



Proyecto Fin de Máster
Máster en Ingeniería de Sistemas y Control
Curso 2015-2016

Sistema de detección de momentos clave en vídeos mediante el análisis de la interrelación de señales electroencefalográficas (EEG)

Eduardo Mariano Horna Prat

Convocatoria de Septiembre de 2016

Directores: Raquel Dormido Canto, Natividad Duro Carralero y Ángel Mur
Departamento de Informática y Automática
Universidad Nacional de Educación a Distancia (UNED)





Proyecto Fin de Máster
Máster en Ingeniería de Sistemas y Control

Clase de proyecto: B - Proyecto específico propuesto por el alumno

Título: Sistema de detección de momentos clave en vídeos mediante el análisis de la interrelacón de señales electroencefalográficas (EEG)

Estudiante: Eduardo Mariano Horna Prat

Directores: Raquel Dormido Canto, Natividad Duro Carralero y Ángel Mur

Todo hombre puede ser, si se lo propone,
escultor de su propio cerebro

Santiago Ramón y Cajal

Sistema de detección de momentos clave en vídeos mediante el análisis de la interrelación de señales electroencefalográficas (EEG)

RESUMEN

La detección de momentos clave en vídeos es un campo de investigación que está creciendo a medida que el vídeo es una herramienta multimedia cada vez más usada para la comunicación entre usuarios, organizaciones o empresas. Este crecimiento se ha visto impulsado por plataformas de vídeo online como Youtube, Vimeo, Dailymotion, etc. Esta detección de momentos de interés está orientada principalmente en dos sentidos: 1) Obtener un etiquetado automático de las características del vídeo de forma que se puede mejorar la clasificación y los sistemas de recomendación automáticos (por ejemplo en plataformas como Netflix). 2) Seleccionar los momento más relevantes en un vídeo para construir previsualizaciones, algo muy importante para captar la posible atención al vídeo por parte de los potenciales usuarios (por ejemplo previews de vídeos en Facebook). La versión más clásica de la aproximación a este problema es mediante la encuesta subjetiva a distintos sujetos. Por otra parte, los sistemas de detección automática actuales están basados en características multimedia del propio vídeo como el ritmo, el movimiento, el audio, etc. En este proyecto se propone afrontar el problema introduciendo las respuestas fisiológicas emocionales (no subjetivas) que se producen en un conjunto de individuos para seleccionar dichos momentos clave. Se intentará buscar los momentos del vídeo que producen una respuesta más coordinada entre un grupo de usuarios y desarrollar un sistema que obtenga dichos momentos clave del vídeo. El trabajo estará dividido en las siguientes fases:

1) Toma de datos de la respuesta electroencefalográfica ante el visionado de una serie de vídeos.

2) Preprocesado de las señales de EEG (series temporales de alta dimensionalidad) mediante algoritmos como ICA (Independent Component Analysis) o ASR (Artifact Subspace Reconstruction) de forma que se pueda crear una serie de filtros automáticos.

3) Implementación de un sistema que modele la respuesta agregada de todas las señales cerebrales de distintos sujetos (mediante algoritmos de relación de series multivariables, ya sea en tiempo o en frecuencia) y que sea capaz de destacar aquellos momentos clave como salida.

4) Evaluación del sistema.

Palabras clave: electroencefalograma, EEG, vídeos, emociones, interrelación

Índice

1. Introducción **1**
1.1. Contexto . . . 1
1.2. Estado del arte . . . 1
1.3. Motivación . . . 3
1.4. Objetivo y Alcance del proyecto . . . 4
1.5. Organización . . . 5
1.6. Herramientas . . . 5

2. EEG y emociones **6**
2.1. Introducción . . . 6
2.2. Señal EEG . . . 6
2.3. Experimento . . . 8
2.4. Adquisición de los datos . . . 9
2.5. Emociones . . . 10
2.5.1. Teorías de las emociones . . . 11
2.5.2. Mapa emocional y estímulos elegidos . . . 11

3. Preprocesado **13**
3.1. Introducción . . . 13
3.2. Referenciado . . . 13
3.3. Filtro paso banda . . . 15
3.4. ASR . . . 16
3.5. ICA . . . 17
3.6. MARA . . . 18

4. Relación entre sujetos **21**
4.1. Introducción . . . 21
4.2. Métodos temporales . . . 21
4.2.1. mCCA . . . 21
4.2.2. cERPv1 y cERPv2 . . . 23
4.3. Métodos frecuenciales . . . 25
4.4. Multimedia . . . 28

5. Resultados **29**
5.1. Introducción 29
5.2. Momentos clave del vídeo 29
5.3. Métodos temporales 29
5.3.1. mCCA 29
5.3.2. cERPv1 y cERPv2 30
5.4. Métodos frecuenciales 32
5.5. Multimedia 34

6. Conclusiones y líneas futuras **67**

Bibliografía **69**

Índice de figuras

1.1. Esquema del problema 3
1.2. Esquema de desarrollo del sistema 4

2.1. Tipos de señal EEG 8
2.2. Esquema del protocolo 9
2.3. Electrodos utilizados 10
2.4. Ejemplo de señal EEG 10
2.5. Modelo emocional de Russell 12
2.6. Posicionamiento emocional de los estímulos elegidos 12

3.1. Cadena de filtros 13
3.2. Referenciado EEG 14
3.3. Respuesta frecuencial del filtro paso banda 15
3.4. Filtrado paso banda EEG 16
3.5. Filtrado ASR 17
3.6. Resultado del filtrado ASR 17
3.7. Filtrado ICA 18
3.8. Componentes ICA 20
3.9. Características MARA 20

4.1. Esquema del sistema mCCA 23
4.2. Esquema del sistema cERPv1 25
4.3. Esquema del sistema cERPv2 26
4.4. Objetivo del algoritmo SSD 27
4.5. Esquema del sistema cSPoC 28

5.1. Resultado mCCA vídeo REC 35
5.2. Resultado mCCA vídeo CocaCola 36
5.3. Resultado mCCA vídeo Esquí 36
5.4. Resultado mCCA vídeo Prinzenzeit 37
5.5. Resultado mCCA vídeo Paranormal 37
5.6. Resultado cERPv1 vídeo REC 38
5.7. Resultado cERPv2 vídeo REC 38
5.8. Resultado cERPv1 vídeo CocaCola 39
5.9. Resultado cERPv2 vídeo CocaCola 39

5.10. Resultado cERPv1 vídeo Esquí 40
5.11. Resultado cERPv2 vídeo Esquí 40
5.12. Resultado cERPv1 vídeo Prinzenzeit 41
5.13. Resultado cERPv2 vídeo Prinzenzeit 41
5.14. Resultado cERPv1 vídeo Paranormal 42
5.15. Resultado cERPv2 vídeo Paranormal 42
5.16. Activaciones cerebrales cSPoC vídeo REC banda alfa 43
5.17. Envolvente banda alfa vídeo REC 43
5.18. Resultados cSPoC vídeo REC banda alfa 44
5.19. Activaciones cerebrales cSPoC vídeo REC banda delta 44
5.20. Envolvente banda delta vídeo REC 45
5.21. Resultados cSPoC vídeo REC banda delta 45
5.22. Activaciones cerebrales cSPoC vídeo REC banda theta 46
5.23. Envolvente banda theta vídeo REC 46
5.24. Resultados cSPoC vídeo REC banda theta 47
5.25. Activaciones cerebrales cSPoC vídeo CocaCola banda alfa 47
5.26. Envolvente banda alfa vídeo CocaCola 48
5.27. Resultados cSPoC vídeo CocaCola banda alfa 48
5.28. Activaciones cerebrales cSPoC vídeo CocaCola banda delta 49
5.29. Envolvente banda delta vídeo CocaCola 49
5.30. Resultados cSPoC vídeo CocaCola banda delta 50
5.31. Activaciones cerebrales cSPoC vídeo CocaCola banda theta 50
5.32. Envolvente banda theta vídeo CocaCola 51
5.33. Resultados cSPoC vídeo CocaCola banda theta 51
5.34. Activaciones cerebrales cSPoC vídeo Esquí banda alfa 52
5.35. Envolvente banda alfa vídeo Esquí 52
5.36. Resultados cSPoC vídeo Esquí banda alfa 53
5.37. Activaciones cerebrales cSPoC vídeo Esquí banda delta 53
5.38. Envolvente banda delta vídeo Esquí 54
5.39. Resultados cSPoC vídeo Esquí banda delta 54
5.40. Activaciones cerebrales cSPoC vídeo Esquí banda theta 55
5.41. Envolvente banda theta vídeo Esquí 55
5.42. Resultados cSPoC vídeo Esquí banda theta 56
5.43. Activaciones cerebrales cSPoC vídeo Prinzenzeit banda alfa 56
5.44. Envolvente banda alfa vídeo Prinzenzeit 57
5.45. Resultados cSPoC vídeo Prinzenzeit banda alfa 57
5.46. Activaciones cerebrales cSPoC vídeo Prinzenzeit banda delta 58
5.47. Envolvente banda delta vídeo Prinzenzeit 58
5.48. Resultados cSPoC vídeo Prinzenzeit banda delta 59
5.49. Activaciones cerebrales cSPoC vídeo Prinzenzeit banda theta 59
5.50. Envolvente banda theta vídeo Prinzenzeit 60
5.51. Resultados cSPoC vídeo Prinzenzeit banda theta 60
5.52. Activaciones cerebrales cSPoC vídeo Paranormal banda alfa 61
5.53. Envolvente banda alfa vídeo Paranormal 61

5.54. Resultados cSPoC vídeo Paranormal banda alfa 62
5.55. Activaciones cerebrales cSPoC vídeo Paranormal banda delta 62
5.56. Envolvente banda delta vídeo Paranormal 63
5.57. Resultados cSPoC vídeo Paranormal banda delta 63
5.58. Activaciones cerebrales cSPoC vídeo Paranormal banda theta 64
5.59. Envolvente banda theta vídeo Paranormal 64
5.60. Resultados cSPoC vídeo Paranormal banda theta 65
5.61. Resultado análisis multimedia REC 65
5.62. Resultado análisis multimedia CocaCola 65
5.63. Resultado análisis multimedia Esquí 66
5.64. Resultado análisis multimedia Prinzenzeit 66
5.65. Resultado análisis multimedia Paranormal 66

Índice de tablas

2.1. Tabla de ritmos cerebrales 8

5.1. Secuencias del vídeo REC 30
5.2. Secuencias del vídeo CocaCola 31
5.3. Secuencias del vídeo Esquí 32
5.4. Secuencias del vídeo Prinzenzeit 33
5.5. Secuencias del vídeo Paranormal 34

Lista de acrónimos

ASR Artifact Subspace Reconstruction - Reconstrucción del subespacio de artefactos

BVP Blood Volume Pulse - Volumen de pulso sanguíneo

CAR Common Average Reference - Referencia media común

CCA Canonical Component Analysis - Análisis de componentes canónicas

cSPoC Canonical Source Power Correlation Analysis - Análisis de la correlación de las fuentes de potencia canónicas

EDA ElectroDermal Activity - Actividad electrodérmica

EEG Electroencefalograma

ERP Event-related Potentials - Potenciales relacionados con un evento

FIR Finite Impulse Response - Respuesta al impulso finita

IAF Individualized Alpha Frequency - Frecuencia alfa individualizada

ICA Independent Component Analysis - Análisis de componentes independientes

LPM Linear Programming Machine - Máquina de programación lineal

MARA Multiple Artifact Rejection Algorithm - Algoritmo para eliminar múltiples artefactos

PCA Principal Component Analysis - Análisis de componentes principales

SNR Signal to Noise Ratio - Relación señal a ruido

SSD Spatio-Spectral Decomposition - Descomposición espacio-frecuencial

1. Introducción

1.1. Contexto

Este proyecto fin de máster se ha llevado a cabo dentro del Máster de Ingeniería de Sistemas y Control de la Universidad Nacional de Educación a Distancia (UNED).

1.2. Estado del arte

Los contenidos multimedia están invadiendo todos los ámbitos de nuestra vida. Dentro de ellos, los vídeos se han convertido en una forma más rápida y directa de transmitir información que los medios impresos, que eran el pincipal canal de transmisión hasta la revolución digital. Para una empresa (o cualquier organización) un vídeo permite transmitir de una forma concisa unos valores y productos que llegan al consumidor rápidamente. Como la cantidad de información a la que tenemos acceso cada día es mayor, un vídeo condensa de forma más eficaz una cantidad de información que necesita de muchas palabras para definirlo.

Todos nuestros dispositivos digitales de uso cotidiano como televisión, ordenador, tablets o móviles son capaces de gestionar y procesar vídeos. Anuncios, series, películas, charlas, noticias, casi toda la información digital va siempre acompañada de un vídeo o directamente su formato es audiovisual. La cantidad de información que recibimos diariamente es mayor que la que nuestro cerebro puede procesar, y por lo tanto la gente prefiere un documento gráfico y multimedia a un texto plano ya que condensa mucha información en un tiempo más breve.

