



Universidad
Zaragoza



Proyecto Final de Carrera
Ingeniería de Telecomunicaciones
Curso 2010-2011

Diseño de una Base Normativa de EEG Basal en Adultos

Mónica Aguilar Herrero

Marzo de 2011

Director: Javier Mínguez Zafra

Departamento de Informática e Ingeniería de Sistemas
Centro Politécnico Superior
Universidad de Zaragoza

Si el cerebro fuese tan sencillo para que pudiéramos comprenderlo
nosotros seríamos tan simples que no lograríamos entenderlo.

Albert Einstein

Diseño de una Base Normativa de EEG Basal en Adultos

RESUMEN

El electroencefalograma constituye una de las técnicas de adquisición no invasivas más ampliamente utilizadas para el registro de actividad cerebral. El uso de Bases Normativas de EEG resulta de gran interés para determinar un patrón de referencia en cuanto a señal EEG se refiere.

En el presente proyecto se diseña, desarrolla y valida una Base Normativa de EEG basal en adultos, con el objetivo de automatizar el diagnóstico, tratamiento y evaluación de las enfermedades neurológicas.

Actualmente la problemática existente en el desarrollo de una Base Normativa de EEG viene determinada por la alta variabilidad interpersonal de la señal EEG. Este aspecto ha sido estudiado mediante la relación de esta variabilidad con factores fisiológicos llevando a cabo una estratificación de sujetos clasificados por dichos factores. Esta técnica ha resultado útil en el caso de Bases Normativas de adultos, pero no en el caso de Bases Normativas infantiles. Este proyecto plantea una alternativa mediante la estratificación de la Base Normativa de forma explícita sobre los cambios significativos de la propia señal EEG. Además, se han elaborado unas herramientas matemáticas que permiten aplicar la Base Normativa en el diagnóstico de una patología neurológica y en el análisis de la evolución de un sujeto bajo una terapia establecida.

La ventaja del conjunto de técnicas desarrolladas frente a las existentes sobre Bases Normativas radica en la automatización tanto del proceso de inclusión de un sujeto a la base, como en la aplicación de las herramientas que determinan la *normalidad* en la señal EEG, evitando cualquier intervención humana en el diseño de la Base Normativa.

El conjunto de procesos desarrollados se centran en favorecer la capacidad de discriminación entre sujetos sanos y patológicos, en función de unos patrones extraídos de la señal EEG, y analizar estadísticamente estos patrones para determinar su robustez frente a la variabilidad interpersonal. La búsqueda de los patrones de referencia se orientó hacia descriptores de la energía en las bandas del espectro de potencias de la señal EEG los cuales se ha demostrado su relación con diferentes patologías neurológicas.

Finalmente, se lleva a cabo la validación global de las técnicas diseñadas tomando la Base Normativa desarrollada como una prueba diagnóstica bajo un contexto clínico, mediante el uso de datos EEG cuasi-sintéticos.

Índice

1. Introducción	2
1.1. Contexto	2
1.2. Motivación y trabajo relacionado	2
1.3. Objetivo y Alcance del proyecto	4
1.4. Organización	6
2. Registro y procesado de señal EEG	7
2.1. Introducción	7
2.2. Electroencefalograma	7
2.3. Descripción de los datos y Registro	8
2.4. Procesado de señal	9
2.4.1. Filtrado de artefactos	9
2.4.2. Extracción de patrones	10
3. Fiabilidad	14
3.1. Introducción	14
3.2. Método 1: Test-Retest	15
3.2.1. Resultados	16
3.3. Método 2: Split&Half	17
3.3.1. Resultados	18
3.4. Ampliación de señal EEG	22

3.4.1. Resultados	23
4. Análisis Estadístico	27
4.1. Análisis estadístico del Periodograma	27
4.2. Transformación	29
4.3. Algoritmo Genético	33
4.3.1. Diseño del Algoritmo Genético	34
4.3.1.1. Codificación	34
4.3.1.2. Parámetros	35
4.3.1.3. Operadores	36
4.3.2. Función objetivo	37
4.3.3. Resultados	39
5. Clusterización	40
5.1. Introducción	40
5.2. Clusterización y Métodos	41
5.2.1. Distancia de similitud	42
5.3. Resultados	43
6. Validación	46
6.1. Estudio de la Fiabilidad	46
6.1.1. Resultados	47
6.2. Estudio del Análisis del Diagnóstico	48
7. Conclusiones y líneas futuras	50
A. Desarrollo	53
A.1. Hitos del proyecto	53
A.2. Diagrama de Gantt	54
B. Sistema BCI	55

B.1. Introducción. Sistema 10-20	55
B.2. Artefactos	57
B.3. Infraestructura	60
B.3.1. Electrodo activos	60
B.3.2. Amplificador	60
B.3.3. Gorro	61
B.3.4. Geles	61
B.3.5. Otros	62
B.3.6. Software: BCI2000	62
C. Análisis estadístico del PSD	63
D. Métodos	64
D.1. Mezcla de gaussianas (GMM)	64
D.2. Distancias en los métodos de clustering	65
D.3. Z-test	65
D.4. Contraste para la diferencia de medias	65
D.5. Distancia Bhattacharyya	66

Índice de figuras

1.1. Aplicaciones de una Base Normativa.	3
1.2. Clusterización de una Base Normativa.	5
1.3. Esquema creación base normativa	6
2.1. Muestra de señal EEG.	8
2.2. Periodograma típico de una señal EEG.	11
2.3. Potencias Absolutas frecuenciales sobre el periodograma de EEG	12
3.1. Test-retest	15
3.2. Resultados de fiabilidad mediante la técnica de <i>Test-Retest</i> para todos los participantes.	17
3.3. Método Split& Half	18
3.4. Rho en función del número de repeticiones	19
3.5. Error SplitHalf	21
3.6. Resultados de <i>Split & Half</i> sobre la señal EEG original para el conjunto de participantes	21
3.7. Añadir epochs	24
3.8. Resultados <i>Split & Half</i>	25
3.9. Resultado test K-S sobre las distribuciones.	25
3.10. Longitud en epochs señal Original & Señal Ampliada	26
3.11. Variación media	26

4.1. Comparación de los histogramas de una variable de potencia absoluta y su versión ampliada mediante <i>epochs</i> 'virtuales'	29
4.2. Variabilidad del valor de λ en la transformación Box-Cox medida sobre todos los participantes de la Base Normativa	31
4.3. Comparativa entre las funciones de transformación aplicadas para mejorar la gaussianidad de una variable para el conjunto de participantes.	32
4.4. Comparativa entre las funciones de transformación aplicadas para mejorar la gaussianidad de una variable en toda la Base Normativa.	32
4.5. Algoritmo Genético	34
4.6. Codificación	35
4.7. Cruce	36
4.8. Mutación	37
4.9. Filtro	38
4.10. Función Objetivo	39
5.1. Ejemplo dendograma.	42
5.2. Cálculo suma de la función <i>Angle</i>	44
5.3. Cálculo codo.	45
5.4. Dendograma en forma circular.	45
6.1. Comparativa sujeto patológico frente al cluster número 1.	49
6.2. Comparativa sujeto patológico frente al cluster número 2.	49
6.3. Comparativa sujeto patológico frente al cluster número 3.	49
A.1. Diagrama de Gantt del proyecto.	54
B.1. Posiciones según el Sistema Internacional 10-20	56
B.2. Montaje con el Sistema Internacional 10-20.	57
B.3. Artefacto por un parpadeo.	58
B.4. Artefacto por un parpadeo.	59
B.5. GAMMABox y electrodos	60

B.6. Amplificador 61

B.7. Gorro 61

D.1. Contraste de la diferencia de medias 66

Índice de tablas

2.1. Bandas frecuenciales de la señal EEG.	12
3.1. Matriz comparativa de los resultados en la aplicación las cuatro técnicas <i>Test-Retest</i>	16
6.1. Resultados de la sensibilidad parcial sobre la Base Normativa.	48

1. Introducción

1.1. Contexto

El presente proyecto fin de carrera se ha realizado dentro del Grupo Robótica, Percepción y Tiempo real (RoPeRT) del Instituto de Investigación en Ingeniería de Aragón (I3A) de la Universidad de Zaragoza. En particular se ha desarrollado junto con el equipo de investigación de Interfaces Cerebro-Computador (BCI), compuesto por un grupo multidisciplinar conformado por profesionales en áreas como la Informática, la Psicología y las Telecomunicaciones.

El trabajo desarrollado se enmarca en un proyecto orientado a la aplicación de Neurofeedback mediante una Interfaz Cerebro-Computador en el tratamiento de trastornos psiquiátricos como el Trastorno de Déficit de Atención (TDA).

1.2. Motivación y trabajo relacionado

La evolución de la tecnología moderna ha transformado completamente el diagnóstico, pronóstico y tratamiento de las enfermedades neurológicas en las últimas décadas situando al electroencefalograma (EEG) entre las técnicas más utilizadas en aplicaciones clínicas. En esta dirección, una herramienta que está ganando peso son las Bases Normativas de EEG como una herramienta de referencia para el diagnóstico clínico. Una Base Normativa de EEG se obtiene a partir de una colección de señales electroencefalográficas que han sido obtenidas usando el mismo protocolo de adquisición sobre un conjunto de personas sanas. La Base Normativa en sí, proporciona un patrón de referencia o comparativo que permite definir el estado de *normalidad* en una muestra de población sana. A su vez, la Base Normativa tiene asociadas unas herramientas matemáticas que permiten comparar a un sujeto frente a una población con el objetivo de identificar las medidas que se desvían de la normalidad, así como la magnitud de esta desviación. De tal forma, una Base Normativa puede usarse en las siguientes aplicaciones:

- Evaluar el estado neurológico de un paciente y **verificar un diagnóstico psicológico**. Un profesional evalúa el historial clínico, síntomas y dolencias, y posteriormente

hace uso de las herramientas de medida para comparar el paciente con la Base Normativa, añadiendo de esta forma precisión al diagnóstico clínico.

- Aumentar la eficiencia de un tratamiento mediante la comparación del EEG antes, durante y después del tratamiento, analizando la **evaluación** del mismo.

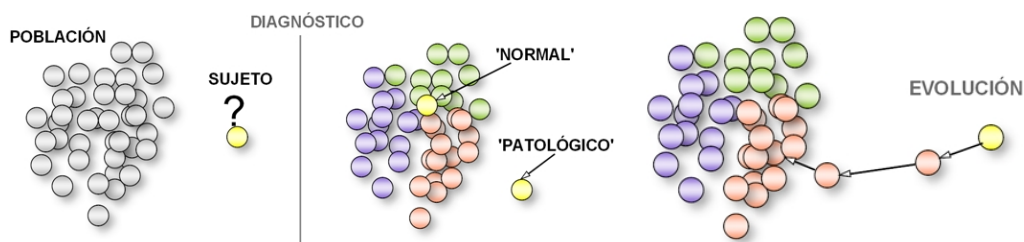


Figura 1.1: Aplicaciones generales de una Base Normativa. (Izquierda) Diagnóstico de un sujeto mediante la comparación del mismo frente a la Base Normativa. (Derecha) Análisis de la evaluación de un tratamiento o terapia realizado sobre un sujeto externo a la Base Normativa.

Se distinguen dos tipos de Bases Normativas según las condiciones de adquisición de EEG: (a) Las Bases Normativas Pasivas (BNP) [1] en el caso que el sujeto permanezca en estado basal¹; y (b) Las Bases Normativas Activas (BNA) [2] cuando el sujeto realiza algún tipo de tarea cognitiva o perceptual durante el registro de datos. Las BNP proporcionan simplicidad y relativa uniformidad en las condiciones de registro de EEG, dado que el registro se realiza en unas condiciones estándar. Esto permite comparar BNP procedentes de diferentes laboratorios y/o muestras de población con alta fiabilidad. Por su parte las BNA resultan más difíciles de homogeneizar dado que el proceso se complica debido a la variabilidad de las condiciones del *recording* (la intensidad de estímulos, la distancia estímulo sujeto) y otros aspectos difíciles de calcular como el nivel de involucración del sujeto. A estos motivos se suma la problemática asociada al filtrado de artefactos de una señal EEG adquirida en condiciones de tarea activa (en EEG basal los artefactos son estándar y se estudian en cursos básicos de Neurofeedback² por medio de atlas [3]). Dadas las dificultades añadidas que presenta una BNA, y a la poca experiencia del grupo investigación en el campo de las Bases Normativas se decidió desarrollar una Base Normativa de EEG basal en ojos cerrados.

Un aspecto clave en el desarrollo de una Base Normativa e independientemente de las condiciones de adquisición de EEG es la alta variabilidad interindividual del registro de EEG. Este aspecto ha sido estudiado de forma explícita [4], demostrando que tal variabilidad podría ser atribuible a la presencia de un factor de escala multiplicativo asociado con la amplitud del EEG. Es decir, los registros del EEG están condicionados por un Factor de Escala Global (FEG) que permanece constante en cada individuo en todas las frecuencias,

¹Estado basal: el sujeto permanece con los ojos abiertos o cerrados en reposo sin realizar ninguna tarea mental en particular

²Neurofeedback: también denominado neuroterapia, es una técnica que se basa en medir la actividad cerebral y enseñar a las personas su propia regulación por aprendizaje condicionado.

derivaciones y diferentes estados funcionales (estado de reposo, tarea activa...) pero cambia entre individuos. Además se atribuye una posible dependencia del FEG con aspectos fisiológicos del individuo (como la edad), y por tanto con propiedades de maduración del volumen conductor del cerebro, como las características del hueso, la geometría del cráneo y las conductancias de la piel y el cráneo. En consecuencia, el desarrollo de una Base Normativa requiere o bien un estudio de la variabilidad del EEG causado por motivos fisiológicos (los cuales no han sido completamente demostrados actualmente, como la edad, la geometría del cráneo, etc), o bien una estratificación de sujetos de la Base Normativa que tenga en cuenta explícitamente los cambios significativos en su EEG. Hasta la fecha todas las Bases Normativas realizan una estratificación por edades [5] (aunque se ha visto que es útil en sujetos maduros dado que el EEG es relativamente invariable este criterio no se satisface en el caso de Bases Normativas de niños) y otros autores han ajustado estos patrones dependiendo además de la región del cerebro analizada [6].

A estos inconvenientes se suma que la contribución del FEG sobre la varianza total del EEG apenas alcanza el 42 % de los datos corregidos por la edad. Por tanto, actualmente en muchos de los estudios de EEG se trata de disminuir esta variabilidad que existe entre individuos de la misma edad [7]. Más allá de la restricción mencionada, los problemas principales asociados a la estratificación del EEG en base a la edad se resume en:

- No linealidad de las relaciones entre el EEG y la edad del individuo.
- Heteroscedasticidad³ de la predicción del error en la no linealidad mencionada.
- Vulnerabilidad en errores de calibración.

Todos estos aspectos apuntan a que la variabilidad interpersonal es una limitación en la construcción de las Bases Normativas actuales. Este proyecto propone una alternativa a la estratificación de señal EEG por edades, la cual está basada en la aplicación de una técnica de clusterización de individuos en función de un conjunto de características de señal extraídas sobre el EEG. Como consecuencia, la población normal queda dividida en clusters caracterizados por propiedades de la señal. La gran ventaja de la técnica es que permite la automatización completa del proceso de generación de la Base Normativa. A su vez, realiza una estratificación explícita basada en los cambios significativos del EEG.

1.3. Objetivo y Alcance del proyecto

El objetivo de este proyecto fin de carrera es el diseño, desarrollo y validación de una Base Normativa de EEG basal de 84 adultos sanos (comprendidos entre 18 y 30 años) en condiciones de adquisición de ojos cerrados. En esta dirección se ha diseñado un proceso innovador de creación de una Base Normativa mediante el cual los sujetos que componen

³Heteroscedasticidad: Se produce cuando la varianza de las perturbaciones no es constante a lo largo de las observaciones.

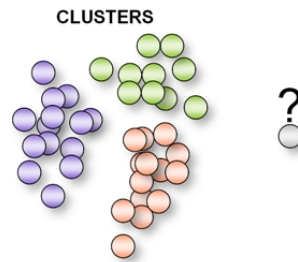


Figura 1.2: Clusterización de una Base Normativa

la Base Normativa se dividen en clases aplicando un algoritmo paramétrico construido sobre características del EEG. A su vez se han creado las herramientas automáticas para comparar un individuo frente a la Base Normativa, y determinar si se encuentra dentro de una *normalidad* definida. Si el sujeto no concuerda con la Base Normativa, es posible determinar de que manera se desvía de la normalidad y la magnitud de esta desviación. La ventaja del proceso creado es que es automático con lo que la inclusión de un individuo es directa y evita la intervención humana en el proceso.

La figura 1.3 muestra el proceso desarrollado y las etapas que se han llevado a cabo en este proyecto. Las cuales se describen a continuación: el desarrollo de la Base Normativa comienza con la etapa de adquisición de EEG, la cual no ha formado parte del presente proyecto. En concreto, se ha trabajado con datos de señal EEG recogidos por la empresa *Nova Tech EEG (NTE)*. A esta etapa le sigue un tratamiento de señal específico para la extracción de las características en las que sustenta la Base Normativa. Estos parámetros son analizados mediante la técnica de *Split&Half* para estudiar su fiabilidad. Posteriormente, se realiza un análisis estadístico de dichas características con el objetivo de obtener una distribución estadística común a todas ellas, determinándose la distribución gaussiana como función objetivo. La siguiente etapa define un método de clusterización de los sujetos en clases siguiendo un criterio de similitud en cuanto a características de señal EEG. Por último, se concluye con una etapa de validación, en la cual se analiza la sensibilidad de la Base Normativa y las herramientas de medida diseñadas como una prueba de diagnóstico.

