

DPTO. DE TEORÍA DE LA SEÑAL Y COMUNICACIONES  
UNIVERSIDAD CARLOS III DE MADRID



DOCTORAL TESIS FOR THE  
DEGREE OF DOCTOR OF PHILOSOPHY  
NEW APPROACHES FOR EEG SIGNAL PROCESSING:  
ARTIFACT EOG REMOVAL  
by ICA-RLS SCHEME  
AND  
TRACKS EXTRACTION METHOD

Author: Carlos Andrés Guerrero Mosquera  
Supervisors: Dr. Ángel Navia Vázquez  
Dr. Armando Malanda Trigueros

LEGANÉS, OCTOBER 2011



TESIS DOCTORAL:

New approaches for EEG signal processing: Artifact EOG removal by ICA-RLS scheme and Tracks extraction method

Autor:  
Carlos Andrés Guerrero Mosquera

Directores.  
Dr. Ángel Navia Vázquez  
Dr. Armando Malanda Trigueros

Firma del Tribunal Calificador.

Presidente:

Vocal:

Vocal:

Vocal:

Secretario:

Calificación:

Leganés, a



## Agradecimientos

En primer lugar me gustaría agradecer a los profesores Aníbal Figueiras y Ángel Navia Vázquez por la oportunidad que me dieron de poder conocer, trabajar, aprender, mejorar y vivir una etapa de mi vida en la Universidad Carlos III de Madrid. Sin vuestra confianza no hubiera sido posible el haber llegado a Madrid. De igual manera quiero agradecer todo el apoyo recibido en estos años por parte del profesor Armando Malanda, a quien le debo de igual forma mucho.

Aníbal, quiero extender mi respuesta a tú adecuada pregunta: seguiré investigando y estudiando. Ángel, gracias por secundar mis experimentos. Armando, gracias también por esa ayuda que me diste cuando se me era importante un diciembre. Te informo que estaré siempre agradecido por tu incondicional apoyo.

A los tres: ¡muchas gracias de corazón!

Esta tesis (como muchas) tiene una historia personal ligada. Han pasado muchos años desde que dejé mi Iruña para coger rumbo al Reino Unido. Recuerdo que ese viaje lo hice con un ánimo derrotado y con ganas de regresar a Colombia, dando por hecho que mi trecho de investigador caducaba. Ahora que escribo estas líneas y que ha pasado la tormenta, quiero agradecer a toda mi familia, desde mi abuela hasta la última bisnieta. Gracias Ofe, MariaPao, mamá, tias, tios, primas y primos.

Obviamente todos estos años he estado rodeado de personas muy especiales a quien quiero dar las gracias y espero no omitir a nadie:

A Ainarica, navarrica, a la que quiero mucho ya que me apoyó en todos los momentos difíciles y me acompañó en un 90% del doctorado: Eskerrik asko denagatik Ainara, muchas gracias por todo ese apoyo. Te merecés todo lo lindo que pueda suceder en la vida. Emozioz eta musikaz beteriko bizitza opa dizut eta maite zaitut.

A mis primeros colegas en Pamplona que me acompañaron y me obsequiaron su amistad. Gracias a ellos, ese primer invierno europeo no me fué tan frío bajo ese abrigo fraternal: Juliantxo Parada, Heller, Ignacio Rodriguez y Armando. Os recuerdo un montón.

A Aurelie Viry por todo su amor, apoyo y confianza en mí. Gracias por

esos meses en Paris y ese par de años al lado mío.

A mis amigos de Edinburgh, con quien compartí cuatro estaciones de ilusión: Iván Charvert, Akshay, Stephanie, Pavlos “malaka” y Batiste. Aún recuerdo nuestras charlas en la cocina observando esa lluvia no pausada de Escocia.

A mis segundo grupo de colegas en Pamplona que conocí a mi vuelta del Reino Unido. Aquel piso de Ermitagaña...Andrés Zazpe con su “cencerrín”, Imanol “el artista”, Antonio “el triki”, Patxi “Francesco”, Mario, Javi, Víctor, Plutín y a toda la cuadrilla de chicas y chicos que me alegraron de igual forma los fines de semana en Pamplona: Clara Zunzarren, Maribel, Jess, Maki, Joana, Ana, Borja, David, Iñigo...¡Gora San Fermín y Aupa Osasuna!

A esas primeras personas que me recibieron con una sonrisa en mi periplo en la UC3M. Vanessa, quien me ayudó mucho con TDS y en llevar de mejor proceder el trabajo con Aníbal; Manu con quien tuve muchas charlas interesantes; Javi y Jaisiel quienes me hacían reír un montón; Azpi que me alegra hasta hoy las largas jornadas del laboratorio; Carlos Martínez con quien compartí esos cafés tempraneros; Marilight por enseñarme el mundo del flamenco y darme mucho cariño; Rocío por su ayuda en trámites burocráticos tan ineludible; Sara por su siempre sonrisa; Edu por sus apariciones súbitas muy “spicy”; Rosa y sus temas tan profundos; Rubén y Sergio por su compañía “desde el otro lado” estos últimos fines de semana; y a Darío, a quien echo en falta ahora que está por Australia.

Quiero mencionar a los profesores del departamento de quienes me llevo un recuerdo especial: Emilio Parrado por sus charlas de pasillo tan interesantes y jocosas. Harold por su eficiencia. Jerónimo por su funcionalidad. Jesús por su brillantez. David Luengo por nuestro PFC. Antonio Artés por su amabilidad. Mati por su super tobleron que siempre regalaba y Marcelino por su energía. Gracias a todos.

Al profesor Michel Verleysen por enseñarme a llevar ecuaciones teóricas a la praxis y por toda su amabilidad que me ofreció en Louvain.

A Jorge Iriarte, neurofisiólogo de la Clínica Universitaria de Navarra. Gracias por toda esa sabiduría y paciencia que tuvo para conmigo.

A los dos revisores externos que gentilmente dieron su visto bueno a la tesis: Gaël de Lannoy y Lars Kain Hansen, Thank you very much!

A los del grupo barrita con tomate: Luis, Soufi, Efra, Robert, Adil y

Marta. Gracias por vuestras historias, chistes, sonrisas que viene tan bien cuando se está investigando, pero en especial quiero agradecer a la Sita Gestora de Proyectos, Marta, que ha estado ahí, dándome su luz y enseñándome tanto. Si soy ahora un poco mejor, es gracias a ella.

A mis amigos foráneos como yo: Diego F., Julián, Neila, Ricardo, Marilea, Darwin, Jair, Leo y mi Moysen. Sin vosotros hubiera perdido muchas cosas de mi País y de Latinoamérica. Gracias de corazón. Moysen, te agradezco por esa sabiduría espiritual que me das y toda esa alegría que transmites. DTB.

A Ingrid por esa fuerza que desde Colombia me ha dado y por todo su amor.

A mi Dios@, que espero sea vuestro@, llámese como se llame. Lo importante no es creer en su existencia sino en que EL/ELLA crea en la vuestra.

A todos, muchas gracias.



# Abstract

Localizing the bioelectric phenomena originating from the cerebral cortex and evoked by auditory and somatosensory stimuli are clear objectives to both understand how the brain works and to recognize different pathologies. Diseases such as Parkinson's, Alzheimer's, schizophrenia and epilepsy are intensively studied to find a cure or accurate diagnosis.

Epilepsy is considered the disease with major prevalence within disorders with neurological origin. The recurrent and sudden incidence of seizures can lead to dangerous and possibly life-threatening situations. Since disturbance of consciousness and sudden loss of motor control often occur without any warning, the ability to predict epileptic seizures would reduce patients' anxiety, thus considerably improving quality of life and safety.

The common procedure for epilepsy seizure detection is based on brain activity monitorization via electroencephalogram (EEG) data. This process consumes a lot of time, especially in the case of long recordings, but the major problem is the subjective nature of the analysis among specialists when analyzing the same record. From this perspective, the identification of hidden dynamical patterns is necessary because they could provide insight into the underlying physiological mechanisms that occur in the brain.

Time-frequency distributions (TFDs) and adaptive methods have demonstrated to be good alternatives in designing systems for detecting neurodegenerative diseases. TFDs are appropriate transformations because they offer the possibility of analyzing relatively long continuous segments of EEG data even when the dynamics of the signal are rapidly changing. On the other hand, most of the detection methods proposed in the literature assume a clean EEG signal free of artifacts or noise, leaving the preprocessing problem opened to any denoising algorithm.

In this thesis we have developed two proposals for EEG signal processing: the first approach consists in electrooculogram (EOG) removal method based on a combination of ICA and RLS algorithms which automatically cancels the artifacts produced by eyes movement without the use of external "ad hoc" electrode. This method, called ICA-RLS has been compared with other techniques that are in the state of the art and has shown to be a good alternative for artifacts rejection. The second approach is a novel method in EEG features extraction called tracks extraction (LFE features). This method is based on the TFDs and partial tracking. Our results in pattern extractions related to epileptic seizures have shown that tracks extraction is appropriate in EEG detection and classification tasks, being practical, easily applicable in medical environment and has acceptable computational cost.



## Resumen extendido en español

En este resumen se pretende dar a conocer los objetivos de esta tesis así como también una breve descripción de la metodología empleada y la organización del documento. De igual manera detallaremos las aportaciones originales, conclusiones y el trabajo futuro que esta tesis pueda originar.

### Introducción

A día de hoy, la neuroingeniería es una nueva línea de investigación que emerge de la integración entre la neurociencia clínica y la ingeniería. Este nuevo campo utiliza la ingeniería, la simulación computacional, el análisis matemático, las técnicas de imagen y el diseño hardware para solucionar diferentes problemas en la neurociencia clínica. El objetivo es utilizar las nuevas propuestas en la ingeniería para descubrir los principios de la función neuronal y utilizarlos en el diseño de sistemas de soporte al diagnóstico médico y terapéutico. La neuroingeniería clínica tiene una gran variedad de campos activos con objetivos tales como potenciar tratamientos basados en herramientas de ingeniería y crear nuevos enfoques para el diagnóstico de enfermedades neuronales, tal y como describe [Thakor and Tong \[2006a\]](#).

Descubrir e interpretar los fenómenos bioeléctricos procedentes de la corteza cerebral y la evocada por estímulos auditivos y somatosensoriales, tienen como finalidad intentar entender cómo funciona el cerebro y reconocer sus diferentes patologías. Enfermedades como parkinson, demencia, alzhéimer, esquizofrenia y epilepsia son motivo de análisis con el objetivo de encontrar una cura o un diagnóstico médico preciso.

En cuanto a la epilepsia, actualmente la investigación está orientada a aislar, identificar y localizar los focos epilépticos empleando generalmente señales electroencefalográficas (EEG), procedimiento previo a un posible tratamiento quirúrgico en los casos en que el medicamento no ha sido capaz de detener la enfermedad. Otra posibilidad para los casos resistentes a medicamentos es la detección temprana de crisis por métodos automáticos basados en señales EEG. Detectar o anticiparse a una crisis, son problemas de gran interés clínico y con alta complejidad teórica y práctica, para lo cual sólo existen hasta hoy soluciones parciales probadas en un número pequeño de pacientes o en bases de datos no muy grandes.

La mayoría de los métodos de detección de epilepsia con señales EEGs, suponen una señal limpia de artefactos, o proponen un filtrado de la señal basado en métodos convencionales tales como el filtrado temporal o métodos adaptativos temporales, siendo habitual el uso de señales de referencia como el electrooculograma (EOG). Por otra parte, dentro del procesamiento del

EEG hay otro proceso importante relacionado con la información que se va a extraer de la señal, procedimiento conocido como extracción de características. Se ha de resaltar, que esas características extraídas del EEG dependen directamente del método empleado, que por lo general son transformaciones de la señal a otros dominios con el objetivo de facilitar la extracción de “información oculta” en el EEG. Dichas transformaciones nos ofrecen subconjuntos de características que pueden ocasionar problemas en el desempeño de un detector, porque existe la posibilidad de extraer características parecidas o no relevantes. Por esa razón, una adecuada selección de características es de igual manera tan importante como la extracción de estas, ya que este proceso elimina las redundantes que pueden ser consideradas ruido introducido al sistema.

En el análisis de señales EEG, las distribuciones tiempo frecuencia (TFD, del inglés Time-Frequency Distributions) y los métodos automáticos de eliminación de artefactos, han demostrado ser alternativas a seguir en el diseño de sistemas de detección automática de enfermedades neurodegenerativas. Las primeras son transformaciones apropiadas para señales no estacionarias y dinámicas como el EEG, y los segundos evitan el proceso de una inspección visual de formas de ondas o de energía espectral (proceso poco preciso y costoso que se basa en un análisis subjetivo de las señales y que además hace necesaria la figura de un especialista en el equipo de trabajo).

En esta tesis se proponen dos algoritmos para el procesado EEG: uno de ellos está orientado a la extracción de características empleando las TFDs y el otro es un método de eliminación automática de artefactos producidos por el movimiento de ojos. Ambas propuestas han sido evaluadas frente al estado del arte actual y han sido probados con diferentes bases de datos. Los resultados son prometedores para el ambiente clínico ya que son prácticos y fiables.

## Motivación

A lo largo de los últimos años, los departamentos de Teoría de la señal y Comunicaciones de la UC3M, Ingeniería Eléctrica y Electrónica de la UPNA, y el Departamento de Neurología y Neurocirugía de la Clínica Universitaria de Navarra vienen colaborando en sucesivas investigaciones centradas en descubrir e interpretar los fenómenos bioeléctricos originados en la corteza cerebral, con el objetivo de aplicar estos estudios al análisis de distintas patologías, como la enfermedad del Parkinson, Alzheimer o la epilepsia.

Las anomalías que ocurren en la corteza del cerebro pueden clasificarse, de forma general, según su etiología (por ejemplo, trauma, enfermedad, toxinas o infecciones). Aunque no siempre existe una respuesta eléctrica cere-

bral fácil de detectar, la mayoría de las enfermedades neurológicas se manifiestan desde niveles celulares y moleculares con actividad magnética medible fuera del cuero cabelludo, lo cual permite localizar la zona dañada del cerebro, la causa y su posible función fisiológica.

La actividad neuronal en el cerebro humano se inicia desde las primeras etapas de desarrollo prenatal. Como ya se sabe, las señales generadas por el encéfalo son eléctricas en naturaleza y representan no solo la función cerebral, sino también el estado del cuerpo entero. Una manera de registrar los potenciales eléctricos generados en el encéfalo -ya sea de forma espontánea o evocada por algún estímulo- es a través del electroencefalograma (EEG), exploración que proporciona estimaciones de la acción sináptica a grandes escalas y relacionadas con el comportamiento y la cognición.

Un caso particular de registro inmediato en el EEG frente a un “estallido” repentino y asíncrono de flujo de corriente eléctrica entre neuronas, sucede con la epilepsia. Existen formas de ondas cerebrales clasificadas que permiten diagnosticar la clase de epilepsia además de su localización en el cerebro. El análisis del EEG ha demostrado asimismo cambios de altas a bajas frecuencias en zonas del cerebro de pacientes afectados de alzhéimer.

El EEG tiene información que depende tanto del procesamiento como del momento en que se hace el registro. De igual forma requiere registros con largos períodos de tiempo, personal especializado, bases de datos con alta capacidad de almacenamiento e implica una alta carga computacional.

Los primeros resultados han mostrado la necesidad de limpiar la señal EEG de una forma eficiente y de reducir la cantidad de información debido a dos razones fundamentales: (i) el EEG es una señal multi-canal ya que se registra desde electrodos situados en diferentes posiciones estandarizadas sobre la cabeza; (ii) el EEG puede convertirse a una señal multi-dimensional, sobre todo cuando se aplica un método de extracción de características. La extracción de información compromete una transformación de la señal a otros dominios con el objetivo de obtener un número de coeficientes representativos del problema. Cada coeficiente es una dimensión adicional que se agrega al problema y que afecta directamente al coste computacional.

En relación con las tareas mencionadas anteriormente, los algoritmos basados en aprendizaje máquina ofrecen soluciones viables para este tipo de señales en áreas de detección, clasificación, extracción y selección de características, con una variedad de soluciones y de alternativas en el procesado EEG.

Ya en el ambiente médico, los fundamentos para un diseño de análisis, detección o clasificación EEG son simples: (i) una señal de EEG se registra, (ii) la señal EEG se limpia utilizando métodos de eliminación de ruido y algoritmos de rechazo de artefactos, (iii) se extraen características del EEG por algún método o transformación de la señal y finalmente, (iv) las características relevantes son seleccionadas.

## Objetivos

Inicialmente se ha propuesto el siguiente objetivo general para esta tesis: diseñar un método de detección automática y de análisis para señales EEG con epilepsia. El diseño ha de ser práctico, sencillo y fiable para ser aplicado en el entorno médico. Para acercarnos a este objetivo general, se hace necesario entonces llevar a cabo diferentes tareas específicas:

- Recolección de datos en el hospital y búsqueda de bases de datos utilizados por otros autores.
- Eliminar artefactos y ruido presente en el EEG empleando el análisis de componente independiente (ICA, del inglés Independent Component Analysis).
- Diseñar un sistema de rechazo de artefactos en el EEG.
- Analizar las diferentes propuestas en TFDs para señales EEG con epilepsia.
- Proponer un método sencillo de extracción de características basado en las TFDs.
- Comparar el desempeño de las características propuestas para tareas de detección y clasificación EEG.
- Aplicar métodos de selección de características para reducir el coste computacional.
- Diseñar un sistema de detección automática de epilepsia basado en señales EEG.

## Organización de la tesis

El capítulo 2 sirve de introducción a los conceptos fundamentales sobre el sistema nervioso y la generación de la señal EEG. Se describen además los conceptos relacionados al análisis visual EEG, artefactos, características anormales del EEG y las diferentes ondas cerebrales. Este capítulo también incluye las diferentes aplicaciones que se pueden desarrollar empleando las

señales EEG, como son el soporte al diagnóstico médico, detección de enfermedades neurodegenerativas, detección de los potenciales relacionados a eventos (ERPs, del inglés event related potentials) y los trastornos del sueño.

En el capítulo 3 se introduce el modelo de la señal EEG. Esto incluye revisar conceptos de segmentación de la señal, filtrado y eliminación de ruido, extracción de características y selección de las mismas. El modelo de la señal se aborda desde el punto de vista físico y no lineal. En la sección de segmentación de señales, se explica la forma de separar segmentos para señales invariantes en el tiempo y para señales variables en el tiempo. En la sección de filtrado y eliminación de ruido se hace una breve introducción a las diferentes técnicas en la limpieza del EEG como los filtros digitales, los métodos de filtrado adaptativo y el análisis de componentes independientes (ICA). También este capítulo introduce los métodos de extracción de características y de selección.

En el capítulo 4 se describen las dos nuevas propuestas que esta tesis aporta al procesado de señal EEG. Una de ellas es un nuevo método de eliminación de artefactos producidos por los ojos basado en un esquema ICA-RLS. La otra propuesta es un nuevo método de extracción de características que emplea el plano tiempo-frecuencia de las distribuciones tiempo-frecuencia (TFDs) para extraer información.

ICA-RLS tiene dos características importantes: la primera, no utiliza señales de referencia dedicadas, como por ejemplo, el electrooculograma (EOG); y la segunda, el filtrado adaptativo se hace en el dominio de las fuentes ICA. Respecto al método de extracción, este capítulo describe las tres características que se extraen : duración ( $L$ ), frecuencia ( $F$ ) y energía ( $E$ ) de un parcial principal que se sigue sobre el plano tiempo-frecuencia. Para finalizar, el capítulo describe la complejidad computacional empleada por los métodos propuestos, ya que usualmente estas órdenes de complejidad no suelen reflejarse en el estado del arte de detección para EEG.

En el capítulo 5 se evalúan los métodos propuestos en tareas de clasificación de segmentos EEG. Para ello, proponemos diferentes problemas de clasificación utilizando bases de datos EEG seleccionados en la clínica Universitaria de Navarra (CUN), así como otras bases de datos utilizadas por otros autores en el estado del arte. De igual manera este capítulo evalúa un método de reducción de dimensionalidad basado en la información mutua (MI, del inglés mutual information) y el método de búsqueda de adelante-atrás (FB, del inglés forward-backward).

El Capítulo 6 concluye este trabajo con un resumen de las contribuciones y una breve discusión sobre la labor futura. Finalmente se incluyen cuatro

apéndices que sirven de referencia teórica adicional, descripción de las bases de datos y de experimentos adicionales que se han realizado.

## Conclusiones y líneas futuras

En esta tesis se han desarrollado dos propuestas para el procesado de señal EEG con el objetivo mejorar las prestaciones en escenarios como la detección y clasificación EEG con epilepsia. Uno de los métodos está diseñado a cancelar de forma automática los artefactos producidos por el movimiento de ojos. Este método designado como ICA-RLS, está basado en una combinación del algoritmo ICA con el algoritmo adaptativo de mínimos cuadrados recurrente (RLS, del inglés Recursive Least Square). La técnica ha sido comparada con otros métodos propuestos en el estado del arte y ha demostrado ser una buena alternativa para la eliminación de artefactos producidos por el movimiento de los ojos. El otro método, denominado extracción de parciales, es una propuesta novedosa de extracción de características EEG que emplea las distribuciones tiempo frecuencia (TFDs). Este método realiza sobre el plano tiempo-frecuencia un seguimiento de parciales. Los resultados en extracción de parciales, muestran como esta técnica es apropiada en escenarios de detección y clasificación EEG, debido a su sencillez en la implementación, su coste computacional no muy alto y la relevancia de la información que extrae para la detección y clasificación de la epilepsia, i.e., duración ( $L$ ), frecuencia ( $F$ ) y energía ( $E$ ) del parcial principal.

Todos los resultados en clasificación EEG obtenido de los dos métodos, se han obtenido a partir de una serie de simulaciones con bases de datos reales y con problemas de clasificación propuestos en el estado del arte que están diseñados para evaluar el desempeño de algoritmos que se proponen en clasificación EEG.

A continuación, resumiremos las aportaciones de la tesis y su comparación con el panorama existente en detección EEG. Queremos enfatizar que en los capítulos anteriores ya hemos mostrado las ventajas significativas que los métodos introducen al estado del arte. Esta sección solo repasa de una forma general todas las ventajas y desventajas que los métodos guardan. Finalmente, concluiremos con las posibles líneas de trabajo que esta tesis ha generado.

## Aportaciones originales

- **Eliminación EOG basado en un esquema ICA-RLS.** La eliminación de artefactos producidos por el movimiento de ojos reflejado en la señal EEG, es un problema con diferentes alternativas de solu-

ción pero que en su mayoría presentan las siguientes desventajas: las soluciones propuestas en la literatura con el método ICA, siguen empleando una inspección visual de la forma de onda en las fuentes, así se determina si una fuente ICA está o no relacionada con el movimiento de los ojos. Y los métodos basados en filtrado adaptativo siguen haciendo uso del electrooculograma (EOG) como señal de referencia. Tenemos entonces como inconvenientes: la inspección visual requiere del conocimiento de un experto, y la utilización del EOG como señal de referencia implica hardware adicional que registre el movimiento de ojos (equipo no siempre disponible en los hospitales). Teniendo en cuenta lo anterior, el método propuesto introduce un importante cambio en el estado del arte de eliminación de artefactos.

Por otra parte, ICA-RLS utiliza el error cuadrático medio (MSE, del inglés mean square error) y los mapas topográficos cerebrales. Esto hace que **no sea necesario conocer las diferentes formas de onda** que se registran en el EEG por el movimiento de ojos porque solo la fuente con bajo error y situada en la zona frontal del cerebro se elimina, por lo que **no es necesario la utilización o registro del electrooculograma (EOG)**.

Los resultados del método ICA-RLS han mostrado ser eficientes tanto en señales EEG muy contaminadas por artefactos oculares como también en señales con poca presencia de las mismas. Un posible inconveniente es el coste computacional que puede ser disminuído si se utiliza otros métodos existentes para ICA en la literatura. La optimización del método ICA-RLS se ha dejado como trabajo futuro.

- **Seguimiento de parciales (LFE)**. Utilizando las TFDs, el nuevo método de extracción ofrece tres características **a partir de un parcial principal**: energía, frecuencia y duración del parcial. En el capítulo 4 se ha mostrado la efectividad del método en tareas de detección de crisis epilépticas. Los resultados mostrados en el capítulo 5 demostraron su buen desempeño en tareas de clasificación EEG y a la vez se ha comprobado experimentalmente la posibilidad de combinar las características propuestas con otras del estado del arte como son las transformadas wavelets y la Fraccional de Fourier.

Además del buen desempeño que ha mostrado el seguimiento de parciales, otras características adicionales se han conseguido:

- **Adaptabilidad del método a cualquier TFD**. La extracción de parciales está basada en el seguimiento o “tracking” de parciales sobre el plano tiempo-frecuencia, lo cual hace que el método sea independiente de la distribución que se esté utilizando. Por esa

- razón, puede extenderse a otras aplicaciones que tienen distribuciones adecuadas a un problema. Por ejemplo, se puede extender el método en el análisis de lesiones de rodilla, donde se utiliza la distribución Pseudo Wigner-Ville (PWV) o para extraer información acústica bajo el agua, donde se utiliza habitualmente la distribución Wigner-Ville (WV).
- **Solución de problemas en clasificación EEG empleando pocas características.** El seguimiento de parciales LFE resuelve diferentes problemas en detección y clasificación EEG basándose en tres características extraídas a partir de un parcial principal. El capítulo 5 ha mostrado varios resultados de un problema de clasificación resuelto con pocas dimensiones. Por ejemplo, se hizo uso de dimensiones  $D=2$  o  $D=3$  para el problema N1 y para los pacientes 2, 3 y 5 respectivamente. Por otra parte, en los problemas más complejos como el N2 y N3, las características extraídas a partir del método de seguimiento de parciales fueron escogidas como relevantes, mostrando una mejora tanto en el desempeño del clasificador como en la reducción de la dimensión en la matriz de características EEG.
  - **Coste computacional no muy alto para clasificación EEG.** Teniendo en cuenta que el método extrae tres características, por lo que se tiene una matriz de características con dimensión  $D=3$ , el cómputo general en un esquema de clasificación es bajo. Otros métodos como la transformada de Fourier, FrFT o wavelets generalmente utiliza muchos más coeficientes ( $> 10$ ) para este tipo de problema.
  - **Posibilidad de combinar las características LFE con otras, para mejorar el desempeño del clasificador.** El EEG se caracteriza por ser una señal muy dinámica que presenta estallidos energéticos o surgimiento de frecuencias de poca duración. Eso hace que en determinados problemas, el método LFE no pueda detectar de forma eficiente esos eventos, haciendo necesario el empleo de otros métodos que son más apropiados para este tipo de escenarios. Wavelets y la FrFT permiten registrar información adicional que el seguimiento de parciales LFE no puede seguir, ya que ambos métodos se basan en un análisis multinivel. En el capítulo 5 se ha podido observar que el seguimiento de parciales LFE ofrece información relevante en tareas de clasificación y que mejora el desempeño del clasificador cuando se combina con estos métodos.
  - **Un método estable.** El capítulo 5 ha mostrado que los valores  $\Delta$ , épocas del EEG, tamaño de la ventana de análisis, etc., presentan una buena estabilidad en un rango de valores bastante

amplio sin alterar considerablemente el desempeño del detector a clasificador. Esto hace que el método de seguimiento de parciales presente una buena estabilidad frente a cualquier variación de estos parámetros libres.

- **Reducción de dimensionalidad EEG empleando información mutua (MI) y el procedimiento de búsqueda delante-atrás (FB).** La introducción teórica expuesta en el capítulo 4 junto con los experimentos desarrollados en el capítulo 5, dan soporte a la eficacia de este método en la reducción de dimensiones para clasificación EEG. Dos observaciones se pueden derivar de estos experimentos:
  - **Las características LFE aportan información importante en problemas de clasificación EEG.** En el capítulo 3 se describió un método de selección de características con un criterio de relevancia basado en la MI. Los resultados experimentales han mostrado que el seguimiento de parciales LFE aporta tres características importantes que pueden resolver problemas de clasificación empleando una, dos o las tres dependiendo de la complejidad del problema. Se puede entonces concluir que no hay dependencia entre las características y que no existe colinealidad entre ellas, por lo que las características LFE introducen información relevante en extracción de información para señales EEG.
  - **El método de selección de características basada en MI y una búsqueda adelante-atrás (FB) es una buena alternativa para la reducción de dimensión de señales EEG.** Como se ha podido ver en el capítulo 5, muchos problemas de clasificación presentaron una buena reducción en la dimensión de la matriz de características junto con una mejora en el desempeño del clasificador.

## Líneas futuras

A continuación presentamos las posibles líneas de trabajo que los dos métodos propuestos en esta tesis abren en el campo del procesamiento de señal EEG, como también las posibles extensiones que se pueden derivar de los algoritmos propuestos:

- **Extensión del método ICA-RLS a otros artefactos.** Aparte de los artefactos producidos por los ojos, existe otro artefacto que contamina la señal EEG como es el movimiento de músculos. La extensión implica analizar este artefacto desde el plano tiempo-frecuencia y extraer información en bandas de frecuencia diferentes para encontrar las características más relacionadas al artefacto. Por otra parte, se ha de analizar los mapas topográficos para encontrar las áreas activas en

el cerebro que correspondan al artefacto y determinar los canales de entrada que han de ser utilizados en el esquema adaptativo.

- **Automatización completa del método ICA-RLS.** Como se ha expuesto en el capítulo 4, el esquema propuesto hace uso de los mapas topográficos cerebrales y del valor MSE para seleccionar la fuente correspondiente al movimiento de ojos. A continuación presentamos los pasos a seguir para lograr una automatización completa de nuestro método: (i) construir un conjunto de datos EEG con la información de localización de canales, (ii) ajustar un modelo de dipolo eléctrico para las componentes ICA, (iii) obtener para cada componente su equivalente del dipolo junto a sus coordenadas y, finalmente, (iv) calcular las distancias empleando un electrodo de referencia para así determinar la fuente que está más cerca de los ojos. Un trabajo inicial de ajuste de modelos dipolo a las fuentes de ICA se describe en [Delorme and Makeig \[2004\]](#).
- **Extensión del método LFE a otros escenarios como BCI.** La detección de potenciales relacionados a eventos (ERPs) es una tarea fundamental en las aplicaciones BCI. Por lo general, la forma de onda ERP es cuantitativamente caracterizada por la amplitud, latencia y su propagación en el cuero cabelludo. Esa información puede ser analizada desde el plano tiempo-frecuencia de una TFD y además se puede extraer los parciales correspondientes al ERP mediante la extracción de características LFE. Por esta razón, creemos que el método de extracción de parciales puede jugar un papel interesante en detección y análisis de ondas ERP.
- **Clasificación de diferentes tipos de epilepsia.** Hay una extensiva clasificación tanto de crisis epilépticas como del tipo de epilepsia. Una forma sencilla de clasificarlas es teniendo en cuenta la forma de como se origina la crisis en la primera fracción de segundo. Si la crisis comienza en una región focal del cerebro es llamada parcial, y si se inicia por todas las partes del cerebro al mismo tiempo, se denomina generalizada. Dentro de cada grupo hay diferentes tipos de epilepsia con características diferentes entre ellas. Un análisis de cada epilepsia puede basarse empleando el método de extracción LFE. Ese método permitiría obtener información natural de cada epilepsia como bandas de frecuencia, duración por crisis y las zonas del cerebro más activas en crisis. Con esa información sería posible clasificar cada tipo de epilepsia con miras a un posible tratamiento, diagnóstico o cirugía.
- **Seguimiento de pacientes bajo tratamiento observando la desaparición del parcial dominante.** En la presentación del método de extracción de características LFE, se observó la aparición de parcial

principal de larga duración sobre el plano tiempo-frecuencia cuando un paciente sufre una crisis epiléptica. Este parcial puede ser utilizado para monitorear el efecto de los medicamentos en los pacientes, bajo la suposición de que el parcial principal tiende a desaparecer en el tiempo con la medicación.

- **Anticipación de las crisis epilépticas.** Una persona que sufre de epilepsia tiene problemas en su vida diaria que afectan a su entorno familiar, laboral y personal trayendo como consecuencia una baja autoestima. Todo esto sucede debido a la inseguridad diaria que ocasiona el no saber en qué momento puede sufrir una crisis. Por medio de un sistema de anticipación al parcial principal que avise o que haga alguna actividad (chip intracraneano, sistema de automedicación) se podría mejorar considerablemente la calidad de vida en un paciente que sufra de epilepsia.

Ya para finalizar, queremos agregar la importancia de lograr una **integración del EEG con otras técnicas como la fMRI**. En el capítulo 2 se describieron las ventajas y desventajas de las señales EEG. El EEG tiene una resolución espacial baja porque está limitada al número de electrodos. La técnica MEG presenta una mejor resolución temporal, pero sufre de igual manera la misma desventaja que el EEG y es económicamente cara. Con fMRI se solucionaría el problema de los dos métodos anteriores porque esta técnica presenta una alta resolución espacial ya que las áreas activas del cerebro pueden ser localizadas milimétricamente. La integración de estas técnicas es de vital importancia en los estudios de neurociencia, porque permite mejorar la detección de otras enfermedades neurodegenerativas como alzhéimer, parkinson, depresión o demencia senil.



# Contents

Acknowledgment	i
Abstract	v
Resumen extendido en español	vii
Contents	xix
<b>1 Introduction to the detection problem in EEG signals</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Problem statement . . . . .	3
1.3 Summary of previous work in epileptic detection on EEG signals	4
1.4 Motivation . . . . .	5
1.5 Objectives . . . . .	7
1.6 Overview of the rest of the thesis . . . . .	7
<b>2 Outline of Electroencephalography</b>	<b>9</b>
2.1 Introduction . . . . .	9
2.2 The nervous system . . . . .	10
2.2.1 Neural activities . . . . .	11
2.2.2 Cerebral cortex . . . . .	12
2.3 The EEG machine: An Overview . . . . .	13
2.3.1 Electrodes . . . . .	15
2.3.2 Jackbox and Montage Selector . . . . .	16
2.3.3 EEG filters . . . . .	19
2.3.4 Amplifiers . . . . .	20
2.3.5 Analog and digital conversion . . . . .	20
2.3.6 The signal display: oscilloscope and computer . . . . .	21
2.4 Introduction to the EEG analysis . . . . .	21
2.4.1 Brain rhythms and waveforms . . . . .	22
2.4.2 Artifacts . . . . .	24
2.4.3 Abnormal EEG patterns . . . . .	25
2.5 EEG applications . . . . .	26
2.5.1 Seizures and epilepsy . . . . .	26

2.5.2	Brain computer interface (BCI)	27
2.5.3	Sleep disorders	28
2.5.4	Event related potential (ERP)	29
2.6	Summary and conclusions	29
<b>3</b>	<b>EEG signal processing</b>	<b>31</b>
3.1	Introduction	31
3.2	Modelling and segmentation	32
3.2.1	EEG signal modelling	32
3.2.2	Signal segmentation	33
3.3	Denoising and filtering	33
3.3.1	Lowpass filtering	34
3.3.2	Independent component analysis (ICA)	34
	ICA fundamentals	35
3.3.3	Adaptive filtering	37
	Least Mean Square algorithm (LMS)	39
	Recursive Least Squares Algorithm (RLS)	40
3.4	Feature extraction	42
3.4.1	Classical signal analysis tools	42
3.4.2	Time-frequency distributions (TFD)	44
3.4.3	Wavelet coefficients	46
3.4.4	Fractional Fourier transform	47
3.5	Feature selection	49
3.5.1	Subset relevant assessment	50
3.5.2	Mutual information (MI)	51
	MI estimation based on Parzen windows	52
	Kraskov's estimator of mutual information	53
	MI estimation for classification problems	54
3.5.3	Search procedure	55
	Forward-backward algorithm	56
3.6	Classification algorithms for EEG signals	57
	Support vector classification	58
3.7	Summary and conclusions	62
<b>4</b>	<b>Proposed methods</b>	<b>63</b>
4.1	Introduction	63
4.2	Artifacts elimination using adaptive filtering and ICA	64
4.2.1	The ICA-RLS method	65
	Independent component analysis (ICA)	65
	Adaptive filtering	66
4.2.2	Experiment 1: EOG removal on EEG epochs	68
	Data and experimental setup	68
	Results	70

4.2.3	Experiment 2: Comparison of the algorithm with other techniques . . . . .	72
	Method . . . . .	73
	Results . . . . .	74
	ICA-kurtosis-modified . . . . .	75
4.2.4	Discussions and conclusion . . . . .	76
4.3	The tracks extraction method (LFE features) . . . . .	77
4.3.1	The EEG model . . . . .	78
4.3.2	Discrete Time-Frequency Distributions . . . . .	79
4.3.3	Local peaks estimation and linking . . . . .	81
4.3.4	Feature matrix construction using tracks extraction . . . . .	82
4.3.5	Experiment 1: Time-frequency analysis of epileptic EEG signals using tracks extraction . . . . .	85
	Data and setting . . . . .	85
	Results . . . . .	85
4.3.6	Experiment 2: Epilepsy seizure detection using LFE features . . . . .	88
	Data and setting . . . . .	88
	Results . . . . .	90
4.3.7	Discussions and conclusions . . . . .	93
4.4	Summary and conclusions . . . . .	94
<b>5</b>	<b>EEG feature selection and classification using SVMs</b>	<b>97</b>
5.1	Introduction . . . . .	97
5.2	Parameter analysis for the proposed methods . . . . .	99
5.2.1	ICA-RLS method: $\lambda$ and filter order . . . . .	99
5.2.2	LFE tracks extraction: distributions, EEG epoch and overlapping . . . . .	100
	Selection of time-frequency distributions (TFDs) . . . . .	101
	Overlapping and EEG epochs . . . . .	104
	Thresholds: “thr” and $\Delta$ . . . . .	104
	Selection of K value for K-NN algorithm and SVM classifier . . . . .	108
5.3	Clinical evaluation of the proposed methods . . . . .	109
5.4	Dimensionality reduction of EEG features matrix by forward-backward algorithm and mutual information (MI) . . . . .	114
5.5	Comparison of classification accuracy . . . . .	116
5.6	Summary and conclusions . . . . .	117
<b>6</b>	<b>Conclusions and future work</b>	<b>121</b>
6.1	Contributions . . . . .	122
6.2	Future work . . . . .	124

<b>A</b>	<b>Filtering concepts</b>	<b>127</b>
A.1	Stochastic process . . . . .	127
A.1.1	Statistics of stochastic process . . . . .	127
A.1.2	Stationary process . . . . .	128
A.2	The filtering problem . . . . .	128
A.3	The matrix inversion lemma . . . . .	129
<b>B</b>	<b>Statistical pattern concepts</b>	<b>131</b>
B.1	A brief introduction to Machine Learning . . . . .	131
B.2	Classification algorithms . . . . .	132
B.2.1	Traditional techniques . . . . .	132
B.2.2	Large margin algorithms . . . . .	133
B.3	Model selection . . . . .	134
B.3.1	Model selection for RBF kernels . . . . .	134
B.4	Bootstrap resampling . . . . .	135
B.4.1	Confidence intervals by percentile bootstrap . . . . .	136
B.5	Receiver Operating Characteristics (ROC) . . . . .	137
<b>C</b>	<b>Time-frequency Distributions (TFDs)</b>	<b>141</b>
C.1	EEG analysis using Time-frequency Distributions (TFDs) . . . . .	141
C.1.1	From Fourier to quadratic TFDs . . . . .	141
C.1.2	The affine class . . . . .	143
C.1.3	The reassignment method . . . . .	144
C.1.4	The Reduced Interference Distribution (RID) . . . . .	144
C.1.5	Optimal Kernel Design (OKD) . . . . .	145
C.1.6	Other alternatives in time-Frequency analysis . . . . .	145
C.1.7	Experiments . . . . .	145
	Data collection . . . . .	145
	Results . . . . .	146
C.1.8	Discussions and conclusion . . . . .	149
<b>D</b>	<b>Databases</b>	<b>151</b>
D.1	Database 1 . . . . .	151
D.2	Database 2 . . . . .	152
D.3	Database 3 . . . . .	153
	<b>Bibliography</b>	<b>155</b>

# List of figures

1.1	EEG signal in time domain (left) and frequency domain (right).	2
1.2	Research line proposed by University Hospital of Navarra.	6
2.1	Presynaptic and postsynaptic activities in the neurons.	11
2.2	Cerebral cortex and its four lobes.	13
2.3	Block diagram of the electroencephalographic machine.	14
2.4	Resistivities and thicknesses ( $\Omega$ =ohms) of the brain	16
2.5	Neurophysiology 10-20 System.	17
2.6	Typical set of EEG signals.	18
2.7	Different effects of using filters in the EEG	19
2.8	Typical normal brain waves in the EEG	24
2.9	Block diagram of the brain computer interface	28
3.1	General ICA process applied to EEG signals	36
3.2	Block diagram of adaptive filter.	38
3.3	Chirp signal using time domain $x(t)$ (upper) and frequency domain $X(\omega)$ (bottom).	43
3.4	Spectrogram representation of the chirp	44
3.5	The relation of fractional domain $(u, v)$ with traditional time-frequency plane $(t, w)$ rotated by an angle $\alpha$ .	48
3.6	Simple prediction problem.	51
3.7	Support vector machines (SVMs).	58
3.8	A binary classification toy problem example.	60
3.9	The nonseparable case.	61
4.1	General scheme of automatic EOG noise cancellation using adaptive filtering and ICA.	67
4.2	Raw EEG data and its ICA decomposition.	70
4.3	An example of EOG artifact rejection using ICA-RLS and RLS.	71
4.4	Topographic map of the components with their MSE values.	73
4.5	Comparison of different EOG artifact removal methods on the EEG using and ICA-kurtosis method.	74
4.6	Comparison of different EOG artifact removal methods on the EEG using and ICA-kurtosis modified.	75

4.7	Time-frequency distributions (TFDs). . . . .	79
4.8	Spectral peaks. . . . .	82
4.9	The tracks extraction method. . . . .	83
4.10	Tracks extraction using a record with a seizure. . . . .	84
4.11	Results of three time-frequency representations of an epileptic EEG segment. . . . .	87
4.12	Three different EEG segments and tracks extraction method. . . . .	87
4.13	Features extraction for an epileptic EEG register. . . . .	89
4.14	Evaluation of the dataset size effect in the N2 detection problem. . . . .	92
5.1	R vs. filter order for RLS algorithm with multiple EOG reference channels. . . . .	100
5.2	First burst of epilepsy seen on the time-frequency plane using an atomic Gabor decomposition (adopted from <a href="#">Guerrero-Mosquera [2011]</a> ). . . . .	101
5.3	Overlapping effect on the SVM classifier for each TFD. . . . .	102
5.4	EEG epochs effect on the SVM classifier for each TFD. . . . .	105
5.5	Linking process on the SPWV distribution (background) for different values of $\Delta$ . . . . .	106
5.6	Effect of $\Delta$ in the classifier performance. . . . .	107
5.7	K-NN performance using a RBF-SVM classifier. . . . .	109
B.1	A basic ROC curve . . . . .	138
C.1	Beginning of a seizure. . . . .	146
C.2	Some TFDs of EEG during an epileptic seizure. . . . .	147
C.3	TFds with 2-D filtering. . . . .	147
C.4	TF representation of the reassignment method with N=1000. . . . .	148
C.5	The RID distribution with different kernels. . . . .	148
C.6	Epileptic seizure with the S-Transform and Ridges method. . . . .	149
D.1	General 10-20 System scheme used for our databases. . . . .	152
D.2	Characteristics of electrodes used for dataset 3. . . . .	153

# List of Tables

2.1	Essential features of EEG analysis. . . . .	22
4.1	ICA-RLS Algorithm . . . . .	69
4.2	Normalized correlation coefficient between the raw EEG and the EEG filtered by ICA-RLS. . . . .	72
4.3	Computational cost figures for all the implemented methods. . . . .	76
4.4	Tracks extraction algorithm (LFE features) . . . . .	86
4.5	LFE features of different EEG's during a seizure. . . . .	90
4.6	Sensitivities and Specificities of EEG's in different patients. . . . .	91
5.1	$F_{score}$ evaluation of different TFDs for N1 problem. . . . .	103
5.2	Feature evaluations for database 2. . . . .	111
5.3	Feature evaluations for database 1 using raw EEG and RID. . . . .	112
5.4	Feature evaluations for database 1 using a preprocessed EEG and RID. . . . .	112
5.5	$F_{score}$ values average and standard deviations correspond to 1000 bootstrap runs using the TFD SPWV. . . . .	113
5.6	$F_{score}$ values average and standard deviations correspond to 1000 bootstrap runs using the TFD RID. . . . .	113
5.7	$F_{score}$ values average and standard deviations correspond to 1000 bootstrap runs using the TFD RSPWV. . . . .	113
5.8	$F_{score}$ evaluations achieved by three different MI estimations and SPWV. . . . .	115
5.9	$F_{score}$ evaluations achieved by three different MI estimations and RID. . . . .	115
5.10	$F_{score}$ evaluations achieved by three different MI estimations and RSPWV. . . . .	115
5.11	Feature selection using three different MI estimations. . . . .	117
5.12	Comparison of classification accuracy (in percent) obtained by LFE approach for epileptic seizure detection. . . . .	119
C.1	Some distributions and their kernels . . . . .	143
C.2	Computation time of different TFD's using rectangular non-overlapping windows. . . . .	149

D.1	Main characteristics of the database 1. . . . .	152
D.2	Main characteristics of the database 2. . . . .	153
D.3	Main characteristics of the database 3. . . . .	154

# Acronyms

AUC	Area under a ROC curve
BCI	Brain Computer Interface
DFT	Discrete Fourier transform
DSPWV	Discrete Smooth Pseudo Wigner-Ville distribution
EEG	Electroencephalogram
EOG	Electrooculogram
EP	Evoked Potential
FB	Forward-backward algorithm
fMRI	Functional Magnetic Resonance Imaging
FrFT	Fractional Fourier transform
FT	Fourier transform
ICA	Independent Component Analysis
JADE	Joint Approximate Diagonalization of Eigen-matrices
k-NN	k-nearest neighbor algorithm
LFE	Length, Frequency and Energy features
LMS	Least mean square
MEG	Magnetoencephalogram
MI	Mutual Information
MSE	Mean square error
PWV	Pseudo Wigner-Ville distribution
RBF	Radial Basis Function

RID	Reduced Interference distribution
RLS	Recursive Least Squares
ROC	Receiver Operating Characteristics
RPWV	Reassignment Pseudo Wigner-Ville distribution
SPWV	Smooth Pseudo Wigner-Ville distribution
SVM	Support Vector Machine
TFD	Time frequency distribution
WT	Wavelets transform
WV	Wigner-Ville distribution

# Chapter 1

## Introduction to the detection problem in EEG signals

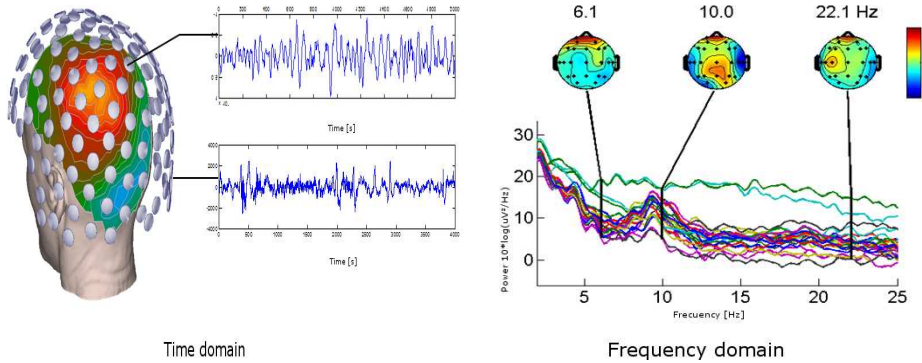
### 1.1 Introduction

An emergent research line arises from the integration between clinical neuroscience and engineering and is called clinical neuroengineering. This new field uses engineering, computational simulation, mathematical analysis, imaging techniques, and hardware based modeling to solve different problems in clinical neurosciences. The goal is to use engineering approaches to uncover the principles of neural function and to use these principles in the design of diagnostic and therapeutic systems; research is then necessary from bench to bedside. Clinical neuroengineering has a variety of active fields with objectives such as new treatments based on new engineering tools and new approaches to neural disease diagnosis, as introduced in [Thakor and Tong \[2006a\]](#).

One important tool for the diagnosis and treatment of mental and brain diseases and abnormalities is the electroencephalogram (EEG). More than eight decades have passed since Hans Berger measured the first EEG in humans and still nowadays EEG is used extensively to evaluate neurological disorders in the clinic and to investigate brain function in the laboratory. [Fig.1.1](#) shows an EEG in the time domain (left) and frequency domain (right). Time and frequency domains are popular methods in EEG signal analysis.

The identification and interpretation of the bioelectric phenomena originating from the cerebral cortex and evoked by auditory and somatosensory stimuli are clear objectives to both understand how the brain works and to recognize different pathologies. Diseases such as Parkinson's, dementia, Alzheimer's, schizophrenia and epilepsy are studied to find a cure or accurate diagnosis. Regarding epilepsy, research is aimed at isolating, identifying and

locating epileptic foci from EEG signals, procedure prior to possible surgical treatment in cases where the medication has not been able to stop the disease. Another possibility in drug-resistant cases is the early detection of seizures by automated methods based on EEG signals. This is a problem of great clinical interest due to its great complexity, theoretical and practical, for which only partial solutions exist up to day, and they have been tested on very small numbers of patients or databases.



**Figure 1.1:** EEG signal in time domain (left) and frequency domain (right). Time and frequency domains are popular methods in EEG signal analysis.

Time-frequency distributions (TFDs) and adaptive methods have demonstrated to be good alternatives in designing systems for detecting neurodegenerative diseases. Abnormal EEG are considered dynamic signals which exhibits non-stationary behavior with focal or multifocal activity, spikes, sharp waves, and focal mono-rhythmic discharges. TFDs are appropriate transformations for this type of signals because they offer the possibility of analyzing relatively long continuous segments of EEG data even when the dynamics of the signal are rapidly changing. On the other hand, EEG records usually vary, not only in different patients but also from a single patient because it depends on the registration time. Therefore we have to consider adaptive methods that consider every scenario.

This chapter will introduce our proposals for EEG feature extraction and preprocessing used in the detection scheme for EEG signals with epilepsy. We also present the problem statement and the state of the art in EEG epilepsy detection. Finally, the objectives, previous work and motivation for this Thesis will be given.

This chapter is organized as follows: Section 1.2 deals with the problem statement, previous work in epileptic detection on EEG signals are reviewed in Section 1.3, in Section 1.4 the motivation of this Thesis is presented, Section 1.5 describes the objectives of this Thesis and in Section 1.6 we outline the contents of the rest of the Thesis.

## 1.2 Problem statement

Epilepsy is considered the disease with major prevalence within disorders with neurological origin. The recurrent and sudden incidence of seizures can lead to dangerous and possibly life-threatening situations. Since disturbance of consciousness and sudden loss of motor control often occur without any warning, the ability to predict epileptic seizures would reduce patients' anxiety, thus improving quality of life and safety considerably.

Intractable epilepsy is one of the most physically and emotionally destructive neurological disorders affecting population of all ages. It is generally accepted that surgical resection of epileptic foci is the best solution. However, before conducting neurosurgery, it is necessary to study the presence of epileptiform activity, which is distinct from background EEG activity. The analysis of EEG data and the extraction of information is not an easy task. EEG recording may be contaminated by extraneous biologically generated (human body) and externally generated signals (power line, electrode movement etc.). The presence of this kind of noise or "artifacts" makes it difficult to discriminate between original brain waves and noise. This problem motivates a preprocessing step to obtain clean signals before the detection task.

Another important problem in EEG processing is to figure out which kind of information or "patterns" we want to extract from the signal. This procedure is known as feature extraction. Extracted features depend considerably on the method used, which are usually transformations to other domains that permit the extraction of hidden information in the signal. Care has to be taken not to extract similar or irrelevant features that could reduced the detector performance or increase the computational load. Therefore, a feature selection procedure is also necessary to complement the features extraction procedure.

Other important task in the medical environment to diagnose, classify or detect abnormalities, is to obtain ictal and interictal patterns. This usually involves monitoring of the patient during several weeks. Continuous observation or patient monitoring is a care activity that requires time and expensive work, being necessary specialized personnel for alerting of possible changes that a patient may have. When information is stored, there is another activity equally important: the analysis of the EEG registers. The specialists have to analyze waveforms, spectrum and peaks, and based on this analysis try to determine the pathology that the patient suffers. Usually they use a video unit. In many instances, there are disagreements among specialists about the same record due to the subjective nature of the analysis.

The introduction of new techniques and mathematical algorithms in the EEG analysis can be helpful to design new supporting methods in medical

decision and diagnosis, thus avoiding tedious analysis of long-term records and doubts about the brain pathology that a patient suffers.