Con el desarrollo de internet se han creado plataformas para compartir y distribuir vídeos como pueden ser Youtube o Netfilx. Conocer los gustos y preferencias de los consumidores se ha vuelto una tarea necesaria ya que de la cantidad de posibles estímulos que una persona recibe, nos interesa mostrar en primer lugar aquellos que se ajusten más a las necesidades y gustos de cada individuo. Por ejemplo Google utiliza los datos que son cedidos por los usuarios para poder adaptar los anuncios que introduce en Youtube a las características y gustos de cada persona. Netflix también utiliza los datos que recoge

de la forma de visionado y los gustos de los usuarios para recomendarles aquellas series o películas que cree que más encajan con sus preferencias.

Hasta la aparición del big data y la posibilidad de recoger información de forma online, la forma más común de conocer los efectos de los vídeos que se emiten era mediante el uso de encuestas. Estas respuestas por parte de los usuarios son subjetivas ya que son resultado de un procesado consciente antes de ser respondidas. El avance de las técnicas para registrar señales fisiológicas está aumentado el interés por conocer qué respuestas inconscientes se producen ante nuevos estímulos. En un mundo tan competitivo como por ejemplo la publicidad, cada segundo cuenta, y por lo tanto nos puede interesar acortar aquellas zonas de nuestro vídeo que no produzcan interés en los usuarios. Por ello surge la necesidad de poder conocer de antemano si un estímulo audiovisual va a provocar las reacciones que se desean en la fase de diseño. Sería de una gran utilidad para empresas de investigación de mercados, para diseñadores de spots publicitarios o para creadores de películas o series.

En los últimos años ha aumentado el número de estudios relacionados con la medición de variables emocionales ante estímulos externos [1]. Se contempla como una forma de acceder a las respuestas inconscientes de los usuarios. Estas respuestas están relacionadas con la carga afectiva o las emociones que suscitan. Este nuevo campo se ha pasado a conocer como neuromarketing, que consiste en aplicar conocimientos del campo de las neurociencias a ámbitos de marketing como pueden ser el branding, el packaging, el diseño de producto o la publicidad. En este proyecto buscaremos un sistema que ayude a seleccionar momentos clave, lo cual es muy interesante desde el punto de vista de publicidad audiovisual o creación de resúmenes de un vídeo de mayor duración.

Esta aproximación psicofisiológica a la medición de emociones se basa principalmente en conocer las respuestas corporales que se producen ante estímulos emocionales externos. Principalmente se basa en la medición de las respuestas del sistema nervioso, ya sean del sistema periférico (EDA, BVP) o del sistema central (EEG). En el presente proyecto nos centraremos en el EEG ya que es uno de los actores principales en la creación de emociones y sentimientos. Una propiedad de las variables fisiológicas (y entre ellas el EEG) es su alta variabilidad entre sujetos, por lo que la respuesta de uno sólo de ellos no es suficiente para conocer qué provoca un vídeo en una persona. Muy recientemente se ha abierto un nuevo campo de investigación conocido como hyperscanning que consiste en usar la información de cerebros de distintos usuarios para buscar los patrones comunes y de esta forma llegar a conclusiones acerca de qué partes del cerebro están involucradas en qué tareas concretas. Sin embargo su aplicación con contenido multimedia ha sido muy reducida. Por lo tanto en este proyecto se intentará usar la información conjunta de varios cerebros mientras visionan el mismo estímulo para buscar momentos de máxima concordancia que nos llevará a detectar momentos clave en los vídeos, lo cual podría ser usado en el diseño de campañas de publicidad, en la creación de escenas para series o películas o en la creación de resúmenes de vídeos para plataformas como Facebook, Netflix, Youtube, etc.

1.3. Motivación

No existe mucha literatura acerca de conocer las respuestas afectivas mediante el registro de las señales fisiológicas comunes entre varios observadores de un estímulo visual. Como hemos comentado anteriormente, el conocer qué momentos son clave en un vídeo, puede ser algo útil por ejemplo para decidir cómo montar un anuncio, qué partes quitar de una película o conocer cuáles son los instantes que más llaman la atención a un grupo de observadores. Incluso podría ser interesante para poder clasificar los vídeos de acuerdo a su nivel emocional. En resumen, la primera motivación es el usar información fisiológica objetiva para poder mejorar el diseño y guiar posibles modificaciones de estímulos de vídeo.

Por otra parte el desarrollo de herramientas matemáticas para conocer la respuesta fisiológica común del cerebro entre distintos sujetos se ha estado desarrollando en los últimos años. Usar estas herramientas para el análisis emocional de estímulos es algo que no se ha hecho en profundidad [2, 3, 4] , hasta donde llega el conocimiento del autor, y que puede ser interesante explorar (figura 1.1).

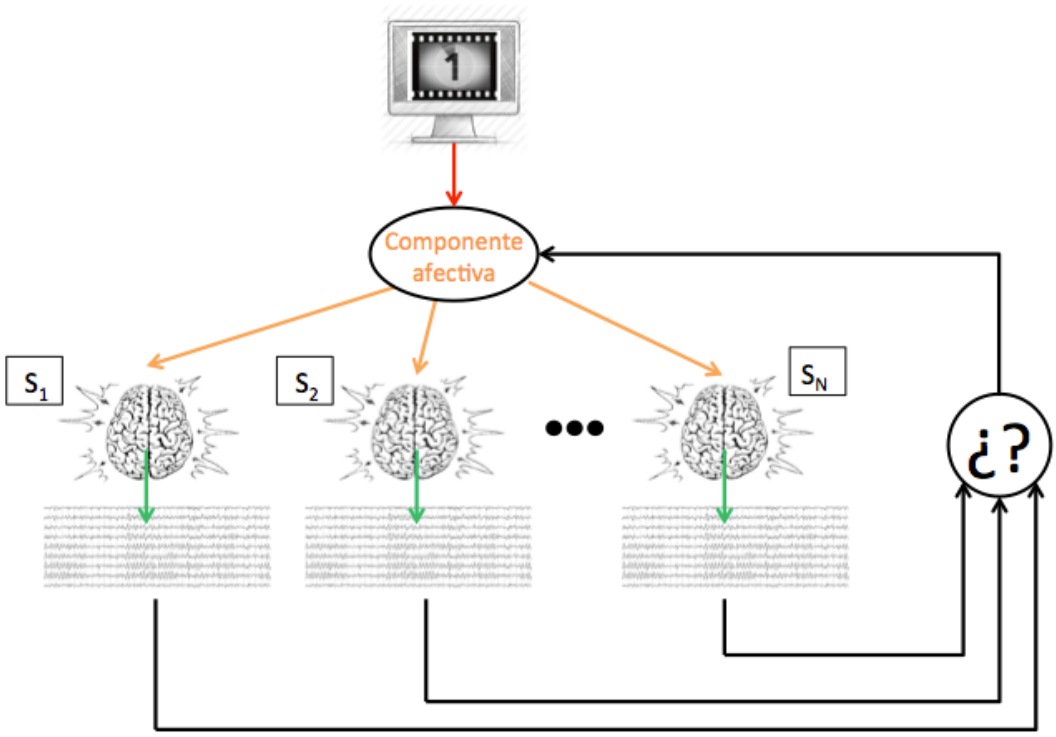


Figura 1.1: Esquema del problema a resolver

1.4. Objetivo y Alcance del proyecto

El objetivo de este trabajo fin de máster es el desarrollo de un sistema automático que nos permita detectar aquellos momentos que son de especial relevancia emocional para un conjunto de sujetos que observan el mismo estímulo. Para ello se llevan a cabo otros objetivos más específicos como pueden ser:

- Diseño de un experimento exploratorio para poder analizar la respuesta fisiológica electroencefalográfica de distintas personas al mismo estímulo.
- Grabación de los datos de acuerdo al protocolo diseñado.
- La capacidad de filtrar el ruido en una señal de muy pequeña amplitud y fácilmente contaminada como puede ser el EEG de una forma totalmente automática.
- Aplicar herramientas de interrelación entre sujetos usando filtros lineales con señales multivariantes.

El proceso para conseguir estos objetivos se expone en la figura 1.2. Las distintas partes del sistema serán explicadas a lo largo de esta memoria del modo comentado en el apartado 1.5:

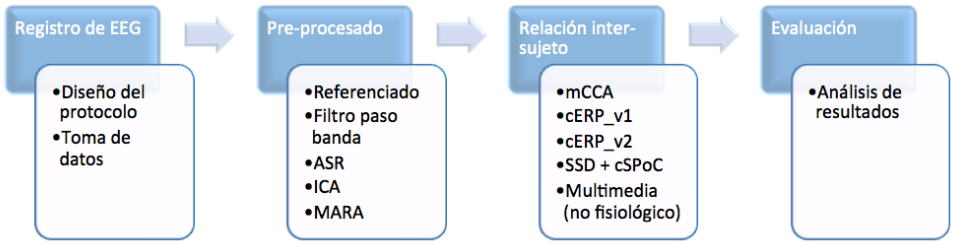


Figura 1.2: Partes en las que se divide el sistema diseñado.

1.5. Organización

La memoria del trabajo fin de máster está dividida en seis capítulos, de los cuales este primero es el de introducción, donde se presenta la motivación y objetivo del proyecto. A partir del capítulo 2 y hasta el capítulo 5 se explicarán cada una de las partes del sistema de detección de momentos clave que se ajustan al diagrama expuesto en la figura 1.2:

- **Capítulo 2 Señal de EEG y emociones**: en esta sección se explicarán el protocolo seguido y la grabación de los datos así como conceptos básicos de la señal EEG registrada, y una breve aproximación a las teorías emocionales.
- **Capítulo 3 Prefiltrado**: en el mismo se presentan los diferentes pasos seguidos para la eliminación automática de artefactos, explicando cada filtro usado para limpiar la señal de EEG.
- **Capítulo 4 Métodos inter-sujeto**: en este apartado se presentan los distintos métodos usados para buscar la relación en la respuesta fisiológica de los distintos sujetos cuando observan un estímulo común.
- **Capítulo 5 Resultados**: se presentan los resultados obtenidos para los distintos métodos y se hace una evaluación de los mismos.

Por último en el capítulo 6 se expondrán las conclusiones que resumen el trabajo realizado y se proponen líneas de investigación futuras para continuar con el desarrollo del sistema.

1.6. Herramientas

Para la ejecución del presente proyecto se utilizó el software MATLAB© desarrollado por Mathworks. La grabación de los datos y el diseño del protocolo se realizó en un software propio de la empresa Bit&Brain Technologies. El registro de los datos es explicado en el capítulo 2 de esta memoria.

2. EEG y emociones

2.1. Introducción

En este capítulo se explican algunos conceptos básicos sobre la señal electroencefalográfica como pueden ser su naturaleza y origen fisiológico así como sus principales características tanto en el dominio temporal como en el frecuencial (sección 2.2). Se expone un breve resumen de cómo se llevó a cabo la toma de datos. En primer lugar se explica el protocolo utilizado (sección 2.3) y a continuación las herramientas usadas para la grabación de los datos (sección 2.4). Por otro lado se hace un breve resumen del concepto de emoción, las diferentes teorías emocionales y dónde se encuentran los estímulos elegidos en este proyecto dentro de un mapa emocional (sección 2.5).

2.2. Señal EEG

El electroencefalograma (EEG) es el nombre que se le da a la señal registrada de forma no invasiva que recoge la actividad eléctrica cerebral mediante una serie de sensores o electrodos colocados sobre el cuero cabelludo. Las oscilaciones del EEG son el resultado de la actividad de grandes poblaciones de neuronas. Las neuronas son el elemento principal del sistema nervioso. Son células nerviosas excitables eléctricamente que reciben, procesan y envían información. La actividad de una sola neurona cortical no puede ser medida debido al grosor de las capas de tejido (fluidos, huesos y piel) que componen la corteza cerebral y que provocan una atenuación de la señal propagada hasta el electrodo. Sin embargo la actividad conjunta de millones de neuronas corticales, localizadas a una profundidad de varios milímetros, producen un campo eléctrico suficientemente fuerte para ser medido en la superficie cerebral, siendo esta profundidad dependiente de la potencia de la fuente neuronal.

