante tareas con el paradigma de evento raro, donde se encontró que la detección del estímulo atendido está relacionada con actividad en las bandas θ , α y β , a latencias de 300 ms post-estímulo [20] [21].

Respecto a esos estudios, los rasgos candidatos encontrados en este trabajo a latencias cercanas a 300 ms (7 de 10 sujetos), son importantes principalmente para épocas de filas. Sin embargo aquí resultó que la banda de frecuencia δ fue la más relevante, seguida de α , y solo para un sujeto (DLP) fueron las bandas θ y β (Tabla 5.1). Para épocas de columnas, solo tres sujetos (DMA, LAG e IZH) presentan rasgos candidatos a latencias de 300 – 500 ms (5.2), en los cuales las bandas de frecuencia de interés predominantes son β , α y θ , precisamente las reportadas anteriormente [20], [21].

La presente discusión, soportada por los resultados, sugiere que los mecanismos cognitivos y atencionales, que dan origen al potencial P300 ante la presentación de estímulos visuales en el Deletreador de Donchin, también generan actividad en el EEG que está codificada en el dominio tiempo-frecuencia (espectral), y que es susceptible de cuantificarse mediante la CWT.

Además, se observó que la metodología propuesta en este trabajo es capaz de seleccionar y aprovechar esa información espectral para optimizar la clasificación de las épocas de EEG en la mayoría de sujetos y condiciones, alcanzando una mayor eficacia para cla-

sificar épocas de columnas que de filas. El promedio entre los 10 sujetos, del valor de $AUROC_{nob}(N)$ para $N = 10$ es de 0.88 ± 0.08 para épocas de filas, y 0.80 ± 0.08 para épocas de columnas).

La Fig. 5.1 y la Tabla 5.4 muestran precisamente, que la información espectral del EEG, codificada en la magnitud de los coeficientes CWT, brinda información importante para la identificación de las épocas atendidas y no atendidas, pues además, hay una clara tendencia a tener mayores valores de $AUROC_{prom}(N)$ y $AUROC_{nob}(N)$, conforme se incrementa el valor de N (el número de épocas promediadas).

Por otro lado, la fase es una característica espectral que puede codificar información que no es susceptible de observarse con medidas de magnitud o potencia espectral de la señal [25]. Aunque esta posibilidad se exploró en una etapa preliminar de este trabajo con el PLV (*Phase Locking Value* [63]), no se incluyó finalmente porque al inicio se encontraron resultados preliminares menos alentadores con información de fase que con información de magnitud espectral (wavelet).

Además, el análisis se complica por el elevado número de combinaciones a explorar entre canales, bandas de frecuencia, etc, y finalmente, se optó por enfocar este trabajo solo a la magnitud espectral.

La literatura indica que el reinicio de fase es un posible mecanismo de generación de los ERPs que puede contener información complementaria a la que sugiere la teoría aditiva de generación de los ERPs [26] [64]. Probablemente, como se ha encontrado para

trabajos sobre sistemas BCI basados en imaginación motora [65], al combinar información de magnitud espectral, como la magnitud CWT, con información de fase, como el *PLV* u otras medidas, pueda elevarse el desempeño de clasificación de las épocas de EEG en el Deletreador de Donchin.

7.1. Conclusiones

Se lograron los objetivos planteados al inicio del trabajo:

- **Objetivo General:**

Se implementó un método para clasificar las épocas de EEG atendidas y no atendidas de un Deletreador de Donchin, a partir de información espectral (coeficientes CWT) y un clasificador SVM-RBF.

- **Objetivos Específicos:**

Para cada sujeto, se implementó un procedimiento para seleccionar una combinación óptima de hiperparámetros de la SVM-RBF y un subconjunto de rasgos, que maximizan el desempeño de clasificación para cada sujeto.

Se entrenó y evaluó el desempeño de un clasificador óptimo para cada valor de N y tipo de épocas (filas o columnas).

Se identificaron 2 factores que parecen influir en el desempeño de clasificación. Primero, se observó que para optimizar la clasificación de épocas de filas y de columnas, se requieren rasgos con diferencias notables en frecuencia, latencia y canales de registro, que parecen apuntar a dos mecanismos cerebrales distintos:

- Mecanismo cognitivo o de reconocimiento consciente: para épocas de filas, en la mayoría de los sujetos predominan los rasgos en las bandas de frecuencia δ y α , a latencias de 300 – 500 ms, y en canales frontales, centrales y parieto-occipitales (Fz, C3, C4, Oz, PO8, PO7).
- Mecanismo sensorial(visual): para épocas de columnas predominan los rasgos en las bandas α , θ y β , a latencias de 100 – 200 ms, y en canales parieto-occipitales (Oz, P08, PO7).

El segundo factor que parece influir en el desempeño de clasificación es la homogeneidad de los subconjuntos de rasgos evaluados:

- Cuando los subconjuntos de rasgos son homogéneos (épocas de columnas), en la mayoría de casos se seleccionan como óptimos aquellos con el mayor número de elementos ($r_{opt} = \{50, 100\}$), y cuando los subconjuntos de rasgos son poco homogéneos (épocas de filas), se seleccionan aquellos con pocos rasgos ($r_{opt} = \{1, 20\}$).
- Los sujetos con mayor desempeño de clasificación tanto en entrenamiento, como en la prueba con datos no-observados, sin importar si se trata de épocas de filas o de

columnas, tuvieron rasgos más homogéneos, concentrando la totalidad, o al menos la mayor parte de sus rasgos candidatos, en uno o un par de canales de registro, una sola banda de frecuencia y un intervalo estrecho de latencia.

Finalmente se tiene que:

- Se lograron identificar dos mecanismos cerebrales predominantes para épocas de filas y de columnas:

Uno relacionado con el componente visual (sensorial) de la tarea, con rasgos en electrodos parietales/occipitales (Oz, PO8, PO7), frecuencias en las bandas α , θ y β , y latencias de 100 – 200 ms (épocas de columnas).

Otro relacionado con el reconocimiento consciente de los estímulos (componente cognitivo), con rasgos en las bandas de frecuencia δ y θ , latencias de 300 – 500 ms, dispersos en los canales de registro Fz, C3, C4, Oz, PO8 y PO7 (épocas de filas).

- A partir de rasgos espectrales (coeficiente CWT) relacionados con estos dos mecanismos, el método es capaz de clasificar las épocas de EEG del Deletreador de Donchin.
- Para épocas de columnas, el método tiene un buen desempeño de clasificación, pues los rasgos de los subconjuntos seleccionados como óptimos son en general más homogéneos entre sí, y en la mayoría de los casos, aprovecha todos los rasgos disponibles ($r_{opt} = 100$). Esto resulta en una alta capacidad de generalización de los clasificadores $SVM_{opt}(N)$ para los 10 sujetos.

$$\text{AUROC}_{\text{prom}}(N = 10) = 0.88 \pm 0.06.$$

$$\text{AUROC}_{\text{nob}}(N = 10) = 0.88 \pm 0.08.$$

El mejor sujeto, LAC, obtuvo un $\text{AUC}_{\text{ROC}_{\text{nob}}}(N = 10) = 0.97$.

- Para épocas de filas se tienen rasgos menos homogéneos (que para columnas), de los cuales se seleccionan como óptimos los subconjuntos con los rasgos más similares entre ellos ($r_{\text{opt}} = \{1, 20, 50\}$). Esto resulta en una buena capacidad de generalización de los clasificadores $\text{SVM}_{\text{opt}}(N)$ en 7 de los 10 sujetos.

$$\text{AUROC}_{\text{prom}}(N = 10) = 0.84 \pm 0.05.$$

$$\text{AUROC}_{\text{nob}}(N = 10) = 0.80 \pm 0.08.$$

El mejor sujeto, LAC, obtuvo un $\text{AUC}_{\text{ROC}_{\text{nob}}}(N = 10) = 0.95$.

Así, se puede concluir que el método propuesto es útil para clasificar las épocas de EEG atendidas/no-atendidas en el Deletreador de Donchin, a partir de los coeficientes de la CWT y un clasificador SVM-RBF, pues es capaz de identificar y seleccionar los rasgos e hiperparámetros que optimizan el desempeño de clasificación para cada sujeto y condición.

A partir de los rasgos espectrales de este método, se identificó la participación de un mecanismo sensorial y un mecanismo cognitivo que, probablemente actúan de manera combinada durante la operación de la BCI, pero que tienen una contribución diferente para las épocas correspondientes a estimulaciones de filas y de columnas en la matriz del

Deletreador de Donchin.

Finalmente, se sugiere que la contribución de estos mecanismos, y la homogeneidad de los rasgos espectrales seleccionados, son dos factores que influyen en las diferencias del desempeño de clasificación entre épocas de filas y columnas.

7.2. Trabajo a Futuro

Primero, se puede explorar un espacio más denso de valores de los hiperparámetros de la SVM-RBF, para buscar un mayor desempeño de clasificación, no sin considerar el riesgo de sobre-entrenar el clasificador si la búsqueda es muy detallada, además del incremento de tiempo en el procedimiento de búsqueda.

También, es necesario utilizar más registros para tener un mayor número de épocas con las cuales entrenar y probar los clasificadores óptimos, pues con el número de épocas atendidas utilizadas en este experimento (315 de filas y 315 de columnas), al llegar al caso de $N = 10$ se tenían solamente 31 elementos de cada clase (promedios), a repartir entre los conjuntos de entrenamiento y de prueba.

Teniendo en cuenta que para muchos sujetos se seleccionaron rasgos concentrados en 1, 2 ó máximo 4 canales de registro, puede ser útil incrementar el número de rasgos candidatos ($r \gg 100$) y de subconjuntos de rasgos, para dar al método más opciones para elegir subconjuntos con mejores desempeño de clasificación. Con el método actual solo se consideran 100 rasgos candidatos y 4 subconjuntos de ellos, aunque después se observó que en

algunos casos había literalmente miles de rasgos candidatos, repartidos entre la mayoría y a veces en los 10 canales de registro.

Falta aún probar la metodología propuesta en el contexto de la operación real del *Deletreador de Donchin*, tomando las épocas de EEG tal como se generan durante la operación de la BCI (de acuerdo a las secuencias de estimulación presentadas al sujeto, correspondientes a cada símbolo deletreado), acumulando y promediando los coeficientes de la CWT (los rasgos espectrales), y clasificando con los SVM_{opt} . En tal caso, la métrica de desempeño ya no sería el AUROC, sino la sensibilidad y especificidad, o bien el % de aciertos al determinar los símbolos objetivo (atendidos) en la matriz del Deletreador.

Ahora bien, un procedimiento de selección de rasgos en base a un algoritmo de búsqueda que considere la correlación entre conjuntos de rasgos, como la *información mutua*, probablemente ofrezca beneficios al desempeño del método, sobre todo si se encuentran rasgos que solo cuando están combinados resulten útiles para la tarea de clasificación en cuestión. En este trabajo, el principal criterio de la selección preliminar de rasgos (los valores p y H que resultan de las pruebas estadísticas) está orientado a maximizar únicamente la separabilidad entre clases con cada rasgo individual.

De implementar los cambios planteados en el párrafo anterior, se puede agregar a la metodología algún método de selección de canales, como *SVM-RFE* [66] o alguno de filtrado espacial, como el algoritmo de *patrones espaciales comunes*, que ha sido utilizado con buenos resultados en sistemas BCI basados en P300 [67] [68].

La idea de los dos párrafos anteriores, es que al considerar la actividad EEG de varios canales en diferentes bandas de frecuencia y latencias, se puedan identificar otras combinaciones de rasgos y/o mecanismos atencionales que no se encontraron en este trabajo, y que resulten en mejoras al desempeño de clasificación.

Aunque es poco probable que mejoren significativamente el desempeño de clasificación por sí mismos, se pueden probar otros algoritmos de clasificación, como SVM lineal o LDA, entre otros. Estos tendrían la ventaja de no tener que optimizar simultáneamente dos hiperparámetros, como en la SVM-RBF, acortando el tiempo empleado para la optimización del modelo de clasificación. Sin embargo, el parámetro importante para elegir el algoritmo de clasificación siempre debe estar en función del desempeño de clasificación y no del tiempo de entrenamiento, pues una vez entrenado el clasificador, éste simplemente se utiliza con nuevos datos, justo como se hizo en la última etapa de la metodología en este trabajo.

Así, se puede concluir que el método propuesto es útil para clasificar las épocas de EEG atendidas/no-atendidas en el Deletreador de Donchin, a partir de los coeficientes de la CWT y un clasificador SVM-RBF, pues es capaz de identificar y seleccionar los rasgos e hiperparámetros que optimizan el desempeño de clasificación para cada sujeto y condición.

A partir de los rasgos espectrales de este método, se identificó la participación de un mecanismo sensorial y un mecanismo cognitivo que, probablemente actúan de manera

combinada durante la operación de la BCI, pero que tienen una contribución diferente para las épocas correspondientes a estimulaciones de filas y de columnas en la matriz del Deletreador de Donchin.

Finalmente, se sugiere que la contribución de estos mecanismos, y la homogeneidad de los rasgos espectrales seleccionados, son dos factores que influyen en las diferencias del desempeño de clasificación entre épocas de filas y columnas. Estos factores probablemente estén relacionados con inconvenientes característicos del diseño de la matriz del Deletreador Donchin (Paradigma Fila-Columna).

Escalogramas p y H por sujeto

En este Anexo se presentan los resultados de la prueba estadística realizada a cada rasgo espectral individual, con las épocas de EEG de los 4 registros de deletreo dirigido de cada sujeto, calculadas como se explicó a detalle en la sección 4.4. Estos resultados se presentan como gráficos del *escalograma-p* y el *escalograma-H*.

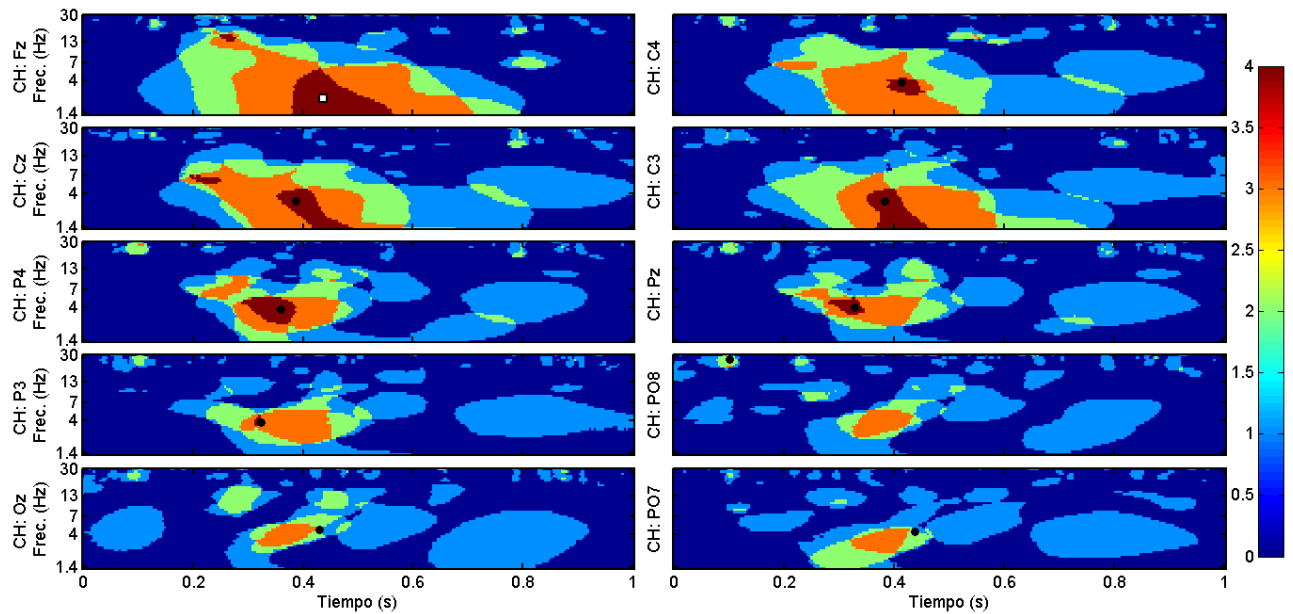
El *escalograma-H* muestra, para cada combinación de canal, frecuencia y latencia (ch, s y m , sección 4.3), el número de registros en los que la prueba estadística rechazó la hipótesis nula (es decir, se obtuvo $H = 1$), esto es, cuando los valores de $MC_{cut}(ch, s, m)$ presentan diferencias estadísticamente significativas entre épocas atendidas y no-atendidas. Así, este es un mapa de rasgos relevantes (aquellos con mayor *valor-H* acumulado en los 4 registros), a partir de los cuales se eligen los 100 rasgos *candidatos* (sección 4.4, de los cuales se toman los subconjuntos a evaluar en la etapa de *selección de rasgos e hiperparámetros óptimos* (sección 4.5).

El *escalograma-p* por su parte, muestra el *valor-p* promedio (entre los 4 registros de EEG) obtenido de la prueba estadística realizada con los valores de $MC_{cut}(ch, s, m)$ de las épocas atendidas (con P300) y no-atendidas (sin P300), en cada combinación de ch, s y m

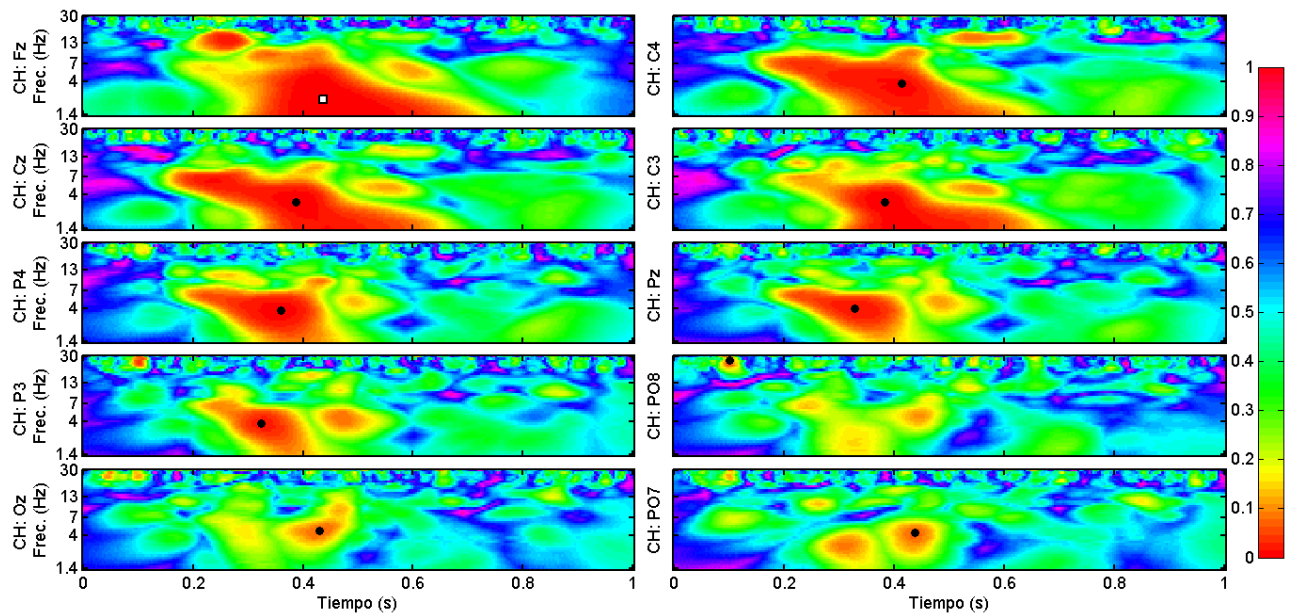
(sección 4.3). La información de este escalograma se utiliza para elegir y ordenar, de entre los rasgos *candidatos*, aquellos 100 con menor *valor-p* promedio en los 4 registros de EEG.

La información de los escalogramas de *valor-p* y *valor-H*, que pueden tomar valores de 0 a 4, se codifican de acuerdo a la escala de colores mostrada en el extremo derecho de las figuras. Para cada sujeto se presenta un gráfico para épocas de filas y otra para épocas de columnas, tanto para el *valor-H* como para el *valor-p*.

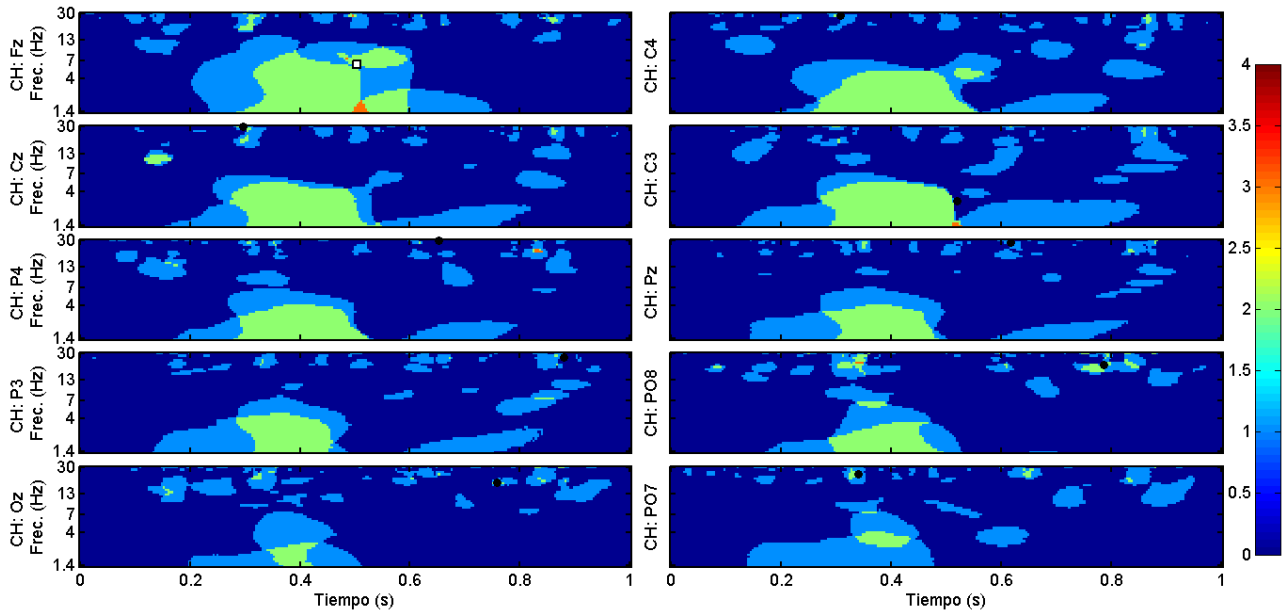
Además, en cada figura se presentan 10 escalogramas, uno por canal de registro, y en cada uno de éstos se indica mediante un pequeño círculo de color negro el punto tiempo-frecuencia con el menor *valor-p*, que se cambia por un cuadrado de color blanco con contorno negro en el caso del mínimo absoluto de *valor-p* entre los 10 canales, pues este es el punto a partir del cual se ordenan ascendentemente los 100 rasgos candidatos seleccionados para cada sujeto y tipo de época (de filas o de columnas).