Nowadays there are many published studies about neurological diseases detection but these results are very focused on private institutional databases or rely on impractical numerical methods which are difficult to implement in a hospital environment. Therefore, the implementation and design of practical and reliable detection systems are very important in hospitals. This doctoral thesis, tries to narrow the gap that exists between EEG signal theory and practical implementation for the medical practice.

### 1.3 Summary of previous work in epileptic detection on EEG signals

Some methods of seizure detection were based on detecting strong rhythmic movements of the patient, but these methods had a limitation: seizures do not always present strong movements. This limitation led the detection problem to methods based on EEG signal analysis, for example, detection of large seizures discharges in several EEG channels by amplitude discrimination was described by [Ives et al. \[1974\]](#); [Babb et al. \[1974\]](#) designed an electronic circuit for seizures detection from intracranial electrodes. However, some seizures do not present EEG changes, therefore seizure detection only based on EEG analysis was not at all reliable and it was necessary to combine it with other methods. For example, [Prior et al. \[1973\]](#) identified on the EEG signal a large increase followed by a clear decrease in the amplitude and at the same time by large electromyogram (EMG) activity; [Murro et al. \[1991\]](#) described a method based on spectral parameters and discriminant analysis.

New alternatives for this detection problem are addressed from the point of view of pattern recognition. [Gotman \[1982\]](#) presented an automatic detection system based on seizure patterns. The drawback of this method is the necessity of traditional visual inspection of the patterns, being necessary a careful examination of them by a specialist.

Presently, EEG epileptic detectors have evolved including new techniques such as neural networks, non-linear models, independent component analysis (ICA), Bayesian methods, support vector machines and variance-based methods, as described in [Guerrero-Mosquera et al. \[2010a\]](#). Other group of methods potentially useful for detecting and analyzing non-stationary signals are time-frequency distributions (TFDs) [Cohen \[1995\]](#). These methods allow us to visualize the evolution of the frequency behavior during some non-stationary event by mapping a one dimensional (1-D) time signal into a two-dimensional (2-D) function of time and frequency. Therefore, from the time-frequency (TF) plane it is possible to extract relevant information using

methods such as peak matching, filter banks, energy estimation, etc.

On the other hand, most of the detection methods proposed in the literature assume a clean EEG signal free of artifacts or noise, leaving the preprocessing problem open to any denoising algorithm such as digital filters, independent component analysis (ICA) or adaptive schemes using the electrooculogram (EOG) as reference signal, in [Guerrero-Mosquera and Navia-Vazquez \[2009\]](#). In [Chapter 3](#) we will show the importance of the noise removing procedure and we also introduce adaptive filtering, which is necessary to eliminate the dependence on visual analysis that generally leads to problems with the interpretation.

## 1.4 Motivation

During the last years the Signal Theory and Communications department of the University Carlos III of Madrid, Electrical and Electronic Engineering department in the Public University of Navarra and the Neuroscience Unit Neurophysiology in the University Hospital of Navarra have been working in successive researches concerning the analysis of different bioelectric phenomena originated at the cerebral cortex with the objectives of applying these studies to describe new approaches to neural disease diagnosis and treatment developed through the use of new engineering tools and techniques. These neurological diseases includes Parkinson, Alzheimer and epilepsy.

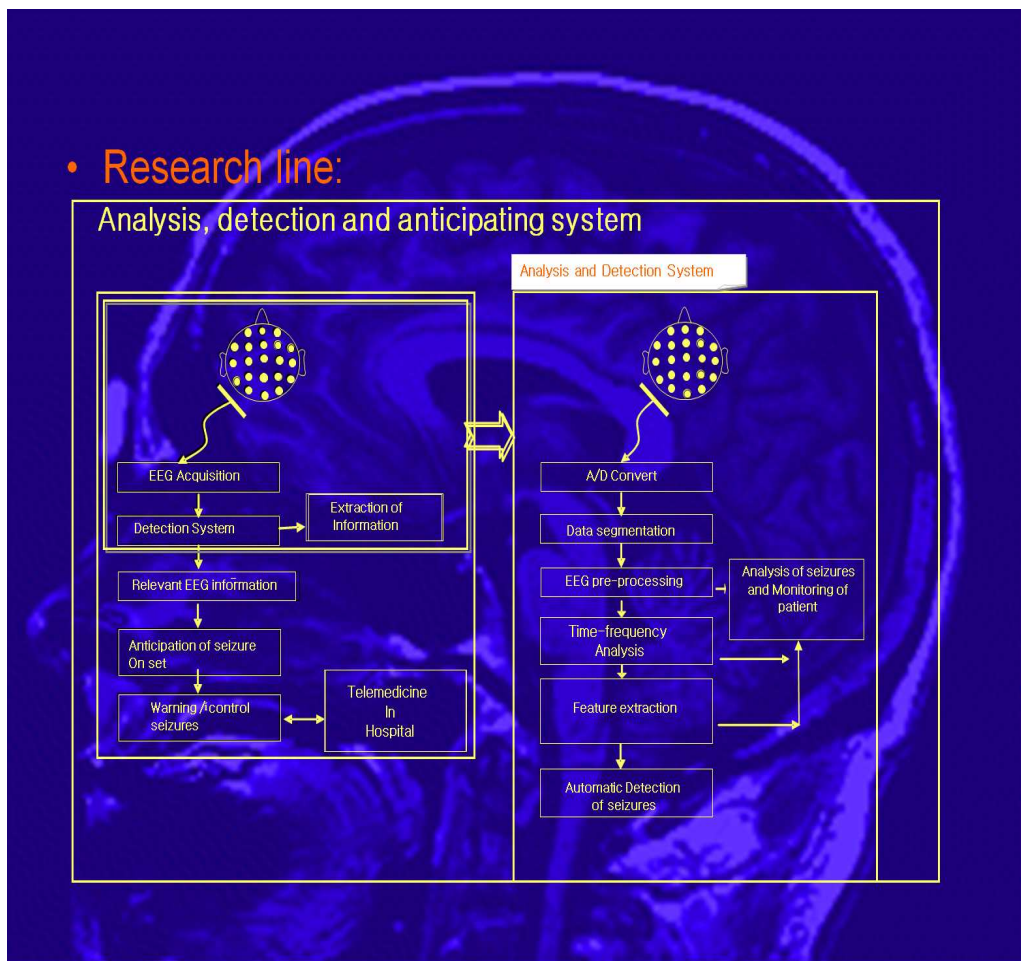
EEG signals change continuously and processing every single one requires a long period of time, more personal, high storage databases, and computational cost. To improve efficiency, adequate algorithms may be used to highlight the most pertinent features of epileptic signals. This allows to identify and classify epileptic seizures faster, and also to see which parts of the brain are the most affected.

On first results showed a need to reduce the amount of information obtained from EEGs due to the fact that the analysis of all the data requires a great deal of time and large processing capacity. In order to achieve this objective, the detection scheme has to allow the extraction of the most relevant features related to epilepsy. Thus, we need algorithms that allow us to detect, classify or identify epilepsy with a small amount of information and then improve the efficiency of detection or classification tasks carried out by analyzing small amounts of information through the use of the most important features extracted from the EEG signals.

The research team proposed a research line that was approved at Neuroscience Unit Neurophysiology in the University Hospital of Navarra. This design, summarized in [Fig.1.2](#) (left panel), consists in developing different algorithms to analyze, detect and anticipate epilepsy seizures with the possi-

bility of applying this design to other neurodegenerative pathologies such as Parkinson, Alzheimer or the analysis of different sleep disorders. This PhD thesis studies analysis and detection methods to detect epileptic events in EEG signals (right panel in Fig.1.2), leaving seizure anticipation methods as future work.

The design foundations are simple: (i) an EEG signal is recorded, (ii) the EEG signal is cleaned by some denoising and artifact rejection algorithm, (iii) features from EEG signal are extracted and finally (iv) relevant features are selected. Next section explains in detail all objectives proposed in the design.



**Figure 1.2:** Research line proposed (left panel) and our approach in analysis and detection methods to epileptic EEG signals developed in this Thesis (right panel).

## 1.5 Objectives

We have proposed the following overall objective in this work: Design detection and analysis algorithms for EEG signals. This design has to be practical, simple and efficient to be applied in the medical environment. To achieve this objective, it is necessary to perform different specific tasks:

- Data collection in the hospital and databases searching used by other authors.
- Remove artifacts and EEG denoising by independent component analysis (ICA).
- Design of adaptive EEG artifact removal system.
- Analysis of different time-frequency distributions (TFDs) for epileptic EEG signals.
- New approach for feature extraction from TFDs to find seizure patterns.
- EEG seizure detection by proposed features.
- Performance comparison of features proposed with different features extraction methods for EEG signal detection and classification.
- Selection of relevant features for improving detection and classification tasks by dimensionality reduction methods.
- Design of automatic EEG detection and classification system.

We will restrict ourselves to the development of tools for EEG preprocessing and features extraction algorithm with the objective of improving detection or classification performance in EEG signals with epilepsy. These methods could be extended to other scenarios in EEG processing and their applications but for this it will be necessary to adapt this methods to the problems which we want to solve.

## 1.6 Overview of the rest of the thesis

The computational complexity introduced for the methods proposed will be shown in Chapter 4. These complexity orders are rarely shown in the state-of-the-art of EEG detection, therefore we think that our results are more reliable from the practical point of view.

In Chapter 2 we introduce the fundamental concepts about the nervous system and EEG generation. Concepts related to visual analysis of the EEG, artifacts, abnormal EEG patterns, brain rhythms are described. This chapter also includes EEG applications such as epilepsy detection, brain computer interface (BCI), sleep disorders and event related potentials (ERPs).

In Chapter 3 concepts in signal modelling, signal segmentation, filtering and denoising, feature extraction, feature selection and classification algorithms are reviewed. Signal modelling includes physical model, linear and non-linear models. Signal segmentation tells us how to separate meaningful segments in time-invariant signals and time-varying signals. Filtering and denoising briefly introduce to different techniques for obtaining clean EEGs, such as digital filtering, adaptive methods and independent component analysis (ICA). Feature extraction and feature selection introduce to transformations which we could apply to EEG signal to describe in a clear way the data and the principle of choosing smaller number of variables among the original ones.

In Chapter 4 we present two new proposals for EEG signal processing: a new method for eliminating eyes artifacts from EEG based on ICA-RLS, and an approach in feature extraction for EEG signals based on a sinusoidal model. ICA-RLS method has two important characteristics: first, it is not necessary to use EOG dedicated electrodes, and second, every ICA projection data is fed into an adaptive filtering. On the other hand, tracks extraction (LFE features) is a novel approach to EEG detection which is based on track measurements such as length, frequency and energy ( $L, F, E$ ). We also present results about both eye artifact rejection and epilepsy detection. These methods could be used in other scenarios such as BCI and EEG seizure classification.

In Chapter 5 we evaluate the performance of the proposed methods in classification tasks for EEG segments. For this, we propose different classification problems using EEG databases recorded at University Hospital of Navarra as well as others used in previous research. This chapter also evaluates a method for dimensionality reduction for EEG based on mutual information and forward-backward procedure.

Chapter 6 concludes this work with a summary of the contributions and a brief discussion about further work.

# Chapter 2

## Outline of Electroencephalography

### 2.1 Introduction

Neural activity in the human brain starts from the early stages of prenatal development. This activity or signals generated by the brain are electrical in nature and represent not only the brain function but also the status of the whole body.

At the present moment, three methods can record functional and physiological changes within the brain with high temporal resolution of neuronal interactions at the network level: the electroencephalogram (EEG), the magnetoencephalogram (MEG), and functional magnetic resonance imaging (fMRI); each of these has advantages and shortcomings. MEG is not practical for experimental work when subjects may move freely, because of the large size of magnetic sensors. For image sequences, fMRI has a time resolution very low and many types of EEG activities, brain disorders and neurodegenerative diseases cannot be recorded. On the other hand the spatial resolution of the EEG is limited to the number of electrodes, as described in [Ebersole and Pedley \[2003\]](#), [Sanei and Chambers \[2007\]](#).

Much effort has been made to integrate information of multiple modalities during the same task in an attempt to establish an alternative high-resolution spatiotemporal imaging technique. The EEG provides an excellent tool for the exploration of network activity in the brain associated to synchronous changes of the membrane potential of neighboring neurons. Understanding of neuronal functions and neurophysiological properties of the brain together with the mechanisms underlying the generation of biosignals and their recordings is important in the detection, diagnosis, and treatment of brain disorders.

Cerebral sources of electroencephalography potentials are three-dimensional volumes of cortex. These sources produce three-dimensional potential fields within the brain. From the surface of the scalp, these can be recorded as two-dimensional fields of time-varying voltage. The physical and functional factors that determine the voltage fields that these sources produce could be appreciated in order to locate and characterize cortical generators of the EEG.

Electroencephalography enables clinician to study and analyze electrical fields of brain activity recorded with electrodes placed on the scalp, directly on the cortex (e.g., with subdural electrodes), or within the brain (with depth electrodes). For each type of recording, the specialist attempts to determine the nature and location of EEG patterns and whether they correspond to normal or abnormal neural activity.

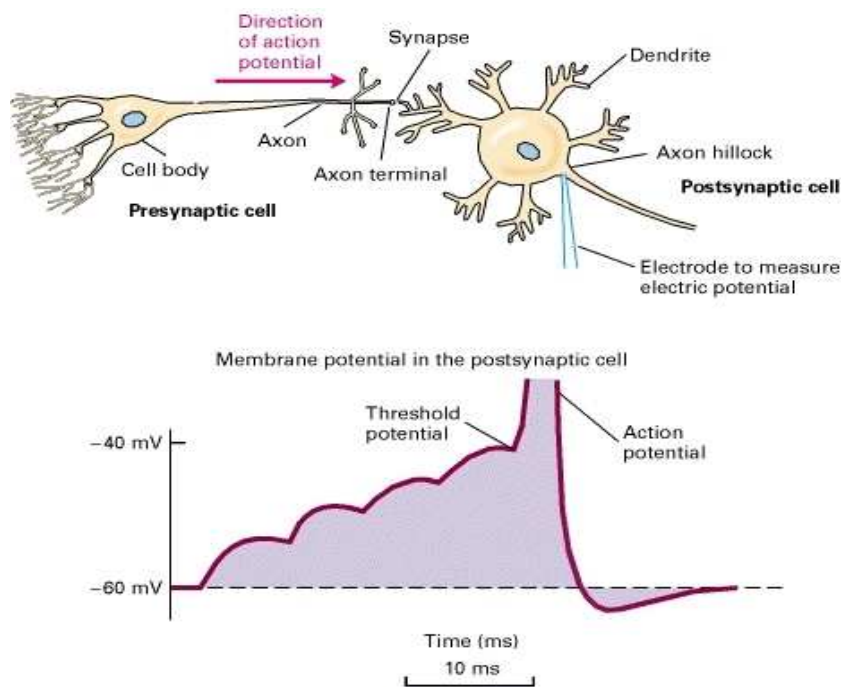
This chapter first presents a brief description of the nervous system and the electrical activity of the brain (Section 2.2); further details can be found in comprehensive descriptions of the human brain in [Ebersole and Pedley \[2003\]](#), [Fisch \[1999\]](#), [Sanei and Chambers \[2007\]](#). Section 2.3 describes the EEG machine, which introduces the electrodes, different EEG montages and filters. A variety of common brain rhythms and waves are presented in Section 2.4 which are of special interest in the chapter on brain rhythms, artifacts and abnormal EEG patterns. In Section 2.6 the summary and conclusions are given.

## 2.2 The nervous system

The nervous system is an organ system containing a network of specialized cells called neurons that gathers, communicates, and processes information from the body and send out both internal and external instructions that are handled rapidly and accurately. In most animals the nervous system is divided in two parts, the central nervous system (CNS) and the peripheral nervous system (PNS). CNS contain the brain and the spinal cord, and the PNS consists of sensory neurons, grouping of neurons called ganglia, and nerves cells that are interconnected and also connect to the CNS. The two systems are closely integrated because sensory input from the PNS is processed by the CNS, and responses are sent by the PNS to the organs of the body. Neurons transmit electrical potentials to other cells along thin fibers called axons, which cause chemicals called neurotransmitters that permit the neuronal function called synapses. These electrical potentials, called as “action potentials” is the information transmitted by a nerve that, in one cell, cause the production of action potentials in another cell at the synapse. A potential of 60-70 mV with some polarity may be recorded under the mem-

brane of the cell body. This potential changes with variations in the synaptic process. In this sequence, the first cell to produce actions potentials is called the *presynaptic cell*, and the second cell, which responds to the first cell across the synapse, is called the *postsynaptic cell*. Presynaptic cells are typically neurons, and postsynaptic cells are typically other neurons, muscle cells, or gland cells. A cell that receives a synaptic signal may be inhibited, excited or otherwise modulated. The Fig.2.1 shows the synaptic activities schematically.

The CNS is a major site for processing information, initiating responses, and integrating mental processes. It is analogous to a highly sophisticated computer with the ability to receive inputs, process and store information, and generate responses. Additionally, it can produce ideas, emotions, and other mental processes that are not automatic consequences of the information input.



**Figure 2.1:** Presynaptic and postsynaptic activities in the neurons. An action potential that travels along the fibre ends in an excitatory synapse. This process causes an excitatory postsynaptic potential in the following neuron.

### 2.2.1 Neural activities

Cells of the nervous system include neurons and nonneural cells. *Neurons* or *nerve cell* communicate information to and from the brain. They are

organized to form complex networks that perform the functions of the nervous systems. All nerve cells are collectively referred to as neurons although their size, shape, and functionality may differ widely. Neurons can be classified with reference to morphology or functionality. Using the latter classification scheme, three types of neurons can be defined: sensory neurons, connected to sensory receptors, motor neurons, connected to muscles, and interneurons, connected to other neurons.

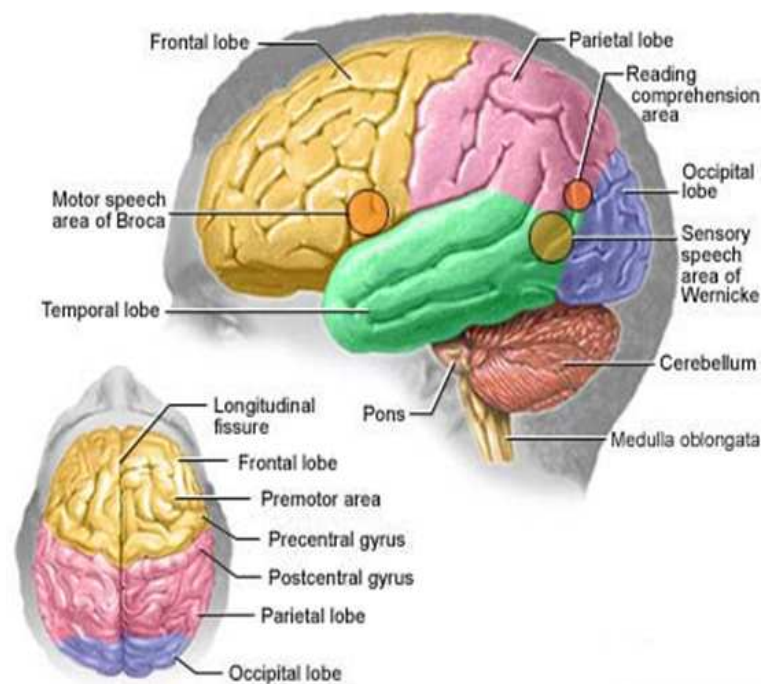
The cell body is called the *soma*, from which two types of structures extend: the *dendrites* and the *axon*. Dendrites are short and consist of as many as several thousands of branches, with each branch receiving a signal from another neuron. The axon is usually a single branch which transmits the output signal of the neuron to various parts of the nervous system. Each axon has a constant diameter and can vary in size from a few millimeters to more than 1 m in length; the longer axons are those which run from the spinal cord to the feet. Dendrites are rarely longer than 2 mm. and are connected to either the axons or dendrites of other cells. These connexions receive impulses from other nerves or relay the signals to other nerves. The human brain has approximately 10,000 connexions between one nerve and other nerves, mostly through dendritic connections.

Neurons are, of course, not working in splendid isolation, but are interconnected into different circuits (“neural networks”), and each circuit is tailored to process a specific type of information.

### 2.2.2 Cerebral cortex

The cerebral cortex constitutes the outermost layer of the cerebrum and physically it is a structure within the brain that plays an important role in memory, perceptual awareness, attention, thought, consciousness and language. Normally, it is called “grey matter” for its grey color and it is formed by neurons and “gray fibers” covered by a dielectric called myelin. Myelinated axons are white in appearance, this characteristic is the origin of the name “white matter,” and it is localized below the grey matter of the cortex. Their composition is formed predominantly by myelinated axons interconnecting different regions of the nervous central system.

The human cerebral cortex is 2-4 mm thick. The cortical surface is highly convoluted by ridges and valleys of varying sizes and thus increases the neuronal area; the total area is as large as 2.5 m<sup>2</sup> and includes more than 10<sup>10</sup> neurons. The cortex consists of two symmetrical hemispheres—left and right—which are separated by the deep longitudinal fissure (the central sulcus). Each cerebral hemisphere is divided into lobes, which are named for the skull bones overlying each one: the frontal lobe, involved with decision-making, motor speech, problem solving, and planning; temporal lobe, involved with

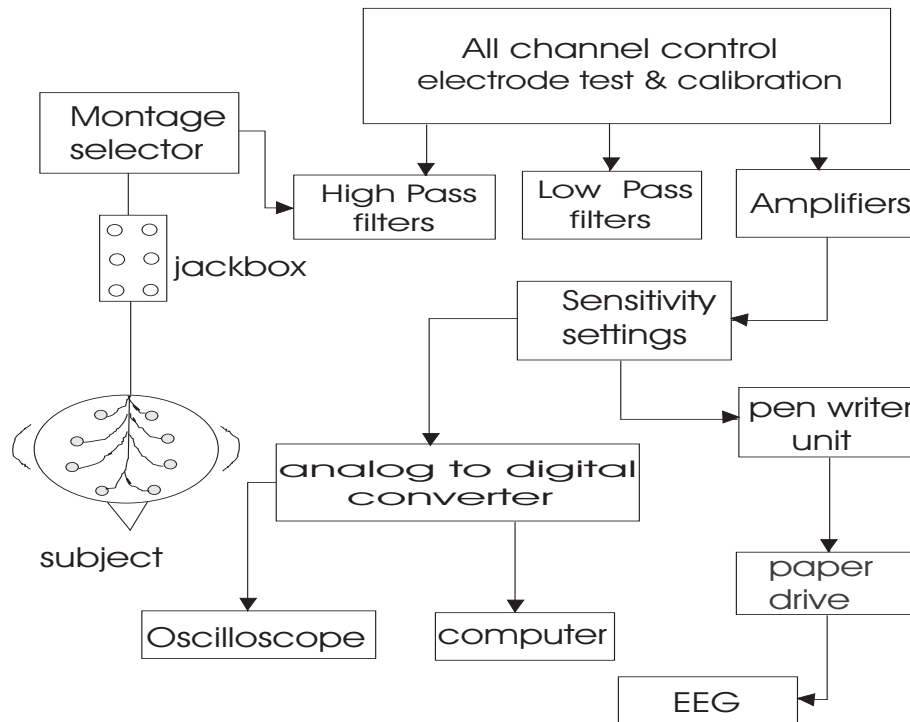


**Figure 2.2:** Cerebral cortex and its four lobes.

memory, sensory speech, emotion, hearing, and language; the parietal lobe, involved in the reception, reading comprehension and processing of sensory information from the body; and the occipital lobe, involved with vision, see Fig.2.2.

## 2.3 The EEG machine: An Overview

Since the 1920s, the EEG machine has been the subject of study and development, and its use is common in medical practice. Although the basics elements are similar to those employed in the days of Hans Berger, many of the components and circuits have been vastly improved by fully computerized systems. The EEG machines are equipped with many signal processing tools, delicate and accurate measurement electrodes with enough memory necessary to very long-term recording (several hours), channel machine expansion (8, 16, 20, 24, 32, 64, etc., even more channels) and invasively or noninvasively EEG record. Fig.2.3 shows a block diagram of the circuit containing the patient and EEG machine, which consists of both digital and analogue components. There are two important factors (one biophysical and one physiological) that limit the EEG interpretation. First, for any scalp recorded EEG signal there are an infinite number of sources within the vol-



**Figure 2.3:** Block diagram of the electroencephalographic machine (with permission of Ebersole and Pedley [2003]).

ume of the brain that can explain, or “fit,” the scalp recorded signal. Therefore, one or more generators in different locations in the brain can produce the same EEG findings at the scalp. This means it is theoretically impossible to know the location of the EEG generation in the brain with only scalp recorded information. This is referred to as the *inverse problem*. In contrast, if the anatomical source, intensity and orientation of the electrical generators in the brain are known, then the EEG findings on scalp electrodes can be accurately predicted. This is referred to as the *forward problem*. However, it is the inverse problem that the specialist is confronted with in clinical practice. Because the localization of EEG sources within the brain is so important, the search for methods to help solve the inverse problem (referred to as *source localization*) is currently a central theme in EEG research. Fortunately, in routine EEG practice a formally trained electroencephalographer can greatly narrow the number of possible solutions of the inverse problem.

The second factor that limits interpretation is that EEG signal abnormalities sometimes do not appear completely localized in the area of the brain where the main pathological condition resides. That is, an abnormal cortical signal may occasionally appear distant from the most prominent functional or structural damage. For example, the substance of a large structural lesion (e.g., tumor, stroke, ect.) is typically electrically silent. However, border-

ing tissue involved in EEG generation produces the abnormal activity seen. Indeed, in some cases EEG abnormalities recorded over the middle and anterior temporal areas may occur in the setting of a structural abnormality that more directly involves deep hemispheric structures or the frontal, parietal or posterior temporal lobes, as discussed in [Fisch \[1999\]](#).

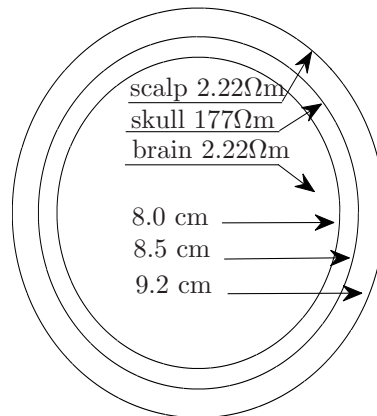
Despite the two factors explained above, the EEG is at the present time the tool with more diagnostic applications in clinical environment, because it has several strong advantages as a tool of exploring brain activity; for example, its time resolution is close to milliseconds (very high), it is relatively low cost and widespread used in clinical environment. Other methods of looking at brain activity, such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) have time resolution in order of seconds or minutes. EEG measures the brain's electrical activity directly, while other methods record changes in blood flow (e.g., fMRI, single photon emission computed tomography (SPECT)) or metabolic activity (e.g., PET), which are indirect measurements of the brain electrical activity. Taking advantage of the high-temporal-resolution from the EEG and combining with the high-spatial-resolution data from the fMRI, it is possible to use both techniques simultaneously, however, it is necessary a good synchronization between them because otherwise the data sets do not necessarily represent exactly the same brain activity. Moreover, there are technical difficulties associated with combining these two modalities such as the need to remove the MRI gradient artifact present during MRI acquisition and the presence of artifacts (i.e. the ballistocardiographic artifact that results from the pulsatile motion of blood and tissue) from the EEG. Furthermore, currents can be induced in moving EEG electrode wires due to the magnetic field of the MRI.

However, EEG and MEG are a good complimentary methods due to high-time-resolution techniques and the possibility to obtain data with high resolution.

### 2.3.1 Electrodes

Following [Sanei and Chambers \[2007\]](#), only large populations of active neurons can generate potentials high enough to be recordable using the scalp electrodes. This potentials travel through the human head that consists of three principals layers: the scalp, skull, brain and many other thin layers in between (see [Fig.2.4](#)). Every layer attenuates the EEG signal, specially the skull, that approximately attenuates one hundred times more than the soft tissue. Scalp electrodes are applied after determining their precise scalp location and after preparing the scalp to reduce electrical impedance.

Electrodes consist of a conductor attached to a wire that leads to a plug that is inserted into the input of the “all channel control” (see [Fig.2.3](#)). Elec-



**Figure 2.4:** Resistivities and thicknesses ( $\Omega = \text{ohm}$ ) of the three main layers of the brain (adopted from [Sanei and Chambers \[2007\]](#)).

trodes conduct electrical potentials from the head to the EEG machine with the help of a conductive paste that connects skin with electrodes. The paste is a electrical conductor whose functions range from the helping of transmission of potentials and decreasing the movement artifacts.

Electrodes are made of metal such as tin, silver and gold. Proper electrodes must be good electrical conductors and must be in good contact with the electrolyte paste that covers the skin. Voltage at any given instant is obtained as the difference in voltage between 2 electrodes sites on the body, at least one of which is placed on the scalp.

Another class of electrodes are the depth EEG electrodes. These electrodes are used to define the targets for surgical destruction and as stimulating electrodes for treatment of movement disorders. EEG recording depth electrodes consist usually of a bundle of fine wires that terminate at different cylindrical contacts along the length of the depth electrode and thus allow for recording from different depths.

Apart of these, there are a number of highly specialized electrodes, when shape and feature depend on the localization or application, for example, nasopharyngeal, sphenoidal, subdural and epidural electrodes are used to analyze, detect and localize epileptiform activity.

### 2.3.2 Jackbox and Montage Selector

Since EEG was first recorded from humans by Hans Berger in 1929, who used two electrodes applied on the front and back of the head, various systems have been used over the years. The EEG signal is transmitted from the patient through conducting electrodes to the electrode board, also called the *electrode box* or *jackbox*. Electrodes are labeled on the jackbox in accordance

to the Committee of the International Federation of Societies for Electroencephalography and Clinical Neurophysiology (IFSECN), that recommended a specific system of electrode placement under standard conditions for use in EEG laboratories. Specific measurements from body landmarks are used to determine the placement of electrodes. From these anatomical landmarks, specific measurements are made, and then 10% to 20% of a specified distance is used as the electrode interval. This enables replication consistently over time and between laboratories. The American Clinical Neurophysiology Society (formerly the American Electroencephalography Society) has recommended using a minimum of 21 electrodes in a system called as *the international 10-20 system*. Odd-numbered electrodes are placed on the left side of the head, and even-numbered electrodes, on the right side of the head. Electrodes are identified according to the brain region proximate to their location: on the frontopolar region the electrodes are labeled as “Fp;” frontal, “F;” central, “C;” temporal, “T;” parietal, “P;” and occipital, “O” (See Fig.2.5).

In many applications such as Brain Computer Interface (BCI) and study of mental activity it is necessary to use at least a small number of electrodes related with the regions that correspond to the movement, and the electrode arrangement usually is based on 10-20 setting system. Fig.2.6 illustrates a typical set of EEG signals during approximately five seconds of normal adult brain activity. Note in this figure that even though the EEG signal contains brain information, sometimes some displayed channels correspond to electrocardiogram (ECGp and ECGn).

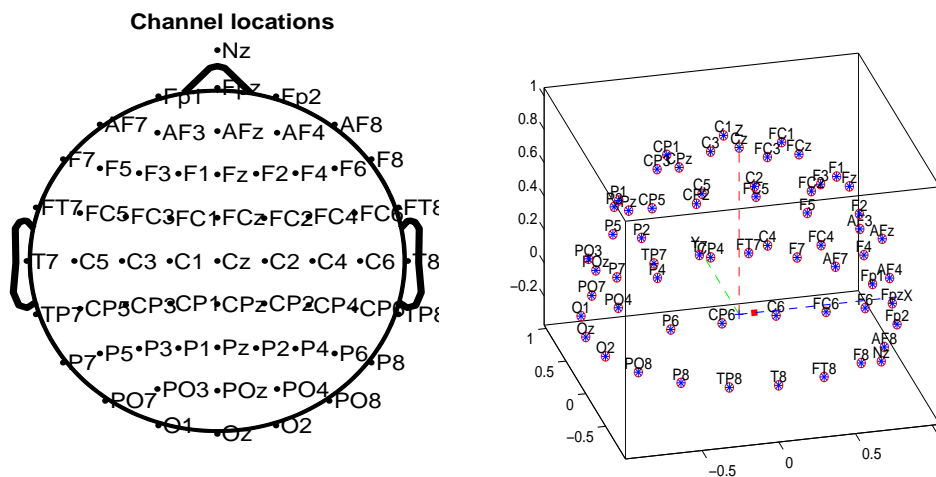
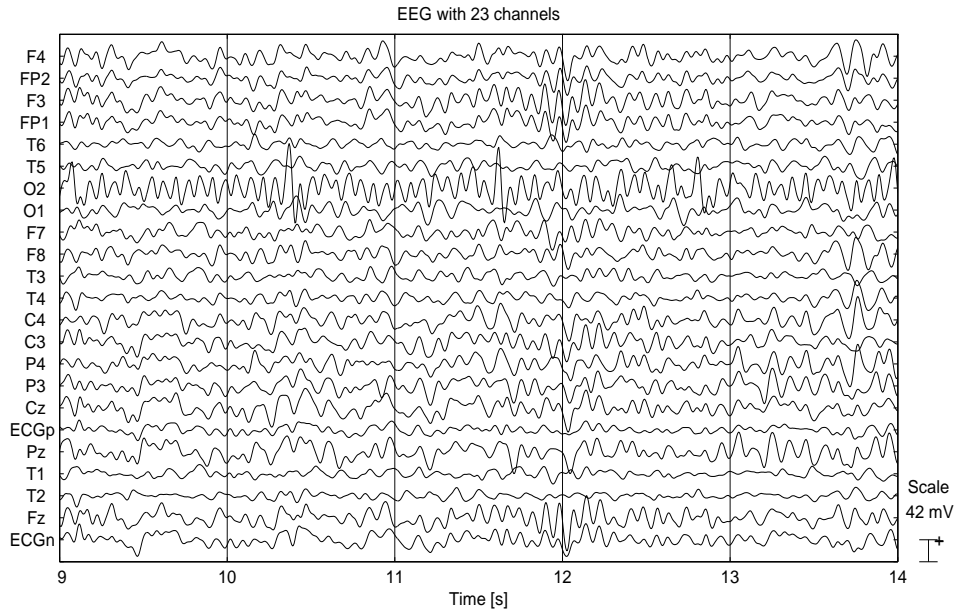


Figure 2.5: Neurophysiology 10-20 System in 2-D (left) and 3-D (right).

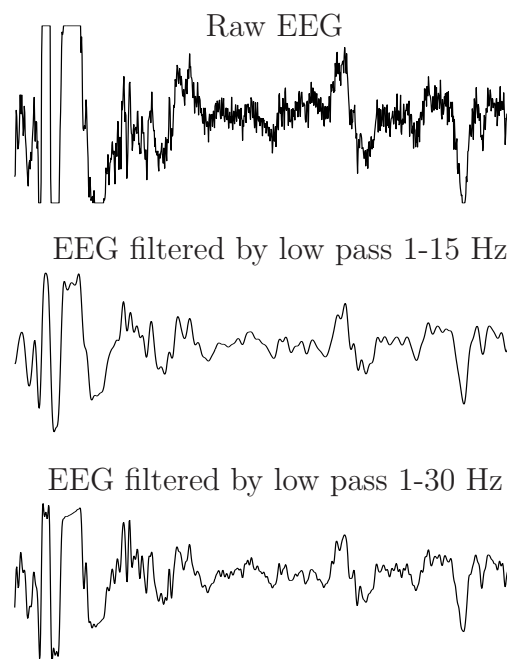


**Figure 2.6:** Typical set of EEG signals during approximately five seconds of normal adult brain activity. Note in this figure the existence of two channels corresponding to electrocardiogram (ECGp and ECGn).

On the other hand, EEG Montages are arrangements of electrode derivations designed to enhance recognition of EEG patterns. Common montage types are:

- Bipolar montage. Each channel recorded is obtained by the difference between two adjacent electrodes.
- Referential montage. Each channel recorded is the difference between a certain electrode with a designated reference electrode. Do not exist standard reference position.
- Average reference montage. There is an averaged signal that is used as the common reference for each channel. This average is obtained from the average of the outputs of all the amplifiers.
- Laplacian montage. Each channel represents the difference between an electrode and a weighted average of the surrounding electrodes (“nearest neighbors”).

Each of these montages has different advantages and disadvantages presented in [Ebersole and Pedley \[2003\]](#).



**Figure 2.7:** Different effects of using filters in the EEG

### 2.3.3 EEG filters

Filters are used to exclude waveforms of relatively high or low frequency from the EEG so that waveforms can be recorded clearly and without distortion. After EEG signals are recorded and amplified, the output is filtered to eliminate specified frequency components, as described earlier. The high-pass filter removes components of the signal that have frequencies less than a specified value and keep the amplitude of the fast waves; the low-pass filter attenuates components with frequencies higher than a certain value and keep the amplitude of slow waves. A special filter, a 60 Hz (50 Hz in Europe) notch filter, is necessary to remove electrical interference generated by electrical power line. It is desirable in EEG recording to minimize the use of typical digital filters with low cutoff frequency of 0.1 Hz and a high cutoff frequency of 70 Hz. Filters distort both the amplitude and the interchannel phase of signals. Sometimes, artifacts originated from muscle (commonly high-frequencies) and others artifacts from movement or sweat potentials (low-frequencies) recommend the use of more stringent filtering. Fig.2.7 shows how filtering can modify a signal into a nearly uninterpretable tracing. It is necessary a good documentation in filtering before applying on the EEG recording, so that the specialist can interpret their possible influence.

### 2.3.4 Amplifiers

In general, EEG machines contain amplifiers that are compound devices whose function is not only to increase voltage. EEG amplifiers also contain filters, voltage dividers, input and output jacks, and calibration devices. Owing to the fact that EEG signal amplitudes are very small, the amplifiers' sensitivity, defined as the amount of voltage required to deflect the recording pens a given distance, has units that oscillate between millivolts per centimeter or microvolts per millimeter. A typical sensitivity value for the EEG is  $7 \mu\text{V}$  per millimeter, leading to pen deflections of 30 to 20 mm for typical EEG input voltages.

The frequency response of the EEG amplifier is considered as flat over a wide range of input values. In practice, the settings that the clinician chooses for the filters determine the range of linear frequency response. To avoid distortion, it is advisable to choose an amplifier with linear frequency response, both high and low frequency components, over the expected range of input voltages.

Signals from each electrode are led to a differential amplifier that amplifies a signal difference. This signal difference is a graph of voltage that depends of the ground electrode on the scalp (type of montage) and another signal in relation to the same reference. This process eliminates noise voltages present in both electrodes, and in addition it isolates the two inputs for each channel from system ground.

### 2.3.5 Analog and digital conversion

All digital instruments contain analog amplifiers and filters that produce an EEG signal suitable for digital processing. Digital processing begins with the transformation of the continuous *analog* signal into a *digital* signal. The digital signal consists of a series of discrete, discontinuous data points separated by equal intervals of time. This transformation is performed by an analog to digital converter, or ADC. It has three key attributes that determinate how accurately the analog signal will be reproduced in digital form:

- the sampling rate (i.e., samples per second according to Nyquist theorem)
- the number of amplitude levels (amplitude resolution in terms of *bits*) that can be resolved, and
- the input voltage range (the range of voltage coming from the analog amplifiers that the ADC is set to analyze)

### 2.3.6 The signal display: oscilloscope and computer

Although digital systems can easily output the recorded signal to a variety of display media, the most practical way of viewing the EEG is on a computer (monitor or flat panel display). According to the American Clinical Neurophysiology Society the display should have a minimum of 2 pixels of resolution per vertical millimeter, but there are higher resolution monitors available.

In summary, EEG enables clinicians to study and analyze electrical fields of brain activity by recording amplified voltage differences between electrodes placed on the scalp, directly on the cortex (e.g., with subdural electrodes), or within the brain (with depth electrodes). For each electrical field, the specialist tries to determine the nature, localization, and configuration of the generator of EEG patterns and whether there are normal or abnormal neural discharges.

## 2.4 Introduction to the EEG analysis

Most of the brain disorders are diagnosed by visual inspection of EEG signals and the analysis is a rational and systematic process requiring a series of orderly steps characterizing the recorded electrical activity in terms of specific descriptors or *features* and measurements as viewed in Table.2.1. For example, an EEG from an 8 year old child, some 2 Hz waves are identified in the awake EEG. This activity must then be characterized according to their location, voltage, waveform, manner of occurrence, frequency, amplitude modulation, synchrony and symmetry. A change in any of these features might entirely change the significance of the 2 Hz waves finding this difference as abnormal.

Some clinical information is required before the EEG analysis is begun, by example the patient's *age* and *state*. Both age and birth date should be part of the EEG record. For example, there are clearly defined differences between the EEG of a premature infant with a conceptional age of 36 weeks, but there are no important or sharply delineated differences between the EEG of a 3 year old child and that 4 year old child described in [Ebersole and Pedley \[2003\]](#).

The clinical experts in the fields are familiar with manifestation of brain rhythms in the EEG signals and it is important to recognize that the identification of a particular activity or phenomenon may depend on its "reactivity" (see Table.2.1). An important element of the recording and its analysis is the testing of the reactions, or responses, of the various components of the EEG to certain physiological changes.

**Table 2.1:** Essential features of EEG analysis described in Ebersole and Pedley [2003].

- 
1. Frequency or wavelength
  2. Voltage
  3. Waveform.
  4. Regulation
    - a. Frequency
    - b. Voltage
  5. Manner of occurrence (random, serial, continuous)
  6. Locus
  7. Reactivity (eye opening, mental calculation, sensory stimulation, movement, affective state)
  8. Interhemispheric coherence (homologous areas)
    - a. Symmetry
      - i. Frequency
      - ii. Voltage
    - b. Synchrony
      - i. Wave
      - ii. Burst
- 

Specification of the reactivity of a given activity, rhythm or pattern is essential for the identification and subsequent analysis of the activity and may clearly differentiate it from another activity with similar characteristics. For example, in healthy adults, the amplitudes and frequencies of brain rhythms change from one state of the human to another, such as wakefulness and sleep. Similarly, a series of rhythmic, high voltage 3 to 4 Hz waves in the prefrontal leads (just over the eyes) occurring in association with arousal in a young child may be normal, but a similar burst occurring spontaneously and not associated with arousal may be abnormal.

### 2.4.1 Brain rhythms and waveforms

The electrical activity of the cerebral cortex is often called as *rhythm* because this recorded signals exhibit oscillatory, repetitive behavior. The diversity of EEG rhythms is enormous and depends, among many other things, on the mental state of the subject, such as the degree of attentiveness, waking, and sleeping. The rhythms usually are conventionally characterized by their frequency range and relative amplitude.

On the other hand, there are five brain waves characterized by their frequency bands. These frequency ranges are alpha ( $\alpha$ ), theta ( $\theta$ ), beta ( $\beta$ ), delta ( $\delta$ ), and gamma ( $\gamma$ ) and their frequencies range from low to high fre-

quencies respectively. The alpha and beta waves were introduced in 1929 by Berger. In 1938, Jasper and Andrews found waves above 30 Hz that labeled as “gamma” waves. A couple years before, in 1936, Walter introduced the delta rhythm to designate all frequencies below the alpha range and he also introduced theta waves as those frequencies within the range of 4-7.5 Hz. In 1944 the definition of a theta wave was introduced by Wolter and Dovey in [Sanei and Chambers \[2007\]](#).

Alpha waves are over the occipital region of the brain and appear in the posterior half of the head. The normal range for the frequency of the occipital alpha rhythm in adults is usually given as 8 to 13 Hz, and commonly appears as a sinusoidal shaped signal. However, sometimes it may manifest itself as sharp waves. In such cases, the alpha wave consist of a negative and positive component that appears to be sharp and sinusoidal respectively. In fact, this wave is very similar to the morphology of the brain wave called rolandic mu ( $\mu$ ) rhythm.

Delta waves lie within the range of 0.5-4 Hz. These waves appear during deep sleep and have a large amplitude. It is usually not encountered in the awake, normal adult, but is indicative of, e.g., cerebral damage or brain disease.

Theta waves are the electrical activity of the brain varying the range of 4-7.5 Hz and its name might be chosen to origin assumption from thalamic region. The theta rhythm occurs during drowsiness and in certain stages of sleep or consciousness slips towards drowsiness. Theta waves are related to access to unconscious material, creative inspiration and associated to deep meditation.

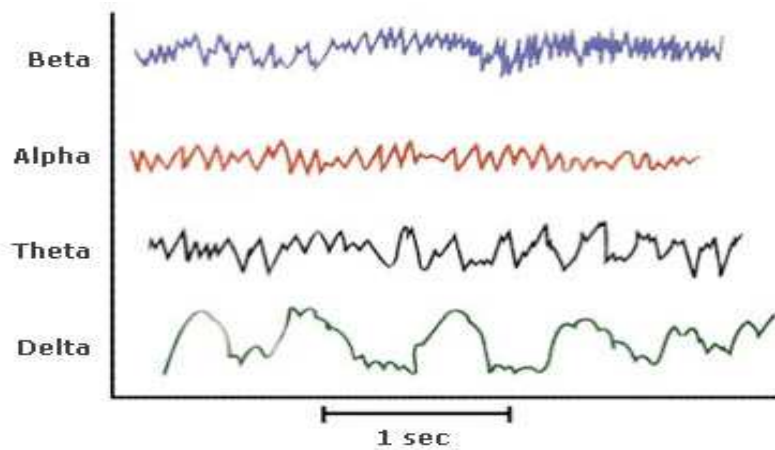
Beta waves are within the range of 14-26 Hz and consists in a fast rhythm with low amplitude, associated with an activated cortex and observed during certain sleep stages. This rhythm is mainly present in the frontal and central regions of the scalp.

Gamma waves (sometimes called the fast beta waves) are those frequencies above 30 Hz (mainly up to 45 Hz) related to a state of active information processing of the cortex. The observation of gamma rhythm during finger movement is done simply by using an electrode located over the sensorimotor area and connected to a high-sensitivity recording system.

Other waves frequencies much higher than the normal activity range of EEG have been found in the range of 200-300 Hz. The localization of these frequencies take place in cerebellar structures of animals, but they do not play any role in clinical neurophysiology. Most of the above rhythms may persist up to several minutes, while others occur only for a few seconds, such as the gamma rhythm.

[Fig.2.8](#) shows typical normal brain waves. There are also less common

rhythms introduced by researchers such as Phi ( $\varphi$ ), Kappa ( $\kappa$ ), Sigma ( $\sigma$ ), Tau ( $\tau$ ), Chi ( $\chi$ ), Lambda ( $\lambda$ ) and transient waveforms associated to two sleep states, commonly referred to as non-REM (Rapid Eye Movement) and REM sleep: vertex waves, sleep spindles, and K complexes described in Sanei and Chambers [2007] and Sörnmo and Laguna [2005].



**Figure 2.8:** Typical normal brain waves in the EEG

It is often difficult to understand and detect the brain rhythms and waves from the scalp EEGs, even with trained eyes. New applications in advanced signal processing tools, however, should enable analysis and separation of the desired waveforms from the EEGs. Definitions such as foreground and background EEG are very subjective and totally depends on the abnormalities and applications. Possibly it is more useful to divide the EEG signal into two general categories: the spontaneous brain activity (the “background EEG”); and brain potentials which are evoked by various sensory and cognitive stimuli (evoked potentials, EPs).

### 2.4.2 Artifacts

Analysis of EEG activity usually raises the problem of differentiating between genuine EEG activity and that which is introduced through a variety of external influence. These *artifacts* may affect the outcome of the EEG recording. Artifacts originate from a variety of sources such as eyes movement, the heart, muscles and line power. Their recognition, identification, and eventual elimination are a primary responsibility of the EEG expert. Even the most experienced neurophysiologist cannot always eliminate all artifacts in EEG records. However, it is always a major goal to identify the artifactual activity and be sure that it is not of cerebral origin and should not be misinterpreted as such.

Following [Ebersole and Pedley \[2003\]](#) and [Fisch \[1999\]](#), artifacts are generally divided into two groups: physiological and non-physiological. Physiological artifacts usually arise from generator sources within the body but not necessarily the brain, for example, eye movements; electrocardiographic and electromyographic artifacts, galvanic skin response and so on. Biological generators present in the body may produce artifacts when an EEG recording is made directly from the surface of the brain. Nonphysiological artifacts come from a variety of sources such as instrumental and digital artifacts (electronic components, line power, inductance, etc.), electrode artifacts, environment, etc.

As technology expands and additional equipment is developed and put into clinical use, novel artifacts will appear. Then, a correct artifact filtering strategy should on the one hand eliminate unnecessary amount of information that has to be eliminated, and on the other hand maintain or ensure that the resulting information is not affected by undetected artifacts. Sometimes visual artifacts inspections could be a good alternative in cases when the artifacts are relatively easy detected by the EEG experts. However, there is the possibility that during the analysis of EEG databases these patterns from artifacts cause serious misinterpretation and then reduce the clinical usefulness of the EEG recordings.

### 2.4.3 Abnormal EEG patterns

Any variation in EEG patterns for certain states of the subject indicate abnormality. This may be due to many causes such as distortion and loss of normal patterns, increased occurrence of abnormal patterns, or disappearance of all patterns. In most abnormal EEGs, the abnormal EEG patterns do not entirely replace normal activity: they appear only intermittently, only in certain head regions, or only superimposed on a normal background.

An EEG is considered abnormal if it contains (a) generalized intermittent slow wave abnormalities, commonly associated in the delta wave range and brain dysfunctions, (b) bilateral persistent EEG, often associated with impaired conscious cerebral reactions, and (c) focal persistent EEG usually associated with focal cerebral disturbance.

The classification of the three categories presented before is not easy and needs to be extended to several neurological diseases and any other available information. A precise characterization of the abnormal patterns leads to a clearer insight into some specific neurodegenerative diseases such as epilepsy, Parkinson, Alzheimer, dementia and sleep disorders, or specific disease processes, for example Creutzfeldt-Jakob disease (CJD) described in [Sanei and Chambers \[2007\]](#). However, following [Fisch \[1999\]](#), recent studies have demonstrated that there is correlation between abnormal EEG patterns,

general cerebral pathology and specific neurological diseases.

## 2.5 EEG applications

Nowadays there are many applications that have been developed based on the EEG monitoring and especially in the detection of neurodegenerative diseases, brain computer interface (BCI) and sleep disorders.

This section presents the most relevant applications developed in the literature without forgetting that the EEG analysis, as important part within neuroengineering, is a field that grows day after day in new applications.

### 2.5.1 Seizures and epilepsy

Epilepsy is a neurological disorder characterized by sudden recurrent and transient perturbations of mental function (e.g. speech impairment, behavioral disturbances, body movements) that result from excessive neural discharges of a group of brain cells. Patients who are suspected of suffering epileptogenic foci in their brain are usually subjected to an electroencephalogram (EEG) recording in the hospital. This process records the electrical activity of the neurons in the brain, and can find or indicate abnormalities in the nervous central system. Even though the procedure of this technique has essentially remained the same, EEG data analysis has profoundly evolved during the last three decades. The common procedure for epilepsy seizure detection is based on brain activity monitorization via EEG data. This usually involves identifying sharp repetitive waveforms or rhythmic patterns in the EEG data that indicate the beginning of the seizure. This process consumes a lot of time, especially in the case of long recordings, but the major problem is the subjective nature of the analysis among specialists when analyzing the same record, as discussed in [Acir et al. \[2005\]](#). From this perspective, the identification of hidden dynamical patterns is necessary because they could yield insight into the underlying physiological mechanisms that occur in the brain. In addition, these analyzes could characterize the non-stationary behavior and isolate abnormal activity in the EEG described in [James and Lowe \[2000\]](#) with the final objective of developing automatic seizure detection systems.

Automatic detection of electroencephalographic seizures has been investigated for years. However, so far, no technique has demonstrated to have competitive sensitivity and specificity values. The presence of artifacts makes the detection more difficult. The availability of a good algorithm for seizure detection would simplify the review of hours and hours of EEG recordings. It

would also be of great value if the detector could help to distinguish between real epileptic seizures and artifacts during non-epileptic events.

### 2.5.2 Brain computer interface (BCI)

Brain-Computer-Interface (BCI) provides a new tool for communication between the human brain and the external world without any physical contact and it is oriented to patients who suffer from severe motor impairments (i.e. head trauma, late stage of Amyotrophic Lateral Sclerosis (ALS), severe cerebral palsy and spinal injuries). BCI system may be used as an alternative form of communication by mental activity.

Following the work of Hans Berger in 1929, in the 1970s, researchers developed a primitive control system based on electrical activity recorded from the head. Jacques Vidal from UCLA's Brain-Computer Interface Laboratory designed a communication channel effective to control a cursor through a two-dimensional maze using a single-trial visual-evoked potentials presented in [Sanei and Chambers \[2007\]](#). Since 1990s, BCI took a great leap forward, owing to more detailed knowledge of the EEG signal, other imaging techniques and rapid advances in computer technology, described in [Vallabhaneni et al. \[2005\]](#).