El campo eléctrico es principalmente generado por corrientes que fluyen durante exci-

taciones sinápticas[1] de las dendritas[2], la excitación post-sináptica. La señal registrada es no estacionaria y de carácter oscilatorio por lo que es común hablar de ritmos cerebrales. La diversidad de ritmos EEG es inmensa y depende, entre otras cosas, del estado mental del sujeto (grado de atención, de relajación, etc). Los ritmos están convencionalmente caracterizados por:

- **La amplitud de EEG**: está relacionada con el sincronismo con el cual las neuronas de la corteza cerebral interaccionan. Una excitación sincronizada produce una señal de mayor amplitud medible en el cuero cabelludo ya que las señales originadas por cada neurona se suman de forma coherente en fase produciendo como resultado una señal de amplitud elevada. Por el contrario, una excitación asíncrona de las neuronas provoca un EEG irregular de baja amplitud. La amplitud es variable, de 20-60 μV (50 μV por término medio).
- **La frecuencia de EEG**: viene definida por el comportamiento rítmico de las interacciones coordinadas de activación de las neuronas de la corteza. [5]

En la actualidad, el registro de EEG se lleva a cabo mediante la colocación de un gorro sobre el cuero cabelludo, en el cual se colocan un conjunto de electrodos. La disposición de los electrodos sobre el cuero cabelludo sigue el estándar descrito por el Sistema Internacional 10-10 o 10-20. Cada par de electrodos registra las variaciones de potencial eléctrico definiendo un canal de EEG. Cada una de estas señales son amplificadas mediante el uso de una amplificador diferencial previo a la etapa de digitalización. La figura 2.1 (izda.) representa una muestra de encefalograma compuesta por 30 canales.

La actividad eléctrica cerebral es normalmente conocida como ritmo porque las señales medidas a menudo presentan un comportamiento oscilatorio y repetitivo como hemos podido comprobar en la figura 2.1 (izda). Los ritmos cerebrales son normalmente clasificados en cinco bandas de frecuencia diferentes. No es la única clasificación pero es la más utilizada por su utilidad clínica. Esta clasificación se puede observar en la tabla 2.1.

La investigación con EEG se centra en dos tipos de señales: EEG espontáneo, también conocido como EEG continuo y potenciales evocados relacionados con un evento (ERP). En este proyecto vamos a usar ambos. El EEG espontáneo es simplemente la actividad cerebral a lo largo del tiempo. Casi toda la potencia de la señal se centra entre 1 Hz. y 40 Hz. como se ha visto en la tabla anterior. Por otro lado los potenciales evocados son el resultado de promediar respuestas neuronales al mismo estímulo externo. Es por lo tanto una actividad no espontánea y forzada por el protocolo. Se verá en más detalle en la sección 4.2.2 donde se explica un método para intentar capturar dicho proceso neuronal. En la figura 2.1 (dcha.) vemos en la primera columna distintas repeticiones de la respuesta

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

[2]Dendrita: Prolongación ramificada de una célula nerviosa, mediante la que ésta recibe estímulos externos

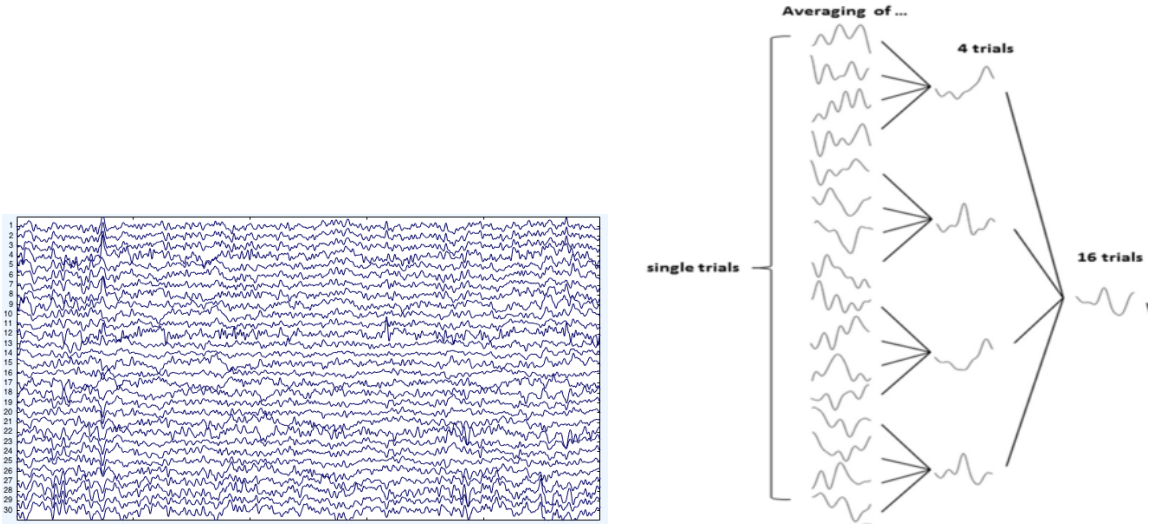


Figura 2.1: (Izquierda) Ejemplo de señal EEG espontánea (Derecha) Ejemplo de formación de un ERP

Ritmo	Frecuencias
Delta	< 4 Hz.
Theta	4-7 Hz.
Alpha	8-13 Hz.
Beta	14-30 Hz.
Gamma	> 30 Hz.

Tabla 2.1: Tabla de ritmos cerebrales

a un estímulo. Como se puede observar no se aprecia un patrón claro, pero cuando vamos promediando las repeticiones llegamos a esa forma ondulatoria que es la típica de un potencial evocado.

2.3. Experimento

32 voluntarios participaron en la toma de datos con edades comprendidas entre los 18 y los 45 años. Todos eran diestros.

Los sujetos se sentaron en una silla confortable delante del montaje experimental. El experimento constaba de 2 fases (figura 2.2):

- Fase de registro de EEG basal: en esta fase se grababan 2 minutos con el sujeto con los ojos cerrados y 2 minutos con el sujeto con los ojos abiertos mirando a un punto fijo. De esta forma conocemos cómo es el EEG individualizado para cada persona

ya que como hemos comentado anteriormente, esta señal varía considereablemente entre sujetos.

- Fase de estimulación: en esta fase los sujetos visionaban 5 vídeos, aleatorizados entre sujetos. Cada trial empezaba con una cruz que permanecía en la pantalla durante 3 segundos y después comenzaba la reproducción del vídeo. Después de cada vídeo había un periodo de descanso de duración indeterminada para que los sujetos pudieran descansar si lo necesitaban.

Este proceso duraba en torno a 15 minutos sin contar con otros 15-20 minutos de colocación del gorro con los sensores de EEG previo a la toma de datos. La única instrucción que se les daba a los participantes era que visionaran los vídeos libremente y que minimizasen en la medida de lo posible los movimientos bruscos. Como hemos visto en la sección 2.2 la amplitud de la señal EEG es muy pequeña por lo que se ve afectada considerablemente por parpadeos o movimientos musculares, lo que es conocido como artefactos de la señal.

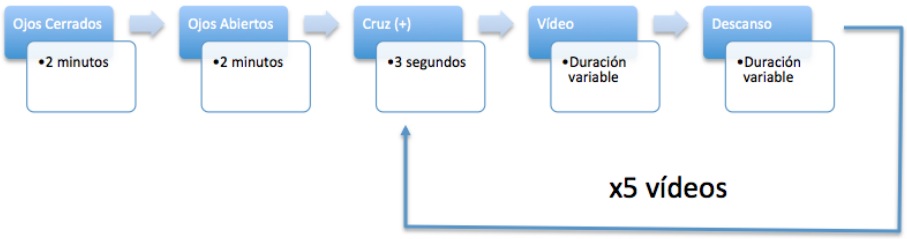


Figura 2.2: Esquema del protocolo experimental

2.4. Adquisición de los datos

La actividad EEG fue registrada con un amplificador REFA de la marca holandesa TMSI. Se registró la actividad de 32 electrodos distribuidos de acuerdo con el sistema internacional 10/10 (Figura 2.3). La tierra estaba colocada en la muñeca izquierda mediante una muñequera y la referencia en los dos lóbulos de ambas orejas. Los canales de EEG grabados corresponden, en el siguiente orden, desde el canal 1 al canal 32 con las siguientes posiciones: Fp1, Fpz, Fp2, F7, F3, Fz, F4, F8, Fc5, Fc1, Fc2, Fc6, T7, C3, Cz, C4, T8, Cp5, Cp1, Cp2, Cp6, P7, P3, Pz, P4, P8, Poz, O1, Oz, O2, oreja izquierda y oreja derecha. Las señales EEG fueron digitalizadas con una frecuencia de muestreo de 256 Hz. No se usó ningún filtro adicional en la toma de datos por lo que la señal nos llegó totalmente cruda. Como se comentará en el capítulo siguiente la señal registrada por este dispositivo aplica un filtro por hardware conocido como CAR (Common Average Reference) o filtro de referencia media común que resta a cada canal grabado la media

de los 32 canales. Por ello cuando se comenta que se usarán las orejas como referencias es porque al final se referenciará a las mismas aunque la señal grabada tenga otro tipo de referencia.

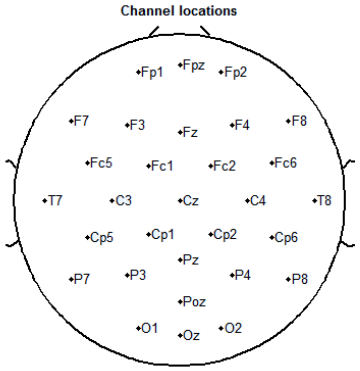


Figura 2.3: Mapa craneal con la posición de los electrodos utilizados para la adquisición de la señal EEG

Se usó un software propio de la empresa Bit&Brain Technologies para la toma de datos (figura 2.4). Toda la sincronización entre los estímulos y la señal fue recogida en ficheros que contenían el EEG usado en este proyecto.

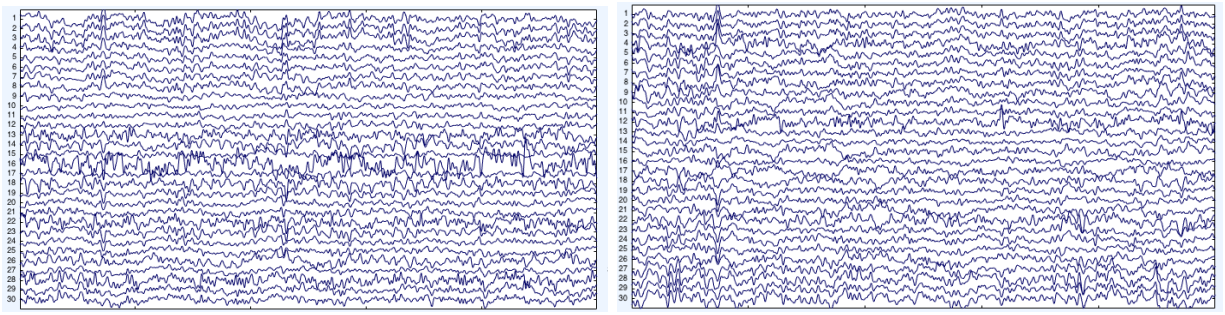


Figura 2.4: Fragmentos de señal EEG de dos sujetos del experimento

2.5. Emociones

Los términos emoción y sentimiento están estrechamente relacionados. Emoción se refiere a los cambios bruscos y puntutales que se producen en nuestro organismo como respuesta a un evento externo. Tienen caracter evolutivo y de supervivencia. De este modo, por ejemplo, si vemos un animal peligroso el cuerpo reacciona, sube la frecuencia cardiaca, aumenta la sudoración y tensa los músculos para preparar una posible huida. Nuestra especie, el Homo Sapiens, sin embargo, ha desarrollado una gran capa cerebral conocida como córtex que permite dar un significado y un sentido a dichas emociones y convertirlas en lo que conocemos como sentimientos, que no son más que las representaciones conscientes de las emociones que estamos sintiendo. La relación entre emociones

(respuestas fisiológicas) y sentimientos por lo tanto es muy estrecha y ha suscitado numerosa literatura [6, 7].

2.5.1. Teorías de las emociones

En esta sección analizaremos brevemente las principales teorías afectivas o de las emociones [8]. Las primeras teorías emocionales surgen a finales del siglo XIX. Darwin escribió en su libro *The expression of emotions in man and animals* [9] que existían ciertas emociones básicas que compartíamos con otras especies. Casi al mismo tiempo, William James escribió su famoso artículo What is an emotion [10] donde expone una teoría que hasta el día de hoy es una de las más aceptadas. En ella James expone que las emociones son fruto del análisis de los cambios fisiológicos que se producen en el cuerpo ante un evento externo. Así, por ejemplo, cuando vemos una amenaza sube el ritmo cardiaco y el reconocimiento de este cambio es lo que conocemos como la emoción del miedo. Esta misma idea fue propuesta por Lange [11] por lo que se conoce como la teoría de James-Lange. A principios del siglo XX Cannon y Bard escribieron un artículo donde exponían una crítica a la teoría de James-Lange [12] . Como ejemplo exponían que hay varias emociones que pueden provocar los mismos cambios fisiológicos. Y que no sólo los cambios fisiológicos son los que determinan la emoción sino que también entra en juego la parte cognitiva. Basados en estas teorías se han creado varios modelos para la clasificación de emociones. Los hay unidimensionales como por ejemplo el modelo de las emociones básicas de Ekman [13, 14] o bidimensionales como el modelo circunflejo de Russell [15] o Plutchnick [16]. En este proyecto haremos uso del modelo de Russell ya que permite modelar un rango amplio de emociones creando un mapa bidimensional que presentaremos en la siguiente sección.