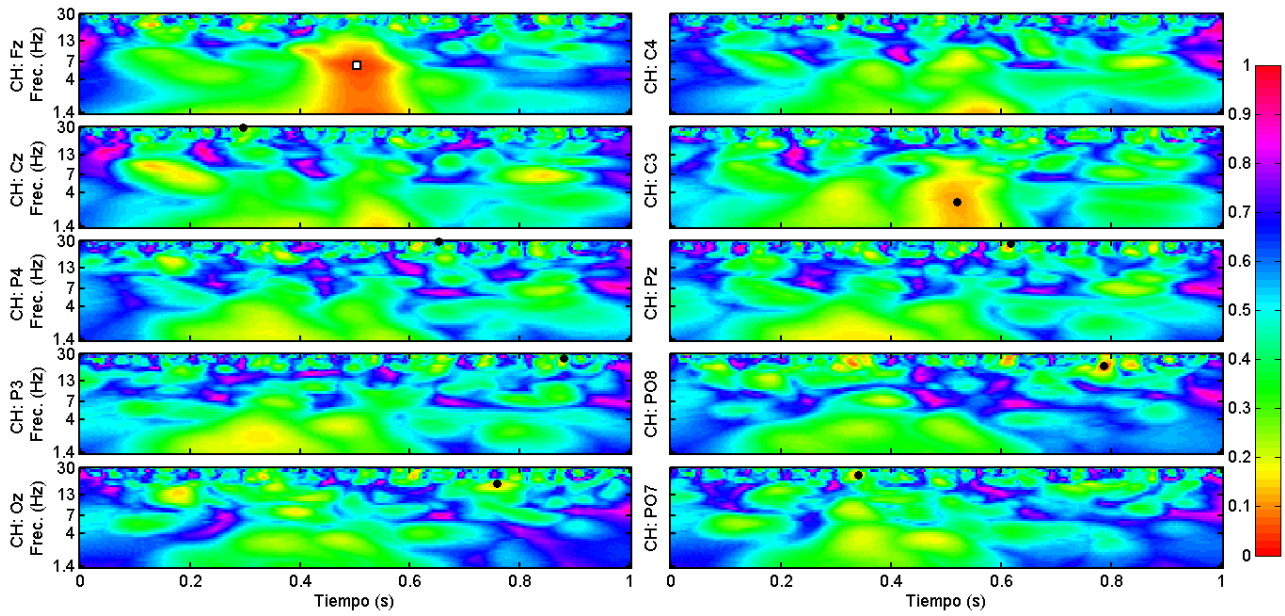
Escalogramas-H de épocas de columnas del sujeto DLP.



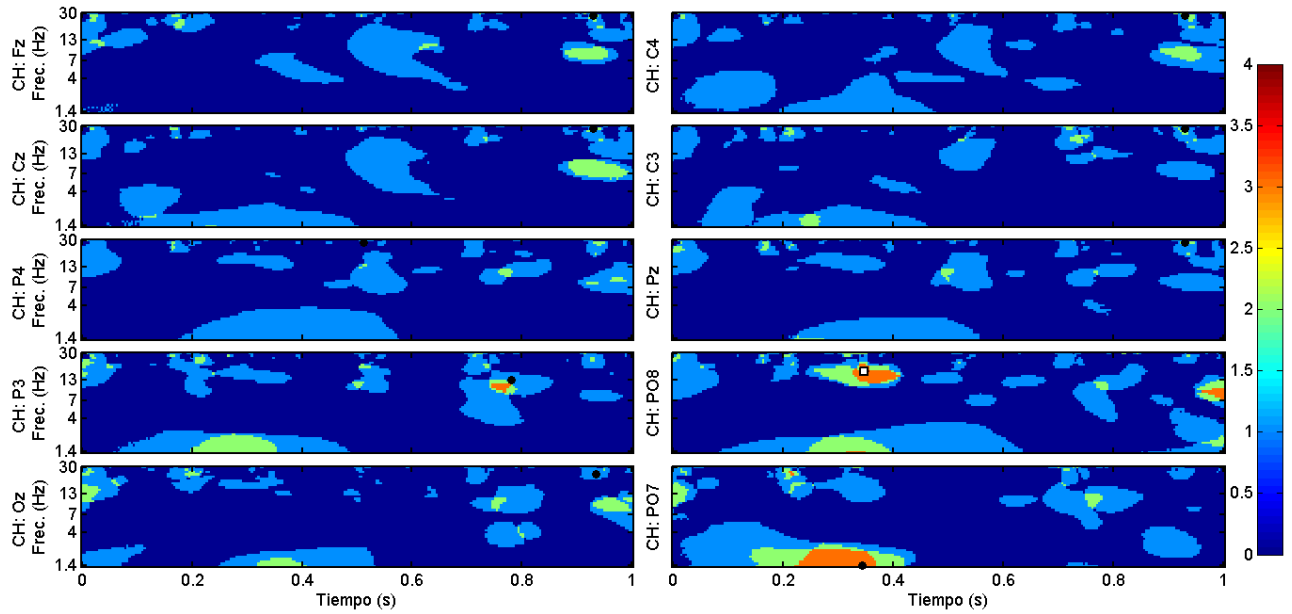
Escalogramas-p de épocas de columnas del sujeto DLP.



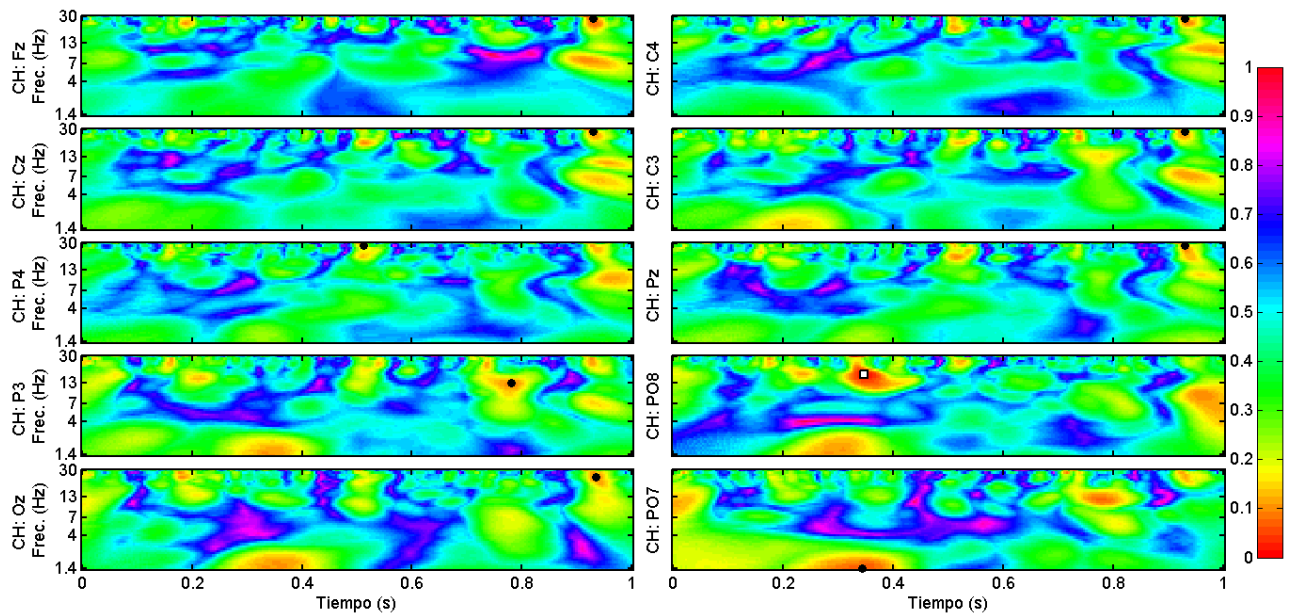
Escalogramas-H de épocas de filas del sujeto DLP.



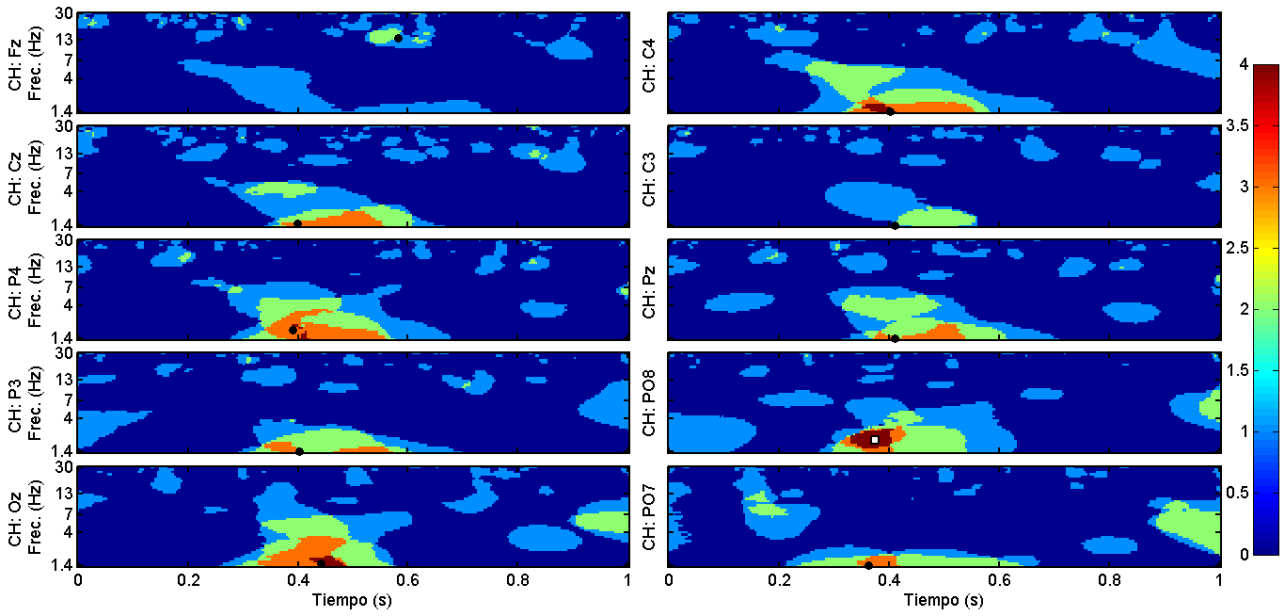
Escalogramas-p de épocas de filas del sujeto DLP.



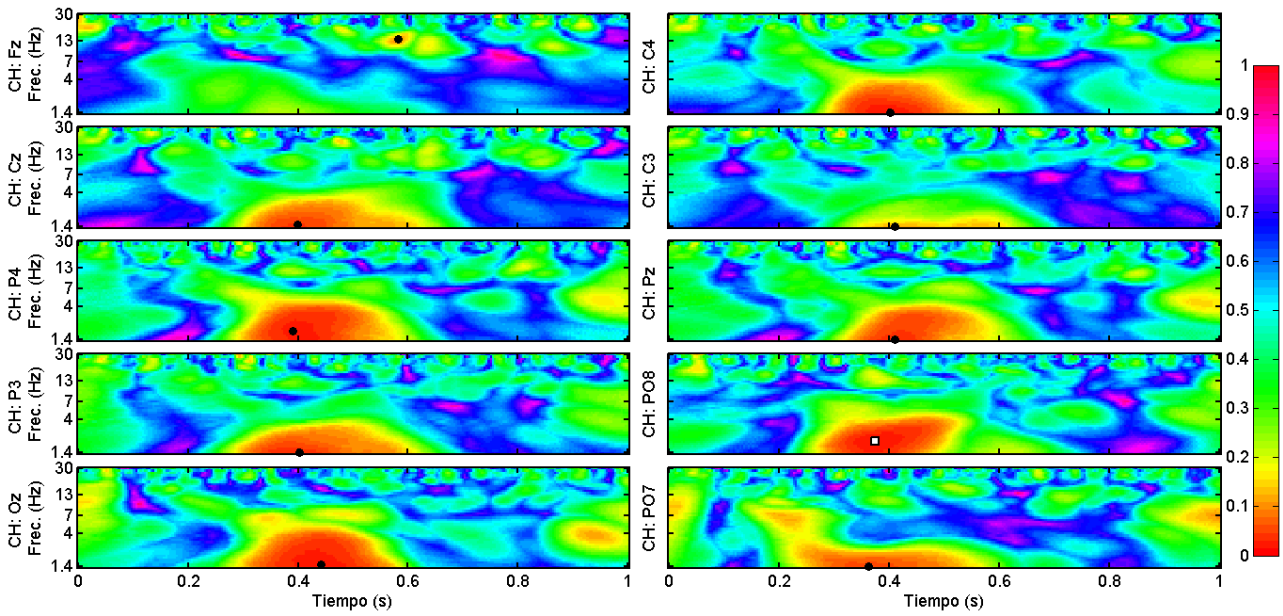
Escalogramas-H de épocas de columnas del sujeto DMA.



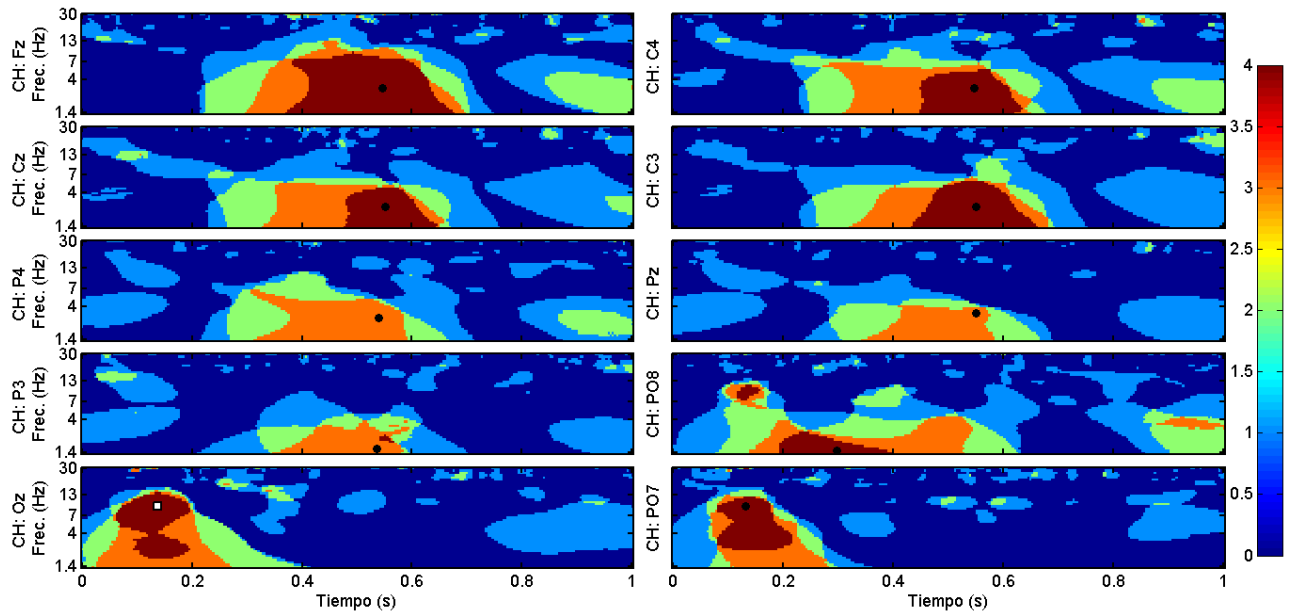
Escalogramas-p de épocas de columnas del sujeto DMA.



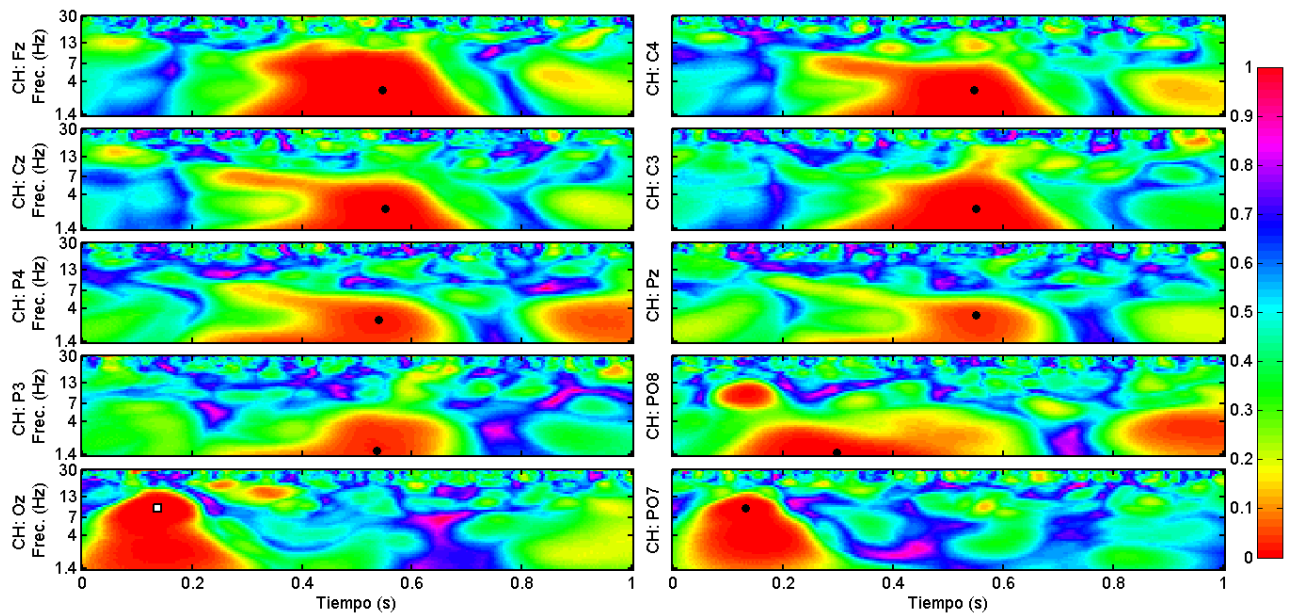
Escalogramas-H de épocas de filas del sujeto DMA.



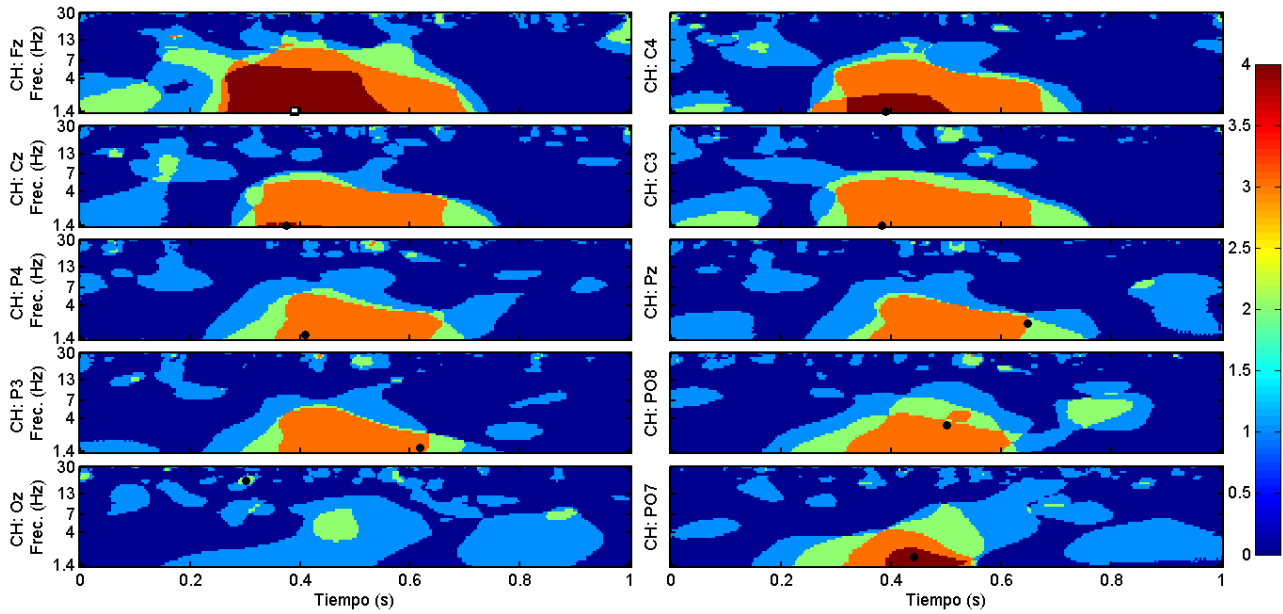
Escalogramas-p de épocas de filas del sujeto DMA.



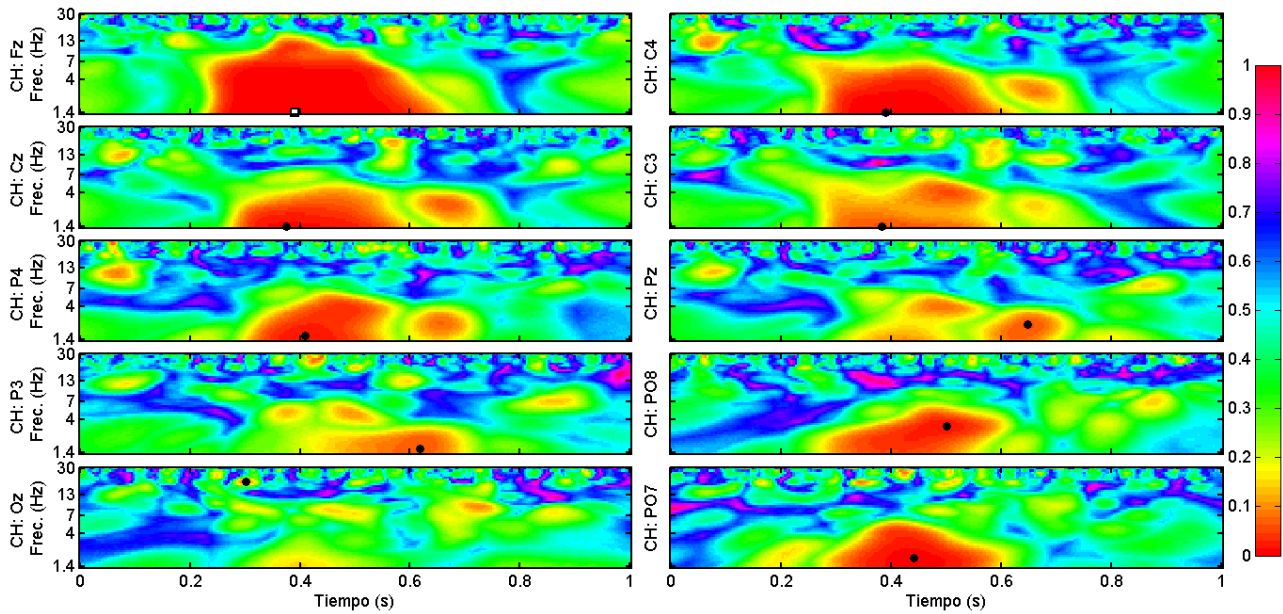
Escalogramas- H de épocas de columnas del sujeto ASG.