From a practical point of view, BCI systems consist of a series of implementations that generally allow the user to select from a menu of options, icons or characters on a screen, whether in motion or stopping a cursor of a dropdown list of options. However, there is a non-trivial problem in the generation and detection of the EEG control signal.

Other problem is related to non-stationarity in EEG signals coming from changes of task that could negatively affect the performance of the BCI systems. Although there are methods proposed to alleviate this problem, such as the Regularization of Common Spatial Patterns (CSP), described in [Wojcikiewicz et al. \[2011\]](#), there are still more research to do. But the major problem in BCI is separating the control signals from the background EEG because in addition to artifacts and electrical noise there are other undesired signals such as physiological noise. Meanwhile, the identification of cortical connectivity, related to different brain activities (or tasks), has to be studied and exploited.

There are two fundamental and interrelated steps to the design and use of a BCI: one is the mental process of the user which encodes commands in the EEG signal, and another is the BCI which, by employing EEG signal processing (preprocessing, feature extraction, classification), translates patterns into commands which control the device. [Fig.2.9](#) outlines the basic components of a BCI. Since the BCI must operate in real time, it is very important that the signal processing does not introduce large time delays.

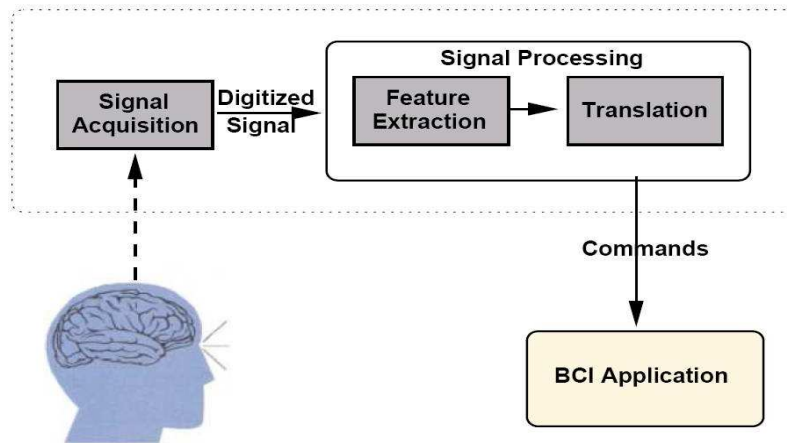


Figure 2.9: Block diagram of the brain computer interface

### 2.5.3 Sleep disorders

The sleep state is one of the most important items of evidence for diagnosing mental disease. Sleep is characterized by a reduction on voluntary body movement, an increased rate of anabolism (the synthesis of cell structures), a decreased rate of catabolism (the breakdown of cell structures), and decreased reaction to external stimuli. Sleep is a dynamic process that consists in two distinct states that reflect different levels of neural activity where each state is characterized by a different type of EEG activity. Sleep consists of nonrapid eye movement (NREM) and REM sleep. NREM is further subdivided into four stages: I (drowsiness), II (light sleep), III (deep sleep), and IV (very deep sleep). Since NREM and REM are alternating cycles, a complete sleep cycle, from the beginning of stage I to the end of REM sleep, is necessary for the brain to remain healthy. Sleep deprivation makes a person drowsy and unable to carry out mathematical calculations, concentrate, physical good performance and other tasks. The rapid growth of sleep disorder medicine has led to an increasing role of the EEG analysis oriented to evaluation of sleep and sleep disorders.

Following [Sörnmo and Laguna \[2005\]](#), sleep disorders may be caused by several conditions of medical and physiological origin. It is possible to classify the sleep disorders in four groups: insomnia (impossibility of maintaining sleep), hypersomnia (excessive sleep or somnolence), circadian rhythm disorders (i.e. “jet lag”) and parasomnia (deviations in the normal sleep pattern). Each of the different types of sleep disorder presents certain manifestations in the EEG, however, not only the EEG is used to sleep analysis. Clinical studies of sleep disorders relies primarily on two major techniques: the overnight polysomnogram (PSG) and the Multiple Sleep Latency Test (MSLT) described in [Ebersole and Pedley \[2003\]](#). These two techniques serve as the

gold standard for sleep evaluation that, in combination with computer based analysis, make it possible to quantify similarities that may exist between different types of signals and open the possibility of classifying the sleep disorders.

### 2.5.4 Event related potential (ERP)

Event related potential (ERPs) are EEG recordings that reflect the responses of the brain to events external or internal to the organism, and have remained as a useful diagnostic tool, in both psychiatry and neurology, being widely useful in BCI.

ERPs are voltage fluctuations in the EEG whose origin are located in the brain and are associated in time with a physical or mental event. These potentials are recorded through electrodes on the human scalp EEG and currently extracted through filtering and signal averaging, which manifest as a form of transient waveform whose morphology depends on the stimulation.

Individual ERP presents a very low amplitude with positive deflections (P) and negative deflections (N), such as the P300 or N400 waves. The digits indicate the time in terms of milliseconds after the stimuli, i.e. N400 has the negative deflection near 400 milliseconds after stimulus onset. Evoked potentials resulting from *auditory* (AEP), visual (VEP), and *somatosensory* (SEP) stimulation are the most commonly used in clinical environment. The P300 (or P3) component is the most important and studied component of the VEPs. P300 is a widely used test for cognitive function but its clinical use remains controversial. Several studies have reported significant differences between normal and patient populations regarding conditions such as dementia, head injury, and multiple sclerosis.

One of the principal methods for detecting the P300 wave has been EEG signal averaging. By averaging, the background EEG activity is canceled and it may be modeled like random noise while the P300 wave remains basically unaltered. However, there are limitations to the averaging technique and applications for which it is not suitable. Other techniques have been employed such as principal component analysis, independent component analysis and source localization. Each one with advantages and disadvantages in separating the P300 components presented in [Elting et al. \[2003\]](#).

## 2.6 Summary and conclusions

In this chapter the fundamental concepts in the nervous system and the EEG generation have been briefly explained. Several concepts in visual analysis of the EEG, brain rhythms, artifacts and abnormal EEG patterns, including

EEG applications such as epilepsy detection, brain computer interface (BCI), sleep disorders and event related potentials (ERPs) have been reviewed.

EEG is widely used as a diagnostic tool in clinical routine with an increasing develop of both analytical and practical methods. Its simplicity, low cost and higher temporal resolution of EEG maintains this tool to be considered in applications such as epilepsy seizures detection, sleep disorders and BCI.

Future work implies the design of new EEG artifacts elimination methods, feature extraction to obtain possible hidden information and dimensionality data reduction.

# Chapter 3

## EEG signal processing

### 3.1 Introduction

Since the integration between classical and modern biomedical signal processing with the engineering<sup>1</sup>, new fields have been activated in a new area known as “neuroengineering.” As discussed in [Thakor and Tong \[2006b\]](#), Clinical neuroengineering has active fields such as neural prosthesis, brain computer interface (BCI), new clinical imaging techniques and treatment tools with EEG, evoked potentials (EPs), MEG and fMRI. Nowadays, there are several processing methods, tools and algorithms for processing EEG signals for helping devising in new treatments, obtaining new measurements of brain activity and detecting brain diseases based on EEG signals.

EEG signals not only represent the brain function but also the status of the whole body, i.e., a simple action as blinking the eyes introduces oscillations in the EEG records. Then, the EEG is a direct way to measure neural activities and it is important in the area of biomedical research to understand and develop new processing techniques.

In EEG analysis, there are crucial considerations to take into account:

- It is a dynamic signal which exhibits a non-stationary behavior, and it could be necessary to use signal segmentation or EEG epochs analysis.
- Some abnormal EEG patterns may be normal at younger ages being necessary detection and classification algorithms.

Apart from EEG pre-processing techniques such as filtering and denoising,

---

<sup>1</sup>Examples of classical methods are: discrete Fourier transform (DFT), power spectral density (PSD) or short-time Fourier transform (STFT). And modern methods such as wavelet transform, time-frequency distributions (TFDs), independent component analysis (ICA), amongst others.

EEG processing could be considered as a “pattern recognition system” consisting of two tasks: feature extraction and classification<sup>2</sup>. The aim of the first ones is to identify “patterns” of brain activity and their results could be used as input to the classifier. The performance of a pattern recognition system depends on both the features and the classification algorithm employed.

In this chapter we will introduce methods for EEG signal pre-processing and processing. The chapter is organized as follows: Section 3.2 is an overview of different alternatives in EEG signal modelling, Section 3.3 reviews some algorithms oriented to eliminate artifacts and noise, sections 3.5 and 3.4 introduce a more theoretical view of algorithms in feature extraction, Section 3.6 provides some classification algorithms especially with applications to EEG signals and finally, Section 3.7 gives a summary and conclusions of this chapter.

## 3.2 Modelling and segmentation

### 3.2.1 EEG signal modelling

Modelling the brain activities is not an easy task as compared with modelling any other organ. First literature related to EEG signal generation includes physical model such as the model proposed by Hodgkin and Huxley, linear models such as autoregressive (AR) modelling, AR moving average (ARMA), multivariate AR (MVAR), Prony methods and so on. There are also methods based on no-linear models such as autoregressive conditional heteroskedasticity (GARCH), Wiener modeling and local EEG model method (LEM). More details about the methods described above can be found in [Sanei and Chambers \[2007\]](#), [Celka and Colditz \[2002\]](#).

Following [Senhadji and Wendling \[2002\]](#), other model relates a sampled EEG signal  $X(n)$  with relevant activities as elementary waves, background activity, noise and artifacts as:

$$X(n) = F(n) + \sum_{i=1}^{n_p} P_i(n - t_{pi}) + \sum_{j=1}^{n_a} R_j(n - t_{aj}) + B(n) \quad (3.1)$$

where  $F(n)$  is the background activity; the  $P_i$  terms represent brief duration potentials corresponding to abnormal neural discharges; the  $R_j$  terms are related to artifacts (discussed later in section 2.2) and  $B(n)$  is the measurement noise which is modeled as a stationary process. This model shows all

---

<sup>2</sup>For readers unfamiliar with statistical pattern recognition concepts, it is recommend to read Appendix B.

the EEG information including the abnormal EEG signal. This is a mathematical model rather than an EEG generation signal model, but facilitates the manipulation of concepts that are introduced in the next sections.

### 3.2.2 Signal segmentation

Signal segmentation is a process that divides the EEG signal by segments of similar characteristics that are particularly meaningful to EEG analysis. Traditional techniques of signal analysis, for example, spectrum estimation techniques, assume time-invariant signals but in practice, this is not true because the signals are time-varying and parameters such as amplitude, frequency and phase change over time. Furthermore, the presence of short time events in the signal causes a nonstationarity effect.

Non-stationary phenomena are present in EEG usually in the form of transient events, such as sharp waves, spikes or spike-wave discharges which are characteristic for the epileptic EEG, or as alternation of relatively homogenous intervals (segments) with different statistical features (e.g., with different amplitude or variance). The transient phenomena have specific patterns which are relatively easy to identify by visual inspection in most cases, whereas the identification of the homogeneous segments of EEG, known as quasi-stationary, requires a certain theoretical basis. Usually each quasi-stationary segment is considered statistically stationary with similar time and frequency statistics. This eventually leads to a dissimilarity measurement denoted as  $d(m)$  between the adjacent EEG frames, where  $m$  is a integer value indexing the frame and the difference is calculated between the  $m$  and  $(m - 1)$ th (consecutive) signal frames.

There are different dissimilarity measures such as autocorrelation, high-order statistics, spectral error, autoregressive (AR) modelling and so on, presented in [Sanei and Chambers \[2007\]](#). These methods are effective in EEG analysis but can not be efficient for detection of certain abnormalities due to the impossibility of obtaining segments completely stationary. It is then necessary to take into account a different group of methods potentially useful for detecting and analyzing non-stationary EEG signals where the segmentation does not play a fundamental role such as the time-frequency distributions (TFDs) (section [3.4.2](#)).

## 3.3 Denoising and filtering

Biomedical signals in general, but more particularly EEG signals, are subject to noise and artifacts which are introduced through a variety of external

influences. These undesired signals may affect the outcome of the recording procedure, being necessary a method that appropriately eliminates them without altering original brain waves. EEG denoising methods try to reject artifacts originated in the brain or body such as ocular movements, muscle artifacts, ECG etc.

Filtering is a signal processing operation whose objective is to process a signal in order to manipulate the information contained in the signal. In other words, a filter is a device that maps an input signal to an output signal facilitating the extraction (or elimination) of information (or noise) contained in the input signal. In our context, the filtering process is oriented to eliminate electrical noise generated by electrical power line or extracting certain frequency bands.

### 3.3.1 Lowpass filtering

Section 2.3.3 presented in Chapter 2 introduced the concept of filtering in EEG signals and effects that may result in the EEG when filters are used. Most frequently EEG signals contain neuronal information below 100 Hz, for example, epileptic waves lie below 30 Hz, it is possible to remove frequency components above this value simply using lowpass filters. In the cases where the EEG data acquisition system is unable to remove electrical noise as 50 or 60 Hz line frequency, it is necessary to use a notch filter to remove it. Although digital filters could introduce nonlinearities or distortions to the signal in both of amplitude and phase, there are digital EEG process that allow corrections of these distortions using commercial hardware devices. However, it should be better to know the characteristics of the internal and external noises that affect the EEG signals but these information usually is not available.

### 3.3.2 Independent component analysis (ICA)

ICA is of interest to scientists and engineers because it is a mathematical tool able to reveal the driving forces which underlie a set of observed phenomena. These phenomena may well be the firing of a set of neurons, mobile phone signals, brain images such as fMRI, stock prices, or voices, etc. In each case, a set of complex signals are measured, and it is known that each measured signal depends on several distinct underlying factors, which provide the driving forces behind the changes in the measured signals. These factors or *source signals* (that are primary interest) are buried within a large set of measured signals or *signal mixtures*. Following Stone [2004], ICA can be used to extract the source signals underlying a set of measured signal mixtures.

ICA belongs to a class of blind source separation (BSS) methods for estimating or separating data into underlying informational components. The term “blind” is intended to imply that such methods can separate data into source signals using only the information of their mixtures observed at the recording channels. BSS in acoustics is well explained in the “cocktail party problem,” which aims to separate individual sounds from a number of recordings in an uncontrolled environment such as a cocktail party. So, simply knowing that each voice is statistically unrelated to the others suggests a strategy for separating individual voices from mixtures of voices. The property of being unrelated is of fundamental importance, because it can be generalized to separate not only mixtures of sounds, but mixtures of other kind of signals such as biomedical signals, images, radio signals and so on.

The informal notion of unrelated signals can be associated to the more precise concept of *statistical independence*. If two or more signals are statistically independent of each other then the value of one signal provides no information regarding the value of the other signals. ICA works under this assumption and this concept plays a crucial role in separating and denoising the signals.

### ICA fundamentals

The basic BSS problem that ICA attempts to solve assumes a set of  $m$  measured data points at time instant  $t$ ,  $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_m(t)]^T$  to be a combination of  $n$  unknown underlying sources  $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_n(t)]^T$ . The combination of the sources is generally assumed to be linear and fixed, and the mixing matrix describing the linear combination of  $\mathbf{s}(t)$  is given by the full rank  $n \times m$  matrix  $\mathbf{A}$  such that

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad (3.2)$$

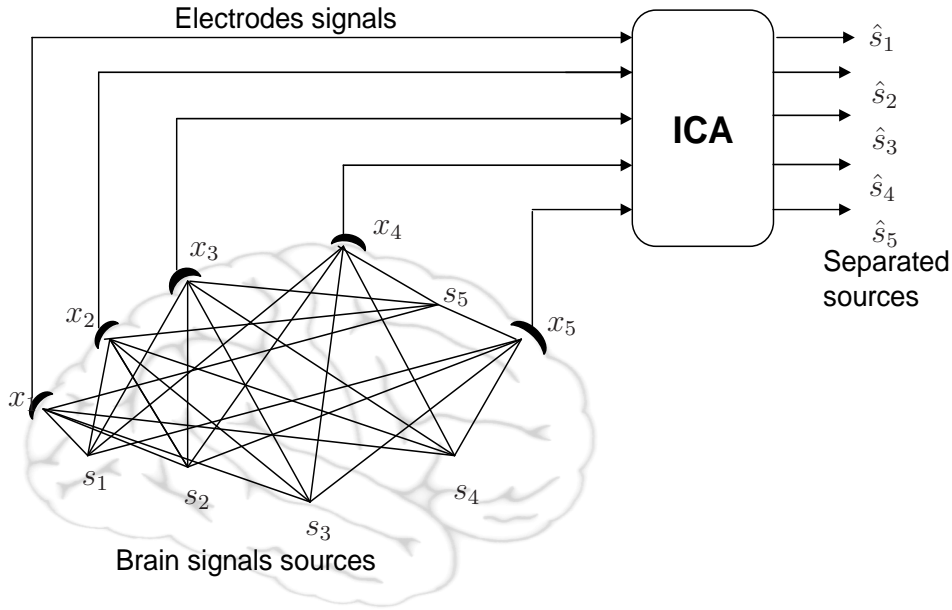
It is also generally assumed that the number of underlying sources is less than or equal to the number of measurement channels ( $n \leq m$ ).

The task of the ICA algorithms is to recover the original sources  $\mathbf{s}(t)$  from the observations  $\mathbf{x}(t)$  and this is generally equivalent to that of finding a separating (de-mixing matrix)  $\mathbf{W}$  such that

$$\hat{\mathbf{s}}(t) = \mathbf{W}\mathbf{x}(t) \quad (3.3)$$

given the set of observed values in  $\mathbf{x}(t)$  and where  $\hat{\mathbf{s}}(t)$  are the resulting *estimates* of the underlying sources. This idealistic representation of the ICA problem is described in Fig.3.1.

In reality the basic mixing model assumed in Eq.3.2 is simplistic and assumed for the ease of implementation. In fact, a perfect separation of the



**Figure 3.1:** General ICA process applied to EEG signals

signals requires taking into account some assumptions and the structure of the mixing process:

- *Linear mixing:* The first traditional assumption for ICA algorithms is that of linear mixing, a realistic model can be formulated as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \quad (3.4)$$

where  $\mathbf{A}$  is the linear mixing matrix described earlier and  $\mathbf{n}(t)$  is additive sensor noise corrupting the measurements  $\mathbf{x}(t)$  (generally assumed to be i.i.d. spatially and temporally white noise, or possibly temporally colored noise), as described in [James and Hesse \[2005\]](#).

In a biomedical signal context, linear mixing assumes (generally instantaneous) mixing of the sources using simple linear superposition of the attenuated sources at the measurement channel.

- *Noiseless mixing:* If observations  $\mathbf{x}(t)$  are noiseless (or at least the noise term  $\mathbf{n}(t)$  is negligible) then Eq.3.4 reduces to Eq.3.2. Whilst this is probably less realistic in practical terms, it allows ICA algorithms to separate sources of interest even if the separate sources themselves remain contaminated by the measurement noise.
- *Square mixing matrix:* So far it has been assumed that the mixing matrix  $\mathbf{A}$  may be non-square ( $n \times m$ ); in fact most classical ICA algorithms assume a *square-mixing* matrix, i.e.  $m = n$ , this makes the

BSS problem more tractable. From a biomedical signal analysis perspective the square-mixing assumption is sometimes less than desirable, particularly in situations where high-density measurements are made over relatively short periods of time such as in most MEG recordings or fMRI.

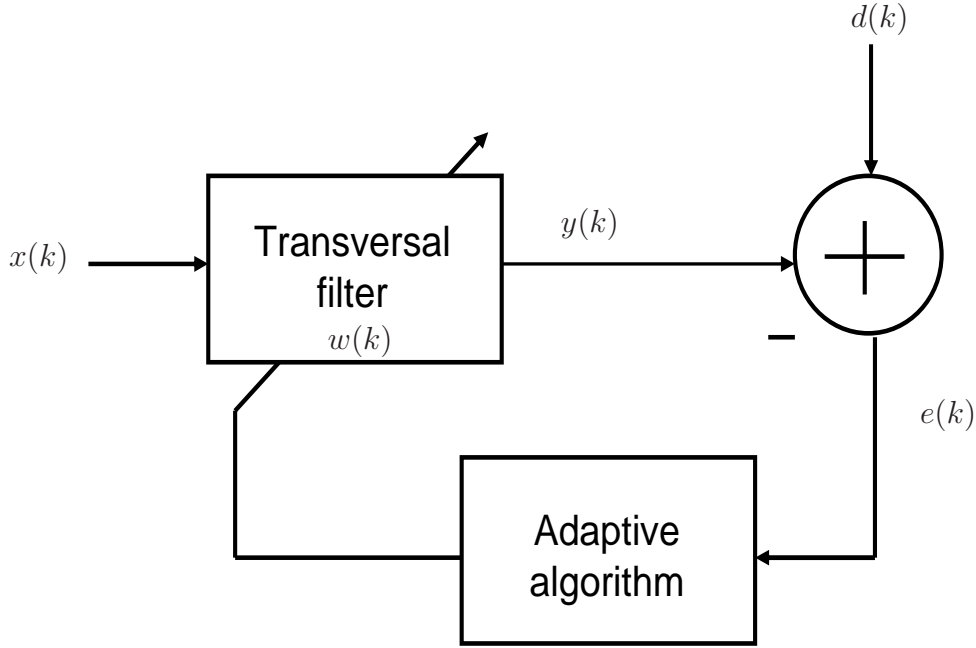
- *Stationary mixing*: Another common assumption is that the statistics of the mixing matrix  $\mathbf{A}$  do not change with time. In terms of biomedical signals this means that the physics of the mixing of the sources as measured by the sensors is not changing.
- *Statistical independence of the sources*: The most important assumption in ICA is that the sources are mutually independent. Two random variables are statistically independent if there is a joint distribution of functions of these variables. This means, for example, that independent variables are uncorrelated and have no higher order correlations. In the case of time-series data, it is assumed that each source is generated by a random process which is independent of the random processes generating the other sources.

### 3.3.3 Adaptive filtering

Digital filters design with fixed coefficients requires well defined prescribed specifications or the physical knowledge of the problem. However, there are situations where the specifications are not available, or are time varying and the optimum design of the filter is not possible. The solution in these cases is to employ a digital filter with adaptive coefficients, known as adaptive filters. An introduction of filtering concepts and statistics can be found in Appendix A.

Adaptive filters are employed in many areas of telecommunications for such purposes as adaptive equalization, echo cancellation, speech and image encoding, and noise and interference reduction. Adaptive filters can be implemented by the form of the impulse response such as finite duration response (FIR) and infinite duration impulse response (IIR) and there are various algorithms to adjust the coefficients. The characterization of adaptive filters is given by a set of adjustable coefficients and a recursive algorithm which updates these coefficients as further information (i.e. the statistics) acquired from the relevant signals, whose choice is normally a tricky task that depends on the application.

Adaptive filters are considered nonlinear systems and seem to be appropriate to non-stationary scenarios such as EEG signals; moreover they are a good choice on scenarios where the statistical information of the signal is not well known.



**Figure 3.2:** Block diagram of adaptive filter.

The general scheme is illustrated in Fig.3.2 where  $k$  is the iteration number,  $x(k)$  denotes the input signal,  $y(k)$  is the adaptive-filter output signal, and  $d(k)$  defines the desired signal. The error signal  $e(k)$  is calculated as  $d(k) - y(k)$ . This error is then used as an objective function that is required by the adaptation algorithm in order to determine the appropriate updating of the filter coefficients,  $\mathbf{w}(k)$ . The minimization of the objective function implies that the adaptive-filter output signal is matching the desired signal in some sense. The filter operation can be resumed in three steps:

- From a sample  $\mathbf{x}(k) = [x(k - N), \dots, x(k)]$  is done a prediction of desired signal  $y(k) = \mathbf{w}(k)^T \mathbf{x}(k)$ .
- The error of the filter is calculated as  $e(k) = d(k) - y(k)$ .
- From  $e(k)$  is obtained a coefficient actualization  $\mathbf{w}(k + 1)$  which it will use to process the next sample  $\mathbf{x}(k + 1)$ .

The coefficient actualization to  $\mathbf{w}(k)$  is based in minimizing the error  $e(k)$  by an iterative process. A way to quantify this error is by Mean Square Error (MSE).

Let  $J(\mathbf{w})$  denote the MSE function. The goal here is to find the set of filter coefficients  $\mathbf{w}$ , which minimize the MSE produced by the filter, i.e.

$$\begin{aligned} \min J(\mathbf{w}) &= \min\{E\{e^2(k)\}\} \\ &= \min\{E\{(d(k) - y(k))^2\}\} \end{aligned} \quad (3.5)$$

where  $d(k)$  is defined as desired signal and  $y(k) = \mathbf{w}(k)^T \mathbf{x}(k)$ . Eq.3.5 states a problem that implies several considerations to take into account, full optimal conditions, stationarity or statistical assumptions. In optimal conditions, i.e., noise is uncorrelated with the desired signal, the solution of Eq.3.5 is solved with the optimal Wiener filter, but from an adaptive point of view, it is necessary an iterative procedure that involves: a) computing the output of a FIR filter produced by the inner product of a set of filter coefficients, b) generating an estimation error by comparing the output of the filter to a desired response and finally adjusting the filter coefficients (tap weights) based on estimation error thereby closing the feedback loop. There are two important adaptive algorithms that will be discussed below.

### Least Mean Square algorithm (LMS)

Unlike the solution offered by the optimal Wiener filter, the Least Mean Square algorithm (LMS), introduced by Widrow and Hoff in 1960, does not require statistical measures such as correlation functions or matrix inversions (see Appendix A.3). Following Haykin [1996], the LMS algorithm also includes a *step size parameter*  $\mu$  that must be selected properly to control stability and convergence speed, mostly because the algorithm suffers random variations around the optimum value of the error surface  $J(\mathbf{w})$ . The parameter  $\mu$ , also known as *weighting constant*, is an inherent value of the iterative method that represents how fast the algorithm moves toward the minimum. A simple analysis of the LMS algorithm with its structure and operation is shown below.

Based on steepest descent algorithm, the update value of the tap weight vector  $\mathbf{w}(k+1)$  may be computed using the simple recursive relation

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu [\mathbf{p} - \mathbf{R}\mathbf{w}(k)], \quad k = 0, 1, 2, \dots \quad (3.6)$$

where  $\mathbf{p}$  (correlation vector) and  $\mathbf{R}$  (correlation matrix) are calculated using their *instantaneous estimates* that are based on samples values of the tap input vector  $\mathbf{x}(k)$  and desired response  $d(k)$ . Then, estimations are calculated as follows:

$$\hat{\mathbf{R}}(k) = \mathbf{x}(k)\mathbf{x}^T(k) \quad (3.7)$$

and

$$\hat{\mathbf{p}}(k) = \mathbf{x}(k)d(k) \quad (3.8)$$

Correspondingly, the new recursive relation for updating the tap weight vector is as follows:

$$\begin{aligned} \hat{\mathbf{w}}(k+1) &= \hat{\mathbf{w}}(k) + \mu \mathbf{x}(k) [d(k) - \mathbf{x}^T(k)\hat{\mathbf{w}}(k)] \\ &= \hat{\mathbf{w}}(k) + \mu \mathbf{x}(k)e(k) \end{aligned} \quad (3.9)$$

where

$$y(k) = \mathbf{w}(k)^T \mathbf{x}(k) \quad \text{filter output} \quad (3.10)$$

$$e(k) = d(k) - y(k) \quad \text{error} \quad (3.11)$$

$$\mathbf{w}(k) = [w_0(k) \ w_1(k) \ \dots \ w_{M-1}(k)]^T \quad \text{filter taps at time } k \quad (3.12)$$

$$\mathbf{x}(k) = [x(k) \ x(k-1) \ x(k-2) \ \dots \ x(k-M+1)]^T \quad \text{input data} \quad (3.13)$$

The LMS algorithm can be expressed in three basic relations showed in equations Eq.3.9 to Eq.3.11, being a member of the family of stochastic gradient algorithms with low computational complexity and high robustness, as described in Poularikas and Ramadan [2006].

Some modifications of the LMS algorithm conceived for improving convergence behavior, reducing computational cost and decreasing the steady state mean square error are described in Haykin [1996].

### Recursive Least Squares Algorithm (RLS)

Another way to solve the problem of linear filtering is based on the Least Squares (LS) method, which can be seen as an alternative to Wiener filters.

Basically, the recursive implementation of the LS method starts with known initial conditions and new data samples to update the old estimates. This process has the objective of finding the best solution to the minimization problem by minimizing the sum of squares of the difference between the desired response  $d(k)$  and the filter output  $y(k)$  defined as  $e(k) = d(k) - y(k)$ . It starts so as to minimize a cost function

$$\mathcal{E}(n) = \sum_{k=1}^n \beta(n, k) |e(k)|^2 \quad (3.14)$$

where  $n$  is the variable length of the observable data and  $\beta(n, k)$  is a *weighting factor* or *forgetting factor*. Note in Eq.3.14 that it is necessary to know all the past samples of the input signal “a priori,” especially if a *exponential weighting factor* defined as  $\beta(n, k) = \lambda^{n-k}$  for  $k = 1, 2, \dots, n$  is used<sup>3</sup>.

Thus, the *method of exponentially weighted least squares* uses the cost function

$$\mathcal{E}(n) = \sum_{k=1}^n \lambda^{n-k} |e(k)|^2 \quad (3.15)$$

---

<sup>3</sup> The errors are weighted by a factor  $\lambda^{n-k}$  with  $0 < \lambda \leq 1$ , thus, the error has less influence in future samples with more distance. The RLS algorithm is based on the LS estimation of the filter coefficient  $w(n-1)$  by computing its estimate at iteration  $n$  using new data samples to update the old estimates.

The minimization of Eq.3.15 gives

$$\nabla_w \mathcal{E}(n) = -2 \sum_{k=1}^n \lambda^{n-k} e(k) \mathbf{x}(k) \quad (3.16)$$

Replacing  $e(k)$  in the above equation, equating to zero and writing it in vector form leads to

$$\mathbf{R}(n) \mathbf{w}(n) = \mathbf{p}(n) \quad (3.17)$$

where

$$\mathbf{R}(n) = \sum_{k=1}^n \lambda^{n-k} \mathbf{x}(k) \mathbf{x}^T(k) \quad (3.18)$$

and

$$\mathbf{p}(n) = \sum_{k=1}^n \lambda^{n-k} d(k) \mathbf{x}(k) \quad (3.19)$$

Finally, Eq.3.17 may be rewritten as follows

$$\mathbf{w}(n) = \mathbf{R}^{-1}(n) \mathbf{p}(n) \quad (3.20)$$

A recursive equation can be developed for updating the LS estimate  $\hat{\mathbf{w}}(n)$  for the tap weight vector at the current time  $n$ :

$$\mathbf{p}(n) = \lambda \mathbf{p}(n-1) + d(n) \mathbf{x}(n) \quad (3.21)$$

$$\mathbf{R}(n) = \lambda \mathbf{R}(n-1) + \mathbf{x}(n) \mathbf{x}^T(n) \quad (3.22)$$

Furthermore, if the inverse of Eq.3.22 is simplified using the matrix inversion lemma, the next expression is obtained:

$$\mathbf{R}^{-1}(n) = \frac{1}{\lambda} \left[ \mathbf{R}^{-1}(n-1) - \frac{\mathbf{R}^{-1}(n-1) \mathbf{r}(n) \mathbf{r}^T(n) \mathbf{R}^{-1}(n-1)}{\lambda + \mathbf{r}^T(n) \mathbf{R}^{-1}(n-1) \mathbf{r}(n)} \right] \quad (3.23)$$

At the end, the recursive equation for updating the tap weight vector can be written as

$$\hat{\mathbf{w}} = \mathbf{w}(n-1) + \mathbf{R}^{-1}(n) \mathbf{x}(n) \alpha(n) \quad (3.24)$$

where  $\alpha(n)$  is the *innovation* defined by

$$\alpha(n) = d(n) - \hat{\mathbf{w}}^T(n-1) \mathbf{x}(n) \quad (3.25)$$

and the error  $e(n)$  after each iteration is recalculated as

$$e(n) = d(n) - \hat{\mathbf{w}}^T(n) \mathbf{x}(n) \quad (3.26)$$

The computation of  $\mathbf{R}^{-1}(n)$  in Eq.3.23 increases computation complexity although RLS shows improved behavior over LMS when inputs are correlated.

## 3.4 Feature extraction

Feature extraction consist in finding a set of measurements or a block of information with the objective of describing in a clear way the data or an event presents in a signal. These measurements or *features* are the fundamental basis for detection, classification or regression tasks in biomedical signal processing and is one of the key steps in the data analysis process.

Features constitute a new form of expressing the data, and can be binary, categorical or continuous, and also represent attributes or direct measurements of the signal. For example, features may be age, health status of the patient, family history, electrode position or EEG signal descriptors (amplitude, voltage, phase, frequency, etc.).

More formally, feature extraction assumes we have for  $N$  samples and  $D$  features, a matrix  $N \times D$ , where  $D$  represents the dimension of the feature matrix. That means, at the sample  $n$  from the feature matrix, we could obtain an unidimensional vector  $x = [x_1, x_2, \dots, x_D]$  called as “pattern vector.” Several methods in EEG feature extractions can be found in the literature, see [Guyon et al. \[2006\]](#).

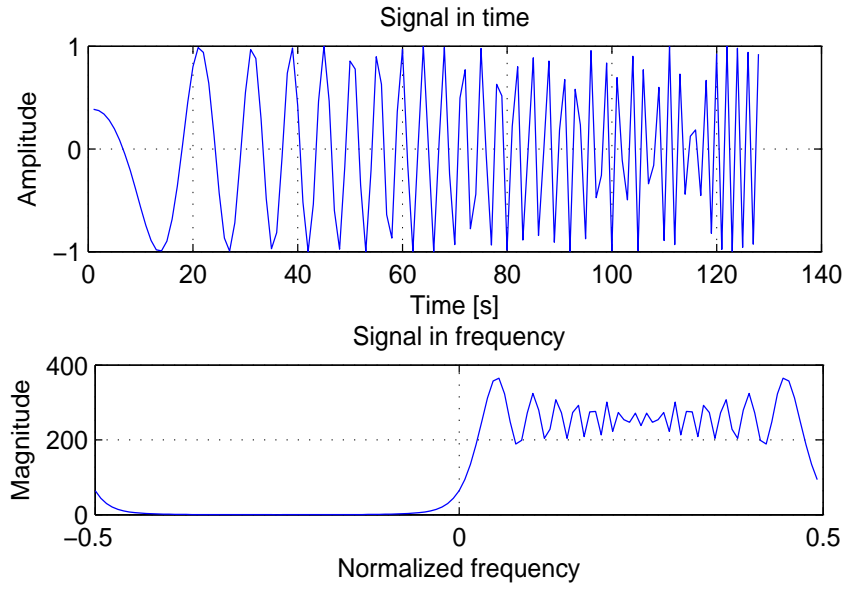
More specifically in EEG detection and classification sceneries, features based on power spectral density are introduced in [Lehmann et al. \[2007\]](#); Lyapunov exponents are introduced in [Güler and Übeyli \[2007\]](#); wavelet transform are described in [Subasi \[2007\]](#), [Hasan \[2008\]](#), [Lima et al. \[2009\]](#) and [Xu et al. \[2009\]](#); sampling techniques are used in [Siuly and Wen \[2009\]](#) and time frequency analysis are presented on [Tzallas et al. \[2009b\]](#), [Boashash \[2003\]](#), [Guerrero-Mosquera et al. \[2010a\]](#) and [Boashash and Mesbah \[2001\]](#). Other approach in feature extraction based in the fractional Fourier transform is described in [Guerrero-Mosquera et al. \[2010b\]](#). It is important to add that features extracted are directly dependent on the application and also to consider that there are important properties of these features to have into account, such as noise, dimensionality, time information, nonstationarity, set size and so on ([Lotte et al. \[2007\]](#)).

This section emphasizes methods oriented to frequency analysis, without excluding the time domain that permits to justify the importance of the frequency analysis and their shortcomings in front of non-stationary signals like the EEG.

### 3.4.1 Classical signal analysis tools

A signal could be represented in different forms being for example in time and frequency. While time domain indicates how a signal changes over time, frequency domain indicates how often such changes take place. For example,

let us consider a signal with a linear frequency modulation varying from 0 to 0.5 Hz and with constant amplitude (see Fig.3.3). Looking at the time domain representation (Fig.3.3 upper) it is not easy to say what kind of modulation is contained in the signal; and from the frequency domain representation (see Fig.3.3 bottom), nothing can be said about the evolution in time of the frequency domain characteristics of the signal.



**Figure 3.3:** Chirp signal using time domain  $x(t)$  (upper) and frequency domain  $X(\omega)$  (bottom).

The two representations are related by the *Fourier transform* (FT) as:

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt \quad (3.27)$$

or by the *inverse Fourier transform* (IFT) as:

$$x(t) = \int_{-\infty}^{\infty} X(\omega)e^{-j\omega t} d\omega \quad (3.28)$$

Eq.3.28 indicates that signal  $x(t)$  can be expressed as the sum of complex exponentials of different frequencies, whose amplitudes are the complex quantities  $X(\omega)$  defined by Eq.3.27.

The squared magnitude of the Fourier transform,  $|X(\omega)|^2$ , is often taken as the frequency representation of the signal  $x(t)$ , which allows in some sense easier interpretation of the signal nature than its time representation.

Better interpretation is obtained using a domain that directly represents frequency content while still keeping the time description parameter. This

characteristic is the aim of time frequency analysis. To illustrate this, let us represent the chirp signal explained above using the spectrogram (more details about this in the following). Note how it is possible to see the linear progression with time of the frequency components, from 0 to 0.5 (Fig.3.4).

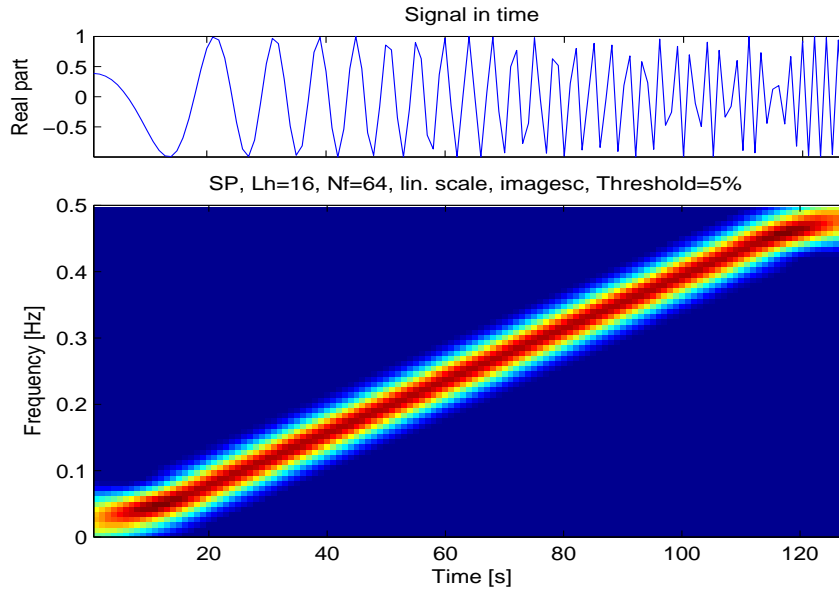


Figure 3.4: Spectrogram representation of the chirp

### 3.4.2 Time-frequency distributions (TFD)

In a series of papers (Cohen [1995], Akay [1996]), Cohen generalized the definition of time-frequency distributions (TFDs) in such a way that a wide variety of distributions could be included in the same framework. Specifically the TFD of a real signal  $x(n)$  is computed as:

$$P(t, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(\theta, \tau) \Phi(\theta, \tau) e^{-j\theta t - j\omega \tau} d\theta d\tau \quad (3.29)$$

where,

$$A(\theta, \tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} x(u + \frac{\tau}{2}) x^*(u - \frac{\tau}{2}) e^{j\theta u} du \quad (3.30)$$

is the so-called ambiguity function and the weighting function  $\Phi(\theta, \tau)$  is a function called the kernel of the distribution that, in general, may depend on time and frequency.

If  $\Phi(\theta, \tau) = 1$  in Eq.(3.29), we have

$$P(t, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(u + \frac{\tau}{2})x^*(u - \frac{\tau}{2})e^{-j\omega\tau} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\theta(t-u)} d\theta dud\tau \quad (3.31)$$

where

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\theta(t-u)} d\theta = \delta(t - u) \quad (3.32)$$

and we know that

$$\int_{-\infty}^{\infty} x(u + \frac{\tau}{2})x^*(u - \frac{\tau}{2})\delta(t - u)du = x(t + \frac{\tau}{2})x^*(t - \frac{\tau}{2}) \quad (3.33)$$

If we substitute the Eq.(3.32) and Eq.(3.33) in Eq.(3.31), then we have the Wigner-Ville distribution (WV) defined as:

$$WV(\omega, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} x(t + \frac{\tau}{2})x^*(t - \frac{\tau}{2})e^{-j\omega\tau} d\tau \quad (3.34)$$

Following [Hammond and White \[1996\]](#), the recurrent problem of the WV is the so-called crossterm interference, due to bilinear nature of its definition. These crossed terms tend to be located mid-way between the two auto terms and are oscillatory in nature.

When  $\Phi(\theta, \tau) = 1$ , we have the Wigner-Ville distribution  $WV(t, \omega)$ . The Smooth Pseudo Wigner-Ville (SPWV) distribution is obtained by convolving the  $WV(t, \omega)$  with a two-dimensional filter in  $t$  and  $\omega$ . This transform incorporates smoothing by independent windows in time and frequency, namely  $W_w(\tau)$  and  $W_t(t)$ :

$$SPWV(t, \omega) = \int_{-\infty}^{\infty} W_w(\tau) \left[ \int_{-\infty}^{\infty} W_t(u - t)x(u + \frac{\tau}{2})x^*(u - \frac{\tau}{2})du \right] e^{-j\omega\tau} d\tau \quad (3.35)$$

Eq.(3.35) provides great flexibility in the choice of time and frequency smoothing, but the length of the windows should be determined empirically according to the type of signal analyzed and the required cross term suppression, as discussed in [Afonso and Tompkins \[1995\]](#).

As proved in [Hlawatsch and Boudreaux-Bartels \[1992\]](#), the SPWV does not satisfy the marginal properties (see Eq.3.35), that is, the frequency and time integrals of the distribution do not correspond to the instantaneous signal power and the spectral energy density, respectively. However, it is still

possible for a distribution to give the correct value for the total energy without satisfying the marginals, as described in Cohen [1995, 1989]. Therefore the total energy can be a good feature to detect signal events in the SPWV representation because the energy in EEG seizure is usually larger than the one during normal activity.

The TFDs offer the possibility of analyzing relatively long continuous segments of EEG data even when the dynamics of the signal are rapidly changing. Taking the most of these, it can extract features from the time frequency plane such as ridges energy, frequency band values and so on. However, three considerations have to be taken, presented in Cohen [1995, 1989] and Durka [1996]:

- A TFD will need signals as clean as possible for good results.
- A good resolution both in time and frequency is necessary and as the “uncertainty principle” states, it is not possible to have a good resolution in both variables simultaneously.
- It is also required to eliminate the spurious information (i.e. cross-term artifacts) inherent in the TFDs.

The first consideration implies a good pre-processing stage to eliminate artifacts and noise. Second and third considerations have motivated the TFD selection or design, then it is important and necessary to choose a suitable TFD for seizure detection in EEG signals as well as for a correct estimation of frequencies on the time-frequency plane. Indeed, it is desirable that the TFD has both low cross-terms and high resolution. Choosing a distribution depends on the information to be extracted and demands a good balance between good performance, low execution time, good resolution and few and low-amplitude cross terms.

One consideration before using the TFD is to convert each EEG segment into its analytic signal for a better time-frequency analysis. The analytic signal is defined to give an identical spectrum to positive frequencies and zero for the negative frequencies, and shows an improved resolution in the time-frequency plane, discussed in Cohen [1989]. It associates a given signal  $x(n)$  to a complex valued signal  $y(n)$  defined as:  $y(n) = x(n) + jHT\{x(n)\}$ , where  $y(n)$  is the analytic signal and  $HT\{.\}$  is the Hilbert transform.

### 3.4.3 Wavelet coefficients

The EEG signals can be considered as a superposition of different structures occurring on different time scales at different times. As presented in

Latka and Was [2003], the Wavelet Transform (WT) provides a more flexible way of time-frequency representation of a signal by allowing the use of variable size windows and can constitute the foundation of a relatively simple yet effective detection algorithm. Selection of appropriate wavelets and the number of decomposition levels is very important in the analysis of signals using the WT. The number of decomposition levels is chosen based on the dominant frequency components of the signals. Large windows are used to get a finer low-frequency information and short windows are used to get high-frequency resolution. Thus, WT gives precise frequency information at low frequencies and precise time information at high frequencies. This makes the WT suitable for EEG analysis of spikes patterns or epileptic seizures.

Wavelets overcome the drawback of a fixed time-frequency resolution of short time Fourier transforms. The WT performs a multiresolution analysis,  $W_{\Psi}f(a, b)$  of a signal,  $x(n)$  by convolution of the mother function  $\Psi(n)$  with the signal, as given in Latka and Was [2003], and Mallat [2009] as:

$$W_{\Psi}x(b, a) = \sum_{n'=0}^{N-1} x(n')\Psi^*\left(\frac{n' - b}{a}\right) \quad (3.36)$$

$\Psi(t)^*$  denote the complex conjugate of  $\Psi(n)$  (basis function),  $a$  the scale coefficient,  $b$  the shift coefficient and  $a, b \in \mathfrak{R}$ ,  $a \neq 0$ .

In the procedure of multiresolution decomposition of a signal  $x(n)$ , each stage consists of two digital filters and two downsamplers by 2. The bandwidth of the filter outputs are half the bandwidth of the original signal, which allows for the downsampling of the output signals by two without losing any information according to the Nyquist theorem. The downsampled signals provide detail D1 and approximation A1 of the signal, this procedure is described in Hasan [2008].

Once the mother wavelet is fixed, it is possible to analyze the signal at every possible scale  $a$  and translation  $b$ . If the basis function  $\Psi(n)$  is orthogonal, then the original signal can be reconstructed from the resulting wavelet coefficients accurately and efficiently without any loss of information. The Daubechies' family of wavelets is one of the most commonly used orthogonal wavelets to non-stationary EEG signals presenting good properties and allowing reconstruction of the original signal from the wavelet coefficients, as described Mallat [2009].

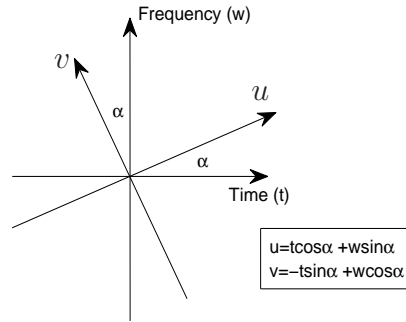
### 3.4.4 Fractional Fourier transform

Fourier analysis is undoubtedly one of the most used tools in signal processing and other scientific disciplines and this technique uses harmonics for the decomposition of signals with time-varying periodicity. Similarly, TFDs are

very frequently used in signal analysis especially when it is necessary to eliminate the windowing dependence on non-stationary signals.

In 1930, Namias employed the fractional Fourier transform (FrFT) to solve partial differential equations in quantum mechanics from classical quadratic Hamiltonians<sup>4</sup>. The results were later improved by McBride and Kerr in [Tao et al. \[2008\]](#). They developed operational calculus to define the FRFT. The FrFT is a new change in the representation of the signal which is an extension of the classical Fourier transform. When fractional order gradually increases, the FrFT of a signal can offer much more information represented in an united representation than the classical Fourier transform and it provides a higher concentration than TFDs, avoiding the cross terms components produced by quadratics TFDs.

FrFT has established itself as a potential tool for analyzing dynamic or time-varying signals with changes in very short time and it can be interpreted as the representation of a signal in neutral domain by means of the rotation of the signal by the origin in counter-clockwise direction with rotational angle  $\alpha$  in time-frequency domain as shown in [Fig.3.5](#).



**Figure 3.5:** The relation of fractional domain  $(u, v)$  with traditional time-frequency plane  $(t, w)$  rotated by an angle  $\alpha$ .

The FrFT with angle  $\alpha$  of a signal  $x(t)$ , denoted as  $X_\alpha(u)$  is defined in [Almeida \[1994\]](#) as:

$$X_\alpha(u) = \int_{-\infty}^{\infty} x(t) K_\alpha(t, u) dt \quad (3.37)$$

<sup>4</sup>A development based on a concept called *fractional operations*. For example, the  $n$ -th derivative of  $f(x)$  can be expressed as  $d^n f(x)/dx^n$  for any positive integer  $n$ . If another value derived is required, i.e. the 0.5-th derivative, it is necessary to define the operator  $d^a f(x)/dx^a$ , where the value  $a$  could be any real value. The function  $[f(x)]^{0.5}$  is the square root of the function  $f(x)$ . But  $d^{0.5} f(x)/dx^{0.5}$  is the 0.5-th derivative of  $f(x)$  ( $a = 0.5$ ),  $(df(x)/dx)^{0.5}$  being the square root of the derivative operator  $d/dx$ . As it can be seen, fractional operations is a concept that goes from the whole of an entity to its fractions.

where  $K_\alpha(u, t)$  is a linear kernel function continuous in the angle  $\alpha$ , which satisfies the basic conditions for being interpretable as a rotation in the time-frequency plane. The kernel has the following properties

$$K_\alpha(t, u) = K_\alpha(u, t) \quad (3.38)$$

$$K_{-\alpha}(t, u) = K_{-\alpha}^*(t, u) \quad (3.39)$$

$$K_\alpha(-t, u) = K_\alpha(t, -u) \quad (3.40)$$

$$\int_{-\infty}^{\infty} K_\alpha(t, u) K_\beta(u, z) du = K_{\alpha+\beta}(t, z) \quad (3.41)$$

$$\int_{-\infty}^{\infty} K_\alpha(t, u) K_\alpha^*(t, u') dt = \delta(u - u') \quad (3.42)$$

The FrFT is given by

$$X_\alpha(u) = \begin{cases} \sqrt{\left(\frac{1 - j \cot \alpha}{2\pi}\right)} e^{j \frac{u^2}{2} \cot \alpha} \int_{-\infty}^{\infty} x(t) e^{j \frac{t^2}{2} \cot \alpha} e^{j u t \csc \alpha} dt, & \text{if } \alpha \text{ is not a multiple of } \pi \\ x(t), & \text{if } \alpha \text{ is multiple of } 2\pi \\ x(-t), & \text{if } \alpha + \pi \text{ is multiple of } 2\pi \end{cases}$$

More detailed definitions, proof and further properties of the kernel can be found in [Almeida \[1994\]](#).

In summary, the FrFT is a linear transform, continuous in the angle  $\alpha$ , which satisfies the basic conditions for being interpretable as a rotation in the time-frequency plane.

### 3.5 Feature selection

After feature extraction, it is necessary to select the subset of features that present better performance or are most useful for a problem at hand, such as regression, classification or detection. The data acquisition in environments such as biomedical signals leads to define each problem by hundreds or thousands of measurements leading to obtain high dimensional data with high computational cost.

As discussed in [Guyon and Elisseeff \[2003\]](#), feature selection is based on the principle that choosing a smaller number of variables among the original ones, leads to an easier interpretation. In fact, under the assumption that reducing the *training data*<sup>5</sup> might improve the performance task, the feature

<sup>5</sup>Concept related to the fact of using a data set (also called data points, samples, patterns or observations) in order to gain knowledge, learn a task associated with desired outcomes.

selection methods also allows a better data understanding and visualization together with reduction in data measurement and storage.

Feature selection could be summarized in two main tasks: choosing the relevant features and searching the best feature subset. The first one tries to solve the question: is a feature (or subset of features) relevant for the problem? And the second one tries to search the best feature subset among all the possible subsets extracted from the initial task<sup>6</sup>. The application of these two tasks to high dimensional data causes a reduction in the data dimension, process known as dimensionality reduction.

Besides feature selection, there is another set of methods known as *projection methods* that perform the same task but in practice could retain the problems suffered by high dimensional data, presented in Rossi et al. [2007, 2006]. Typical projection algorithms are Principal Component Analysis (PCA), Sammon's Mapping, Kohonen maps, Linear Discriminant Analysis (LDA), Partial Least Squares (PLS) or Projection pursuit, amongst others ( see Duda et al. [2009]).