2.5.2. Mapa emocional y estímulos elegidos

Profundizaremos un poco más en el modelo de emociones propuesto por Russell ya que es el más usado dentro del campo de la detección de emociones mediante señales fisiológicas. El motivo es que divide de forma simple e intuitiva las emociones con respecto a dos ejes. El eje de abscisas está constituido por lo que se conoce como valencia, que no es más que el grado de positividad/acercamiento o negatividad/rechazo que provoca un estímulo. Por lo tanto su rango va desde lo muy negativo (más a la izquierda) a lo más positivo (más a la derecha). El eje de ordenadas está formado por la dimensión conocida como arousal que es el grado de excitación que provoca un estímulo. Su rango va desde lo más calmado (abajo) a una excitación muy elevada (más arriba). En la figura 2.5 se puede ver un ejemplo del mapa bidimiensional de Russell donde se han proyectado distintas emociones discretas.

Se realizó una pequeña encuesta entre los trabajadores de la compañía (los cuales no participaron en el experimento) para conocer previamente si los vídeos elegidos se

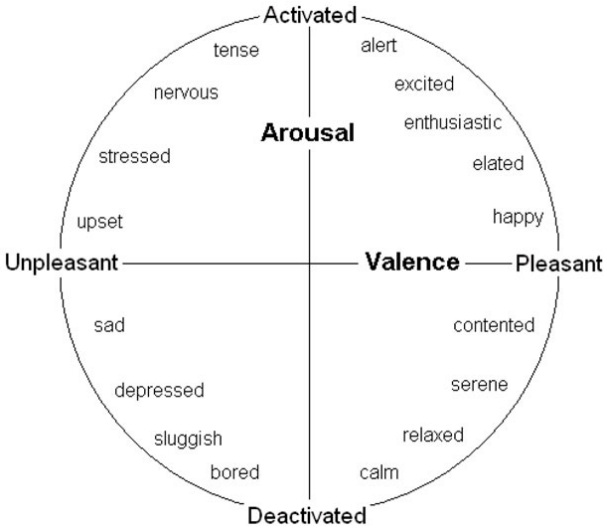


Figura 2.5: Modelo emocional de Russell

posicionaban en los cuadrantes deseados. Se eligieron más vídeos de alto arousal (4) que de bajo arousal (1) ya que nos interesaba crear un sistema que detecte aquellos puntos con un mayor impacto o relevancia común entre todos los sujetos. En la figura 2.6 se encuentran los vídeos en las posiciones marcadas por la encuesta. Se puede observar que tenemos 4 vídeos de alto arousal, 2 de ellos con valencia positiva y 2 de ellos con valencia negativa. Por otro lado tenemos un vídeo de bajo arousal y valencia positiva. Cabe destacar de este modelo que no ocupa todo el espacio bidimiensional dado que es muy complicado tener, por ejemplo, un estímulo muy negativo pero que no nos produzca excitación. Los nombres usados en la figura 2.6 serán los usados durante el resto de la memoria para referirnos a los distintos estímulos.

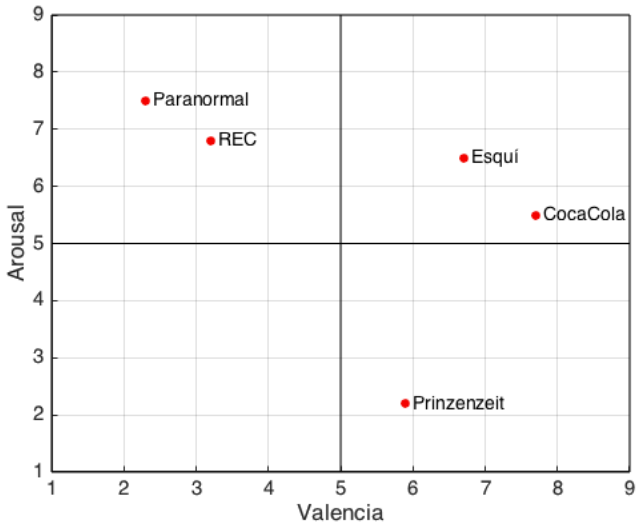


Figura 2.6: Posicionamiento emocional de los estímulos elegidos

3. Preprocesado

3.1. Introducción

La señal EEG tiene una amplitud del orden de los microvoltios por lo que se ve afectada por ruido y artefactos. Los más comunes son interferencias de la propia fisiología como pueden ser los movimientos oculares o la actividad eléctrica de los músculos del cuello. Ambas pueden tener una amplitud que duplica o triplica la amplitud de la actividad registrada por las neuronas. Por ello, siempre que se trabaja con electroencefalografía hace falta una primera fase para preprocesar los datos. En este trabajo propondremos una cadena de filtros (figura 3.1) que permitirán tener una señal limpia para trabajar posteriormente y además este procesado será totalmente automático. El capítulo está dividido en varias secciones donde se comentarán cada uno de los pasos, explicando su función y la necesidad de los mismos.

Antes de pasar por la serie de filtros, todo el EEG grabado de cada sujeto se concatenó en una única señal de 32 canales por el tiempo total del experimento. Una vez filtrada la señal se cortó de acuerdo al protocolo, quedando al final 5 vídeos limpios y filtrados de duración igual al vídeo más tres segundos previos que correspondían con la presentación de la cruz.

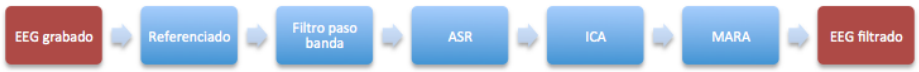


Figura 3.1: Serie de filtros para el preprocesado automático

3.2. Referenciado

El hardware usado para la grabación de datos utiliza un filtrado electrónico de referencia común o CAR por sus siglas en inglés (Common Average Reference). Esto quiere decir que a cada uno de los canales (en este caso 32 canales) se le sustrae la media de

todos ellos. Por lo tanto, para recuperar la actividad específica registrada por cada sensor hace falta referenciar la señal a otro punto. Para este propósito se registraron dos canales con la actividad eléctrica en las orejas (que debería ser nula o casi nula). Referenciaremos cada canal a la media de los dos canales de las orejas. Matemáticamente la operación para recuperar la actividad de un sensor quedaría así. Sea X_{i-hw} la actividad registrada por el dispositivo para el canal i y sea X_{media} la actividad media de todos los canales. Podemos reescribir la actividad registrada por el hardware para el canal i como $X_{i-hw} = X_{i-real} - X_{media}$. Sea $X_{ear1-hw}$ la señal grabada en la oreja izquierda y sea $X_{ear2-hw}$ la señal grabada en la oreja derecha. Usaremos como nueva referencia el promedio de ambas señales, es decir, $X_{ear-hw} = (X_{ear1-hw} + X_{ear2-hw})/2$ para que sea una medida menos contaminada por algún ruido. De forma más general, la actividad grabada en las orejas sería $X_{ear-hw} = X_{ear-real} - X_{media}$. Para hacer el referenciado a las orejas, restamos la actividad registrada en el canal i menos la actividad grabada en las orejas y nos quedará $X_{i-hw} - X_{ear-hw} = X_{i-real} - X_{media} - (X_{ear-real} - X_{media}) = X_{i-real} - X_{ear-real}$. Si consideramos que la señal cerebral en las orejas es nula, haciendo esta resta o referenciado obtenemos la señal real en dicho punto del cuero cabelludo. En la figura 3.2 podemos ver el resultado[1] de referenciar las señales. Se puede ver claramente cómo el parpadeo (distorsión observada en torno al segundo 32.5 en la figura 3.2), antes de referenciar, se mete en todos lo canales y después de referenciar lo recuperamos únicamente en la zona frontal (canales del 1 al 12) que es donde se debería ver con mayor intensidad ya que la distorsión que produce el ojo es más próxima a dichos canales. Se observa también cómo los dos canales últimos (oreja izquierda y derecha) no contienen información de EEG y por tanto son eliminados después de referenciar.

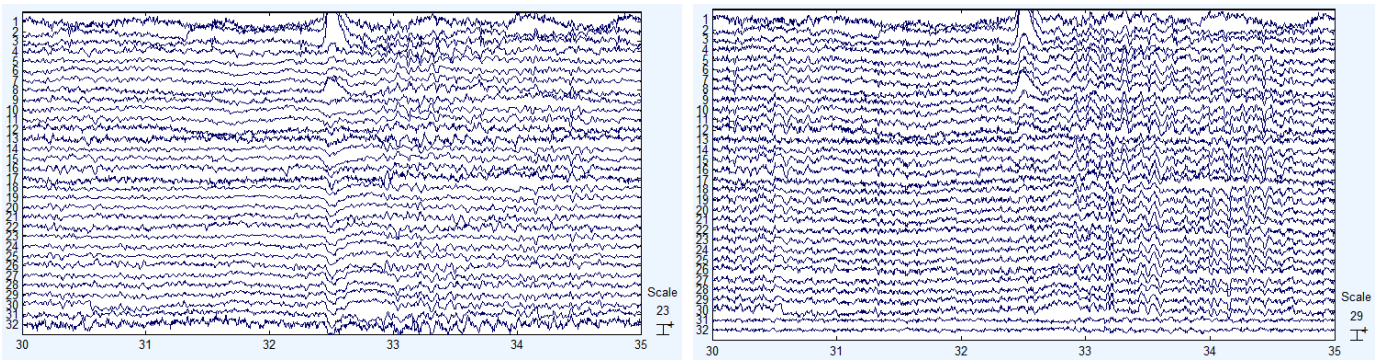


Figura 3.2: Efecto de referenciar el EEG. Antes de referenciar (izda.) y después de referenciar (dcha.)

[1]Para una mejor visualización del resultado se ha filtrado la señal cruda registrada por el dispositivo con un filtro paso alto a partir de 0.5 Hz. De esta forma se elimina la señal continua y podemos ver el efecto con mayor claridad. Esto se ha hecho sólo para poder representar la utilidad del referenciado pero en el procesado del proyecto se referenció directamente sobre la señal cruda.

3.3. Filtro paso banda

La señal fue grabada en crudo por lo que el siguiente paso en el preprocesado es eliminar las componentes frecuenciales que no nos son de interés. La señal continua (frecuencia = 0 Hz.) no lleva información neuronal por lo que filtraremos por encima de 1 Hz. para eliminar dicha componente. Además, como hemos visto en el capítulo anterior, la mayor parte de la información frecuencial del EEG viene en las bandas por debajo de 39 Hz. Así que estableceremos este límite como frecuencia superior. La banda gamma es muy propensa a artefactos y por lo tanto no es de interés en este trabajo. Con este filtrado además eliminaremos la distorsión a la frecuencia de la línea eléctrica, en nuestro caso de 50 Hz. al ser la grabación realizada en España. Para este filtrado haremos uso de un filtro FIR (o filtro de respuesta finita al impulso) de orden 876. Este tipo de filtros permiten una respuesta en fase lineal (figura 3.3 abajo) y por lo tanto se evita la distorsión de fase (distintos retrasos para frecuencias distintas) pero necesitan de un orden alto por lo que son más costosos computacionalmente hablando. Las frecuencias de corte son las explicadas anteriormente (1 Hz. - 39 Hz.). En la figura 3.3 podemos ver la respuesta frecuencial del filtro utilizado. Este filtro paso banda va a diminuir la potencia de la señal en las bandas frecuenciales correspondientes a las bandas eliminadas por el filtro dejando pasar aquellas componentes frecuenciales que corresponden a la banda de paso (figura 3.3 arriba). El resultado del filtrado se puede ver en la figura 3.4, donde se observa[2] cómo ahora la señal está mucho más limpia sobre todo de ruido de alta frecuencia, la cual no proviene de fuentes neuronales.

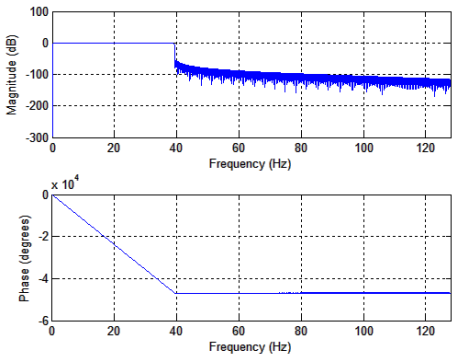


Figura 3.3: Respuesta frecuencial del filtro paso banda. Arriba en magnitud y abajo en fase

[2]De nuevo para poder ver el efecto del filtro no partimos directamente de la señal cruda referenciada sino de la señal referenciada y filtrada paso alto con una frecuencia de corte de 0.5 Hz. En el procesado real se parte directamente de la señal grabada referenciada.