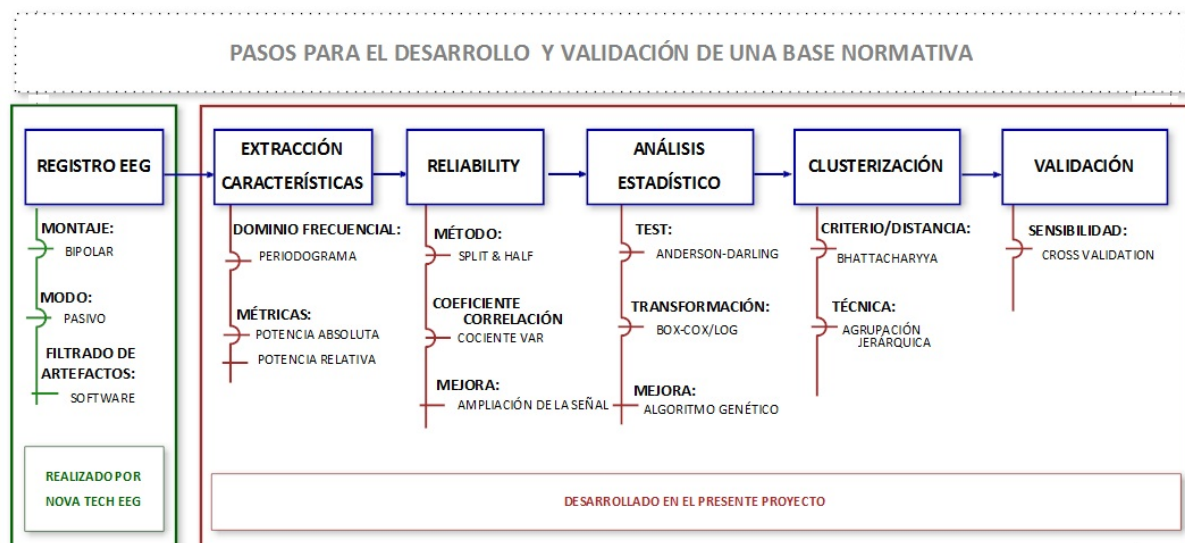


Figura 1.3: Procedimientos para el desarrollo y validación de una base normativa de EEG.

1.4. Organización

El presente documento está estructurado en siete capítulos, siendo este primero la introducción; el capítulo dos presenta los conceptos básicos sobre el electroencefalograma además detalla el registro y procesamiento de señal EEG llevado a cabo para la extracción de los patrones que caracterizan la Base Normativa; el capítulo tres detalla los estudios realizados sobre la fiabilidad de los datos adquiridos; el capítulo cuatro describe el análisis estadístico realizado sobre la distribución de las muestras utilizadas; el capítulo cinco expone el proceso de segmentación realizado sobre los sujetos de la Base Normativa según criterios de similitud en parámetros de la señal EEG; el capítulo seis describe la validación de las herramientas de medida diseñadas junto con la Base Normativa y por último en el capítulo siete se recopila la valoración sobre el trabajo realizado, y se exponen las posibles extensiones del proyecto.

2. Registro y procesado de señal EEG

2.1. Introducción

El presente capítulo describe la señal empleada en el desarrollo de este proyecto, su adquisición y procesado. La sección 2.2 presenta una introducción del electroencefalograma, describiendo el origen fisiológico de la señal EEG y de sus características temporales y espectrales. A continuación, la sección 2.3 describe los datos EEG utilizados en este proyecto y el registro llevado a cabo. Y por último, en la sección 2.4 se detalla el tratamiento de señal realizado, una primera etapa de filtrado de artefactos y un procesado más concreto para la extracción de los parámetros que recogen la Base Normativa de este proyecto.

2.2. Electroencefalograma

El electroencefalograma (EEG) es el registro no invasivo de la actividad eléctrica cerebral mediante electrodos colocados en la superficie del cuero cabelludo. Al conjunto de la actividad eléctrica se le conoce como ritmo debido a su comportamiento oscilatorio. Concretamente la actividad de una sola neurona cortical no puede ser medida debido al grosor de las capas de tejido que componen la corteza cerebral. Sin embargo la actividad conjunta de millones de neuronas corticales, localizadas a una profundidad de varios milímetros, producen un campo eléctrico suficientemente fuerte para ser medido en la superficie cerebral, siendo esta profundidad dependiente de la potencia de la fuente neuronal. El campo eléctrico es principalmente generado por corrientes que fluyen durante excitaciones sinápticas¹ de las dendritas, la excitación post-sináptica. La diversidad de ritmos EEG es inmensa y depende, entre otras cosas, del estado mental del sujeto (grado de atención, de relajación, etc). Los ritmos están convencionalmente caracterizados por:

- **La amplitud de EEG:** está relacionada con el grado de sincronización con el cual las neuronas corticales interaccionan. Una excitación sincronizada se produce

¹Sinapsis: unión intercelular producida por una descarga química que origina una corriente eléctrica en la membrana de la célula sináptica (emisora) una vez que este impulso alcanza el extremo del axón la propia neurona segrega neurotransmisores.

2. Registro y procesamiento de señal EEG

2.3 Descripción de los datos y Registro

cuando las señales originadas por cada neurona se suman de forma coherente en fase produciendo como resultado una señal de amplitud elevada. Por el contrario, una excitación asíncrona de las neuronas provoca un EEG irregular de baja amplitud. La amplitud es variable, de 20-60 μV (50 μV por término medio).

- **La frecuencia de EEG:** viene definida por la actividad proveniente del Tálamo. Esta parte del cerebro está formada por neuronas que poseen propiedades marcapasos, es decir, tienen la capacidad intrínseca para generar un patrón rítmico.

En la actualidad, el registro de EEG se lleva a cabo mediante la colocación de un gorro sobre el cuero cabelludo, en el cual se colocan un conjunto de electrodos. La disposición de los electrodos sobre el cuero cabelludo sigue el estándar descrito por el Sistema Internacional 10-20, ver en anexo B.1. Cada par de electrodos registra las variaciones de potencial eléctrico definiendo un canal de EEG. Cada una de estas señales son amplificadas mediante el uso de un amplificador diferencial previo a la etapa de digitalización. La figura 2.2 representa una muestra de encefalograma compuesta por 16 canales. Tal y como se observa, la señal es segmentada en fragmentos temporales denominados *epochs* (divididos por líneas verticales en la figura).

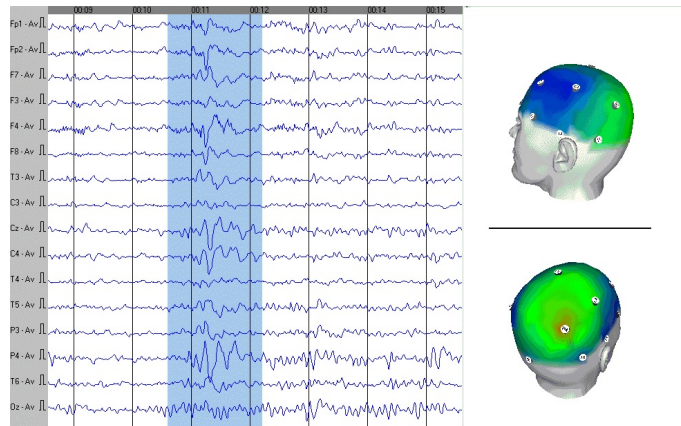


Figura 2.1: (Izquierda) Ejemplo de señal EEG adquirida sobre 16 posiciones del cuero cabelludo. Cada fila representa un canal. (Derecha) Representación visual en forma de mapa de color de la zona sombreada en la imagen izquierda

2.3. Descripción de los datos y Registro

Los datos empleados en este proyecto fueron adquiridos por la empresa *NTE*. Éstos se componen de la señal EEG registrada en 84 adultos sanos, con edades comprendidas entre los 18 y 30 años. La selección de los sujetos siguió unos criterios de exclusión basados en el historial psiquiátrico, tanto de los sujetos como de sus familiares, además de un análisis sobre posibles abusos de drogas, discapacidades físicas o lesiones cerebrales. El registro

se realizó durante una sesión (o trial) continua de unos 3 a 5 minutos de duración, en la cual el sujeto se encontraba en condiciones de reposo, con los ojos cerrados y sentado cómodamente en una habitación silenciosa, con luz tenue. El protocolo de montaje junto con el registro de señal EEG fue llevado a cabo por *NTE* (en el anexo B.3 se detalla de forma más extensa la infraestructura de montaje utilizado). Para la adquisición de señal EEG se utilizaron 19 electrodos cuya colocación siguió el estándar prescrito por el Sistema Internacional 10-20.(FP1, FP2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2). Se realizó un montaje bipolar, en el cual, la señal de referencia es adquirida en los lóbulos de las orejas, denominado ‘linked-ears reference’. Se comprobó que la impedancia de los electrodos estuviera por debajo de 5 Kohmios asegurando una relación señal a ruido apropiada. El amplificador diferencial utilizado es el sistema de adquisición-24 NeuroSearch empleando un conversor de 12-bit Analógico-Digital (Lexicor Medical technology, Inc., boulder, CO). El cual permite un filtrado analógico para eliminar la interferencia debido al ruido ambiental que proviene de las lámparas fluorescentes y otros dispositivos que emiten ruido a través de ondas de 60Hz, utilizando un filtro elimina-banda tipo Notch. En el sistema de digitalización (previo a la conversión Analógico-Digital) se realizó un prefiltrado mediante un filtro antialiasing: filtro paso bajo con frecuencia de corte de 50Hz. Se seleccionó una f_{Δ} de sampleo de 128 Hz para garantizar que la f_{Δ} es superior a dos veces la máxima frecuencia presente en la señal continua y así satisfacer el Teorema de Muestreo de Nyquist:

$$f_{\Delta} > 2f_{max} \quad (2.1)$$

2.4. Procesado de señal

2.4.1. Filtrado de artefactos

La señal EEG se caracteriza por su reducida magnitud (del orden de las decenas de uV), por lo que es fácilmente contaminada por ruido e interferencias. De tal manera, un aspecto especialmente importante en el tratamiento de señal EEG consiste en la eliminación de ruido y artefactos para mejorar la relación señal a ruido (en inglés Signal to Noise Ratio SNR o S/N). Un correcto filtrado de artefactos permite, además de eliminar señales no deseadas, minimizar la variabilidad interpersonal del EEG. Este es un aspecto crítico en el desarrollo de una Base Normativa, dado que en definitiva una BN agrupa señales EEG provenientes de diferentes sujetos con el objetivo de definir un patrón de referencia.

Un artefacto es un registro de actividad no cerebral que aparece dentro del EEG alterando en gran medida la señal adquirida debido a que su amplitud suele ser mucho mayor que la propia actividad cerebral. Los artefactos se presentan inevitablemente en todo registro de EEG pudiendo tener un origen biológico, instrumental o medioambiental. Para más información ver el anexo ???. El filtrado de artefactos requiere de un importante conocimiento y experiencia. Este proceso consiste en un exhaustivo reconocimiento de artefactos mediante la ayuda de un *software* de visualización, procediendo manualmente

a eliminar aquellos *epochs* afectados por un posible artefacto. Como consecuencia, la longitud de señal EEG original se ve sensiblemente reducida. El filtrado de artefactos realizado por *Nova Tech EEG* se compone de dos fases. La primera de ellas, consiste en una inspección visual del EEG digitalizado en una pantalla de alta-resolución para identificar los *epochs* que contengan artefactos visibles. A continuación, se procede a un filtrado más exhaustivo mediante la aplicación Eureka ², ayuda a la identificación de artefactos más sutiles la cual necesita siempre la intervención humana para la eliminación.

2.4.2. Extracción de patrones

Una Base Normativa recoge un conjunto de patrones para caracterizar el EEG de un sujeto sano. La extracción de patrones relevantes es particularmente crucial cuando el propósito es diseñar un sistema que permita discriminar entre un EEG 'normal' y uno patológico, como es una de las aplicaciones de este proyecto. Una vez la señal EEG está libre de artefactos se prosigue con un tratamiento más específico para la extracción de características de la señal. Los patrones más utilizados en las investigaciones mediante EEG se clasifican en patrones espectrales y patrones de conectividad. Los primeros se definen a partir del análisis del espectro de frecuencias en una ventana temporal y describen la intensidad del campo electromagnético de esa localización. Los patrones de conectividad son medidas más complejas que consideran características espacio-temporales, que implican varias localizaciones simultáneamente y describen la intensidad de la conexión entre esas regiones cerebrales.

Los patrones de señal EEG más comunes desde un punto de vista clínico son los patrones espectrales de frecuencia. El tratamiento frecuencial de EEG comienza con el cálculo de la densidad espectral de potencias (PSD). La PSD de la señal EEG se ha llevado a cabo mediante el estimador espectral de potencia, denominado periodograma. El cálculo del periodograma asume que la señal es estacionaria³ y periódica, es decir, que se puede identificar un patrón que se va repitiendo con cierta periodicidad (a estos segmentos se les denomina *epochs*). Se seleccionó una longitud de *epoch* igual a un segundo, coincidiendo con la duración elegida en la limpieza de artefactos. Además, esta longitud es usada habitualmente en anteriores estudios que analizan la estacionariedad del EEG. Debido a la propiedad no estacionaria del EEG, se asume en su análisis que el EEG está compuesto por segmentos consecutivos con propiedades estacionarias.

El procesado se realizó sobre las 19 señales EEG filtradas de artefactos por el proceso descrito en la subsección anterior. La estimación del PSD se llevó a cabo mediante el periodograma modificado, en el cual cada una de las señales es inventanada en segmentos de un segundo. Se utilizó la ventana de *Tukey*, o coseno alzado truncado, cuya expresión se define para una longitud L :

²Aplicación para el análisis y filtrado de artefactos en la señal EEG, de *Nova Tech EEG*.

³Señal estacionaria: señal cuyos parámetros estadísticos permanecen constantes sobre el tiempo.

$$W_{tukey}(i) = \begin{cases} \frac{1}{2}(1 + \cos(\frac{2\pi}{\alpha}[x - \alpha/2])) & 0 \leq x < \frac{\alpha L}{2} \\ 1 & \frac{\alpha L}{2} \leq x < L(1 - \frac{\alpha}{2}) \\ \frac{1}{2}(1 + \cos(\frac{2\pi}{\alpha L}[x - 1 + \alpha/2])) & (1 - \frac{\alpha}{2}) \leq x \leq L \end{cases} \quad W_{tukey}(i) = \frac{1}{2}(1 + \cos \frac{\pi}{T}t) \quad (2.2)$$

El valor de longitud de ventana L se tomó igual a 128 puntos, los cuales representan un segundo de la señal EEG en tiempo muestreada a una frecuencia de 128 Hz. El cálculo de la Transformada Discreta de Fourier (*Discrete Fourier Transform, DFT*) se llevó a cabo mediante la aplicación de *zero-padding*⁴, lo que permitió emplear un total de 2048 puntos en el cálculo de la DFT. Obteniéndose una resolución frecuencial de:

$$\Delta_f = \frac{Fs}{N} = \frac{128Hz}{2048muestras} = 0,0625Hz/muestra \quad (2.3)$$

La densidad espectral de potencias del EEG se caracteriza por la forma representada en la figura 2.2. Aproximadamente el 98 % de la potencia en la señal se distribuye dentro del intervalo 0.5 y 30Hz, por lo que éste intervalo es el analizado en este proyecto.

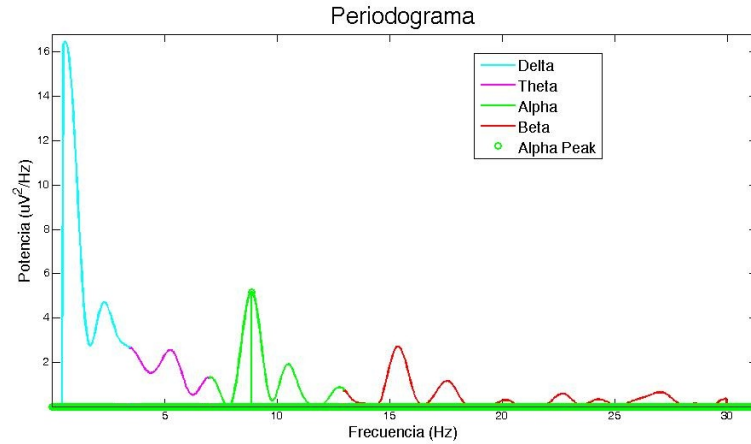


Figura 2.2: Periodograma típico de una señal EEG, el cual representa la distribución de la potencia de la señal sobre el dominio frecuencial. A través del código de colores se representan las bandas de frecuencia típicas del espectro de EEG: Delta, Theta, Alpha y Beta.

En general, la densidad espectral de potencia de EEG es fuertemente variable de forma intrapersonal, es decir, la distribución de potencia en el dominio frecuencial varía dependiendo de cada persona. Por este motivo, no resulta útil en términos de comparación tomar un patrón o (parámetro) de señal EEG extraído de una única frecuencia o bin del espectrograma, sino que usualmente se agrupan en bandas frecuenciales. Tradicionalmente

	Delta	Theta	Alfa	Beta
Intervalo	[0.5-3.5 Hz]	[3.5- 7 Hz]	[7-13 Hz]	[13-30 Hz]

Tabla 2.1: Definición bandas frecuenciales del espectro en la señal EEG.

se han definido cuatro bandas frecuencia, mostradas en la tabla 2.1.

A pesar de que estas bandas de frecuencia han sido estudiadas durante décadas y la gran mayoría de trabajos con el EEG están descritos en base a ella, no existe consenso absoluto respecto a los límites exactos de estos ritmos frecuenciales. Como resultado, existe pluralidad en la definición tanto del número de bandas frecuenciales como de las frecuencias limítrofes. En este proyecto, se emplean las bandas definidas anteriormente, las cuales han sido definidas en anteriores Bases Normativas [1].