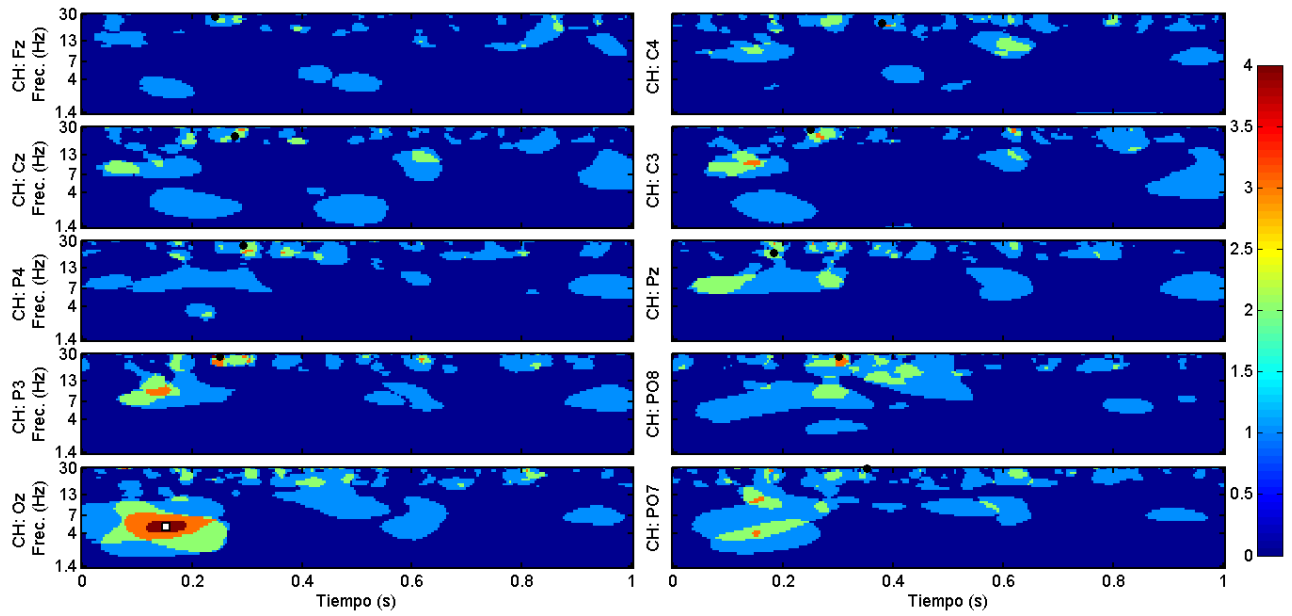
Escalogramas- p de épocas de columnas del sujeto ASG.



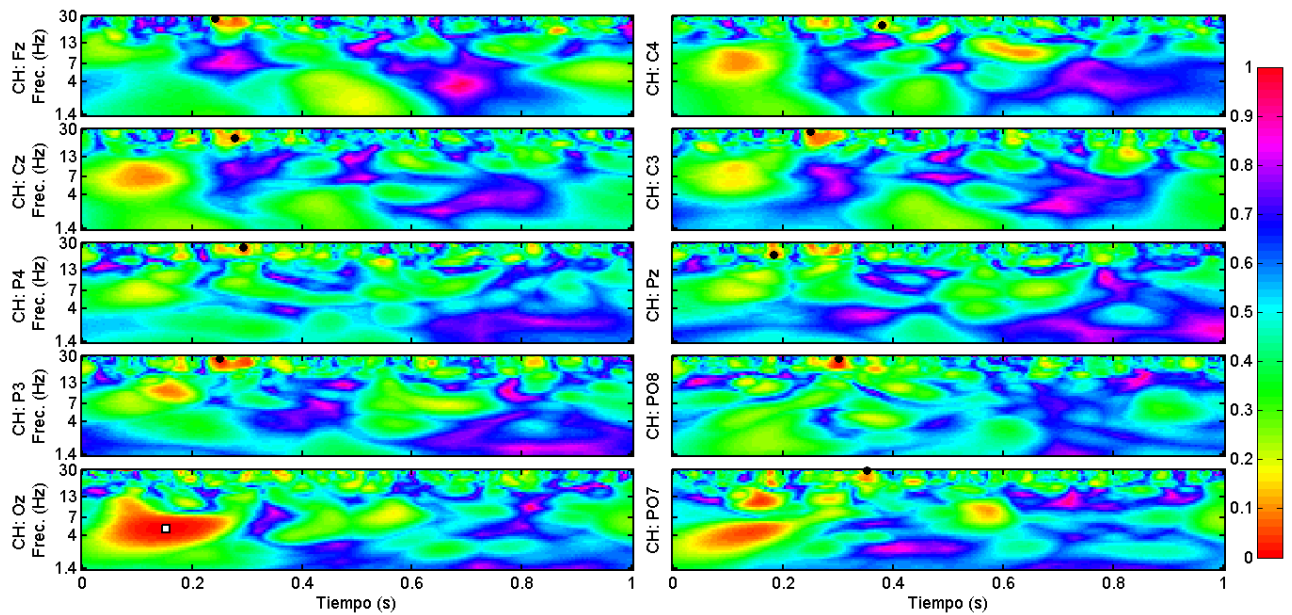
Escalogramas-H de épocas de filas del sujeto ASG.



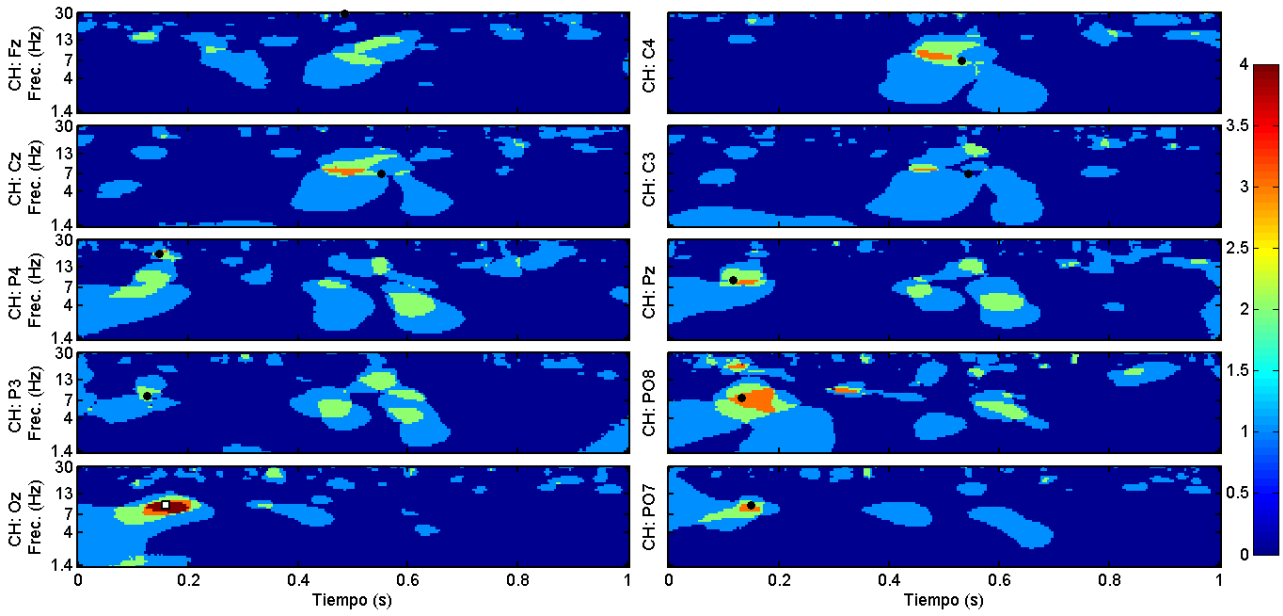
Escalogramas-p de épocas de filas del sujeto ASG.



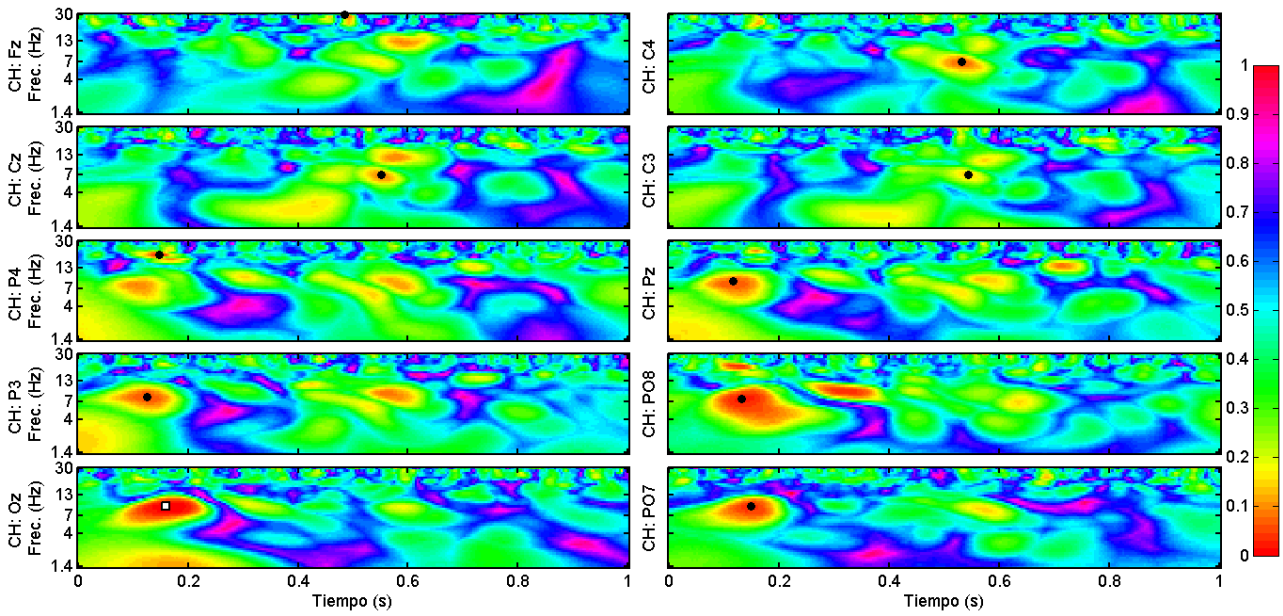
Escalogramas-H de épocas de columnas del sujeto CLL.



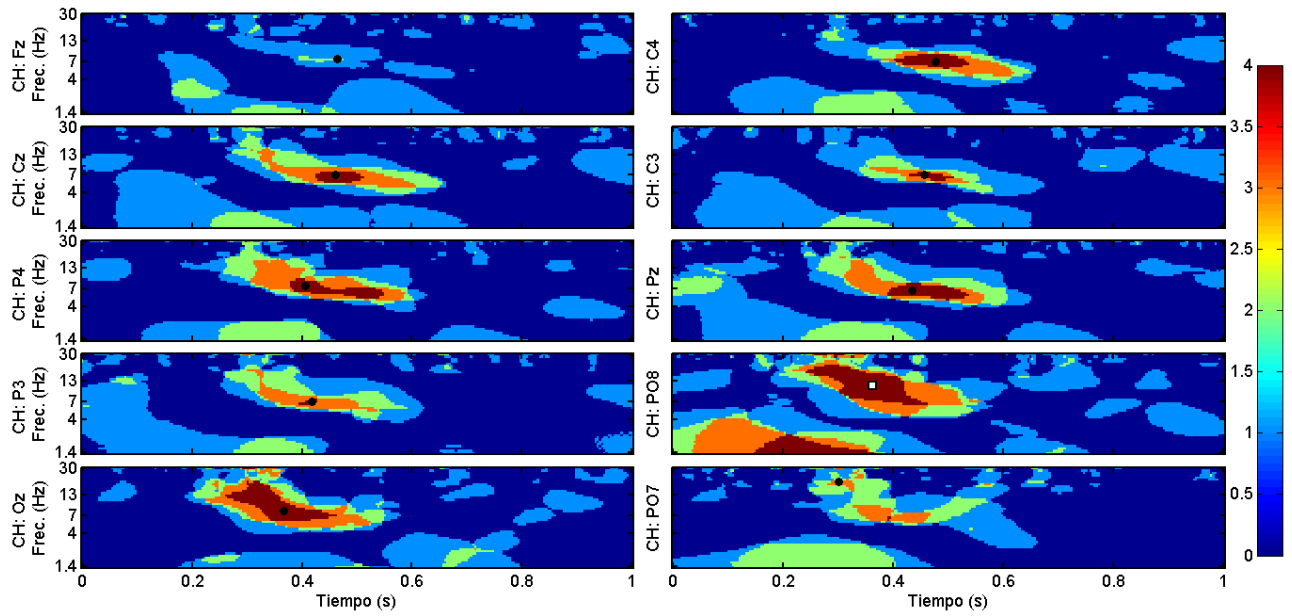
Escalogramas-p de épocas de columnas del sujeto CLL.



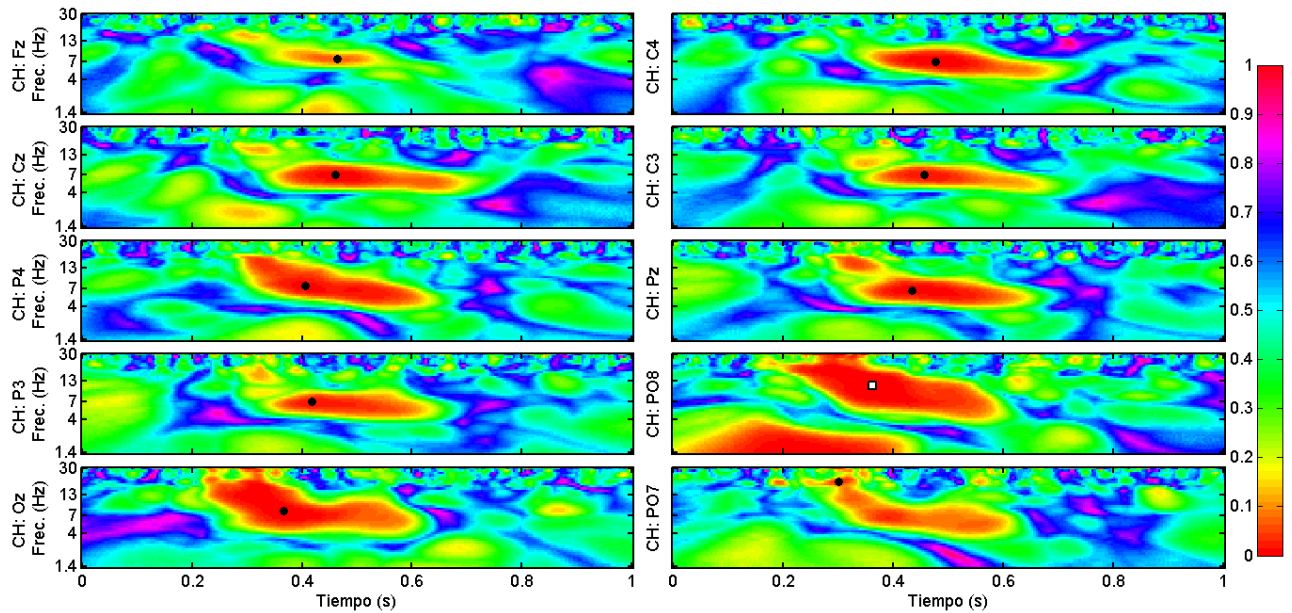
Escalogramas-H de épocas de filas del sujeto CLL.



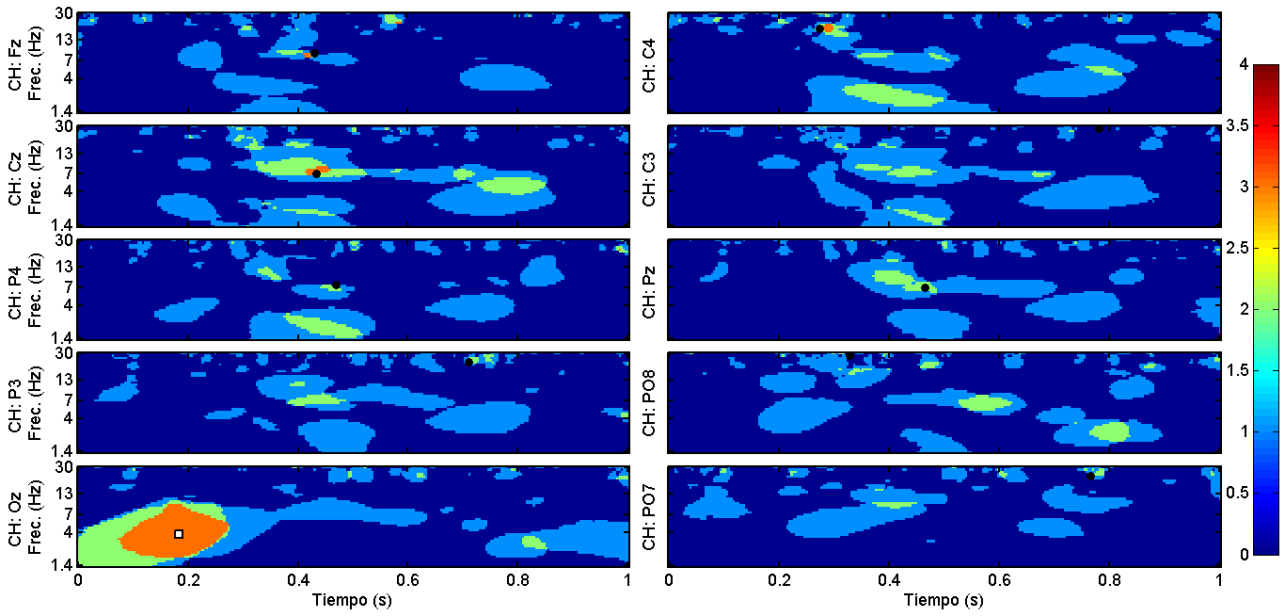
Escalogramas-p de épocas de filas del sujeto CLL.



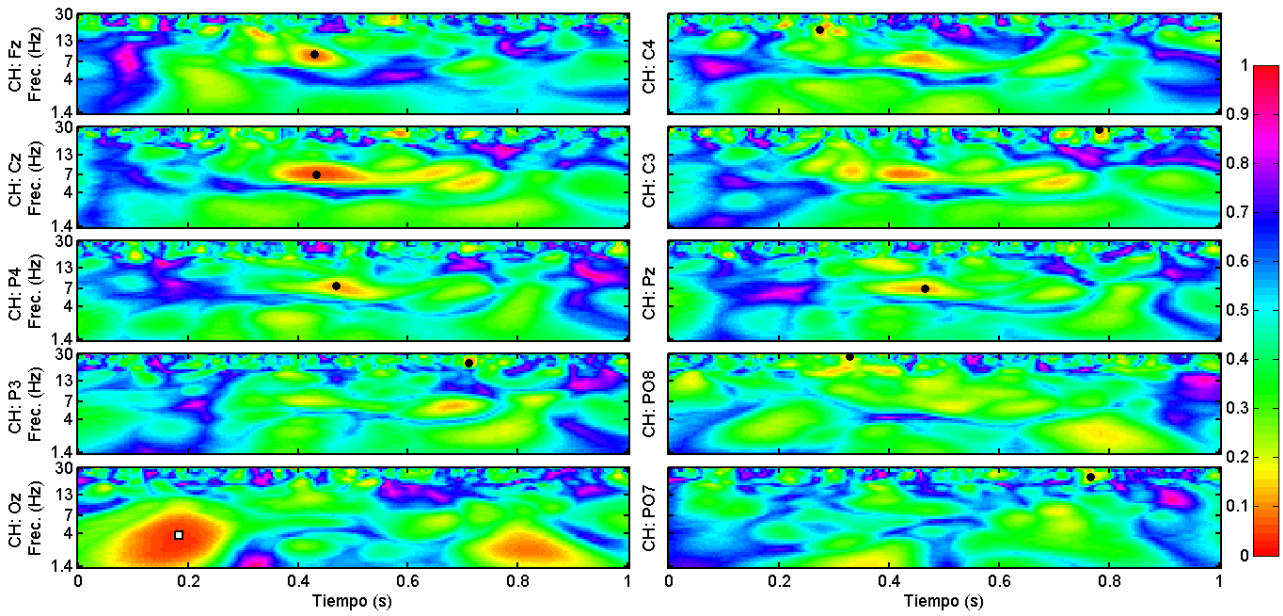
Escalogramas- H de épocas de columnas del sujeto LAG.



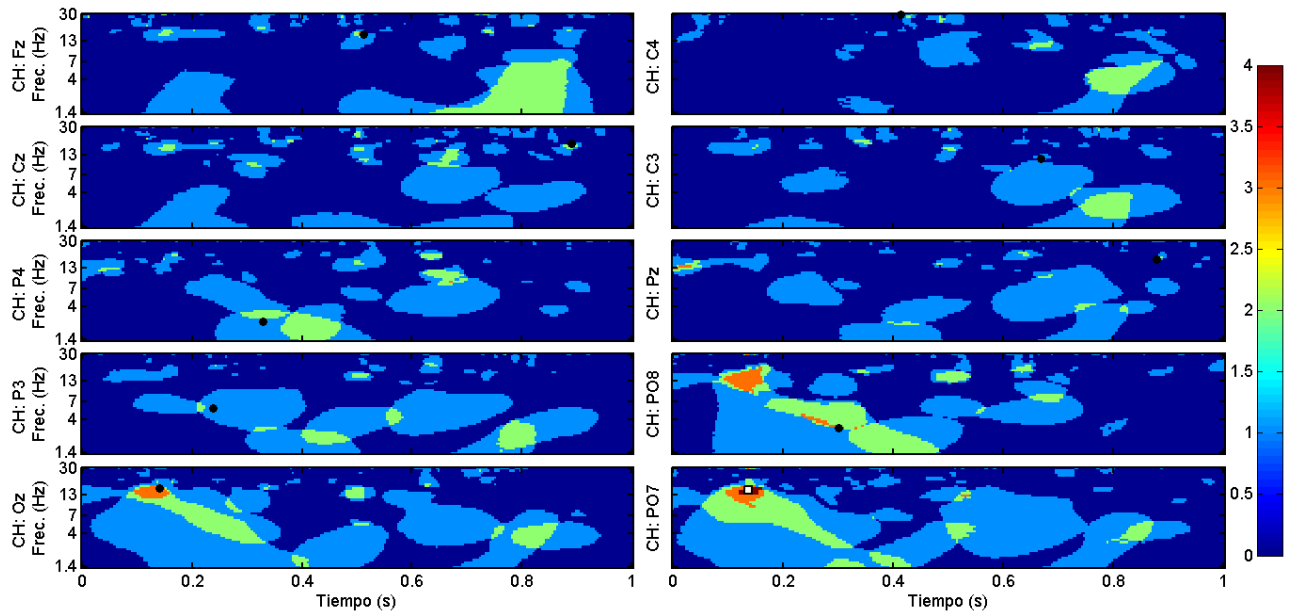
Escalogramas- p de épocas de columnas del sujeto LAG.