### 3.5.1 Subset relevant assessment

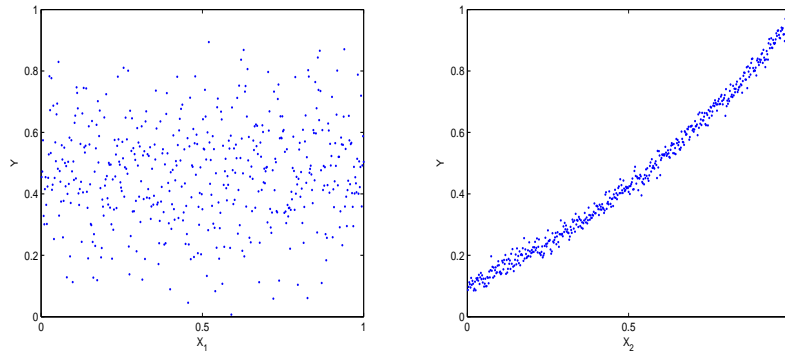
This step is mainly based on a *relevance criterion* that has to be chosen by some measurement. The best choice for the criterion is certainly to estimate the performances of the model itself, i.e., an individual feature ranking could be appropriate at scenarios where the features provide a good performance by itself and there is the possibility of choosing features associated to high ranks.

The idea of the “individual relevance ranking” can be clarified by the following example: Fig.3.6 shows a situation where the feature  $X_2$  is more relevant individually to predict the output  $Y$  than the feature  $Y_1$ . Notice the importance of choosing the right features to improve the performance of a task, which in this example is related to prediction of  $Y$ .

There are different alternatives in relevance criteria, such as the Pearson correlation coefficient, mutual information (MI) and wrapper methodology. Although each method has its advantages and disadvantages, mutual information has proven to be an appropriate measure in several applications such as selection of spectral variables, spectrometric nonlinear modelling and functional data classification, see Gomez-Verdejo et al. [2009], Rossi et al.

---

<sup>6</sup>Although feature extraction and feature selection are different aspects of the pattern recognition process, it is important to distinguish the difference between them. The first one aims at building a good feature representation based on several measurements, and the second one tries to reduce the feature matrix by selecting subsets of features more useful in determined tasks.



**Figure 3.6:** Simple prediction problem. The horizontal axis represents the feature and the vertical axis the output. It can be seen that feature  $X_2$  (left) is more relevant individually than feature  $X_1$  (right) in this simple prediction problem.

[2007, 2006]. Moreover, as discussed in Cover and Thomas [1991], correlation does not measure nonlinear relations among features and wrapper approach presents a high computational load. Furthermore, MI could be seen as a correlation measure applied to determine the nonlinearity among features.

Next section focuses on the well-known concept of MI and shows why this relevance criterion is applicable for feature selection.

### 3.5.2 Mutual information (MI)

Mutual information (MI) measures the relevance between a group of features  $X$  and the variable or output  $Y$ . This relationship is not necessarily linear. As described in Cover and Thomas [1991], the mutual information between two variables is the amount of uncertainty (or entropy) that is lost on one variable when the other is known, and vice-versa. The variables  $X$  and  $Y$  could be multidimensional, solving the drawback in correlation measurements that are based on individual variables.

Let  $p_X(x)$  and  $p_Y(y)$  be the marginal of probability density function (pdf) of  $X$  and  $Y$  respectively, and the joint probability density function of  $X$  and  $Y$  is  $p_{X,Y}(x, y)$ . If  $X$  has  $\mathcal{X}$  alphabets, the entropy of  $X$  is defined as

$$H(X) = - \sum_{x \in \mathcal{X}} p_X(x) \log p_X(x) \quad (3.43)$$

The base of the logarithm determines the units in which information is measured. Particularly, if the logarithm is base 2 the entropy is expressed in bits.

The joint entropy  $H(X, Y)$  of a pair of discrete random variables  $(X, Y)$  with a joint distribution  $p_{X,Y}(x, y)$  is defined as

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{X,Y}(x, y) \log p_{Y|X}(y|x) \quad (3.44)$$

and the MI between two variables is calculated as

$$I(X, Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{X,Y}(x, y) \log \frac{p_{X,Y}(x, y)}{p_X(x)p_Y(y)} \quad (3.45)$$

Eq.3.45 gives the relation between  $X$  and  $Y$ , meaning that  $I(X, Y)$  is large (small) the variables are closely (not closely) related. The MI and entropy have the following relation, see [Cover and Thomas \[1991\]](#):

$$I(X, Y) = H(Y) - H(Y|X) \quad (3.46)$$

For continuous variables, the entropy and MI are defined as

$$H(X) = - \int_{-\infty}^{\infty} p_X(x) \log p_X(x) dx \quad (3.47)$$

$$I(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{X,Y}(x, y) \log \frac{p_{X,Y}(x, y)}{p_X(x)p_Y(y)} dx dy \quad (3.48)$$

Note in Eq.3.45 and Eq.3.48 that it is necessary to know the exact pdf's for estimating the MI and this is the most sensitive part in the MI estimation. Several methods have been proposed in the literature to estimate such joint densities, see [Lotte et al. \[2007\]](#), [Duda et al. \[2009\]](#). Next, details of three estimators that will be used in the following of this section will be introduced briefly: MI based on Parzen windows, the Kraskov method and other MI estimators derived from Kraskov method and oriented to classification problems.

### MI estimation based on Parzen windows

The Parzen window density estimate can be used to approximate the probability density functions in Eq.3.45 and Eq.3.48 providing a smoother and more reliable estimate. Hence, in MI estimation of a multi-class classification problem, where  $\mathbf{x}$  is a multi-dimensional observation with  $N$  values ( $1 \leq n \leq N$ ) that has to be classified among  $C$  classes with class labels  $y_c (1 \leq c \leq C)$  is calculated as

$$\hat{I}(X, Y) = \hat{H}(Y) - \hat{H}(Y|X) \quad (3.49)$$

$$\begin{aligned} &= - \sum_{c=1}^C p_Y(y = y_c) \log p_Y(y = y_c) \\ &\quad + \sum_{n=1}^N \frac{1}{N} \sum_{c=1}^C p_{Y|X}(y = y_c|\mathbf{x}) \log(p_{Y|X}(y = y_c|\mathbf{x})) \end{aligned} \quad (3.50)$$

As  $C$  is a discrete random variable, its entropy  $H(Y)$  can be easily calculated as

$$p_Y(y = y_c) = \frac{n_c}{N} \quad c = 1, \dots, C, \quad (3.51)$$

where  $n_c$  is the number of observations in class  $c$  and  $C$  the number of classes. The entropy  $H(Y|X)$  is more difficult to calculate because it is necessary to estimate  $p_{Y|X}(y = y_c|\mathbf{x})$ . Following Kwak and Choi [2002], this conditional probability from a Parzen estimator with Gaussian window becomes

$$\hat{p}_{Y|X}(y = y_c|\mathbf{x}) = \frac{\sum_{i \in y_c} \exp\left(-\frac{(\mathbf{x}-\mathbf{x}_i)^T \Sigma^{-1} (\mathbf{x}-\mathbf{x}_i)}{2h^2}\right)}{\sum_{c=1}^C \sum_{i \in y_c} \exp\left(-\frac{(\mathbf{x}-\mathbf{x}_i)^T \Sigma^{-1} (\mathbf{x}-\mathbf{x}_i)}{2h^2}\right)} \quad (3.52)$$

where  $h$  is the window width and  $\Sigma$  the data covariance matrix.

### Kraskov's estimator of mutual information

The Kraskov's method is based on entropy estimates from  $k$ -NN statistics<sup>7</sup>. Therefore the MI estimation is done directly without first estimating the pdf's. A good property of this estimator is that it can be used easily for sets of features, something necessary in the feature selection procedure. MI calculation is obtained by estimating  $H(X)$ ,  $H(Y)$  and  $H(X, Y)$  separately and using the next equation

$$I(X, Y) = H(X) + H(Y) - H(X, Y) \quad (3.53)$$

Let  $Z = (X, Y)$  be a random variable with pdf  $p_{X,Y}(X, Y)$  and a set of  $N$  input-output measurements  $z_i = (x_i, y_i)$ ,  $1 \leq i \leq N$ , which are assumed independent and identically distributed (i.i.d) realizations of  $Z$ . Following Kraskov et al. [2004], it is then possible to rank for each point  $z_i$  its neighbour by distance  $d_{i,j} = \|z_i - z_j\| : d_{i,j_1} \leq d_{i,j_2} \leq \dots$ ,  $X$  and  $Y$  having values in the spaces  $\mathcal{R}$  or in  $\mathcal{R}^P$ . The algorithm will therefore use the natural norm in these spaces, i.e., the Euclidean norm<sup>8</sup>.

Measurement pairs are compared through the maximum norm: if  $z = (x, y)$  and  $z' = (x', y')$ , then

$$\|z - z'\|_\infty = \max(\|x - x'\|, \|y - y'\|) \quad (3.54)$$

The basic idea of Kraskov method is to estimate  $I(X, Y)$  from the average distances in the  $X$ ,  $Y$  and  $Z$  spaces from  $z_i$  to its  $k$ -NN, averaged over all  $z_i$ . If  $z^{k(i)} = (x^{k(i)}, y^{k(i)})$  is the  $k$ -th nearest neighbour of  $z^i$  (according to

<sup>7</sup>Explanations about this algorithm are in Appendix B.

<sup>8</sup>The notation for all metrics will be the same. For example, the norm of  $v$  is denoted  $\|v\|$

the maximum norm) and  $x^{k(i)}$  and  $y^{k(i)}$  are the input and output parts of  $z^{k(i)}$  respectively, a distance called  $\epsilon(i)/2$  could be measured from  $z_i$  to its  $k$ -th neighbour. For example,  $\epsilon_x(i)/2$  and  $\epsilon_y(i)/2$  are the distances between the same points projected into the  $X$  and  $Y$  subspaces. Obviously,  $\epsilon(i) = \max\{\epsilon_x(i), \epsilon_y(i)\}$ . Thus, it is counted the number  $\tau_x(i)$  of points  $x_j$  whose distance from  $x_i$  is strictly less than  $\epsilon(i)/2$  and similarly for  $y$  instead of  $x$ .

After calculating the entropy estimator and some mathematical manipulations, two different MI estimators are describe in [Kraskov et al. \[2004\]](#) as

$$\hat{I}^{(1)}(X, Y) = \psi(N) + \psi(K) - \frac{1}{N} \sum_{n=1}^N [\psi(\tau_x(n) + 1) + \psi(\tau_y(n) + 1)], \quad (3.55)$$

$$\hat{I}^{(2)}(X, Y) = \psi(N) + \psi(K) - \frac{1}{K} - \frac{1}{N} \sum_{n=1}^N [\psi(\tau_x(n)) + \psi(\tau_y(n))] \quad (3.56)$$

where  $K$  is the number of nearest neighbours (a parameter of the algorithm),  $N$  is the number of samples in the data set and  $\psi(\cdot)$  is the digamma function given by

$$\psi = \frac{\Gamma'(t)}{\Gamma(t)} = \frac{d}{dt} \ln \Gamma(t), \quad (3.57)$$

with

$$\Gamma(t) = \int_0^{\infty} u^{t-1} \exp^{-u} du \quad (3.58)$$

The quality of the estimators presented before is related with the values of  $k$ . With a small value of  $k$ , the estimator has a large variance and a small bias, whereas a large value of  $k$  leads to a small variance and a large bias, as discussed in [Kraskov et al. \[2004\]](#).

### MI estimation for classification problems

In the case of classification tasks, [Gomez-Verdejo et al. \[2009\]](#) presented one approach to use the information given by the training set and to use the conditional entropy to estimate MI as

$$\hat{I}(X, C) = \hat{H}(X) - \sum_{c=1}^C p_Y(y = y_c) \hat{H}(X|Y = y_c) \quad (3.59)$$

Using the Kozachenko-Leonenko estimator, an expression to estimate the entropy can be obtained (more details in [Kraskov et al. \[2004\]](#)):

$$\hat{H} = -\psi(K) + \psi(N) + \log C_d + \frac{d}{N} \sum_{n=1}^N \log \epsilon(n, K) \quad (3.60)$$

where  $N$  is the number of samples in the data set,  $K$  is the number of nearest neighbours,  $C_d$  is the volume of a unitary ball,  $d$  is the dimensionality of  $X$  and  $\epsilon(n, K)$  is twice the distance from  $x_i$  to its  $k$ -th neighbor.

Substituting Eq.3.60 in Eq.3.59 it is obtained

$$\begin{aligned}
\hat{I}(X, Y) &= -\psi(K) + \psi(N) + \log C_d + \frac{d}{N} \sum_{n=1}^N \log \epsilon(n, K) \\
&\quad - \sum_{c=1}^C p_Y(y = y_c) \left[ -\psi(K) + \psi(n_c) + \log C_d + \frac{d}{n_c} \sum_{n \in y_c} \log \epsilon_c(n, K) \right] \\
&= -\psi(K) + \psi(N) + \log C_d + \frac{d}{N} \sum_{n=1}^N \log \epsilon(n, K) - \psi(K) - \log C_d \\
&\quad - \sum_{c=1}^C p_Y(y = y_c) \left[ \psi(n_c) + \frac{d}{n_c} \sum_{n \in y_c} \log \epsilon_c(n, K) \right] \tag{3.61}
\end{aligned}$$

After a few manipulations described in [Gomez-Verdejo et al. \[2009\]](#), the MI estimator is given by

$$\begin{aligned}
\hat{I}(X, Y) &= \psi(N) - \frac{1}{N} \sum_{c=1}^C n_c \psi(n_c) \\
&\quad + \frac{d}{N(K_{max} - K_{min} + 1)} \left( \sum_{k=K_{min}}^{K_{max}} \left[ \sum_{n=1}^N \log \epsilon(n, K) - \sum_{c=1}^C \sum_{n \in y_c} \log \epsilon_c(n, K) \right] \right) \tag{3.62}
\end{aligned}$$

where  $K_{min}$  and  $K_{max}$  determine the range of  $K$  values which is determined by cross-validation.

### 3.5.3 Search procedure

Several search strategies presented in [Guyon and Elisseeff \[2003\]](#) could be used for finding the most adequate subset of features, such as best-first, branch-and-bound, simulated annealing and genetic algorithms. Greedy search strategies such as forward selection, backward elimination or any combination of them are the most popular. The forward selection method starts from an empty set and progressively add features one by one according to some criterion. In a backward elimination procedure one starts with all the features and progressively eliminates the least useful ones.

With  $D$  input features, there are  $2^{D-1}$  possible subsets that should be studied but this evaluation is unfeasible for large  $D$  due to its high computational cost.

### Forward-backward algorithm

The combination of forward and backward procedures could alleviate the curse of dimensionality by avoiding the evaluation of features with large dimension  $D$  (high computational cost). As described in Rossi et al. [2006] this algorithm works in the following steps:

- (1) The first feature is selected by MI maximization between the original features  $[X_1, \dots, X_D]$  and the output variable  $Y$ :

$$X_1^{sel} = \arg \max_{X_j} \left\{ \hat{I}(X_j, Y) \right\}, \quad 1 \leq j \leq D \quad (3.63)$$

where  $\hat{I}(X, Y)$  is the MI estimation of  $I(X, Y)$  and  $X_1^{sel}$  is the first variable selected.

- (2) Once  $X_1^{sel}$  is selected, the next components must be selected taking this first variable into account. The second variable  $X_2^{sel}$  is the one that maximizes the MI in conjunction with the first one and the output variable  $Y$ :

$$X_2^{sel} = \arg \max_{X_j} \left\{ \hat{I}(\{X_1^{sel}, X_j\}, Y) \right\}, \quad 1 \leq j \leq D, \quad X_j \neq X_1^{sel} \quad (3.64)$$

The next steps consisting in selecting the variable  $X_{st}^{sel}$  in the  $t$ -th step given a subset of already selected features

$$S = [X_1^{sel}, X_2^{sel}, \dots, X_{s(t-1)}^{sel}]$$

then  $X_{st}^{sel}$  is chosen according to

$$X_{st}^{sel} = \arg \max_{X_j} \left\{ \hat{I}(\{S, X_j\}, Y) \right\}, \quad 1 \leq j \leq D, \quad X_j \notin S \quad (3.65)$$

- (3) Assuming that  $t$  variables have been selected after the step  $t$ , which means that the last variable selected is  $X_{st}^{sel}$ , the backward procedure consists in checking one by one what happens with the MI when a variable is removed from the subset  $S$ . The variable chosen is the one ( $X_t^{rem}$ ) that increases the estimation of the MI when eliminated. In other words, we apply the next maximization rule is applied after the forward step  $t$ :

$$X_t^{rem} = \arg \max_{X_j^{sel}} \left\{ \hat{I}(\{X_1^{sel}, \dots, X_{j-1}^{sel}, X_{j+1}^{sel}, \dots, X_{t-1}^{sel}\}, Y) \right\}, \quad 1 \leq j \leq t,$$

if

$$\hat{I}(\{X_1^{sel}, \dots, X_{j-1}^{sel}, X_{j+1}^{sel}, \dots, X_t^{sel}\}, Y) > \hat{I}(\{X_1^{sel}, \dots, X_t^{sel}\}, Y) \quad (3.66)$$

Search methods must have a stopping criterion. The backward elimination is more intuitive than the forward selection. For the latter one could use a permutation test among the features chosen to evaluate if a new variable presents a significant increase of MI. Another way is to use a ranking algorithm rather than a selection one. These methods are described in Rossi et al. [2006], Gomez-Verdejo et al. [2009] respectively. However, studying stopping criteria for this problem exceed the scope of this Phd Thesis and is left as future work.

### 3.6 Classification algorithms for EEG signals

Unlike many theoretical approaches that solve certain problems using some model or formula, many classifiers are based on statistical learning. In such cases the system should be trained to obtain a good classifier taking into account that, under the following considerations described in Sanei and Chambers [2007], classification algorithms do not perform efficiently when:

- the number of features is high,
- there is limited execution time for a classification task,
- the classes or labels from feature matrix are unbalanced,
- there are nonlinearities between inputs and outputs,
- data distribution is unknown,
- there is no convergence guarantee to best solution (problem not convex or monotonic).

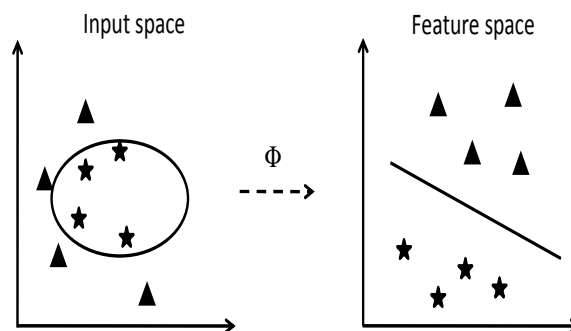
Up-today, several algorithms in EEG signal classification and detection have been proposed in the literature. For example, Multiple signal classification (MUSIC) combining EEG and MEG for EEG source localization described in Mosher and Leahy [1998]; classification of patients with Alzheimer using Support Vector Machine (SVM) and neural networks (NNs) described in Lehmann et al. [2007]; Güler and Übeyli [2007] introduced the multiclass SVM for EEG. Lotte et al. [2007] describes several applications for BCI using methods such as Hidden Markov Modelling (HMM), Linear Discriminant Analysis (LDA) and fuzzy logic; Chiappa and Barber [2006] used the Bayes's rule to discriminate mental tasks; detection of ERPs using SVM described in Thusalidas et al. [2006]; Fuzzy SVM (FSVM) is utilized in Xu et al. [2009]; Fisher's discriminant is introduced in Müller et al. [2003]. Applications in epilepsy classification such as Artificial Neural Networks (ANN) described in Subasi [2007];  $k$ -NN classifier and logistic regression with TFDs are used in

Tzallas et al. [2009b]; Least Square SVM (LS-SVM) in Siuly and Wen [2009]; Learning Vector Quantization with NN (LVQ-NN) described in Hasan [2008]; Mixture of Experts (ME) and Multilayered Perceptron (MLP) in Subasi [2007]; an automatic EEG signal classification using Relevance Vector Machine (RVM) is proposed by Lima et al. [2009].

As showed in Guyon et al. [2006], SVM and its variants have many applications in different classification scenarios and are a powerful approach for pattern recognition, showing to be a good alternative for EEG signal classification due to their high performance, good generalization compared to other methods such as NN. The next section will introduce the basic concepts of SVM classifiers.

### Support vector classification

As described by Hearst et al. [1998], the support vector machines (SVMs) are a promising tool for data classification because of two aspects: first, SVM have proved to be one of the most appropriate alternative for solving classification problems and their solution is supported in statistical learning theory, see Vapnik [2000], Schölkopf and Smola [2002]. Second, SVM presents high performance in different practical sceneries. With these aspects, SVM could be considered as the intersection point between learning theory and practice, that means that the statistical learning theory can identify rather precisely the factors necessary to solve the problem in question. Then, with an appropriated nonlinear mapping  $\Phi(\cdot)$ , the basic SVM idea is to map the training data nonlinearly into a higher dimensional feature space via  $\Phi$  maximizing the separation between classes as shown in Fig.3.7.



**Figure 3.7:** The basic idea of the SVM is to map the training data into a high dimensional space and find a separating hyperplane with the maximal margin via  $\Phi$ .

The concept of SVM was initiated in 1979 by Vapnik [2000]. Support vectors classifiers are based on the class of hyperplanes

$$(\mathbf{w} \cdot \mathbf{x}) + b = 0 \quad \mathbf{w} \in \mathfrak{R}^N, \quad b \in \mathfrak{R} \quad (3.67)$$

corresponding to decision function

$$f(\mathbf{x}) = \text{sign}((\mathbf{w} \cdot \mathbf{x}) + b) \quad (3.68)$$

where  $\mathbf{w}$  determines the orientation of a discriminant hyperplane. The assumption of separability means that there exists some set of values  $\mathbf{w}$  and  $b$ , such that the following constraints hold for all input vectors, given that the classes are labeled +1 and -1:

$$\mathbf{w} \cdot \mathbf{x}_i + b \geq +1, \quad \forall y_i = +1 \quad (3.69)$$

$$\mathbf{w} \cdot \mathbf{x}_i + b \leq -1, \quad \forall y_i = -1 \quad (3.70)$$

or

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0, \quad \forall i \quad (3.71)$$

The number of possible hyperplanes that could correctly classify the training data is infinite but there is an “optimal hyperplane” that maximizes the margin of separation between two classes (see Fig.3.8).

The construction of the hyperplane  $(\mathbf{w} \cdot \mathbf{x}_i) + b = 0$  can be solved by quadratic optimization problem, where  $\mathbf{w}$  has an expansion  $\mathbf{w} = \sum_i v_i \mathbf{x}_i$  in terms of a subset of training patterns (called support vectors) that lie on the margin and have all relevant information about the classification problem. Therefore, it is possible to maximize the margin of the classifier by minimizing  $\|\mathbf{w}\|$ , subject to the constraints given in Eq.3.71. Thus the problem of training the SVM can be stated as follows: find  $w$  and  $b$  such that the resulting hyperplane correctly classifies the training data and the Euclidean norm of the weight vector is minimized.

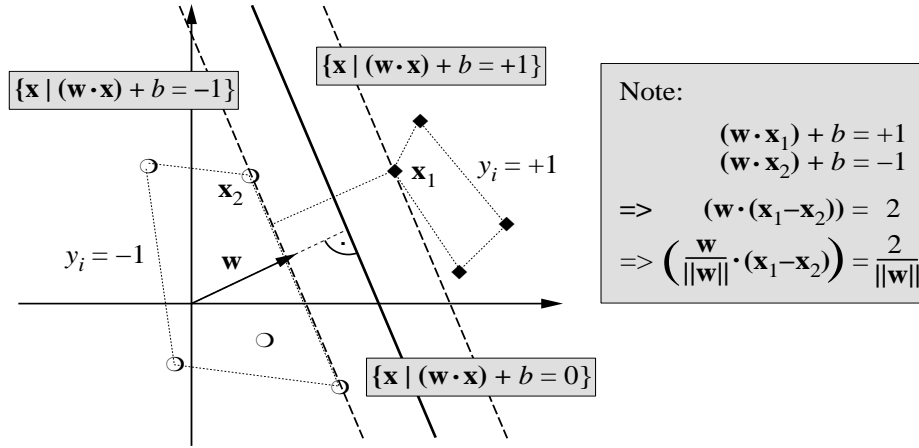
To solve the problem described above, one solves the following optimization problem:

$$\text{minimize} \quad \tau(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 \quad (3.72)$$

$$\text{subject to} \quad y_i((\mathbf{w} \cdot \mathbf{x}_i) + b) \geq 1, \quad i = 1, \dots, m. \quad (3.73)$$

This constrained optimization problem is typically reformulated as a Lagrangian optimization problem. In this reformulation, nonnegative Lagrange multipliers  $A = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$  are introduced, yielding the Lagrangian

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^m \alpha_i (y_i((\mathbf{x}_i \cdot \mathbf{w}) + b) - 1). \quad (3.74)$$



**Figure 3.8:** A binary classification toy problem consisting in separating diamonds from balls, each class with labels  $\{+1, -1\}$  respectively. The “optimal hyperplane” is orthogonal to the shortest line connecting the convex hulls of the two classes (dotted), and intersects it half-way between the two classes. If the problem is separable, there exists a weight vector  $\mathbf{w}$  and a threshold  $b$  such that  $y_i((\mathbf{w} \cdot \mathbf{x}_i) + b) > 0$  ( $i = 1, \dots, m$ ). Rescaling  $\mathbf{w}$  and  $b$  such that the point(s) closest to the hyperplane satisfy  $|(\mathbf{w} \cdot \mathbf{x}_i) + b| = 1$ , it obtains a form  $(\mathbf{w}, b)$  of the hyperplane with  $y_i((\mathbf{w} \cdot \mathbf{x}_i) + b) \geq 1$ . Note that in this case, the margin, measured perpendicularly to the hyperplane, equals  $2/\|\mathbf{w}\|$  (in [Schölkopf and Smola \[2002\]](#)).

The Lagrangian  $L$  has to be minimized with respect to the primal variables  $\mathbf{w}$  and  $b$  and simultaneously maximized with respect to the dual variables or Lagrangian multipliers  $\alpha_i$ .

Differentiating with respect to  $\mathbf{w}$  and  $b$

$$\frac{\partial}{\partial b} L(\mathbf{w}, b, \alpha) = 0, \quad \frac{\partial}{\partial \mathbf{w}} L(\mathbf{w}, b, \alpha) = 0, \quad (3.75)$$

and applying the results to the Lagrangian yields two conditions of optimality,

$$\sum_{i=1}^m \alpha_i y_i = 0 \quad (3.76)$$

and

$$\mathbf{w} = \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i \quad (3.77)$$

The optimal weight vector  $\mathbf{w}_0$  is described in terms of a subset of the training patterns, called as Support Vectors, and only those training examples whose corresponding Lagrange multipliers are non-zero contribute to  $\mathbf{w}_0$ . This result is obtained from the Karush-Kuhn-Tucker (KKT) conditions described in [Schölkopf and Smola \[2002\]](#).

Substituting the optimality conditions, Eq.3.76 and Eq.3.77, into Eq.3.74 yields the Wolfe dual of the optimization problem, consisting in finding multipliers  $\alpha_i$  such that

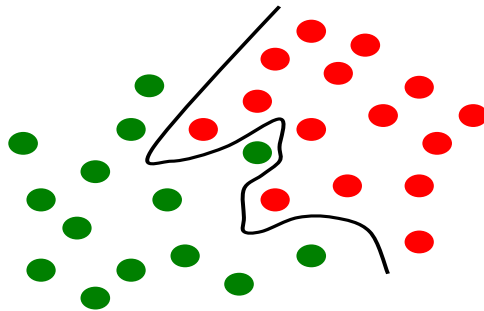
$$\text{maximize} \quad W(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \quad (3.78)$$

$$\text{subject to} \quad \alpha_i \geq 0, \quad i = 1, \dots, m, \quad \text{and} \quad \sum_{i=1}^m \alpha_i y_i = 0. \quad (3.79)$$

yielding a decision function of the form,

$$f(\mathbf{x}) = \text{sgn} \left( \sum_{i=1}^m y_i \alpha_i \cdot (\mathbf{x} \cdot \mathbf{x}_i) + b \right) \quad (3.80)$$

Eq.3.78 and Eq.3.79 can be solved mathematically using quadratic programming (QP) algorithms<sup>9</sup>. However, in many practical situations the datasets show overlap in the feature space and they are not separable. Fig.3.9 shows the problem described above. The solution consist in “relaxing” the constraints, this means consider as support vector the points that subsequently fall on the wrong side of the margin due to their lower influence on the location of the hyperplane. This classifier is known as *soft margin classifier* described in Schölkopf and Smola [2002], Shawe-Taylor and Cristianini [2000].



**Figure 3.9:** The nonseparable case, although exists a nonlinear boundary that can solve the task.

A very important way to generalize the SVMs to nonlinear decision function is introducing the concept of *kernel function*. It is known that the basic idea of SVMs is mapping the data into some other dot product space called

<sup>9</sup>Many practical algorithms are found in <http://www.support-vector.net> or <http://www.kernel-machines.org>.

feature space ( $F$ ) via nonlinear mapping  $\Phi(\mathbf{x})$ ,  $\Phi(\mathbf{x}) : \mathbb{R}^N \rightarrow F$ , then the SVM optimization problem would consist of dot products in this higher dimensional space,  $\Phi(\mathbf{x}_i)$ ,  $\Phi(\mathbf{x}_j)$ . At the end, the optimization problem only requires the evaluation of dot products in the feature space computable with kernel function  $k(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$ .

There are two aspects to consider: First, the evaluation of the kernel will be very expensive to compute if  $F$  is high-dimensional, and second, the mapping  $\Phi(\mathbf{x})$  would not appear in the optimization problem and would never need to be calculated, or even known. The first consideration can be solved efficiently by Mercer's theorem described in [Schölkopf and Smola \[2002\]](#) and sometimes called reproducing kernel Hilbert spaces (RKHSs). Some examples of popular suitable RKHS functions are:

- Polynomial kernel

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \cdot \mathbf{x}_j + c)^p \quad (3.81)$$

- Radial Basis Function Kernel

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x}_i - \mathbf{x}_j\|^2\right) \quad (3.82)$$

The second consideration guarantees the classifier performance due to its high dimensionality that operates the SVM. As shown in [Cover \[1965\]](#), this affirmation responds to Cover's theorem on the separability of patterns, which essentially says that data cast nonlinearly into a high dimensional feature space is more likely to be linearly separable there than in a lower dimensional space.

## 3.7 Summary and conclusions

In this chapter several concepts in the pre-processing and processing of EEG signals, including signal modelling, signal segmentation, filtering and denoising, feature extraction, feature selection and classification have been reviewed. Although all methods have been described in a brief way, they are introduced to give a good theoretical grounding in EEG processing and to better understand the methods proposed and their performance. Signal processing algorithms for EEG applications have specific requirements in filtering, feature extractions and selections.

The chapter also provides key references for further reading in the field of EEG signal processing.

# Chapter 4

## Proposed methods

### 4.1 Introduction

In the previous chapter we have seen the importance of feature construction in the data analysis process largely conditioning the performance of any classifier. More exactly, one should beware of not losing information at the feature construction stage.

Usually, feature extraction is based on different methods that permit to combine the temporal information or spatial-temporal information obtained from techniques such as EEG, MEG, fMRI etc., with the objective of better analyzing the brain functions. Generally speaking, most of the features consist in measures expressed in coefficients obtained by signal transformation such as Fourier, wavelets, Lyapunov, fractional Fourier amongst others, but this information only shows coefficients or values extracted from some transformation that could cause a loss of information from the physical nature of the problem.

Given the importance of knowing the characteristics of the signal, for example, if the signal does not change in some sense (stationary), if it is very contaminated by noise, lasts a short time or it is explicitly known (deterministic); it is very important to use a suitable method that allows to clean the signal, extract information and at the same time describe the nature of the signals.

It is well known that time and frequency analysis by themselves do not fully describe the nature of signals and linear filtering does not offer appropriate solutions for EEG signals. A good alternative is to have processing methods that enable to clean the EEG signal without removing important information and extract information from distributions that represent time and frequency together. From this it could be known which frequency contents are changing in time and from it to develop physical and mathematical

ideas to understand the nature of problem.

Time-Frequency Distributions (TFDs) are robust and compact signal representations that provide means for isolating various signal characteristics of interest in the time-frequency plane. However, TFDs suffer of cross-terms and the EEGs signals are normally contaminated by eyes or muscle movements, which difficults the interpretation of the time-frequency plane. In this chapter we introduce two approaches: one of them is a novel method to eliminate artifacts originated from eyes movements and the other one is a new approach to feature extraction based on three measurements extracted from TFDs. This chapter is organized as follows: Section 4.2 and 4.3 describe the theoretical fundamentals of both approaches: the first one introduces a new method for EEG pre-processing based on adaptive filtering and independent component analysis (ICA), and the second one describes a new approach in EEG feature extraction that is called *tracks extraction*. Each section is accompanied by a series of experiments using EEG signals from epileptic patients and also discussions and conclusions obtained for each experiment. In Section 4.4 the summary and general conclusions are given.

## 4.2 Artifacts elimination using adaptive filtering and ICA

This section describes an adaptive filtering approach in ICA space for eliminating EOG contamination<sup>1</sup>. The principal difference with other methods for ocular artifacts removal is the use of ICA components as reference inputs corresponding to the noise that wants to be eliminated. The adaptive filtering works under ICA domain using as a reference signal the EEG electrodes localized close to the eyes. The experiments are oriented to further analyze the normalized correlation coefficient between raw EEG and EEG filtered by ICA-RLS and also the correspondence of these electrodes with ocular artifacts using the scalp topographic map as shown in Li et al. [2006]. Similarly, this approach is compared with other techniques.

Signal generated by eye movements and blink is known as Electrooculogram (EOG). This signal frequently appears in the recorded EEG as an interference, causing serious problems in EEG interpretation and analysis. To remove the EOG from the EEG, it is convenient to discriminate between artifacts and brain waves without altering important information of EEG activity.

On the other hand, many applications such as brain computer interface (BCI) require online and real-time processing of the EEG signal. The poten-

---

<sup>1</sup>This section is based on Guerrero-Mosquera and Navia-Vazquez [2011]

tial of optimal filtering based on adaptive methodologies that search very efficiently the optimal solution could be used in the EEG signal to optimally perform in real time tasks or improve the performance in detection or classification for EEG signals as described in Guerrero-Mosquera and Navia-Vazquez [2009], Erdogmus and Principe [2006] and He et al. [2007].

The standard approach for eliminating interferences in EEG recording is filtering (typical values: low cutoff frequency of 0.1 Hz and a high cutoff frequency of 70 Hz). However, linear filtering could distort both amplitude and interchannel phase of signals overlapping frequency bands of artifacts with the frequency bands of the EEG. Following Ebersole and Pedley [2003], the use of filters should always be documented on the EEG recording, so that the specialist can interpret their possible influence. In general, the use of filters should be reduced as much as possible, and even better, avoided at all.

Taking these requirements into account, several authors have published different methods for automatic removal of EEG artifacts alternative to linear filtering, using Independent Component Analysis (ICA), e.g.: Iriarte et al. [2003], Romero et al. [2008], Ghandeharion and Erfanian [2006]; etc. ICA allows to separate components in complex signals with the possibility of discriminating between artifacts and brain waves. This method is widely used as a tool to eliminate artifacts with the possibility of combining it with other methods such as Bayesian classifier or high-order statistics, as described in Van et al. [2006] and Delorme et al. [2007].

### 4.2.1 The ICA-RLS method

#### Independent component analysis (ICA)

The ICA technique appears ideally suited for removing artifacts from EEG in domains where, (i) the sources are independent, that means, if we employ  $M$  sensors we can separate at least  $M$  independent sources, (ii) the propagation delays from the sources to the electrodes are negligible and (iii) the summation of potentials arising from different parts of the brain, scalp, and body is linear at the electrodes, as described Delorme and Makeig [2004]. According to the work by Makeig et al. [1996], in EEG source analysis, just assumption (i) is questionable, since we do not know the effective statistical independence among the sources obtained from EEG data registered from the scalp. Moreover, the nature between ocular artifacts and brain signals are different therefore ICA method can be used to separate the ocular artifacts and EEG brain activity into separate components, just as shown the works by Romero et al. [2008], Ghandeharion and Erfanian [2006], Van et al. [2006], Delorme et al. [2007] and Iriarte et al. [2003]. Let us assume that EEG data

$\mathbf{X}$  is arranged in a matrix of  $M$  sensors or electrodes (rows) by  $N$  time points (columns) data values. The value of  $M$  depends on the number of electrodes used. The objective of the ICA algorithm is to find a separating or unmixing matrix  $\mathbf{W}$  such that we estimate the sources as  $\mathbf{S}' = \mathbf{W}\mathbf{X}$ . The  $M \times M$  matrix  $\mathbf{W}$  obtained by ICA is the linear combination of the used channels. The columns of the inverse matrix  $\mathbf{W}^{-1}$  contain the relative weights of the respective components at each of the scalp sensors. These weights give the scalp topographic map of each component and could be a good indicator for selecting EOG artifacts, as described in Li et al. [2006] and Joyce et al. [2004].

A “filtered” EEG can be derived as  $\mathbf{X}' = \mathbf{W}^{-1}\mathbf{S}''$ , where  $\mathbf{S}''$  is the matrix  $\mathbf{S}'$  with the row representing artifact source set to 0. We do not know “a priori” which row must be set to zero, but the linear filtering scheme in Fig.4.1 will let us know, as will be shown in what follows. It is important to know that the spatial order in  $\mathbf{S}'$  does not correspond to the spatial order in  $\mathbf{X}$ .

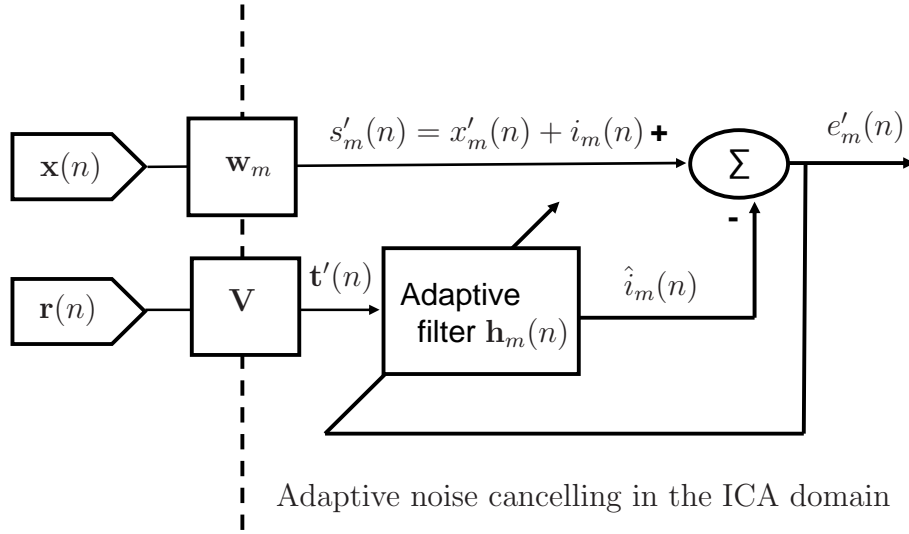
There are many well known ICA algorithms that have proven to be capable of isolating both artifacts and brain signals, for instance those based on Fast-ICA or kernel-ICA, Infomax ICA, SOBI, fastICA and JADE, in Li and Sun [2005], Delorme and Makeig [2004]. In this PhD thesis, we used the Joint Approximate Diagonalization of Eigen-matrices (JADE) that is based on the diagonalization of cumulant matrices, see Cardoso [1998]. This algorithm has been successfully applied to processing of real data sets and EEGs<sup>2</sup>. An extensive study to identify if other ICA algorithm is more suitable exceeds the scope of this thesis and is left as further work.

### Adaptive filtering

In conventional adaptive noise cancellation systems, the primary input signal is a combined signal  $x(n) + i(n)$  where  $x(n)$  represents the clean signal and  $i(n)$  is the interference. This assumes the availability of a reference signal  $r(n)$  expected to be correlated with  $i(n)$  and uncorrelated with  $x(n)$  and the goal is to obtain an output signal  $e(n)$  that is the residual after subtracting from  $x(n) + i(n)$  the best Least Squares estimation of  $i(n)$ ,  $\hat{i}(n)$ .

The method proposed here cannot assume that  $x(n)$  and  $i(n)$  are completely uncorrelated because we use reference channels that could contain information of the brain. Initially, ICA projections are obtained from both EEG data ( $\mathbf{W}$  matrix in  $\mathbf{S}' = \mathbf{W}\mathbf{X}$ ) and reference data ( $\mathbf{V}$  matrix in  $\mathbf{T}' = \mathbf{V}\mathbf{R}$ ), where  $\mathbf{R}$  is a  $P \times N$  matrix that stores measures from the  $P$

<sup>2</sup>JADE Matlab code and tutorial paper is available in <http://perso.telecom-paristech.fr/~cardoso/guidesepsou.html>



**Figure 4.1:** General scheme of automatic EOG noise cancellation using adaptive filtering and ICA. Processing of signal from sensor “ $m$ ” is shown, this scheme has to be run  $M$  times in parallel to process all EEG data.

reference electrodes (Fp1, Fp2, F7 and F8, the ones closest to the eyes) and  $N$  time points. These electrodes could register vertical and horizontal eye movements in the EEG caused by eye blink, saccades, eyes opening and closing that produce changes of potentials at frontal areas of the brain as described in Fisch [1999]. Although EOG dedicated electrodes could have also been used, the use of the Fp1, Fp2, F7 and F8 electrodes as reference has proved to be a reasonable approach and a more direct method, since dedicated electrodes are not always available.

Next, every ICA projection data is fed into an adaptive filtering scheme in Fig.4.1, to be run  $M$  times (possibly in parallel), one for every EEG ICA channel to calculate the  $M$  filter outputs. In Fig.4.1,  $r(n)$  is a  $P \times 1$  vector storing measures from reference electrodes at time  $n$ ,  $\mathbf{h}_m(n)$  is the transversal filter coefficient vector (also  $P \times 1$ ).  $\mathbf{x}(n)$  is a  $M \times 1$  vector storing measures from the EEG electrodes, and  $\mathbf{w}_m$  is the  $m$ -th column of the matrix  $W$ , and also with size  $M \times 1$ . The adaptive filter with weights  $\mathbf{h}_m(n)$  aims at estimating the interfering component  $\hat{i}_m(n)$  present in the  $m$ -th ICA channel in a Least Squares sense, from the reference signal  $\mathbf{t}'(n)$ . The filter operates in ICA domain, and the residual signal is:

$$e'_m(n) = s'_m(n) - \hat{i}_m(n) \quad (4.1)$$

where

$$\hat{i}_m(n) = \mathbf{h}_m^T(n) \mathbf{t}'(n) \quad (4.2)$$

Eq. (4.2) represents a transversal spatial filter with four tap weights. We

adjust the coefficients of the filter by solving:

$$\min_{\mathbf{h}_m(n)} \left\{ \sum_{i=1}^n \lambda^{n-i} (s'_m(i) - \mathbf{h}_m^T(i) \mathbf{t}'(i))^2 \right\} \quad (4.3)$$

The solution of Eq.(4.3) is given by the well known Recursive Least Squares (RLS) algorithm. The use of the forgetting factor  $\lambda$ , where  $0 < \lambda \leq 1$ , allows to use the algorithm in non-stationary situations described in Haykin [1996].

Finally, the component selected is expected to concentrate most EOG interference contribution, and must be eliminated by setting to zero the corresponding row in matrix  $\mathbf{S}'$ , to obtain matrix  $\mathbf{S}''$ . It would also be possible to eliminate more than one ICA component, but that study exceeds the scope of this paper and will be proposed as further work.

We have summarized in Table 4.1 the pseudo code of EEG adaptive filtering using RLS and ICA.

Two initial experiments are presented below for evaluating the performance of the method proposed in removing EOG artifacts. The first one consists in evaluating artifact elimination using EEG epochs, and the other one compares this approach to other techniques under different signal to noise ratios (SNRs). Chapter 5 will show other results of ICA-RLS filtering in classification tasks.

## 4.2.2 Experiment 1: EOG removal on EEG epochs

The aim of this section is to evaluate the performance of the proposed method in removing EOG artifacts in different problems. Several experiments under different signal to noise ratios (SNRs), correlation analysis, comparison with other techniques and EEG segment classification have been performed.

### Data and experimental setup

An EEG record containing mainly eye movements activity was selected. The data was collected from 23 scalp electrodes placed according to International 10-20 system. The sampling frequency was 200 Hz and all registers are from adults. This EEG presents a high interference on 3 channels and the artifact ICA component was chosen according to the visual inspection of an experienced neurophysiology. Using the 10-20 International System and following Fisch [1999], the electrodes with major information about eyes movements are Fp1, Fp2, F7 and F8, therefore,  $P = 4$  and  $M = 19$  in this experiment. The electrodes that record the largest potential change in the presence of vertical eye movements are Fp1 and Fp2 because they are placed directly above the eyes. The electrodes that record the largest potential change when

**Table 4.1:** ICA-RLS Algorithm

```

*****
Inputs:  $\mathbf{X}, \mathbf{R}, \lambda$ 
Output:  $\mathbf{X}'$  (filtered EEG)
Comment: ***** ICA pre-processing using JADE *****
*****

 $\mathbf{W} = \text{jade}(\mathbf{X})$ 
 $\mathbf{V} = \text{jade}(\mathbf{R})$ 
 $\mathbf{S}' = \mathbf{W}\mathbf{X}$ 
 $\mathbf{T}' = \mathbf{V}\mathbf{R}$ 
*****

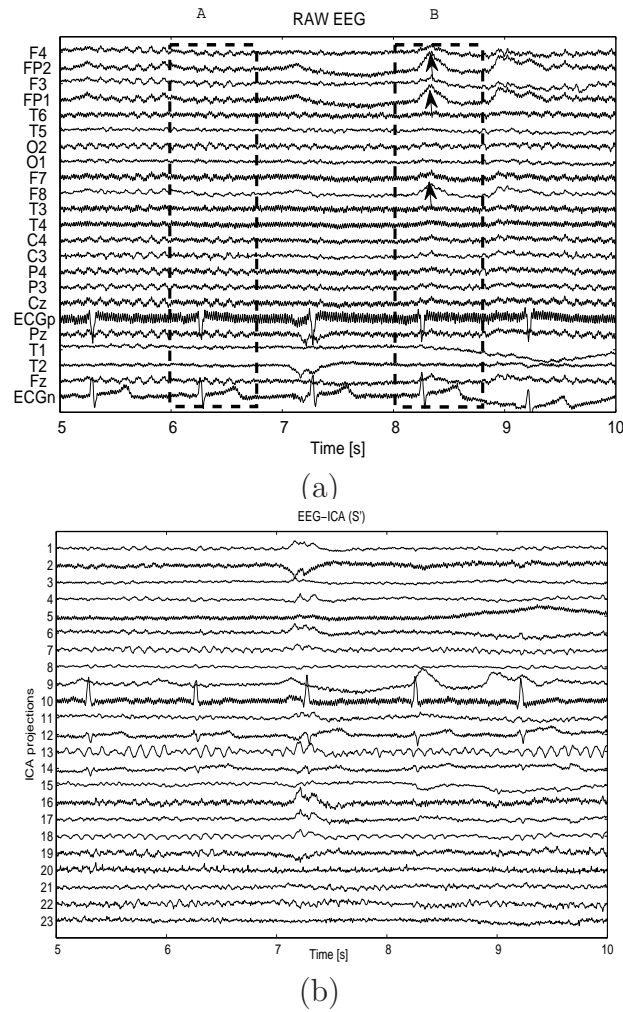
Comment: Noise cancellation in every channel  $m = 1, \dots, M$ 
Comment: ***** RLS Initialization *****
*****

 $\mathbf{P}(0) = 10^4\mathbf{I}$ 
 $\mathbf{h}_m(0) = \mathbf{0}$ 
for  $n \rightarrow 1$  to  $N$ 
   $\begin{cases} \mathbf{t}'(n) = \mathbf{V}\mathbf{r}(n) & \text{ICA projection} \\ \boldsymbol{\pi}(n) = \mathbf{t}'^T(n)\mathbf{P}(n-1) \\ \mathbf{k}(n) = \boldsymbol{\pi}^T(n)/(\lambda + \boldsymbol{\pi}(n)\mathbf{t}'(n)) \\ s'_m(n) = \mathbf{w}_m^T\mathbf{x}(n) & \text{ICA projection} \\ \alpha(n) = s'_m(n) - \mathbf{h}_m^T(n-1)\mathbf{t}'(n) \\ h_m(n) = h_m(n-1) + \alpha(n)\mathbf{k}(n) \\ P(n) = (P(n-1) - \mathbf{k}(n)\boldsymbol{\pi}(n))/\lambda \end{cases}$ 
Comment: *** Recovery of filtered EEG ***
 $j = \arg \min_m \left\{ \sum_{n=1}^N e_m'^2(n) \right\}$ 
Comment: * Set the  $j$ -th row in  $\mathbf{S}'$  to zero to obtain  $\mathbf{S}''$  *
return  $(\mathbf{X}' = \mathbf{W}^{-1}\mathbf{S}'')$ 

```

horizontal (lateral) eye movements are produced are F7 and F8 because they are approximately lateral to the eyes. These electrodes will be our reference signals to build  $\mathbf{R}$ .

A Pentium III with Matlab©(V.7) was used for the implementation of the algorithm in Table 4.1. We explored different values of  $\lambda$  and we observed that this parameter is not critical for the performance of the algorithm (we use the value  $\lambda = 0.9$  through out the experiments). Although it is an adaptive method oriented to real-time applications, in this work we just present off-line results, since to fully extend these results to a time varying scenario, an adaptive ICA algorithm with epoch processing should be used, and this has been left as further work.

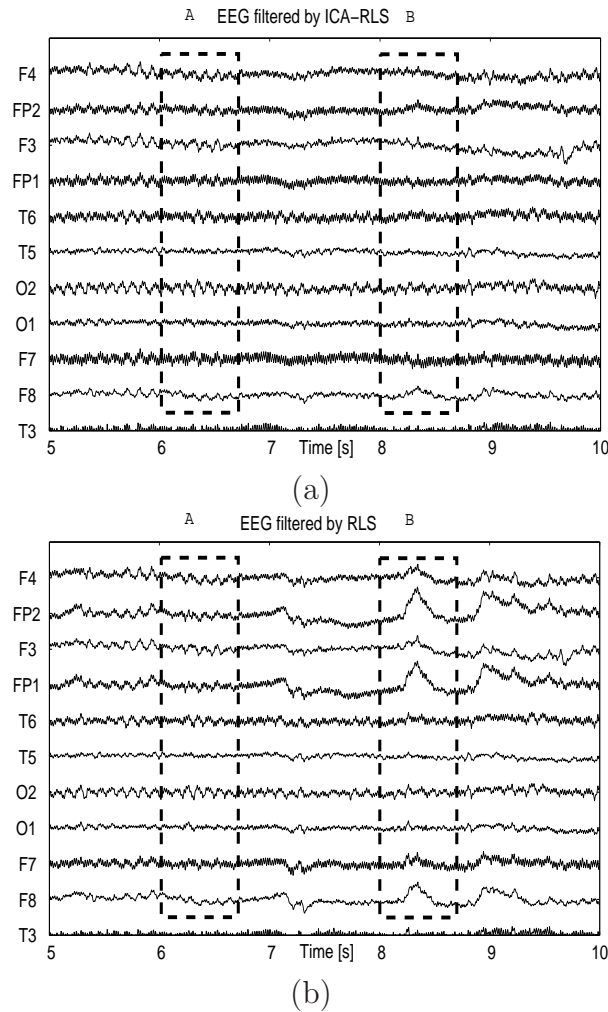


**Figure 4.2:** Raw EEG data (a) and its ICA decomposition (b). Note in (a) that the region marked as “B” is affected by ocular movements, marked with the arrows. In the ICA projections in (b), the EOG contribution has been mainly concentrated in component 9.

## Results

It is shown in Fig.4.2(a) the original EEG with EOG peaks (marked with arrows in the dotted box B) caused by eye movements on the electrodes Fp1, Fp2, F7 and F8. Fig.4.2(b) represents the ICA projections of the same EEG data. Note how it is possible to observe that ICA has been able to separate the Electrooculogram (EOG) contribution, mainly represented in this case by the component number 9.

Fig.4.3(a) shows how the ICA-RLS algorithm has been able to eliminate the artifacts with minor modification of the EEG signals. In fact, as discussed in Iriarte et al. [2003], Li et al. [2006], ICA has demonstrated minimal dis-



**Figure 4.3:** An example of EOG artifact rejection using ICA-RLS and RLS. The result from ICA-RLS algorithm in (a) shows how the algorithm rejects the positive pulse corresponding to eye opening and the negative deflection close to peak since it corresponds to eye closing (dotted box B). Note also the poor performance of RLS algorithm (b) when applied without ICA preprocessing and how our method does not introduce significant changes in the absence of ocular artifacts (dotted box A in (a)).

tortion using measures such as minimal correlation analysis or average waveform similarity. On the other hand, the results without ICA preprocessing (i.e. RLS applied to the EEG signals) are not satisfactory, since the EOG interference is still present, as shown in Fig.4.3(b) which proves the usefulness of ICA. Furthermore, the proposed ICA-RLS method does not affect those parts of the EEG signals where the EOG is not present (zone A, for instance).