Para determinar los parámetros de señal que recogen la Base Normativa de este proyecto, se estudiaron los diferentes patrones patológicos conocidos en el EEG [8]. Finalmente, se determinó por su relevancia en la caracterización mayoritaria de las patologías estudiadas, la potencia absoluta. La potencia absoluta (PA) se define como la cantidad media de potencia (en μV^2) en una banda frecuencial y es calculada como el área representada bajo la curva de PSD en el intervalo definido por dicha banda. En la figura 2.3 se representa el área correspondiente a la potencia absoluta de cada banda frecuencial. En consecuencia,

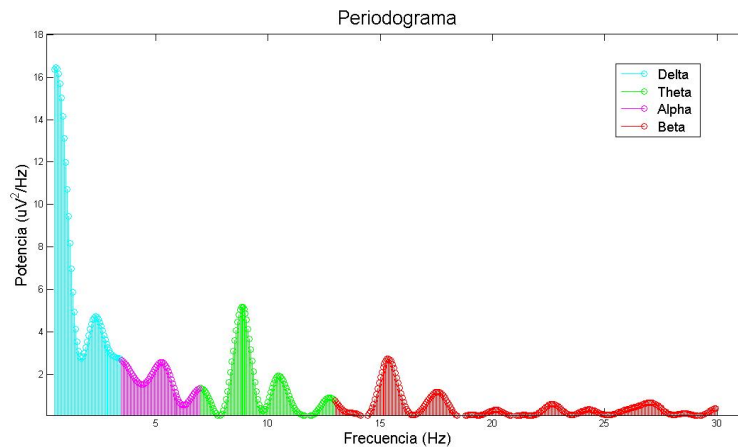


Figura 2.3: Potencias Absolutas frecuenciales sobre el periodograma de EEG

cada *epoch* de un canal de EEG es analizado en el dominio frecuencial para la extracción de la potencia absoluta en cada una de las bandas: delta, theta, alpha y beta. Como resultado se obtienen 4 conjuntos de valores por cada cálculo de potencia en los canales EEG, es decir, $19 \times 4 = 76$ variables de potencia por sujeto.

En las próximas secciones se tomará la siguiente nomenclatura:

⁴*Zero-Padding*: consiste en rellenar con ceros la señal en tiempo a partir de su último valor hasta alcanzar el tamaño deseado para mejorar la resolución espectral en su Transformada de Fourier.

- P_{δ}^i : Variable de potencia en la banda δ , del canal i.
- P_{θ}^i : Variable de potencia en la banda θ , del canal i.
- P_{α}^i : Variable de potencia en la banda α , del canal i.
- P_{β}^i : Variable de potencia en la banda β , del canal i.
- $\{P_{EEG}\} = \{P_{\delta}^1, P_{\theta}^1, P_{\alpha}^1, P_{\beta}^1, \dots, P_{\beta}^n\}$: Representa el conjunto de las variables de potencia absoluta, con n=19 canales.

3. Fiabilidad

3.1. Introducción

Este capítulo describe el estudio sobre la fiabilidad de la señal EEG adquirida en cada sujeto, (fiabilidad intra-personal). La fiabilidad se define como la consistencia de un conjunto de medidas realizadas sobre el mismo fenómeno. Por tanto, la fiabilidad es un indicador sobre la calidad y reproducibilidad de una medida. En este capítulo se estudiará si los datos EEG adquiridos sobre el mismo sujeto, $\{P_{EEG}\}$, mantienen una coherencia a lo largo de la sesión registrada. En la literatura, la fiabilidad de la señal EEG ha sido ampliamente estudiada, empleándose los métodos de *Test-Retest* y *Split & Half*. La sección 3.2 detalla el método de *Test-Retest* y modificaciones empleadas para su adaptación al presente contexto. Sin embargo los resultados obtenidos no consiguieron alcanzar unos valores aceptables en cuanto a fiabilidad lo que se achaca principalmente a la adaptación necesaria dado que no se cumplen de forma estricta las condiciones de aplicación *Test-Retest*. En la sección 3.3 se plantea un segundo método (*Split & Half*) con el que si se tienen las condiciones de aplicabilidad. Aunque con este método se obtuvieron resultados de fiabilidad acordes a la literatura no se alcanzan los resultados deseados. Esto es debido a que *Split & Half* es un método sensible a la longitud de la señal y la cantidad de datos disponibles $\{P_{EEG}\}$ es relativamente reducida. Para paliar este problema, en la sección 3.4 se propone una técnica para aumentar la longitud de $\{P_{EEG}\}$ y mejorar de este modo la fiabilidad de los mismos. Para ello se diseña un algoritmo que permite aumentar el número de valores de señal. Con un efecto no apreciable en el rango de frecuencia de interés que permite aumentar la fiabilidad. Como resultado se obtuvo una fiabilidad del conjunto de distribuciones de potencia $\{P_{EEG}\}$ en torno al 89 %.

3.2. Método 1: Test-Retest

La mayoría de estudios sobre fiabilidad coinciden en estudiar el método denominado *Test-Retest*, en el cual se administra el mismo test sobre un *item* en dos ocasiones de tiempo. La cantidad de tiempo transcurrido entre tests es crítica, debido a que cuando se realiza una medida sobre el mismo *item* dos veces, la correlación entre las dos observaciones dependerá en parte del tiempo transcurrido entre ellas.

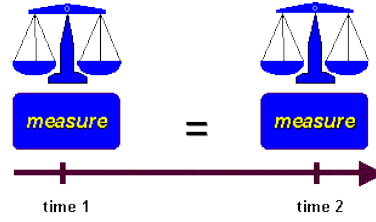


Figura 3.1: Representa el estudio de fiabilidad mediante la técnica de *Test-Retest*. Este método consiste en comparar las medidas resultantes de aplicar un test sobre el mismo *item* en dos instantes de tiempo diferentes.

La fiabilidad viene cuantificada a través de una medida de correlación dada por el coeficiente de correlación de Pearson o el coeficiente de correlación de Spearman:

- El coeficiente de correlación de Pearson (r) es un índice que mide el grado de asociación lineal entre dos variables X e Y . Viene definido por el cociente:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (3.1)$$

donde σ_{xy} indica la covarianza de (X, Y) , σ_x y σ_y representan las desviaciones típicas de las distribuciones marginales.

- El coeficiente de correlación de Spearman (ρ) es una medición no paramétrica de correlación, asume una función monótona arbitraria para describir la relación entre dos variables sin hacer ninguna asunción sobre la distribución de frecuencia de las variables X e Y . Para calcular ρ , los datos son ordenados previamente. Su cálculo viene dado por:

$$\rho = 1 - \frac{6 \sum D^2}{N(N^2 - 1)} \quad (3.2)$$

siendo D la diferencia entre los correspondientes valores de $x - y$. N es el tamaño de la muestra expresada en parejas.

3.2.1. Resultados

Anteriores estudios han demostrado la fiabilidad intra-personal de las potencias espectrales en la señal EEG mediante la técnica de *Test-Retest* en condiciones de reposo y sobre señales adquiridas en la misma persona en diferentes momentos de tiempo [9] y con una separación de varios meses o incluso años [10]. En estos estudios hacen uso tanto del coeficiente de correlación de Pearson como el de Spearman [11]. Además otros estudios describen resultados altos de fiabilidad incluso con muestras de señal EEG de duración menor a las empleadas con anterioridad [12], lo cual demuestra valores de fiabilidad alrededor del 92 % en muestras de señal de 60 segundos.

En relación al estudio de fiabilidad en este proyecto, existe unas limitaciones impuestas por la metodología del experimento, dado que consistió en la adquisición de una única sesión de EEG por participante. Bajo esta restricción, no es posible satisfacer las condiciones para aplicar la técnica de *Test-Retest* al no poder analizar dos señales de EEG adquiridas en dos instantes de tiempo diferentes. Como alternativa, se plantearon dos soluciones:

1. Técnica I: Dividir la sesión adquirida de EEG en dos mitades y calcular la correlación existente entre ellas. De esta manera, se estudia la fiabilidad entre la señal EEG del comienzo y final de la sesión.
2. Técnica II: Segmentar la sesión de EEG, agrupar los fragmentos pares e impares en dos conjuntos y calcular la correlación de las dos agrupaciones. Esta solución estudia la fiabilidad de forma más global a lo ocurrido en la sesión adquirida.

Se aplicaron ambas propuestas combinadas con los dos coeficientes de correlación Pearson y Spearman. En el cálculo de ambos coeficientes se tomó como resultado su valor absoluto de éstos dado que únicamente se requiere el grado de correlación (sin tener en cuenta si la correlación es positiva o negativa). Es decir, los valores r y ρ toman valores en el intervalo $[0,1]$, siendo el valor de 0 (fiabilidad igual a 0) el resultado de correlación nula y el valor de 1 correspondiente a la correlación máxima (fiabilidad igual al 100 %). Como resultado se obtuvieron cuatro técnicas en la aplicación de *Test-Retest* cuyos resultados se muestran en la tabla 3.1

	Técnica I	Técnica II
Pearson	0.1255	0.1658
Spearman	0.1270	0.1631

Tabla 3.1: Muestra los valores de fiabilidad calculados aplicando la técnica *Test-Retest* mediante el promedio del coeficiente de correlación en los 84 participantes.

La Figura 3.2 describe los resultados de fiabilidad por participante para cada una de las técnicas de *Test-Retest* aplicadas. Notar que los resultados mostrados corresponden al

promedio de los coeficientes de correlación resultantes en cada una de las 76 distribuciones de potencia que componen $\{P_{EEG}\}$.

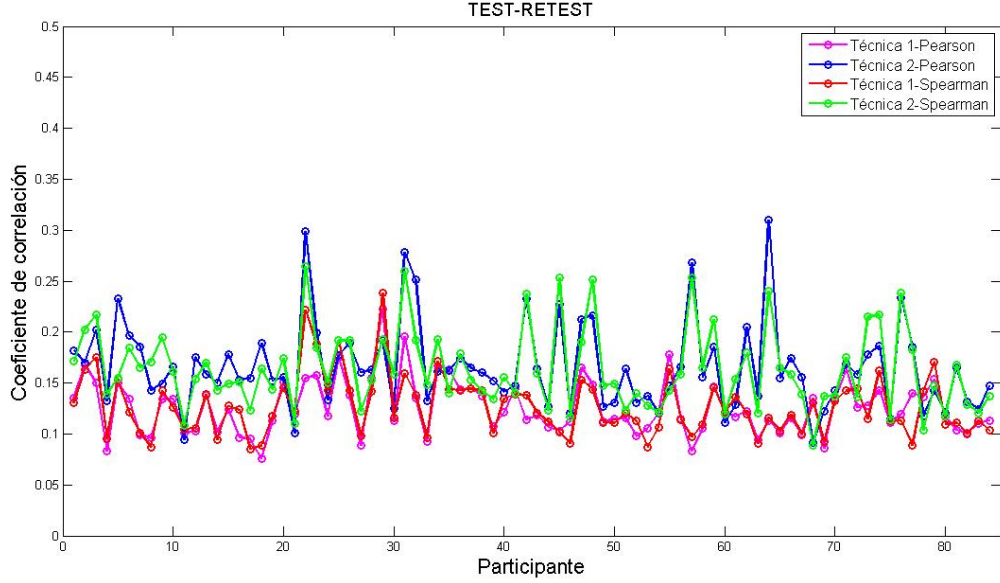


Figura 3.2: Representa los resultados de fiabilidad obtenidos mediante las cuatro técnicas de *Test-Retest* por participante. Para cada participante se muestra el valor medio de los coeficientes de correlación resultantes en cada variable de potencias absolutas.

Los resultados en términos de correlación no superan el 0.4 en ningún caso (30 % de fiabilidad) encontrándose muy alejados en comparación a los conseguidos por los autores mencionados. Los motivos se relacionan por un lado con la longitud de señal EEG con la que se trabaja, y por otro, por no satisfacer las condiciones en las cuales se aplica la técnica de *Test-Retest*.

3.3. Método 2: Split&Half

Debido a que el estudio previo no alcanza unos resultados de fiabilidad satisfactorios como lo han logrado otros autores, se planteó otro estudio de fiabilidad denominado *Split & Half* basado en la consistencia interna de un experimento. Este método juzga la fiabilidad de un experimento estimando cómo de consistentes son los resultados para diferentes segmentos de la medida. La Figura 3.3 representa el cálculo *Split & Half*, que consiste en segmentar el experimento en *items*, reordenarlos de forma aleatoria para después dividir en dos mitades y calcular la correlación entre éstas. La correlación de las dos agrupaciones resultantes de una medida X , viene dada por el siguiente cociente:

$$r_{xx} = \frac{S_1^2}{S_2^2} \quad (3.3)$$

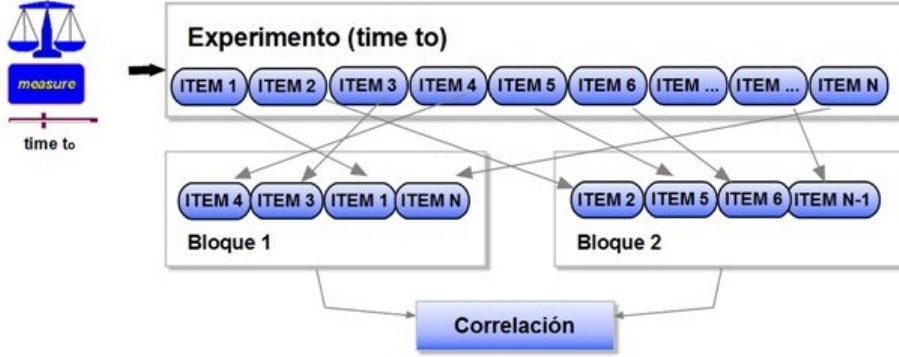


Figura 3.3: Representa el procedimiento para la aplicación del estudio de fiabilidad mediante el método *Split & Half* sobre un experimento realizado. Éste consiste en la segmentación del propio experimento en *items*, para después reordenarlos de forma aleatoria y dividirlos en dos mitades. El resultado de fiabilidad viene dado por la correlación de las dos mitades.

donde S_1^2 es la estimación de la varianza para la medida en el bloque uno, mientras que S_2^2 es la varianza de la medida en el bloque dos. El coeficiente de fiabilidad representa la proporción de varianza obtenida que es válida o fiable.

3.3.1. Resultados

Resultados anteriores [13] han alcanzado valores de fiabilidad entre el 93 % y 98 % mediante la aplicación de *Split & Half* sobre la potencia en las distintas bandas frecuenciales de la señal EEG registrada en condiciones de reposo.

En la aplicación de la técnica de *Split & Half* en este problema, un ítem de la medida se corresponde con el valor de potencia extraído en un fragmento de señal EEG, *epoch*. La duración de un *epoch* se determinó previamente, consistiendo en un segundo de adquisición de EEG. El resultado de fiabilidad *Split & Half* depende de las mitades resultantes al aleatorizar los datos y dividir la señal en dos. Para eliminar esta dependencia, se propone la siguiente alternativa: recomputar el algoritmo *Split & Half* un número determinado de repeticiones y promediar el coeficiente de correlación resultante en cada iteración. Obteniéndose como resultado final el valor de fiabilidad dado por:

$$\bar{r}_{xx} = \frac{\sum_{i=1}^M r_{xx}^i}{M} \quad (3.4)$$

donde r_{xx}^i representa el coeficiente de correlación de la iteración i . Y M es el número de veces que se recomputa el coeficiente de correlación. Para determinar el número de

iteraciones (M) se analizó la influencia de este valor con el resultado final de correlación, \bar{r}_{xx} de forma experimental. En la figura 3.4 se representa la variación de \bar{r}_{xx} en función del número de iteraciones (M) del algoritmo de *Split & Half*, en las distribuciones de potencia de las cuatro bandas frecuenciales calculadas sobre un sujeto en el canal Cz. Se puede observar como \bar{r}_{xx} tiende a estabilizarse en torno a 800 iteraciones \bar{r}_{xx} en las cuatro bandas frecuenciales.

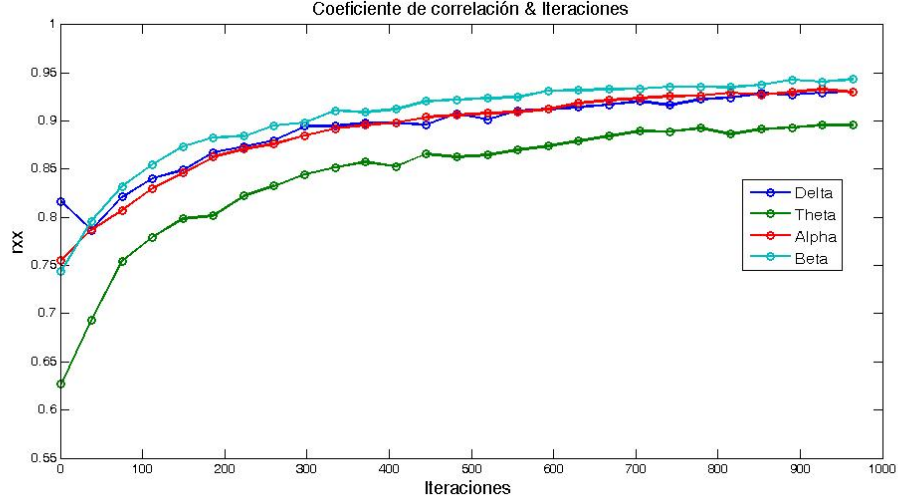


Figura 3.4: Representa el resultado del coeficiente de correlación \bar{r}_{xx} en función del número de iteraciones aplicadas en el método de *Split & Half*. Muestra el resultado obtenido en las distribuciones de potencia de las cuatro bandas frecuenciales: Delta, Theta, Alpha y Beta, de la señal EEG adquirida en el canal Cz del sujeto número 4.

Para determinar de forma analítica el valor del parámetro M, que resultara adecuado de forma generalizada para las señales $\{P_{EEG}\}$ de cada participante, se definió la siguiente métrica: $E(i)$ que cuantifica el error entre dos valores de \bar{r}_{xx} para un número de iteraciones $i = n$ e $i = n - 1$.

$$E(i) = \sqrt{\sum_{c=1}^{19} [r_{dd}^c(i) - r_{dd}^c(i-1)]^2 + [r_{tt}^c(i) - r_{tt}^c(i-1)]^2 + [r_{aa}^c(i) - r_{aa}^c(i-1)]^2 + [r_{bb}^c(i) - r_{bb}^c(i-1)]^2} \quad (3.5)$$

siendo:

c : El canal de la señal EEG.

i : El número de iteraciones computado del algoritmo *Split&Half*.

$r_{\delta\delta}^c(i), r_{\theta\theta}^c(i), r_{\alpha\alpha}^c(i), r_{\beta\beta}^c(i)$: Los coeficientes de correlación de la potencia en las bandas: δ , θ , α y β respectivamente, en la iteración i .