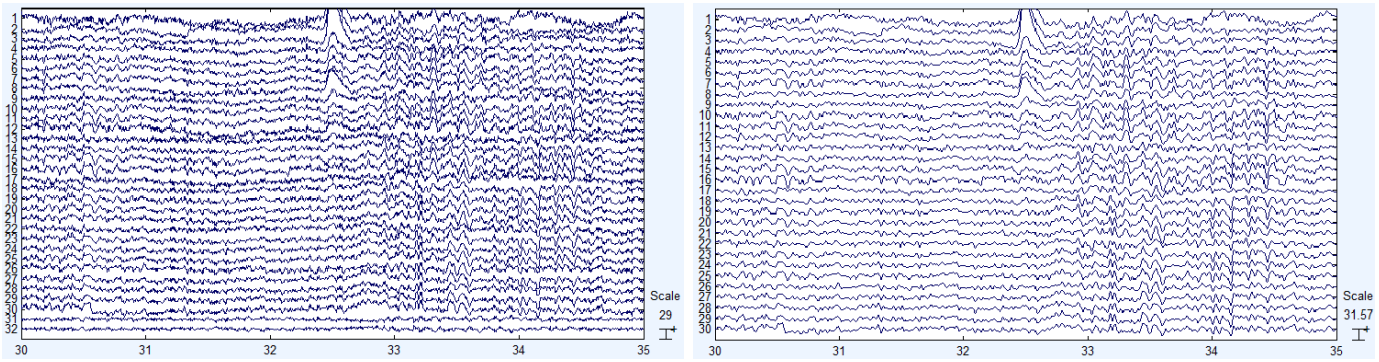


Figura 3.4: Efecto de filtrar paso banda el EEG. Antes de filtrar (izda.) y después de filtrar (dcha.)

3.4. ASR

La señal de EEG suele estar contaminada por elementos conocidos como poppings que se observan como grandes saltos en la señal y están relacionados a pérdidas momentáneas de un buen contacto electrodo-piel. Para detectar dichos saltos haremos uso de un filtrado estadístico en el dominio de las componentes PCA. Este método es útil para detectar cambios no estacionarios de alta varianza. Además normalmente en EEG lo que se hace es marcar las zonas contaminadas por artefactos y luego eliminarlas. Este método, sin embargo, reconstruye los canales afectados de una forma robusta mediante un filtro espacial que tiene en cuenta la forma del cerebro y sus propiedades electromagnéticas. En la figura 3.6 se puede ver el resultado de un filtrado ASR sobre una señal.

El algoritmo se puede resumir en los siguientes pasos (la matemática en profundidad está descrita en [17]):

- Se necesita 1 minuto de EEG limpio como calibración. Este minuto se puede introducir como entrada en el filtro y ser independiente de la señal a filtrar, o se puede obtener de la propia señal a filtrar, tomando ventanas con baja varianza hasta formar un minuto de EEG.
- Se toma una ventana de EEG. Esta es la ventana que se quiere filtrar.
- Se calculan sus componentes principales mediante PCA.
- Se eliminan aquellas componentes que correspondan con artefactos, eligiendo dichas componentes si superan una varianza umbral obtenida de los datos de calibración.
- Se reconstruye cada componente eliminada como una combinación lineal de las componentes sin artefactos.
- Por último se proyecta de vuelta al espacio del EEG obteniendo una ventana limpiada y reconstruída.

En la figura 3.5 se puede observar la operación matemática utilizada para realizar los pasos anteriores para cada ventana. Se puede ver cómo todos los pasos anteriores se pueden colapsar en una operación matricial sobre $x(t)$ (que es un momento puntual dentro de la ventana) para obtener la reconstrucción filtrada $\hat{x}$. El detalle matemático de cada paso se puede encontrar en [17] como se ha indicado anteriormente.

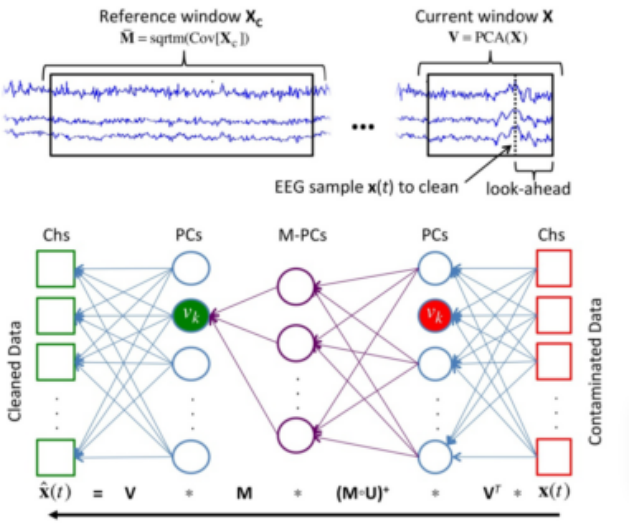


Figura 3.5: Filtrado ASR

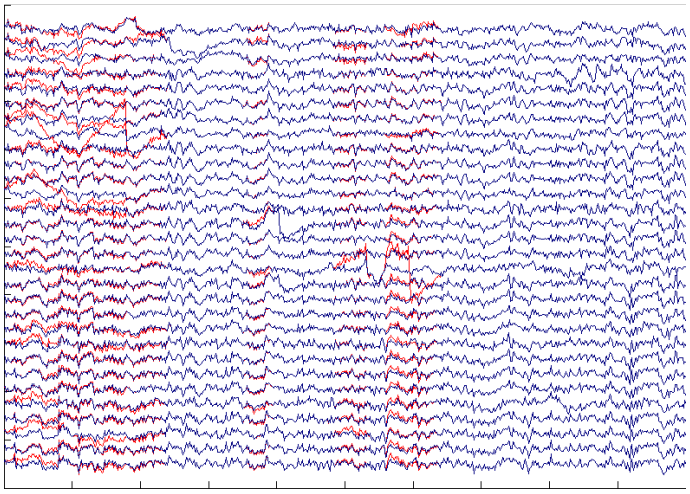


Figura 3.6: Ejemplo de EEG antes (rojo) y después de filtrarlo con ASR (azul)

3.5. ICA

El siguiente paso en la serie de filtros es ICA. Este filtro busca una combinación lineal que separe la señal en componentes que sean independientes entre ellas. De esta forma

estamos intentando localizar las fuentes que han compuesto la señal registrada. ICA se puede clasificar dentro del apartado de separación ciega de fuentes ya que no introducimos ninguna información a priori en el procesado. La señal registrada por cada electrodo es el resultado de la interacción de millones de neuronas y la suma de la interacción de muchas fuentes. Con este método intentaremos separar dichas componentes para después eliminar aquellas que no tengan una razón de ser neuronal o incluso no tengan explicación fisiológica. Las componentes ICA fueron calculadas mediante el algoritmo $fastICA$ [18]. Matemáticamente el filtrado se puede expresar de este modo [19]:

$$ICA_{comp} = W * EEG$$

$$ICA'_{comp} = ICA_{comp}\ sustituyendo\ por\ ceros\ las\ filas\ que\ no\ representan\ señales\ EEG$$

$$EEG_{filtrado} = W^{-1} * ICA'_{comp}$$

La selección de las componentes que son eliminadas vendrá dado por el siguiente paso del procesado (MARA). En la figura 3.7 se puede ver el ciclo completo del cálculo de componentes de ICA y su proyección de vuelta al espacio del EEG. En esta figura no se eliminaría ninguna componente y por lo tanto obtendríamos el mismo EEG de vuelta.

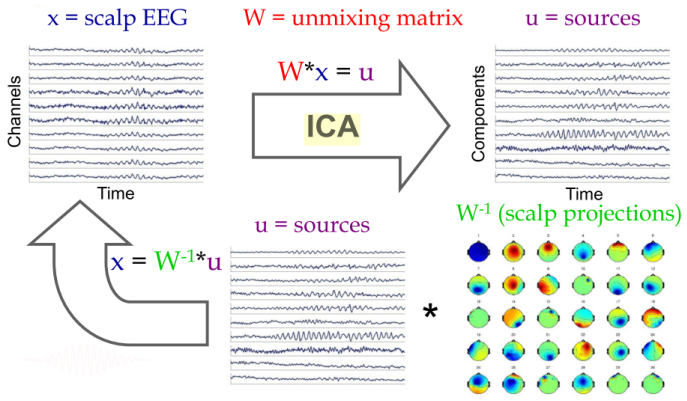


Figura 3.7: Filtrado ICA

3.6. MARA

MARA es una técnica construida sobre ICA que permite determinar de una forma automática aquellas componentes de ICA que no provienen de actividad eléctrica del cerebro. Para ello MARA usa un clasificador (LPM Linear Programming Machine) que fue entrenado con un conjunto de datos tal como se explica en [20]. El clasificador está basado en 6 características básicas que permiten discriminar una componente de ICA cuyo origen no sea fisiológico. Estas características son fijas y su elección viene determinada por los resultados del artículo citado. Las 6 características son las siguientes:

- **Norma de la densidad de corriente:** ICA no aporta información acerca de la localización de las fuentes de las componentes calculadas. Sin embargo, los mapas sobre la cabeza de las componentes de ICA se pueden interpretar como potenciales de EEG para los que la localización de sus fuentes puede ser estimada. MARA usa un modelo de fuentes que sólo representa fuentes de actividad cerebral, por lo que aquellas componentes que no tengan una fuente cerebral deberían tener un modelado muy complejo de fuentes. La característica utilizada es la norma L2 de estas fuentes.

- **Rango dentro de un patrón:** Se calcula el logaritmo de la diferencia entre las activaciones máxima y mínima en un mapa de los electrodos sobre el cuero cabelludo. Se sabe que la actividad típica de artefactos como los músculos del cuello o los parpadeos tienen un patrón espacial determinado. Con esta característica también se pueden localizar patrones que representan un canal suelto.

- *Skewness* **media local:** La media de la *skewness* local absoluta de intervalos de tiempo de 15 segundos. Se usa para detectar valores atípicos o *outliers* en una serie temporal. Puede por lo tanto ser últil para detectar artefactos como poppings.

- λ **y Error de Ajuste**: Estas dos características describen la desviación del espectro de las componentes con respecto a un típica curva 1/f. Los parámetros de la curva se obtienen de: (1) La potencia logarítmica en 2 Hz., (2) La potencia logarítmica en 3 Hz., (3) el punto del mínimo local en la banda 5-13 Hz., (4) el punto un hertzio por debajo del tercer punto de soporte (5) el punto del mínimo local en la banda 33-39 Hz., y (6) el punto un hertzio por debajo del quinto punto de soporte. MARA usa el logaritmo de λ y el error cuadrático medio de la aproximación de f al espectro real entre 8-15 Hz. como características.

- **Promedio de potencia logarítmica en la banda alfa:** Esta característica es útil para detectar el pico típico en la banda entre 8 y 13 Hz. característico de las componentes de origen neuronal.

En las siguientes figuras se puede observar el proceso de selección automática de componentes con artefactos. El clasificador, previamente entrenado gracias a los datos obtenidos en el artículo mencionado, le asigna una probabilidad de que no tengan razón neuronal mediante la clasificación atendiendo a las características previas. En la imagen 3.8 se observan las proyecciones sobre el cuero cabelludo en el espacio del EEG de cada una de las componentes (se calculan 30 pero se muestran sólo las primeras 24 ordenadas de mayor a menor varianza) además de su espectro. Esta información es útil cuando queremos eliminar componentes por inspección visual. Por ejemplo la componente 2, es un claro ejemplo de parpadeos, ya que su proyección en el espacio del EEG se encuentra localizada en la zona de la frente y además no se observa un pico en banda alfa. Por este motivo el análisis de EEG suele requerir de expertos con conocimientos de la fisiología y de tratamiento de señales biológicas. Como hemos comentado, MARA sirve para saltar este paso, calculando 6 características de cada componente (figura 3.9) que introducidas al clasificador previamente entrenado, nos da una probabilidad de que la componente sea

un artefacto. De esta forma definiendo el límite de probabilidad (en este proyecto de 0.5) se seleccionan automáticamente aquellas componentes de ICA a eliminar antes de hacer la proyección de vuelta al espacio del EEG.