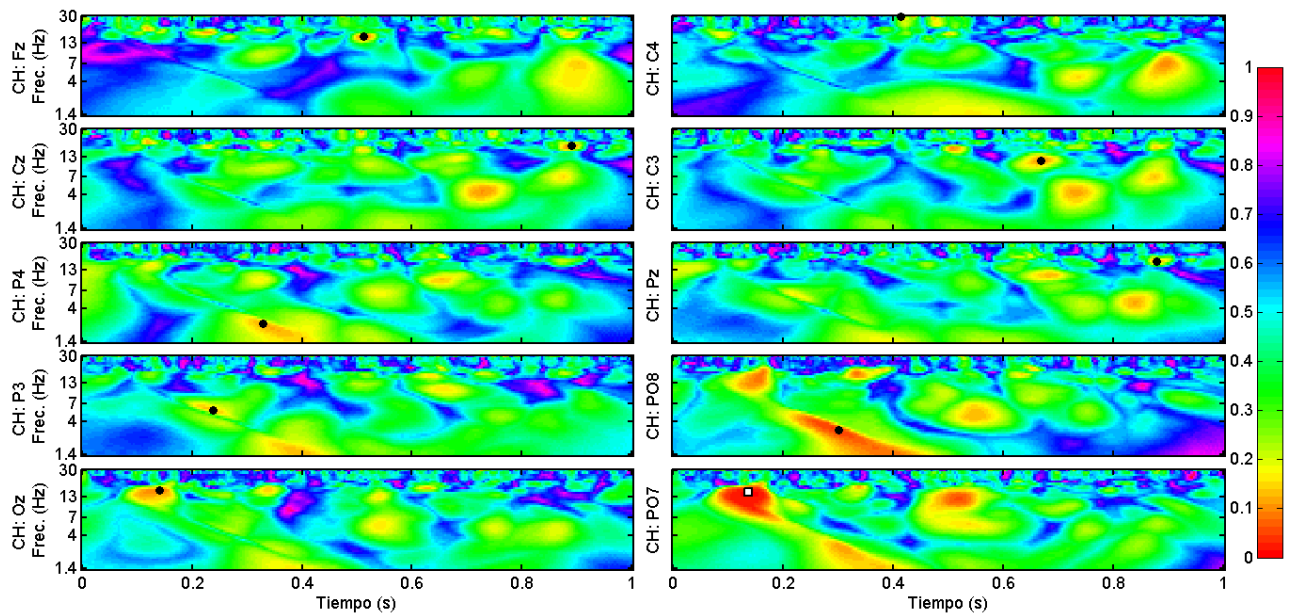
Escalogramas-H de épocas de filas del sujeto LAG.



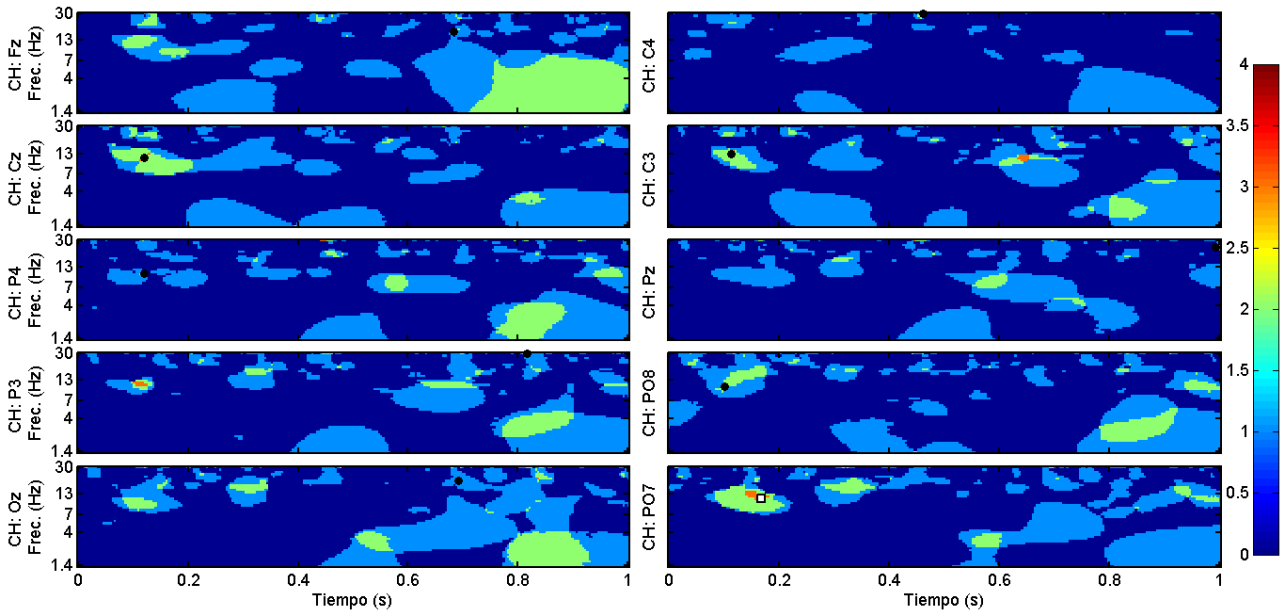
Escalogramas-p de épocas de filas del sujeto LAG.



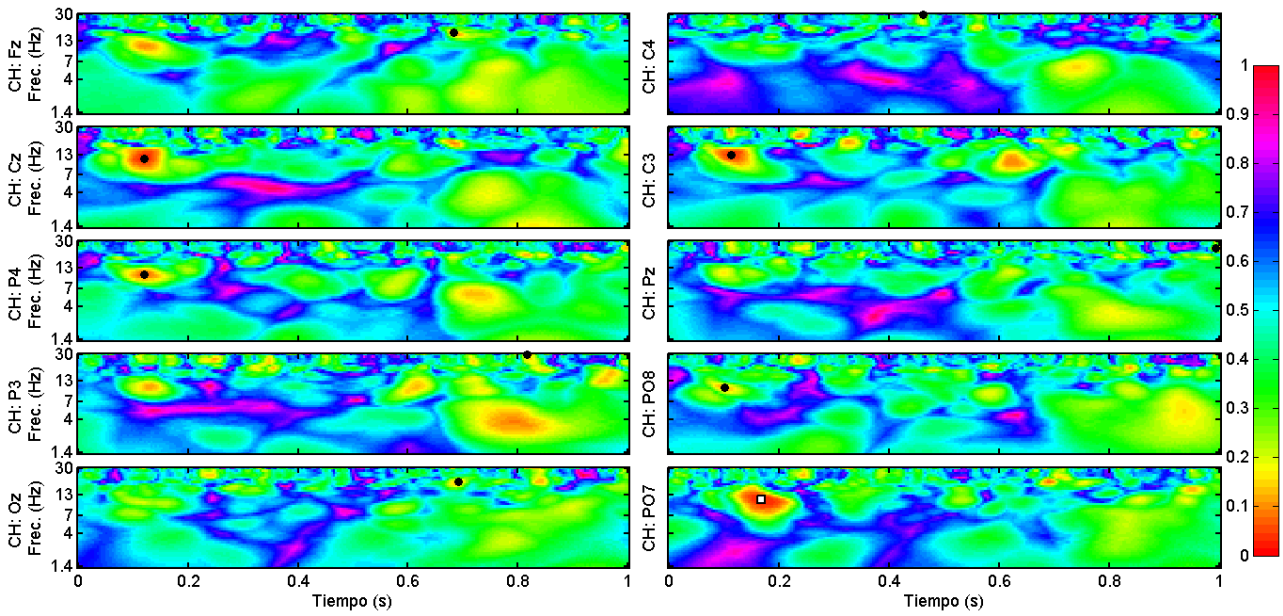
Escalogramas-H de épocas de columnas del sujeto LGP.



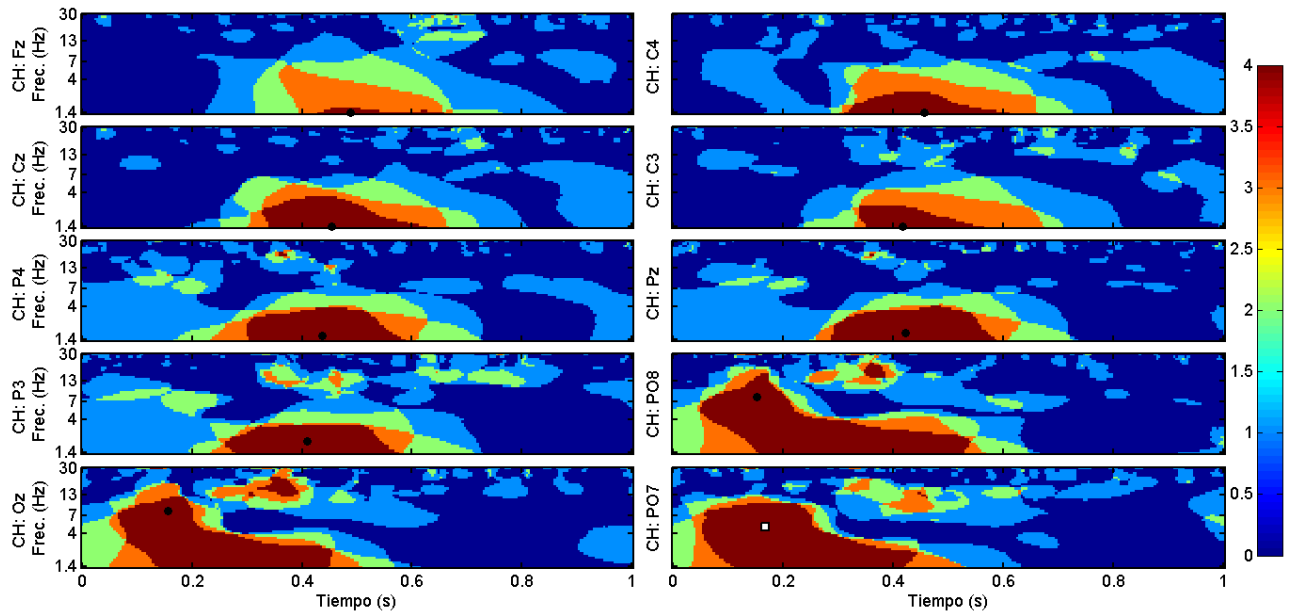
Escalogramas-p de épocas de columnas del sujeto LGP.



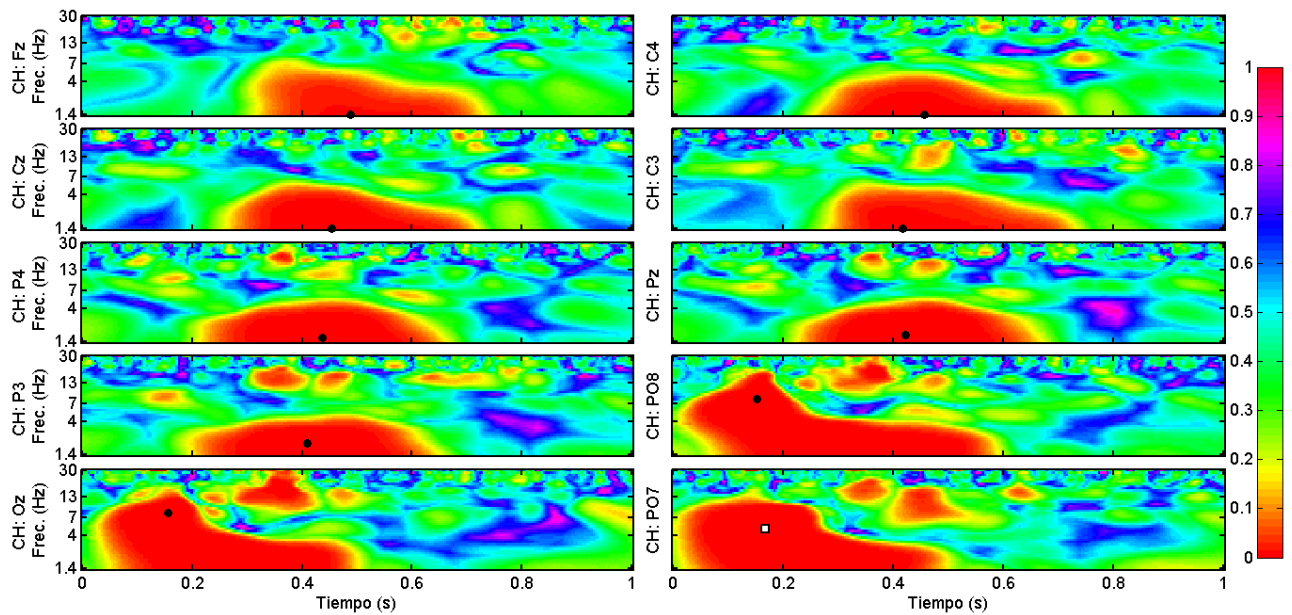
Escalogramas-H de épocas de filas del sujeto LGP.



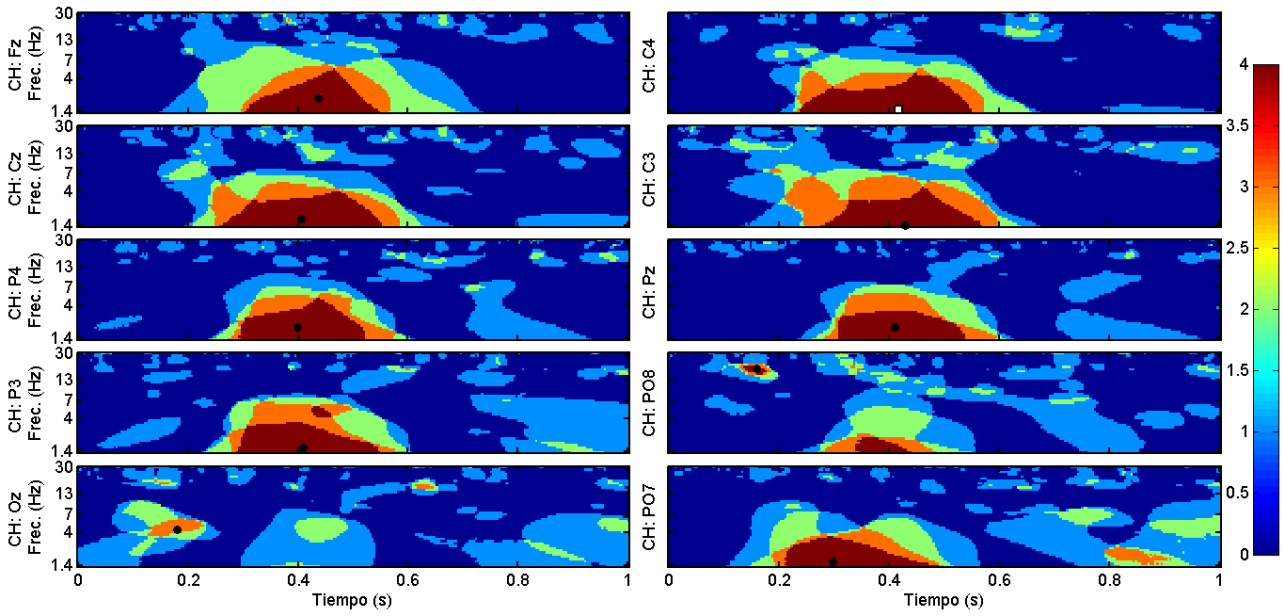
Escalogramas-p de épocas de filas del sujeto LGP.



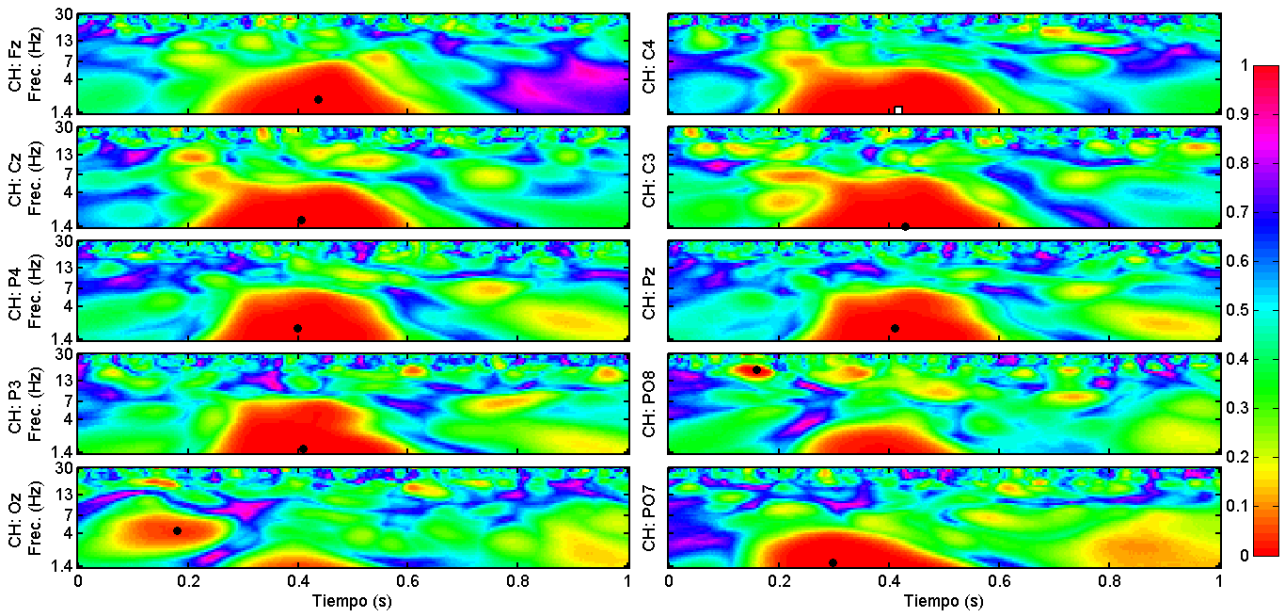
Escalogramas-H de épocas de columnas del sujeto LAC.



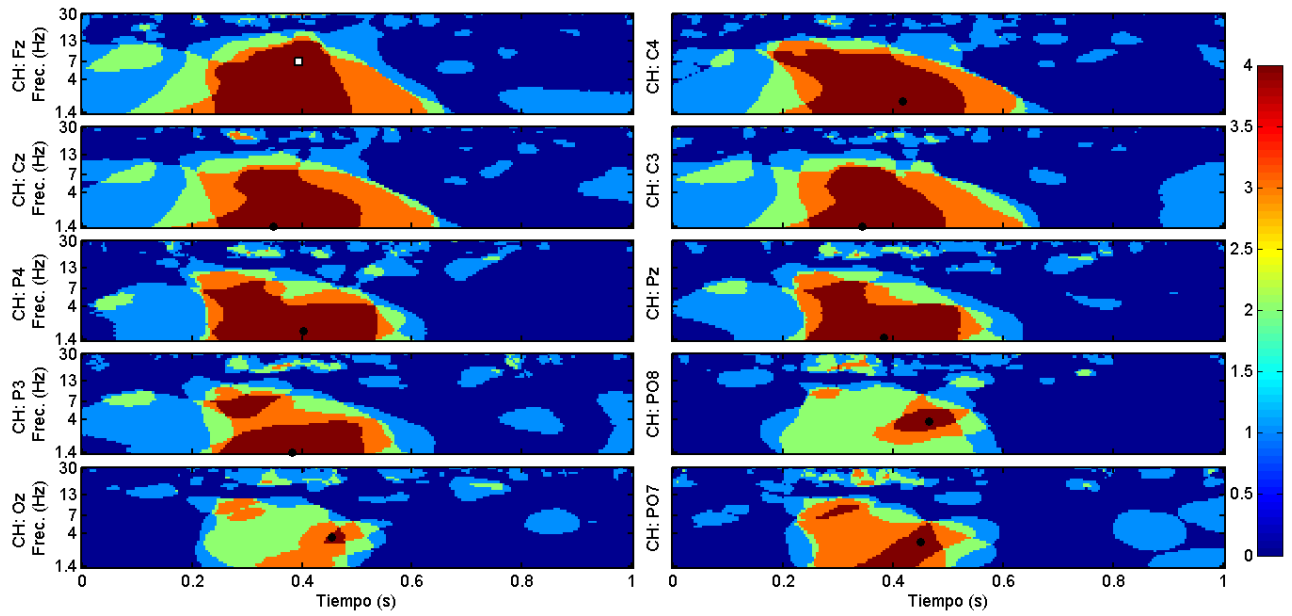
Escalogramas-p de épocas de columnas del sujeto LAC.



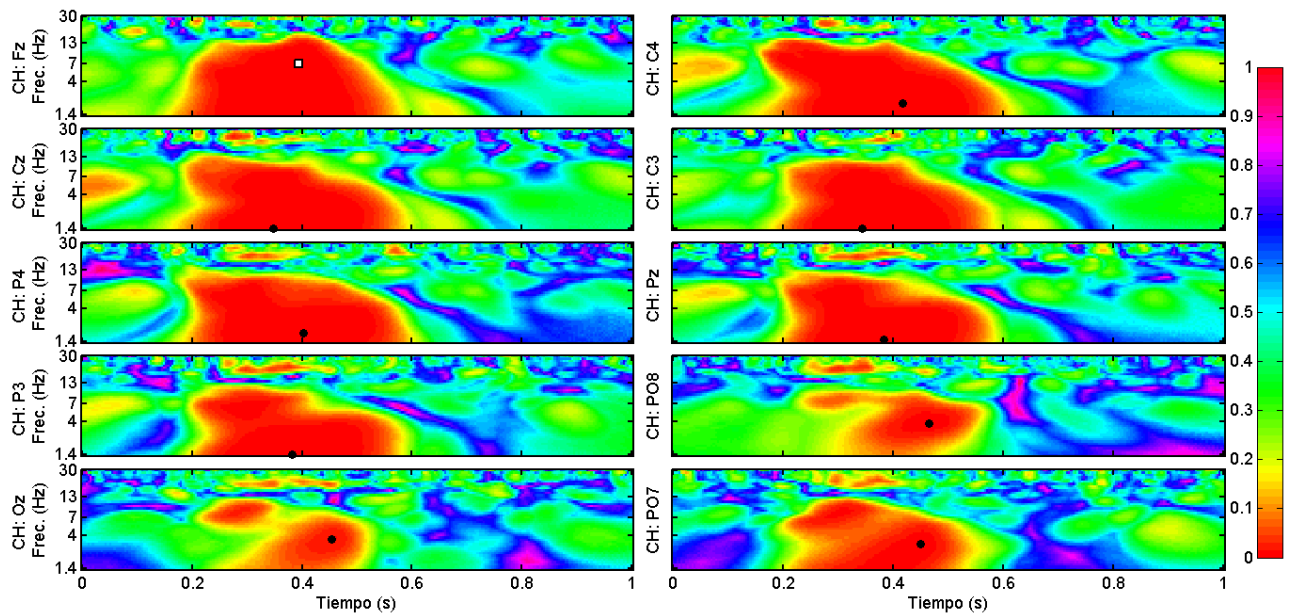
Escalogramas-H de épocas de filas del sujeto LAC.



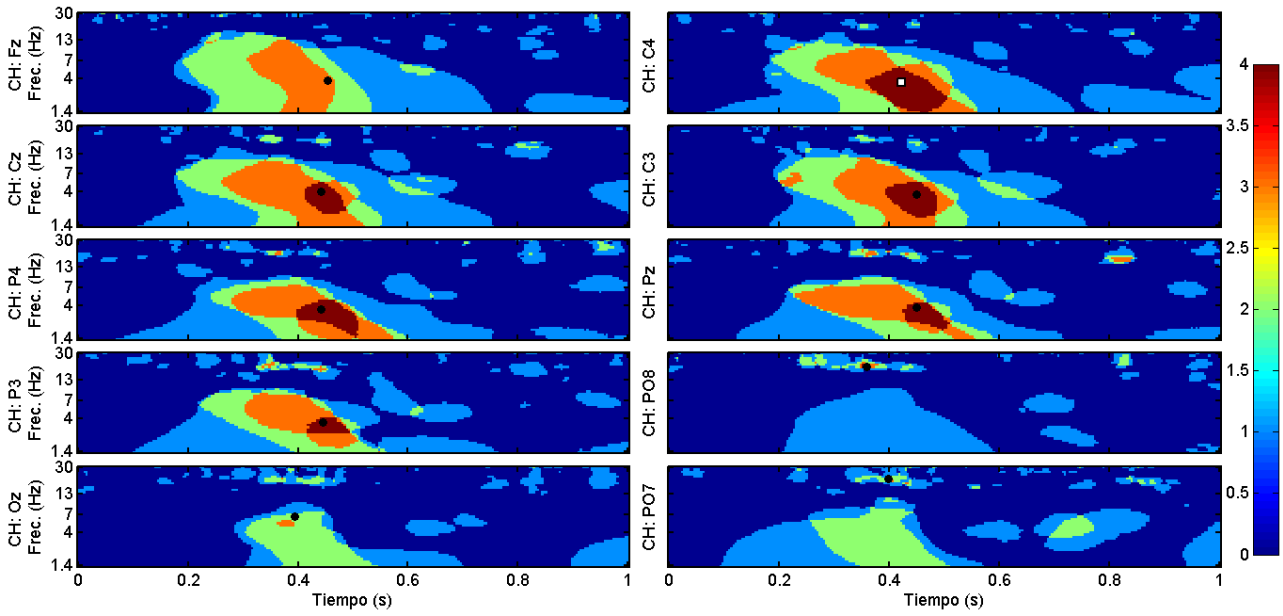
Escalogramas-p de épocas de filas del sujeto LAC.