Se fijó un error máximo de umbral $E_{max} = 0,005$, el cual representa un error global en el cálculo de fiabilidad por participante. La correspondencia con el error cometido en

cada una de las 76 distribuciones de potencia que caracterizan a un participante, sería la siguiente: denotando $E_r = r_{xx}^c(i) - r_{xx}^c(i - 1)$ la diferencia del coeficiente de correlación \bar{r}_{xx} como resultado de computar *Split&Half* con i e $i - 1$ iteraciones sobre una variable banda-canal denotada por x :

$$E(i) = \sqrt{19[E_r^2 + E_r^2 + E_r^2 + E_r^2]} \quad (3.6)$$

$$E(i) = \sqrt{76 * E_r^2} \quad (3.7)$$

Siendo $E_{max}(i)=0.005$:

$$E_r < \sqrt{\frac{0,005^2}{76}} = 0,0081 \quad (3.8)$$

Es decir, este valor de error impone una cota máxima de la diferencia del coeficiente de correlación \bar{r}_{xx} como resultado de computar *Split&Half* con i e $i - 1$ iteraciones sobre una variable banda-canal denotada por x por debajo de las milésimas.

A continuación se muestran los resultados del cálculo de la fiabilidad. La figura 3.5 representa el valor del parámetro $E(i)$ en función del número de iteraciones del algoritmo *Split & Half* superponiendo de la solución de cada participante. Las líneas rojas representan el valor de error máximo admisible $E_{max} = 0,005$. Éste valor se cumple entre las 800 y 1000 iteraciones para el conjunto de participantes. Por tanto, se toma como parámetro M del algoritmo de la ecuación 3.4 el valor de 1000 iteraciones.

En la figura 3.6 se muestra los resultados de fiabilidad de cada participante. En la cual se representa la distribución de los valores de r_{xx} resultante en las 76 variables de potencia absoluta. Se concluye que los resultados de fiabilidad se encuentran en torno a $r_{xx} = 0,7$. A diferencia de los resultados anteriores, se consigue aumentar significativamente el valor de fiabilidad, obteniéndose una fiabilidad ajustada con la literatura.

Sin embargo, los resultados obtenidos no alcanzan un nivel de fiabilidad admisible en aplicaciones clínicas. Por tanto, se requiere de una técnica que favorezca esta fiabilidad. Teóricamente la fiabilidad viene limitada gravemente por la duración de la señal analizada. En particular, el presente análisis trabaja con datos de longitud muy reducida (entre 60 y 120 muestras, *epochs* de un segundo). En consecuencia, la longitud de la señal EEG es un valor crítico y determinante en lo que a fiabilidad se refiere.

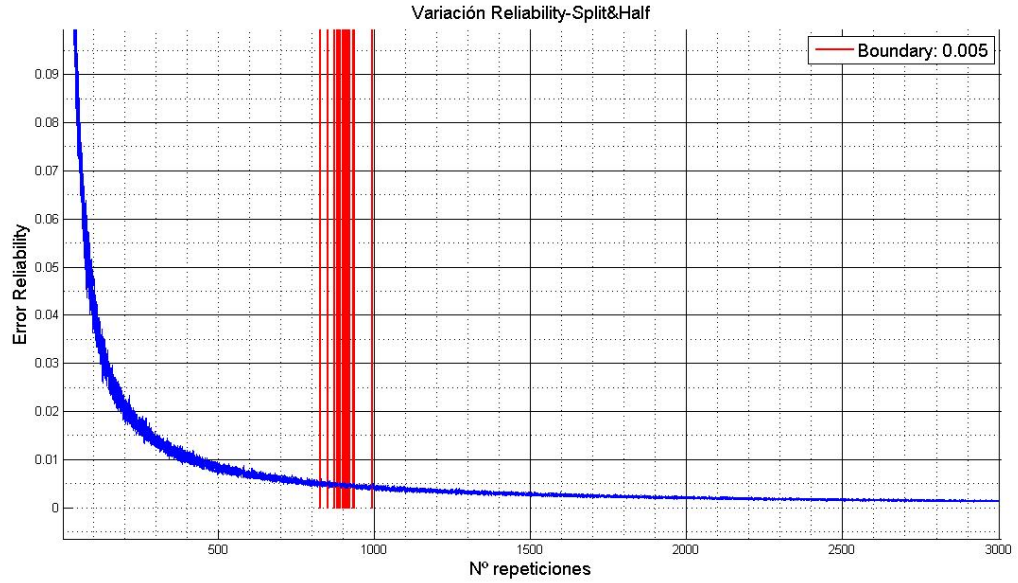


Figura 3.5: Representa el valor del parámetro $E(i)$ en función el número de iteraciones i , el cual define el error cometido en el computo del algoritmo *Split & Half* al tomar $i - 1$ iteraciones en vez de i . Las líneas verticales rojas representan el valor de error máximo admisible $E_{max} = 0,0005$. En ella se ha superpuesto la función $E(i)$ junto con la cota E_{max} para cada uno de los 84 participantes.

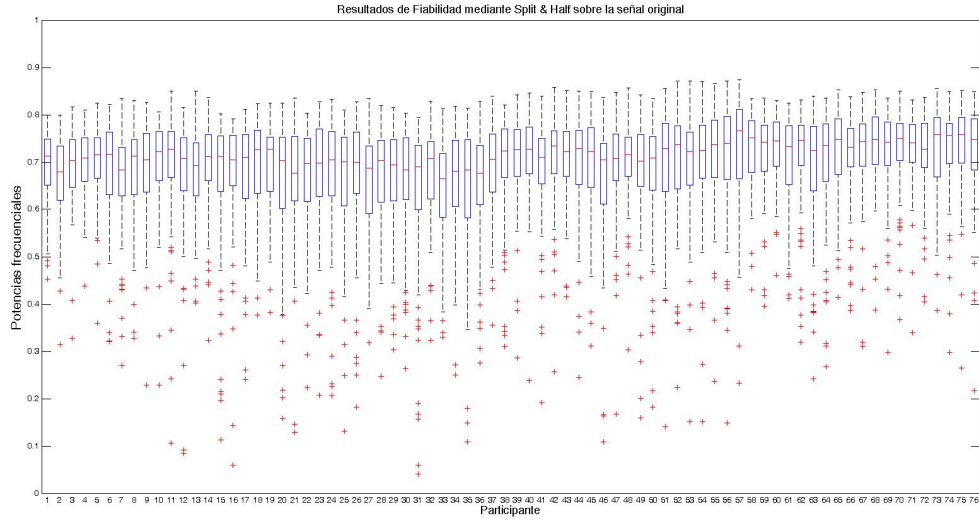


Figura 3.6: Representa los resultados de fiabilidad al aplicar *Split & Half* con un valor del parámetro M igual a 1000 iteraciones. Los resultados se representan mediante el diagrama estadístico denominado *Box-Plot*. El cual es un gráfico representativo la distribución de la variable ρ calculada sobre cada participante. La caja central indica el rango en el que se concentra el 50% central de los datos, sus extremos son el primer y tercer cuartil de la distribución y la línea central en la caja es la mediana. Los extremos delimitan el 95% central de los datos, y las muescas en rojo representan los datos considerados como espurios en la distribución.

3.4. Ampliación de señal EEG

Debido a una restricción en la longitud de la señal EEG adquirida y por tanto del número de epochs, los resultados de fiabilidad obtenidos no fueron los deseados. Por tanto, fue necesario encontrar una estrategia para aumentar la duración de la señal EEG de cada uno de los participantes añadiendo epochs ‘virtuales’ con el objetivo de mejorar los valores de fiabilidad intra-personal.

Un primer diseño contemplaba el uso de ventanas temporales deslizantes de un segundo de duración aplicando un porcentaje de solapamiento entre ventanas consecutivas. De este manera, se obtiene un valor de potencia por cada ventana aplicada. Al fijar la longitud de ventana junto con el porcentaje de solapamiento entre ellas, el número de ventanas resultantes queda limitado por la duración de la señal temporal en segundos. Este primer diseño fue abandonado porque el número de *epochs* ‘virtuales’ resultantes quedaba limitado a un número insuficiente para su aplicación.

Se diseñó una segunda técnica donde los límites descritos por el método anterior quedaban resueltos. Para ello, se añadió una reordenación aleatoria de la señal original segmentada en *epochs* de un segundo de duración. Un *epoch* ‘virtual’ resulta de unir dos mitades de *epochs* consecutivos de la señal original reordenada, en consecuencia se obtienen $n - 1$ *epochs* ‘virtuales’ sobre una señal de duración n . Este procedimiento se recomputa hasta conseguir una señal con una fiabilidad deseada calculada mediante la técnica de *Split & Half*. En la Figura 3.7 se representa el algoritmo diseñado, el cual está descrito por el siguiente algoritmo iterativo:

1. Reordenar la secuencia original: permutar los epochs iniciales aleatoriamente.
2. Agrupar pares de epochs consecutivos de dos en dos.
Ejemplo: $[EPOCH_i - EPOCH_{i+1}]$.
3. Dividir cada pareja en cuatro fragmentos cortando cada *epoch* por la mitad.
Ejemplo: $[EPOCH_i(a) - EPOCH_i(b) - EPOCH_{i+1}(a) - EPOCH_{i+1}(b)]$.
4. Crear el nuevo *epoch* ‘virtual’ seleccionando aleatoriamente entre las siguientes casos: uniendo la segunda mitad del primer epoch con la primera mitad del segundo, la primera mitad del primer *epoch* con la segunda mitad del segundo.
Ejemplo: $[EPOCH_i(b) - EPOCH_{i+1}(a)]$ o bien $[EPOCH_i(a) - EPOCH_{i+1}(b)]$.
5. Volver al punto 1 hasta alcanzar un valor de fiabilidad estipulado, en cada una de las características extraídas por separado.

Este procedimiento no altera la distribución de la señal en el tiempo, debido a que la adquisición se produce en un estado continuo de reposo. Además, como resultado del

filtrado de artefactos se introducen discontinuidades de tiempo, debido a la eliminación de epochs alterados con artefactos, obteniéndose una secuencia de epochs con discontinuidades de tiempo entre ellos.

3.4.1. Resultados

En el diseño del algoritmo se determinó como criterio de parada el mínimo valor de fiabilidad obtenido en las potencias de las cuatro bandas frecuenciales de los 19 canales que componen los parámetros de EEG. Se fijó un valor del coeficiente de correlación mínimo de $r_{xx} = 0,8$ (umbral mínimo para estudios clínicos). Es decir, en cada una de las 76 señales de potencias que caracterizan el EEG se obtuvo una fiabilidad de al menos el 80 %. En la figura 3.8 se muestra el resultado de fiabilidad obtenido por participante en cada una de las 76 distribuciones en particular. Como resultado se ha obtenido una fiabilidad media de $r_{xx} = 0,894$.

A continuación, se describen estudios relacionados con la aplicación de este método. El algoritmo se aplicó en cada participante por separado, como resultado se obtuvo unas señales ampliadas diferentes para da participante. La figura 3.10 muestra la ampliación resultante para el conjunto de participantes. El método utilizado fue diseñado con el objetivo de mejorar fiabilidad. Éste debe cumplir el propósito para el cual fue diseñado, sin modificar la distribución de la señal original. En un primer análisis, se comprobó gráficamente la variación que sufre la media para diferentes longitudes de señal. En la figura 3.11 se muestra esta variación para longitudes $N=56, 606$ y 1101 epochs. Además, se realizó un análisis estadístico sobre la distribución de las características antes y después del proceso de ampliación de epochs, mediante el análisis estadístico K-S. La gráfica 3.9 muestra el número de distribuciones de potencia por participante que mantienen su distribución. Se observa unos resultados positivos, de las 76 distribuciones de potencia que componen una participante en la mayoría de participantes se mantiene su 90 % de sus distribuciones.

En resumen, en este capítulo se ha alcanzado unos resultados de fiabilidad sobre los datos de potencias en las bandas frecuenciales : Delta, Theta, Alpha y Beta, para los 19 canales adquiridos por encima del 80 %. Estos resultados demuestran una fiabilidad intrapersonal aceptables en un contexto clínico.

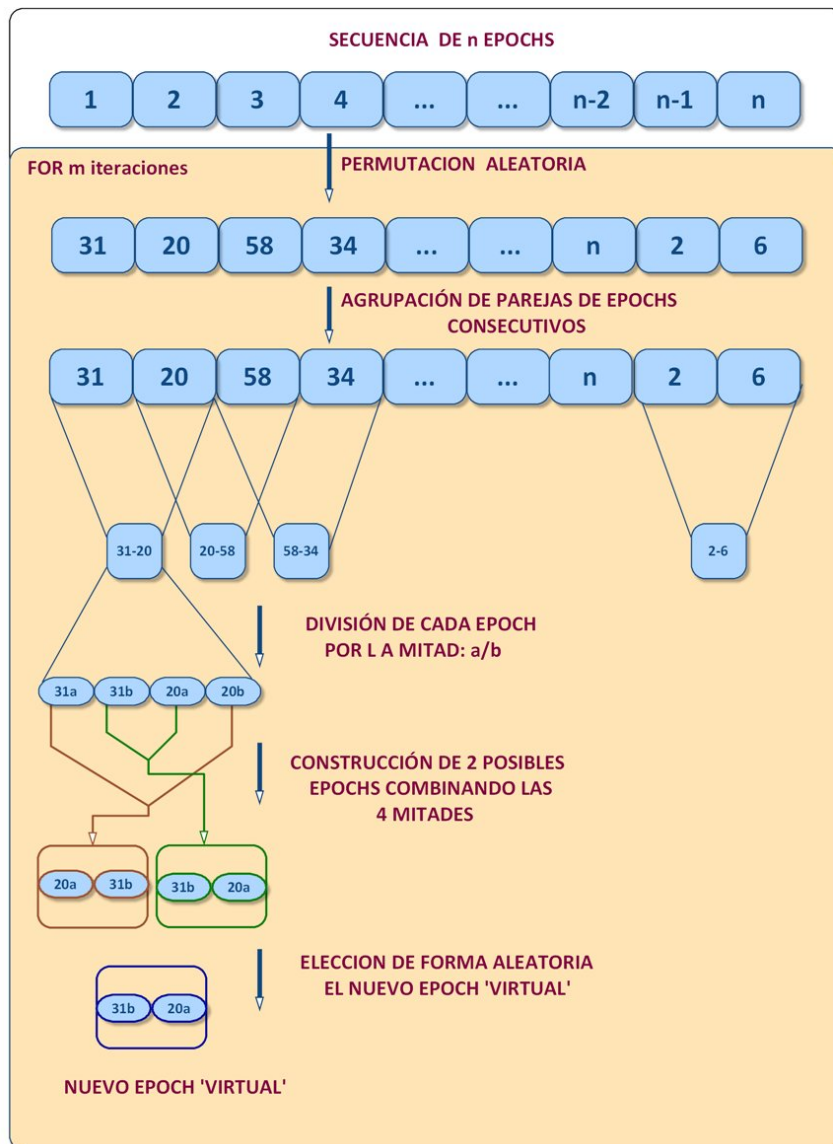


Figura 3.7: Algoritmo para añadir epochs virtuales al conjunto inicial.

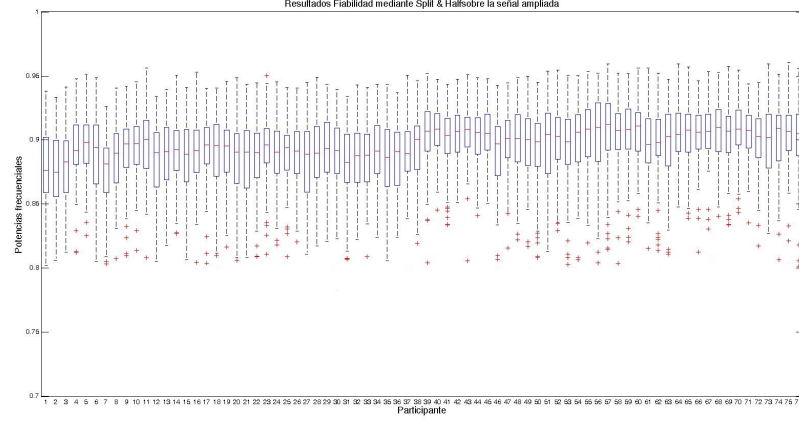


Figura 3.8: Resultados de *Split & Half* para todo el conjunto de participantes. Cada fila representa la potencia banda-canal, de arriba a bajo: $P_D^1, P_T^1, P_A^1, P_B^1, P_D^2 \dots P_B^1 9$

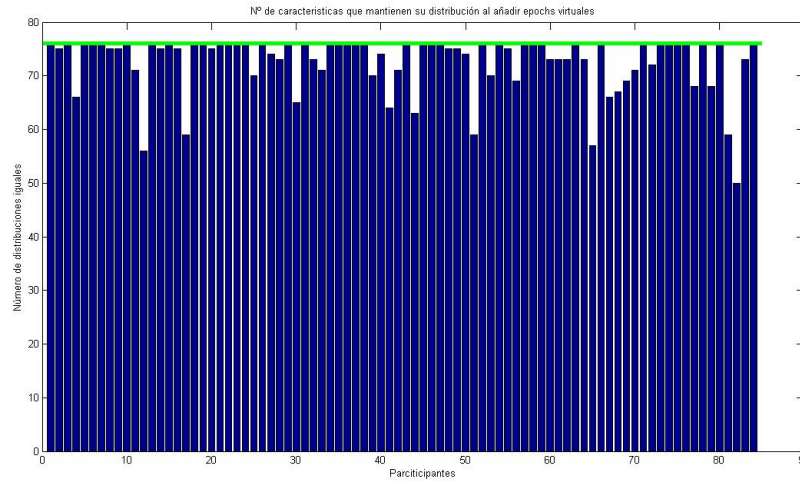


Figura 3.9: Representa el número de variables de las 76 que componen: $\{P_{EEG}\}$ que mantienen su distribución estadística tras la aplicación del método de ampliación mediante *epochs* 'virtuales' en cada uno de los participantes. La línea horizontal en verde muestra la cota máxima de esta gráfica situada en el valor 76

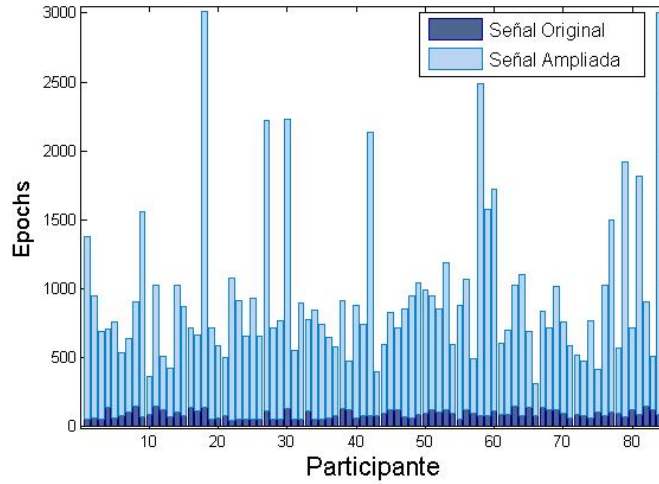


Figura 3.10: Muestra la longitud de la señal EEG en número de *epochs* adquirida en cada uno de los sujetos, junto con el número de *epochs* resultantes tras la aplicación del método de ampliación mediante *epochs* 'virtuales', empleado para conseguir un valor de fiabilidad de *Split & Half* superior a 0.8.