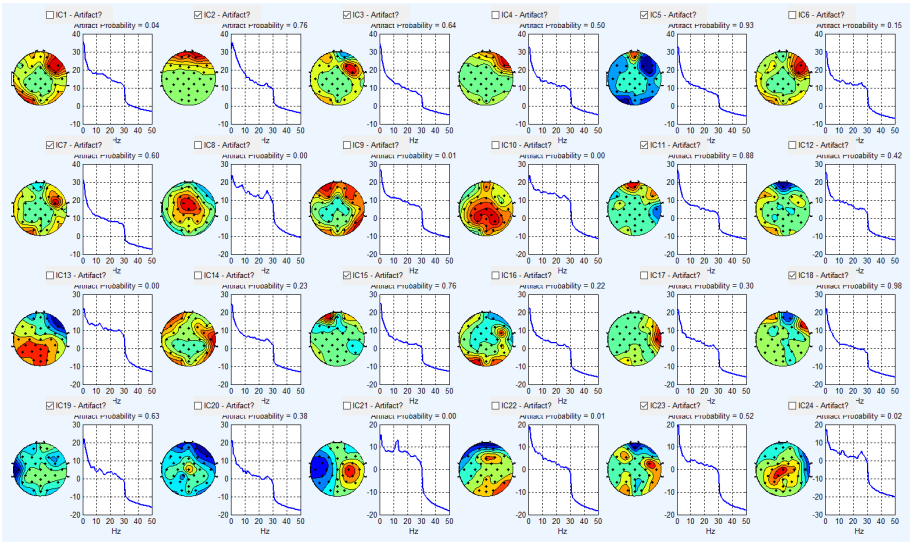


Figura 3.8: Primeras 24 componentes de ICA

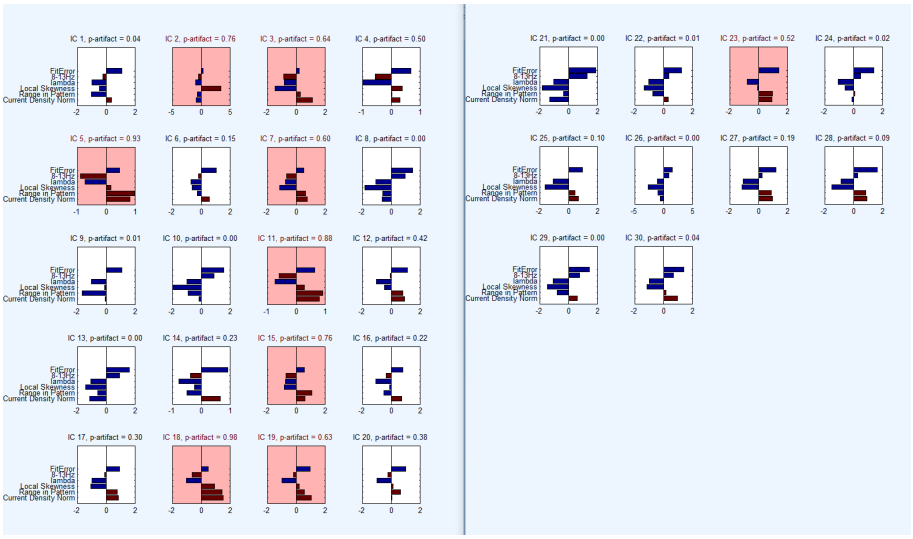


Figura 3.9: Características MARA

4. Relación entre sujetos

4.1. Introducción

Una vez que la señal ha sido registrada y convenientemente filtrada se procede al procesado de los datos. En este capítulo presentaremos cuatro métodos posibles para intentar buscar una relación en las respuestas fisiológicas de los sujetos cuando visionaban el mismo estímulo. El capítulo está dividido en tres secciones principales. En primer lugar aquellas basadas en el aspecto temporal de las señales (sección 4.2). El segundo presenta aquellas técnicas o filtros relacionados con las características frecuenciales del EEG (sección 4.3). Y por último presentaremos un método que nos puede servir de referencia y que no está relacionado directamente con señales fisiológicas sino que es un modelo de predicción de arousal mediante las características multimedia del vídeo (sección 4.4).

4.2. Métodos temporales

Esta sección contiene los métodos utilizados para la búsqueda de patrones comunes entre las respuestas cerebrales de los distintos participantes haciendo hincapié en el dominio temporal, es decir, buscamos maximizar relaciones entre distintas señales multidimensionales en cada instante temporal. Presentaremos dos métodos utilizados para tal objetivo. En la sección 4.2.1 se expone el Análisis de Componentes Canónicas (CCA) en su versión para señales multivariantes que denominaremos mCCA y en la sección 4.2.2 se presenta un método que lo que intenta es buscar potenciales evocados por eventos (ERP en sus siglas en inglés) que se presenten de forma común entre todos los sujetos. Denominaremos el método como cERP, siendo la c en este caso referente que es un ERP común entre participantes.

4.2.1. mCCA

El Análisis de Componentes Canónicas es una técnica de principios del siglo XX que surge de la necesidad de buscar relaciones entre series temporales. Lo que se busca con este

método es encontrar un filtro lineal que maximice la correlación entre distintas señales. En nuestro caso tenemos 32 señales (sujetos) multivariantes con 30 canales cada una (la actividad de las orejas se ha eliminado por no relevante) y duración igual a la grabación. Habría que utilizar alguna modificación sobre la técnica original al tratarse de señales multivariantes. El método para dos señales se podría resumir matemáticamente de la siguiente forma [21, 22]:

Sean $X_{s1} \in \mathbb{R}^{CxT}$ y $X_{s2} \in \mathbb{R}^{CxT}$ las señales multivariantes que corresponden con los EEG del sujeto 1 y del sujeto 2 respectivamente. Las dimensiones de ambos son de CxT, es decir, canales por tiempo. El objetivo de Análisis de Componentes Canónicas es buscar un par de filtros lineales espaciales (se denominan espaciales porque corresponden con la dimensión de los canales y por lo tanto tienen una relación con la distribución de los electrodos) $w_{s1} \in \mathbb{R}^{(Cx1)}$ y $w_{s2} \in \mathbb{R}^{(Cx1)}$ de forma que se maximice la correlación entre ambas señales de la siguiente forma:

$$\underset{w_{s1},w_{s2}}{\text{argmax}}\left(\frac{w_{s1}^{\mathsf{T}} X_{s1} X_{s2}^{\mathsf{T}} w_{s2}}{\sqrt{w_{s1}^{\mathsf{T}} X_{s1} X_{s1}^{\mathsf{T}} w_{s1} w_{s2}^{\mathsf{T}} X_{s2} X_{s2}^{\mathsf{T}} w_{s2}}}\right)$$

que es lo mismo que maximizar

$$\underset{w_{s1},w_{s2}}{\text{argmax}}\, w_{s1}^{\mathsf{T}} C_{s1s2} w_{s2}$$

sujeto a

$$w_{s1}^{\mathsf{T}} C_{s1s1} w_{s1} = 1$$
$$w_{s2}^{\mathsf{T}} C_{s2s2} w_{s2} = 1$$

La resolución del problema se puede encontrar en [21, 22]. En el mismo artículo también se expone cómo se amplía este problema al caso que nos concierne donde no sólo tenemos dos señales multivariantes sino que son 32 (sujetos). La diferencia básica está en la maximización que ahora busca maximizar la suma de todas las correlaciones por pares de forma que obtenemos a la salida un filtro lineal para cada sujeto.

El sistema completo se presenta en la figura 4.1. Comienza desde las señales de EEG de cada participante. A continuación se realiza el preprocesado de cada una de las señales como se ha comentado en el capítulo anterior. Para facilitar el trabajo a la maximización de la correlación se pasa cada señal filtrada por un Análisis de Componentes Principales. Este método es un filtro lineal que maximiza la varianza de forma que en cada componente se represente la mayor parte de la varianza de la señal y que cada componente sea ortogonal a las demás. De esta forma podemos por ejemplo reducir la dimensión de 30 canales a 4 componentes que explican casi el 100 % de la varianza de los datos, es decir, de la información de utilidad de la señal[1]. A las señales fruto de las componentes PCA de cada sujeto se le aplica el método de mCCA explicado anteriormente de forma que obtenemos un filtro lineal espacial por cada sujeto que maximiza la correlación entre dicho

[1]En el capítulo siguiente se presentan también los resultados sin esta reducción de dimensionalidad con PCA

sujeto y todos los demás. Se multiplican las señales de cada sujeto por el filtro resultante en mCCA. En este punto tenemos una señal de dimensión $1xT$ por cada sujeto que maximiza la correlación con el resto. Al final lo que nos interesa es tener una representación de dimensión $1xT$ a lo largo del tiempo de los momentos de máxima similitud pero la correlación es una operación que da como resultado un punto por lo que para tener una representación temporal, se calcula una correlación enventanada por pares de sujetos. Ya que el filtro de cada sujeto está pensado en maximizar la correlación, éste es el motivo de que la métrica que obtenemos para cada sujeto se base en la correlación con el resto de sujetos. Por último promediamos todas estas correlaciones y obtenemos una señal que sería nuestra métrica.

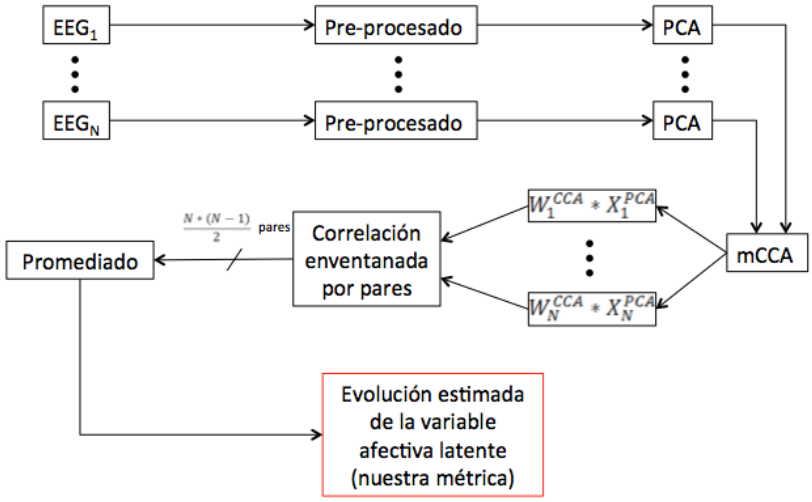


Figura 4.1: Esquema del sistema mCCA

4.2.2. cERPv1 y cERPv2

Este método se basa en la búsqueda de momentos que hayan provocado un potencial común entre todos los participantes. Como ya se comentó en el capitulo 2 la investigación con EEG se ha basado en dos tipos de ondas, las conocidas como espontáneas o continuas (con los ritmos asociados en la tabla 2.1) y lo que se conoce como potenciales evocados. Estos potenciales se asocian a respuestas neuronales provocadas por un estímulo externo que puede ser por ejemplo visual, auditivo, somatosensorial u olfativo. Estas respuestas son débiles y a simple vista no se pueden detectar o identificar si miramos un EEG en crudo. La forma típica de afrontar el problema de la detección de estos procesos cognitivos consiste en grabar muchas repeticiones del ERP para poder incrementar el SNR (relación señal a ruido) de forma considerable promediando entre las distintas repeticiones. Al ser el EEG un proceso no estacionario no podemos asegurar que el EEG esté en la misma fase o sincronizado temporalmente de forma perfecta por lo que el promedio de muchas repeticiones o *trials* es la única manera de poder obtener una señal ERP de calidad. En

nuestro problema no podemos contar con varias repeticiones de un mismo estímulo ya que estamos hablando de vídeos y obtener muchas repeticiones supone muchos visionados y esto requeriría un tiempo excesivamente largo (aparte de otras consideraciones como que la carga afectiva ya no sería la misma). Sin embargo partimos de la hipótesis de que si un evento del vídeo produce un potencial evocado de forma consistente para un número elevado de sujetos es que dicho evento es relevante para la muestra. De esta forma podríamos decir que nuestras repeticiones estarían compuestas por las respuestas neuronales de un conjunto de sujetos. Vamos a presentar dos posibles formas de obtener los resultados.

Empezaremos explicando el sistema paso a paso para la primera versión que denominaremos cERPv1 (figura 4.2). Partimos del EEG de N sujetos. Cada uno de estos EEGs es preprocesado como se explica en el capítulo anterior. A continuación cada EEG preprocesado se filtra paso banda entre 2 y 8 Hz. Se sabe que el ERP tiene una respuesta frecuencial en torno a la banda theta por lo que filtraremos en un rango próximo a dicha banda para facilitar los siguientes pasos y evitar contaminaciones innecesarias en el proceso. A continuación, y como si de varios trials realizados por el mismo sujeto se tratara, se promedian las señales filtradas de todos los sujetos de forma que tenemos la respuesta y los posibles ERP de un metasujeto. Después, para cada canal, se calculan los máximos y los mínimos locales. Partimos de la base de que al promediar y siendo el EEG una señal no estacionaria, su media tenderá a cero excepto en aquellas zonas de posibles potenciales que se verían ligeramente resaltadas sobre la media. Se calcula a continuación la amplitud de cada oscilación restando a cada máximo local su mínimo local anterior. El siguiente paso será una normalización de todos los posibles potenciales evocados detectados. Se usará una normalización *z-score* asumiendo la *gaussianidad* de las amplitudes detectadas. Es decir, del vector de posibles amplitudes de ERPs se resta la media y se divide por su desviación típica. De esta normalización se elige un umbral de 2 que significa que nos quedaremos con aquellos ERP cuya amplitud esté por encima de dos desviaciones típicas de la media de ERPs. De esta forma estamos filtrando estadísticamente aquellos ERP menos significativos. La señal que tenemos en este punto es una señal impulsiva para cada canal. Se sabe que los potenciales evocados son máximos en la línea central del cerebro por lo que se promediarán los canales Fz, Cz y Pz. Antes de realizar el promediado se optó por un convolucionado de la señal impulsiva con una gaussiana de forma que se da una cierta incertidumbre a la localización del estímulo. Por último se realiza el promediado que dará la salida del sistema y será nuestra métrica de interés.