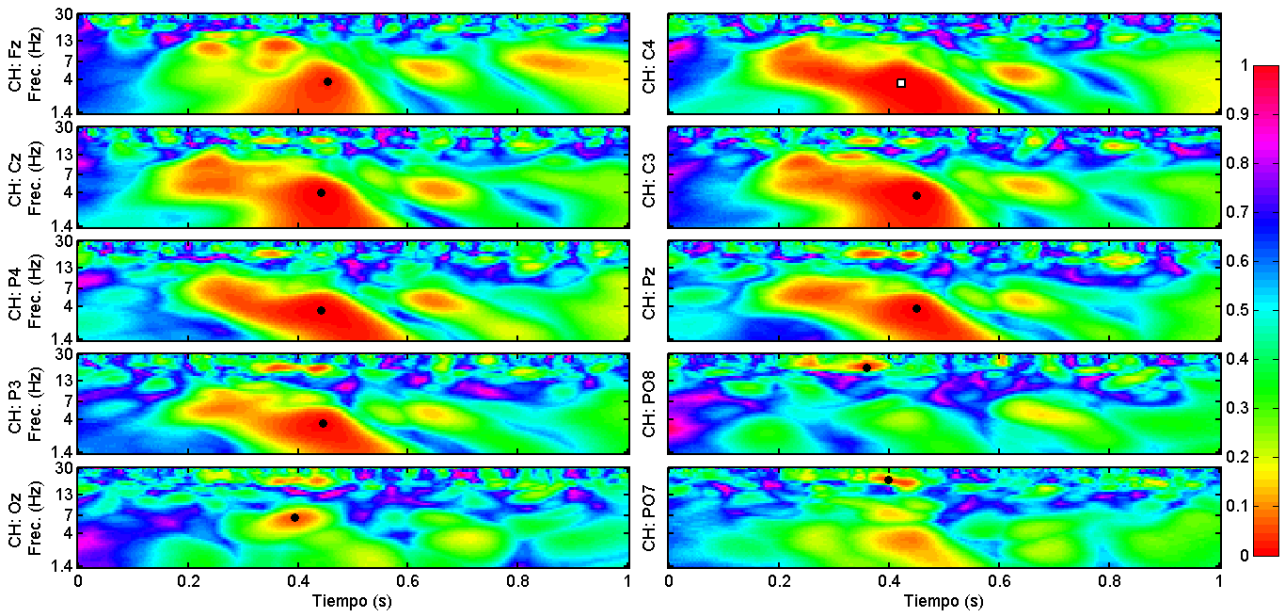
Escalogramas-H de épocas de columnas del sujeto IZH.



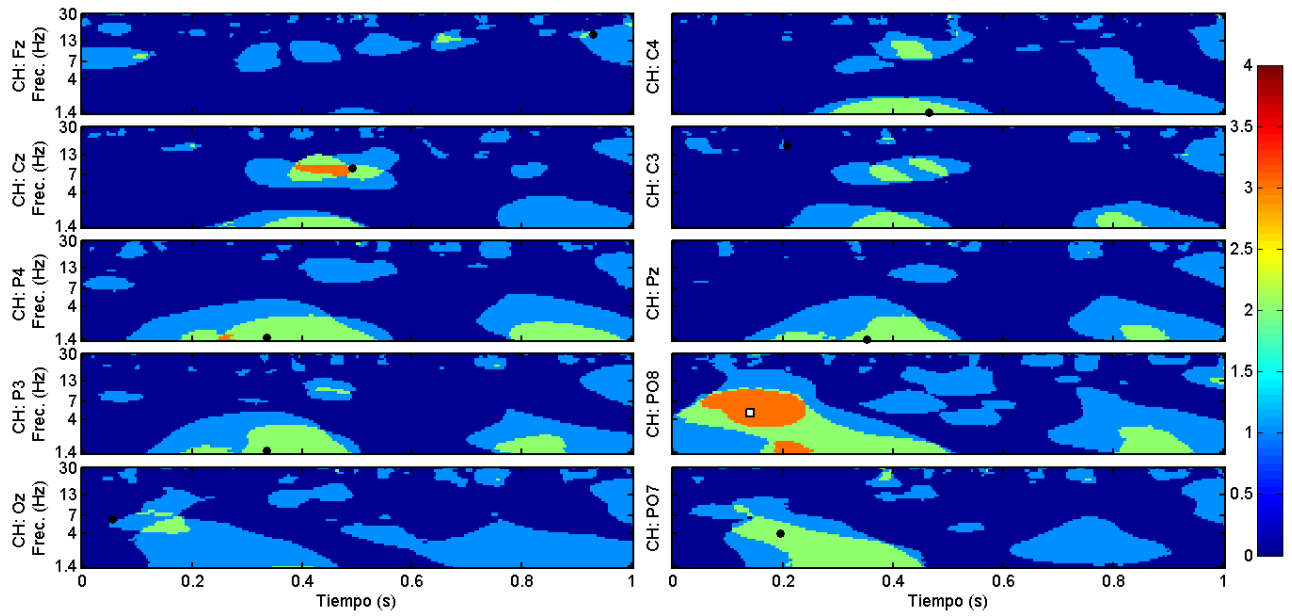
Escalogramas-p de épocas de columnas del sujeto IZH.



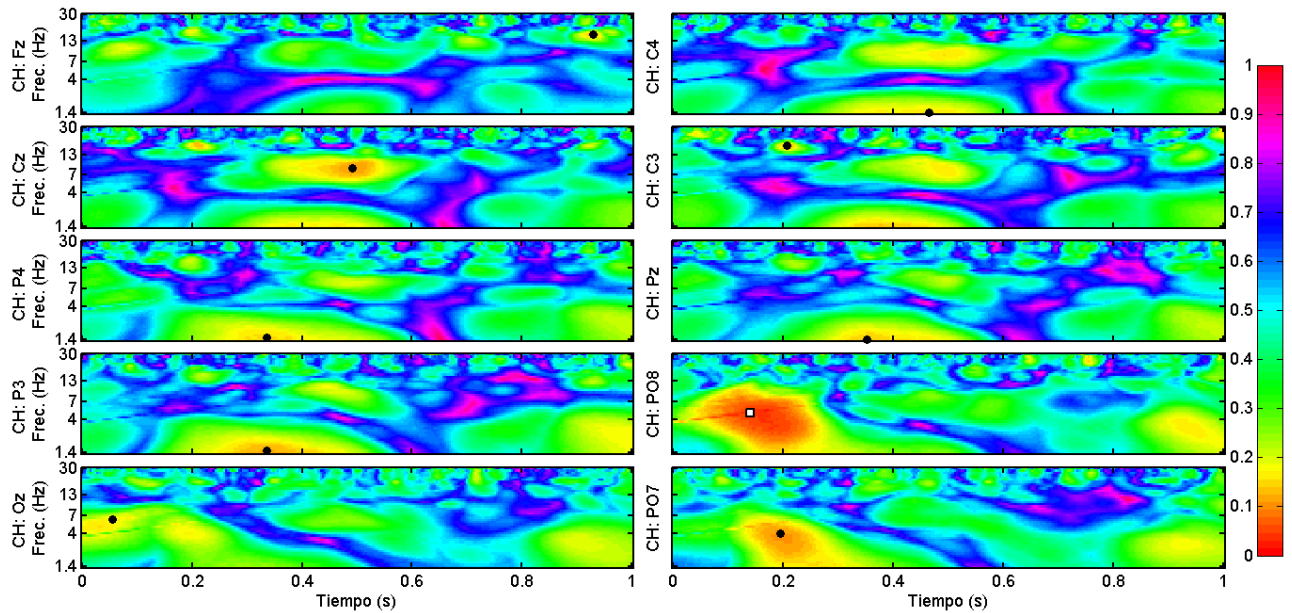
Escalogramas-H de épocas de filas del sujeto IZH.



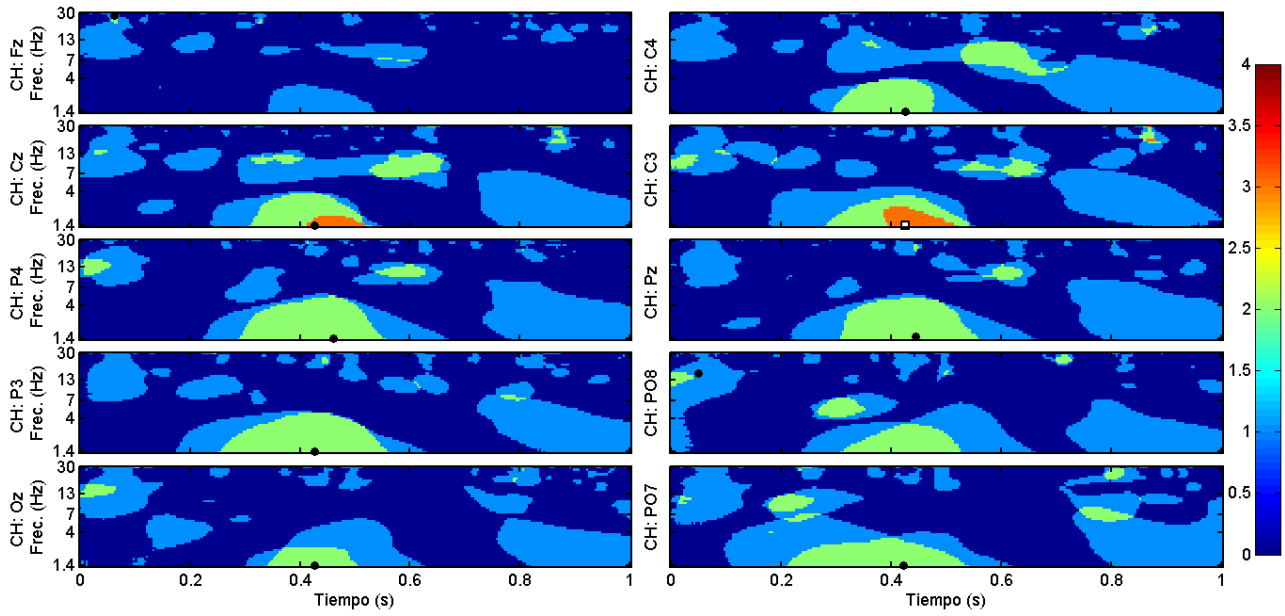
Escalogramas-p de épocas de filas del sujeto IZH.



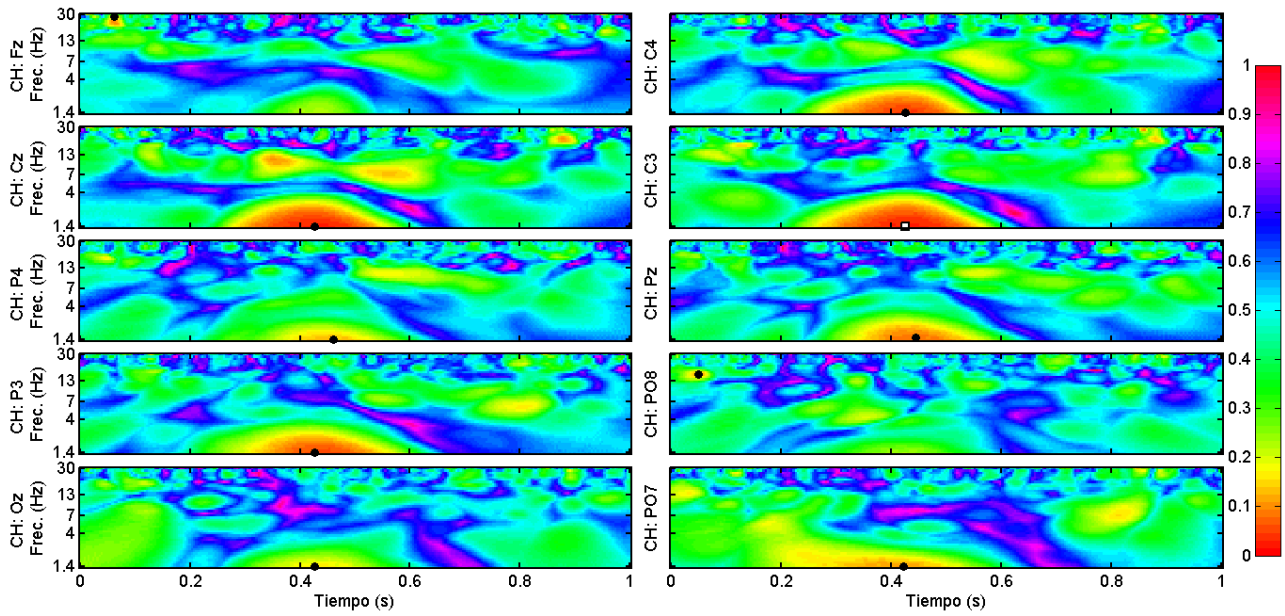
Escalogramas-H de épocas de columnas del sujeto GCE.



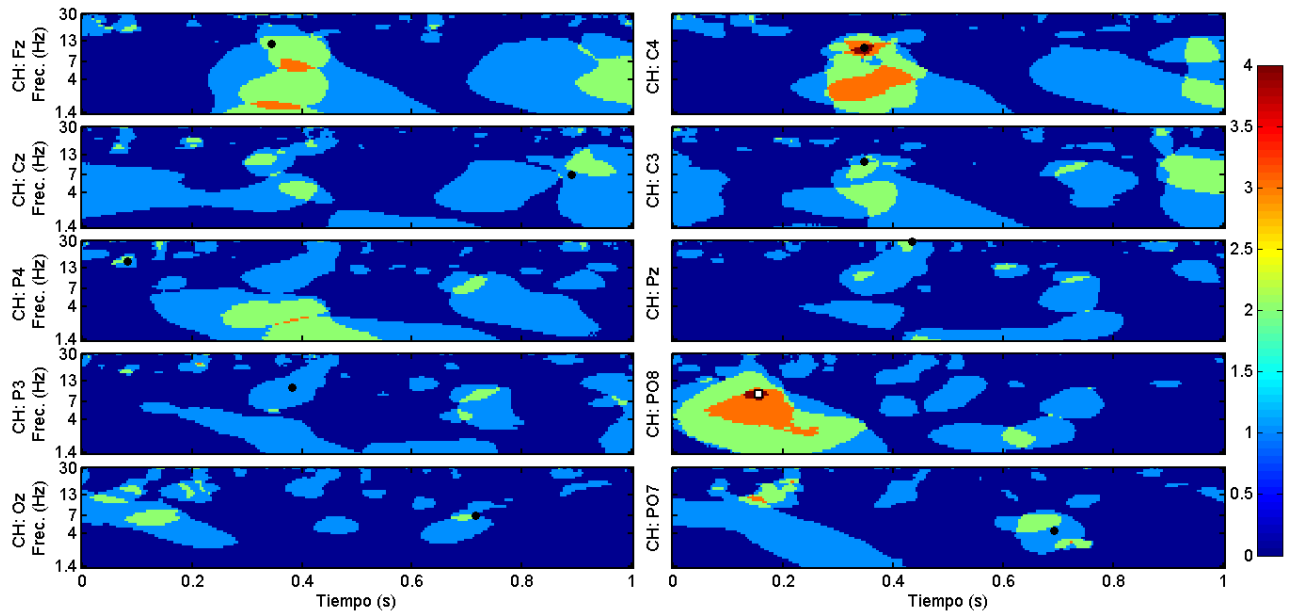
Escalogramas-p de épocas de columnas del sujeto GCE.



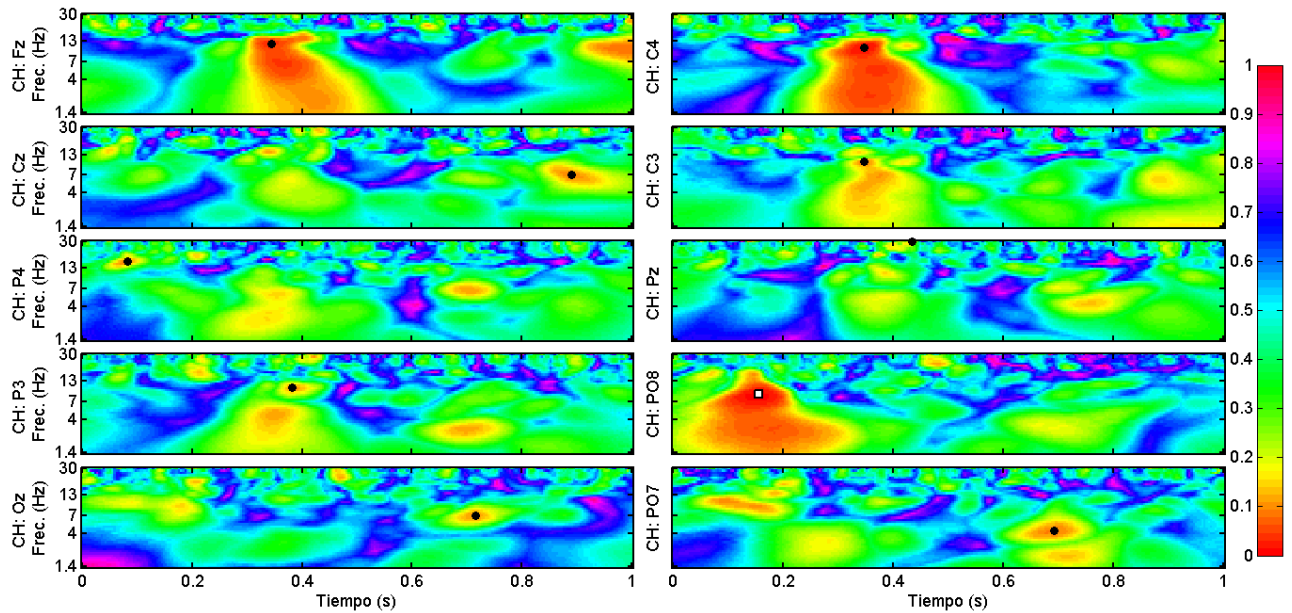
Escalogramas-H de épocas de filas del sujeto GCE.



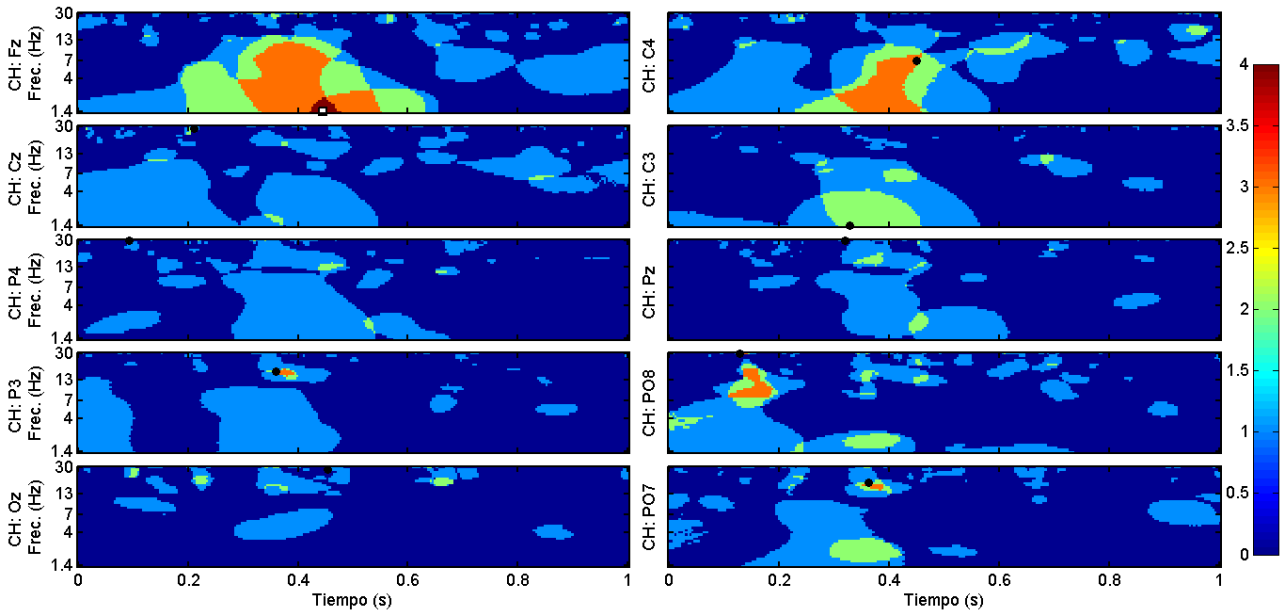
Escalogramas-p de épocas de filas del sujeto GCE.



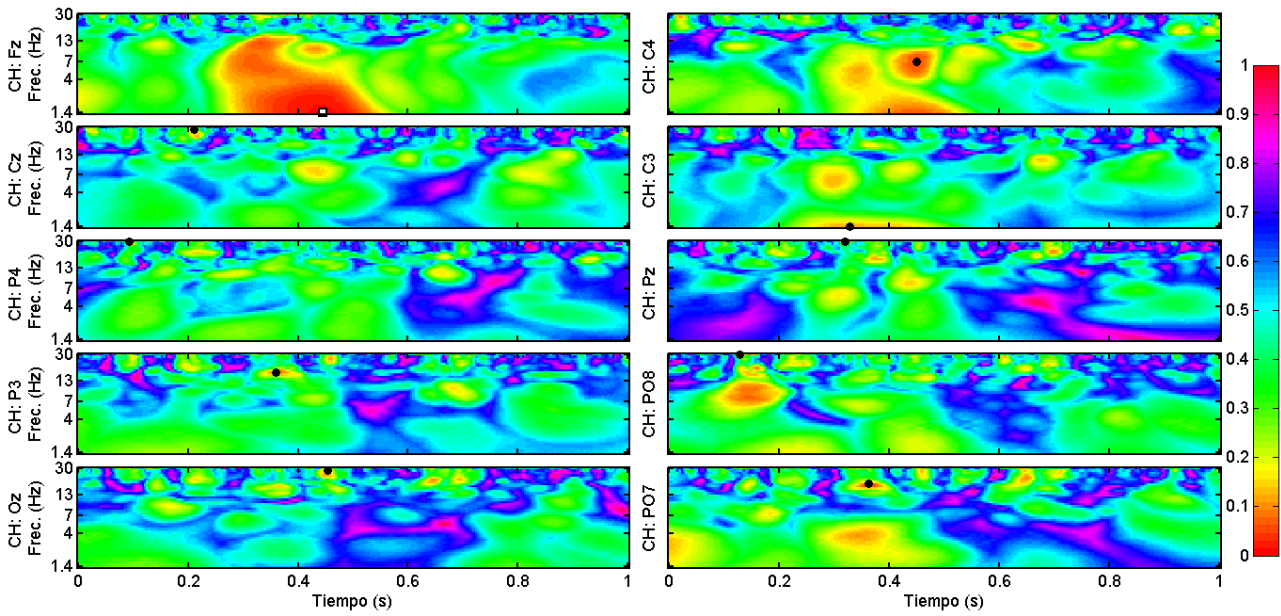
Escalogramas-H de épocas de columnas del sujeto PGA.