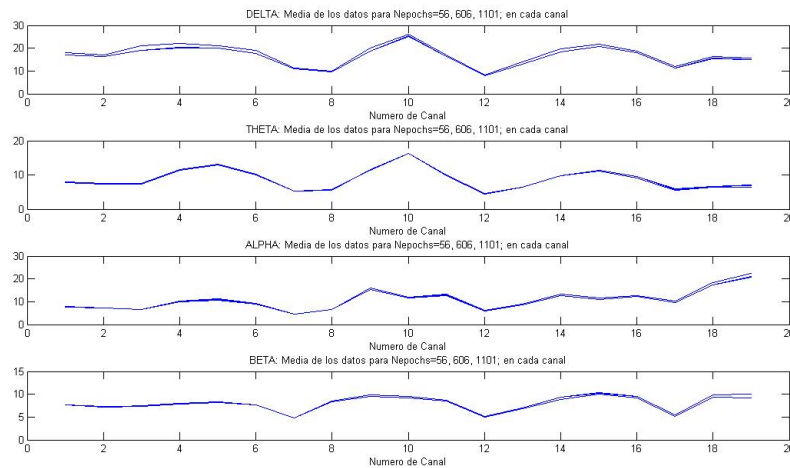


Figura 3.11: Representa la la media en las cuatro variables de potencia en cada uno de los 19 canales, de las siguientes tres señales EEG: el EEG original adquirido en el participante número uno de duración 56 epochs; la señal anterior ampliada con una longitud de 606 epochs y la anterior señal ampliada hasta 1101 epochs. De este modo, se puede estudiar si la media de las distribuciones varían mediante la técnica de ampliación mediante *epochs* 'virtuales'.

4. Análisis Estadístico

Una vez alcanzados unos niveles aceptables sobre la fiabilidad en las características extraídas en la señal EEG, el siguiente proceso consiste en analizar estadísticamente los patrones espectrales extraídos. El objetivo es conocer la función de distribución de las mismas y facilitar en la medida que sea posible los siguientes procesos.

En sección 4.1 se describe la distribución de probabilidad de los datos de partida. Las potencias absolutas en las diferentes bandas frecuenciales. Para ello se realiza un análisis estadístico del periodograma, y en particular para cada una de las potencias absolutas, concluyendo que una distribución que no son modeladas a través de ninguna distribución conocida. Con el objetivo de facilitar los cálculos posteriores y alinearse con los actuales estándares en Bases Normativas [14] es ajustar las distribuciones de potencias a la distribución gaussiana. El proceso de aproximación de cada una de las distribuciones hacia una gaussiana se lleva a cabo en dos etapas: la primera consiste en una transformación de los datos de partida (sección 4.2), y la segunda un filtrado de los datos que acerque la distribución a la gaussianidad mediante un algoritmo genético diseñado ad hoc (sección 4.3). Como resultado, se obtiene para cada participante, un conjunto de 76 distribuciones normales unidimensionales, (definidas por los 19 canales, y 4 bandas frecuenciales).

4.1. Análisis estadístico del Periodograma

Cuando se hace uso de una señal estocástica como el EEG, una de las principales cuestiones es identificar la función de densidad de distribución (PDF, *Probability Density Function*) que permita una caracterización adecuada de la señal. En particular, en este epígrafe se analiza la PDF de las características extraídas a partir de la densidad espectral de potencia del EEG. Estas características están formadas por las potencias absolutas y relativas de las cuatro bandas de frecuencias δ , θ , α y β . Tal y como se describe en la sección 2.4.2, una potencia absoluta es el área representada bajo la curva de PSD en una banda frecuencial. El cálculo de la PA se ha llevado a cabo integrando de forma trapezoidal la función PSD en un intervalo de frecuencias.

El espectro de potencia (PSD) es no negativo por definición y según estimadores espectrales de señales estocásticas no está normalmente distribuido. Ha sido demostrado

teóricamente que la parte real e imaginaria de la DFT de una señal $y(t)$ estacionaria pueden ser considerados independientes y modelados como una distribución gaussiana de media cero. Por tanto, el módulo de la DFT (periodograma) puede ser modelado como la suma de estas variables al cuadrado. Bajo esta asunción cada bin del periodograma $|Y(\lambda, k)|^2$ está distribuido por una función Chi-Cuadrado con dos grados de libertad, $\chi^2(2)$ (Para una información más detallada consultar el anexo C). Esta distribución puede ser interpretada como una distribución exponencial:

$$f_{|Y(\lambda, k)|^2}(x) = \frac{U(x)}{\sigma_Y^2(\lambda, k)} \exp(-x/\sigma_Y^2(\lambda, k)) \quad (4.1)$$

Debido a la ampliación de la señal EEG realizada mediante epochs virtuales, se estudió de forma experimental si los bins de los periodogramas resultantes cumplían la distribución caracterizada por una función $\chi^2(2)$. Para ello, se hizo uso de la prueba de bondad de ajuste de Kolmogórov-Smirnov (K-S) mediante la cual se comprobó la distribución $\chi^2(2)$ en cada bin de frecuencia de 0.5 a 30 Hz.

Una vez caracterizados los *bins* de la PSD, se procede a estudiar la distribución estadística de los patrones espectrales en particular, es decir, de cada una de las potencias en las cuatro bandas frecuenciales calculadas en cada canal EEG. Tal y como se ha indicado anteriormente, cada una de las potencias calculadas, resultan de integrar de forma trapezoidal la curva PSD en una banda frecuencial determinada. La distribución de las potencias no se conoce de forma directa. En consecuencia, el objetivo siguiente se centra en modelar las distribuciones de potencia mediante una distribución conocida.

Con el objetivo de facilitar los cálculos posteriores y alinearse con los actuales estándares en Bases Normativas se ha optado por ajustar las distribuciones de potencias a la distribución gaussiana. En la práctica es muy frecuente buscar la gaussianidad en datos resultantes de una medición. Debido a la premisa básica de que la mayoría de los fenómenos tienen una distribución normal. El teorema Central de Límite es una evidencia teórica sobre este supuesto cuando se suman un conjunto de variables independientes. Aunque no es aplicable en el caso de potencias porque son definidas positivas. Dadao que no se conoce la distribución real de las potencias, se eligió la gaussianidad como distribución objetivo a alcanzar, por su simplicidad y porque facilita la aplicación de numerosas herramientas matemáticas (notar que el segundo objetivo en la construcción de una Base Normativa es proporcionar las herramientas matemáticas para poder comparar sujetos frente a la Base Normativa). Además, la importancia de aproximar parámetros de señal EEG hacia la normalidad ha sido enfatizada anteriormente por varios autores [15],[16]. Para alcanzar la distribución gaussiana se estudió transformar los datos originales, como se verá en la siguiente sección.

4.2. Transformación

El asunto de alcanzar distribuciones normales en un conjunto de observaciones de distribución desconocida, ha sido usualmente resuelto mediante la transformación de los mismos. Esta opción tiene la propiedad de que las conclusiones obtenidas con los datos transformados también se aplican a los datos originales siempre y cuando la transformación sea invertible. Este proceso requiere el conocimiento a priori sobre la distribución de partida para determinar el tipo de transformación más adecuada.

En primer lugar, se analizó gráficamente el comportamiento estadístico de las distribuciones de potencias en cada banda frecuencial. A modo de ejemplo, la Figura 4.1 representa el histograma de la potencia absoluta en la banda Delta en el canal F3 de un participante, de la señal original (izquierda), y de la señal ampliada (derecha). La distribución representada se caracteriza por una fuerte asimetría positiva (las frecuencias más altas se encuentran en el lado izquierdo de la media, mientras que en el derecho hay frecuencias más pequeñas)

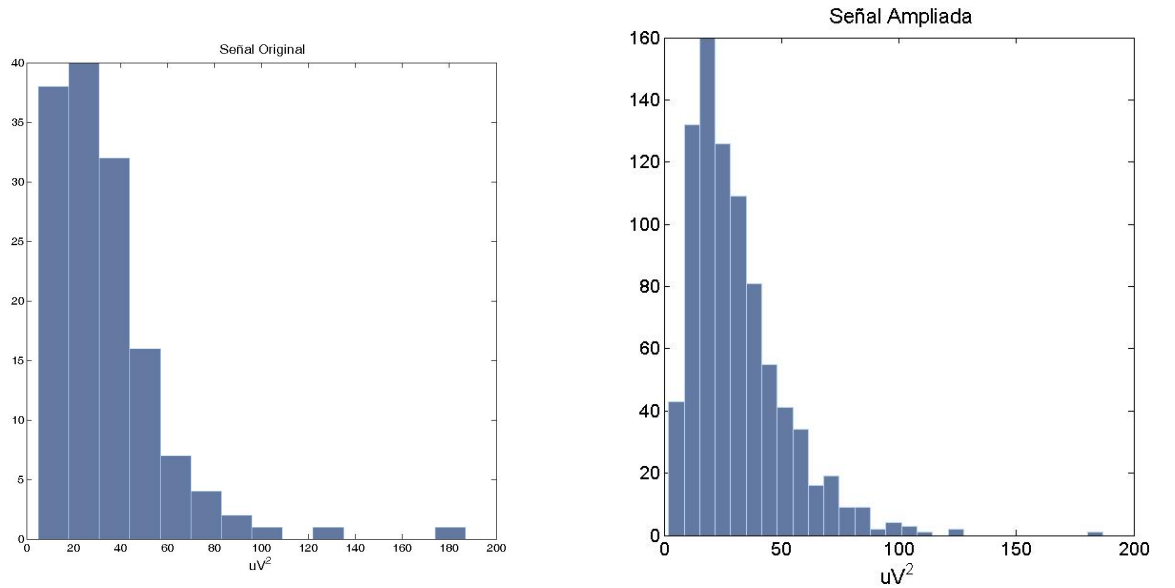


Figura 4.1: El histograma de la izquierda representa la distribución de la potencia absoluta en la banda δ calculada sobre el canal F3 de la señal EEG adquirida en el sujeto número 4. El histograma de la derecha representa la distribución de la potencia en la banda δ calculada sobre la señal ampliada mediante *epochs* 'virtuales' sobre el canal F3 del sujeto número 4.

En conclusión a los resultados obtenidos gráficamente, la función de transformación aplicada sobre estas distribuciones de datos, tiene que corregir la fuerte asimetría positiva con el objetivo de aproximar la distribución a una simétrica como la normal. En la literatura existen diversas funciones de transformación usadas en parámetros de la señal EEG [17]. Para el caso de distribuciones de datos con asimetría positiva se utilizan las

siguientes transformaciones: $\log(x)$, $\log(x+a)$, \sqrt{x} , $\sqrt[3]{x}$, $1/\sqrt{x}$. Algunos autores enfatizan que la distribución más empleada es la logarítmica [15]. Además de las transformaciones nombradas anteriormente que pueden corregir específicamente la falta de normalidad en distribuciones con asimetría positiva existe un método denominado Box-Cox útil en la búsqueda de gaussianidad. El Box-Cox ofrece un método simple para elegir la transformación más apropiada sobre un conjunto de datos para alcanzar la gaussianidad, mediante el método de máxima verosimilitud (se estima el parámetro λ , que define la transformación). La familia de transformaciones viene dada por:

$$y^\lambda = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \ln(y) & \text{si } \lambda = 0 \end{cases} \quad (4.2)$$

En la aplicación de Box-Cox, el valor λ depende de la propia distribución de los datos, de tal forma se obtiene un valor diferente según para los de cada sujeto. Para determinar una única transformación para cada distribución unidimensional de la potencia absoluta en una banda y canal determinado, se estudió la variación del valor de λ_p^i que resulta al aplicar el método Box-Cox en el participante p , en la distribución de potencias i , definida por una banda canal (ver figura 4.2). Para obtener un valor único de cada una de las 76 distribuciones para toda la Base Normativa donde $N=84$ es el número de participantes, se optó por calcular el promedio de los valores λ_p^i resultantes para el conjunto de participantes:

$$\lambda_m^i = \frac{\sum_{p=1}^N \lambda_p^i}{N} \quad (4.3)$$

En consecuencia, se obtienen 76 valores de lambda diferentes, λ_m^i , uno para cada distribución de potencia absoluta determinada por su banda y canal. De este modo, se determina una transformación diferente en cada potencia-banda-canal.

Una vez obtenida la transformación se decide evaluar la proximidad de los datos resultantes a la gaussianidad. El nivel de gaussianidad se cuantificó mediante el test de gaussianidad Anderson-Darling,(AD). El test AD es una prueba de bondad de ajuste que determina si una serie de datos sigue una distribución conocida, en particular se puede aplicar para una distribución gaussiana. Su elección frente al conjunto de pruebas de bondad utilizadas para este propósito se debió a las ventajas demostradas sobre ellas [18]. Este resultado se comprobó experimentalmente por medio de una comparativa entre las transformaciones más populares en estudios anteriores aplicadas a los presentes datos. Los resultados se muestran en la gráfica 4.3, donde se representa la media a través de todos los participantes y la proporción en % de distribuciones gaussianas de las 76 variables descritas por las cuatro bandas frecuenciales y 19 canales. Como se puede observar en la gráfica 4.4, los resultados más favorables en general se obtienen mediante la transformación Box-Cox, dado que aproximadamente consigue que un 50 % de distribuciones para el test de gaussianidad.

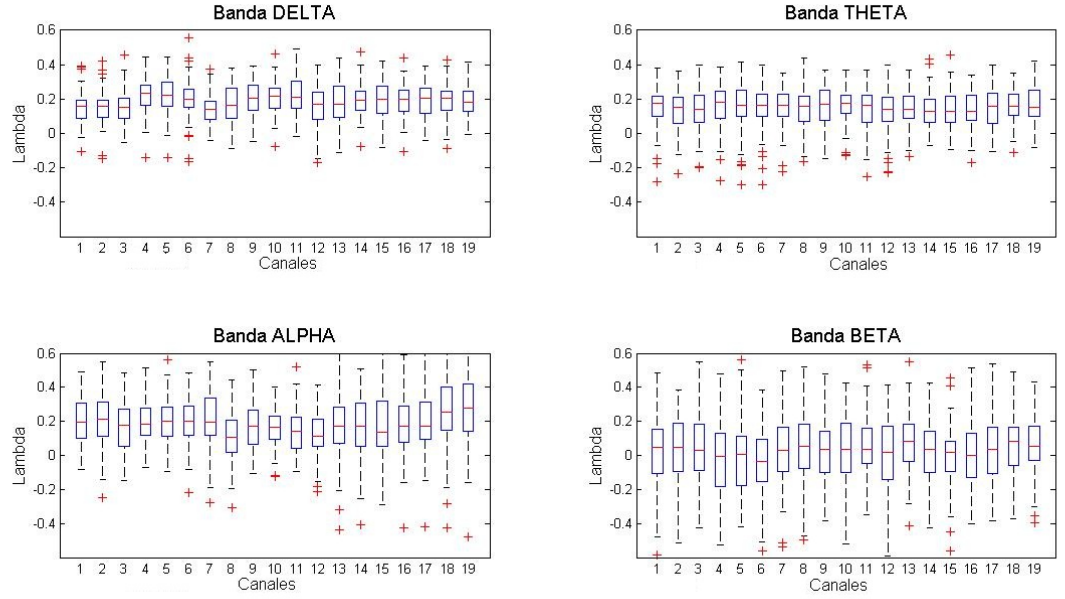


Figura 4.2: Describe la distribución de los valores del parámetro λ en la transformación Box-Cox aplicada sobre las distribuciones de potencias absolutas en las cuatro bandas frecuenciales δ , θ , α y β sobre los 19 canales de señal EEG. La distribución se representa mediante el diagrama estadístico denominado *Box-Plot*. El cual es un gráfico representativo la distribución de la variable ρ calculada sobre cada participante. La caja central indica el rango en el que se concentra el 50 % central de los datos, sus extremos son el primer y tercer cuartil de la distribución y la línea central en la caja es la mediana. Los extremos delimitan el 95 % central de los datos, y las muescas en rojo representan los datos considerados como espurios en la distribución.

En resumen se ha demostrado que la transformación Box-Cox es la más adecuada con el objetivo de mejorar la gaussianidad de los datos. Notar que en esta transformación hay una dependencia de los datos con la transformación paramétrica, de tal forma los 76 valores de λ resultantes son únicos para esta Base Normativa. La ventaja es que la transformación está adaptada a los datos para favorecer la gaussianidad. En el caso de que se deseara añadir un nuevo sujeto a la Base Normativa, estos valores deben ser recalculados.