La versión número 2 (figura 4.3) de este algoritmo se basa en los mismos pasos de procesamiento pero el promediado no se hace hasta el final. Por lo que las señales de cada EEG (cada sujeto) son prefiltradas, filtradas paso banda entre 2 y 8 Hz., calculados sus mínimos y máximos locales, sus amplitudes impulsivas, se normalizan, se filtran por umbral y se convolucionan. Después de realizar esto para cada sujeto se promedia entre todos los sujetos y se promedian los canales Fz, Cz y Pz[2].

[2]En los resultados se presenta una versión haciendo PCA para reducir dimensiones o sin PCA

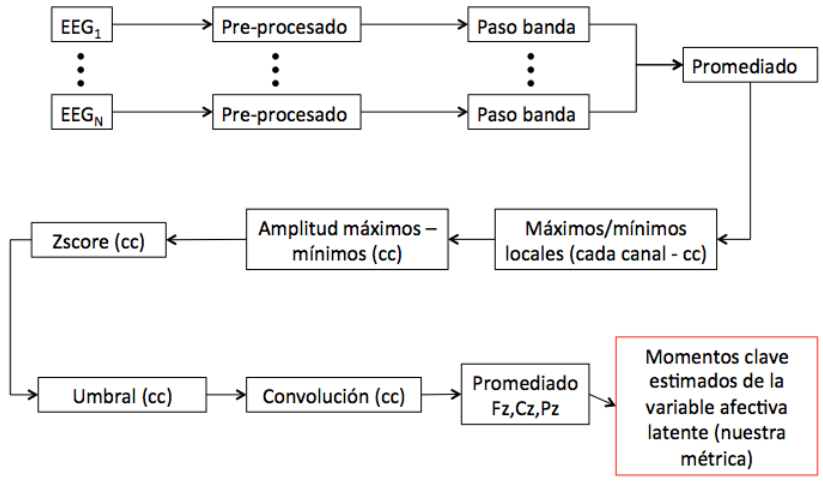


Figura 4.2: Esquema del sistema cERPv1

4.3. Métodos frecuenciales

Los dos métodos anteriores hacían uso de la evolución temporal de las señales de EEG. En este caso se va a intentar buscar una relación entre las componentes frecuenciales de determinadas bandas. Como estamos interesados en encontrar los momentos clave a lo largo del tiempo, buscaremos la relación de las envolventes de señales previamente filtradas en ciertas bandas frecuenciales. El sistema completo lo podemos ver en la figura 4.5.

El proceso comienza de nuevo con un preprocesado del EEG de cada participante. De nuevo usaremos la técnica de PCA para reducir el número de dimensiones de nuestro problema[3]. Por otro lado calcularemos el IAF (pico máximo de la banda alfa individualizado) para cada sujeto. La división del ritmo neuronal en diferentes bandas como en la tabla 2.1 es un modelo teórico y el situar 10 Hz. como el pico común en alfa funciona si hablamos del promedio de cerebros. Pero si queremos individualizar nuestro sistema debemos calcular lo que se conoce como IAF o Individualized Alpha Frequency [23]. Para ello haremos uso del EEG grabado con ojos cerrados que nos permite tener una muestra más clara de dónde se encuentra el pico en alfa para cada individuo.

Antes de usar la técnica de cSPoC hay que elegir en qué banda frecuencial nos interesa centrarnos. Usaremos la técnica con tres bandas frecuenciales distintas como pueden ser la banda delta, theta y alfa. Para mejorar la calidad de la señal en dichas bandas se usó la técnica SSD [24] que podemos resumir de la siguiente manera. Buscaremos un filtro lineal que maximice la potencia en una determinada banda mientras que se reduce la potencia de las bandas colindantes (figura 4.4). Se asume que la señal está contaminada en todo su espectro con un ruido (el ruido en EEG suele ser del tipo ruido blanco o ruido rosa

[3]En los resultados se presentar también la versión sin PCA

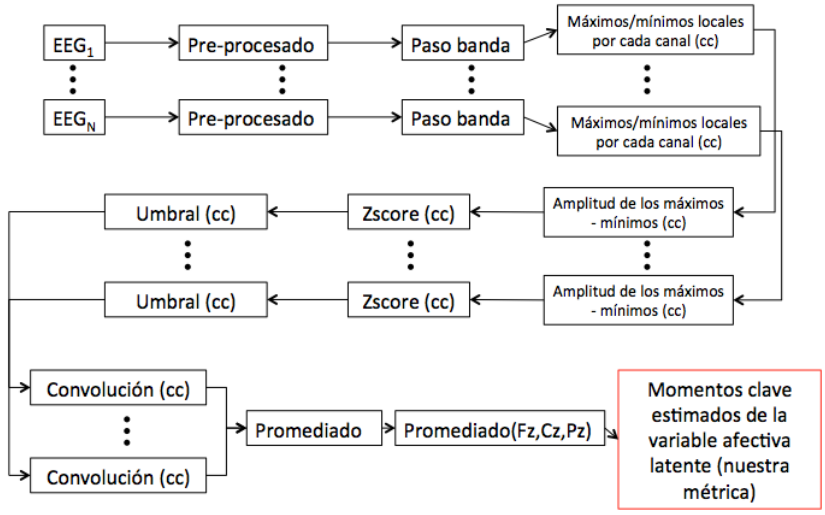


Figura 4.3: Esquema del sistema cERPv2

1/f) y que dicho ruido está incorrelacionado con la señal de interés. Matemáticamente se puede explicar así:

Sea una señal de EEG m que se puede dividir en señal de interés y en ruido, es decir $m = s + n$. Lo que nos interesa maximizar el la SNR entre la señal de interés y el ruido. $SNR = \frac{P_s(f)}{P_n(f)}$. Si sólo estamos interesados en obtener la potencia alrededor de un determinado pico y considereando Δf suficientemente pequeño podemos escribir que $P_n(f) \approx \frac{P_m(f-\Delta f)+P_m(f+\Delta f)}{2}$. Por otra parte si tomamos la transformada de Fourier sabemos que $\langle|F_m(f)^2|\rangle = \langle|F_s(f)^2|\rangle + \langle|F_n(f)^2|\rangle + 2\Re\langle F_s(f)F_n^*(f)\rangle$ y teniendo en cuenta la suposición de incorrelación entre el ruido y la señal podemos simplificar la ecuación anterior como $\langle|F_m(f)^2|\rangle = \langle|F_s(f)^2|\rangle + \langle|F_n(f)^2|\rangle$ o lo que es lo mismo en terminología de potencias $P_n(f) = P_s(f) + P_n(f)$. El objetivo es maximizar el ratio entre la señal a una frecuencia con respecto a las frecuencias de alrededor por lo que reescribiendo con las fórmulas anteriores $\frac{P_m(f)}{P_m(f-\Delta f)+P_m(f+\Delta f)} \approx \frac{P_s(f)+P_n(f)}{2P_n(f)} = 0{,}5(\frac{P_s(f)}{P_n(f)} + 1)$ por lo que podemos afirmar que incrementando el término de la izquierda podemos incrementar la señal a ruido de interés en dicha banda. Partiendo de esta idea se busca un filtro lineal que maximice la idea anterior de forma que maximicemos la señal a ruido en dicha banda. La matemática se encuentra descrita en [24].

Una vez que tenemos las señales filtradas y maximizada la potencia de dichas señales en la banda frecuencial de interés se aplica el algoritmo cSPoC que consiste en buscar mediante optimización aquel filtro lineal que maximice la correlación entre las envolventes de las señales. Mediante este método se encuentra un filtro espacial por cada sujeto que maximiza su correlación con el resto de sujetos. Es una idea similar al CCA, lo expondremos para el caso de dos sujetos [25].

Sean $X_{s1} \in \mathbb{R}^{CxT}$ y $X_{s2} \in \mathbb{R}^{CxT}$ las señales multivariantes que corresponden con los EEG del sujeto 1 y del sujeto 2 respectivamente. Las dimensiones de ambos son de CxT,

es decir, canales por tiempo. El objetivo de cSPoC es buscar un par de filtros lineales espaciales $w_{s1} \in \mathbb{R}^{(Cx1)}$ y $w_{s2} \in \mathbb{R}^{(Cx1)}$ de forma que se maximice la correlación entre las envolventes de las señales. La maximización en términos matemáticos queda así: Definimos las envolventes de las proyecciones (cuya correlación se querrá maximizar) como

$$\Phi[(w_{s1}^{\mathsf{T}} X_{s1})] = \sqrt{(w_{s1}^{\mathsf{T}} X_{s1})^2 + (w_{s1}^{\mathsf{T}} H[X_{s1}])^2}$$

$$\Phi[(w_{s2}^{\mathsf{T}} X_{s2})] = \sqrt{(w_{s2}^{\mathsf{T}} X_{s2})^2 + (w_{s2}^{\mathsf{T}} H[X_{s2}])^2}$$

donde $H[\cdot]$ significa transformada de Hilbert. La función a maximizar es la correlación entre ambas envolventes. Denoratemos $\Phi[(w_{s1}^{\mathsf{T}} X_{s1})]$ como Φ_{s1} y $\Phi[(w_{s2}^{\mathsf{T}} X_{s2})]$ como Φ_{s2}

$$\underset{w_{s1}, w_{s2}}{\operatorname{argmax}} Corr[\Phi_{s1}, \Phi_{s2}] = \underset{w_{s1}, w_{s2}}{\operatorname{argmax}} \frac{Cov[\Phi_{s1}, \Phi_{s2}]}{\sqrt{Var[\Phi_{s1}] Var[\Phi_{s2}]}}$$

En el artículo [25] se expone la extensión del proceso cuando tenemos más de dos sujetos como es nuestro caso. A la salida de este proceso por lo tanto tenemos un filtro lineal por cada participante que maximiza la función anterior. El siguiente paso es usar el filtro calculado para cada sujeto para proyectar los datos al espacio donde la correlación de las envolventes es máxima entre sujetos. El siguiente paso es, como se hizo en el método mCCA, enventanar la señal y calcular la correlación de las envolventes por pares de sujetos. Y por último promediamos todas estas correlaciones para obtener la métrica final.

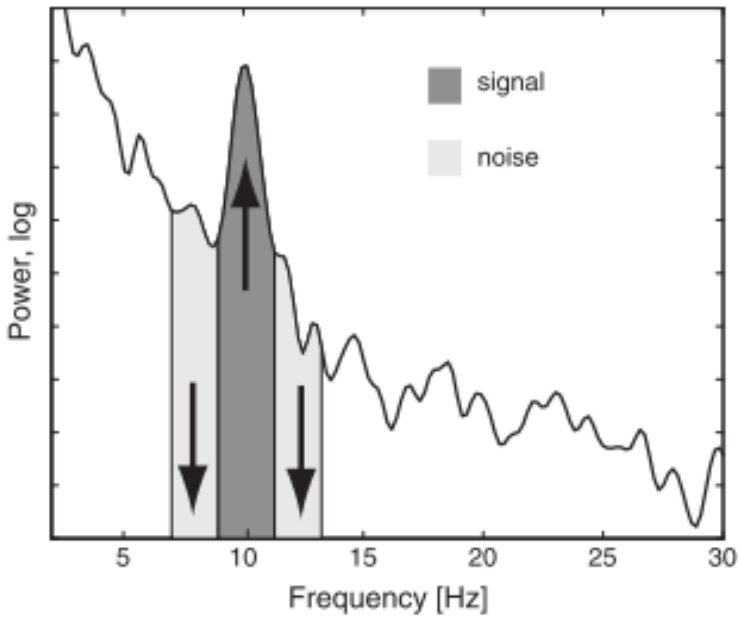


Figura 4.4: Objetivo del algoritmo SSD

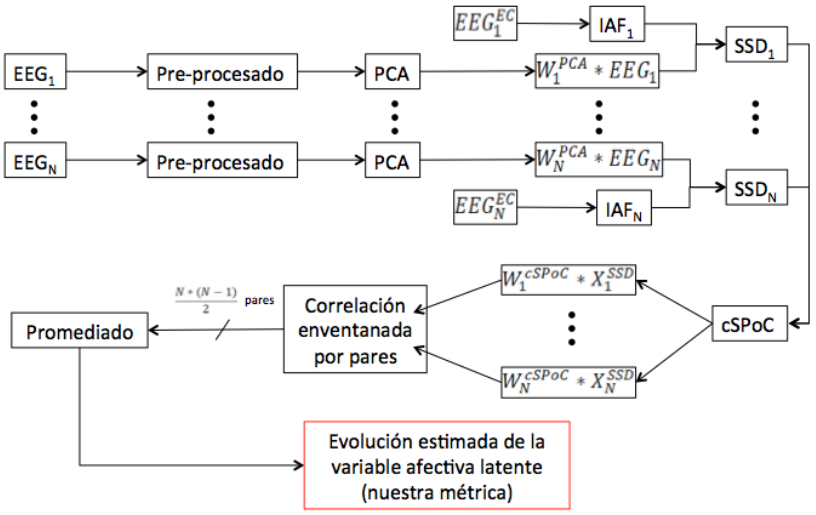


Figura 4.5: Esquema del sistema cSPoC

4.4. Multimedia

Por último y como complemento a los métodos fisiológicos que se presentan en este trabajo se implementó también un sistema basado en características multimedia de los propios vídeos. La idea es crear un modelo que represente la dimensión del arousal para poder detectar aquellos momentos de mayor impacto emocional.