Escalogramas-p de épocas de columnas del sujeto PGA.



Escalogramas-H de épocas de filas del sujeto PGA.



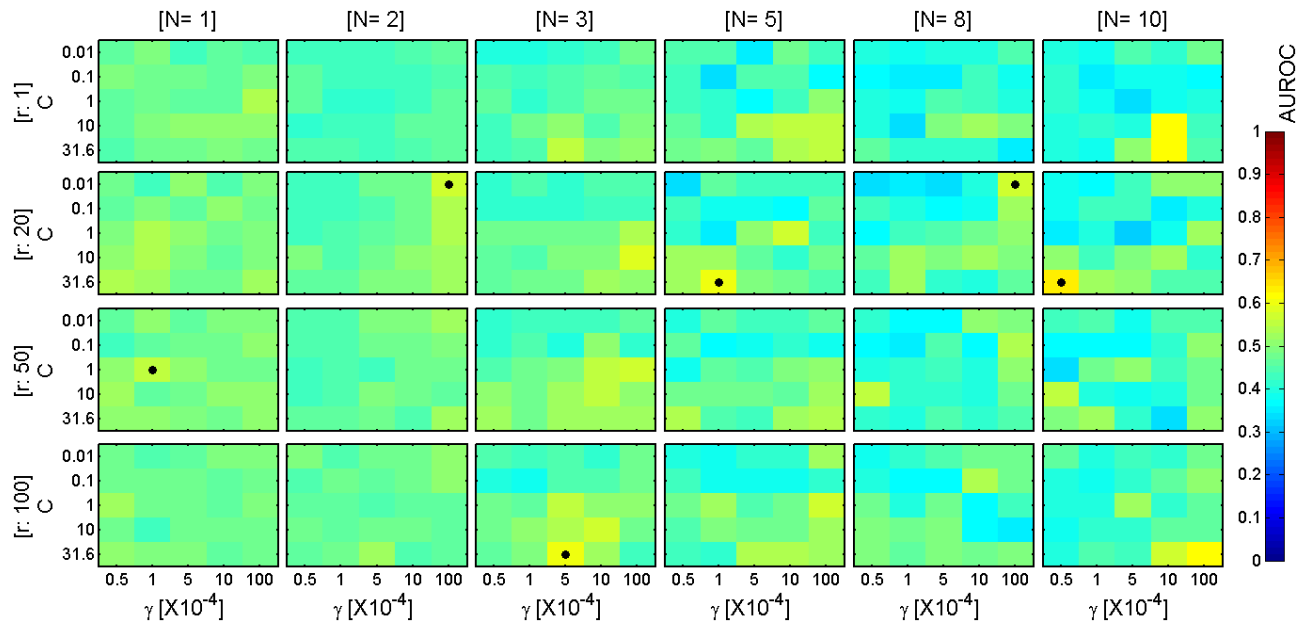
Escalogramas-p de épocas de filas del sujeto PGA.

Selección de rasgos e hiperparámetros óptimos por sujeto

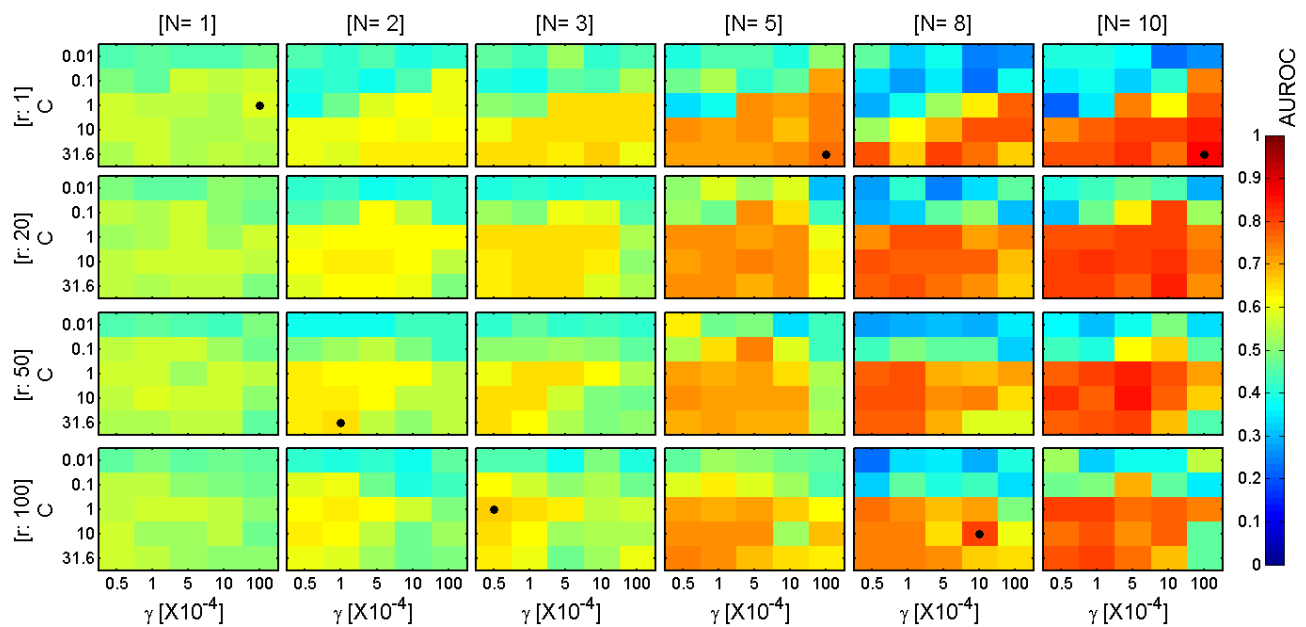
En este Anexo se ilustra el resultado del proceso de *Selección de rasgos e hiperparámetros óptimos* (sección 4.5) para cada uno de los 10 sujetos. Cada figura consiste en 24 gráficos, correspondientes a las combinaciones a evaluar de 6 valores de N y 4 subconjuntos de rasgos. A su vez, cada gráfico consiste en una matriz de 25 elementos, que indican el valor de $\text{AUROC}_{\text{prom}}(C, \gamma, r, N)$, para cada combinación de 5 valores de C y 5 valores de γ , los hiperparámetros del clasificador SVM-RBF.

El valor del parámetro $\text{AUROC}_{\text{prom}}$, que toma valores de 0.0 a 1.0, está codificado en la escala de colores de la barra del extremo derecho en cada figura. Para cada sujeto se presenta una figura para épocas de filas y otra para épocas de columnas.

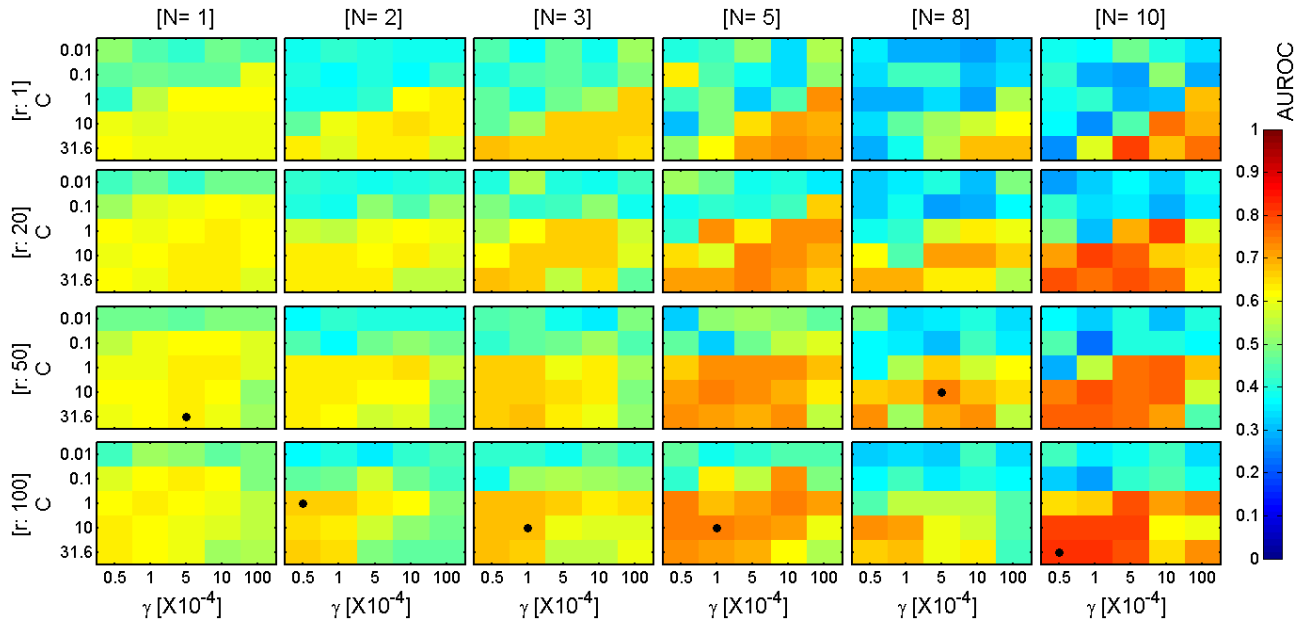
Así, para cada valor de N se elige una única combinación que maximiza el valor de $\text{AUROC}_{\text{prom}}$: $(C_{\text{opt}}, \gamma_{\text{opt}}, r_{\text{opt}})$, como se explica a detalle en la sección (4.5). Estas combinaciones óptimas, que se indican en las figuras con un círculo de color negro para cada valor de N (columnas), se usan para entrenar los *clasificadores óptimos* ($\text{SVM}_{\text{opt}}(N)$) en la sección 4.6.



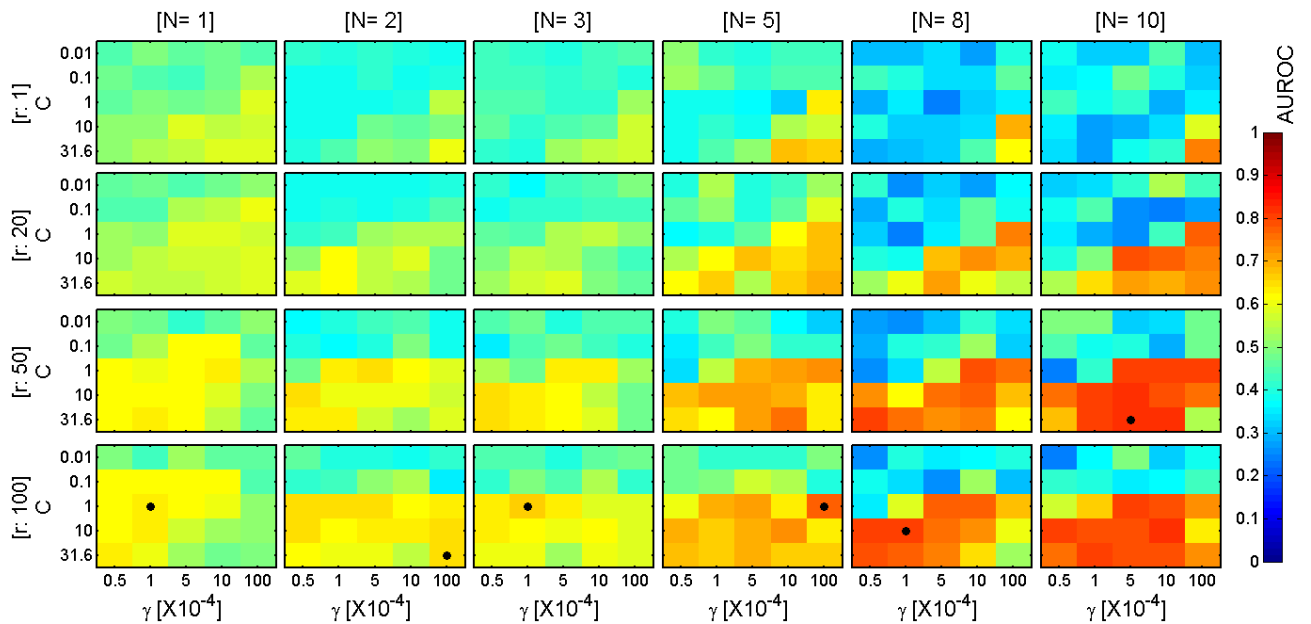
Selección de rasgos e hiperparámetros óptimos del sujeto DLP. Épocas de filas.



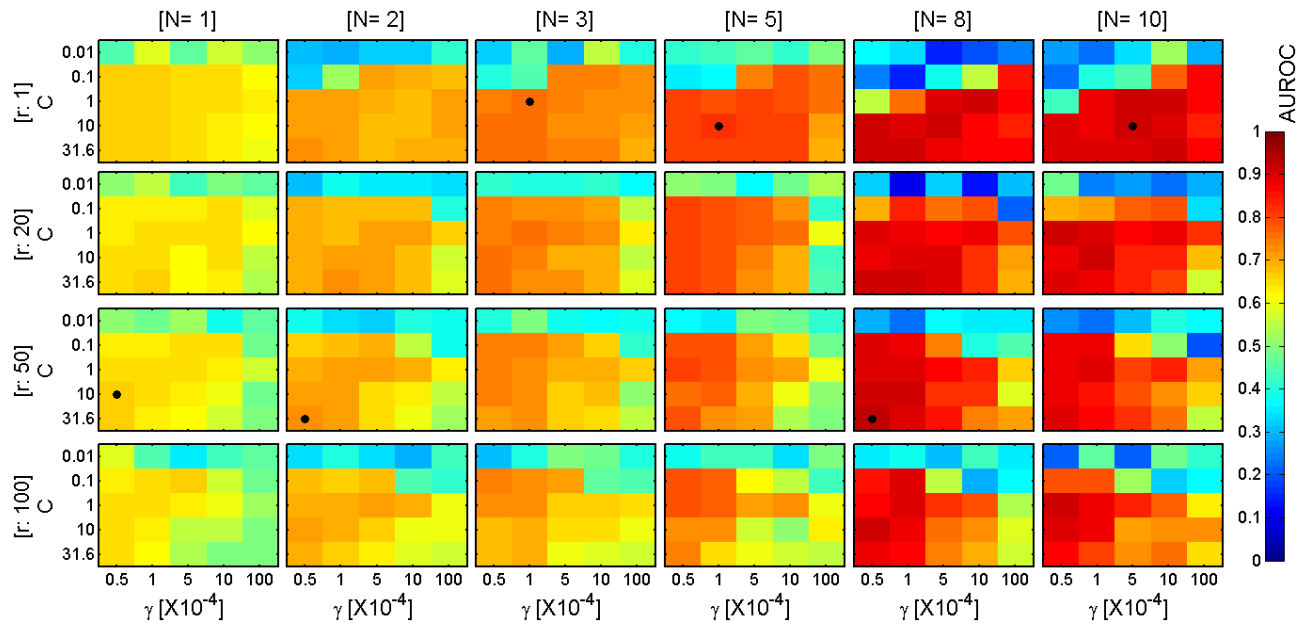
Selección de rasgos e hiperparámetros óptimos del sujeto DLP. Épocas de columnas.



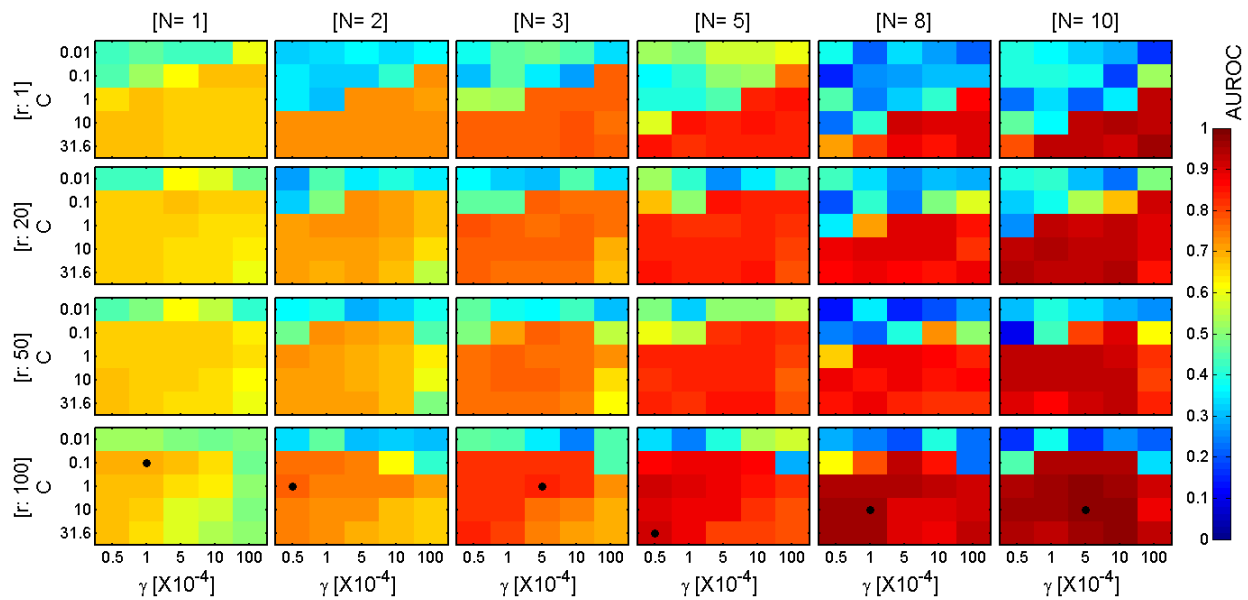
Selección de rasgos e hiperparámetros óptimos del sujeto DMA. Épocas de filas.



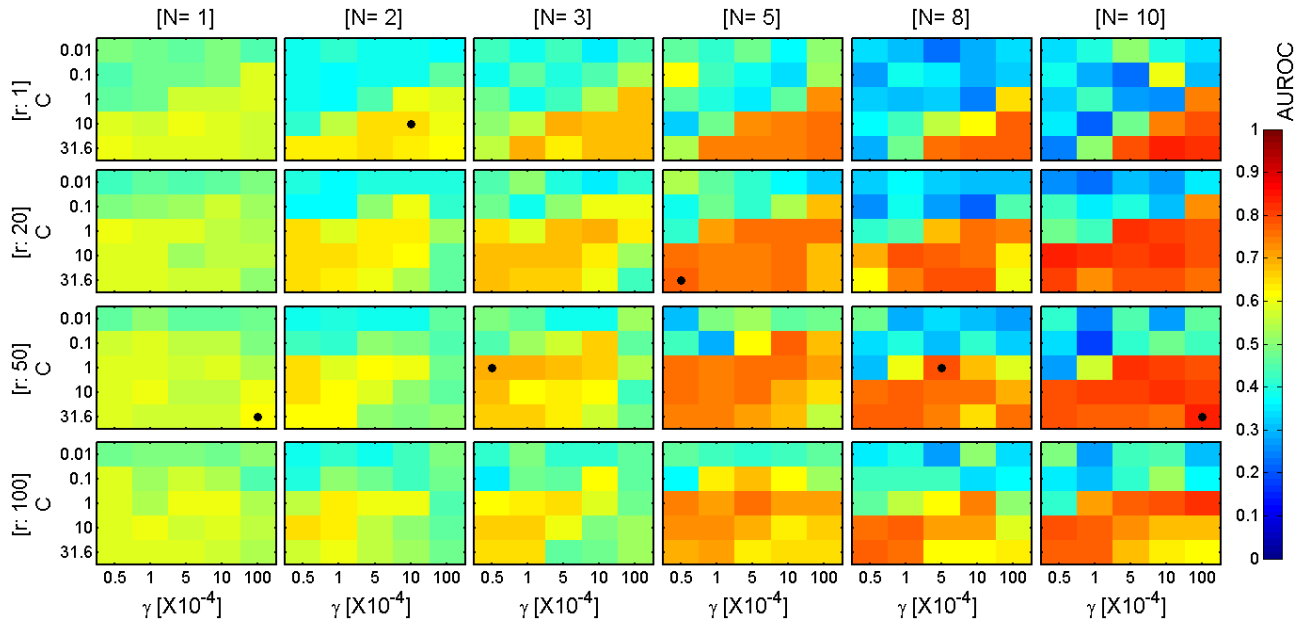
Selección de rasgos e hiperparámetros óptimos del sujeto DMA. Épocas de columnas.



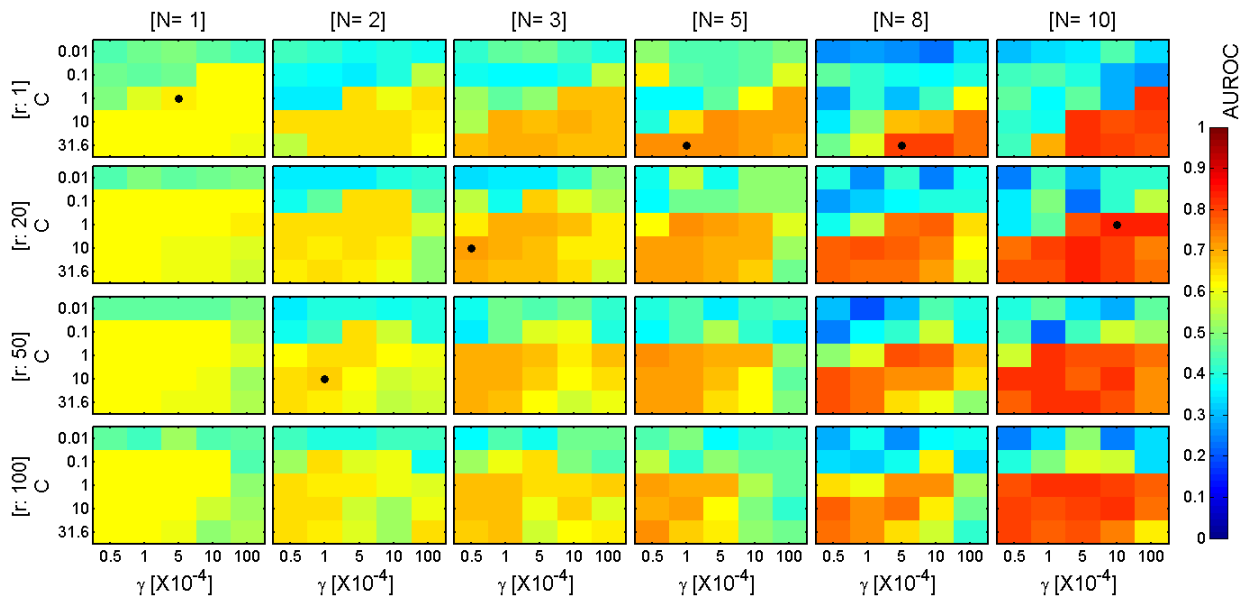
Selección de rasgos e hiperparámetros óptimos del sujeto ASG. Épocas de filas.