To illustrate the performance of the algorithm, we calculate the normal-

ized correlation coefficient between the raw EEG and the EEG filtered by ICA-RLS for each selected segment A and B in Fig.4.3. This coefficient informs about the changes in each channel after removing the artifacts described in Iriarte et al. [2003]. Table 4.2 shows the correlation between preelimination and postelimination of the eye artifact for segment A (without eye artifact) and segment B (in the presence of eye artifact). Note in this table the high correlation for the segment A for all electrodes which denotes that ICA-RLS is not modifying the EEG, and the low correlation in the frontal electrodes Fp1, Fp2, F3 and F8 in segment B, which indicates that the EOG has been filtered out.

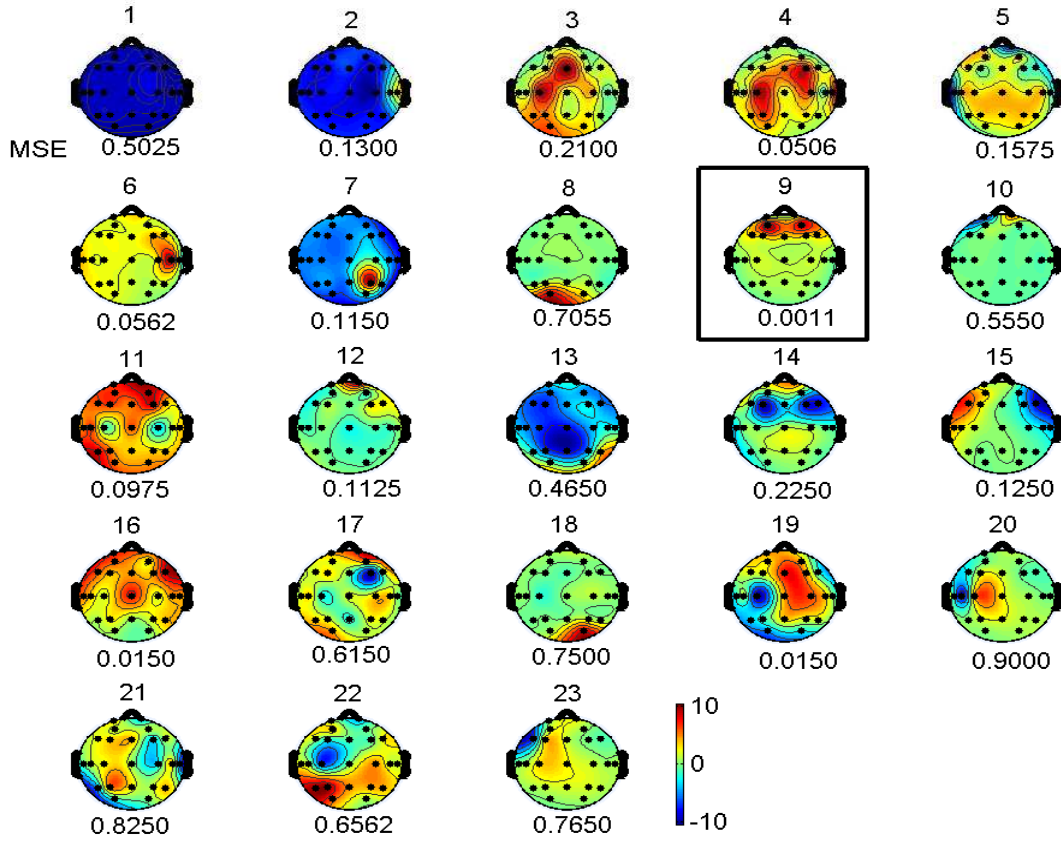
To further validate the results, we analyze using the topographic scalp map the projections corresponding to every ICA component. In Fig.4.4 we have depicted such projections, as well as the MSE value obtained after the ICA-RLS interference cancellation process ( $MSE = E\{e_m'^2(n)\}$ ). Observe that the component number 9 presents the minimum MSE and its projection presents a maximum activity in the frontopolar region.

**Table 4.2:** Normalized correlation coefficient between the raw EEG and the EEG filtered by ICA-RLS for each selected segment A and B in Fig.4.3.

Channel	ICA-RLS		RLS	
	Seg. A	Seg. B	Seg. A	Seg. B
F4	0.9978	0.5297	0.8992	0.9562
Fp2	0.9792	<b>0.2596</b>	0.9659	0.8633
F3	0.9991	<b>0.3425</b>	0.8999	0.8454
Fp1	0.9977	<b>0.3006</b>	0.9999	0.8076
T6	0.9986	0.7901	0.9114	0.8399
T5	0.9951	0.9511	0.8999	0.9009
O2	0.9942	0.9834	0.9102	0.9787
O1	0.9991	0.8913	0.9023	0.8888
F7	0.9946	0.5374	0.8699	0.8362
F8	0.9850	<b>0.3450</b>	0.9899	0.7963
T3	0.9998	0.9601	0.9974	0.9052

### 4.2.3 Experiment 2: Comparison of the algorithm with other techniques

For further validation, the ICA-RLS algorithm is compared against other techniques. (a) ICA-Least Mean Squares (LMS) algorithm with step adaptation equal to  $10^{-3}$ . (b) Visual analysis using topographic maps of the components. (c) ICA-kurtosis method described in Ghandeharion and Erfanian



**Figure 4.4:** Topographic map of the components with their MSE values. Each figure represents the component activity for each projection. Note that the component number 9, with the minimum MSE, presents a maximal activity in the frontopolar region.

[2006] and (d) RLS algorithm applied to vertical EOG component as a reference signal described in He et al. [2004].

## Method

Following Clercq et al. [2006], after separating the brain signal  $B$  from eye artifact  $M$  according to the visual ICA source analysis of an expert, the EEG signal  $X$  can be expressed as follows:

$$X(\alpha) = B + \alpha M \quad (4.4)$$

when  $\alpha$  is a factor that permits to increment the contribution of the eye artifact signal  $M$ . The Root Mean Squared (RMS) value is then:

$$\text{RMS}(B) = \sqrt{\frac{1}{KN} \sum_{k=1}^K \sum_{n=1}^N B^2(k, n)} \quad (4.5)$$

with  $N$  equal to the number of time samples and  $K$  equal to the number of EEG channels. The signal to noise ratio (SNR) measure is defined as:

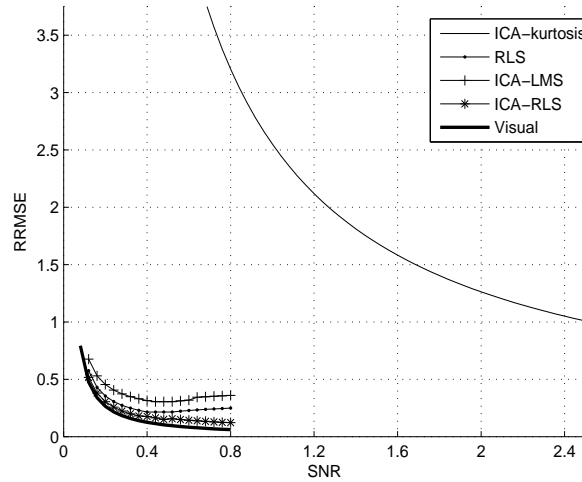
$$\text{SNR} = \frac{\text{RMS}(B)}{\text{RMS}(\alpha M)} \quad (4.6)$$

Modifying the  $\alpha$  factor we can find different SNRs and eye artifact estimations,  $\hat{B}$ . The performance of this estimations could be expressed in terms of the relative root-mean-square-error (RRMSE) defined in Clercq et al. [2006] as:

$$\text{RRMSE} = \frac{\text{RMS}(B - \hat{B})}{\text{RMS}(B)} \quad (4.7)$$

## Results

Fig.4.5 shows the RRMSE as a function of SNR for different EOG removal techniques using  $\alpha$  values from 0 to 2. Following He et al. [2004], the RRMSE values from visual analysis, ICA-RLS, RLS using vertical EOG and ICA-LMS were 0.06, 0.12, 0.24 and 0.35 respectively. Visual analysis method outperformed all methods for all SNRs but requires a human expert to operate.



**Figure 4.5:** Comparison of different EOG artifact removal methods on the EEG using the RRMSE as a function of SNR.

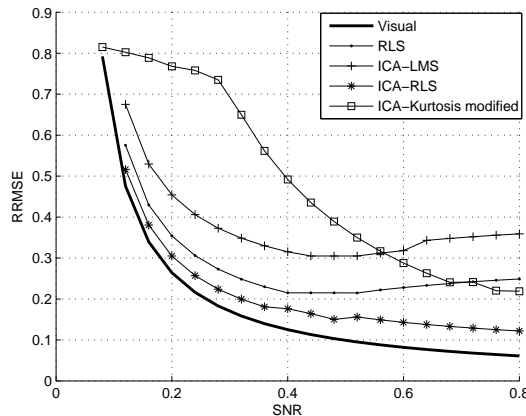
Among all the automated techniques the ICA-RLS method presented the best performance. The ICA-kurtosis method, when it is exactly implemented as described in Ghandeharion and Erfanian [2006], does not present a good performance, although its computation time was fast. However, some changes in the ICA-kurtosis approach are introduced (named as ICA-kurtosis-modified from now on), adapting it to our particular problem, it can obtain the results shown in Fig.4.6.

### ICA-kurtosis-modified

The changes that we have introduced in the ICA-kurtosis-modified method are:

- Daubechies 8 (db8) is used as mother wavelet (original paper uses biorthogonal 4.4). This family of wavelets is one of the most commonly used orthogonal wavelets to process non-stationary EEG signals.
- Fp1, Fp2, F3, F4, F7, F8 and C3 are used as EEG channels (original paper uses F3, F4, Fz, Pz, C3, Cz and FP1). Following [Fisch \[1999\]](#), the proposed subset of electrodes contains more information about eye movement than those proposed by [Ghandeharion and Erfanian \[2006\]](#).
- Kurtosis has been calculated as  $k(x) = E(x - \mu)^4 / \sigma^4$  where  $\sigma$  is the standard deviation of  $x$  and  $\mu$  its mean. (the kurtosis formula in [Ghandeharion and Erfanian \[2006\]](#) is  $k(x) = E(x^4) - 3[E(x^2)]^2$ ). We used the equation implemented in Matlab<sup>©</sup> because it is more accurate since it does not ignore the mean  $\mu$  of the data.
- We used the ICA method based on JADE. Although this method presents a high computational cost described in [Cardoso \[1998\]](#), it could be used in real datasets if the extra computation is not a problem (off-line processing, not real time). Matlab optimized code is available in the web.

The changes mentioned above lead to an improvement in the performance of the algorithm (see Fig.4.6) but among all the automated techniques the ICA-RLS method presented the best performance.



**Figure 4.6:** Comparison of different EOG artifact removal methods on the EEG using the RRMSE as a function of SNR.

The computational cost was also calculated from two points of view: theoretical computational complexity estimation and measured computational time. The computational cost figures are summarized in Table 4.3:

**Table 4.3:** Computational cost figures for all the implemented methods, where  $N$  is the number of samples.

	Computational complexity	Time execution [secs]
RLS	$\mathcal{O}(7N^2)$	45
ICA-LMS	$\mathcal{O}(N^4) + \mathcal{O}(7N)$	41
ICA-RLS	$\mathcal{O}(N^4) + \mathcal{O}(7N^2)$	76
ICA-kurtosis	$\mathcal{O}(N^3) + \mathcal{O}(5N \log_2 N)$	26

Computational complexity was estimated assuming that the wavelet analysis corresponds to decomposition series 5, the adaptive algorithms used 7 reference channels and the ICA method was JADE for RLS, ICA-LMS and ICA-RLS methods and Infomax for ICA-kurtosis. Execution time was estimated by running the algorithm on a PC computer with Intel(R) Core(TM) 2 processor, 2.66GHz, 4 Gb RAM and Matlab©V7.b. It can be observed that the proposed method is the most computationally demanding method, but it provides the best performance. Therefore, if the computational cost is a hard limitation, its use has to be re-considered. If the extra computation is affordable, then the proposed method should be used.

#### 4.2.4 Discussions and conclusion

ICA-RLS approach gives a new alternative method for eliminating noise without calibration. Furthermore, it is easy to implement, very robust and does not need expert supervision. By “robust” it means that all the steps used in the algorithm (ICA decomposition, RLS filtering, etc.) are not facing any ill conditioning problem or critically depend on a particular parameter choice. Basically, ICA processing is able to separate the interfering contributions and grouping them in a separate channel, and the RLS-based interference cancellation is very successful in the identification of the main interfering ICA channel.

Eq.4.3 adjusts the coefficients of the ICA filter based on MSE criteria and helps us to choose the component to be eliminated. There is the possibility of assuming that the frontal activity could also be subtracted by the algorithm but note in Fig.4.1 that all EEG information is contained in the signal  $x(n)$ . Then, the algorithm does not fail when there are no artifacts, that is, there is a high correlation between filtered EEG signal and raw EEG (see Table 4.2).

In the comparison of ICA-RLS method with other techniques, an improvement in the performance of the ICA-kurtosis modified algorithm could be seen. These changes are due to a priori knowledge of different algorithms in EEG signal processing. After the experiment, three conclusions can be extracted: the first is that the modified algorithm has good performance from the computational cost point of view compared to ICA-RLS method; the second is that ICA-RLS is the best method in elimination of artifacts produced by the eye compared with ICA-kurtosis modified. We think that the decision rule used in [Ghandeharion and Erfanian \[2006\]](#) (total number of maximums obtained from nine measures) is not adequate for applications where we are eliminating ICA sources; and the third is that it should be noted how the modified algorithm may be appropriate for EEGs with less presence of artifacts (SNRs with high values in Fig.4.6).

As a final conclusion, the results show that methods based on independent component analysis (ICA) and Recursive Least Squares (RLS) adaptive filtering are able to eliminate eye movement artifacts and efficiently rejects artifacts produced by eye movements. Future work is the performance evaluation in larger data set with different types of ocular artifacts and other activities such as epileptic seizures (next chapter discusses how RLS-ICA method is suitable as EEG pre-processing step within a classifier).

### 4.3 The tracks extraction method (LFE features)

As previously discussed in earlier sections, the EEG is composed of several signals with their respective amplitudes and phases. More specifically, the EEG can be considered as the sum of several monocomponent signals and a noise component.

Then, it is necessary a method that permits to separate frequency components on the time-frequency plane and estimate the values of dominant frequencies in a reliable manner, extracting the important information.

Next, we present a new approach for EEG feature extractions based on the *tracks extraction* method that is based on the possibility of applying a sinusoidal model to EEG signals.

### 4.3.1 The EEG model

As already explained in Section 3.2, an EEG signal can be expressed as follows:

$$X(n) = F(n) + \sum_{i=1}^{n_p} P_i(n - t_{pi}) + \sum_{j=1}^{n_a} R_j(n - t_{aj}) + B(n) \quad (4.8)$$

where  $F(n)$  is the background activity; the  $P_i$  terms represent brief duration potentials corresponding to abnormal neural discharges; the  $R_j$  terms are related to artifacts and  $B(n)$  is measurement noise which is modeled as a stationary process.

The goal is to obtain neural discharge information (i.e.  $P_i$  and  $t_{pi}$ ) corresponding to epileptic seizures from the signal  $X(n)$ .

If noise and artifacts are successfully eliminated, Eq.4.8 can be approximated as:

$$X(n) \approx F(n) + S(n) \quad (4.9)$$

where

$$S(n) = \sum_{i=1}^{n_p} P_i(n - t_{pi}) \quad (4.10)$$

Based on the results of Blume et al. [1984], Freeman [1963], EEG waves represent the combined activity of many neuronal cells which can manifest as oscillatory waves. In this sense the EEG signal may be modeled as a collection of sinusoidal components of arbitrary amplitude, frequency and phase, such that the elementary wave part in Eq.4.10 can be written as:

$$S(n) = \sum_{\ell=1}^L A_{\ell} \exp[jn\Psi_{\ell}] \quad (4.11)$$

where  $A_{\ell}$  and  $\Psi_{\ell}$  represent the amplitude and frequency of the  $\ell$ -th component (from the  $L$  components (waves) conforming the EEG signal), respectively. The amplitudes and frequencies of these components are implicitly related to the  $P_i$  terms of Eq.4.10 and a direct form to calculate the Instantaneous Frequency (IF) based on the phase estimation of  $\Psi_{\ell}$  is:

$$f_i(t) = \frac{1}{2\pi} \frac{d\Psi_{\ell}}{dt} \quad (4.12)$$

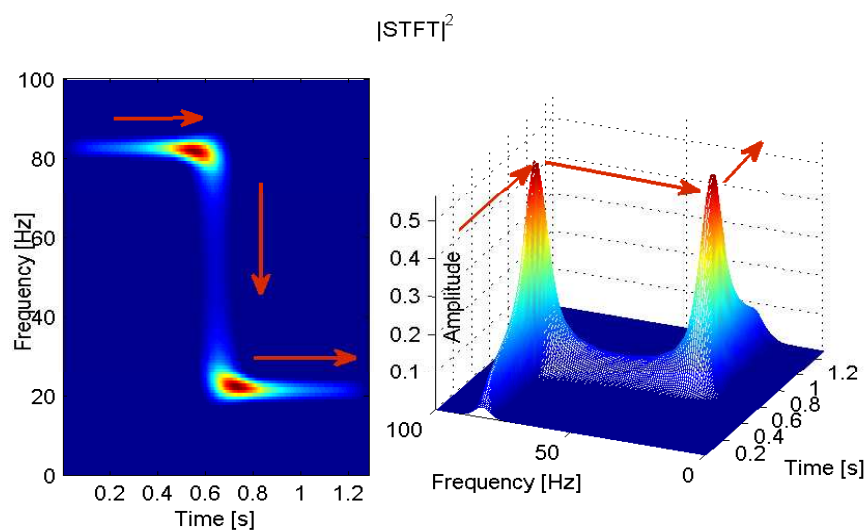
There are several approaches in the literature oriented to find the dominant frequency of a signal or extracting its IFs, for example, calculating the local averages using the marginals distributions and densities of time and frequency as stated Boashash [1992] or Gandetto et al. [2004]; detecting spectral peaks in terms of the frequency reassignment for speech signals, in Zivanovic [2011];

other alternatives apply combinations of different algorithms such as Markov Chain Monte Carlo method, wavelet ridge extraction and Hough transform, as showed in [Carmona et al. \[1999\]](#) and [Barbarossa and Lemoine \[1996\]](#) respectively; using image processing techniques and peaks extractions as described in [Rankine et al. \[2007\]](#), [Malarvili et al. \[2007\]](#) respectively. These methods have been developed in different scenarios such as biomedical engineering or telecommunications.

In the following section, we introduce another technique for IFs estimation using the time-frequency plane from the discrete TFDs.

### 4.3.2 Discrete Time-Frequency Distributions

TFDs provide means for describing various signal characteristics of interest using the most important and fundamental variables in signal processing such as time and frequency. The time-frequency plane represents the energy density of the signal, where the most valuable information is encoded in the instantaneous frequency (IF) that shows the regions where the signal energy is concentrated and how this density changes with time (see [Fig.4.7](#)).



**Figure 4.7:** TFDs represent the energy density in both time and frequency axes. The energy time evolution is showed by red arrow. Left: Time-frequency plane represented by two-dimensional map. Right: Same signal by represented by three-dimensional representation.

The figure above is just one example of a FM signal in which the peaks of energy on the time-frequency plane are easy to “follow”. It is well known that EEG signals usually present different components and it seems adequate to

develop a method to be able to identify dominant frequencies on the time-frequency plane in a simple, robust and reliable way in order to be used in the clinical setting.

The implementation starts converting the signal  $X(n)$  of Eq.4.9 in its analytic signal  $X_a(n)$  and mapping this signal in a time-frequency domain by TFDs.

Next, the Cohen-class TFD is obtained by two steps: first, smoothing the ambiguity function by the weighting function  $\Phi(\theta, \tau)$  and second, applying the two dimensional (2-D) Fourier transforms to the result. In other words, the TFD is a 2-D filtering in the ambiguity function that makes possible to improve the results in IF identification.

Since our EEG records are discrete time signals, the discrete and general version of TFDs over discrete frequency samples  $k = 0, 1, \dots, 2N - 1$  can be calculated using the discrete Fourier Transform (DFT). Although there are distributions that present a good performance for EEG applications such as RID, RSPWV, ridges, etc., see Tzallas et al. [2009a], Papandreou-Suppappola [2003]; it is not our objective to show all the implementations for each TFD. Here, we present our implementation of the Discrete Smooth Pseudo Wigner-Ville (DSPWV). More implementations are in Boashash [2003]<sup>3</sup>.

A simple way to understand the implementation of Discrete SPWV (DSPWV) is to start with the Discrete Wigner-Ville (DWV) distributions, that can be expressed as follows:

$$DWV(n, m) = 2T_e \sum_{\ell=-\infty}^{\infty} X_a(n + \ell)X_a^*(n - \ell)e^{-j2\pi\ell m} \quad (4.13)$$

where  $X_a$  is the analytic signal of Eq.4.9,  $X_a^*$  is its complex conjugate and  $n$  and  $m$  are the time and frequency indices respectively. The discrete-time expressions of the ambiguity function and the weighting function in Eq.3.29 are obtained by sampling at a period  $T_e$  such that  $t = nT_e$  and  $\tau = mT_e$ . For simplicity and without loss of generality, we assume  $T_e = 1$ .

The definition of Eq.4.13 requires the calculus of the quantity  $X_a(n + \ell)X_a^*(n - \ell)$  from infinite values, which can be a problem in practice. We can construct a windowed version using a symmetric, normal data window  $h(m)$  of finite interval  $2N - 1$  as follows:

$$h(k) = \begin{cases} g(k), & -N + 1 \leq m \leq N - 1 \\ 0, & \text{otherwise} \end{cases}$$

Then, the Discrete Pseudo Wigner-Ville (DPWV) distribution of a discrete

---

<sup>3</sup>Experiments and comparisons among TFDs can be seen in Appendix C

time signal is given by:

$$DPWV(n, m) = 2 \sum_{\ell=-N+1}^{N-1} |g(\ell)|^2 X_a(n + \ell) X_a^*(n - \ell) e^{-j2\pi\ell m/N} \quad (4.14)$$

As it is well known, DPWV has a directly dependency in time by the window  $g(\ell)$  that it can be improved by adding a greater degree of freedom. For this, the discrete smooth PWV (DSPWV, whose notation will be  $\vartheta(n, m)$ ) allows independent control of both time and frequency through a separable smoothing function:

$$\vartheta(n, m) = 2 \sum_{\ell=-N+1}^{N-1} |g(\ell)|^2 \sum_{p=-N+1}^{N-1} X_a(n + \ell + p) X_a^*(n - \ell - p) e^{-j2\pi\ell m/N} \quad (4.15)$$

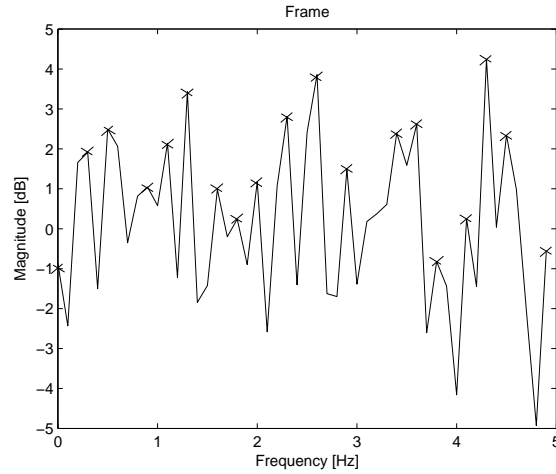
Eq.4.15 shows a especial difference with the other distributions that keep a compromise between time and frequency resolution. This compromise is now replaced by a new compromise between the joint time-frequency resolution and the level of the cross terms.

### 4.3.3 Local peaks estimation and linking

From the time frequency plane  $\vartheta(n, m)$  obtained from the analytic signal  $X_a(n)$ , a procedure starts for localizing the local peaks and linking the frequencies values. Spectral peaks have to be estimated to determine which frequencies are selected to represent the waveform on the time-frequency plane. As each TFD has a correspond resolution and its inherent cross-terms, the local peaks estimation involves identifying a shape using enough points for a given resolution. A practical solution is to detect as many peaks as possible in a frame, as illustrated in Fig.4.8.

The peak detection procedure uses a magnitude threshold “thr” that could be found in an empirical way and its value should ensure that practically all the peaks detected are relevant on the frame.

Once the peaks of the different frames are detected, we proceed to estimate the values of their magnitude and frequency using quadratic interpolation, consisting in quadratically interpolating the maximum peaks per frame and estimating the peak frequency. This adjustment is done because in the sinusoidal model it is necessary that the amplitudes and frequencies are stable, that means, the spectral interpolation looks for samples with maximum magnitude values and with good frequency representation (the partials). As discussed in Zölzer [2002], the spectral interpolation uses only samples immediately surrounding the maximum magnitude sample and it is very efficient.



**Figure 4.8:** Spectral peaks whose locations are denoted by the crosses.

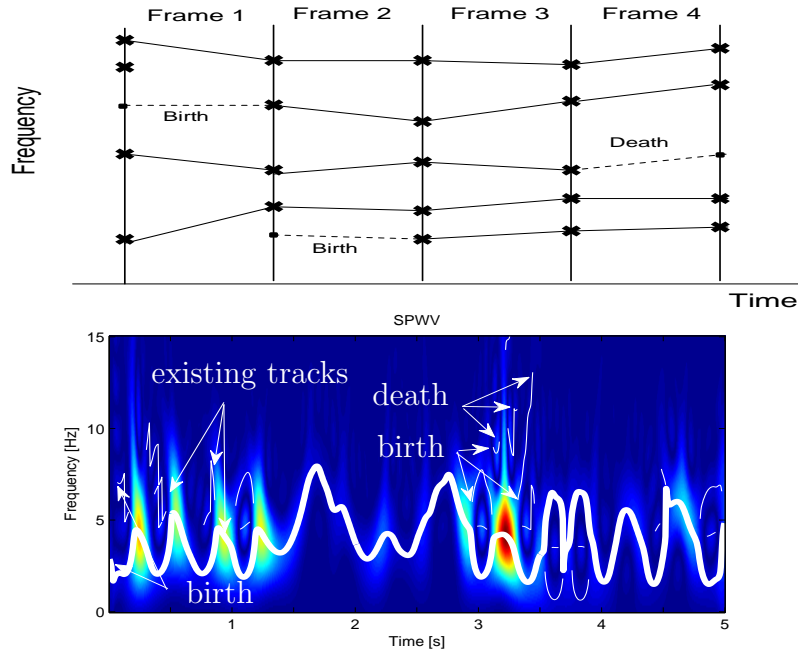
In order to account for rapid variations in the spectral peaks and even unequal number of peaks from frame  $t$  to frame  $t + 1$ , we use the concepts of *birth* and *dead* introduced by McAulay and Quatieri [1986]. This concept is used to account for the appearance or disappearance of spectral peaks between frames, such that tracks are formed by connecting peaks between contiguous frames (see Fig.4.9, upper). By linking peaks which occur at similar frequencies, it is possible to define tracks along the time-frequency plane.

A new track is born if the frequency of a peak in the current frame does not appear in the  $\pm\Delta$  interval of the frequency of that peak in the previous frame. Similarly, a track dies when a peak in the current frame is not followed by another peak in the  $\pm\Delta$  frequency interval in the next frame.

Fig.4.9 illustrates the birth and death of frequency tracks formed by connecting peaks of similar frequencies between frames (upper) and the result of applying this method to an EEG seizure segment using the SPWV distribution (bottom).

#### 4.3.4 Feature matrix construction using tracks extraction

Tracks obtained from time-frequency plane represent relevant information to be used for tasks such as detection or classification. To this end we propose to use three features based on Length, Frequency and Energy of the principal track ( $L, F, E$ ). Fig.4.9 (bottom) shows the existence of a principal track in the seizure corresponding to non-normal activity. Similarly in another EEG record with a duration of 75 secs. (see Fig.4.10), we can observe a longer



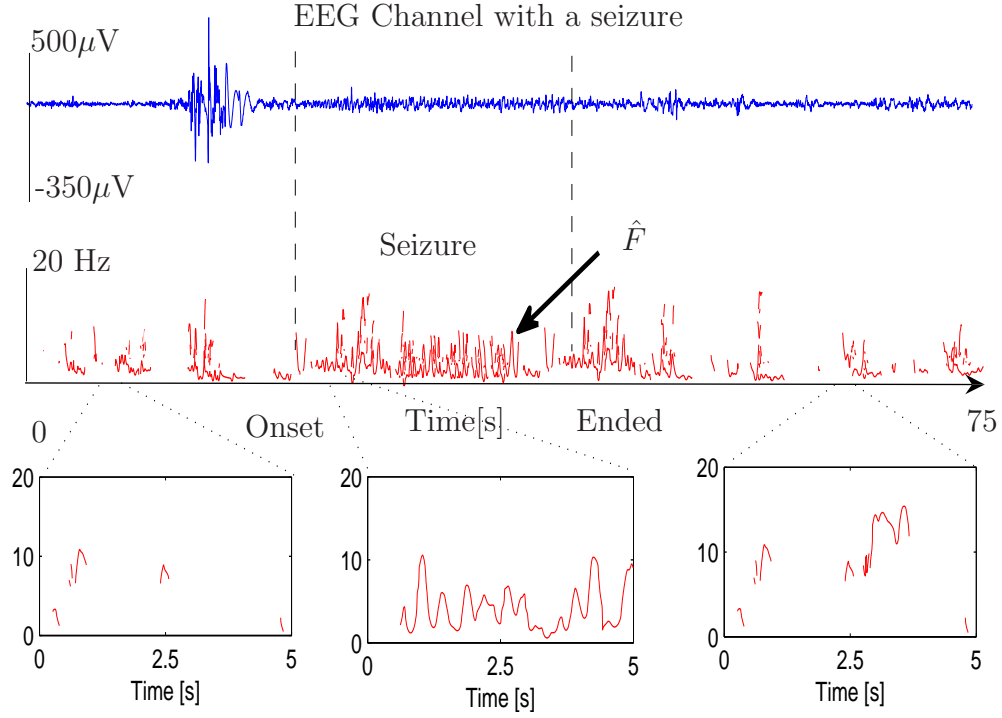
**Figure 4.9:** Upper: Frequency matching process for determining frequency tracking in a TFD window. Each path in the graph is called a track. The birth of a track occurs when there is no partial in the previous frame to connect a peak in the current frame. Conversely, death occurs when a partial does not exist in the next frame to connect a peak in the current frame. Bottom: Peak-matching on the SPWV from a real EEG segment in a seizure. There is a principal track (largest length), marked with a thick white line, and other minor tracks, marked with thinner lines. These tracks serve to summarize the spectral content on the time-frequency plane calculated by the SPWV distribution.

track  $\hat{F}$  clearly visible during the seizure. These appreciations make it possible to introduce a new feature based on the duration of the principal track and use it in the detection task.

Apart from the duration of the principal track, it also becomes necessary to measure other characteristics such as energy and frequency to bring better information about the principal track.

After splitting the segmented EEG into  $K$  segments, we obtain the values  $L_k$ ,  $F_k$  and  $E_k$  from each EEG segment and we subsequently construct a three dimensional feature vector for each segment.

The procedure to be applied to each segment is explained below. The method is based on a discretized version of the  $k$ -th segment in the time frequency plane,  $\vartheta_k(n, m)$ , such that the tracks extraction procedure identifies the coordinates of every track with a dummy variable that is equal to 1 in



**Figure 4.10:** Tracks extraction using a record with a seizure. The figure shows the EEG in the time domain (upper) and the time frequency domain using track extraction (middle). The length of the register is 75 secs. Taking zoom in a window (5 secs.) on three different EEG parts, it can be observed how a dominant and sustained frequency  $F$  appears when there is a seizure, while tracks appear discontinuous in the non-seizure periods.

the points that belong to the track:

$$T_{k,\ell}(n, m) = \begin{cases} 1, & \text{if } \vartheta_k(n, m) \text{ belongs to the } \ell\text{-th track} \\ 0, & \text{otherwise} \end{cases}$$

The length of the track is computed as:

$$L_{k,\ell} = \sum_n \sum_m T_{k,\ell}(n, m) \quad (4.16)$$

the average frequency is

$$F_{k,\ell} = (\sum_n \sum_m T_{k,\ell}(n, m)m) / L_{k,\ell} \quad (4.17)$$

and the energy is

$$E_{k,\ell} = (\sum_n \sum_m T_{k,\ell}(n, m)\vartheta_k(n, m)) / L_{k,\ell} \quad (4.18)$$

We identify the principal track in segment  $k$  as the largest track:

$$\ell' = \arg \max_{\ell} \{L_{k,\ell}\} \quad (4.19)$$

such that the final features for segment  $k$  are:

$$L_k = L_{k,\ell'} \quad (4.20)$$

$$F_k = F_{k,\ell'} \quad (4.21)$$

$$E_k = E_{k,\ell'} \quad (4.22)$$

*Remark:* If there is more than one track with the same length, the principal track is chosen by a largest energy criterion.

Table 4.4 summarizes the pseudo code of EEG tracks extraction using DSPWV distribution. The computational cost introduced by LFE features extractions is in order of  $\mathcal{O}(N^2)$ . At the end, if we have an EEG segment of  $N$  points, the total computational cost (TFD + tracks extraction) will be  $\mathcal{O}(N^2 \log N) + \mathcal{O}(N^2)$ .

Two experiments are developed in order to show the performance of the method both time-frequency analysis of epileptic EEG signals and the epilepsy seizure detection by LFE feature extraction.

### 4.3.5 Experiment 1: Time-frequency analysis of epileptic EEG signals using tracks extraction

#### Data and setting

Five EEGs signals from one epileptic patient were used in this study experiment. The EEGs were recorded at the Clinica Universitaria de Navarra, Department of Neurophysiology (Pamplona, Spain) and all of them presented a focal epileptiform activity. Raw EEG was sampled at 200 Hz and a low pass filter was applied, with a cutoff frequency at 20 Hz.

#### Results

The elimination of undesirable information of the EEG improved our task detection or EEG feature extraction. For example, Fig. 4.11 shows the results of three time-frequency representations of an epileptic EEG segment: SPWV of the raw EEG in the upper figure, the SPWV of the preprocessed EEG in the middle and tracks extraction using preprocessed EEG at the bottom. Note how the ICA preprocessed EEG produces a SPWV transform that highlights the non-stationary signal in an epileptic episode permitting

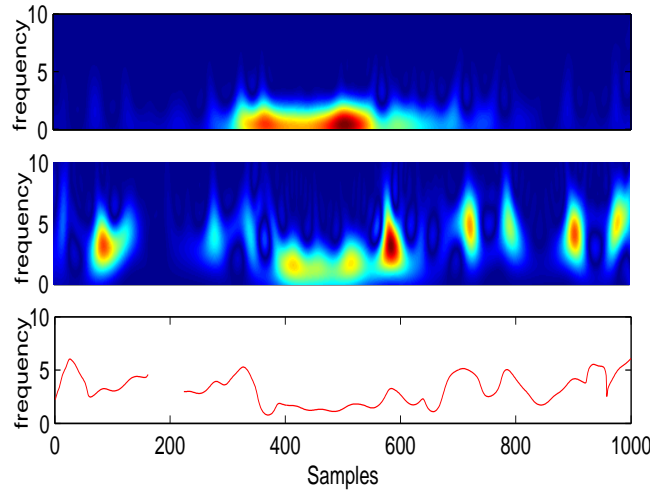
**Table 4.4:** Tracks extraction algorithm (LFE features)

```

*****
Inputs:  $x(t)$ ,  $M$ , thr,  $\Delta$ 
Output:  $\mathbf{X}$  (Feature matrix  $K \times 3$ )
Comment: ***** Initialization *****
*****

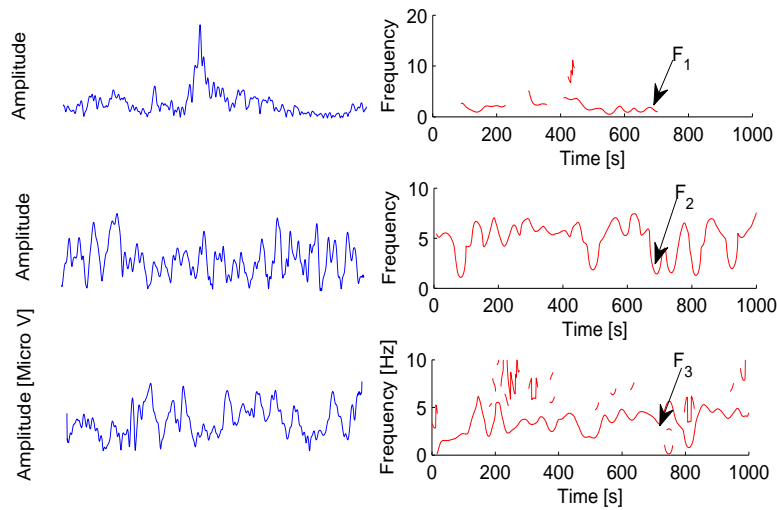
Comment: ** Segmentation and transformation **
Comment: ***  $HT\{.\}$  is the Hilbert transform ****
 $y(k) = x(k + (i - 1) * M), 0 \leq k \leq M - 1, 1 \leq i \leq t$ 
 $y_a(k) = y(k) + jHT\{y(k)\}$ 
Comment: Time-frequency transformation for every segment  $k = 1, \dots, K$ 
 $h =$  Kaiser windows
 $g =$  Kaiser windows
 $track = 0$ 
for  $k \rightarrow 1$  to  $K$ 
 $\vartheta_k(n, m) = DSPWV(y_a(k), h, g)$  Time – frequency plane
{
Comment: Local peak estimation
for  $kk \rightarrow 1$  to  $m$ 
{
 $\vartheta_{max}(kk) = \text{Find-peaks}(\vartheta_k(1 : n, kk), \text{thr})$ 
for each  $\ell$  peak
{
 $\mathcal{A}(\ell) = \text{Quadratic-interpolation}(\log \vartheta_{max}(kk))$ 
then  $\mathcal{F}(\ell)$ 
Comment: Linking based on birth and dead
if  $track == 0$ 
 $track = \max(\mathcal{A})$  born a track
else for each  $\ell$  of  $\mathcal{A}(\ell)$ 
{
 $\mathcal{F}(\ell)$  close to  $\mathcal{F}(\ell - 1) \pm \Delta?$ 
Yes, extend the track
Not, removal actual peak  $\Rightarrow track = 0$  dead a track
}
}
}
}
Comment: *** Feature matrix construction ***
 $T_{k,\ell}(n, m) = \begin{cases} 1, & \text{if } \vartheta_k(n, m) \text{ belongs to the } \ell\text{-th track} \\ 0, & \text{otherwise} \end{cases}$ 
Comment: * Feature extractions *
 $L_{k,\ell} = \sum_n \sum_m T_{k,\ell}(n, m)$ 
 $\ell' = \arg \max_{\ell} \{L_{k,\ell}\}$ 
 $L_k = L_{k,\ell'}$ 
 $F_k = F_{k,\ell'}$ 
 $E_k = E_{k,\ell'}$ 
return ( $\mathbf{X} = [L, F, E]$ )
*****

```



**Figure 4.11:** EEG segment in a seizure.  $N=1000$ , and Kaiser 2-D filter (15,54) was used in SPWV. Upper: SPWV of the raw EEG. Middle: SPWV of preprocessed EEG. Bottom: Track extraction from SPWV using preprocessed EEG. Note how it is easier to obtain a principal track (bottom) using a preprocessed EEG (middle) than a raw EEG (upper). It can also see the improvement in resolution obtained by this method highlighting the non-stationary behavior of the seizure.

Tracks extractions from different EEG segments



**Figure 4.12:** Three different EEG segments (5 secs. length)(right) and tracks extraction (left). The frequencies  $F_1$ ,  $F_2$  and  $F_3$  correspond to the higher track  $F_k = F_{k,\ell'}$  applying Eq.4.20. For this case of 3 segments, the largest tracks are stored into a feature vector  $F = \{F_1, F_2, F_3\}$ .

to better identify the tracks. This improvement is due to the elimination of considerable contribution of noise, background and artifacts that hide important information from seizure activity. This enables the task to extract features and detection of seizure activity, as will be shown in what follows.

Fig. 4.12 shows the results on 3 EEG segments for frequencies extraction using tracks extraction. Note the correspondence between a larger track and its oscillatory frequency for each segment, demonstrating the typical non-stationary behavior during epileptic episodes. For  $k=1,2,3$  we extract  $F_k = F_{k,\ell}$  applying Eq.4.20. At the end, we obtain a frequency vector  $F = \{F_1, F_2, F_3\}$ .

Fig. 4.13 shows the feature vector in an epileptic EEG record for  $k = 58$  segments, consisting in  $L$ ,  $F$  and  $E$  (upper, middle and bottom panels respectively) and illustrates how these features grow during the seizure (segment between the arrows).

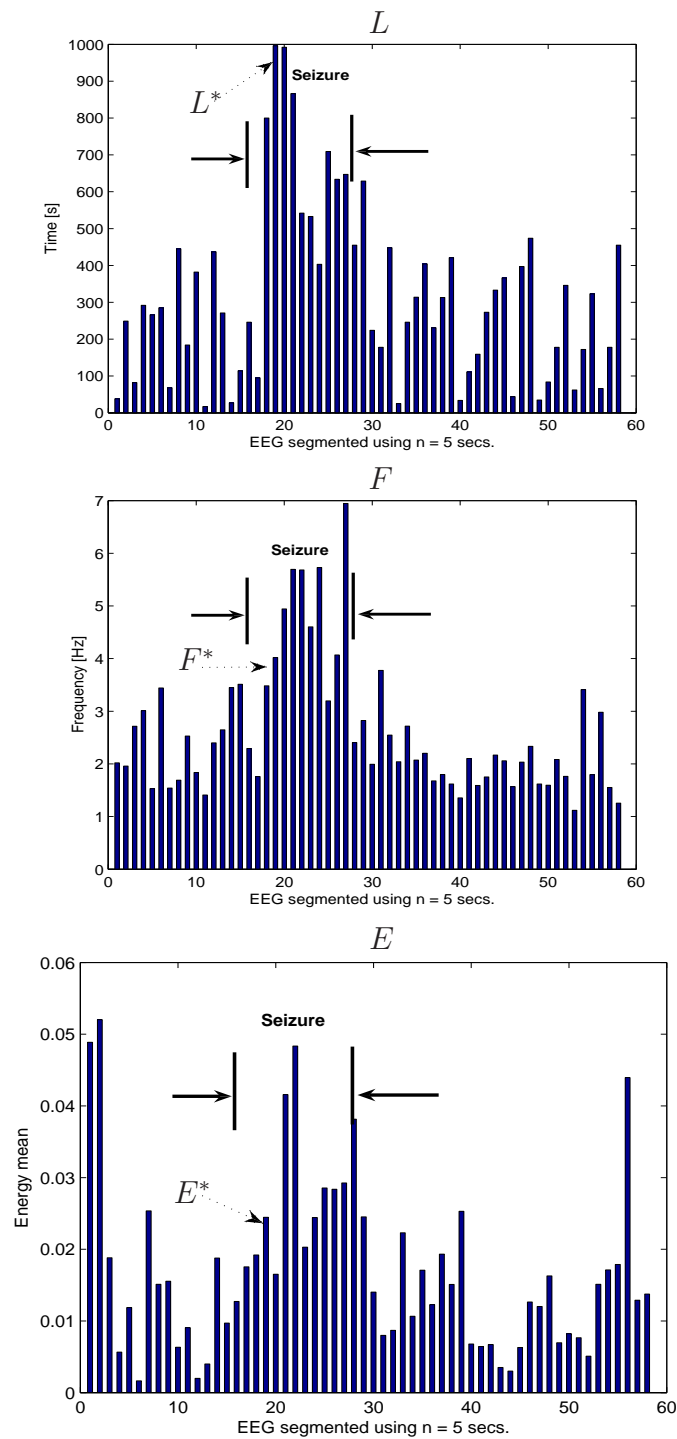
### 4.3.6 Experiment 2: Epilepsy seizure detection using LFE features

#### Data and setting

This section uses two EEGs databases: one of them consisting in 7 adult epileptic patients obtained in a restful wakefulness stage and recorded at the Clinica Universitaria de Navarra, Department of Neurophysiology (Pamplona, Spain). All of them contained focal epileptiform activity, according to experienced neurologists. 17 EEG records of 24 min. are used taken from 23-th and 25-th channels using the 10-20 International System of Electrode Placement with additional anterotemporal electrodes T1/T2. The seizure lasted for a few minutes. In practice, raw EEG data were digitized at a sample rate of 200 Hz using a “DAD-32” equipment (La Mont Medical) and were filtered by a digital low-pass filter with cut-off frequency of 20 Hz. All computation was carried out off-line in a Pentium III computer, using the Matlab©V.7.5 programming environment.

The other database described in Tzallas et al. [2009a] was used to get more generalization. A first detection task called N1, used both normal and seizure EEG segments. An EEG signal of 32.5 min. length was also used to evaluate the effect of the number of samples in the performance of our detector. This problem is called N2.

To obtain the threshold values and apply them at the proposed decision scheme, 6 random EEG records from patient 1 were taken as the training data. In every EEG register  $L^*$  was identified as the largest track, and  $F^*$  and  $E^*$  the corresponding frequency and energy values, at the same time



**Figure 4.13:** Features extraction for an epileptic EEG register ( $k = 58$  segments). The seizure is localized between the arrows and the EEG was segmented using  $n = 5$  secs. (Upper) Vector  $L$ . (Middle) Vector  $F$ . (Bottom) Vector  $E$ . The feature vector  $F$  give us information about the relevant frequency components in a seizure. We can visually choose the larger  $L^*$  value (dotted arrow) with its corresponding values in frequency  $F^*$  and energy  $E^*$  to test the classification algorithm in new EEG datasets.

position as  $L^*$  in the record. The threshold values selected were the median values of all  $L^*$ ,  $F^*$  and  $E^*$  measures on the training set. Their  $L^*$ ,  $F^*$  and  $E^*$  values are depicted in Table 4.5. The obtained thresholds were  $\bar{L} = 2.7$  secs.,  $\bar{F} = 4.13$  and  $\bar{E} = 24\%$  and it was used an empirically value  $\Delta$  equal to 0.5 Hz. The rest of the EEG recordings from that patient, plus data from patients 2 to 7, plus the N1 data collection were used as test data.

**Table 4.5:** LFE features of different EEG's from patient 1 during a seizure (training data).

EEG	$L^*$ [secs.]	$F^*$ [Hz]	$E^*$ [%]
1	2	1.7	26
2	3	5.9	37
3	2.5	4.4	28
4	3.2	2.5	32
5	2.6	6.5	1
6	3	3.8	27

## Results

Table 4.6 presents the results of sensitivity and specificity, which are defined as follows:

- *Sensitivity*: Percentage of EEG segments containing seizure activity correctly classified.
- *Specificity*: Percentage of EEG segments not containing seizure activity correctly classified.

We also used another measure of performance of our detector as a function of dataset size called "*F score*" and defined as:

$$F_{score} = 2 * sensitivity * specificity / (sensitivity + specificity) \quad (4.23)$$

Note the good performance of our method when it is tested with different EEG data (patients number 2-7 and N1 problem) and how this performance is also good when it tries to detect epileptic activity from the same patient (5 EEG's records from patient number 1).

Since the dataset used in the Table 4.6 is quite small (the larger EEG is 15.01 min. that correspond to 901 samples), a new larger EEG database (N2) was used to evaluate the effect of the dataset size in the detector. To evaluate this effect we computed the Receiver Operating Characteristics (ROC) and

**Table 4.6:** Sensitivities and Specificities of EEG's in different patients (test data).

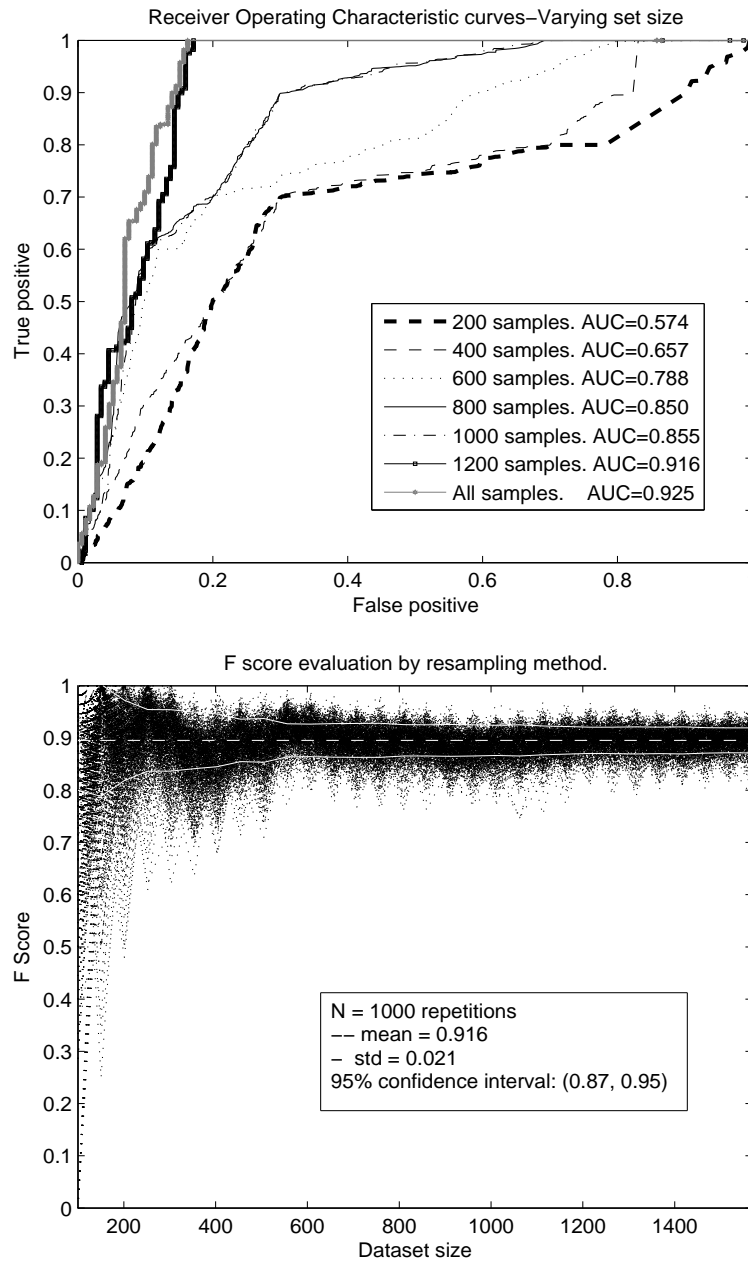
Patient	EEG	Seizure	Sensitivity [%]	Specificity [%]	F-score
1	03:02	00:31	89	97	92.8
	00:40	00:11	90	99	94.2
	15:01	00:22	80	89	84.2
	00:58	00:29	30	100	46.1
	01:34	00:13	77	94	84.6
2	04:54	00:42	72	99	83.3
3	05:24	01:15	88	93	90.4
4	06:45	01:46	56	97	71
5	05:36	00:44	90	99	94.2
6	10:52	01:43	66	100	79.5
7	04:53	01:31	30	100	46.1
N1	00:46	00:23	97	85	90.6
Average			72.1	96	82.3

The duration of EEG records and epilepsy episode are given in minutes

area under a ROC curve (AUC) for a varying data size. A value of AUC of 0.5 indicates random detections, and a value of 1 indicates perfect detection. Note how the AUC increases when more data is used, up to a maximum value of 0.925 (see Fig.4.14, upper).

Following Harrel [2001], a 95% confidence interval for the  $F_{score}$  is estimated using  $N$  bootstrap datasets<sup>4</sup>. Each bootstrap dataset is a simple random sample from 50 to  $n$  values selected with replacement from the original EEG data (the increment step is 50). Because a bootstrap dataset is drawn with replacement, some of the original observations are repeated more than once. The statistics are estimated for each bootstrap dataset and bootstrap confidence interval was computed as the percentile confidence, where the endpoints of the 95% confidence interval are given by the 25th and 975th sorted bootstrap values. For the N2 problem, using  $N = 1000$  and  $n = 1569$  samples, the interval is (0.87, 0.95). Fig.4.14 (bottom) shows the evolution of  $F_{score}$  for  $N = 1000$ , the mean  $m = 0.91$  (dashed line) and standard deviation  $\sigma = 0.0217$  (solid line curves represent  $m \pm \sigma$ ). Note how the value of  $F_{score}$  is more stable when the size of the training dataset increase and it presents a good percentile bootstrap confidence. Although the percentile bootstrap illustrated here is one of the simplest bootstrap confidence interval methods, this experiment in large EEG data discards the hypothesis that our results are overfitted to the data.