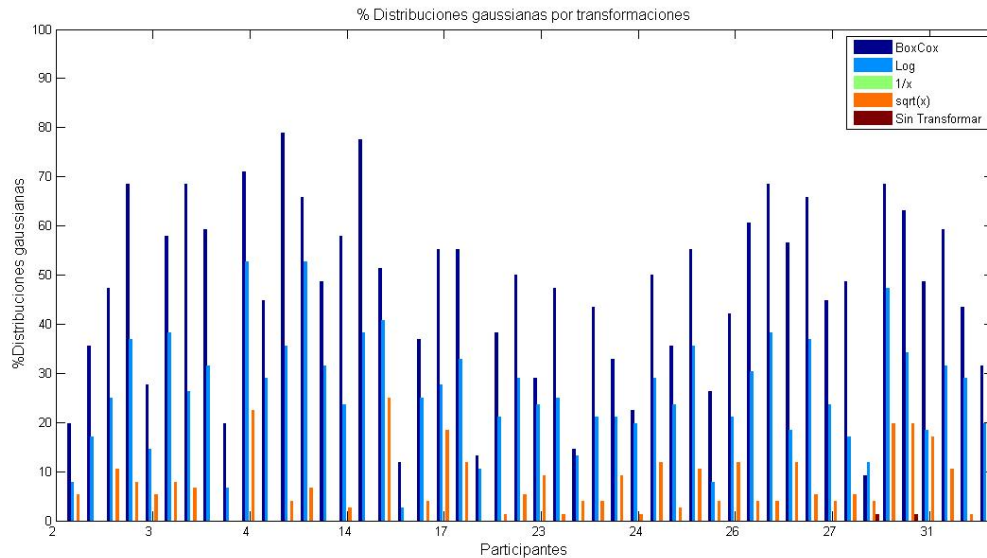


Figura 4.3: Representa una comparativa sobre las siguientes funciones de transformación: (a) familia de transformaciones Box-Cox, (b) $y = \text{Log}(x)$, (c) $y = 1/x$, (d) $y = \sqrt{x}$ aplicadas sobre las 76 distribuciones de potencias absolutas en cada participante. Para cada participante se describe la proporción en % de las distribuciones de potencias con un valor positivo tras aplicar el test de gaussianidad Anderson-Darling.

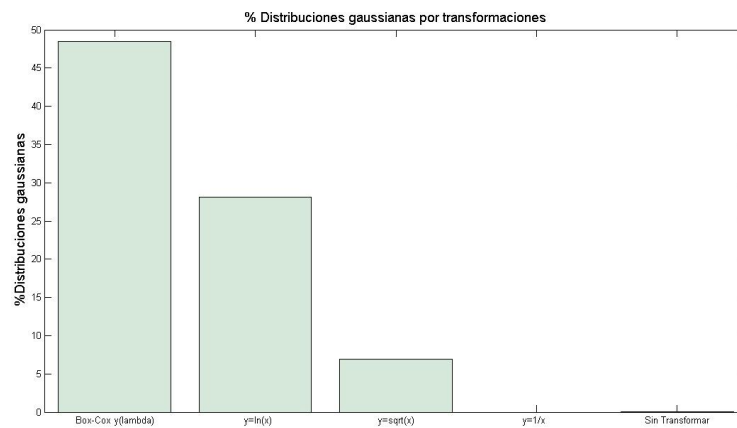


Figura 4.4: Describe una comparativa sobre las funciones de transformación:(a) familia de transformaciones Box-Cox, (b) $y = \text{Log}(x)$, (c) $y = 1/x$, (d) $y = \sqrt{x}$ mediante los resultados medios sobre el conjunto de participantes de la Base Normativa. Se representa el porcentaje medio de distribuciones que han resultado positivas mediante el test de gaussianidad Anderson-Darling.

4.3. Algoritmo Genético

Como se ha detallado en la sección anterior, tras aplicar la transformación Box-Cox se ha conseguido obtener gaussianidad alrededor del 50 % de distribuciones gaussianas de entre las 76 que componen el conjunto de datos de un participante. Sin embargo, el objetivo final es alcanzar gaussianidad en la totalidad de los datos para todos los sujetos que componen la Base Normativa. El problema a resolver tiene como objetivo para cada uno de los participantes seleccionar los *epochs* para los cuales sus valores correspondientes de potencia sigan distribuciones gaussianas (para todas las bandas frecuenciales y cada canal simultáneamente). Es por tanto un problema compuesto por 76 subproblemas unidimensionales.

La estrategia a estudiar debía contemplar la eliminación de datos 'espurios' en términos de gaussianidad (aquellos que empeoran la gaussianidad en la distribución). Es decir, sería necesario la eliminación de aquellos datos que no forman parte de la distribución gaussiana que la caracteriza. En un primer estudio, se propuso eliminar aquellos datos con mayor desviación con respecto a la media para descartar los datos más discordantes en la distribución. Sin embargo, se rechazó esta propuesta porque no asegura alcanzar distribuciones gaussianas. De tal forma se buscó una técnica para filtrar los datos 'espurios' de forma que maximicen la gaussianidad de la distribución resultante. Este problema planteado carece de un método analítico para resolverlo. Para ello, se plantea aplicar algoritmos de búsqueda heurísticos que permitan encontrar la solución al problema propuesto. Entre los métodos de búsqueda heurística se encuentran: el recocido simulado, la búsqueda tabú y las estrategias evolutivas entre otros.

Se eligió aplicar un Algoritmo Genético diseñado ad-hoc. Los Algoritmos Genéticos son algoritmos de búsqueda heurística inspirados en la evolución natural de poblaciones. Los algoritmos evolutivos son métodos robustos de búsqueda que permiten tratar problemas que minimizan o maximizan una función de adaptación (*fitness*). Estos algoritmos operan con una población de individuos $P(t) = \{x'_1, \dots, x'_n\}$ para la iteración t , donde cada individuo x_i se evalúa según una función de adaptación $f(x_i)$. Esta función permite ordenar del mejor al peor los individuos de la población en un continuo de grados de adaptación. La población inicial evoluciona sucesivamente hacia mejores regiones del espacio de búsqueda mediante procesos probabilísticos de: (a) selección de los individuos más adaptados en la población a mayor grado de adaptación mayor probabilidad de dejar descendencia) y (b) modificación por recombinación y/o mutación de los individuos seleccionados. La estructura del algoritmo evolutivo se describe en la Figura 4.5.

Para aplicar este esquema se necesita una codificación o representación del problema que resulte adecuada al mismo. Además se requiere de una función adaptación al problema, la cual determina la capacidad de adaptación de cada individuo. Durante la ejecución del algoritmo, es necesario seleccionar parejas de individuos para la reproducción, a continuación dichos padres seleccionados se cruzarán generando un individuo, sobre el que actuará la mutación. El resultado de la combinación de las anteriores funciones sería un

conjunto de individuos (posibles soluciones al problema), los cuales en la evolución del Algoritmo Genético formarían parte de la siguiente población. Para trasladar el problema presente el problema actual en primer lugar es necesario identificar el modo de codificación de los sujetos que forman la población, además de una función de *fitness* que se desea maximizar o minimizar. Dado que el objetivo de aplicar el Algoritmo Genético es realizar un filtrado de muestras para alcanzar la gaussianidad de los mismos, la función *fitness* vendrá determinada por el nivel de gaussianidad de los datos y un individuo será codificado mediante los valores de potencias absolutas medidas sobre una banda frecuencial y un canal de señal EEG. Notar que se aplicará un Algoritmo Genético por cada sujeto de la Base Normativa. En particular, cada sujeto se caracteriza por 76 distribuciones unidimensionales, en las cuales se desea alcanzar gaussianidad simultáneamente. A continuación se describe el diseño del Algoritmo Genético, la función de *fitness* que sustenta el algoritmo y los resultados finales.

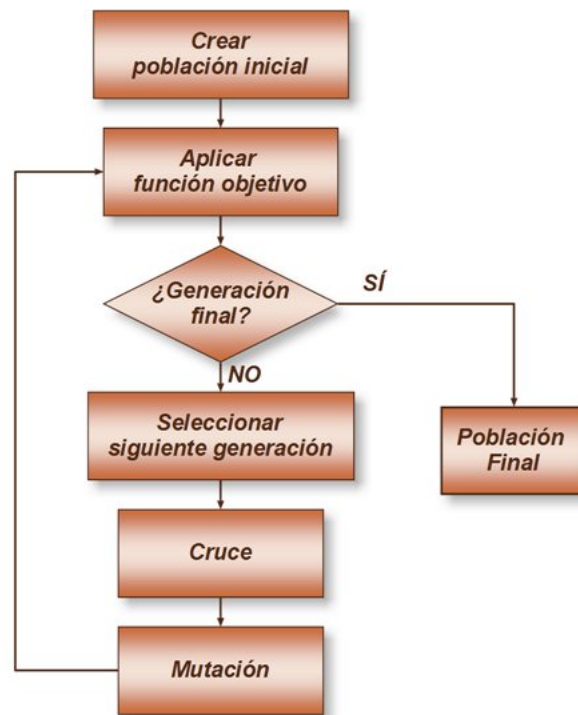


Figura 4.5: Representa la estructura de un Algoritmo Genético general.

4.3.1. Diseño del Algoritmo Genético

4.3.1.1. Codificación

Cada individuo de la población representa las 76 distribuciones de potencias unidimensionales que caracterizan a un sujeto de la Base Normativa en particular. Para ello, el individuo (x'_i) es codificado como una serie de índices $\{i_1, i_2, i_3, \dots, i_N\}$ con $i = 1..N$ (siendo

N el número de epochs de la señal EEG) que determinan la ordenación de los *epochs* de la secuencia original de la señal EEG. Por tanto, la población de individuos $P = x'_1, \dots, x'_n$ está formada por distintas ordenaciones de la secuencia original de *epochs*. La población está compuesta por un número de individuos (M), este valor se definió a partir del número de *epochs* siendo $M = N/2$.

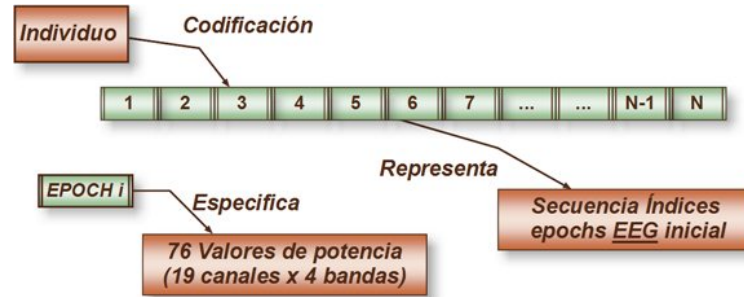


Figura 4.6: Codificación de un individuo de la población mediante una secuencia de índices que determina la ordenación de los *epochs* de la señal EEG adquirida en un sujeto de la Base Normativa.

4.3.1.2. Parámetros

Los parámetros necesarios para el diseño del Algoritmo Genético son:

- Tamaño de la población (M): Número de individuos que forman la población.
- Porcentaje de Elitistas (Nelite): Número de elitistas en la población. Los elitistas son los individuos con mayor capacidad de adaptación. Su adaptación viene cuantificada por la función de *fitness*.
- Porcentaje de supervivencia 'buenos' (Nsuper): Número de individuos 'buenos' que conforman la población. Los 'buenos' son los individuos con capacidad de adaptación por debajo de los elitistas.
- Porcentaje de supervivencia 'malos' (Nmalos): Número de individuos 'malos' que conforman la población. Los 'malos' son los individuos con peor capacidad de adaptación.
- Tamaño del individuo(N): Longitud de *epochs* de la señal EEG, permanece constante durante el algoritmo.
- Porcentaje de cruce (Pc): Porcentaje de individuos a los cuales se les aplicará el operador de cruce.
- Porcentaje de mutación (Nmutación): Porcentaje de individuos a los cuales se les aplicará el operador de mutación.
- Número de mutaciones (Píndies): Número de mutaciones que se producen sobre un individuo.
- Porcentaje de espurios (%Esp): Porcentaje de espurios en relación al número de *epochs* (N).

4.3.1.3. Operadores

Los operadores de un Algoritmo Genético son empleados para obtener la población de la siguiente generación a partir de la población actual. Éstos son: la selección, el cruce y la mutación. A continuación se describe el diseño llevado a cabo para cada uno de ellos.

- **Selección:** La selección consiste en la elección de los individuos de la población para efectuar el operador de cruce y posteriormente la mutación. En este diseño, la selección viene determinada por el concepto de elitismo, que consiste en la idea de que los (*Nelite*) mejores individuos de la actual generación pasen a la siguiente generación sin modificarse por los operadores de cruce o mutación. Por el contrario, los (*Nmalos*) peores individuos de la población no pasarán a la siguiente generación. Sobre el resto de la población se seleccionan de forma aleatoria cada pareja de padres para llevar a cabo el operador de cruce.
- **Cruce:** Crea una generación de individuos nuevos a partir de la información de sus ancestros: padre y madre. El individuo resultante se forma aplicando la siguiente técnica: de forma alternada se copian un índice del padre y un índice de la madre, comenzando por el extremo izquierdo de la secuencia del padre y por el extremo derecho de la secuencia de la madre. De tal forma se va completando la secuencia del hijo resultante, quedando la primera mitad de la secuencia del padre junto con la segunda mitad de la secuencia de la madre (Figura 4.7). Notar que la secuencia resultante no admite la repetición del mismo índice, en el caso de intentar copiar un índice ya incluido en la secuencia del hijo, se procede a copiar el índice consecutivo.

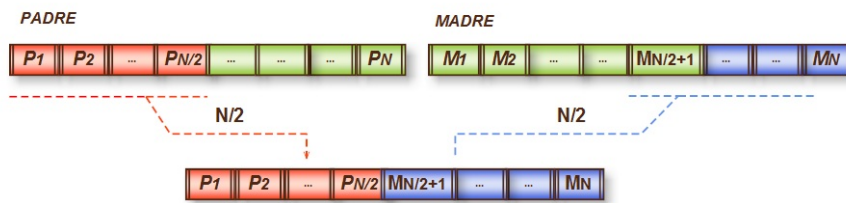


Figura 4.7: Operador de cruce que determina a partir de dos individuos de la población seleccionada un individuo nuevo. Para ello se toma la información de los dos individuos y se copia para el nuevo hijo, tomando la primera mitad de la secuencia del padre y la primera mitad de la secuencia de la madre.

- **Mutación:** Se aplica a cada hijo de manera individual y consiste en el intercambio de posiciones de dos índices seleccionados aleatoriamente (Figura 4.8). El número de mutaciones que se aplica a cada individuo se selecciona de forma aleatoria en cada iteración ($P_{indices}$). Todos los hijos resultantes del cruce tienen la misma probabilidad de mutación, la cual viene dada por el parámetro global *Nmutación*.

Notar que los operadores de cruce y mutación no alteran la longitud de la secuencia que codifica a un individuo, únicamente reordenan los índices de la secuencia original

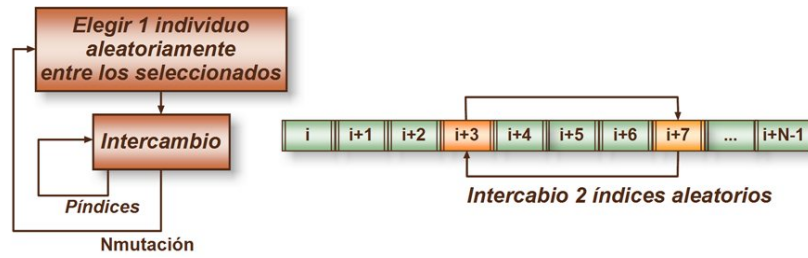


Figura 4.8: Operador de mutación que se realiza sobre los hijos resultantes del operador de cruce. El individuo mutado resulta del intercambio de una pareja de índices seleccionados aleatoriamente. El número de mutaciones que se realiza sobre el mismo sujeto viene determinado aleatoriamente en cada iteración del Algoritmo por *Píndices*.

de *epochs*. El operador de cruce explota las buenas propiedades de los individuos, y sus efectos decrecen con la convergencia del Algoritmo Genético. Por el contrario, la mutación permite escapar de óptimos locales explorando el espacio de búsqueda.

4.3.2. Función objetivo

La función objetivo representa la capacidad de adaptación para cada individuo devolviendo un número real proporcional a su nivel de adaptación. El Algoritmo Genético diseñado tiene como objetivo obtener gaussianidad en las distribuciones de potencias que caracterizan a un sujeto de la Base Normativa, mediante el filtrado de un conjunto de *epochs* de la señal EEG. En primer lugar, es necesario definir un filtro de 'espurios' (desde el punto de vista de la gaussianidad) y a continuación diseñar el cálculo de la función objetivo.

Dado que un individuo es codificado por una secuencia de índices que determina la ordenación de los *epochs* de la señal EEG, el filtrado de *epochs* es aplicado directamente sobre cada individuo de la población. El filtro es representado por una máscara centrada en la secuencia de índices (Figura 4.9). La longitud de la máscara es constante en cada generación de individuos, cuyo valor es igual a la longitud de la secuencia que codifica a cada individuo N (igual al número de *epochs* de la señal EEG) menos el valor de $\%Esp * N$, el cual representa el número de *epochs* que se desean filtrar. Por tanto, a cada lado de la máscara se encuentra la mitad de los índices sin seleccionar $N * \%Espurios / 2$.

El filtrado de índices conlleva un filtrado de *epochs* sobre la señal EEG. A partir de cada *epoch* se ha determinado 76 variables que representan sobre cada canal la potencia absoluta en las cuatro bandas frecuenciales. Por lo tanto, un *epoch* determina una variable multidimensional de 76 dimensiones. En esta sentido, la búsqueda de gaussianidad queda reflejado por la búsqueda de una distribución multigaussiana de 76 dimensiones. Debido al elevado número de dimensiones y por tanto a la complejidad que conlleva alcanzar esta distribución, se decidió simplificar el problema en 76 variables unidimensionales, sin violar la dependencia que existe entre ellos. Es decir, un *epoch* eliminado sobre una distribución



Figura 4.9: Filtro sobre la secuencia de índices. Está determinado por una máscara central sobre la secuencia. El número de índices que no son seleccionados sobre la secuencia viene determinado por $N * \%Espuros$, quedando en los extremos de la máscara la mitad de éstos.

unidimensional de potencias conlleva la eliminación de éste sobre las 75 distribuciones marginales restantes. De este modo se consigue alcanzar gaussianidad simultáneamente en cada una de las distribuciones marginales.