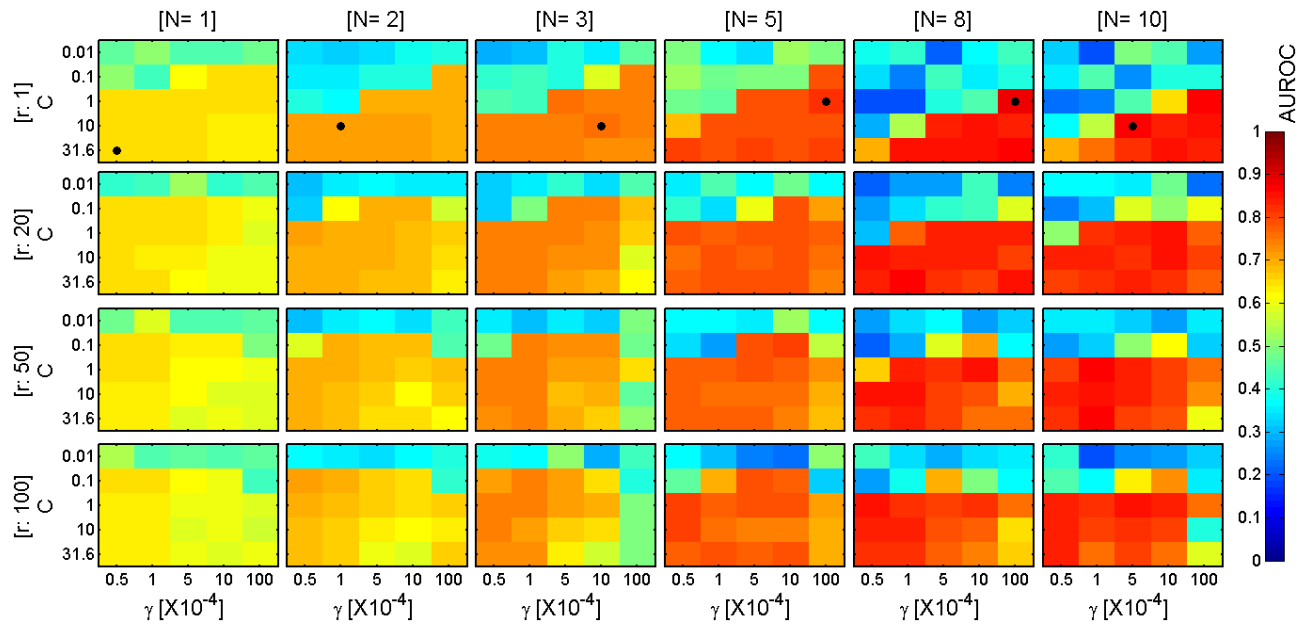
Selección de rasgos e hiperparámetros óptimos del sujeto ASG. Épocas de columnas.



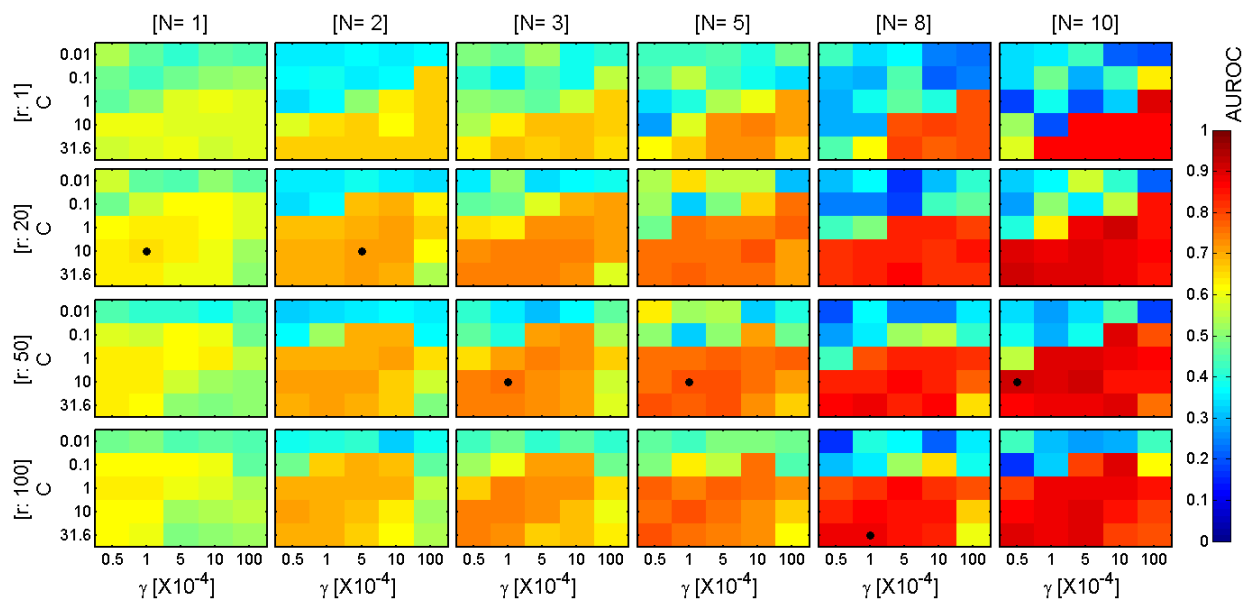
Selección de rasgos e hiperparámetros óptimos del sujeto CLL. Épocas de filas.



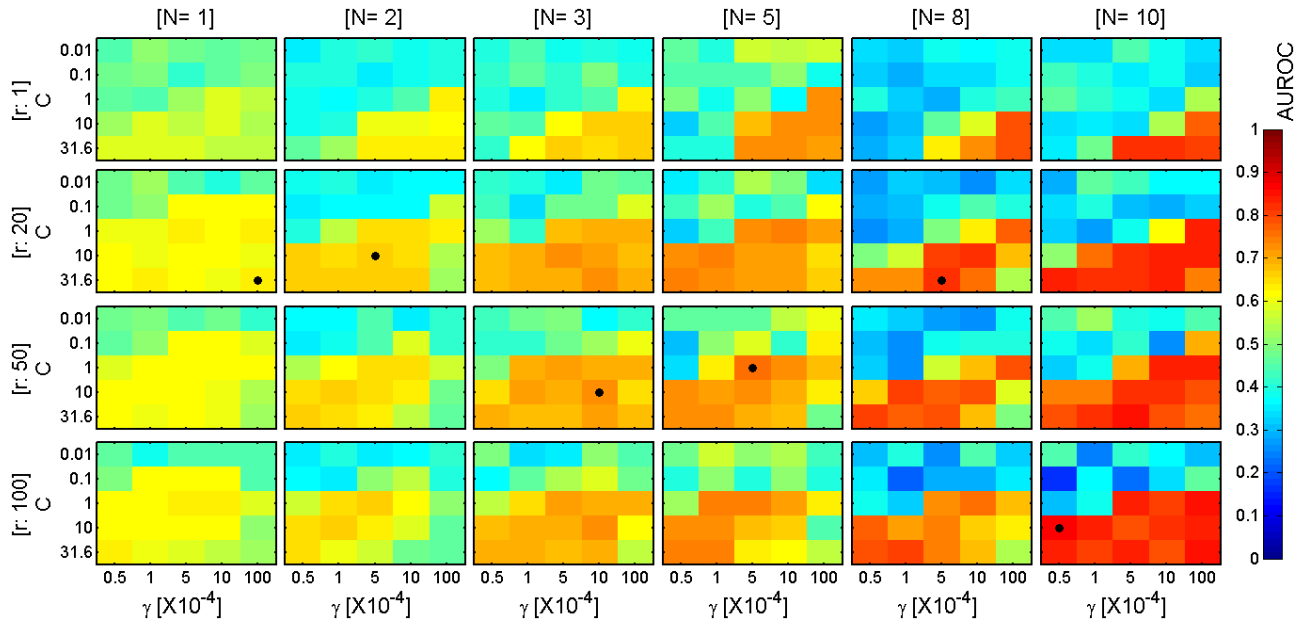
Selección de rasgos e hiperparámetros óptimos del sujeto CLL. Épocas de columnas.



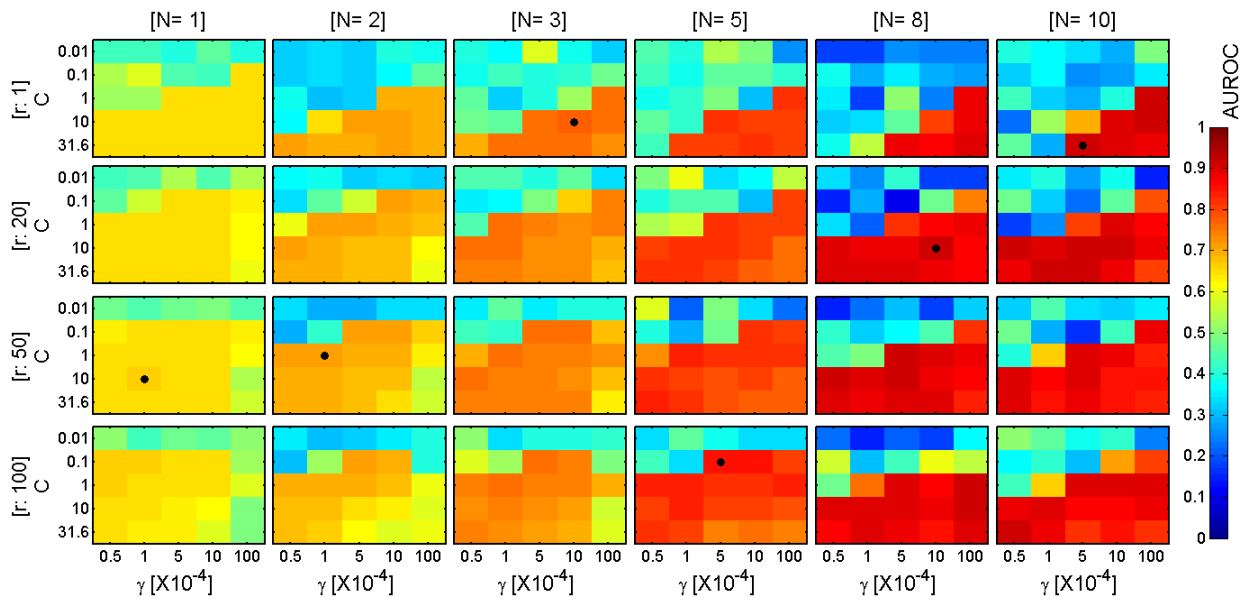
Selección de rasgos e hiperparámetros óptimos del sujeto LAG. Épocas de filas.



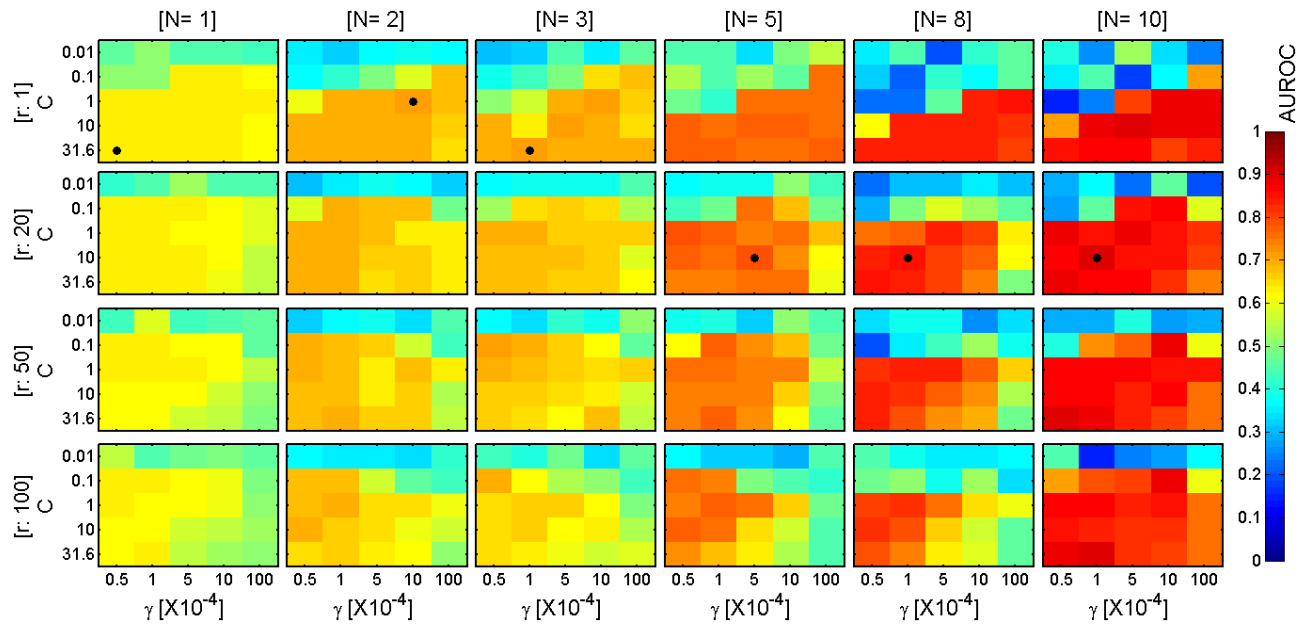
Selección de rasgos e hiperparámetros óptimos del sujeto LAG. Épocas de columnas.



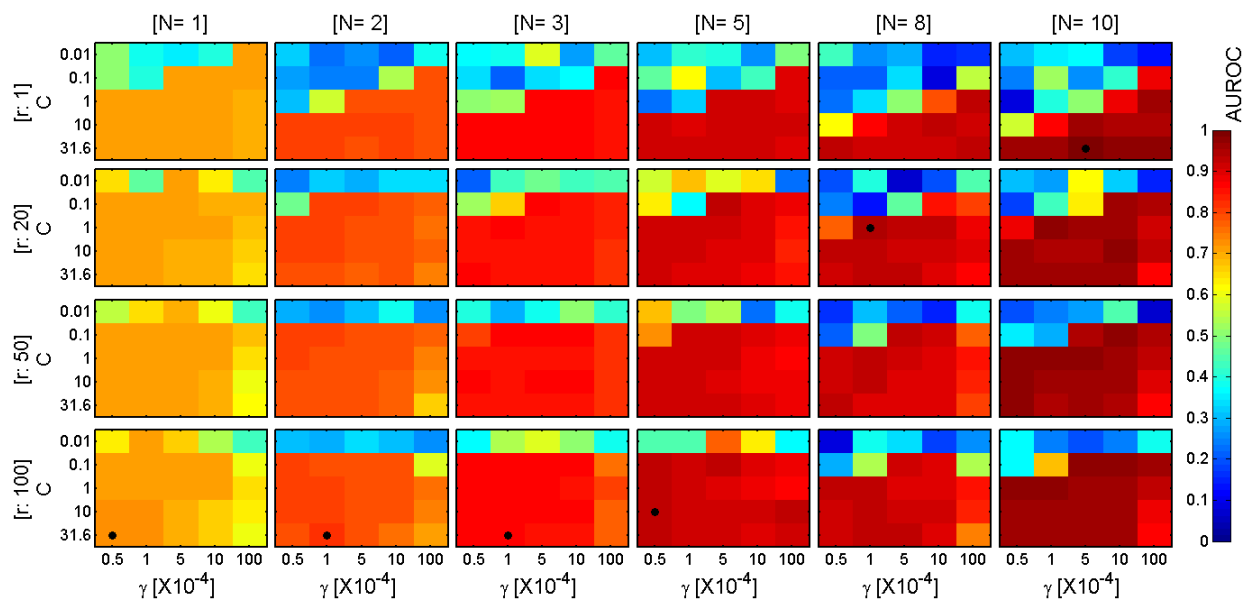
Selección de rasgos e hiperparámetros óptimos del sujeto LGP. Épocas de filas.



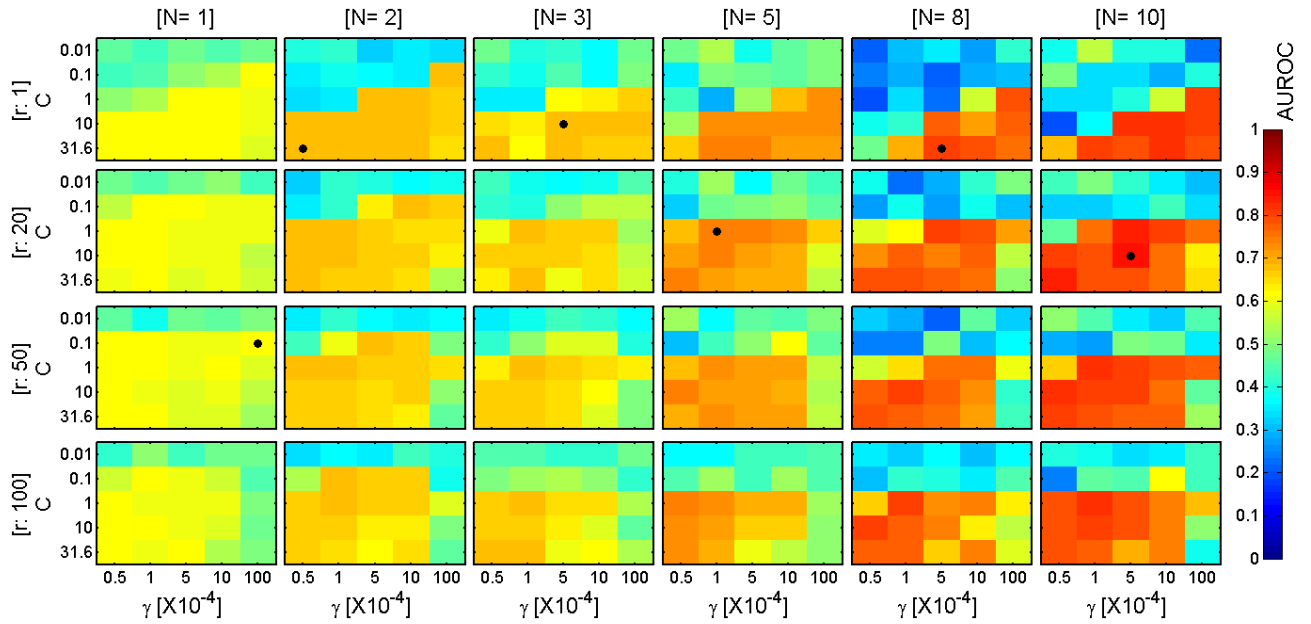
Selección de rasgos e hiperparámetros óptimos del sujeto LGP. Épocas de columnas.



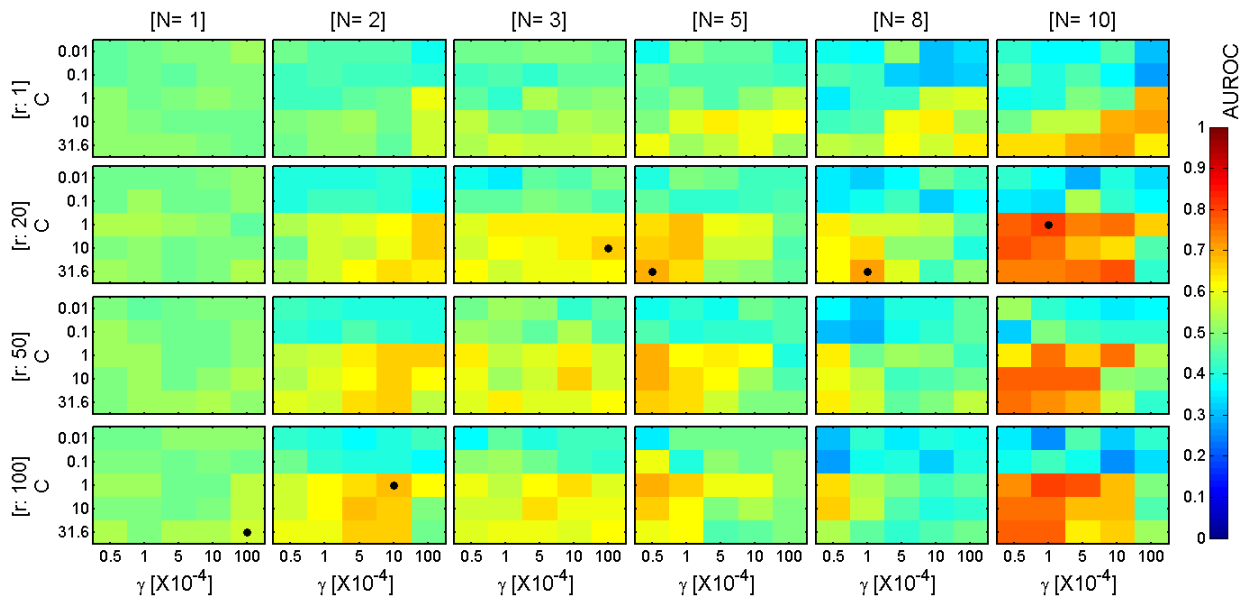
Selección de rasgos e hiperparámetros óptimos del sujeto LAC. Épocas de filas.



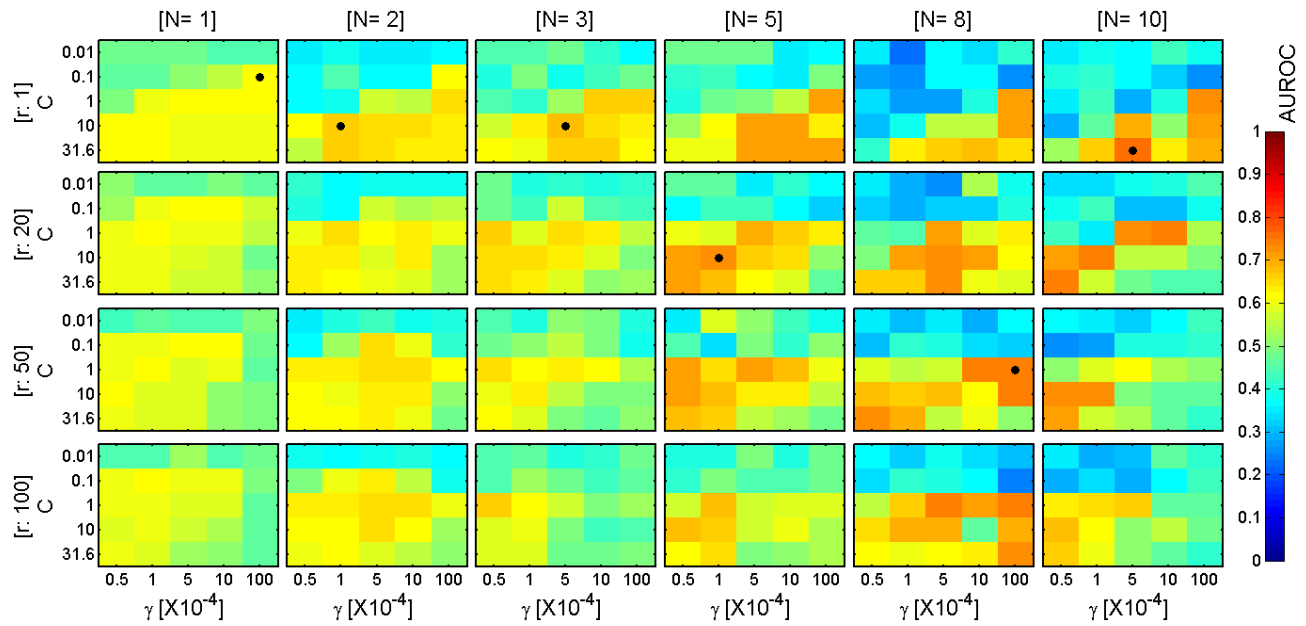
Selección de rasgos e hiperparámetros óptimos del sujeto LAC. Épocas de columnas.