<sup>4</sup>For readers unfamiliar with statistical pattern recognition concepts such as ROC curve or bootstrap methods, we recommend to read Appendix B.



**Figure 4.14:** Evaluation of the dataset size effect in the N2 detection problem. (Upper) Receiver operating characteristics (ROC) curves and area under a ROC curves (AUC) values. Observe the AUC values and note the normal behavior of the detector with the increasing data size. (Bottom) Confidence interval estimation for the  $F_{score}$  using  $N = 1000$  bootstrap samples. Note how the value of  $F_{score}$  is more stable when the size of the training dataset increase and it presents a good percentile bootstrap confidence (0.87, 0.954).

### 4.3.7 Discussions and conclusions

A new feature extraction method in epileptic EEG signals relying on track extraction and analysis in a time-frequency plane is presented. The results suggest that the method proposed is a powerful tool for extracting features in EEG signals. The feature vector based on track measurements such as length, frequency and energy ( $L, F, E$ ) in every segment is simple and useful for the detection task. It gives us  $\ell$  tracks on the time-frequency plane  $\vartheta(n, m)$  representing the true nature of the spectral components and really concentrates and localizes EEG frequencies with low computational cost. This opens up the possibility of classifying epileptic EEG channels in a new and promising way.

In order to account for noise, background, artifacts and seizure activity, the EEG has been preprocessed using the method discussed in Section 4.2. As shows Van et al. [2006], ICA has been reported to isolate multiple ictal components in EEG analysis. Automatic EOG removal has proved to be useful except when EEG is very contaminated with muscle artifacts because ICA is not able to eliminate them totally. Muscle artifacts are more difficult to suppress since its morphology and topography causes a confusion with the abnormal spikes. Moreover, this problem does not considerably affect the performance of the detection task because the seizure information is not too much affected by muscle artifacts. Additionally, a low-pass filter was chosen because it is possible to detect epileptic activity on low frequencies and the EEG typically has a frequency content from 1 to 40 Hz, as shown in Lin and Chen [1996].

We have seen the presence of tracks on the time-frequency plane during seizure events as also observed by other authors as Williams et al. [1995]. Boashash and Mesbah [2001] found a time-frequency seizure criteria based on two calibrations in time and time-frequency domain. However, this section has proposed a new form of extracting features based on the principal track by following the ridges (tracks) on the time-frequency plane and obtaining measures such as duration, frequency and energy. Although some authors such as Boashash and Mesbah [2001], Rankine et al. [2007] previously proposed the length of the ridge of the main time-frequency EEG component in a number of applications, including EEG and features such as energy and other frequency-based features have been widely used in the literature dealing with EEG, the extraction method proposed here combines just three features being much simpler than others previously proposed in the literature, since they need many calibrations or use more features to properly work.

Another important issue is the applicability of the method to any distribution due to its non-dependency to a particular TFD. For example, the Ridges Extraction method, which is a good approach for the reassignment method,

is able to extract relevant information from the time-frequency plane, but it depends of the values obtained by the reassignment method affecting the time computation as posted out by Auger et al. [1996]. This problem is also discussed in Carmona et al. [1999] where the frequency updating is not easy because it is necessary to modify the ridge detection algorithm.

The technique proposed could also be used in any scenario where different types of EEG activity have to be detected and associated to particular events. In brain computer interface (BCI) applications, the model could be adopted to detect “brain actions,” e.g. moving up, left, right or down a cursor on a screen using EEG readings. The detection of other brain disorders could also be tackled as described in Hinrikus et al. [2009], Swarnkar et al. [2009], Abásolo et al. [2008]. However, further research is needed to validate the discriminative capability of the tracks extraction features in these new scenarios.

Since the algorithm takes information from a TFD, it is necessary a suitable distribution for EEG signals, subject to the following compromise: high-quality resolution, good detection and low computation time, as discussed in Boashash [2003]. With a good TFD choice, the localization of both amplitude and frequency peaks is less problematic. In the next chapter we will evaluate the Smooth Pseudo Wigner-Ville (SPWV) as the TFD suitable for EEG signal classification as it provides good resolution, low cross-terms and is computationally efficient.

Although the detector presented a good performance by the evaluation of receiver operating (ROC) curves, “*Fscore*” measure and confidence intervals, another important issue is how to select the threshold to yield high sensitivity. The particular values of magnitude threshold and  $\Delta$  used during tracks extraction algorithm do not appreciably affect the results, but a good choice in these values is required. Likewise, it is necessary a long-term analysis to understand the epilepsy behavior and to account for all possible  $L$  values (maximum and minimum) in seizure, because our EEG data records were not very large.

Future works implies the study of a wide range of machine learning methods to better exploit the features proposed here to finally obtain improved seizure detections.

## 4.4 Summary and conclusions

This chapter has presented two new proposals for EEG signal processing: a new method for eliminating eyes artifacts from EEG based on ICA-RLS, and an approach in feature extractions for EEG signals based on sinusoidal model. The results have shown a good performance both in EEG pre-processing and

---

feature extraction (LFE). These methods open the possibility of using it in other scenarios such as BCI and EEG seizure classification.

In the next chapter the combination of two techniques within a Support Vector Machine (SVM) classifier will be presented and we evaluate our method with other methods in EEG feature extractions for classification. This chapter also evaluates the free parameters for each algorithm such as  $\lambda$ ,  $\Delta$ , overlapping percentage, type of TFD, windowing, amongst others.



# Chapter 5

## EEG feature selection and classification using SVMs

### 5.1 Introduction

Detection of events on EEG signals such as seizure activity, event related potentials (ERPs), slow wave-sleep (SWS) or oxygen deprivation on fetal EEG, is important for the diagnosis of abnormalities that may occur in the brain. Generally, EEG analysis involves examining in detail different waveforms or extracting hidden information on the EEG signal and eventually taking some decisions. The analysis of EEG signals often causes disagreements among specialists and the design of medical support devices that facilitates these decisions would be highly valued.

Aside from facilitating a medical decision support system, an automatic detection system also simplifies EEG database analysis. Usually these records contain EEG with different size, ranging from few minutes to several days. On the other hand, there are several conditions such as type of patient, disease, EEG setup, etc, that require considerable skills and time-consuming procedures in large databases. In addition, EEG detection tasks are not easy to perform because biomedical signals are dynamic in nature and present a non-stationary behavior, especially signals from the brain.

In the previous chapter, we have proposed two new methods for EEG signal processing. One of them is an automatic method for eliminating artifacts caused by the eyes using the Recursive Least Squares algorithm with Independent Component Analysis (ICA-RLS). The other one is an approach for EEG feature extraction that obtains a three-dimensional  $L, F, E$  feature vector by “tracks” extracted from the time-frequency plane. Both methods can be used in conjunction with other machine learning methods with the aim of improving the performance tasks such as detection and classification

of EEG signals.

In this chapter, we evaluate the performance of the methods proposed in classification tasks for EEG segments. For this, we propose different classification problems using EEG databases recorded by ourselves as well as other used by other authors. These databases will be described later. On the other hand, we also evaluate tracks extraction using a feature selection method based on mutual information and forward-backward procedure.

For all problems, we use averages both “*Fscore*” and the area under the ROC<sup>1</sup> (AUC) achieved jointly by 1000 bootstrap runs and the best design using support vector machine (SVM). SVM is a technique accepted as the “state of the art” in machine learning. Bootstrap is a good method to reduce the risk of overfitting (poor test performance) and can be used to estimate the confidence intervals, so that *Fscore* average will be a good indicator to assess the quality of the results obtained by the algorithms proposed (see Appendix B). All the simulations were performed in Linux Gentoo, Kernel 2.6.x 64 bits with Matlab©(V.7b) 64 bits. Although the approaches are oriented to real-time applications, in this PhD Thesis we present off-line results.

This chapter is organized as follows: Section 5.2 presents a parameters analysis of the algorithms proposed: for ICA-RLS method, the forgetting factor  $\lambda$  is discussed. For the tracks extraction algorithm, considerations such as window size, time-frequency distribution to chose, percentage of overlapping, threshold values of “thr” and  $\Delta$  are also discussed. Similarly, we discuss the value of K for K-NN algorithm and the SVM selected. These characteristics are necessary in feature selection and classification of EEG signals. Section 5.3 shows the performance of the proposed method with EEG signals taken from real patients. Based on the results presented in this section, ICA-RLS method will be suggested as a pre-processing step in classification tasks. Likewise, this section also compares our feature extraction method with other most widely used as wavelets and the fractional Fourier transform. Section 5.4 provides an analysis of relevant features by forward backward procedure using mutual information as relevance criteria. Section 5.5 generalizes the classification results obtained in previous sections with those obtained by other authors. Finally, Section 5.6 gives the summary and conclusions of the chapter.

---

<sup>1</sup>ROC is an acronym for Receiver Operating Characteristic, explained in Appendix B

## 5.2 Parameter analysis for the proposed methods

As mentioned in the introduction to this chapter, there are several considerations to take into account for both ICA-RLS filtering and feature extraction. Obviously, the assignment of initial values can affect the accuracy and performance of both methods; therefore the description that these parameters may have on the implementation of the algorithms is a fundamental matter. On other hand, this analysis gives rules of thumb for choosing appropriate values for the initial parameters and to assign reasonable limits to ensure better results. Next, each algorithm and their parameters are analyzed individually.

### 5.2.1 ICA-RLS method: $\lambda$ and filter order

The RLS algorithm has an excellent steady-state performance in a stationary environment and its forgetting factor  $\lambda$ , in the case of a non-stationary environment, uses values  $0 < \lambda < 1$ . Values of  $\lambda$  most of the time is chosen based on a priori knowledge but it is necessary to know its possible influence in the application. Following [Haykin \[1996\]](#), when  $\lambda$  is close to 1, smoother and slower convergence is observed than with lower values of  $\lambda$ . However, the important fact here is that we are comparing the prediction error of a bank of filters (all of them running with the same  $\lambda$ ), and if we reduce or increase  $\lambda$ , the behavior of all of them will be analogously modified, ultimately leading to the selection of the same ICA component as EOG (the one with the smaller prediction error). Therefore, although the value of  $\lambda$  does affect the behavior of the individual filters, the overall behavior of the procedure for identifying and removing the component with lowest prediction error is not affected for a wide range of  $\lambda$  values. The effects of the forgetting factor  $\lambda$  on the convergence of the RLS algorithm from the point of view of weight error vector and number of iterations are detailed in [Ardalan \[1986\]](#).

On the other hand, filter order has proved not to be an important factor in the performance of removing artifacts on the adaptive scheme. Studies in [Puthusserypady and Ratnarajah \[2006\]](#) have shown that for various SNR values, the RLS method shows a good performance in different filter orders. In those experiments, the authors used the ratio of the power of the ocular artifact being removed from the primary signal to the power in the estimated EEG.

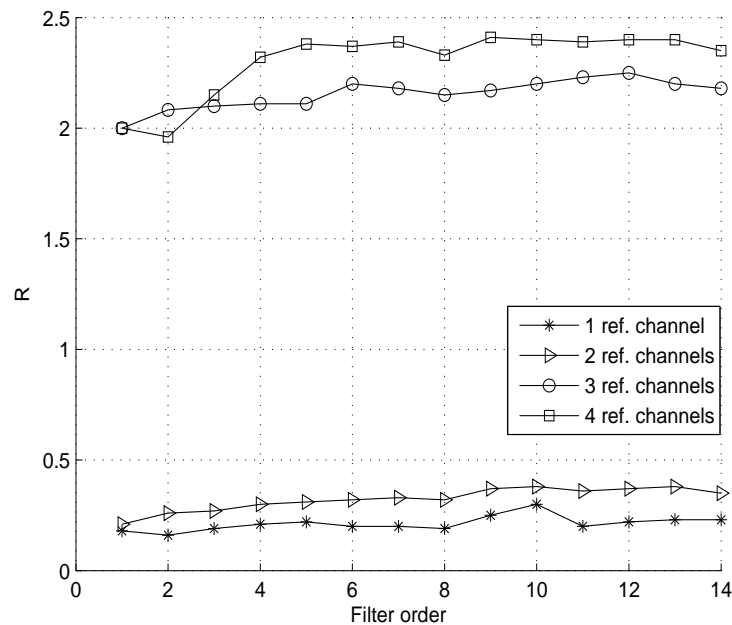
The measurement called as “R” was defined as follows:

$$R^2 = \frac{\sum_{n=1}^N (y(n) - \hat{s}(n))^2}{\sum_{n=1}^N \hat{s}^2(n)} \quad (5.1)$$

Eq.5.1 shows a quantitative comparison of artifact elimination performance and higher values of R correspond to better artifact rejection.

Fig.5.1 shows the corresponding curves for RLS algorithm. It is noted that there is a stability of the RLS algorithm at greater values than 4. Also, there is an improvement in the algorithm when they are using four reference channels. In the same figure, we also see the difference of using one or two reference channels with three or four channels. We think that using only one or two reference channels have less artifacts's information. Eye movements are related to two for the horizontal movements (left and right) and two for vertical movements (up and down) and since the numerator of Eq.5.1 depends directly of the estimation, R values could change considerably.

The ICA-RLS approach uses a 4-th filter order with  $\lambda = 0.9$ .

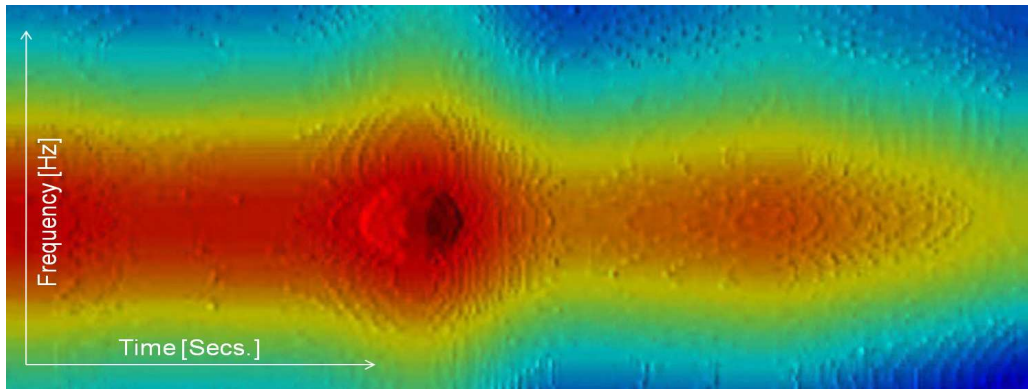


**Figure 5.1:** R vs. filter order for RLS algorithm with multiple EOG reference channels. Higher values of R correspond to better artifact rejection (with permission of Puthusserypady and Ratnarajah [2006]).

## 5.2.2 LFE tracks extraction: distributions, EEG epoch and overlapping

Time-frequency distributions (TFDs) have proved to be a good alternative for non-stationary analysis of EEG signals. Basically, these methods transform a time domain signal to a “snapshot” time-frequency plane where it is possible

to see how the energy and frequencies change over time. For example, Fig.5.2 shows the time evolution of the magnitude of the coefficients of the discrete Gabor transform for a EEG segment with epilepsy. The color spectrum from red to blue corresponds to the magnitude from maximum to minimum. In other words, red areas represent a highest energy concentration than blue areas.



**Figure 5.2:** First burst of epilepsy seen on the time-frequency plane using an atomic Gabor decomposition (adopted from Guerrero-Mosquera [2011]).

To obtain good results on the time-frequency plane it is necessary: (1) to choose an appropriate distribution and this depends on the application, (2) to segment EEG length (epochs), (3) to deal with boundary effects that could be solved by a suitable window analysis and overlap of samples.

Next subsection discusses in detail each consideration in order to take clear the distributions to choose and the adjustments necessary to have good results.

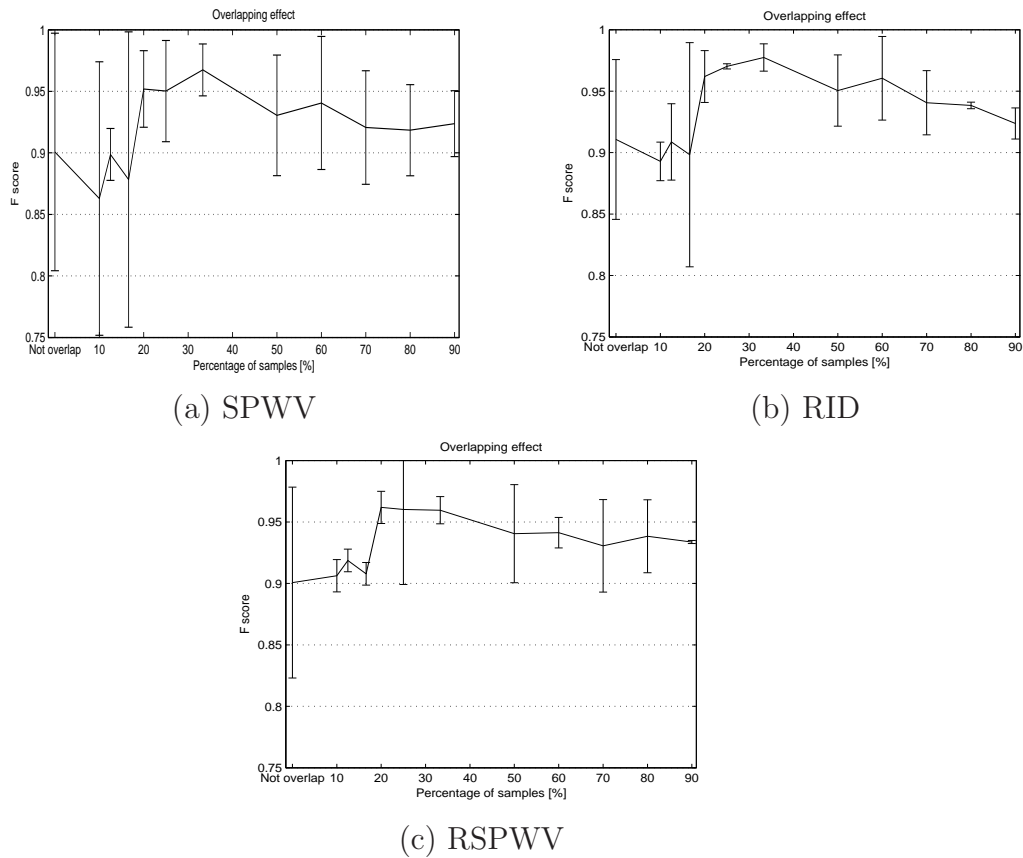
### Selection of time-frequency distributions (TFDs)

As showed in Section 3.4.2, each distribution has different resolutions and suffer cross terms that could significantly affect the results in EEG signal analysis, specially concerning epileptic activity. Although several authors have recommended the TFDs as a good alternative to EEG signals analysis, as described in Papandreu-Suppappola [2003], Boashash [2003], we will explain in more detail the TFDs selection process taking into account characteristics such as resolution, cross terms and computational cost for epileptic classifications tasks.

TFD analysis computes for each segment a discrete Fourier transform (DFT) which in turn drives an analysis window. The windows (usually Hamming, Kaiser or Gaussian) attenuates the leakage effect or spectral edge that is an important inconvenient for feature extraction method because this

method relies on “linking peaks.” An initial analysis is presented in Appendix C, which concluded, in a qualitative way, that distributions based on Wigner-Ville transform, reassignment method and reduced interference distributions are appropriate for epileptic detection. The analysis also concluded that windows such as the Kaiser one presented better resolution both in time and frequency.

Moreover, the features extracted from each TFD are used as inputs into a classifier, therefore the computational cost has to be taken into account and also the performance. The individual analysis for each TFD and its performance gives us more precise conclusions on which distributions may be more suitable for EEG signal classifications.



**Figure 5.3:** Overlapping effect on the SVM classifier for each TFD. (a) SPWV, (b) RID and (c) RSPWV.  $F_{score}$  values are the average on 100 bootstraps runs and vertical lines on the curve represent the standard deviation bar at each % of overlapping. Database used is N1 problem described in Appendix D

Table 5.1 shows the results (average and standard deviation) obtained for each TFD into a classifier scheme.  $F_{score}$  values are the average on 5 bootstrap runs. The database used is named as “problem N1” and is described in

Appendix D. The feature extractions method used is the tracks extraction. For simplicity and following Quiroga [1998], each segment length was 5 secs. and Kaiser windows was used.  $F_{score}$  measurement was used for both the selection of SVM parameters (10-folds cross-validation) and evaluation of each TFD. EEG preprocessing is the raw EEG with artifact rejection by RLS-ICA method.

**Table 5.1:**  $F_{score}$  evaluation of different TFDs for N1 problem.  $F_{score}$  values are the average on 5 bootstrap runs. Features extraction is based on the tracks extraction method. EEG preprocessing correspond to raw EEG with artifact rejection by RLS-ICA method.

TFD distribution	Raw EEG		Preprocessed EEG		Amount of computation per segment [secs]
	Train	Test	Train	Test	
STFT	91.82 $\pm 0.0224$	91.45 $\pm 0.0330$	98.88 $\pm 0.0251$	98.72 $\pm 0.0371$	0.26
Pseudo Wigner-Ville (PWV)	90.95 $\pm 0.0525$	90.28 $\pm 0.0784$	93.18 $\pm 0.0101$	93.91 $\pm 0.0587$	0.29
Spectrogram	89.36 $\pm 0.0774$	89.02 $\pm 0.0139$	89.99 $\pm 0.0222$	89.90 $\pm 0.0452$	0.29
MH Spectrogram	77.45 $\pm 0.0587$	77.28 $\pm 0.0696$	79.18 $\pm 0.0989$	79.11 $\pm 0.0798$	0.36
Wigner-Ville (WV)	87.45 $\pm 0.0556$	87.40 $\pm 0.0451$	98.61 $\pm 0.0133$	98.58 $\pm 0.0310$	0.48
Rihaczek	86.05 $\pm 0.0405$	86.01 $\pm 0.0906$	89.10 $\pm 0.0548$	89.08 $\pm 0.0678$	0.54
Margenau-Hill (MH)	77.05 $\pm 0.0878$	77.03 $\pm 0.0194$	78.44 $\pm 0.0505$	78.40 $\pm 0.0602$	0.57
Page	79.53 $\pm 0.0505$	79.48 $\pm 0.0533$	80.07 $\pm 0.0993$	80.01 $\pm 0.0317$	0.63
Butterworth	89.66 $\pm 0.0300$	89.66 $\pm 0.0143$	92.27 $\pm 0.0588$	92.19 $\pm 0.0967$	0.63
Smooth PWV (SPWV)	97.91 $\pm 0.0274$	97.89 $\pm 0.0125$	100 $\pm 0.0566$	<b>99.36</b> $\pm 0.0084$	<b>4.82</b>
S-transform	90.35 $\pm 0.0778$	90.28 $\pm 0.0261$	91.18 $\pm 0.0317$	91.11 $\pm 0.0287$	5.01
Reduced interference (RID)	97.85 $\pm 0.0230$	97.82 $\pm 0.0545$	100 $\pm 0.0111$	<b>99.57</b> $\pm 0.0068$	<b>8.60</b>
Choi-Williams	95.89 $\pm 0.0796$	95.05 $\pm 0.0680$	98.32 $\pm 0.0809$	98.31 $\pm 0.0666$	9.41
Affine PWV	83.75 $\pm 0.0887$	83.20 $\pm 0.0320$	87.55 $\pm 0.0865$	87.41 $\pm 0.0631$	10.24
Reassigned PWV	92.75 $\pm 0.0565$	92.68 $\pm 0.0531$	98.58 $\pm 0.0378$	98.51 $\pm 0.0223$	14.15
Reassigned SPWV	97.89 $\pm 0.0471$	97.81 $\pm 0.3908$	100 $\pm 0.0194$	<b>99.41</b> $\pm 0.0083$	<b>21.12</b>
Ridge extractions	97.66 $\pm 0.0483$	97.49 $\pm 0.0901$	100 $\pm 0.0296$	<b>99.01</b> $\pm 0.0127$	<b>22.03</b>
Radially Gaussian kernel	79.91 $\pm 0.0163$	79.88 $\pm 0.0199$	82.11 $\pm 0.0249$	82.06 $\pm 0.0661$	32.01

Note in Table 5.1 that there are four distributions with good performance in solving this problem: Smooth Pseudo Wigner-Ville, Reduced Interference (RID), Reassigned SPWV and Ridge extractions. The Ridge extractions method also shows a good performance but has a very high computational cost per segment (22.03 seconds) than the other three distributions. Note also that EEG preprocessing performed better than raw EEG. Section 5.3 will explain in more detail those results.

### Overlapping and EEG epochs

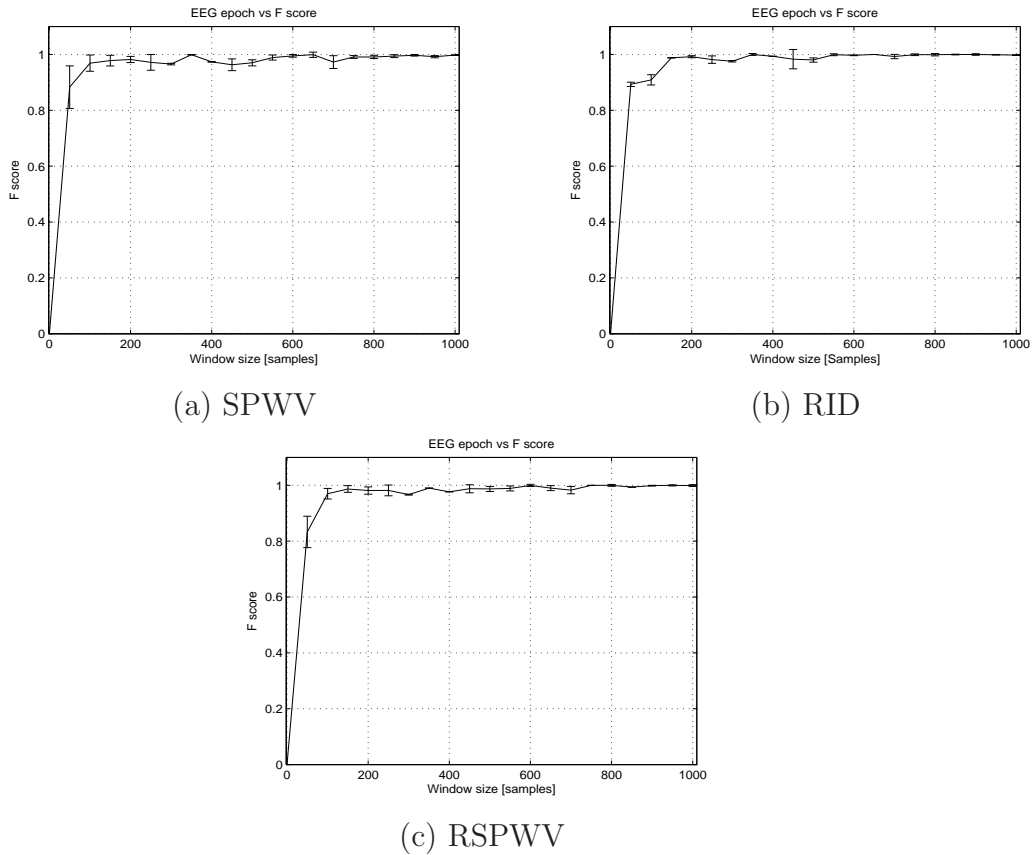
Overlapping is a good option to avoid the boundary effects (spectral leakage) that occurs when we are using slide windows analysis on the time or frequency. Then, it is very important to know what is the appropriate percentage of overlapping because we are possibly introducing redundancy in the problem and this could affect the SVM performance. Fig.5.3 shows the effect on the classifier for different overlapping values for the distributions: (a) SPWV, (b) RID and (c) RSPWV. Note that there is a considerable  $F_{score}$  increase for values between 20% and 40% for all distributions.

Regarding the EEG epochs, several authors have advised EEG segment lengths among 1 to 5 seconds that depend on the model or transformation used, as described in Quiroga [1998], Boashash and Mesbah [2001] and Guerrero-Mosquera et al. [2010a]. It is important to have into account EEG epochs and its effect on the classifier because tracks extraction method uses a feature called as “L” that consists in the partial duration of the tracks. This duration has to take maximum or minimum values that guarantee the classifier performance. In other words, we want to eliminate the assumption that using a segment length of 2 secs. implicitly it assumes that the maximum duration of the seizure is less or equal than 2 sec (after removing boundary effects). Fig.5.4 shows the effect of EEG epochs on the classifier for each distribution chosen: (a) SPWV, (b) RID and (c) RSPWV. Note in the figure that values from 200 samples (1 secs) ensure a stable performance of the classifier for the three distributions.

### Thresholds: “thr” and $\Delta$

The tracks extractions algorithm makes a “linking peaks” on the time-frequency plane using two thresholds: a peak magnitude called in the algorithm as “thr” and a matching interval  $\pm\Delta$  (see Table 4.4).

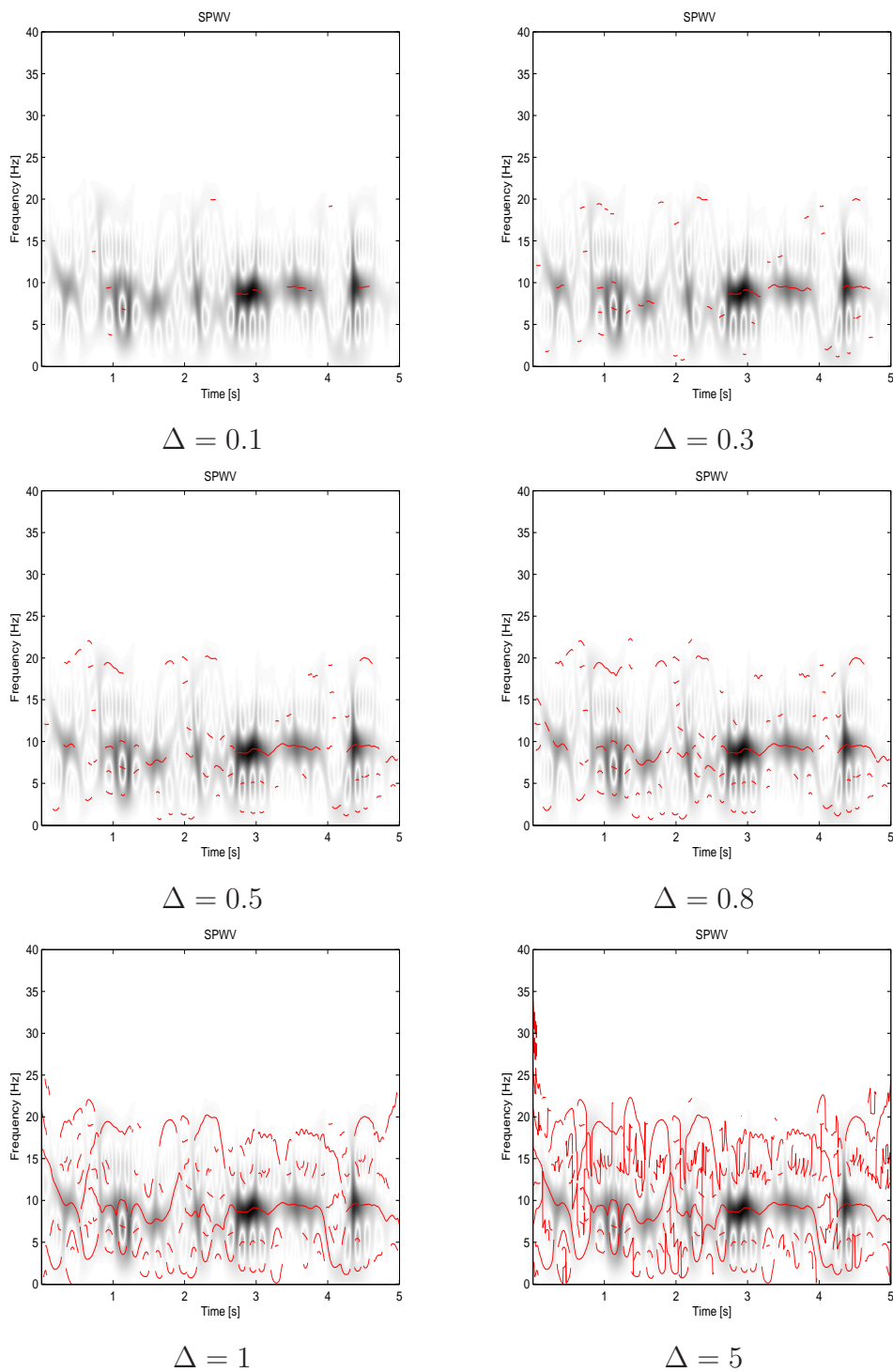
The threshold “thr” is used in the process of finding peaks and defines a spectral peak magnitude. This value determines the total number of candidates per frame that are used in the linking peaks process. Very small values of “thr” can produce an overwhelming number of partials that could



**Figure 5.4:** EEG epochs effect on the SVM classifier for each TFD. (a) SPWV, (b) RID and (c) RSPWV.  $F_{score}$  values are the average on 100 bootstraps runs and vertical lines on the curve represent the standard deviation bar at each epoch length. Database used is N1 problem described in Appendix D

primarily affect the computational cost of the algorithm. For this reason, we have to choose an appropriate value of “thr” that leads to small numbers of partials, while guaranteeing at least enough partials to solve the classification problem.

In addition, it is necessary to find which frequencies are related to these peaks. Finding frequencies on the time-frequency plane is limited by the TFD resolution, so in order to fit the peaks in both magnitude and frequency we use the quadratic interpolation and thus obtain more data for the location for each peak. Quadratic interpolation makes a quadratic fit of all maxima above “thr.” We choose “thr” as the ratio (in dB) between the largest peak in the frame and the absolute lower threshold.

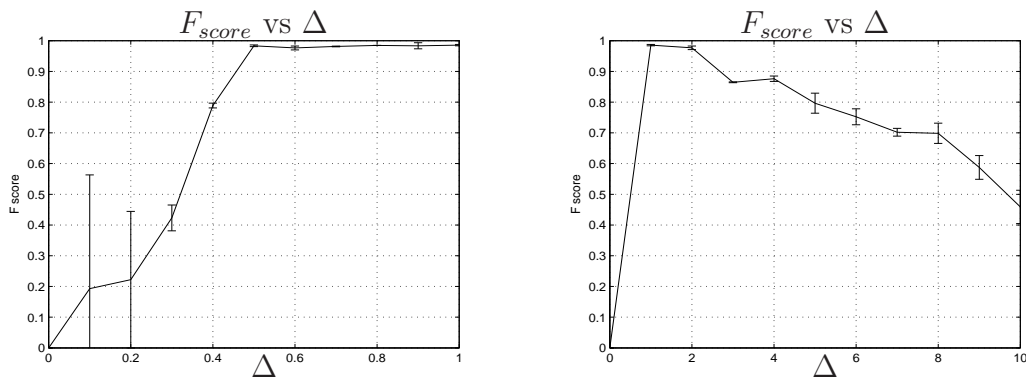


**Figure 5.5:** Linking process on the SPWV distribution (background) for different values of  $\Delta$ , segment length 5 secs and 30% of overlapping. Note the partial increase on the time-frequency plane (red lines) with increasing  $\Delta$ .

On the other hand, the detection algorithm proposed in this Thesis detects activities in the low frequencies. In Chapter 4 we showed several time-frequency planes obtained from epileptic EEGs, where we could see epileptic activity within the range of 0 – 20 Hz. Therefore, since the matching interval  $\pm\Delta$  values define the range where a peak may continue in the next frame and continue as a partial, the value of  $\Delta$  is crucial for birth/death process used in LFE feature extractions.

An initial qualitative analysis is done to observe which is the effect that produces the different values of  $\Delta$  in the creation of partials on the time-frequency plane. Fig.5.5 shows this effect on the SPWV distribution. Note in these figures how there is an increasing number of partials as  $\Delta$  increases. This results show the important role that this parameter has on the classifier performance.

To more clearly assess the working range of  $\Delta$ , we propose to evaluate the effect of this parameter in the output of a classifier (Fig.5.6). The left panel shows the performance of the classifier when  $\Delta$  varies in the range values between 0-1 Hz with a step increment of 0.1 Hz. In the right panel  $\Delta$  values between 0-10 Hz are covered with a step increment of 1 Hz. Note that there is clearly a stable working range of  $\Delta$  between 0.5 and 1 Hz, which leads us to conclude that any value between these limits provides a performance almost equal in the classifier. From here on we will use EEG epochs of 5 secs.,  $\Delta = 0.5$  Hz and 30% of overlapping.



**Figure 5.6:** Effect of  $\Delta$  in the classifier performance. Values between 0-1 Hz with a step increment of 0.1 Hz (left). Values between 0-10 Hz with a step increment of 1 Hz (right). TFD RID and 30% of overlapping.  $F_{score}$  values are the average on 5 bootstraps runs and vertical lines on the curve represent the standard deviation bar at each value of  $\Delta$ . Database used is N1 problem described in Appendix D. Note that there is a stable working range of  $\Delta$  between 0.5 and 1 Hz.

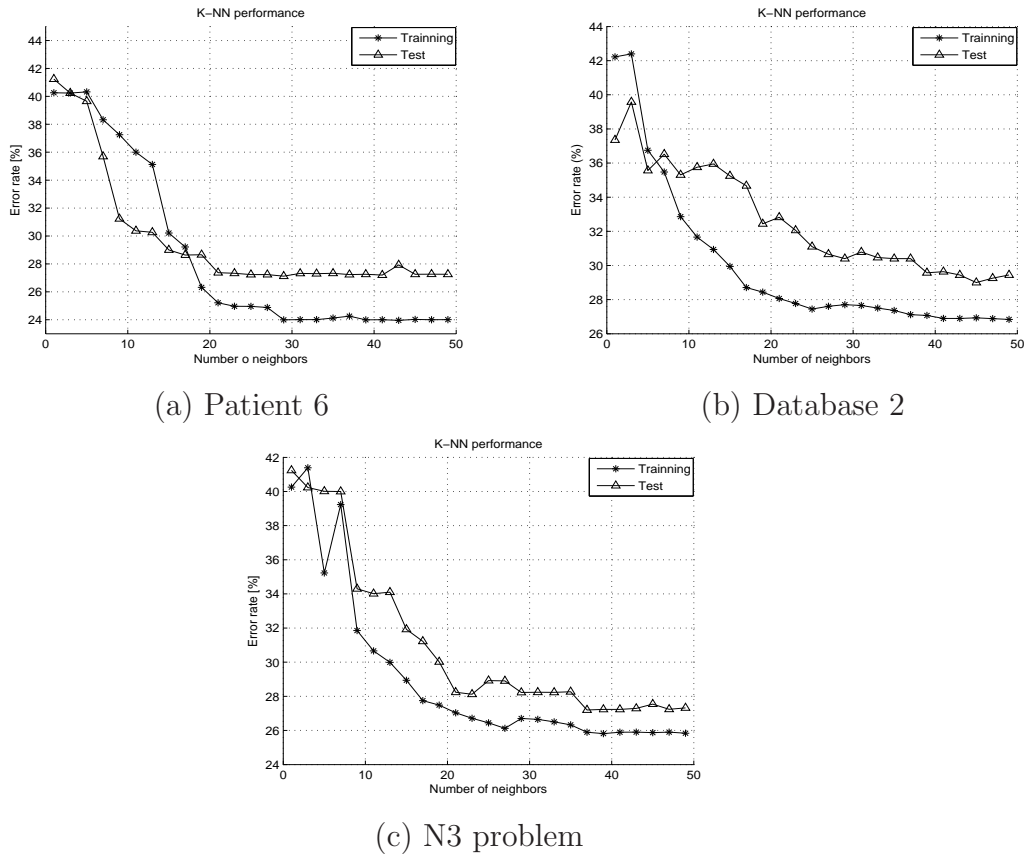
### Selection of K value for K-NN algorithm and SVM classifier

Apart from parameters from tracks extraction and RLS-ICA methods, two issues still remain to be dealt with: parameter  $K$  in for the K-NN algorithm used in estimating the probability density function, and the type of SVM classifier we will use to evaluate the performance of both methods. K-NN estimation is necessary to calculate the mutual information (MI) which is the relevant criteria for the feature selection procedure.

Generally speaking, SVM have proved to be the most appropriate alternative for solving the classification problems and their solution is supported in statistical learning theory Vapnik [2000], Schölkopf and Smola [2002]. In fact, SVM has reached the best results in both linear and nonlinear form due to regularization to outliers effect, its simplicity and robustness with respect to the curse-of-dimensionality.

Based on Liang et al. [2010], Lotte et al. [2007], Güler and Übeyli [2007], SVMs with radial basis function (RBF) have demonstrated a good efficiency in EEG classification. Although there are results with other kernels, for this section we will use SVMs with RBF kernel. Determining the most appropriate SVM kernel for solving the EEG classification problems with epilepsy, is beyond the scope of this thesis and it is not our goal. The parameters  $\sigma$  and  $C$  are calculated by  $K$ -fold cross validation (CV) method which is explained in Appendix B.

On the other hand, it is well know in K-NN estimation that the value of  $K$  is generally dependent on the data. Initially there is no optimal way to choose the value of  $K$  usually this value is found by cross-validation involving an increase in the computational cost. We propose to evaluate the error rate of SVM classifier vs the number of neighbors for three databases. That way we can get  $K$  values which ensure a low error rate and classifier's stability. This experiment has the following setup: RID distribution, tracks extraction method with EEG epochs of 5 secs.,  $\Delta = 0.5$  Hz and 30% of overlapping and RBF-SVM parameters was found by 10-folds cross validation. We chose randomly the databases: patient 6, database 3 and N3 classification problem that are described in Appendix D. Subsequently, each database has been partitioned by using 60% for training and the remaining data for test. Results for each database are showed in Fig.5.7. Note how the error rate decreases rapidly up to  $K = 20$  for all three databases. From this value we observe no notable changes in the error rate. At the end, a value of  $K = 30$  was chosen for all experiments of feature selection.



**Figure 5.7:** K-NN performance using a RBF-SVM classifier. Feature matrix is based on tracks extraction method with RID distribution, EEG epochs of 5 secs.,  $\Delta = 0.5$  Hz and 30% of overlapping. RBF-SVM parameters was found by 10-folds cross validation.

### 5.3 Clinical evaluation of the proposed methods

Since the features extraction method and ICA-RLS provide new alternatives to EEG signal processing, it is important to evaluate the possibilities which these features offer in comparison with other methods and the effectiveness in artifact rejection.

So far, we have seen in the previous section, how the pre-processing step based on ICA-RLS improves the classifier performance (see Table 5.1). This section we will show the use of EEG data from real patients to evaluate the EEG preprocessing by both ICA-RLS and tracks extraction. First, we will use the database 2 (described in Appendix D) which has high ocular artifacts contamination and presence of interictal activity. From this database, we created a feature matrix and individually tested each feature or extraction

method using a SVM classifier. Feature matrix, with dimension (D) equal to 65, has been formed by three feature extraction methods: tracks extraction (D=3), wavelet coefficients (D=45) and fractional Fourier coefficients (D=17). The SVM parameters are found by 10-folds cross-validation and  $F_{score}$  measurement was used for both the selection of SVM parameters and the performance measures of the classifier. Additionally, each  $F_{score}$  value presented in Table 5.2 is the average corresponding to 1000 bootstrap runs. Distributions chosen are SPWV, RID and RSPWV.

Table 5.2 shows clearly the improved performance obtained with a pre-processed EEG by ICA-RLS algorithm in all extraction methods. Note that this improvement is noticeable for the three distributions, but among these, the best results were obtained using the RID distribution. Another way to evaluate the methods proposed is using a database with more diversity. Database 1 (described in Appendix D) has been recorded from different patients, brain regions and class balance. Results showed in Table 5.3 (raw EEG) and Table 5.4 (preprocessed EEG) confirm that ICA-RLS adaptive scheme could be considered as a good alternative for removing artifacts in the EEG. The results also show the relevance of the LFE features in different classification problems (patients 1, 2 and 5). For this database, the AUC measurement was used both in the selection of SVM parameters and in the performance measurements of the classifier.

Looking for more generalization, we also use the database 3 described in Appendix D that is frequently cited in the state of art, in [Andrzejak et al. \[2001\]](#). Maintaining the same features matrix (D=65) and training the SVMs with  $F_{score}$  measurement, 10-folds cross-validation, 1000 bootstrap runs and SPWV, RID and RSPWV distributions, we obtain the results shown in Table 5.5, Table 5.6 and Table 5.7 respectively<sup>2</sup>. Finally, although there are some other methods proposed to EEG extraction in the literature as described in [Sanei and Chambers \[2007\]](#), we have chosen these algorithms for two reasons: first, both wavelets and the fractional show better results than others methods, for example: FrFT has the property that if we increase gradually the order, we can obtain more information in form of coefficients than the Fourier transform. Besides, wavelets improves the time-frequency resolution through multi-resolution analysis. The second reason is that tracks extraction contains three detectors: energy (E), frequency (F) and track length (L), which together can solve complex problems and improve the EEG classification performance apart from offering information closer to nature of the

---

<sup>2</sup>The statistical relevance of the results has been verified by means of a Kruskal-Wallis test, which is a sort of nonparametric ANOVA test that does not assume Gaussianity in the data under study. In all cases (except between Fractional Fourier (FrF) and LFE+Wavelets (W) in the N1 case) a p-value smaller than 0.01 has been obtained, thereby rejecting the null hypothesis that data come from the same distribution. Note in these tables the difference in difficulty among N1 (easy), N2 and N3 (hard) problems.

**Table 5.2:** Feature evaluations for database 2.  $F_{score}$  values average and standard deviations correspond to 1000 bootstrap runs. Each abbreviation corresponds to the feature(s): (L) duration of track; (F) track frequency; (E) track energy; (LF) L+F; (FE) F+E; (LE) L+E; (LFE) L+F+E; (W) wavelets coefficients; (FrF) fractional Fourier coefficients. (+) means a feature combinations.

Features	Dimension	SPWV		RID		RSPWV	
		Raw	Pre-processed	Raw	Pre-processed	Raw	Pre-processed
L	1	54.09 $\pm 0.0036$	54.90 $\pm 0.0022$	54.65 $\pm 0.0014$	55.18 $\pm 0.0089$	54.88 $\pm 0.0067$	55.07 $\pm 0.0050$
F	1	58.84 $\pm 0.0029$	58.92 $\pm 0.0081$	59.26 $\pm 0.0088$	59.32 $\pm 0.0030$	58.67 $\pm 0.0062$	59.01 $\pm 0.0046$
E	1	69.45 $\pm 0.0013$	70.13 $\pm 0.0055$	71.22 $\pm 0.0023$	73.80 $\pm 0.0079$	70.14 $\pm 0.0050$	72.03 $\pm 0.0019$
LF	2	61.35 $\pm 0.0034$	62.36 $\pm 0.0066$	65.21 $\pm 0.0068$	69.87 $\pm 0.0015$	62.07 $\pm 0.0041$	63.28 $\pm 0.0032$
FE	2	92.68 $\pm 4.06 \times 10^{-4}$	93.67 $\pm 0.0306$	94.32 $\pm 6.80 \times 10^{-4}$	95.67 $\pm 0.0427$	93.19 $\pm 0.0101$	93.99 $\pm 0.0856$
LE	2	91.28 $\pm 4.06 \times 10^{-4}$	93.05 $\pm 0.0306$	92.29 $\pm 6.06 \times 10^{-4}$	94.35 $\pm 0.0427$	91.25 $\pm 0.0101$	93.87 $\pm 0.0856$
LFE	3	93.21 $\pm 0.0184$	96.23 $\pm 0.0183$	94.32 $\pm 0.0520$	96.74 $\pm 0.0113$	92.99 $\pm 0.0885$	97.04 $\pm 0.0253$
W	45	93.15 $\pm 0.0191$	93.21 $\pm 0.0185$	94.01 $\pm 0.0021$	94.33 $\pm 0.0548$	93.01 $\pm 0.0665$	93.88 $\pm 0.0221$
FrF	17	90.09 $\pm 0.0263$	91.95 $\pm 0.0287$	91.22 $\pm 0.0525$	92.37 $\pm 0.0002$	91.64 $\pm 0.0113$	93.04 $\pm 0.0699$
LFE+W	48	95.23 $\pm 0.0159$	97.21 $\pm 0.0161$	96.01 $\pm 0.0888$	96.95 $\pm 0.0669$	95.92 $\pm 0.0548$	96.23 $\pm 0.0159$
LFE+FrF	20	93.62 $\pm 0.0177$	97.56 $\pm 0.0169$	94.99 $\pm 0.0754$	98.02 $\pm 0.0958$	93.77 $\pm 0.0952$	98.12 $\pm 0.0224$
W+FrF	62	94.29 $\pm 0.0113$	96.31 $\pm 0.0176$	94.50 $\pm 0.0830$	97.08 $\pm 0.0669$	95.33 $\pm 0.0117$	96.99 $\pm 0.0903$
All	65	96.92 $\pm 0.0143$	98.26 $\pm 0.0142$	97.03 $\pm 0.0607$	98.99 $\pm 0.0055$	96.99 $\pm 0.0159$	98.81 $\pm 0.0362$

**Table 5.3:** Feature evaluations for database 1 using raw EEG and RID. AUC values average and standard deviations correspond to 1000 bootstrap runs.

Dim	L	F	E	LF	FE	LE	LFE	W	FrF	LFE+W	LFE+FrF	W+FrF	All
	1	1	1	2	2	2	3	45	17	48	20	62	65
Patient 1	82.61 ±0.2212	42.01 ±0.8965	46.80 ±0.7485	88.54 ±0.0865	85.33 ±0.0487	86.21 ±0.5454	87.13 ±0.0397	85.07 ±0.2187	85.94 ±0.2856	83.07 ±0.0398	<b>88.64</b> ±0.2231	84.99 ±0.1878	83.35 ±0.0999
Patient 2	87.54 ±0.0951	86.42 ±0.2008	80.03 ±0.2154	85.54 ±0.3652	90.14 ±0.0165	<b>91.92</b> ±0.1010	89.70 ±0.0025	83.66 ±0.1235	83.67 ±0.0258	83.56 ±0.2027	88.10 ±0.0148	78.55 ±0.1550	80.80 ±0.3002
Patient 3	85.77 ±0.0303	69.99 ±0.1896	75.11 ±0.1580	84.94 ±0.2654	87.88 ±0.1778	90.14 ±0.0369	90.12 ±0.0085	88.55 ±0.1791	85.03 ±0.1444	88.01 ±0.0585	<b>90.28</b> ±0.1355	89.07 ±0.0333	88.55 ±0.1111
Patient 4	83.92 ±0.2020	67.15 ±0.0369	87.90 ±0.3020	85.92 ±0.1247	86.58 ±0.1565	91.93 ±0.1058	90.05 ±0.1365	88.41 ±0.0584	90.09 ±0.1990	86.66 ±0.1754	90.04 ±0.1002	90.80 ±0.1452	<b>92.73</b> ±0.1988
Patient 5	86.08 ±0.1511	59.03 ±0.0930	86.60 ±0.1080	83.50 ±0.0505	86.77 ±0.1337	88.42 ±0.0210	<b>89.99</b> ±0.1452	82.09 ±0.1453	84.99 ±0.0225	83.91 ±0.1111	84.44 ±0.1228	84.90 ±0.0054	82.39 ±0.0365
Patient 6	70.21 ±0.0777	35.55 ±0.2001	78.34 ±0.0545	88.28 ±0.0787	87.34 ±0.1012	89.07 ±0.1845	86.48 ±0.1770	86.23 ±0.0666	86.56 ±0.0433	86.75 ±0.1288	84.01 ±0.0182	83.36 ±0.1843	<b>89.16</b> ±0.1121
Average	82.68	60.02	75.79	86.10	97.34	89.61	88.91	85.66	86.04	85.32	87.58	85.27	86.16

**Table 5.4:** Feature evaluations for database 1 using a preprocessed EEG and RID. AUC values average and standard deviations correspond to 1000 bootstrap runs.