Para cuantificar la gaussianidad sobre la secuencia de *epochs* filtrados se hace uso del test de bondad *Anderson-Darling*, el cual devuelve un p-valor que indica la aproximación a una distribución gaussiana. Un p-valor igual o superior a 0.05 indica gaussianidad a un nivel de significancia del 95 %. El test *Anderson-Darling* es aplicado en cada una de las 76 distribuciones unidimensionales que determina la secuencia de *epochs* filtrados que resultan al aplicar la máscara sobre una secuencia de índices.

La función objetivo se diseñó con el propósito de maximizar la gaussianidad de las distribuciones unidimensionales de potencias, que se traduce a maximizar los 76 p-valores resultantes del test *Anderson-Darling* aplicado sobre cada una de las variables de potencia. En un primer diseño, se determinó como función objetivo la suma de los 76 p-valores para maximizar la gaussianidad de cada una simultáneamente. Esta función maximiza el resultado global ponderando cada p-valor de forma equitativa. Sin embargo, el objetivo de este Algoritmo no es aumentar la suma de los p-valores, sino que cada uno de ellos alcance al menos un valor igual a 0.05 (umbral que determina la gaussianidad al 95 % de significancia). Por tanto, se diseñó otra función que contemplara este valor umbral para cada uno de los p-valores. La función objetivo a maximizar consiste en la suma de los p-valores por separado potenciando aquellas distribuciones con un p-valor por debajo del valor umbral de 0.05. Para ello se diseñó la función objetivo representada en la Figura 4.10, en la cual aquellas variables unidimensionales gaussianas (con un $p\text{-valor} \geq 0,05$) se establece su p-valor igual a un valor constante P . Por el contrario si la distribución no es gaussiana (su $p\text{-valor} < 0,05$) se mantiene su p-valor inicial. El valor de P se fijó por el valor máximo que puede tomar la función objetivo si todas las distribuciones menos una han alcanzado la gaussianidad, es decir $0.05 * 3.75$, en la práctica se tomó $P = 4$.

La convergencia del Algoritmo Genético se determina cuando se obtiene una solución óptima. Por tanto se define un criterio de parada definido por la gaussianidad de las 76 variables de potencias unidimensionales, es decir, cuando el resultado de la función objetivo es $\geq 4 * 76$. Como resultado se obtiene una secuencia de *epochs* 'virtuales' que definen 76 distribuciones de potencias modeladas por una distribución gaussiana unidimensional.

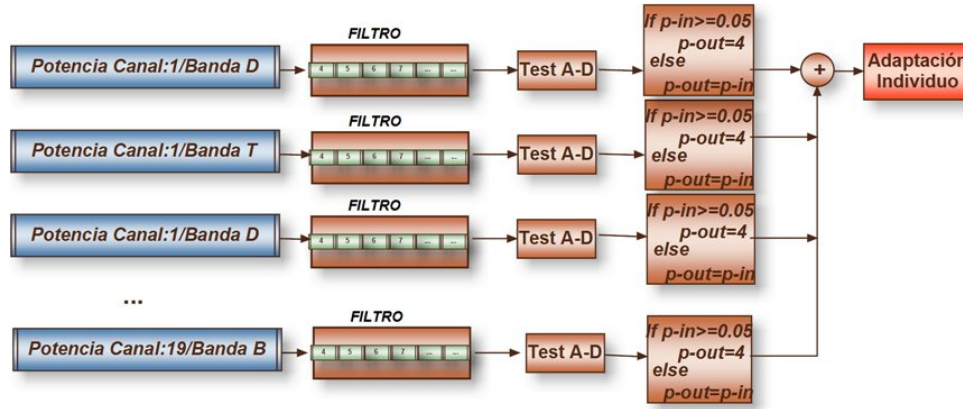


Figura 4.10: Función Objetivo obtenida por el análisis de gaussianidad en cada una de las 76 distribuciones de potencias. La función viene determinada por la suma de los p-valores resultantes al aplicar el test de gaussianidad Anderson-Darling. Si un p-valor es $\geq 0,05$ se fija su valor a 4, si por el contrario es $< 0,05$ su valor permanece constante.

4.3.3. Resultados

El diseño del Algoritmo Genético sufrió diferentes cambios como resultado del estudio de sus parámetros. En particular se estudió la influencia del parámetro $\%Esp$, por su relevancia en los factores como: la velocidad de evolución del algoritmo o tiempo de procesamiento de cada iteración. En las pruebas iniciales se estudiaron valores de $\%Esp$ iguales al 15 %, 25 %, 35 % y 45 %. Además, se analizó la estabilidad del Algoritmo Genético. Para ello se recomputó el algoritmo sobre el mismo sujeto, y se compararon las distribuciones obtenidas mediante el análisis estadístico K-S, para comprobar la consistencia del mismo. Como resultado se obtuvo las mismas distribuciones gaussianas en las 76 variables de potencias, por tanto se concluyó que el Algoritmo Genético diseñado ad hoc resultaba consistente.

Como resultado final se obtuvo una convergencia en los Algoritmos Genéticos en 61 sujetos de los 84 que forman la Base Normativa. Los Algoritmos Genéticos que no han conseguido converger durante la ejecución del algoritmo durante tres meses. Para estos sujetos se podría aplicar una modificación del algoritmo, como podría ser la modificación del parámetro $Nmutación$ que representa la probabilidad de mutación, debido a que aumentando este porcentaje se evita el estado de *stagnation*, problema que ocurre cuando la solución se estanca en un mínimo local. De este modo, se asegura una suficiente exploración del espacio de búsqueda favoreciendo la convergencia del Algoritmo Genético. Esta modificación queda determinada para un trabajo futuro de este proyecto.

5. Clusterización

5.1. Introducción

El presente capítulo propone el desarrollo de técnicas de clusterización para resolver el problema actual que se enfrentan las Bases Normativas sobre la variabilidad interpersonal del EEG. Dicha variabilidad ha sido relacionado con aspectos fisiológicos como la edad del sujeto, y por tanto con propiedades de maduración del volumen conductor del cerebro. En consecuencia, los estudios actuales se centran en determinar la variabilidad del EEG causada por la edad del sujeto, mediante técnicas de regresión por grupos de edades. Los cuales han resultado útiles en el caso de Bases Normativas en adultos, al contrario que en el caso de aplicaciones de Bases Normativas en niños. Este hecho se debe a que en la etapa de niñez la maduración del cerebro varía de forma considerable. A este inconveniente se suma que la variabilidad intrapersonal de EEG no se debe únicamente al factor de la edad del sujeto por lo que los parámetros de regresión han de ser modificados mediante la intervención humana. Como alternativa a los métodos de regresión utilizados actualmente, se propone realizar una estratificación de sujetos de la Base Normativa que tenga en cuenta explícitamente los cambios significativos en su EEG. La técnica de estratificación requiere de unas herramientas matemáticas que determinen un criterio de separación. La clusterización da soporte al uso de estas herramientas matemáticas en la clasificación de los sujetos de la Base Normativa de EEG, ofreciendo un método de separación automática.

En este capítulo se detalla el diseño de la técnica de clusterización basada en cambios significativos del EEG para solventar el problema de variabilidad de la señal EEG en Bases Normativas. La sección 5.2 introduce el concepto de clusterización como una técnica de separación de los datos de una distribución, así como los métodos estudiados y se justifica la selección del método jerárquico como la técnica de clusterización elegida. Esta sección también recoge estudios relacionados con la medida de la distancia del propio método de clusterización, así como el método para seleccionar el punto de corte del gráfico dendograma.

5.2. Clusterización y Métodos

La clusterización permite a partir de una serie de observaciones determinar si existen clases en las que dichas observaciones puedan ser agrupadas. De forma genérica la clusterización es un método de separación no supervisado en el que no se conoce a priori ni el número de clases ni la pertenencia de las observaciones a éstos (en el caso de que se conozca el número de clases el problema se simplifica y si además se conocen las etiquetas que caracterizan cada clase el problema es trivial). El método de clusterización permite resolver la siguiente situación: dado un conjunto de sujetos (de N elementos) caracterizados por la información de n variables X_j , ($j = 1, 2, \dots, n$), se plantea clasificarlos de manera que los individuos pertenecientes a un grupo (cluster) sean tan similares entre sí como sea posible.

El problema de clusterización sobre el desarrollo de una Base Normativa consiste en separar los sujetos de la misma, los cuales están formados por N elementos determinados por el número de *epochs* de EEG de cada uno de ellos. Cada *epoch* está caracterizado por el conjunto de 76 variables de potencias. Por tanto, el problema se presenta mediante una clusterización de sujetos de acuerdo a 76 variables unidimensionales.

Entre las herramientas existentes para desarrollar clusterización se encuentran: métodos basados en particiones, métodos jerárquicos, métodos basados en densidades, etc. Entre los métodos mencionados se estudiaron aquellos elegidos por su aproximación al presente problema. En primer lugar se estudió un método basado en particiones denominado *k-medias*, el cual está basado en la agrupación de observaciones según la proximidad a unos centros de clusters definidos inicialmente, los cuales varían en el proceso de desarrollo. Esta técnica fue desarrollada en cada variable unidimensional, sin embargo los resultados no fueron favorables, dado que los clusters resultantes no eran consistentes sobre las dimensiones. En un segundo estudio, se desarrolló el método de mezclas de gaussianas (en inglés *Gaussian Mixture Model*, GMM). El cual supone que los datos se han generado a partir de una mezcla de k distribuciones multigaussianas. Debido a no poseer variables multigaussianas se realizó una prueba de GMM a partir de una variable de potencias (para una banda y canal determinado), la cual es modelada por una gaussiana unidimensional. Como resultado no se obtuvo una solución consistente, es decir, en cada ejecución de la técnica GMM variaba la agrupación de sujetos resultantes. Para solucionar la falta de fiabilidad de los resultados se decidió aplicar el método de remuestreo *bootstrapping*. En la cual se recomputaba la técnica GMM N veces, y el resultado de la clusterización vino dada por la agrupación con mayor moda de las N repeticiones. Esta técnica se repitió para el resto de variables de potencias absolutas resultando diferentes clusterizaciones. Para evitar este problema se estudió reducir la dimensionalidad de los datos mediante la técnica de síntesis de información: análisis de componentes principales (en inglés, *Principal Component Analysis* PCA). Esta técnica se rechazó porque una reducción de dimensionalidad produce inevitablemente una pérdida indeseada de información.

Finalmente, se decidió la aplicación del clustering ascendente jerárquico por su simpli-

cidad y porque no requiere elección del número de clusters a diferencia de los anteriores estudiados. En el clustering ascendente jerárquico se pretende ir agrupando en cada paso aquellos dos objetos (o conglomerados) más cercanos, para de esta forma ir construyendo una estructura conocida como dendrograma o árbol de clasificación. El dendrograma establece una relación ordenada de los grupos previamente definidos y la longitud de sus ramas es una representación de la distancia entre los distintos nodos del mismo (Figura 5.1) El dendrograma posibilita la obtención de distintas particiones, simplemente variando el nivel de corte de dicha estructura. El diseño del método de clustering esta unido a la elección de una medida de la distancia de similitud entre clusters y a la definición de un umbral que permita cortar el dendrograma para determinar los clusters resultantes.



Figura 5.1: Ejemplo de un árbol de clasificación o dendrograma resultante al aplicar un método de clusterización jerárquico.

5.2.1. Distancia de similitud

La medida de similitud indica la fuerza de la relación entre dos observaciones. La elección de la medida de similitud proporciona el criterio de agrupación entre sujetos de la Base Normativa. A continuación se detallan los estudios realizados para la selección de la medida de distancia para llevar a cabo el clustering ascendente jerárquico.

El primer estudio se realizó en términos de inferencia estadística, mediante la aplicación de una prueba de hipótesis. Entre las pruebas de hipótesis, se barajaron las siguientes pruebas: el Z-test y el contraste para la diferencia de medias de dos poblaciones normales con datos independientes. Para información más detallada consultar el apartado D.2. Sin embargo, ambas métricas fueron rechazadas por su carácter estadístico, cuyos resultados dependían fuertemente de la muestra de población utilizada.

Un segundo diseño contempló el uso de la distancia de Bhattacharyya, la cual mide la semejanza de dos distribuciones de probabilidad discretas y es usada generalmente para medir la posibilidad de separación de clases en la clasificación. Este coeficiente es un número comprendido entre cero y uno que expresa la similitud entre dos histogramas. Concretamente es una medida de aproximación de la cantidad de solapamiento entre dos funciones de distribución. Para una descripción más detallada consultar el anexo D.5.

Para aplicar esta distancia al problema de clusterización presente fue necesario definir una distancia global \overline{DB} , que reuniera la distancia de Bhattacharyya aplicada en cada una de las distribuciones unidimensionales. Dada DB_{ij} , la distancia marginal de Bhattacharyya para el conjunto de potencias del canal i en la banda j , la distancia global viene dada por:

$$\overline{DB} = \sqrt{\sum_{i=1}^{19} \sum_{j=1}^4 DB_{ij}^2} \quad (5.1)$$

Se eligió esta métrica entre otras alternativas como podían ser la mediana, media, o máxima de las distancias marginales, por ser robusta ante espurios además de favorecer la contribución de las distancias marginales DB_{ij} de mayor valor sobre la distancia global \overline{DB} . Es decir, las distancias DB_{ij} con valores altos contribuirán en mucha mayor medida que las de valores bajos.

5.3. Resultados

A continuación se detallan los resultados obtenidos de aplicar el método del *codo* en la selección del punto de corte del dendograma. Además, se describen los resultados obtenidos mediante la clusterización de los sujetos que conforman la Base Normativa.

Se realizó un estudio para determinar el punto de corte del dendograma resultante en la aplicación de método de clusterización ascendente jerárquico sobre los sujetos de la Base Normativa, que define las agrupaciones de los sujetos en clusters. El único criterio para determinar el punto de corte sobre el dendograma es la función que representa la distancia máxima entre clusters frente al número de clusters. Ésta es una función monótona creciente, sobre la cual se podría definir una función de coste. Sin embargo, se propone una alternativa que permite calcular el umbral de forma automática. El método consiste en calcular el *codo* de dicha función. El *codo* de una curva se puede definir como el punto con mayor gradiente dentro de una curva monótona. Sin embargo, en la práctica se trabaja con curvas poco o nada suavizadas, con una gran cantidad de picos y en algunas ocasiones la curva no es monótona en todos sus puntos. Determinar el *codo* de una curva no es un procedimiento trivial, debido a su definición ambigua. Partiendo de un método desarrollado en la literatura [19] se realizaron pequeñas modificaciones para adaptarlo a funciones crecientes. Este método tiene en cuenta la tendencia global de la curva y no únicamente la diferencia entre valores sucesivos en la curva. Dada la función $F(m)$ el *codo* se determina siguiendo el siguiente método:

1. Calcular la diferencia de pendientes consecutivas, teniendo la siguiente función: $DiffFun(m) = F(m-1) + F(m+1) - 2F(m)$, donde $DiffFun$ es la función diferencia.
2. Detectar los n cambios locales significantes en $DiffFun$, seleccionando los n máximos locales de la función $DiffFun(m)$. Una vez localizado estos máximos, ordenarlos de

forma descendente en función del valor de $DiffFun(m)$.

3. Calcular el ángulo que forma la curva $F(m)$ en cada uno de los n máximos locales calculados anteriormente. En la figura 5.2 se muestra el cálculo este ángulo como la suma de α_1 y α_2 siguiendo la expresión:

$$Angle(m) = \arctan(1/|F(m) - F(m-1)|) + \arctan(1/|F(m+1) - F(m)|) \quad (5.2)$$

4. El valor del *codo* viene dado por el primer mínimo de la función $Angle(m)$.

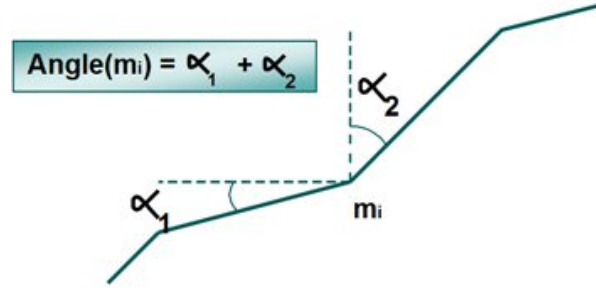


Figura 5.2: Cálculo del ángulo que forma la curva $F(m)$ descrita en el punto 3 del método de cálculo del *codo*.

La Figura 5.3 muestra el resultado de aplicar el cálculo del *codo* sobre la función que describe la distancia entre clusters en función de la distancia entre clusters, resultante en la aplicación del método de clusterización ascendente jerárquico. El *codo* se sitúa en el punto $[3, 8.107]$, éste representa el punto de corte del dendograma en una distancia de $\overline{BC}=8.107$ resultando tres clusters sobre la Base Normativa. Definido el punto de corte a $\overline{BC}=8.107$ sobre el dendograma la Figura 5.4 representa el dendograma resultante en forma circular. Como solución se han obtenido tres clusters representados por los colores (rojo, azul y verde).

En consecuencia mediante la clusterización jerárquica ha sido posible realizar una estratificación sobre los sujetos de la Base Normativa desarrollada mediante los cambios significativos en los parámetros de potencias absolutas de la señal EEG. Como resultado se ha obtenido un total de tres clusters sobre los 61 sujetos que han sido sometidos al método de clusterización.

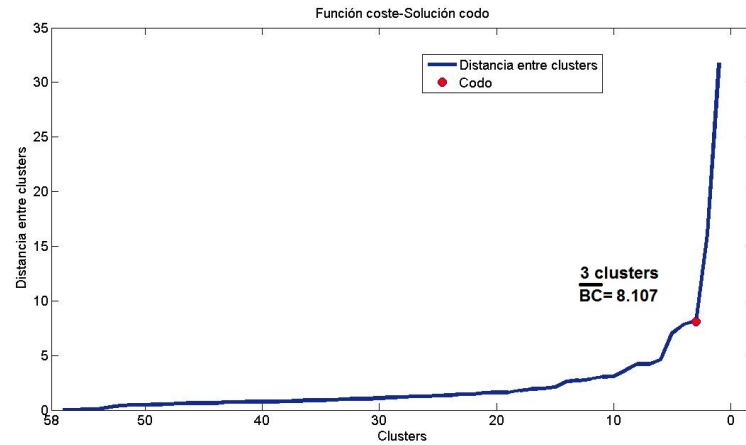


Figura 5.3: Función distancia entre clusters en función de los clusters resultantes del método clusterización ascendente jerárquica aplicado sobre la Base Normativa. El punto en rojo muestra el *codo* de dicha función.