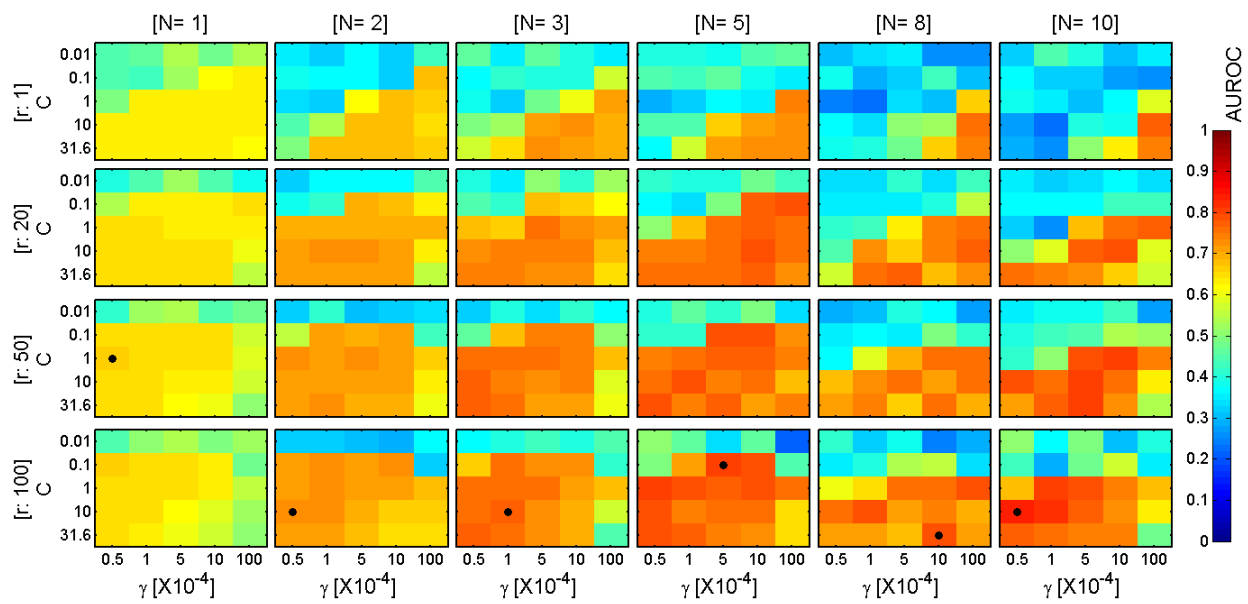
Selección de rasgos e hiperparámetros óptimos del sujeto IZH. Épocas de filas.



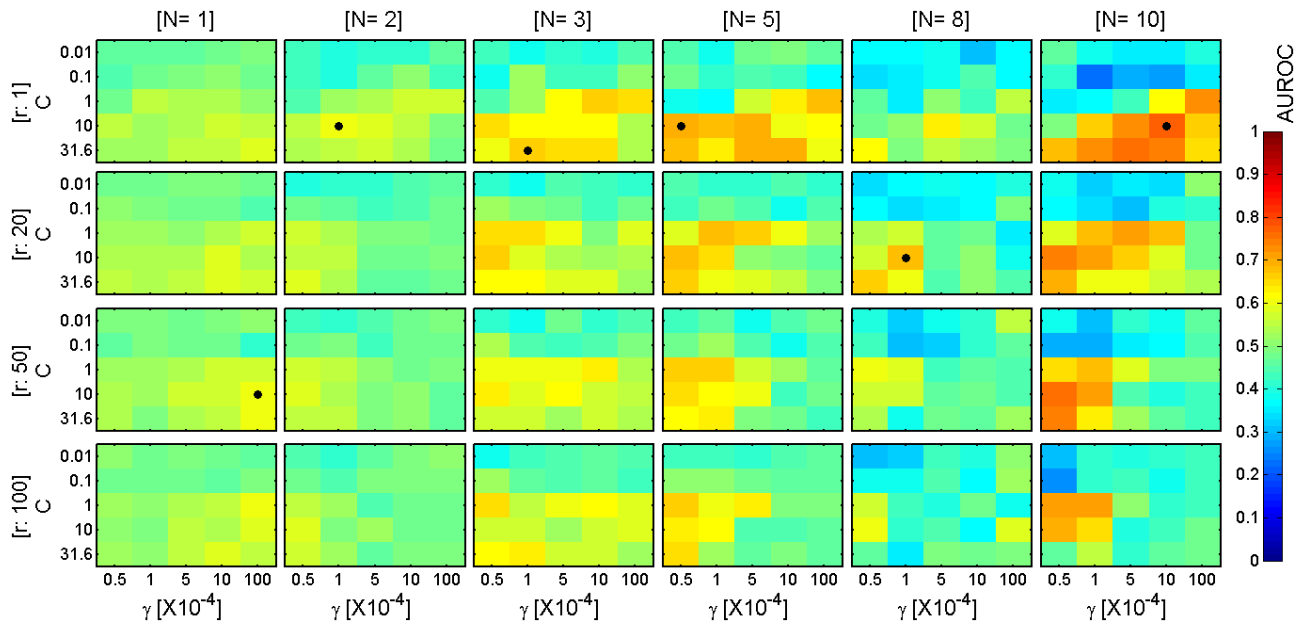
Selección de rasgos e hiperparámetros óptimos del sujeto IZH. Épocas de columnas.



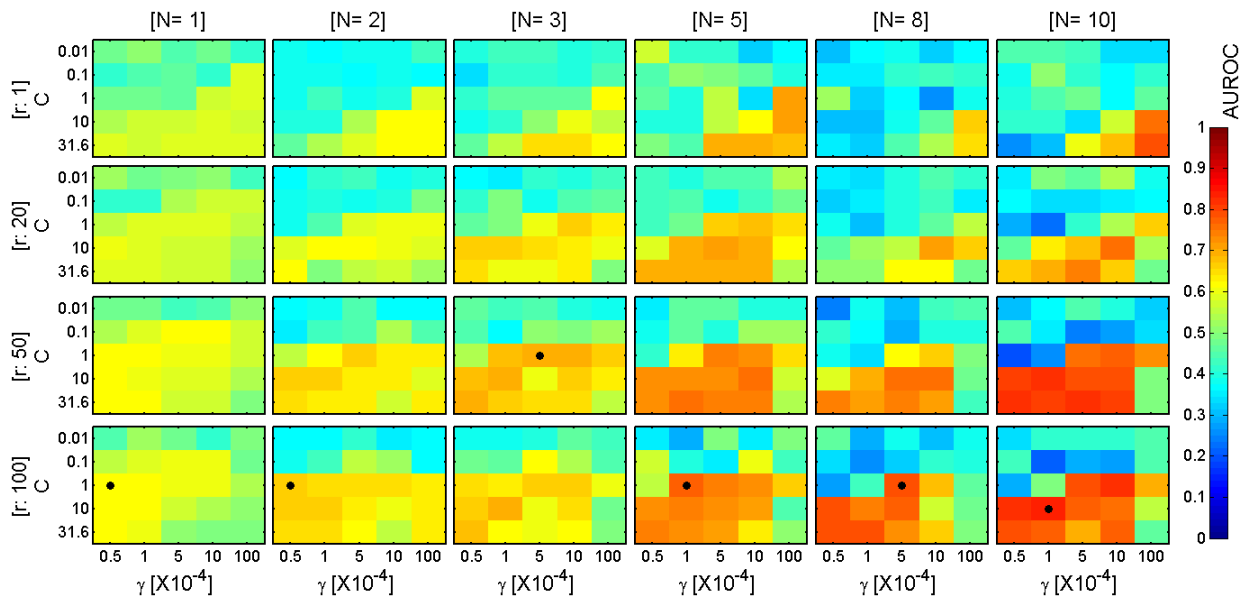
Selección de rasgos e hiperparámetros óptimos del sujeto GCE. Épocas de filas.



Selección de rasgos e hiperparámetros óptimos del sujeto GCE. Épocas de columnas.



Selección de rasgos e hiperparámetros óptimos del sujeto PGA. Épocas de filas.



Selección de rasgos e hiperparámetros óptimos del sujeto PGA. Épocas de columnas.

Bibliografía

- [1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-Computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, pp. 767–791, Marzo 2002.
- [2] J. J. Vidal, "Toward Direct Brain-Computer Communication," tech. rep., Brain Research Institute, University of California, Los Angeles, California, 1973.
- [3] E. Donchin and Y. Arbel, "P300 Based Brain Computer Interfaces: A Progress Report," in *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience* (D. Schmorow, I. Estabrooke, and M. Grootjen, eds.), vol. 5638 of *Lecture Notes in Computer Science*, pp. 724–731, Springer Berlin/Heidelberg, 2009.
- [4] C. Guger, S. Daban, E. Sellers, C. Holzner, G. Krausz, R. Carabona, F. Gramatica, and G. Edlinger, "How Many People Are Able to Control a P300-based Brain-Computer Interface (BCI)?," *Neuroscience Letters*, vol. 462, pp. 94–98, 2009.
- [5] T. Hinterberger, S. Schmidt, N. Neumann, J. Mellinger, B. Blankertz, G. Curio, and N. Birbaumer, "Brain-Computer Communication and Slow Cortical Potentials," *Transactions on Biomedical Engineering*, vol. 51, pp. 1011–1018, Junio 2004.
- [6] L. A. Farwell and E. Donchin, "Talking Off the Top of Your Head: Toward a Mental Prosthesis Utilizing Event-Related Brain Potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, pp. 510–523, 1988.
- [7] J. Polich, "Updating P300: An Integrative Theory of P3a and P3b," *Clinical Neurophysiology*, vol. 118, pp. 2128–2148, Junio 2007.
- [8] C. S. Syan and R. E. Harnarinesingh, "Comparison of Pre-Processing and Classification Techniques for Single-Trial and Multi-Trial P300-Based Brain Computer Interfaces," *American Journal of Applied Sciences*, vol. 7, no. 9, pp. 1219–1225, 2010.
- [9] N. Manyakov, N. Chumerin, A. Combaz, and M. V. Hulle, "On the Selection of Time Interval and Frequency Range of EEG Signal Preprocessing for P300 Brain-Computer

- Interfacing," *XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010*, vol. 29, pp. 57–60, 2010.
- [10] Z. Amini, V. Abootalebi, and M. T. Sadeghi, "A Comparative Study of Feature Extraction Methods in P300 Detection," in *Proceedings of the 17th Iranian Conference of Biomedical Engineering*, IEEE, Noviembre 2010.
- [11] T. Demiralp, A. Ademoglu, M. Schumann, C. Basar-Eroglu, and E. Basar, "Detection of P300 Waves in Single Trials by the Wavelet Transform (WT)," *Brain and Language*, vol. 66, pp. 108–128, 1999.
- [12] V. Bostanov, "BCI Competition 2003-Data Sets Ib and Iib: Feature Extraction From Event-Related Brain Potentials With the Continuous Wavelet Transform and the t-Value Scalogram," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 1057–1061, June 2004.
- [13] S. Costagliola, B. D. Seno, and M. Matteucci, "Recognition and Classification of P300s in EEG Signals by Means of Feature Extraction Using Wavelet Decomposition," in *Proceedings of International Joint Conference on Neural Networks*, pp. 597–603, IEEE, Junio 2009.
- [14] A. Combaz, N. V. Manyakov, N. Chumerin, J. A. K. Suykens, and M. M. V. Hulle, "Feature Extraction and Classification of EEG Signals for Rapid P300 Mind Spelling," in *2009 International Conference on Machine Learning and Applications*, pp. 386–391, IEEE Computer Society, 2009.
- [15] Q. Xiaoyan, L. Douzhe, and D. Youer, "P300 Feature Extraction Based on Parametric Model and FastICA Algorithm," in *Fifth International Conference on Natural Computation*, IEEE, 2009.
- [16] T. Lan, C. Huang, and D. Erdogmus, "A Comparison of Temporal Windowing Schemes for Single-trial ERP Detection," in *Proceedings of the 4th International IEEE EMBS Conference on Neural Engineering*, pp. 331–334, IEEE, 2009.
- [17] Y. Gucluturk, U. Guclu, and A. Samraj, "An Online Single Trial Analysis of the P300 Event Related Potential for the Disabled," in *2010 IEEE 26th Convention of Electrical and Electronics Engineers in Israel*, pp. 338–341, IEEE, 2010.
- [18] J.-A. Guan, "Evoked Potentials Estimation in Brain-Computer Interface Using Support Vector Machine," in *RSKT 2006*, vol. 4062 of *LNAI*, pp. 701–706, Springer-Verlag Berlin Heidelberg, 2006.

-
- [19] C. Wang, C. Guan, and H. Zhang, "P300 Brain-Computer Interface Design for Communication and Control Applications," in *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, IEEE, 2005.
- [20] A. Mazaheri and T. W. Picton, "EEG Spectral Dynamics During Discrimination of Auditory and Visual Targets," *Cognitive Brain Research*, vol. 24, pp. 81–96, Febrero 2005.
- [21] C. S. Hermann and R. T. Knight, "Mechanisms of Human Attention: Event-Related Potentials and Oscillations," *Neuroscience and Behavioural Reviews*, vol. 25, pp. 465–476, 2001.
- [22] I. Gurtubay, M. Alegrea, A. Labargab, A. Malandab, J. Iriarte, and J. Artiedaa, "Gamma Band Activity in an Auditory Oddball Paradigm Studied with the Wavelet Transform," *Clinical Neurophysiology*, vol. 112, pp. 1219–1228, 2001.
- [23] L. Marshall, M. Molle, and P. Bartsch, "Event-related gamma band activity during passive and active oddball tasks," *Neuroreport*, vol. 7, June 1996.
- [24] R. J. Barry, "Evoked Activity and EEG Phase Resetting in the Genesis of Auditory Go/NoGo ERPs," *Biological Psychology*, vol. 80, pp. 292–299, 2009.
- [25] B. W. Jervis, M. J. Nichols, T. E. Johnson, E. Allen, and N. R. Hudson, "A Fundamental Investigation of the Composition of Auditory Evoked Potentials," *IEEE Transactions on Biomedical Engineering*, vol. 30, pp. 43–50, Enero 1983.
- [26] P. Sauseng, W. Klimesch, W. R. Gruber, S. Hanslmayr, R. Freunberger, and M. Doppelmayr, "Are Event-related Potential Components Generated By Phase Resetting Of Brain Oscillations? A Critical Discussion," *Neuroscience*, vol. 146, pp. 1435–1444, 2007.
- [27] V. Abootalebi, M. H. Moradi, and M. A. Khalilzadeh, "A Comparison of Methods for ERP Assessment in a P300-based GKT," *International Journal of Psychophysiology*, vol. 62, pp. 309–320, 2006.
- [28] D. Ming, X. Ana, Y. Xi, Y. Hu, B. Wan, H. Qi, L. Cheng, and Z. Xuea, "Time-Locked and Phase-Locked Features of P300 Event-Related Potentials (ERPs) for Brain-Computer Interface Speller," *Biomedical Signal Processing and Control*, vol. 5, pp. 243–251, Agosto 2010.
- [29] P. Wang, J.-Z. Shen, and J.-H. Shi, "P300 Detection Algorithm Based on Fisher Distance," *Modern Education and Computer Science*, vol. 2, pp. 9–17, 2010.
- [30] V. S. Ramachadran, ed., *Encyclopedia of the Human Brain*. Elsevier Science, 2002.
-

- [31] M. E. Valentinuzzi, *Understanding the human machine. A Primer for Bioengineering*, vol. 4 of *Series on Bioengineering and Biomedical Engineering*. World Scientific Publishing, 2004.
- [32] J. B. Fisch, *Fisch and Spehlman's EEG Primer*. Elsevier Science, third ed., 1999.
- [33] J. D. Kropotov, *Quantitative EEG, Event-Related Potentials and Neurotherapy*. Elsevier, first ed., 2009.
- [34] G. Pfurtscheller and F. H. L. D. Silva, eds., *Event-Related Desynchronization*, vol. 6 of *Handbook of Electroencephalography and Clinical Neurophysiology. Serie Revisada*. Elsevier Science, primera ed., 1999.
- [35] P. Sauseng and W. Klimesch, "What Does Phase Information of Oscillatory Brain Activity Tell us About Cognitive Processes?," *Neuroscience and Behavioural Reviews*, vol. 32, pp. 1001–1013, 2008.
- [36] S. Tong and N. V. Thakor, eds., *Quantitative EEG Analysis Methods and Clinical Applications*. Engineering in Medicine and Biology, Artech House, 2009.
- [37] N. Naseer and K.-S. Hong, "fNIRS-based Brain-Computer Interfaces: a Review," *Frontiers in Human Neuroscience*, vol. 9, 2015.
- [38] S. Amiri, R. Fazel-Rezai, and V. Asadpour, "A Review of Hybrid Brain-Computer Interface Systems," *Advances in Human Computer Interaction*, vol. 2013, 2013.
- [39] BCI2000, "P3SpellerScreen." <http://bci2000.org/wiki/index.php/File:P3SpellerScreen.PNG>, April 2012. BCI2000 Wiki.
- [40] G. Schalk, D. J. McFarland, T. Hinterberg, N. Birbaumer, and J. R. Wolpaw, "BCI2000. A General-Purpose Brain-Computer Interface (BCI) System," *IEEE Transactions on Biomedical Engineering*, vol. 51, June 2004.
- [41] J. D. Bayliss, "Use of the Evoked Potential P3 Component for Control in a Virtual Apartment," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, pp. 113–116, Junio 2003.
- [42] A. Shoeb and G. Clifford, *Biomedical Signal and Image Processing Spring 2005*, ch. 16 - Wavelets; Multiscale Activity in Physiological Signals, pp. 1–29. Harvard-MIT Division of Health Sciences and Technology, 2006.
- [43] C. Liner, "An Overview of Wavelet Transform Concepts and Applications." University of Houston, Febrero 2010.

-
- [44] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, eds., *Feature Extraction. Foundations and Applications*, vol. 207 of *Studies in Fuzziness and Soft Computing*. Springer-Verlag Berlin Heidelberg, 2006.
- [45] T. Navin, O. Chapelle, and B. Scholkopf, *Feature Extraction. Foundations and Applications*, vol. 207 of *Studies in Fuzziness and Soft Computing*, ch. 20, pp. 439–445. Springer-Verlag Berlin Heidelberg, 2003.
- [46] Y.-W. Chen and C.-J. Lin, *Feature Extraction. Foundations and Applications*, vol. 207 of *Studies in Fuzziness and Soft Computing*, ch. 12, pp. 315–324. Springer-Verlag Berlin Heidelberg, 2006.
- [47] G. Dreyfus and I. Guyon, *Feature Extraction. Foundations and Applications*, vol. 207 of *Studies in Fuzziness and Soft Computing*, ch. Assessment Methods, pp. 65–88. Springer-Verlag Berlin Heidelberg, 2006.
- [48] A. Ben-Hur and J. Weston, “A User’s Guide to Support Vector Machines,” in *Data Mining Techniques for the Life Sciences* (O. Carugo and F. Eisenhaber, eds.), vol. 609 of *Methods in Molecular Biology*, ch. 13, pp. 223–239, Humana Press, 2010.
- [49] C. M. Bishop, *Pattern Recognition and Machine Learning*. Information Science and Statistics, Springer-Verlag New York, 2006.
- [50] B. Scholkopf, K. Tsuda, and J.-P. Vert, *Kernel Methods in Computational Biology*, ch. A Primer on Kernel Methods, pp. 35–70. Computational Molecular Biology, MIT Press, 2004.
- [51] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A Training Algorithm for Optimal Margin Classifiers,” in *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, 1992.
- [52] C. Cortes and V. Vapnik, “Support-Vector Networks,” *Machine Learning*, vol. 20, no. 1995, pp. 273–297, 1995.
- [53] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Elsevier Academic Press, second ed., 2003.
- [54] L. Hamel, *Knowledge Discovery with Support Vector Machines*. Wiley Series on Methods and Applications in Data Mining, John Wiley and Sons, 2009.
- [55] T. Fawcett, “An Introduction to ROC Analysis,” *Pattern Recognition Letters*, vol. 27, pp. 861–874, 2006.
-

- [56] K. Hajian-Tilaki, "Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation," *Caspian Journal of Internal Medicine*, vol. 4, no. 2, pp. 627–635, 2013.
- [57] C. Ledesma-Ramirez, E. Bojorges-Valdez, O. Yanez-Suarez, C. Saavedra, and L. Bougrain, "An Open-Access P300 Speller Database," in *Fourth International Brain-Computer Interface Meeting*, May 2010.
- [58] A. Popov and M. Zhukov, "Computation of continuous wavelet transform of discrete signals with adapted mother functions," in *Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2010* (R. S. R. and, ed.), vol. 7502, SPIE- The International Society for Optical Engineering, 2010.
- [59] C.-C. Chang and C.-J. Lin, "LIBSVM : A Library for Support Vector Machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 27, pp. 1–27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [60] R. Fazel-Rezai and W. Ahmad, *Recent Advances in Brain-Computer Interface Systems*, ch. P300 Brain-Computer Interface Paradigm Design. Recent Advances in Human-Computer Interaction, InTech, 2011.
- [61] P. Brunner and G. Schalk, "Toward a gaze Independent matrix speller brain-computer interface," *Clinical Neurophysiology*, vol. 122, pp. 1063–1064, 2011.
- [62] R. Fazel-Rezai, B. Z. Allison, C. Guger, E. Sellers, S. C. Kleih, and A. Kubler, "P300 Brain Computer Interface: current challenges and emerging trends," *frontiers in Neuroengineering*, vol. 5, July 2012.
- [63] J. P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Human Brain Mapping*, vol. 8, pp. 194–208, 1999.
- [64] S. Hanslmayr, W. Klimesch, P. Sauseng, W. Gruber, M. Doppelmayr, R. Freunberger, T. Pecherstorfer, and N. Birbaumer, "Alpha Phase Reset Contributes to the Generation of ERPs," *Cerebral Cortex*, vol. 17, pp. 1–8, 2006.
- [65] E. Gysels and P. Celka, "Phase Synchronization for the Recognition of Mental Tasks in a Brain Computer Interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 12, pp. 406–415, Diciembre 2004.
- [66] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene Selection for Cancer Classification using Support Vector Machines," *Machine Learning*, no. 46, pp. 389–422, 2002.

- [67] Y. Tu, G. Huang, Y. S. Hung, L. Hu, Y. Hu, and Z. Zhang, "Single-trial Detection of Visual Evoked Potentials by Common Spatial Patterns and Wavelet Filtering for Brain-computer Interface," in *35th Annual International Conference of the IEEE EMBS*, IEEE, July 2013.
- [68] G. Pires, U. Nunes, and M. Castelo-Branco, "P300 Spatial Filtering and Coherence-based Channel Selection," in *Proceedings of the 4th International IEEE EMBS Conference on Neural Engineering*, IEEE, April-May 2009.