Dim	L	F	E	LF	FE	LE	LFE	W	FrF	LFE+W	LFE+FrF	W+FrF	All
	1	1	1	2	2	2	3	45	17	48	20	62	65
Patient 1	85.12 ±0.1083	45.38 ±0.2207	48.57 ±0.2143	89.01 ±0.0980	87.61 ±0.0953	88.81 ±0.0840	90.2 ±0.0920	87.2 ±0.1040	87.7 ±0.1030	87.28 ±0.1040	<b>93.20</b> ±0.1050	87.41 ±0.1030	87.63 ±0.1090
Patient 2	89.82 ±0.1040	90.33 ±0.1006	85.86 ±0.1178	87.02 ±0.1197	94.08 ±0.0109	<b>94.23</b> ±0.0797	93.77 ±0.0180	87.4 ±0.1200	88.24 ±0.1230	87.22 ±0.1240	91.16 ±0.1090	82.21 ±0.1120	86.91 ±0.1220
Patient 3	88.94 ±0.0425	73.29 ±0.0664	79.41 ±0.0564	87.41 ±0.0511	91.85 ±0.0372	93.91 ±0.0360	<b>94.1</b> ±0.0320	91.9 ±0.0420	89.3 ±0.0520	93.5 ±0.0360	93.8 ±0.0300	93.9 ±0.0330	92.8 ±0.0410
Patient 4	85.16 ±0.0586	69.04 ±0.0922	89.32 ±0.0526	87.20 ±0.0631	88.21 ±0.0549	92.23 ±0.0448	90.9 ±0.0500	90.6 ±0.0710	92.1 ±0.0480	88.5 ±0.0720	93.4 ±0.0400	95.3 ±0.0360	<b>97.1</b> ±0.0460
Patient 5	89.04 ±0.0627	62.54 ±0.1415	90.33 ±0.0644	88.38 ±0.0764	90.06 ±0.0648	92.10 ±0.0563	<b>93.36</b> ±0.0530	87.6 ±0.0781	88.2 ±0.0741	88.1 ±0.0800	88.4 ±0.0741	87.8 ±0.0766	87.9 ±0.0758
Patient 6	74.69 ±0.0861	38.54 ±0.1455	80.91 ±0.0698	90.19 ±0.0488	89.57 ±0.0529	91.35 ±0.0446	89.15 ±0.0521	89.11 ±0.0557	89.38 ±0.0545	89.55 ±0.0544	89.21 ±0.0541	89.66 ±0.1249	<b>92.31</b> ±0.1454
Average	85.46	63.19	79.07	88.20	90.23	92.27	91.53	88.97	88.97	89.39	91.52	90.27	90.71

**Table 5.5:**  $F_{score}$  values average and standard deviations correspond to 1000 bootstrap runs using the TFD SPWV.

	L	F	E	LF	FE	LE	LFE	W	FrF	LFE+W	LFE+FrF	W+FrF	All
Dim	1	1	1	2	2	2	3	45	17	48	20	62	65
N1	67.03 $\pm 0.0626$	85.68 $\pm 0.0355$	99.46 $\pm 0.0078$	86.42 $\pm 0.0357$	99.74 $\pm 0.0059$	<b>100</b> $\pm 6.89 \times 10^{-4}$	99.36 $\pm 0.0084$	99.89 $\pm 0.0032$	98.70 $\pm 0.0120$	99.40 $\pm 0.0117$	99.23 $\pm 0.0093$	99.74 $\pm 0.0048$	99.66 $\pm 0.0059$
N2	17.19 $\pm 0.1769$	63.99 $\pm 0.1563$	80.33 $\pm 0.0333$	87.27 $\pm 0.0345$	87.35 $\pm 0.0332$	88.42 $\pm 0.0304$	87.45 $\pm 0.0358$	93.28 $\pm 0.0250$	99.18 $\pm 0.0089$	85.01 $\pm 0.0467$	94.16 $\pm 0.0275$	98.36 $\pm 0.0145$	97.96 $\pm 0.0151$
N3	16.22 $\pm 0.0944$	9.04 $\pm 0.0567$	4.88 $\pm 0.0156$	43.85 $\pm 0.1034$	55.69 $\pm 0.0963$	30.44 $\pm 0.2237$	86.27 $\pm 0.0361$	82.54 $\pm 0.0459$	83.59 $\pm 0.0460$	81.35 $\pm 0.0474$	88.18 $\pm 0.0378$	93.23 $\pm 0.0270$	92.25 $\pm 0.0299$
Average	33.48	52.90	61.55	72.51	80.92	72.95	91.03	91.91	93.82	88.35	93.85	97.11	96.62

**Table 5.6:**  $F_{score}$  values average and standard deviations correspond to 1000 bootstrap runs using the TFD RID.

	L	F	E	LF	FE	LE	LFE	W	FrF	LFE+W	LFE+FrF	W+FrF	All
Dim	1	1	1	2	2	2	3	45	17	48	20	62	65
N1	67.24 $\pm 0.0553$	86.95 $\pm 0.0357$	99.38 $\pm 0.0091$	92.29 $\pm 0.0261$	99.82 $\pm 0.0041$	<b>100</b> $\pm 0$	99.57 $\pm 0.0068$	99.89 $\pm 0.0032$	98.70 $\pm 0.0120$	99.91 $\pm 0.0031$	99.06 $\pm 0.0099$	99.74 $\pm 0.0048$	99.36 $\pm 0.0059$
N2	36.84 $\pm 0.1438$	51.99 $\pm 0.2895$	75.81 $\pm 0.0803$	87.07 $\pm 0.0322$	87.41 $\pm 0.0320$	86.19 $\pm 0.0359$	83.51 $\pm 0.0406$	93.28 $\pm 0.0250$	99.18 $\pm 0.0089$	92.55 $\pm 0.283$	<b>99.74</b> $\pm 0.0055$	98.36 $\pm 0.0145$	98.06 $\pm 0.0303$
N3	20.14 $\pm 0.1019$	5.01 $\pm 0.0139$	9.80 $\pm 0.0748$	15.79 $\pm 0.1204$	67.34 $\pm 0.1040$	64.73 $\pm 0.0841$	82.13 $\pm 0.0449$	82.54 $\pm 0.0459$	83.59 $\pm 0.0460$	81.36 $\pm 0.0434$	89.82 $\pm 0.0303$	93.23 $\pm 0.0270$	<b>96.02</b> $\pm 0.0301$
Average	41.40	47.98	61.66	65.05	84.85	83.04	88.40	91.91	93.82	91.27	96.20	97.11	<b>97.81</b>

**Table 5.7:**  $F_{score}$  values average and standard deviations correspond to 1000 bootstrap runs using the TFD RSPWV.

	L	F	E	LF	FE	LE	LFE	W	FrF	LFE+W	LFE+FrF	W+FrF	All
Dim	1	1	1	2	2	2	3	45	17	48	20	62	65
N1	67.21 $\pm 0.0590$	85.39 $\pm 0.0364$	99.75 $\pm 0.0056$	86.56 $\pm 0.0342$	99.44 $\pm 0.0081$	<b>100</b> $\pm 6.89 \times 10^{-4}$	99.41 $\pm 0.0083$	99.89 $\pm 0.0032$	98.70 $\pm 0.0120$	99.55 $\pm 0.0070$	99.22 $\pm 0.0096$	99.74 $\pm 0.0048$	99.76 $\pm 0.0102$
N2	17.73 $\pm 0.1766$	63.72 $\pm 0.1528$	80.03 $\pm 0.0490$	87.49 $\pm 0.0345$	87.21 $\pm 0.0286$	88.14 $\pm 0.0326$	89.71 $\pm 0.0319$	93.28 $\pm 0.0250$	99.18 $\pm 0.0089$	93.69 $\pm 0.0220$	99.46 $\pm 0.0110$	98.36 $\pm 0.0145$	99.16 $\pm 0.0095$
N3	16.51 $\pm 0.727$	8.64 $\pm 0.0618$	4.12 $\pm 0.1058$	57.95 $\pm 0.0863$	58.9 $\pm 0.1057$	63.19 $\pm 0.2312$	86.09 $\pm 0.0395$	82.54 $\pm 0.0459$	83.59 $\pm 0.0460$	84.21 $\pm 0.0405$	88.47 $\pm 0.0379$	93.23 $\pm 0.0270$	95.56 $\pm 0.0265$
Average	33.81	52.58	61.30	77.33	81.85	83.77	91.73	91.91	93.82	92.48	95.71	97.11	98.16

problem. Moreover, this method could be combined with other features and improve results, as show in Table 5.4 (patients 1, 4 and 5) and Table 5.6 (N2 and N3 problems).

Next subsection discusses the results obtained by the experiments proposed and also analyzes the LFE performance with other extraction methods.

## 5.4 Dimensionality reduction of EEG features matrix by forward-backward algorithm and mutual information (MI)

In EEG classification problems, one might think that expanding the features matrix can improve the classifier performance because we are adding new features or information to the problem. However, it is worth noting that the size of that features matrix can not grow indefinitely and also adding new features does not always improve the performance of the classifier. The reason is simple: there is a possibility that instead of adding new information to the problem, we are introducing redundancy and this could possible be taken as noise. Then, the performance of the classifiers may degrade due to overfitting. Moreover, expanding the features matrix involves increasing the computational cost, therefore it must be clear what features we must add to improve the classifier performance.

In practice, we need to know what features are sufficient and appropriate for specific problems but this is usually not easy to know a priori. It is here when feature selection methods play an important role because these techniques could reach the following goals through reduction of the feature matrix improving both the computational cost and performance in the classifier. This means that we do not choose the most potentially relevant, because the design could be suboptimal or conversely, the subset most useful, because it may exclude the most important relevant characteristics, as discussed in [Guyon et al. \[2006\]](#).

This section evaluates each feature (or feature subsets) using a feature selection algorithm based on forward-backward procedure and mutual information (MI) as relevant criteria. We use databases 1 and 3. Several experiments with database 2 showed that it was necessary to use all features to achieve good performance. Although the N1 problem only needs two features (LE) to solve the classification problem, this problem was also included in feature selection analysis to see if there is possibility of finding other subset of features with same performance (100%). Tables 5.8 to 5.10 show the results on feature selection using distributions SPWV, RID and RSPWV. It should

**Table 5.8:**  $F_{score}$  evaluations achieved by three different MI estimations and SPWV.  $F_{score}$  values average and standard deviations correspond to 1000 bootstrap runs.

Forward-backward selection			
	Kraskov	Parzen	Knn
N1	{E, 1 FrF} 99.63 $\pm$ 0.0053	{L,E} <b>100</b> $\pm$ 5.77 $\times 10^{-4}$	{L,F,E, 12 WC, 6 FrF} 99.75 $\pm$ 0.0054
N2	{L,F,E, 44 WC, 11 FrF} <b>99.37</b> $\pm$ 0.0073	{L,F,E} 87.50 $\pm$ 0.0339	{L,F,E, 15 WC, 9 FrF} 99.2 $\pm$ 0.0082
N3	{3 FrF} 85.35 $\pm$ 0.0363	{L,F,E} <b>86.34</b> $\pm$ 0.0441	{F,E,1 WC, 5 FrF} 85.59 $\pm$ 0.0372

**Table 5.9:**  $F_{score}$  evaluations achieved by three different MI estimations and RID.  $F_{score}$  values average and standard deviations correspond to 1000 bootstrap runs.

Forward-backward selection			
	Kraskov	Parzen	Knn
N1	{E, 1 FrF} 99.77 $\pm$ 0.0013	{L,E} <b>100</b> $\pm$ 0	{F,E, 5 WC, 10 FrF } 99.16 $\pm$ 0.0058
N2	{L,F,E, 1 FrF} <b>100</b> $\pm$ 0.0023	{L,F,E} 83.50 $\pm$ 0.0552	{L,F,E, 2 WC, 5 FrF} 99.79 $\pm$ 0.0023
N3	{3 FrF} 85.45 $\pm$ 0.0116	{L,F,E} 83.34 $\pm$ 0.0721	{L,F,E, 2 WC, 6 FrF} <b>99.59</b> $\pm$ 0.0356

**Table 5.10:**  $F_{score}$  evaluations achieved by three different MI estimations and RSPWV.  $F_{score}$  values average and standard deviations correspond to 1000 bootstrap runs.

Forward-backward selection			
	Kraskov	Parzen	Knn
N1	{E, 1 FrF} 99.77 $\pm$ 0.0403	{L,E} <b>100</b> $\pm$ 6.89 $\times 10^{-4}$	{L,F,E, 12 WC, 6 FrF} 99.85 $\pm$ 0.0494
N2	{L,F,E, 10 WC, 14 FrF} 98.73 $\pm$ 0.0063	{L,F,E} 89.01 $\pm$ 0.0179	{L,F,E, 2 WC, 6 FrF} <b>99.12</b> $\pm$ 0.0053
N3	{3 FrF} 85.29 $\pm$ 0.0753	{L,F,E} 86.04 $\pm$ 0.0775	{F,E,1 WC, 5 FrF} <b>90.08</b> $\pm$ 0.0649

be noted that each method selects different features, but in most cases they all have been successful considering the  $F_{score}$  rate.

The following can be noted from these results:

- For the three distributions the problem N1 is definitely solved with only two features obtained by tracks extraction method (LFE features). Parzen method selected the appropriate features.
- Problem N2 with RID distribution has increased from 99.74% with dimension D=20 (see Table 5.6) to 100% with D=4 (see Table 5.9). Similarly, problem N3 from 96.02% with D=65 (see Table 5.6) to 99.59% with D=11 (see Table 5.9). The selection method that showed better performance was the K-NN.
- Comparing Table 5.4 and Table 5.11 for database 1:

- Patient 1: 91.2% with D=20 to 97.52% with D=5. K-NN method.
- Patient 2: 93.2% with D=3 (LFE) to 91.65% with D=16. Feature selection method did not improve performance.
- Patient 3: 94.1% with D=3 (LFE) to 91.17% with D=12. Feature selection method did not improve performance.
- Patient 4: 97.1% with D=65 (all) to 93.13% with D=15. Feature selection method did not improve performance.
- Patient 5: 93% with D=3 (LFE) to 89.52% with D=4. Feature selection method did not improve performance.
- Patient 6: 92.3% with D=65 (all) to 93.23% with D=16.

The classification results clearly show the good performance of the LE detector in N1 problem (100% for all distributions). They also show an increasing of accuracy with less features in N1 and N2 as well as Patients 1 and 6. results in patients 2, 3 and 5, with high accuracy with LFE features, do not improve after the feature selection procedure. In fact, there was a decrease when one of them was eliminated or when the method added more features. This validates the importance of choosing suitable features in each classification problem.

As a conclusion, it is clear that the tracks extraction method together with the ICA-RLS technique proposed in this Thesis offer a new alternative for EEG classification with epilepsy. On the other hand, these method have shown that it can be used with other extraction methods improving the classifier performance.

## 5.5 Comparison of classification accuracy

Table 5.12 shows a comparison between our technique and other methods proposed in the literature. As the merit figure we used the accuracy rate. The distribution used is the RID. Note that N1 and N2 has reached the maximum accuracy similarly as other methods but using a lower dimension of the feature matrix, D=2 and D=4 respectively, and the best approaches were D=4 for both problems. Something similar happens with N3 that has increased in accuracy from 99.28% to 99.59% but with a noticeable dimension reduction (D=11) compared with the best proposal in the literature, which was D=24.

**Table 5.11:** Feature selection using three different MI estimations. AUC values average and standard deviations correspond to 1000 bootstrap runs.

	Forward-backward selection		
	Kraskov	Parzen	Knn
Patient 1	{F,E,10 W,1 FrF} 86.84 ± 0.1073	{L,F,E,13 FrFT} 88.59 ± 0.0968	{ E,2 W,2 FrF} <b>97.52</b> ± 0.0991
Patient 2	{F,5 W, 4 FrF} 91.29 ± 0.0989	{L,F,E, 13 FrF} <b>91.65</b> ± 0.0894	{ 4 W} 88.27 ± 0.0114
Patient 3	{F,E,10 W} <b>91.17</b> ± 0.0039	{L,F,E,12 FrF } 90.54 ± 0.0465	{ 4 FrF} 89.74 ± 0.0521
Patient 4	{1 FrF} 86.15 ± 0.058	{L,F,E,12 FrF} <b>93.13</b> ± 0.040	{5 FrF} 90.77 ± 0.0456
Patient 5	{E,14 W,2 FrF} 87.19 ± 0.0916	{L,F,E,1 FrF} <b>89.52</b> ± 0.0765	{ 3 W,7 FrF} 88.70 ± 0.0725
Patient 6	{L,9 W} 88.22 ± 0.0553	{L,E,F,1 W,12 FrF} <b>93.23</b> ± 0.048	{E,4 FrF} 89.16 ± 0.0530

## 5.6 Summary and conclusions

In this chapter we have shown the classification performance of ICA-RLS filter and tracks extraction method (LFE features). Also it was illustrated the validation of some parameters included in the algorithms such as  $\Delta$ ,  $\lambda$ , TFD, overlapping and EEG epoch. Several experiments have been conducted using clinical datasets from real patients as well as other datasets. Feature selection and comparison with other results obtained from other authors are also presented. Results show that ICA-RLS and tracks extraction method are suitable techniques to EEG processing and more specifically to epilepsy detection and EEG segment classification.

Finally, we have to emphasize: RLS-ICA is an efficient method to eliminate EOG artifacts produced by eyes movements but has a high computational cost:  $\mathcal{O}(N^4) + \mathcal{O}(7N^2)$ . This cost could be optimized using other ICA algorithm or reducing the number of reference channels. However, the use of this algorithm could be restricted just in patients with hard epilepsy detection because the extra time consuming of this method and its improvement in classification performance does not justify its use in all patients. Moreover, if we compare the time spent by an expert manually eliminating the artifacts compared to computational cost of ICA-RLS method, obviously we will choose the second option for the reasons discussed in Chapter 4.

In our medical environment, the steps that we follow in a classification problem are: (i) denoising and artifacts removal, (ii) LFE features extractions, (iii) detection/classification using these features and their combina-

tions, (iv) if we do not have conclusive results, we add features from wavelets and fractional Fourier transform, (v) detection/classification using these features and their combinations, (vi) we apply dimensionality reduction if necessary.

**Table 5.12:** Comparison of classification accuracy (in percent) obtained by LFE approach for epileptic seizure detection.

Authors	Method	Problem	Dimension	Accuracy
Nigam and Graupe [2004]	Non-linear pre-processing filter-Diagnostic neural network	N1	2	97.2
Srinivasan et al. [2005]	Time & frequency domain features-Recurrent neural network	N1	5	99.6
Kannathal et al. [2005b]	Entropy measures-Adaptive neuro-fuzzy inference system	N1	4	92.2
Kannathal et al. [2005a]	Chaotic measures-Surrogate data analysis	N1	4	≈90
Polat and Günes [2007]	Fast Fourier transform-Decision tree	N1	129	98.72
Subasi [2007]	Discrete wavelet transform-Mixture of expert model	N1	4	95
Chua et al. [2010]	High order spectra (HOS)-Gaussian mixture model and Support vector machine	N1	6	93.11
Liang et al. [2010]	Spectral and Entropy analysis-Linear and Non-linear classifiers	N1	16	98.51
Wongsawat [2008]	Phase congruency-Linear discriminant analysis	N1	1	99
Tzallas et al. [2009b]	Time & frequency analysis-Artificial neural network	N1	4	100
Guerrero-Mosquera et al. [2010a]	<b>LFE-Tracks extractions-Support vector machine</b>	<b>N1</b>	<b>2</b>	<b>100</b>
Güler et al. [2005]	Lyapunov exponents-Recurrent neural network	N2	4	96.79
Sadati et al. [2006]	Discrete wavelets transform-Adaptive neural fuzzy network	N2	6	85.9
Liang et al. [2010]	Spectral and Entropy analysis-Linear and Non-linear classifiers	N2	16	98.67
Wongsawat [2008]	Phase congruency-Linear discriminant analysis	N2	1	96.5
Tzallas et al. [2009b]	Time & frequency analysis-Artificial neural network	N2	4	100
Guerrero-Mosquera et al. [2011]	<b>LFE Tracks extractions-Support vector machine and feature selection</b>	<b>N2</b>	<b>4</b>	<b>100</b>
Güler and Übeyli [2005]	Wavelet transform-Adaptive neuro-fuzzy inference system	N3	20	98.68
Güler and Übeyli [2007]	Wavelet transform-Support vector machine	N3	24	99.28
Übeyli and Güler [2007]	Eigenvector method-Modified of mixture of expert model	N3	12	98.60
Liang et al. [2010]	Spectral and Entropy analysis-Linear and Non-linear classifiers	N3	16	85.9
Wongsawat [2008]	Phase congruency-Linear discriminant analysis	N3	1	91
Tzallas et al. [2009b]	Time & frequency analysis-Artificial neural network	N3	4	89
Guerrero-Mosquera et al. [2011]	<b>LFE Tracks extractions-Support vector machine and feature selection</b>	<b>N3</b>	<b>11</b>	<b>99.59</b>



# Chapter 6

## Conclusions and future work

In this thesis we have developed two proposals for EEG signal processing in order to improve the performance in different scenarios such as detection and classification of epileptic signals. The first approach consists in a combination of ICA and RLS algorithms which automatically cancels the artifacts produced by eyes movement without the use of external signals such as electrooculograms (EOG). This method, called ICA-RLS has been compared with other techniques that are in the state of the art and it has shown to be a good alternative for artifacts rejection. The second approach is a novel method in EEG features extraction called tracks extraction (LFE features). This method is based on the time-frequency distributions (TFDs) and partial tracking. Our results in pattern extractions related to epileptic seizures have shown that tracks extraction is appropriate in EEG detection and classification tasks, being practical, easily applicable in medical environment and has acceptable computational cost. Also, the extracted features have important information that can be used in the diagnosis of epilepsy.

These methods have shown improvements in classification problems in the EEG databases provided by Hospital of Navarra and in other datasets proposed in the state of the art. Moreover they could also be considered for other applications based on EEG signals, as well as detection of other neurodegenerative diseases.

In preceding chapters we have shown the significant advantages that our methods have shown in epileptic detection and artifact rejection. In the following we will summarize the main contributions of this thesis, presenting the advantages and disadvantages of the proposed methods, and conclude with possible lines of work that this thesis has generated.

## 6.1 Contributions

- **ICA-RLS EOG removal.** In Chapter 4 we introduced a new method which exploits the capacity of ICA algorithm in discrimination between brain waves and noise. Eye movements are reflected in EEG and its elimination is an important problem. Although there are alternative solutions, most of them present practical disadvantages, for example: methods based on ICA need a visual inspection of the source waveforms to determine if a source is related or not with the eye movement. On the other hand, an artifact rejection design based on adaptive filtering typically makes use of the EOG signal as a reference signal which is not always available in the majority of hospitals. In summary, a visual inspection requires a knowledge of an expert, and for recording an EOG signal it is necessary additional hardware. From these points, the proposed method introduces an important change in the state of the art of artifact rejection.

The method eliminates the source located at the frontal area in the brain and that presents the minimum squared error (MSE). The source with the lowest error and located at the frontal area of the brain is removed. **This makes it not necessary to know the different EOG waveforms** recorded in the EEG caused by eye movements and implies that **it does not requires the use or registration of electrooculograms (EOG) signal.**

The results of ICA-RLS method has proved to be efficient both in EEG strongly contaminated by ocular artifacts and when there is low presence of artifacts. The main drawback is the computational cost which was discussed in Section 5.6. This computational cost could be reduced with further optimization or the use of the algorithm itself could be restricted just in patients with hard epilepsy detection. Testing other ICA methods proposed in the literature is proposed as future work.

- **Tracks extraction (LFE).** A new feature extraction method has been presented for EEG signals. Following on the time-frequency plane a principal track, the novel method extracts three features from this track: energy, frequency and its duration. Chapter 4 showed the effectiveness of the method in epilepsy detection. Moreover, Chapter 5 showed the good performance in classification for EEG epileptic segments. This chapter also explained the possibility of combining LFE features with other features proposed in the state of art such as wavelet transforms and Fractional Fourier, with the goal of improving the classification performance. With the proposed method we have the following advantages:

- **Adaptability of the method to any TFD.** Many applications use specific TFD to solve detection problems. For example, for knee injury diagnostic the Pseudo Wigner-Ville (PWV) distribution has been used, and underwater acoustic signal applications use the WV distribution. Tracks extraction is based on peak tracking on the time-frequency plane without care of the distribution being used. For that reason, the method may extend to other applications because it has the capability of extracting new features and could improve the task performance.
- **Problem solving classification using fewer features.** Tracks extraction method solves different problems in EEG detection and classification based on only three features extracted from the main track. Chapter 5 showed results where LFE features have solved the classification problem (N1 problem and patients 2, 3 and 5) with the best performance. Moreover, more complex problems such as N2 and N3, after selecting procedure the LFE features were included in the relevant subset which solve the classification problem in an optimum way.
- **Not very high computational cost for EEG classification.** Given that the method extracts three features, resulting in an array of features with a dimension  $D = 3$ , all the computational cost into a classification scheme is low. Other methods such as Fourier transform (FT), wavelets or fractional FT (FrFT) usually use more features ( $> 10$ ) for such problems.
- **Possibility of combining LFE features with other features to improve the classification performance in EEG signals.** EEG signals are very dynamic, especially in cases when there are energy bursts of short duration or rapid frequency changes. Then, there are certain problems where LFE method cannot detect these events efficiently, being necessary to add more information from other methods that could be more appropriate for this type of scenario. Wavelets and FrFT could improve the tracks extraction performance because both methods carry out a multilevel analysis. In Chapter 5 it was observed that tracks extraction method provides relevant information for EEG classification and it improves the classifier performance when it is combined with other features.
- **Stability.** Chapter 5 showed that parameters such as  $\Delta$ , EEG epochs, overlapping and windowing, have a good performance in a wide range of operative values. This implies that tracks extraction method presents a good stability against any change of these parameters.

- **Dimensionality reduction using mutual information (MI) and forward-backward procedure (FB).** Chapter 3 discussed the applicability of these methods to EEG signal processing. This PhD thesis has demonstrated the effectiveness of this method in EEG dimensional reduction for EEG classification. Two observations may be derived from these experiments:
  - **LFE features really provide important information in EEG classification problems.** Following the features selection methods described in Chapter 3, the results in this thesis have shown that tracks extractions provide three important features that can solve classification problems. We have seen in Chapter 5 some classification problems solved just using one, two or three features depending on problem complexity. So there is no colinearity among LFE features which means we are avoiding the overfitting in classification.
  - **Feature selection method based on MI and FB procedure are good alternatives for EEG dimensional reduction.** As seen in Chapter 5, many classification problems proposed in this thesis have shown considerable dimensional reduction and also an improvement in the classifier performance.

In summary, this thesis provides a set of new methods to EEG signal theory that can be used in EEG detection or classification. These methods are flexible, simple, and could be successfully extended favorably to other applications.

## 6.2 Future work

Beside the main contributions that ICA-RLS and tracks extraction methods offer to EEG signal processing, we present several extensions or possible changes that might to improve their shortcomings such as computational cost, or extend their applications range. For example:

- **Extension of ICA-RLS method to other artifacts.** Apart from artifacts produced by eyes movement, there are other artifacts which contaminate the EEG signal such as muscle movements or electrodes displacement. This extension implies to work in different frequency bands and in turn selecting the input signals which are to be used in the adaptive scheme.
- **Full automation of ICA-RLS method.** As discussed in Chapter 4, the proposed scheme makes use of topographic maps of the brain

(visually) and MSE value to select the corresponding source to rejected. EOG artifacts have an important energy contribution on frontal zone of the brain. This energy is frequently registered at electrodes closely to eyes. In EEG data, ICA identifies scalp maps nearly fitting the projection of a single equivalent current dipole, and is therefore quite compatible with the projection to the scalp electrodes of synchronous local field activity. To do that we need: (i) build an EEG dataset with channel locations, (ii) fit dipole models to ICA components, (iii) obtain for each component its equivalent dipole and its coordinates and finally, (iv) calculate distances to determine which source is close to the eyes. An initial work in fitting dipole models to ICA sources is described in [Delorme and Makeig \[2004\]](#).

- **Stopping criterion in Forward-Backward (FB) algorithm.** Both forward and backward procedure must be given a stopping criterion. In backward search is quite easily achieved. The purpose being to select the most informative set of variables. On the other hand, forward procedure is more difficult to stop. Indeed, it may be considered as a ranking algorithm (the variables are ordered according to their mutual information with the output) rather than a selection one, as described in [Rossi et al. \[2006\]](#). Moreover, combining the Mutual Information (MI) criterion with a forward feature selection strategy offers a good trade-off between optimality of the selected feature subset and computation time. However, it requires to set the parameter(s) of the mutual information estimator and to determine when to halt the forward procedure. To solve these problems, [Françoise et al. \[2007\]](#) proposed to use resampling methods, a  $K$ -fold cross-validation and the permutation test. There are more alternatives in the literature and all standard search methods such as simulated annealing and genetic algorithms that could be used to find relevant subsets. Additionally, there exist non-greedy methods that maximize the MI between output features and the class labels (e.g. [Hild et al. \[2006\]](#)) which directly will change the stopping criterion in the algorithm. Future work implies a deeply analysis of these methods to compare with the FB algorithm based on MI in EEG signals.
- **Extension of tracks extraction to other scenarios such as Brain-Computer-Interface (BCI).** Detecting event related potentials (ERPs) is a fundamental task in BCI applications. Usually, the ERP waveform is quantitatively characterized by amplitude, latency and scalp distribution. TFDs offer an new alternative of signal analysis and tracks extractions gives us relevant information by extracting LFE features from time-frequency plane. For this reason, we think tracks extraction could plays an interesting role in detecting ERPs.
- **Classification of different types of epilepsy.** There are many types

of epilepsy. A simple way of classifying them is based on the origin of the crisis in the first fraction of a second. If the crisis begins in a focal region of the brain it is called *partial*, if it starts all over the brain at the same time it is called *generalized*. Within each group there are different types of epilepsy with different characteristics between them. An analysis of each epilepsy may be based on the tracks extraction method. This method would yield relevant information for each epilepsy such as frequency bands, duration of crisis and active brain areas in the crisis. This information could be used for epilepsy classification and then support for possible treatment, diagnosis or surgery.

- **Monitoring of treatments.** Principal track analysis on the time-frequency plane for patients with epilepsy and under epileptic medications could be considered as a new tool for medical supporting. The principal track can be used to monitor the effect of drugs on patients, assuming that this track tends to disappear with the treatment.
- **Anticipation of epileptic seizures.** People with epilepsy have problems in their daily lives that affect their family, work and personal relations, resulting in low self-esteem. This problem causes insecurity because they do not know when a crisis may occur. A warning system (intracranial chip, system of self-medication) may be very helpful because this system could significantly improve their quality of life.

An additional potential field of research is the **EEG Integration with other techniques such as fMRI**. Chapter 2 described the advantages and disadvantages of using EEG signals. The principal drawback of the EEG is its low spatial resolution because it depends on the number of electrodes. MEG has a better temporal resolution than EEG but suffers the same disadvantage. fMRI solves this problem and its spatial resolution is on the order of millimeters. The integration of these techniques is of vital importance in neuroscience studies because this could improve the detection of other neurodegenerative diseases like Alzheimer, Parkinson, depression or dementia.

# Appendix A

## Filtering concepts

The definitions showed in this Appendix can be found in most adaptive filter textbooks, e.g. [Haykin \[1996\]](#).

### A.1 Stochastic process

Random signals can only be characterized by probabilistic models or average performance terms. The mathematic tools that allows us to characterize these signals are stochastic processes. Examples of a stochastic processes include biomedical signals, radar signals and noise.

#### A.1.1 Statistics of stochastic process

A stochastic process is a noncountable infinity of random variables, one for each  $t$ . For a specific  $t$ ,  $\mathbf{x}(t)$  is a random variable with distribution

$$F(x, t) = P\{\mathbf{x}(t) \leq x\}$$

This function depends on  $t$ , and it equals the probability of the event  $\{\mathbf{x}(t) \leq x\}$  consisting of all outcomes  $y$  such that, at the specific time  $t$ , the samples  $\mathbf{x}(t, y)$  of the given process do not exceed the number  $x$ . The function  $F(x, t)$  will be called the *probability distribution* of the process  $\mathbf{x}(t)$ . Its derivate with respect to  $x$ :

$$f(x, t) = \frac{\partial F(x, t)}{\partial x}$$

is the probability density of  $\mathbf{x}(t)$ . For the determination of the statistical properties of a stochastic process could be used only certain averages. These quantities are defined as follows:

- Mean:  $\mu_x(t) = E\{\mathbf{x}(t)\}$ , where  $E\{\cdot\}$  is the expected value of the random variable  $\mathbf{x}(t)$
- Variance:  $\sigma_x(t) = E\{(\mathbf{x}(t) - \mu_x(t))^2\}$
- Autocorrelation:  $R_x(t_1, t_2) = E\{\mathbf{x}(t_1)\mathbf{x}(t_2)\}$
- Autocovariance:  $C_x(t_1, t_2) = R_x(t_1, t_2) - \mu_x(t_1)\mu_x(t_2)$

The autocovariance and autocorrelation can be generalized to the case of two stochastic processes,  $\mathbf{x}(t)$ ,  $\mathbf{y}(t)$ , and measure their relationship.

- Cross-correlation:  $R_{xy}(t_1, t_2) = E\{\mathbf{x}(t_1)\mathbf{y}^*(t_2)\}$ , where  $y^*$  denotes the complex conjugate of  $y$ .
- Cross-covariance:  $C_{xy}(t_1, t_2) = R_{xy}(t_1, t_2) - \mu_x(t_1)\mu_y^*(t_2)$

### A.1.2 Stationary process

A stochastic process  $\mathbf{x}(t)$  is stationary in the strict sense if its statistical properties are invariant to a shift of the origin. This means that the process  $x(t)$  and  $x(t + c)$  have the same statistics for any integer  $c$ .

The autocorrelation of a stationary stochastic process has the following properties:

- symmetry pair:  $R_x(t) = R_x(-t)$
- $R_x(0) = E\{x(t)^2\}$
- $\forall k, R_x(k) \leq R_x(0)$

## A.2 The filtering problem

Filters are systems that can be described by a linear equation in differences that relate filter input signal  $x(t)$  and output signal  $y(t)$ :

$$\sum_{k=0}^N a_k y[t - k] = \sum_{\ell=0}^L b_\ell x[t - \ell]$$

In the filtering problem we can identify the following elements:

- Desired signal  $d(n)$ . We want to design a filter that produces an estimate  $\hat{d}(n)$  of the desired signal  $d(n)$  using a linear combination of the data  $x(n)$  such that the mean-square error (MSE) function  $E\{(d(n) - \hat{d}(n))^2\}$ , is minimized.

- Input signal  $x(n)$  which is related to  $d(n)$ .
- Impulse response of the filter  $w(n)$ .
- Measurement error  $e(n) = d(n) - (x(n) * w(n))$ .

The error signal  $e(n)$  is then used to form a performance (or objective) function that is required by the adaptation algorithm in order to determine the appropriate updating of the filter coefficients. The minimization of the objective function implies that the adaptive-filter output signal is matching the desired signal in some sense.

The complete specification of an adaptive system consists of three items:

- **Application:** The type of application is defined by the choice of the signals acquired from the environment to be the input and desired-output signals. The number of different applications in which adaptive techniques are being successfully used has increased enormously during the last two decades. For example: echo cancellation, equalization of dispersive channels, system identification, signal enhancement, adaptive beamforming, noise cancelling, and control.
- **Adaptive-Filter Structure:** The adaptive filter can be implemented in a number of different structures or realizations. The choice of the structure can influence the computational complexity (amount of arithmetic operations per iteration) of the process and also the necessary number of iterations to achieve a desired performance level. Basically, there are two major classes of adaptive digital filter realizations, distinguished by the form of the impulse response, namely the finite-duration impulse response (FIR) filter and the infinite-duration impulse response (IIR) filters.
- **Algorithm:** The algorithm is the procedure used to adjust the adaptive filter coefficients in order to minimize a prescribed criterion. The algorithm is determined by defining the search method (or minimization algorithm), the objective function, and the error signal nature. The choice of the algorithm determines several crucial aspects of the overall adaptive process, such as existence of sub-optimal solutions, biased optimal solution, and computational complexity.

### A.3 The matrix inversion lemma

The matrix inverse lemma is an efficient method for computing the least square solution.

Let  $\mathbf{A}$  and  $\mathbf{B}$  be two positive definite  $M$ -by- $M$  matrices related by

$$\mathbf{A} = \mathbf{B}^{-1} + \mathbf{C}\mathbf{D}^{-1}\mathbf{C}^H \quad (\text{A.1})$$

where  $\mathbf{D}$  is a positive definite  $N$ -by- $N$  matrix and  $\mathbf{C}$  is an  $M$ -by- $N$  matrix. Then,

$$\mathbf{A}^{-1} = \mathbf{B} - \mathbf{B}\mathbf{C}(\mathbf{D} + \mathbf{C}^H\mathbf{B}\mathbf{C})^{-1}\mathbf{C}^H\mathbf{B} \quad (\text{A.2})$$

The proof is obtained by multiplying Eq.A.1 by Eq.A.2, i.e.,  $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ . The matrix inversion lemma is also known as *Woodbury's identity* and states that for a matrix  $\mathbf{A}$  defined as Eq.A.1, its inverse can be calculated by Eq.A.2.

# Appendix B

## Statistical pattern concepts

### B.1 A brief introduction to Machine Learning

The Machine Learning field evolved from the broad field of Artificial Intelligence, which aims to mimic intelligent abilities of humans by machines<sup>1</sup>. In the field of Machine Learning one considers the important question of how to make machines able to “learn”. Learning in this context is understood as *inductive inference*, where one observes *examples* that represent incomplete information about some “statistical phenomenon”. The term *learning machines* encompasses many kinds of computational intelligence systems capable of gathering knowledge by means of data analysis. Learning from data may be treated as searching for the most adequate model (hypothesis) describing a given data. A learning machine is an algorithm which determines a learning model, which can be seen as a function:

$$f : \mathcal{X} \rightarrow \mathcal{Y} \tag{B.1}$$

The function transforms objects from the data domain  $\mathcal{X}$  to the set  $\mathcal{Y}$  of possible target values. The data domain and the set of target values are determined by the definition of the problem for which the  $f$  is constructed.

The learning model  $f$  usually depends on some *adaptive parameters* sometimes also called *free parameters*. In this context, learning can be seen as a process in which a learning algorithm searches for parameters of the model  $f$ , which solve a given task.

The learning algorithm learns from a sequence  $\mathcal{D}$  of data, defined in the

---

<sup>1</sup>This section is based on Rättsch [2004], Hsu et al. [2007] and Guyon et al. [2006]

space  $X$  or in  $X \times Y$ :

$$\mathcal{D} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\} = X \quad (\text{B.2})$$

$$\mathcal{D} = \{\langle \mathbf{x}_1, y_1 \rangle, \langle \mathbf{x}_2, y_2 \rangle, \dots, \langle \mathbf{x}_m, y_m \rangle\} = \langle X, Y \rangle \quad (\text{B.3})$$

Usually  $X$  (called as “pattern”) has a form of a sequence of multidimensional vectors. A pattern is described by its *features*  $\mathbf{x}_i$ . These are the characteristics of the examples for a given problem. For instance, in a face recognition task some features could be the color of the eyes or the distance between the eyes. Thus, the input to a pattern recognition task can be viewed as a two-dimensional matrix, whose axes are the examples and the features.

In *unsupervised learning* (Eq.B.2) one typically tries to uncover hidden regularities (e.g. clusters) or to detect anomalies in the data (for instance some unusual machine function or a network intrusion). Unsupervised learning is used for example in clustering, self-organization, auto-association and some visualization algorithms. In *supervised learning* (Eq.B.3), there is a label associated with each example. This means the learning algorithms use pairs  $\langle \mathbf{x}_i, y_i \rangle$  where  $y_i$  is the desired output value for  $\mathbf{x}_i$ . It is supposed to be the answer to a question about the example. If the label is discrete, then the task is called *classification problem*- otherwise, for real valued labels we speak of a regression problem. Based on these examples (including the labels), one is particularly interested to predict the answer for other cases before they are explicitly observed. Hence, learning is not only a question of remembering but also of generalization to unseen cases.

Pattern classification tasks are often divided into several sub-tasks:

- Data collection and representation.
- Feature selection and/or dimensionality reduction.
- Classification

## B.2 Classification algorithms

Although Machine Learning is a relatively young field of research, there exist more learning algorithms than we can briefly mention in this section.

### B.2.1 Traditional techniques

***k*-Nearest Neighbor (k-NN) Classification.** Here the  $k$  points of the training data closest to the test point are found, and a label is given to the

test point by a majority vote between the  $k$  points. This method is highly intuitive and attains- given its simplicity- remarkably low classification errors, but it is computationally expensive and requires a large memory to store the training data.

**Linear Discriminant Analysis.** This method computes a hyperplane in the input space that minimizes the within-class variance and maximizes the between class distance. It can be efficiently computed in the linear case even with large data sets. However, often a linear separation is not sufficient. Nonlinear extensions by using kernels exist, however, making it difficult to apply it to problems with large training sets.

**Decision Trees.** This is another intuitive class of classification algorithms. These algorithms solve the classification problem by repeatedly partitioning the input space, so as to build a tree whose nodes are as pure as possible (that is, they contain points of a single class). Classification of a new test point is achieved by moving from top to bottom along the branches of the tree, starting from the root node, until a terminal node is reached. Decision trees are simple yet effective classification schemes for small datasets.

**Neural Networks (NN).** NN are perhaps one of the most commonly used approaches to classification. Neural networks are a computational model inspired by the connectivity of neurons in animate nervous systems. A further boost to their popularity came with the proof that they can approximate any function mapping via the Universal Approximation Theorem, in [Haykin \[1999\]](#).

## B.2.2 Large margin algorithms

Machine learning rests upon the theoretical foundation of Statistical Learning Theory, e.g. [Vapnik \[2000\]](#), which provides conditions and guarantees for good generalization of learning algorithms. Within the last decade, large margin classification techniques have emerged as a practical result of the theory of generalization. Roughly speaking, the margin is the distance of the example to the separation boundary and a large margin classifier generates decision boundaries with large margins to almost all training examples.

**Support Vector Machines (SVMs).** This method works by mapping the training data into a feature space by the aid of a so-called “kernel function” and then separating the data using a large margin hyperplane. Intuitively, the kernel computes a similarity between two given examples. Most commonly used kernel functions are Radial Basis Function (RBF) kernels. The SVM finds a large margin separation between the training examples and pre-

viously unseen examples will often be close to the training examples. Hence, the large margin then ensures that these examples are correctly classified as well, i.e., the decision rule *generalizes*. For so-called *positive definite* kernels, the optimization problem can be solved efficiently and SVMs have an interpretation as a hyperplane separation in a high dimensional feature space.

**Boosting.** The basic idea of boosting and ensemble learning algorithms in general is to iteratively combine relatively simple base hypotheses- sometimes called rules of thumb- for the final prediction. One uses a so-called base learner that generates the base hypotheses. In boosting the base hypotheses are linearly combined. In the case of two-class classification, the final prediction is the weighted majority of the votes. The combination of these simple rules can boost the performance drastically. It has been shown that Boosting has strong ties to support vector machines and large margin classification. Boosting techniques have been used on very high dimensional data sets and can quite easily deal with than hundred thousands of examples.

## B.3 Model selection

To determine the generalization ability of a model one would need to measure the average risk for the set of all possible data objects. When doing this we should be aware of the danger of testing models on a single test set (for example resulting from a rigid partition of the set of all available data to the training and test parts). Model selection based on testing trained models on a single test set does not get rid of the danger of overfitting.

A more accurate estimation of the empirical risk can be obtained with  $K$ -fold cross-validation (CV). In this technique we split the set of available data into  $n$  parts and perform  $n$  training and test processes (each time the test set is one of the parts and the training set consists of the rest of the data). The average test risk can be a good estimate of real generalization ability of the tested algorithm, especially when the whole cross-validation is performed several times (each time with different data split) and  $n$  is appropriately chosen. To get a good estimate of generalization ability of a learning machine, it is important to analyze not only the average test error, but also its variance, which can be seen as a measure of stability.

### B.3.1 Model selection for RBF kernels

The RBF kernel nonlinearly maps samples into a higher dimensional space, so it, unlike the linear kernel, can handle the case when the relation between class labels and attributes is nonlinear. Although the linear kernel is a special

case of RBF kernel, there are some situations where the RBF kernel is not suitable. In particular, when the number of features is very large. For this case, one may just use the linear kernel.

There are two parameters while using RBF kernels:  $C$  and  $\sigma$ . It is not known beforehand which  $C$  and  $\sigma$  are the best for one problem; consequently some kind of model selection (parameter search) must be done. The goal is to identify good  $(C, \sigma)$  so that the classifier can accurately predict unknown data (i.e., testing data). Note that it may not be useful to achieve high training accuracy (i.e., classifiers accurately predict training data whose class labels are indeed known). Therefore, a common way is to separate training data to two parts of which one is considered unknown in training the classifier. Then the prediction accuracy on this set can more precisely reflect the performance on classifying unknown data.

In  $K$ -fold cross-validation, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified. Generally it is recommended a “grid-search” on  $C$  and  $\sigma$ , that means pairs of  $(C, \sigma)$  are tried and the one with the best cross-validation accuracy is picked. Researchers have found that trying exponentially growing sequences of  $C$  and  $\sigma$  is a practical method to identify good parameters (for example,  $C = 2^{-5}, 2^{-3}, \dots, 2^{15}, \sigma = 2^{-15}, 2^{-13}, \dots, 2^3$ ).

## B.4 Bootstrap resampling

As described [Zoubir and Iskander \[2004\]](#), the bootstrap is a resampling method for statistical inference. It is commonly used to estimate confidence intervals, but it can also be used to estimate bias and variance of an estimator or calibrate hypothesis tests. The principle of resampling could be summarized as follows: let  $\mathcal{X} = \{X_1, X_2, \dots, X_n\}$  be a sample, i.e., a collection of  $n$  numbers drawn at random from a completely unspecified distribution  $F$ . When we say “at random” we mean that  $X_i$ 's are *independent and identically distributed* (iid) random variables, each having distribution  $F$ . Let  $\theta$  denote an unknown characteristic of  $F$ . It could be the mean or variance of  $F$  or even the spectral density function. The problem we wish to solve is to find the distribution of  $\hat{\theta}$ , an estimator of  $\theta$ , derived from the sample  $\mathcal{X}$ . This is of great practical importance as we need to infer  $\theta$  based on  $\hat{\theta}$ .

One way to obtain the distribution of  $\hat{\theta}$  is to repeat the experiment a sufficient number of times and approximate the distribution of  $\hat{\theta}$  by the so obtained empirical distribution. This is equivalent to Monte Carlo simulations. In many practical situations, however, this is inapplicable for cost reasons or because the experimental conditions are not reproducible.

The bootstrap suggests that we resample from a distribution chosen to be

close to  $F$  in some sense. This could be the sample (or empirical) distribution  $\hat{F}$ , which is that probability measure that assigns to a set  $A$  in the sample space of  $X$  a measure equal to the proportion of sample values that lie in  $A$ . It is known that under some mild assumptions  $\hat{F}$  approaches  $F$  as  $n \rightarrow \infty$ . The principle of the non-parametric bootstrap is based on next steps:

- Conduct the experiment to obtain the random sample

$$\mathcal{X} = \{X_1, X_2, \dots, X_n\}$$

and calculate the estimate  $\hat{\theta}$  from the sample  $\mathcal{X}$ .

- Construct the empirical distribution  $\hat{F}$ , which puts equal mass  $1/n$  at each observation

$$X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$$

- From  $\hat{F}$ , draw a sample

$$\mathcal{X}^* = \{X_1^*, X_2^*, \dots, X_n^*\}$$

called the bootstrap resample.

- Approximate the distribution of  $\hat{\theta}$  by the distribution of  $\hat{\theta}^*$  derived from the bootstrap resample  $\mathcal{X}^*$

With the non-parametric bootstrap, we simply use the random sample  $\mathcal{X} = \{X_1, X_2, \dots, X_n\}$  and generate a new sample by sampling with replacement from  $\mathcal{X}$ . Herein, we create a number  $B$  of resamples  $\mathcal{X}_1^*, \dots, \mathcal{X}_B^*$ . A resample  $\mathcal{X}^* = \{X_1^*, X_2^*, \dots, X_n^*\}$  is an unordered collection of  $n$  sample points drawn randomly from  $\mathcal{X}$  with replacement, so that each  $X_i^*$  has probability  $n^{-1}$  of being equal to any of the  $X_j$ 's. In other terms,

$$\text{Prob}[X_i^* = X_j | \mathcal{X}] = n^{-1}, \quad 1 \leq i, j \leq n$$

That is, the  $X_i^*$  are independent and identically distributed, conditional on the random sample  $\mathcal{X}$ , with this distribution. This means that  $\mathcal{X}^*$  is likely to contain repeats.

### B.4.1 Confidence intervals by percentile bootstrap

The bootstrap distribution  $\hat{F}$  is used to estimate bias, estimate a standard error (SE) or construct a confidence interval for the statistic of interest. The

bootstrap estimates of bias,  $B_b$ , and SE,  $s_b$ , are the empirical estimates calculated from  $m$  bootstrap values:

$$B_b = \frac{\sum_{i=1}^m (\hat{\theta}_i^* - \hat{\theta})}{m}$$

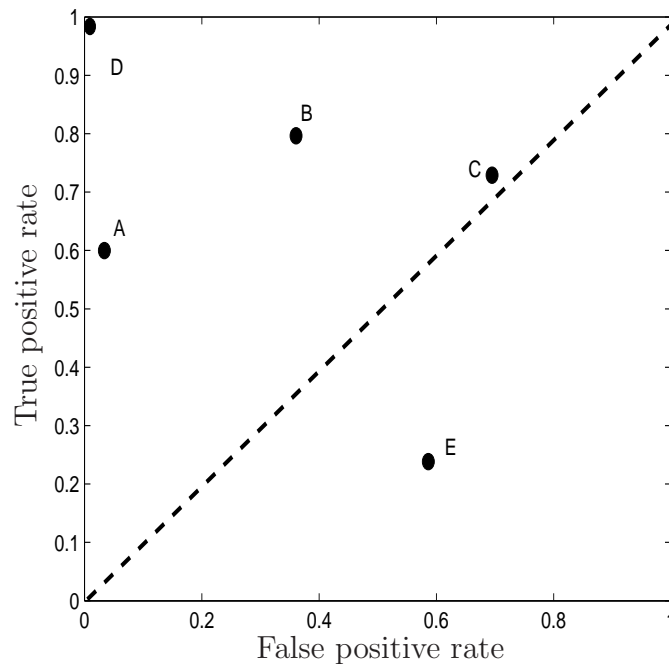
$$s_b = \left[ \frac{\sum_{i=1}^m \left( \hat{\theta}_i^* - \text{mean}(\hat{\theta}) \right)^2}{m - 1} \right]^{1/2}$$

The percentile confidence interval method uses the  $\alpha/2$  and  $-\alpha/2$  quantiles of  $\hat{F}$  as a  $1 - \alpha$  level confidence interval for the parameter. A 95% confidence interval for the mean could be estimated using 1000 bootstrap samples. Each bootstrap sample is a simple random sample selected with replacement from the original observations. Because a bootstrap sample is drawn with replacement, some of the original observations are repeated more than once in the bootstrap sample. The statistic is estimated for each bootstrap sample. Bootstrap confidence intervals can be computed from the set of bootstrap values in a variety of ways. For example, the percentile bootstrap confidence, where the endpoints of the 95% confidence interval are given by the 25th and 975th sorted bootstrap values. Confidence interval coverage is the probability that the confidence interval includes the true parameter, under repeated sampling from the same underlying population. When the coverage is the same as the stated size of the confidence interval (e.g. coverage = 95% for a 95% confidence interval), the intervals are accurate. Empirical and theoretical studies of coverage have shown that the percentile interval is accurate in many situations, as described [Zoubir and Iskander \[2004\]](#).

## B.5 Receiver Operating Characteristics (ROC)

A receiver operating characteristics (ROC) graph is a technique for visualizing, organizing and selecting classifiers based on their performance<sup>2</sup>. ROC graphs have long been used in signal detection theory, which in recent years has seen an increasing use for diagnostic, machine learning, and information-retrieval systems. ROC graphs plot false-positive (FP) rates on the  $x$ -axis and true-positive (TP) rates on the  $y$ -axis. Fig.B.1 shows an ROC graph with five classifiers labeled A through E. The points of the curve are obtained by sweeping the classification threshold from the most positive classification value to the most negative. The AUC is defined as the area under the ROC curve.

<sup>2</sup>This section is based on [Fawcett \[2006\]](#)



**Figure B.1:** A basic ROC graph showing five classifiers

Several points in ROC space are important to note. The lower left point  $(0,0)$  represents the strategy of never issuing a positive classification; such a classifier commits no false positive errors but also gains no true positives. For a fully random classification, the ROC curve is a straight line connecting the origin to  $(1,1)$ . Any improvement over random classification results in an ROC curve at least partially above this straight line. The point  $(0,1)$  represents perfect classification. D's performance is perfect as shown in the figure.

Informally, one point in ROC space is better than another if it is to the northwest (TP rate is higher, FP rate is lower, or both) of the first. Classifiers appearing on the left-hand side of an ROC graph, near the  $x$ -axis, may be thought of as “conservative”: they make positive classifications only with strong evidence so they make few false positive errors, but they often have low true positive rates as well. Classifiers on the upper right-hand side of an ROC graph may be thought of as “liberal”: they make positive classifications with weak evidence so they classify nearly all positives correctly, but they often have high false positive rates. In Fig.B.1, A is more conservative than B. Many real world domains are dominated by large numbers of negative instances, so performance in the far left-hand side of the ROC graph becomes more interesting.