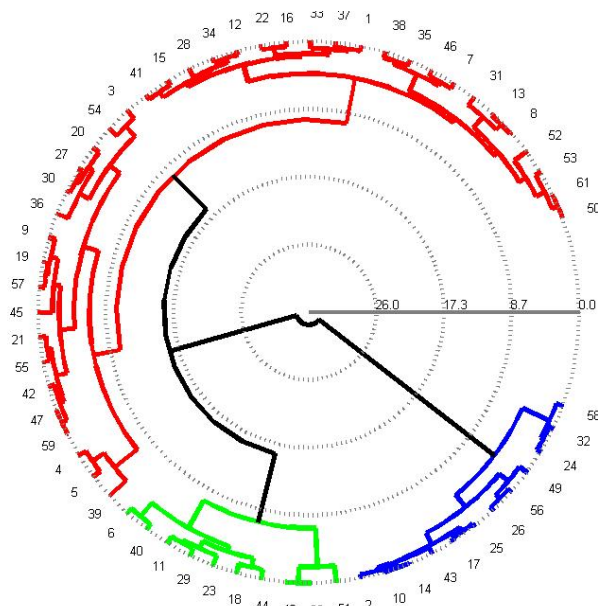


Figura 5.4: Dendrograma circular resultante del método de clusterización ascendente jerárquico aplicado sobre la Base Normativa compuesta por 61 sujetos. El corte de dicho dendrograma se ha realizado sobre el punto 8.107 resultante de aplicar el método del *codo*.

6. Validación

Este capítulo describe el estudio de la validación de la Base Normativa de EEG desarrollada y de las herramientas matemáticas empleadas para determinar los parámetros que caracterizan la normalidad en el EEG. Como resultado se estudia de forma indirecta la validez del conjunto de técnicas desarrolladas en las etapas anteriores en el desarrollo de la Base Normativa.

En la sección 6.1 se realiza un estudio de la fiabilidad del método de clusterización diseñado mediante la técnica de validación cruzada. Además, se detallan los resultados de fiabilidad obtenidos mediante el cálculo de la sensibilidad. En la sección 6.2 se detalla un caso de uso de la Base Normativa mediante datos de EEG cuasi-sintéticos obtenidos directamente de la Base Normativa y modificados con el propósito de simular datos de EEG caracterizados por la patología del Trastorno de Déficit de Atención. De esta forma se analiza la fiabilidad de la Base Normativa aplicada como una prueba de diagnóstico de una patología neurológica.

6.1. Estudio de la Fiabilidad

Para estudiar la fiabilidad de la Base Normativa desarrollada se requiere de la señal EEG adicional adquirida en sujetos sanos bajo las mismas condiciones de registro y aplicarle las mismas técnicas empleadas para el desarrollo de la Base Normativa. Debido a la falta de datos de EEG reales adquiridos en sujetos sanos, se emplean los mismos participantes de la Base Normativa para realizar el análisis de la fiabilidad de la misma. Para ello, se aplica el método de validación cruzada (*Cross-Correlation*), el cual ofrece un resultado robusto para juzgar el funcionamiento de un test como predictor, por su similitud en la evaluación del test en condiciones reales de prueba. La evaluación de este método consiste en desarrollar el clasificador con una muestra del registro total obtenido y evaluarlo con el resto del registro que no se usó para el desarrollo. Dada la reducida longitud de la Base Normativa no es posible dividirla en un conjunto de entrenamiento y uno de test, se ha empleado la técnica *Leave One-Out Cross-Validation*, *LOOCV* (dejar uno fuera). Ésta consiste en quitar un sujeto del conjunto de entrenamiento, entrenar el clasificador con el resto de sujetos y clasificar el que se ha extraído, que constituye el conjunto de test, evaluando si ha sido clasificado correctamente o no. Este proceso se repite para todos los

sujetos del conjunto de entrenamiento cada vez sujeto diferente, de forma que al final se computa el porcentaje de sujetos correctamente.

La aplicación de LOOCV sobre la Base Normativa implica repetir todo el desarrollo de la misma extrayendo un sujeto para compararlo frente a la nueva Base Normativa desarrollada con el resto de sujetos y repetirlo para cada uno de los 61 sujetos que forman la BN. En este sentido, se realizaría una validación global de todo el proceso llevado a cabo. En particular, las primeras etapas como la extracción de características o el cálculo de la fiabilidad intrapersonal son independientes de los sujetos que intervengan en el proceso. Sin embargo, la transformación de los datos mediante la técnica Box-Cox sí depende de los sujetos para determinar la función de transformación (definida por el valor λ). Por tanto, el proceso de LOOCV debería recomputarse a partir de esta etapa. En la práctica, se recomputó únicamente la etapa de clusterización, dado que la variación que supone la eliminación de un sujeto en el cálculo del valor λ no es significativo. Además hay que sumar, la carga computacional y temporal que supone recomputar al Algoritmo Genético para desarrollar 61 Bases Normativas compuestas por 60 sujetos, el número de ejecuciones se elevaría a 3660.

6.1.1. Resultados

A continuación se muestran los resultados obtenidos sobre el estudio de la fiabilidad mediante el método de LOOCV. Como resultado al proceso de LOOCV aplicado sobre el método de clusterización pueden obtenerse dos soluciones: (a) el sujeto excluido enfrentado sobre la nueva Base Normativa no pertenece a la misma, es decir, no pertenece a ninguno de los clusters definidos por la clusterización (*False Negative FN*); (b) el sujeto excluido enfrentado sobre la nueva Base Normativa sí pertenece a la misma, dado que forma parte de uno de los clusters resultantes por la clusterización (*True Positive TP*). La fiabilidad es determinada mediante el índice de sensibilidad dado por:

$$Sensibilidad = \frac{TN}{TN + FP} \quad (6.1)$$

La sensibilidad indica la proporción de sujetos pertenecientes de la Base Normativa que han sido resultantes positivas (pertenecientes a la misma) mediante la técnica de validación cruzada. El resultado obtenido en términos de fiabilidad global resultó *Sensibilidad* = 91,3793 %. Asimismo, se realizó una validación más específica para comprobar la correcta clusterización. Es lo que se denominó *Sensibilidad Parcial* en la cual se verifica si el sujeto se clasifica correctamente en el cluster correspondiente. Sobre los resultados positivos de la prueba anterior se mide el nivel de fiabilidad en cada uno de los clusters. La tabla 6.1 muestra la matriz de confusión sobre los resultados de sensibilidad parcial. Sobre los 91 % de los sujetos que han verificado estar dentro de la Base Normativa (TP) resultantes en el análisis de la sensibilidad global, se ha medido la sensibilidad parcial en cada uno de los clusters, obteniéndose 100 % para los clusters 2 y 3. Dentro del 91 % de los sujetos que han verificado estar dentro de la Base Normativa, todos aquellos pertenecientes a los

	Solución Test		
	Cluster 1	Cluster 2	Cluster 3
Cluster 1	88.88	0	0
Cluster 2	11.11	100	0
Cluster 3	0	0	100

Tabla 6.1: Resultados de la sensibilidad parcial aplicada en la Base Normativa mediante validación cruzada.

clusters 2 y 3 han permanecido en el mismo cluster. Sin embargo, el cluster 1 presenta una fiabilidad del 88.88 %, es decir, el 11.11 % de los sujetos del cluster 1 que verificaban estar dentro de la Base Normativa se identificaron con el cluster 2 en vez del cluster 1. En resumen el método de clusterización diseñado posee un nivel de fiabilidad del 91 %.

6.2. Estudio del Análisis del Diagnóstico

En esta sección se detalla un caso de uso de la Base Normativa que consiste en primer lugar en la obtención de una señal EEG cuasi-sintética adquirida mediante la modificación de la señal EEG de un sujeto sano perteneciente a la Base Normativa. A continuación se emplean las técnicas diseñadas en este proyecto para comparar el EEG patológico obtenido frente a la Base Normativa con el objetivo de verificar la precisión del diagnóstico. Se seleccionó el EEG del sujeto número 4 de la Base Normativa y se modificó el canal Cz multiplicando por un factor igual a 4 la variable que describe la potencia absoluta en la θ , caracterizando un sujeto patológico. Con el propósito de imitar el comportamiento de un EEG bajo la patología del Trastorno de Déficit de Atención (TDAH). Mediante la técnica de clusterización diseñada se realiza una segmentación de la Base Normativa junto con el sujeto patológico para realizar un diagnóstico de éste. Como resultado de la clusterización se obtiene la distancia del sujeto patológico a cada uno de los clusters de la Base Normativa. Dado que cada una de ellas superan el umbral de $\overline{BC}=8.107$, el sujeto patológico queda fuera de la base. En las Figuras 6.1, 6.2 y 6.3 se muestran los resultados de las distancias parciales de el sujeto a los tres clusters. La parte superior de cada figura muestra cuatro matrices, una referida a cada banda frecuencial. Éstas matrices representan las distancias parciales del sujeto patológico a cada uno de los sujetos agrupados en un cluster (por filas) y los 19 canales de EEG (columnas). Gráficamente se puede observar sobre la matriz correspondiente a la banda theta una línea vertical de color rojizo sobre el canal Cz. Esta situación se repite para cada uno de los clusters. De tal forma queda identificada la distribución patológica, indicando la banda y el canal donde se produce.

En consecuencia, se afirma que ha sido posible diagnosticar la patología impuesta sobre el sujeto cuasi-sintético. De esta forma se demuestra la eficiencia de la Base Normativa desarrollada en este proyecto para una aplicación de diagnóstico en la para la patología de TDAH.

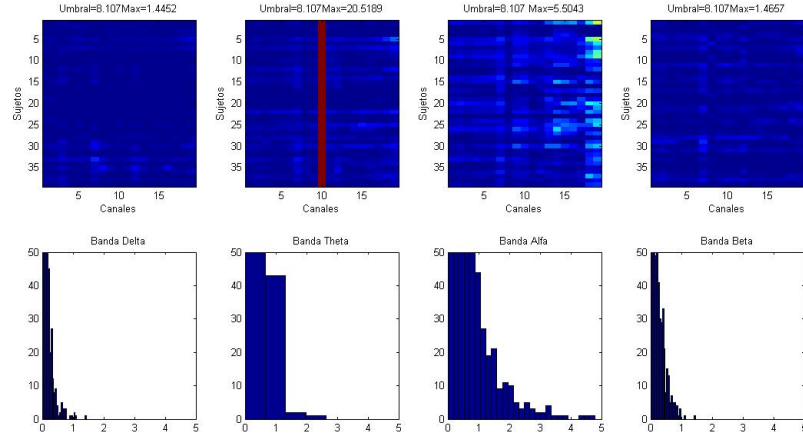


Figura 6.1: Resultados de la prueba de diagnóstico sobre los datos cuasi-sintéticos que representan una patología sobre el canal Cz y banda θ frente al cluster número 1 de la Base Normativa.

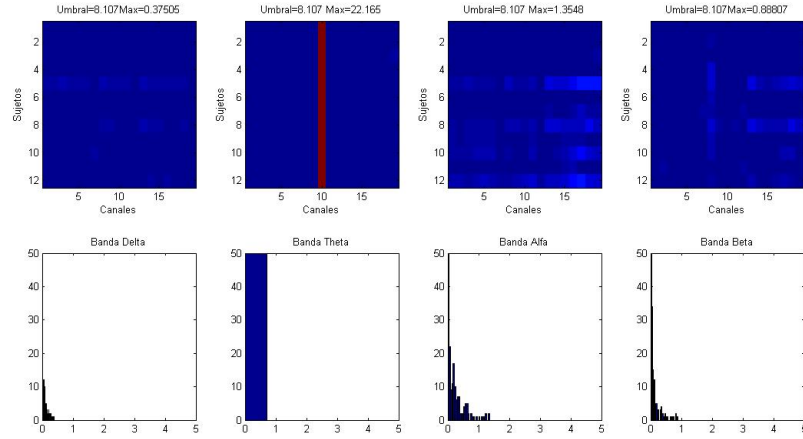


Figura 6.2: Resultados de la prueba de diagnóstico sobre los datos cuasi-sintéticos que representan una patología sobre el canal Cz y banda θ frente al cluster número 2 de la Base Normativa.

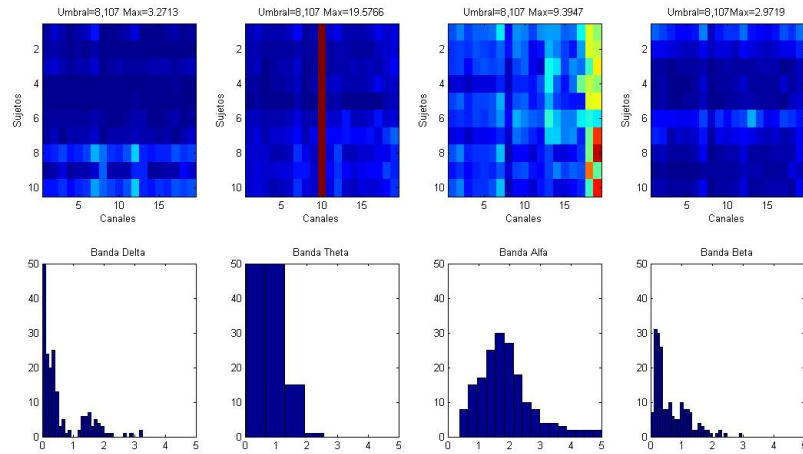


Figura 6.3: Resultados de la prueba de diagnóstico sobre los datos cuasi-sintéticos que representan una patología sobre el canal Cz y banda θ frente al cluster número 3 de la Base Normativa.

7. Conclusiones y líneas futuras

En el presente proyecto se ha diseñado un conjunto de técnicas para el desarrollo de una Base Normativa EEG basal en adultos. Dado que no se ha encontrado constancia de ninguna Base Normativa desarrollada en la misma línea, la comparación no es directa. Únicamente ha sido posible contrastar los resultados en el desarrollo de cada etapa del proceso por separado. Se han estudiado: (a) la fiabilidad intrapersonal en la señal EEG mediante el método *Split&Half* y (b) la normalidad en los parámetros espectrales de potencia de la Base Normativa. Ambos resultados se encuentran en concordancia con la literatura.

Este proyecto propone una alternativa a la estratificación de la señal EEG por edades, la cual está basada en la aplicación de una técnica de clusterización de individuos en función de un conjunto de características de señal extraídas sobre el EEG. La gran ventaja de este método es que permite la automatización completa del proceso de generación de la Base Normativa.

El presente proyecto aporta una nueva técnica en el desarrollo de Bases Normativas EEG, y abre el camino a nuevos trabajos de investigación, entre los que se puede incluir:

- Replicar las técnicas diseñadas en otras métricas de la señal EEG en el dominio temporal tales como la coherencia, fase, simetría.
- Desarrollar una Base Normativa más específica en términos de patología. Determinando una nueva métrica que únicamente recogiera las características más discriminantes (eliminando información redundante) en la clasificación de sujetos sanos y patológicos para una enfermedad en particular.
- Validar la Base Normativa mediante el uso de señal EEG real adquirida en sujetos patológicos.
- Replicar las técnicas diseñadas sobre una población infantil. El EEG en la niñez sufre importantes cambios producidos por la maduración fisiológica del cerebro produciendo una alta variabilidad intrapersonal entre la señal EEG de diferentes sujetos en la edad infantil. En esta línea se podría continuar el trabajo que se llevó a cabo en paralelo con este proyecto, que consistió en el diseño y la realización de experimentos con 45 niños con edades comprendidas entre los 12 y 18 años. Cuyos datos fueron procesados y filtrado de artefactos. Este trabajo no se continuó debido a la dificultad que conlleva el filtrado de artefactos para el desarrollo de una BN.

Bibliografía

- [1] R.W. Thatchert. Normative EEG databases and EEG biofeedback. *Journal of Neurotherapy*, 2(4):8–39, 1998.
- [2] KE. Thornton DP Carmoby. Eyes-Closed and Activation QEEG Databases in Predicting Cognitive Effectiveness and the Inefficiency Hypothesis. *Journal of Neurotherapy*, 13(1):1–21, 2009.
- [3] J. Gunkleman C. Hammond. *The art of artifacting*. Society for Neuronal Regulation, 2001.
- [4] J. L. Hernández P. Valdés R. Biscay T. Virues S. Szava J. Bosch A. Riquenes I. Clark. A global scale factor in brain topography. *International Journal of Neuroscience*, 76(3-4):267–278, 1994.
- [5] J. Wackermann. From the EEG age to a rational scale of brain electric maturation. *Electroencephalography and Clinical Neurophysiology*, 107:415–421, 1998.
- [6] KE. Thornton DP Carmoby. Functional Neuroscience. *Neurometrics: Clinical Applications of Quantitative Electrophysiology*, 2, 1997.
- [7] T. Fernández T. Harmony M. Rodriguez J. Bernal J. Silva A. Reyes E. Marosi. EEG activation patterns during the performance of tasks involving different components of mental calculation. *Electroencephalography and clinical Neurophysiology*, 94(3-4):175–182, 1995.
- [8] Paul L.Nunvez. *Electric fields of the brain*. Oxford University Press, 2006.
- [9] HH. Stassen. Computerized recognition of persons by EEG spectral patterns. *Electroencephalography and clinical Neurophysiology*, 49:190–4, 1980.
- [10] T. Gasser P. Bacher H. Steinberg. Test-retest reliability of spectral parameters of the EEG. *Electroencephalography and clinical Neurophysiology*, 60:312–9, 1985.
- [11] A. Kondacs M. Szabo. Long-term intra-individual variability of the background EEG in normals. *Electroencephalography and clinical Neurophysiology*, 110:1708–16, 1999.