In order to get away from this diagonal into the upper triangular region,

the classifier must exploit some information in the data. In Fig.B.1, C's performance is virtually random. At (0.7, 0.7), C may be said to be guessing the positive class 70% of the time.

Any classifier that appears in the lower right triangle performs worse than random guessing. This triangle is therefore usually empty in ROC graphs. If we negate a classifier-that is, reverse its classification decisions on every instance-its true positive classifications become false negative mistakes, and its false positives become true negatives. Therefore, any classifier that produces a point in the lower right triangle can be negated to produce a point in the upper left triangle. In Fig.B.1, E performs much worse than random, and is in fact the negation of B. Any classifier on the diagonal may be said to have no information about the class. A classifier below the diagonal may be said to have useful information, but it is applying the information incorrectly.



# Appendix C

## Time-frequency Distributions (TFDs)

### C.1 EEG analysis using Time-frequency Distributions (TFDs)

As described in [Feichtinger and Strohmer \[1998\]](#), the main argument of TFDs is that by combining the two different (but actually equivalent) lines of time and frequency, at the point where they meet, we can create a powerful tool for the analysis and representation of complex phenomena.

The TFDs are potentially very useful for detecting and analyzing non-stationary epileptic EEGs. Although visual analysis of raw EEG traces is still the major clinical tool and the point of reference for other methods, we can relate visual analysis to mathematics with a time-frequency description.

As there are several TFDs, it is convenient to know the different characteristics that each one presents, and their relationships between each other based on results applied to EEG. A description and review of many TFDs in the current literature are detailed below<sup>1</sup>.

#### C.1.1 From Fourier to quadratic TFDs

The classical Fourier Transform (FT) analysis is able to achieve infinite frequency resolution, but it does not provide temporal localization information. Proper description of the EEG often requires simultaneous localization in both time and frequency domains.

---

<sup>1</sup>This section is based on [Guerrero et al. \[2005\]](#)

The first solution consisted in a time-frequency representation of the signal:

$$X(t, \omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(\tau) h(\tau - t) e^{-j\omega\tau} dt d\tau \quad (\text{C.1})$$

$X(t, \omega)$  is the Short-Time Fourier Transform (STFT) of the signal  $x(t)$ .  $h(\tau)$  is called the windowing function. The squared magnitude of the STFT is the Spectrogram.

Other method is the Gabor representation that does not assume that the signal is know at arbitrary time and frequency of points, but at a lattice points:  $t = nT$ ,  $\omega = kF$ ; where  $n, k \in \mathbb{Z}$ ;  $T$  is the sampling interval in the time domain; and  $F$  is the sampling interval in the frequency domain. If so, a signal  $x(t)$  can be expressed as:

$$x(t) = \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} G(n, k) h(t - nT) e^{j2\pi kFT} \quad (\text{C.2})$$

where  $G(n, k)$  are the Gabor Coefficients, and  $h(t - nT) e^{j2\pi kFT}$  are the synthesis functions.

In 1932, Wigner derived a representation of the phase space in quantum mechanics and was the first example of joint time-frequency distributions that was different from the spectrogram. In 1948, Ville, searching for an “instantaneous spectrum” and influenced by the work of Gabor, introduced the same transform in signal analysis, introduced in section 3.4.2. The WV distribution is still in use in signal analysis, as described in Papandreu-Suppappola [2003], and it is utilized in a wide range of applications, including biomedical signals such as speech recognition, EEGs, heart sounds and muscle sounds, electrocorticograms, electrogastrogram and temporomandibular joint sounds. In 1966 Cohen provided an overall class of TFDs based on the Wigner Distribution (see Eq.3.29) and emphasized its importance in signal processing. This grouping led to the introduction, by other researchers, of  $\Phi(\theta, \tau)$  called as “kernel function”, and also provided an important model to obtain many different types of time-frequency distributions.

A large number of TFDs have been proposed, each differing only in the choice of the kernel function  $\Phi$ . The lack of a single distribution that is “best” for all applications and the trade-off between time and frequency resolutions has resulted in a proliferation of TFDs (see Table C.1).

Since these distributions are quadratic TFDs (QTFD), i.e. they are based on the properties of covariance by shifts in time or frequency, they introduce cross terms in the time-frequency plane<sup>2</sup>.

---

<sup>2</sup>Matlab toolbox and tutorial are available in <http://tftb.nongnu.org/>

**Table C.1:** Some distributions and their kernels

Name	kernel: $\Phi(\theta, \tau)$	Distribution: $P(t, \omega)$ (Notation: $\int_{-\infty}^{\infty} \equiv \int$ )
General class	$\Phi(\theta, \tau)$	$\frac{1}{2\pi} \int \int A(\theta, \tau) \Phi(\theta, \tau) e^{-j\theta t - j\omega \tau} d\theta d\tau$ where $A(\theta, \tau)$ is the ambiguity function defined by Eq.3.30
Wigner	1	$\frac{1}{2\pi} \int e^{-j\tau\omega} x(t + \frac{\tau}{2}) x^*(t - \frac{\tau}{2}) d\tau$
Choi-Williams	$e^{-\theta^2 \tau^2 / \sigma}$	$\frac{1}{2\sqrt{\pi}} \int \frac{1}{\sqrt{\tau^2 / \sigma}} A(\theta, \tau) e^{(-\sigma(u-t)^2 / \tau^2 - j\tau\omega - j\theta u)} d\tau$
Margenau-Hill	$\cos \frac{1}{2} \theta \tau$	$\text{Re} \frac{1}{\sqrt{2\pi}} x(t) X^*(\omega) e^{-jt\omega}$ , $X^*(\omega)$ is the FT of $x(t)$
Kirkwood-Rihaczek	$e^{j\theta\tau/2}$	$\frac{1}{\sqrt{2\pi}} x(t) X^*(\omega) e^{-jt\omega}$
Bon-Jordan	$\frac{\sin \frac{1}{2} \theta \tau}{\frac{1}{2} \theta \tau}$	$\int_{t- \tau /2}^{t+ \tau /2} \frac{A(\theta, \tau)}{ \tau } e^{-j\tau\omega - j\theta u} du$
Page	$e^{j\theta \tau }$	$\frac{\partial}{\partial t} \left  \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t x(t') e^{-j\omega t'} dt' \right ^2$
Spectrogram	$\int h^*(u - \frac{1}{2}\tau) e^{-j\theta u} h(u + \frac{1}{2}\tau) du$	$\left  \frac{1}{\sqrt{2\pi}} \int e^{-j\omega\tau} x(\tau) h(\tau - t) d\tau \right ^2$
Zhao-Atlas-Marks (ZAM)	$g(\tau)  \tau ^{\frac{\sin a \theta \tau}{a \theta \tau}}$	$\frac{1}{2a} \int_{t- \tau a}^{t+ \tau a} g(\tau) A(\theta, \tau) e^{-j\tau\omega - j\theta u}$

### C.1.2 The affine class

Most of linear or quadratic TDFs can be defined by the principle of covariance, which is used to obtain a time scale that favors the signal average analysis of the properties of translation and dilation in time. The next group of transformations are known as affinity groups, where the action induced on a signal  $x(t)$  is given by

$$x(t) \rightarrow x_{a',b'}(t) = \frac{1}{\sqrt{|a'|}} x\left(\frac{t-b'}{a'}\right) \tag{C.3}$$

and its Fourier transform is

$$X(\omega) \rightarrow X_{a',b'}(\omega) = \sqrt{|a'|} e^{-j\omega b'} X(a'\omega) \tag{C.4}$$

With the transformations above, it can be shown that when a bilinear distribution is scaled in time  $\Omega_x(t, a)$ , then it is invariant to these transformations, i.e.

$$\Omega_{x_{a',b'}}(t, a) = \Omega_x\left(\frac{t-b'}{a'}, \frac{a}{a'}\right) \tag{C.5}$$

An affine transformation can be characterized using a bi-frequency kernel

$$\Omega_x(t, a; \Pi) = \frac{1}{|a|} \iint_{-\infty}^{+\infty} \psi(\nu, \omega) X\left(\frac{\omega - \frac{\nu}{2}}{a}\right) X^*\left(\frac{\omega + \frac{\nu}{2}}{a}\right) e^{-j\omega t/a} d\nu d\omega \quad (\text{C.6})$$

with

$$\psi(\nu, \omega) = \int_{-\infty}^{+\infty} \Pi(t, \omega) e^{-j\omega t} dt \quad (\text{C.7})$$

Using different values to this kernel, we can obtain Bertrand distribution, Scalogram, D-Flandrin distribution, etc. These classes may be appropriate for several applications, but a good interpretation of the time-frequency and time-scale is necessary.

### C.1.3 The reassignment method

In 1995, Auger and Aldrin calculated parameters on the spectral components by tracking peaks on the time-frequency plane. As described [Flandrin \[1999\]](#), the method consisted on extracting a single value to represent a whole distribution and assigning this value to the geometric center of the domain over which the distribution is considered. Under this situation, the centroids calculated at each point TF  $(t, f)$  are displaced to another point  $(\hat{t}, \hat{f})$  and can be interpreted as the local instantaneous frequency (IF) and group delay (GD) of the analyzed signal. Thus, the average energy value of the signal around each point is not assigned to its geometric center, but to its center of gravity, which is much more representative than the local distribution of signal energy.

### C.1.4 The Reduced Interference Distribution (RID)

To eliminate the effects of the cross terms, Choi-Williams introduced the TFD distribution of Cohen Class obtained with a exponential kernel  $K(\theta, \tau) = e^{-\theta^2 \tau^2 / \sigma}$ , where  $(\sigma > 0)$  is the scaling factor and the cross terms are affected by this parameter.

The problem with this distribution is that it may not have time and frequency supports. To overcome this difficulty, a method was proposed that retains high time and frequency resolution and preserves most of the desirable kernel requirements including the time-frequency supports. This method, called the Reduced Interference Distribution (RID), is a cross-shaped low-pass filter, which satisfies  $|K(\theta, \tau)| \ll 1$ .

There are several window functions that estimates the RID.

### C.1.5 Optimal Kernel Design (OKD)

It is important to design a TFD that provides good performance for a variety of signals. Baraniuk and Jones [1993], proposed a radially Gaussian kernel design to produce a TFD that adapts to each signal and presents good performance for a large class of signals.

The optimal kernel  $K_{opt}(m, n)$  which is the discretized version of  $K_{opt}(\theta, \tau)$ , is real, nonnegative and radially nonincreasing.

### C.1.6 Other alternatives in time-Frequency analysis

Other time-frequency representation of a time series is the S-transform described in Stockwell [1999]. This transformation combines elements of wavelet and the STFT, from what the S-transform is similar to the STFT, but with a Gaussian window whose width scales inversely, and whose height scales linearly, with the frequency.

The S-transform has an advantage in that it provides multiresolution analysis while retaining the absolute phase of each frequency. This has led to its application for detection and interpretation of events in time series in a variety of disciplines such as seismic analysis, ECG artifact elimination and medical image processing, as described by Schimmel and Gallard [2005].

Besides wavelets, other approaches include the ambiguity function and the Friedman's instantaneous frequency density (FIFD), that can be derived from the reassigned distribution, as described in Cohen [1995, 1989].

Below we present an experimental results of the new techniques developed in the last decades for time-frequency analysis in the epileptic EEG. The main goal of including these experiments is shows different characteristics in 33 TFDs in events like detection and epileptic seizures.

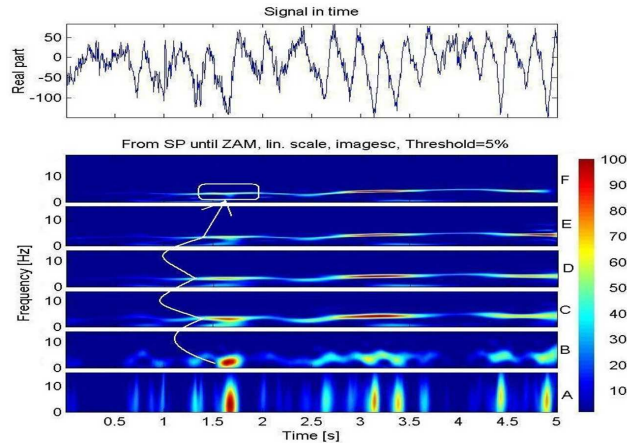
### C.1.7 Experiments

The dynamic nature of epileptic phenomena causes EEG signals to exhibit stochastic and non-stationary behavior and the TFDs are potentially very useful for detecting and analyzing non-stationary epileptic EEGs. We will show next the ability of TFDs to detect spikes in EEG signals from epileptic patients.

#### Data collection

The data setup consists in 6 EEG epileptic records (adult patients) with focal epileptiform activity obtained while restful wakefulness stage. EEG records of 3 min and had previously analyzed by experienced neurologists.

23 channels of EEG have been recorded in each session using the 10-20 International System of Electrode Placement. Raw EEG data were digitized at a sample of 200 Hz using “DAD-32” equipment (LA MONT MEDICAL) and were analyzed using 5 seconds a long segments.



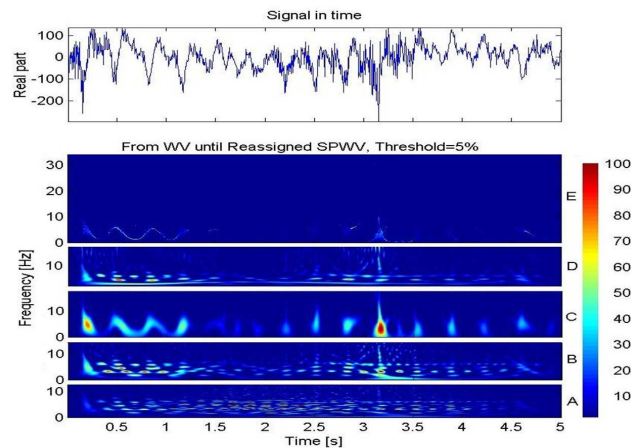
**Figure C.1:** Beginning of a seizure. Bands from A to E correspond to spectrograms with Hamming windows length 17, 101, 171, 251 samples respectively. F: ZAM distributions with Hamming 2-D filter (101,251)

## Results

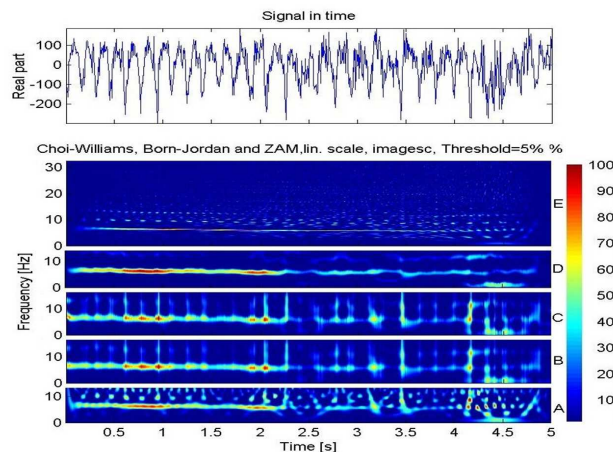
The following figures contain EEG segments obtained with different TFDs. The frequency band is 0-10 Hz. There are distributions that use 2-D filters ( $W_t, W_w$ ) with Hamming and Kaiser windows. Note from Fig.C.1 the disappearance of an event as the beginning of the seizure. In Fig.C.1 (band B) a burst frequency (1.65 sec) can be seen that disappears in the band F. Bands A-E show the windowing effect in the spectrogram when using different windows. The Zhao-Atlas-Marks (ZAM) distribution (band F) for these values displays low temporal resolution.

Fig.C.2 shows the comparison between different TFDs based on the Wigner Ville (WV) distributions. Note that bands A-D keep introducing cross-terms, and the Reassigned method (band E) shows a nonlinear characteristic with a continuous sinusoidal like trace. When the nature of the analyzed signal changes, it is possible to find strong interferences.

The distribution in Fig.C.3 introduce 2-D filtering. It can be observe that Choi-Williams (bands A and B), the parameter  $\sigma$  is an important factor for improving resolution but has presence of artifacts. Similar results have presented Bon-Jordan and ZAM distributions. The Reassignment method provides time-frequency pictures closer to the nature of epileptic signals.



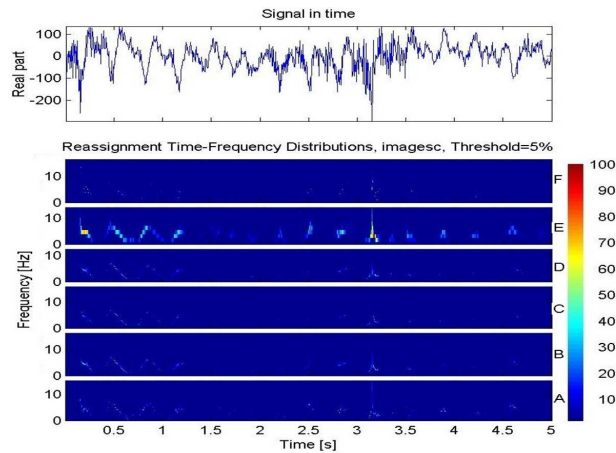
**Figure C.2:** Some TFDs of EEG during an epileptic seizure. Kaiser 2-D filter (15,54) was used and the computed TFDs are: A) Wigner Ville Distribution, B) Pseudo WV windows, C) Smoothed Pseudo WV, D) Smoothed Pseudo affine WV. Length Morlet Wavelet  $L=31$ , 62 and E) Reassigned Smoothed Pseudo WV. We can observe that TFDs A, B and D introduce cross-terms (manifested as small energy bumps spread in the time-frequency plane), while the Reassigned method (E) and SPWV (C) do not present cross-terms and are able to capture in a better way the non-stationary nature of the signal.



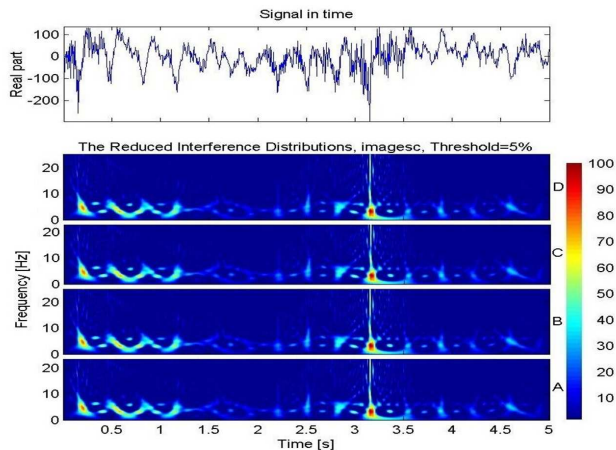
**Figure C.3:** TFDs with 2-D filtering. Figures from A to D have a Kaiser 2-D filter (5,241) ms. A. Choi-Williams Distribution with  $\sigma = 3$ . B. Choi-Williams with  $\sigma = 10$ . C. Born-Jordan Distribution. D. ZAM distribution. E. ZAM with Kaiser 2-D filter (3,241) ms.

Fig.C.4 shows different results of reassignment. The bands A-E show that energy trajectories rapidly change from 0 to 8 Hz (aprox.). Similar results have been obtained with RID (Fig.C.5), but the time-frequency plane reso-

lution is poor.

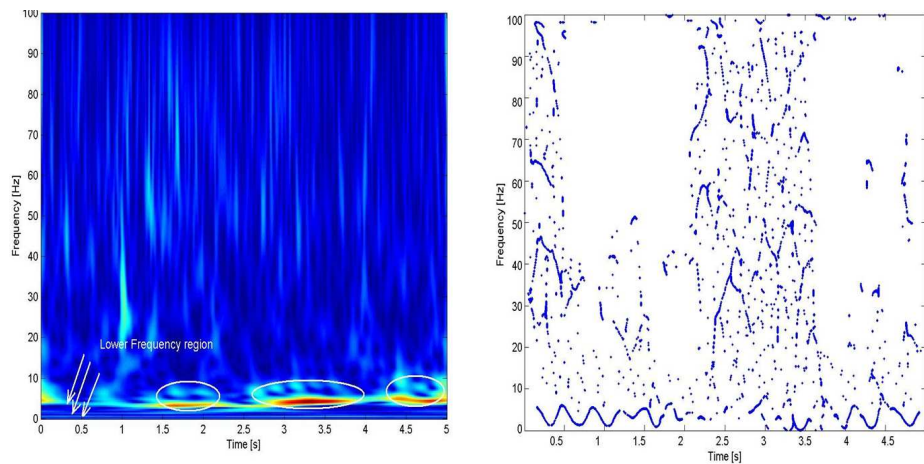


**Figure C.4:** TF representation of the reassignment method with  $N=1000$ . 1-D Kaiser filter of length 25 samples was used. A. Gabor. B. Spectrogram. C. Pseudo WV. D. Pseudo Margenau-Hill. E. Stankovic distribution with 2-D Kaiser filter (7,25). F. Pseudo Page.



**Figure C.5:** The RID distribution with different kernels using a 2-D Kaiser filter (7,151): A. Bessel. B. Binomial. C. Hanning. D. Triangular. The RID's, shows the dynamic behavior of the EEG and presents identical results with different kernels

Other distributions with interesting results are shown in Fig.C.6 using the S-transform and ridges method. In Fig.C.6 (right) we can see several lines between 0-8 Hz, and Fig.C.6 (left) shows a continuous wave in the same band. The result with OKD and FIFD shows similar behavior in low frequencies. Finally, we have analyzed the efficiency of the algorithms introduced before and compared them with others regarding the average execution time with 5 secs. long EEG segment. The software was implemented on a Pentium IV, 1



**Figure C.6:** Epileptic seizure. Right: The S-Transform show several lines in a lower frequency region. Left: Good resolution by “Ridges” method. Note the nonlinear characteristic and continuous wave when the crisis occurred.

Ghz, 164 RAM using Matlab©V6.1. The fastest TFD was the STFT (0.26 sec) and the slowest was the Reassignment Scalogram with Morlet wavelet (256 sec.) per segment. [See Table C.2].

**Table C.2:** Computation time of different TFD’s using rectangular non-overlapping windows. EEG segment length 5 secs.

Time-frequency distribution	Computation time (seconds)
STFT, Spectrogram, WV, PWV, Rihaczeck, Margenau-Hill (and Pseudo), Page (and Pseudo), Butterworth, MH Spectrogram	$0 \leq t \leq 1$
Scalogram, Smoothed Pseudo WV, Cross-Entropy	$0 \leq t \leq 5$
Gabor, Choi-Williams, Born-Jordan, ZAM, S-Transform, RID’s	$5 \leq t \leq 10$
Affine PWV, Rectangular, R. Pseudo WV, R. Page, R. PMH, R. Stankovic	$10 \leq t \leq 15$
D-Flandrin, Unter, R. Gabor, R. Scalo, R. SPWV, Radially Gaussian Kernel, Ridges, Skeleton	$t \geq 15$

### C.1.8 Discussions and conclusion

Before trying to answer what TFD is the best, the question turns to: what information do we need about the distributions? To study the characteris-

tics of the dynamics of epileptic EEG, the Reassignment method, Ridges or RID's distributions are suitable. If we are simply monitoring for presurgery evaluation, the Smoothed PWV gives good resolution. Choi-Williams, RID, SPWV or RSPWV are suitable for events detection such as seizures and although the OKD method is slow (32 seconds) it has an excellent frequency resolution in low bands.

As conclusion, we can say that choosing a distribution depends on the information that it wants to be extracted. There are distributions that give similar results, but they differ in the time required for computation. It is necessary to have a TFD that provides the required information in an efficient way.

# Appendix D

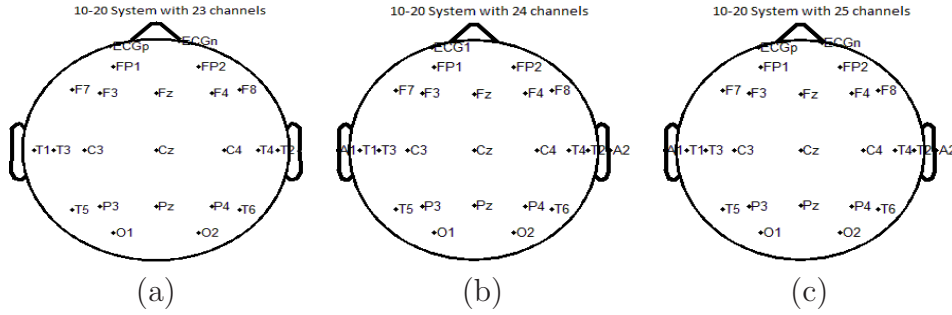
## Databases

The performance of the algorithms proposed in this Thesis are assessed on 3 databases for EEG segments classification problems, which have been selected so that there is sufficient diversity in terms of number of samples, artifacts presence, epileptic activity, unbalanced data and EEG complexity (connection and disconnection EEG in the same register, ictal activity, etc.). Each dataset is obtained from different recording regions and from different physiological and pathological brain states.

The following sections describe the databases used in this Thesis.

### D.1 Database 1

The data collection was obtained taking into account: (i) same type of epilepsy, (ii) activity at different brain areas and (iii) balance of the classes (seizure and non-seizure). The EEG records of 7 adult epileptic patients were obtained in a restful wakefulness stage and recorded at the Clinica Universitaria de Navarra, Department of Neurophysiology (Pamplona, Spain). All of them contained focal epileptiform activity, according to experienced neurologists. EEG data were recorded from 23, 24 and 25 scalp electrodes based on the 10-20 International System of Electrode Placement with additional anterotemporal electrodes T1/T2 (see Fig.D.1). Raw EEG data were digitized at a sample rate of 200 Hz using a “DAD-32” equipment (La Mont Medical) and were filtered by a digital low-pass filter with cut-off frequency of 20 Hz. Table D.1 summarizes the main characteristics of the problem: the notation for each problem, EEG channels where the record was made and the number of data in each class (segments with seizure correspond to labels +1), both in the training set and test.



**Figure D.1:** General 10-20 System scheme used for our databases. (a) System with 23 channels. (b) System with 24 channels. (c) System with 25 channels

**Table D.1:** Main characteristics of the database 1.

EEG Channel	Notation	Train data $n_{+1}/n_{-1}$	Test data $n_{+1}/n_{-1}$
T1, T3, T5	Patient 1	55/151	34/102
F8, T2, T4	Patient 2	39/168	28/110
A2, T4, T6	Patient 3	188/185	139/109
A1, F7, T1	Patient 4	122/165	81/169
A1, F7, T1, T3	Patient 5	85/362	57/240
A2, F8, T2, T4	Patient 6	159/609	101/411

## D.2 Database 2

This dataset consist of a large EEG (10000 samples aprox.) that contains two epileptic seizures, high presence of ocular artifacts, interictal epileptiform discharges and a couple of minutes of disconnection of EEG recording. The EEG recordings were acquired using the standard video-EEG equipment (23-channel digital EEG) at the Clínica Universitaria de Navarra, Department of Neurophysiology (Pamplona, Spain) with LaMont amplifiers “DAD-32” equipment (LaMont Medical, Madison, WI, U.S.A) and Harmonie software (Stellate, Montreal, Quebec, Canada). This register contain generalized epileptiform activity, according to experienced neurologists. Electrodes were placed according to the 10-20 system, with additional anterotemporal electrodes in T1/T2 and with both mastoids as reference. Raw EEG data were digitized at a sample rate of 200 Hz and were filtered by a digital low-pass filter with cut-off frequency of 20 Hz. Table D.2 summarizes the main characteristics of the problem.

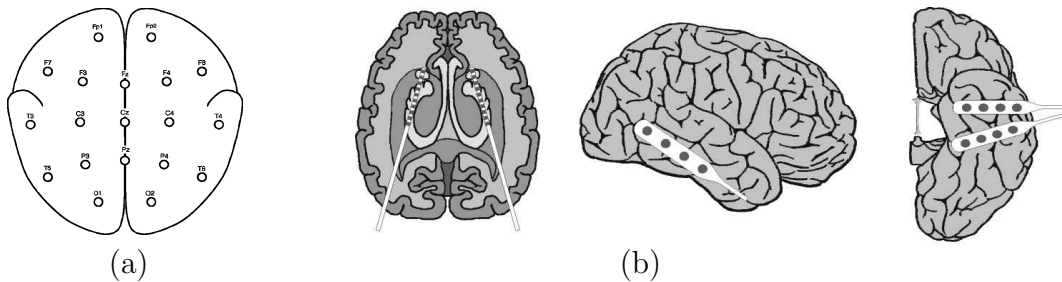
**Table D.2:** Main characteristics of the database 2.

Dataset	Notation	Train data $n_{+1}/n_{-1}$	Test data $n_{+1}/n_{-1}$
Two crisis	Database 2	2268/6284	434/1135

### D.3 Database 3

The dataset consist of sets of EEG time series: surface EEG signals from healthy volunteers with eyes closed and eyes open, and intracranial EEGs from epilepsy patients during the seizure free interval from within and from outside the seizure generating area as well as intracranial EEG recordings of epileptic seizures (described in [Andrzejack et al. \[2001\]](#))).

The five sets ((denoted as Z, O, N, F and S) each one containing 100 single-channel EEG segments each having 23.6 sec duration and sampling rate of 173.61 Hz. Sets Z and O are EEG segments acquired extracranially according to the international 10-20 system from surface EEG recordings of five healthy volunteers, with eyes open (Z) and closed (O), respectively (see Fig.D.2 (a)). Segments in subsets F and N contain seizure-free intervals from five patients in the epileptogenic zone (F) and from the hippocampal formation of the opposite hemisphere of the brain (N). Subset S presents seizure activity, selected from all recording sites exhibiting ictal activity. Sets N, F, and S have been recorded intracranially. More specifically, depth electrodes are implanted symmetrically into the hippocampal formation (see Fig.D.2 (b)). All dataset were recorded with the same 128-channel amplifier system using an average common reference.



**Figure D.2:** Characteristics of electrodes used for dataset 3. Sets Z and O have been recorded extracranially, whereas sets N, F and S have been recorded intracranially. (a) Localizations of electrodes according to the international 10-20 system. (b) Scheme of intracranial electrodes implanted for presurgical evaluation of epilepsy patients (from [Andrzejack et al. \[2001\]](#)).

The three classification problems created were:

- 1) The first problem called N1, two classes are examined: normal (Z) and seizure (S).
- 2) The second classification problem called N2, includes the classes normal (Z), seizure-free (F) and seizure (S).
- 3) In the third problem called N3, all the five classes are used (Z,O,N,F,S).

Table D.3 summarizes the main characteristics of the problems.

**Table D.3:** Main characteristics of the database 3.

Dataset	Notation	Train data $n_{+1}/n_{-1}$	Test data $n_{+1}/n_{-1}$
Z,S	N1	310/305	192/217
Z,F,S	N2	272/651	202/412
Z,O,N,F,S	N3	335/1202	209/815

# Bibliography

- D. Abásolo, J. Escudero, R. Hornero, C. Gómez, and P. Espino. Approximate entropy and auto mutual information analysis of the electroencephalogram in alzheimer's disease patients. *Med. Biol. Eng. Comput.*, 46:1019–1028, 2008.
- N. Acir, I. Oztura, M. Kuntalp, B. Baklan, and C. Guzelis. Automatic detection of epileptiform events in EEG by three-stage procedure based on artificial neural networks. *IEEE Trans. On Biomed. Eng.*, 52:30–40, 2005.
- V. Afonso and W. Tompkins. Detecting ventricular fibrillation. *IEEE Engineering in Medicine and Biology*, 14:152–159, 1995.
- M. Akay. *Detection and estimation methods for biomedical signals*, volume 1. 1 ed Academic Press, 1996.
- L. Almeida. The fractional fourier transform and time-frequency representation. *IEEE. Trans. on Signal Proc.*, 42:3084–3091, 1994.
- R. Andrzejack, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. Elger. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Rev. E*, 64:061907, 2001.
- S. H. Ardalan. Floating-point error analysis of recursive least-squares and least-mean-squares adaptive filters. *IEEE Trans. On Circuits and Systems*, 12:1192–1208, 1986.
- F. Auger, P. Aldrin, P. Goncalves, and O. Lemoine. *Time-frequency toolbox for Matlab, User's guide and reference guide*. CNRS (France) and Rice University (USA), 1996.
- T. Babb, E. Mariani, and P. Crandall. An electronic circuit for detection of EEG seizures recorded with implanted electrodes. *Electroencephalography and Clinical Neurophysiol.*, 37:305–308, 1974.
- R. Baraniuk and D. Jones. A signal-dependent time-frequency representation: optimal kernel design. *IEEE. Trans. on Signal Proc.*, 41:1589–1602, 1993.

- S. Barbarossa and O. Lemoine. Analysis of nonlinear fm signals by pattern recognition of their time-frequency representation. *IEEE Signal Proc. Letters*, 3:112–115, 1996.
- E. D. Übeyli and I. Güler. Features extracted by eigenvector methods for detecting variability of EEG signals. *Pattern Recognit. Lett.*, 28:592–603, 2007.
- W. Blume, G. Young, and J. L. J.F. EEG morphology of partial epileptic seizures. *Electroencephalogr Clin Neurophysiol*, 4:295–302, 1984.
- B. Boashash. *Time Frequency Signal Analysis and processing. A comprehensive reference*, volume 1. Elsevier, 2003.
- B. Boashash. Estimating and interpreting the instantaneous frequency of a signal-part 2: Algorithms and applications. *Proceedings of the IEEE*, 80:540–568, 1992.
- B. Boashash and M. Mesbah. A time-frequency approach for newborn seizure detection. *IEEE Eng. in Med. and Biol. Magazine*, 20:54–64, 2001.
- J. Cardoso. Blind signal separation: statistical principles. *Proceedings of the IEE*, 86:2009–2025, 1998.
- R. Carmona, W. Hwang, and B. Torr sani. Multiridge detection and time-frequency reconstruction. *IEEE. Trans. on Signal Proc.*, 47:480–492, 1999.
- P. Celka and P. Colditz. Nonlinear nonstationary wiener model of infant EEG seizures. *IEEE Trans. On Biomed. Eng.*, 49:556–564, 2002.
- S. Chiappa and D. Barber. EEG classification using generative independent component analysis. *Neurocomputing*, 69:769–777, 2006.
- K. Chua, V. Chandran, U. Acharya, and C. Lim. Application of higher order spectra to identify epileptic EEG. *Journal of Medical Systems*, pages 1–9, 2010.
- W. D. Clercq, A. Vergult, B. Vanrumste, W. V. Paesschen, and S. V. Huffel. Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. *IEEE Trans. On Biomed. Eng.*, 53:2583–2587, 2006.
- L. Cohen. *Time-Frequency Analysis*, volume 1. Prentice Hall, 1995.
- L. Cohen. Time-frequency distributions-a review. *Proceedings of the IEEE*, 77:941–981, 1989.
- T. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. On Electronic Computers*, EC-14:326–334, 1965.

- T. Cover and J. Thomas. *Elements of Information Theory*, volume 2. Wiley, 1991.
- A. Delorme and S. Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134:9–21, 2004.
- A. Delorme, T. Jung, T. Sejnowski, and S. Makeig. Improved rejection of artifacts from EEG data using high-order statistics and independent component analysis. *Neuroimage*, 34:1443–1449, 2007.
- R. Duda, P. Hart, and D. Stork. *Pattern classification*, volume 2. Elsevier, 2009.
- P. Durka. *Time-Frequency analysis of EEG*. Thesis Institute of Experimental Physics, Warsaw University, 1996.
- J. Ebersole and T. Pedley. *Current Practice of Clinical Electroencephalography*, volume 3. Lippincott Williams & Wilkins, 2003.
- J. Elting, T. van Weerden, and et al. P300 component identification using source analysis techniques: Reduced latency variability. *Journal of Clin. Neurophysiol.*, 1:26–34, 2003.
- D. Erdogmus and J. Principe. From linear adaptive filtering to nonlinear information processing - the design and analysis of information processing systems. *IEEE Signal Processing Magazine*, 23:14–33, 2006.
- T. Fawcett. An introduction to roc analysis. *Pattern Recognit. Lett.*, 27: 861–874, 2006.
- H. Feichtinger and T. Strohmer. *Gabor Analysis and algorithms: Theory and applications*, volume 1. Birkhauser, 1998.
- B. J. Fisch. *Fisch and Spehlmann's EEG primer: Basic principles of digital and analog EEG*, volume 3. Springer, 1999.
- P. Flandrin. *Time-Frequency / Time-Scale Analysis*, volume 1. Academic Press., 1999.
- D. Françoise, F. Rossi, W. Wertz, and M. Verleysen. Resampling methods for parameter-free and robust feature selection with mutual information. *Neurocomputing*, 70:1276–1288, 2007.
- W. Freeman. The electrical activity of a primary sensory cortex: Analysis of EEG waves. *International Reviews of Neurobiology*, 5:53–119, 1963.

- M. Gandetto, M. Guainazzo, and C. Regazzoni. Use of time-frequency analysis and neural networks for mode identification in a wireless software-defined radio approach. *EURASIP Journal on Applied Signal Processing*, 12:1778–1790, 2004.
- H. Ghandeharion and A. Erfanian. A fully automatic method for ocular artifact suppression from EEG data using wavelet transform and independent component analysis. *Proceedings of the 28th IEEE Annual EMBS International Conference*, 1:5265–5268, 2006.
- I. Güler and E. Übeyli. Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients. *Journal of Neuroscience Methods*, 148:113–121, 2005.
- I. Güler and E. D. Übeyli. Multiclass support vector machines for EEG-signals classification. *IEEE Trans. on Inf. Tech. in Biomed.*, 11:117–126, 2007.
- N. F. Güler, E. Übeyli, and I. Güler. Recurrent neural networks employing lyapunov exponents for EEG signals classification. *Exp. Syst. Appl.*, 29: 506–514, 2005.
- V. Gomez-Verdejo, M. Verleysen, and J. Fleury. Information-theoretic feature selection for functional data classification. *Neurocomputing*, 72:3580–3589, 2009.
- J. Gotman. Automatic recognition of epileptic seizures in the EEG. *Electroencephalography and Clinical Neurophysiol.*, 54:530–540, 1982.
- C. Guerrero, A. M. Trigueros, and J. I. Franco. Time-frequency EEG analysis in epilepsy: What is more suitable? *Proceedings of the 5th IEEE Symposium on Signal Processing and Info. Tech.*, 1:202–207, 2005.
- C. Guerrero-Mosquera. La bioelectricidad cerebral en las patologías neurológicas. *Mente cerebro y sociedad*, pages 4946–4949, 2011.
- C. Guerrero-Mosquera and A. Navia-Vazquez. Automatic removal of ocular artifacts from EEG data using adaptive filtering and independent component analysis. *Proceedings of the 17th European Signal Processing Conference (EUSIPCO)*, pages 2317–2321, 2009.
- C. Guerrero-Mosquera and A. Navia-Vazquez. Automatic removal of ocular artifacts using Adaptive Filtering and Independent Component Analysis for EEG data. *Next to appear in IET Signal Processing*, 2011.
- C. Guerrero-Mosquera, A. M. Trigueros, J. I. Franco, and A. Navia-Vazquez. New feature extraction approach for epileptic EEG signal detection using time-frequency distributions. *Med. Biol. Eng. Comput.*, 48:321–330, 2010a.

- C. Guerrero-Mosquera, M. Verleysen, and A. Navia-Vazquez. EEG feature selection using mutual information and support vector machine: A comparative analysis. *Proceedings of the 32nd Annual EMBS International Conference*, pages 4946–4949, 2010b.
- C. Guerrero-Mosquera, M. Verleysen, and A. Navia-Vazquez. Dimensionality reduction for EEG classification using Mutual Information and SVM. *Next to appear in Machine Learning Signal Processing Conference (MLSP11)*, 2011.
- I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.
- I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh. *Feature Extraction, Foundations and Applications*, volume 1. Springer, 2006.
- J. Hammond and P. White. The analysis of non-stationary signals using time-frequency methods. *Journal of Sound and Vibration*, 3:419–447, 1996.
- F. Harrel. *Regression Modeling Strategies*. Springer, 2001.
- O. Hasan. Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm. *Signal processing*, 88:1858–1867, 2008.
- S. Haykin. *Adaptive filter theory*, volume 3. Prentice Hall, 1996.
- S. Haykin. *Neural Networks: A comprehensive foundation*, volume 2. Prentice Hall, 1999.
- P. He, G. Wilson, and C. Russel. Removal of ocular artifacts from electroencephalogram by adaptive filtering. *Med. Biol. Eng. Comput.*, 42:407–412, 2004.
- P. He, G. Wilson, C. Russel, and M. Gerschutz. Removal of ocular artifacts from the EEG: a comparison between time-domain regression method and adaptive filtering method using simulated data. *Med. Biol. Eng. Comput.*, 45:495–503, 2007.
- M. Hearst, S. Dumais, E. Osman, J. Platt, and B. Schölkopf. Support vector machines. *IEEE Intelligent Sys. and their applications*, 13:18–28, 1998.
- K. Hild, D. Erdogmus, K. Torkkola, and J. Principe. Feature extraction using information-theoretic learning. *IEEE Trans. On Pattern analysis and Machine Intelligence*, 28:1385–1392, 2006.
- H. Hinrikus, A. Suhhova, and M. Bachmann. Electroencephalographic spectral asymmetry index for detection of depression. *Med. Biol. Eng. Comput.*, 47:1291–1299, 2009.

- F. Hlawatsch and G. Boudreaux-Bartels. Linear and quadratic time-frequency signal representation. *IEEE SP Magazine*, 9:21–67, 1992.
- C. Hsu, C. Chang, and C. Lin. A practical guide to Support Vector Classification. <http://www.csie.ntu.edu.tw/~cjlin>, 2007.
- J. Iriarte, E. Urrestarazu, M. Valencia, M. Alegre, A. Malanda, C. Viteri, and J. Artieda. Independent component analysis as a tool to eliminate artifacts in EEG: A quantitative study. *Journal of Clin. Neurophysiol.*, 20:249–257, 2003.
- J. Ives, C. Thompson, P. Gloor, A. Olivier, and J. Woods. The on-line computer detection and recording of spontaneous temporal lobe epileptic seizures from patients with implanted depth electrodes via radio telemetry link. *Electroencephalography and Clinical Neurophysiol.*, 37:205, 1974.
- C. James and C. Hesse. Independent component analysis for biomedical signals. *Physiol. Meas.*, 26:R15–R39, 2005.
- C. James and D. Lowe. Using independent component analysis and dynamical embedding to isolate seizure activity in the EEG. *Proceedings of the 22nd Annual EMBS International Conference*, 2:1329–1332, 2000.
- C. Joyce, I. Gorodnitsky, and M. Kutas. Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology*, 41:1–13, 2004.
- N. Kannathal, U. Acharya, C. Lim, and P. Sadasivan. Characterization of EEG- A comparative study. *Comput. Methods Prog. Biomed*, 80:17–23, 2005a.
- N. Kannathal, M. Choo, U. Acharya, and P. Sadasivan. Entropies for detection of epilepsy in EEG. *Comput. Methods Prog. Biomed*, 80:187–194, 2005b.
- A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Physical Review E*, 69:066138, 2004.
- N. Kwak and C. Choi. Input feature selection by mutual information based on parzen window. *IEEE Trans. On Pattern analysis and Machine Intelligence*, 24:1667–1671, 2002.
- M. Latka and Z. Was. Wavelet analysis of epileptic spikes. *Physical Rev. E*, 67:052902, 2003.
- C. Lehmann, T. Koenig, V. Jelic, L. Prichep, R. John, L. Wahlund, Y. Dodgee, and T. Dierks. Application and comparison of classification

- algorithms for recognition of alzheimer's disease in electrical brain activity (EEG). *Journal of Neuroscience Methods*, 161:342–350, 2007.
- H. Li and Y. Sun. The study and test of ICA algorithms. *Proc. IEEE Wireless Communications, Networking and Mobile Computing*, 1:602–605, 2005.
- Y. Li, Z. Ma, W. Lu, and Y. Li. Automatic removal of the eye blink artifact from EEG using an ICA-based template matching approach. *Physiol. Meas.*, 27:425–436, 2006.
- S. Liang, H. Wang, and W. L. Chang. Combination of EEG complexity and spectral analysis for epilepsy diagnosis and seizure detection. *EURASIP Journal on Applied Signal Processing*, 2010:853434, 2010.
- C. Lima, A. Coelho, and S. Chagas. Automatic EEG signal classification for epilepsy diagnosis with relevance vector machines. *Expert Systems with Applications*, 36:10054–10059, 2009.
- Z.-Y. Lin and J. Chen. Advances in time-frequency analysis of biomedical signals. *Critical Reviews in Biomedical Engineering*, 24:1–70, 1996.
- F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for EEG based brain-computer-interface. *Journal of Neural Eng.*, 4:R1–R13, 2007.
- S. Makeig, A. Bell, T. Jung, and T. Sejnowski. Independent component analysis of electroencephalogram data. *Adv. in Neural Informatics Processing Sys.*, pages 145–151, 1996.
- M. Malarvili, L. Rankine, M. Mesbah, P. Colditz, and B. Boashash. Heart rate variability characterization using a time-frequency based on instantaneous frequency estimation technique. *Proc. Of IFMBE Int. Conf. On Biomedical Eng.*, 15:455–459, 2007.
- S. Mallat. *A wavelet tour of signal processing, Third edition: The sparse way*, volume 3. Elsevier, 2009.
- R. McAulay and T. Quatieri. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Trans. Acoustic, Speech, and Signal Processing*, 34:744–754, 1986.
- K.-R. Müller, C. Anderson, and G. Birch. Linear and nonlinear methods for brain-computer interfaces. *IEEE Trans. on Neural Systems and Rehabilitation Eng.*, 11:165–168, 2003.
- J. Mosher and R. Leahy. Recursive MUSIC: A framework for EEG and MEG source localization. *IEEE Trans. On Biomed. Eng.*, 45:1342–1354, 1998.

- A. Murro, D. King, J. Smith, B. Gallagher, H. Flanigin, and K. Meador. Computerized seizure detection of complex partial seizures. *Electroencephalography and Clinical Neurophysiol.*, 79:330–333, 1991.
- V. P. Nigam and D. Graupe. A neural network based detection of epilepsy. *Neurol. Res.*, 26:55–60, 2004.
- A. Papandreou-Suppappola. *Applications in Time-Frequency signal processing*, volume 1. CRC Press., 2003.
- K. Polat and S. Günes. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast fourier transform. *App. Math. Comput.*, 32:625–631, 2007.
- A. Poularikas and Z. Ramadan. *Adaptive filtering primer with Matlab*, volume 1. CRC Press, 2006.
- P. Prior, R. Virden, and D. Maynard. An EEG device for monitoring seizure discharges. *Epilepsia*, 14:367–372, 1973.
- S. Puthusserypady and T. Ratnarajah. Robust adaptive techniques for minimization of EOG artefacts from EEG signals. *Signal Processing*, 86:2351–2363, 2006.
- R. Q. Quiroga. Quantitative analysis of EEG signals: Time-frequency methods and chaos theory. 1998.
- L. Rankine, M. Mesbah, and B. Boashash. If estimation for multicomponent signals using image processing techniques in the time-frequency plane. *Signal Processing*, 87:1234–1250, 2007.
- S. Romero, M. Mañanas, and M. Barbanoj. A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: A simulation case. *Comp. Biol. and Med.*, 38:348–360, 2008.
- F. Rossi, A. Lendasse, D. Françoise, W. Wertz, and M. Verleysen. Mutual information for the selection of relevant variables spectrometric nonlinear modelling. *Chemometrics and Intelligent Laboratory Systems*, 80:215–226, 2006.
- F. Rossi, D. Françoise, W. Wertz, M. Meurens, and M. Verleysen. Fast selection of spectral variables with b-spline compression. *Chemometrics and Intelligent Laboratory Systems*, 86:208–218, 2007.
- G. Rätsch. A brief introduction to machine learning. <http://www.tuebingen.mpg.de/~raetsch>, 2004.

- N. Sadati, H. R. Mohseni, and A. Magshoudi. Epileptic seizure detection using neural fuzzy networks. *Proc. Of IEEE Int. Conf. On Fuzzy Syst.*, 29:506–514, 2006.
- S. Sanei and J. Chambers. *EEG signal processing*, volume 1. Wiley, 2007.
- M. Schimmel and J. Gallard. The inverse s-transform in filtes with time-frequency localization. *IEEE. Trans. on Signal Proc.*, 53:4417–4422, 2005.
- B. Schölkopf and A. Smola. *Learning with kernels*, volume 1. The MIT Press, 2002.
- L. Senhadji and F. Wendling. Epileptic transient detection: wavelets and time-frequency approaches. *Neurophysiol Clin.*, 32:175–192, 2002.
- J. Shawe-Taylor and N. Cristianini. *Support Vector Machines and other kernel-based learning methods*, volume 1. Cambridge University Press, 2000.
- Y. Siuly and P. Wen. Classification of EEG signals using sampling techniques and least square support vector machine. *Lectures notes in Computer Science*, 5589:375–382, 2009.
- V. Srinivasan, C. Eswaran, and N. Sriraam. Artificial neural network based epileptic detection using time-frequency domain features. *Journal of Medical Systems*, 29:647–660, 2005.
- L. Sörnmo and P. Laguna. *Bioelectrical signal processing in cardiac and neurological applications*, volume 1. Elsevier Academic Press, 2005.
- R. Stockwell. *S-transform analysis of gravity wave activity from a small scale network of airglow imagers*. Thesis Faculty of Graduate Studies, University of Western Ontario, 1999.
- J. Stone. *Independent Component Analysis: A tutorial introduction*, volume 1. The MIT Press., 2004.
- A. Subasi. EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*, 32:1084–1093, 2007.
- V. Swarnkar, U. Abeyratne, C. Hukins, and B. Duce. A state transition-based method for quantifying EEG sleep fragmentation. *Med. Biol. Eng. Comput.*, 47:1053–1061, 2009.
- R. Tao, B. Deng, W. Zhang, and Y. Wuang. Sampling and sampling rate conversion of band limited signals in the fractional fourier transform domain. *IEEE. Trans. on Signal Proc.*, 56:158–171, 2008.

- N. Thakor and S. Tong. From bench to bedside: An overview of clinical neuroengineering from the guest editors. *IEEE Eng. in Med. and Biol. Magazine*, 25(4):18–19, 2006a.
- N. Thakor and S. Tong. From bench to bedside. *IEEE Eng. in Med. and Biol. Magazine*, 25:18–19, 2006b.
- M. Thusalidas, C. Guan, and J. Wu. Robust classification of EEG signal for brain-computer interface. *IEEE Trans. on Neural Systems and Rehabilitation Eng.*, 14:24–29, 2006.
- A. Tzallas, M. Tsipouras, and D. Fotiadis. Epileptic seizure detection in EEGs using time-frequency analysis. *IEEE Trans. on Inf. Tech. in Biomed.*, 13:703–710, 2009a.
- A. Tzallas, M. Tsipouras, and D. Fotiadis. Epileptic seizure detection in EEGs using time-frequency analysis. *IEEE Trans. on Inf. Tech. in Biomed.*, 13:703–710, 2009b.
- A. Vallabhaneni, T. Wang, and B. He. Brain computer interface. *Neural Engineering*, 2005.
- P. L. Van, E. Urrestarazu, and J. Gotman. A system for automatic removal in ictal scalp EEG based on independent component analysis and bayesian classification. *Clin. Neurophysiol.*, 117:912–927, 2006.
- V. Vapnik. *The Nature of Statistical Learning Theory*, volume 2. Springer-Verlag, 2000.
- W. Williams, H. Zaverly, and J. Sackellares. Time-frequency analysis in electrophysiology signals in epilepsy. *IEEE Engineering in Medicine and Biology*, 14:133–143, 1995.
- W. Wojcikiewicz, C. Vidaurre, and M. Kawanabe. Improving classification performance of BCIs by using Stationary Common Spatial Patterns and unsupervised bias adaptation. *Hybrid Artificial Intelligent Systems*, 6679: 34–41, 2011.
- Y. Wongsawat. Epileptic seizure detection in EEG recordings using phase congruency. *Proceedings of the 30th IEEE Annual EMBS International Conference*, pages 927–930, 2008.
- Q. Xu, H. Zhou, Y. Wang, and J. Huang. Fuzzy support vector machine for classification of EEG signals using wavelet-based features. *Medical Eng. & Physics*, 31:858–865, 2009.

- M. Zivanovic. Detection of non-stationary sinusoids by using joint frequency reassignment and null-to-null bandwidth. *Digital Signal Processing*, 21: 77–86, 2011.
- U. Zölzer. *DAFX, Digital Audio Effects*. Wiley, 2002.
- A. Zoubir and D. R. Iskander. *Bootstrap techniques for signal processing*, volume 1. Cambridge, 2004.