El algoritmo utilizado se basa en tres características principales explicadas en detalle en [26]:

- Componente de movimiento: Esta característica cuantifica la cantidad de movimiento existente entre frames consecutivos. Para cuantificarlo se usa el método de estimación de movimiento basada en bloques. Se entiende que un mayor movimiento entre los frames representa un mayor nivel de arousal.
- Componente del ritmo: Se basa en cuantificar la densidad de cambios de plano. Una mayor densidad de cambios de plano supone un mayor ritmo del vídeo y por lo tanto un mayor arousal.
- Energía del sonido: La última característica está relacionada con el audio y no es más que la energía de sonido del vídeo a lo largo del tiempo. Una mayor energía supone un mayor volumen y por lo tanto se relaciona con un mayor arousal.

Como las tres características anteriores tienen rangos diferentes, se normalizan entre 0 y 100 % para el conjunto de vídeos. Además cada una de las características anteriores se convoluciona con una onda parecida a una gaussiana para suavizar la señal obtenida.

5. Resultados

5.1. Introducción

En el presente capítulo se presentan los resultados de aplicar las diferentes métricas de relación entre sujetos para la obtención de aquellos momentos de mayor interés de los vídeos. Se seguirá la misma estructura que en el capítulo anterior exponiendo los resultados más relevantes para cada una de las técnicas.

5.2. Momentos clave del vídeo

En esta sección presentaremos los momentos clave de los cinco vídeos. Todos los momentos seleccionados son aquellos susceptibles de provocar algún cambio en el estado de las personas. Se indican con números que corresponden con el segundo en el que se produce el evento. Pueden ser desde cambios de planos a sustos, saltos etc. Los momentos de los vídeos se muestran en las tablas 5.1 - 5.5.

5.3. Métodos temporales

5.3.1. mCCA

En esta sección se presentan los resultados del uso de la técnica de mCCA en los cinco vídeos de interés. Se obtuvieron los resultados tanto para una versión del procesado en la que se hacía PCA como para una versión en la que no había reducción de la dimensionalidad de los datos. Lo primero que podemos observar es que existe una diferencia entre las versiones con PCA y sin PCA. Estas diferencias, unidas al hecho de que no parece existir un patrón claro que relacione los momentos clave con los picos de nuestro sistema, nos lleva a pensar que este método no es del todo útil para nuestro propósito. El factor principal que puede influir negativamente es el hecho de que la correlación es un método puntual (punto a punto) y que el EEG (temporal) es una señal de rápida variación por

Segundos	Secuencia
0 - 3	filmax
3 - 13	presentación con varios cambios de plano (5,7,8,10,11)
13 - 23	en la puerta (con susto al final) (cambia plano en 19, 21)
23 - 46	escena señora (27, 30, 41)
45 - 49	letras
49 - 55	equipo de sanidad (51, 53)
55 - 57	letras
57 - 60	deja de grabar!
60 - 64	en negro
64 -70	calma tensa escaleras (susto final)
70 - 83	persecución con letras en (72, 75, 80, 83) y susto en 82
83 - 90	plano cara respirando
90 - 96	logo
96 - 100	próximamente en cines
100-103	créditos

Tabla 5.1: Tabla de secuencias del vídeo REC

lo que no se capturan en el tiempo de forma sencilla relaciones entre distintos EEGs. Los resultados se muestran en las figuras 5.1 - 5.5. Existen momentos de máxima relación entre los sujetos de acuerdo a la métrica obtenida pero estos son difícilmente distinguibles de otros momentos que no podríamos explicar de manera sencilla. Por lo que este sistema puede capturar algo más que el simple arousal pero esto conlleva una dificultad añadida en buscar dicha relación oculta y subjetiva.

5.3.2. cERPv1 y cERPv2

En esta sección se presentan los resultados para los cinco vídeos de las técnicas basadas en potenciales evocados. La teoría bajo este sistema es que los momentos clave (sustos, saltos, eventos, cambios bruscos) podrían ser capturados mediante potenciales evocados en el EEG. No es un potencial evocado al uso ya que un mismo sujeto no visiona varias veces el mismo estímulo y luego se promedia, sino que la idea es que en este caso los diferentes sujetos son repeticiones de un metasujeto y por lo tanto se verá una mayor amplitud en aquellos momentos que sean un potencial evocado en un ficticio metasujeto. La primera observación es que se puede ver que reducir las dimensiones con PCA cambia ligeramente los resultados pero los momentos de máxima coordinación se siguen respetando casi siempre ya sea con PCA o sin PCA. Si comparamos entre los dos métodos (v1 y v2) vemos que existen diferencias. En algunos casos los picos de máxima amplitud se comparten pero en otras ocasiones cambian de sitio. Esto se debe a la naturaleza estadística del método, en uno se promedia primero (por lo que se supone que el nivel de señal a ruido aumentará)

Segundos	Secuencia
0 - 5	hombre mete botellas (3 cambio de plano)
5 - 8	letras
8 - 14	hombre mete botellas (12 cambio de plano)
14 - 16	letras
16 - 42	botellas sin parar (cambios de plano en 19, 21, 22, 23, 25, 29, 30, 32, 33, grito en 35 y cambio en 38)
42 - 51	flores (cambios de plano en 43, 45, 47, 48, 49, 51)
51 - 68	más botellas con sonido en 51 (cambios de plano en 52, 54, 56, 58, 60, 62, 63, 64, 66)
68 - 84	regalos manos (cambios de plano en 70, 73, 75, 77, 80, 82)
84 - 103	pizza (cambios de plano en 87, 89, 91, 92, 93, 95, 96, 98, 100, 103)
103 - 117	bocadillo con sonido al principio (cambios en 106, 108, 111, 114, 116, 117)
117 - 121	letras con logo
121 - 123	chico comiendo bocata

Tabla 5.2: Tabla de secuencias del vídeo CocaCola

y luego se hace la estadística sobre dicha señal. En el segundo método se calculan las estadísticas por sujeto, lo que en principio es peor ya que se necesita del promediado de potenciales para tener la típica respuesta de potencial evocado. Comentando brevemente los momentos que serían más relevantes para cada vídeo (figuras 5.6 - 5.15):

- **REC:** En este vídeo podemos observar diferentes momentos importantes detectados con naturaleza distinta. Podemos observar los picos en 23 segundos, en 41 segundos o en 70 segundos que corresponden con sustos. También podemos observar picos en cambios de plano como en 5 o 7 segundos y algo importante como es la aparición del logo en 90 segundos.

- **CocaCola:** Los momentos se corresponden sobre todo con cambios de planos en la secuencia del bocadillo (segundo 106) o en la secuencia cuando empiezan a salir botellas sin parar (segundos entre 20 y 40)

- **Esquí:** En este vídeo no se observan momentos tan claros, pero podemos observar un pico en 40 segundos que corresponde con un cambio de plano. Y si acudimos al v2, podemos ver un pico en el momento del grito en torno a 81 segundos y otro en los cambios de plano del final en torno a 96 segundos.

- **Prinzenzeit:** En este vídeo de relax se puede observar cómo los picos detectados como máximos corresponden con cambios de planos. Por ejemplo el cambio a los 40 segundos, a los 52 o a los 62.

Segundos	Secuencia
0 - 3	abrochar botas
3 - 5	pasos
5 - 11	golpes nieve
11 - 22	deslizamiento en la nieve
22 - 26	caída (muy suave)
26 - 33	descenso
33 - 35	vista hacia atrás
35 - 40	descenso
40 - 43	vista hacia atrás
43 - 55	descenso
55 - 66	en 55 salta al vacío y abre paracaídas
66 - 74	lo mismo desde otra vista
74 - 94	avalancha (gritos en 81, 91 y 93)
94 - 103	aterrizaje (cambios de plano en 95, 96, 100)

Tabla 5.3: Tabla de secuencias del vídeo Esquí

- **Paranormal:** Se puede observar que los mayores picos coinciden con momentos clave como el sonido a los 9 segundos, el primer susto a los 14 segundos, la aparición de letras en la pantalla, y los sustos en torno a los 30 segundos y a los 52 segundos.

5.4. Métodos frecuenciales

En esta sección analizaremos los resultados obtenidos para el sistema basado en componentes frecuenciales. Se han calculado tres tipos de gráficas por cada banda frecuencial. En primer lugar se muestran la proyecciones de activaciones cerebrales (obtenidas a partir del filtro lineal que maximiza la relación entre sujetos) de los distintos sujetos. Con respecto a estas figuras cabe destacar que la mayor activación (el signo no importa ya que es la proyección en el cerebro de un filtro lineal y por lo tanto podríamos tener pesos de diferente signo indistintamente) parece situarse de forma bastante consistente en las zonas frontales y parietal-occipital. La interacción de las bandas alfa y theta en la zona frontal y parietal se asocia a procesos cognitivos relacionados con las emociones (memoria, atención [23]) y la zona occipital gestiona el procesamiento visual lo cual es lógico ya que estamos viendo vídeos. Luego se presenta una gráfica con la evolución de la envolvente de la potencia media de todos los sujetos (sin pasar por cSPoC) y por último lo que sería el resultado cSPoC. Cabe destacar las pocas diferencias que existen al hacer una reducción de dimensionalidad en cuanto a la componente frecuencial se refiere. Analizaremos cada vídeo por separado centrándonos en el resultado del sistema cSPoC (figuras [5.18, 5.21, 5.24, 5.27, 5.30, 5.33, 5.36, 5.39, 5.42, 5.45, 5.48, 5.51, 5.54, 5.57, 5.60]):

Segundos	Secuencia
0 - 10	mar
10 - 14	respiración
14 - 18	mar con zoom de piedras
18 - 28	mar
28 - 34	rama
34 - 40	bosque
40 - 44	fruto
44 - 48	tela de araña
48 - 52	flores
52 - 55	ramas
55 - 59	caricia al perro
59 - 62	pata del perro
62 - 65	embarcadero
65 - 71	letras con embarcadero

Tabla 5.4: Tabla de secuencias del vídeo Prinzenzeit

- **REC:** En banda alfa podemos observar una sincronización en la escena de la señora (segundos 24-32 aprox.), y podemos detectar la aparición de letras (segundos 45 y 55). Hay una fase larga entre 45 y 64 segundos que parece de alta sincronización que se corresponde con letras, una escena de tensión y una pantalla en negro. Todo ello tiene lógica ya que se tratan de procesos cognitivos (leer, desatención por pantalla en negro) que tienen relación con variaciones en alfa. Las relaciones en delta son más difíciles de explicar ya que es una banda que suele estar contaminada por artefactos sobre todo oculares. Por lo tanto es curioso cómo en momentos con susto (segundo 23, segundo 70) vemos unos picos destacados. Esto se puede deber a un hecho conocido que es parpadear fuertemente cuando vemos algo que nos asusta. En banda theta vuelve a aparecer la escena de la señora, las letras y la escena entre letras.

- **CocaCola:** En este vídeo no se observan picos claros relacionados con momentos exactos. En banda theta es una sucesión casi continua de picos (también es verdad que es un vídeo con muchos cambios de planos). Si hay que destacar algo es en banda alfa con las escenas de botellas y alegría (segundo 60) o con los regalos (en torno a 90 segundos).

- **Esquí:** Empezando por banda alfa vemos una sincronización grande en torno a 80 segundos que corresponde con el grito en la avalancha o en la primera escena de deslizamiento en la nieve (segundos 11-22). En banda delta volvemos a ver el pico en el grito de 80 segundos y también podemos apreciar el salto al vacío con una subida en 55 segundos. Por último en banda theta volvemos a ver el grito de 80, un nuevo grito en 93 y la escena de deslizamiento primera como momentos de máxima

Segundos	Secuencia
0 - 6	green audience rating
6 - 14	pantalla en negro con distorsiones en (7 y 9) y pasos
14 - 23	susto, luego imagen de la misma habitación
23 - 27	letras
27 - 30	cambios de plano a espectadores y casa
30 - 31	susto y reacción
31 - 34	negro con distorsiones
34 - 45	cambios de plano en la casa en (37, 39, 41)
45 - 46	letras
46 - 50	perro ladra
50 - 51	letras
52 - 54	susto al final
55 - 62	logo
63 - 69	página web
69 - 72	logo

Tabla 5.5: Tabla de secuencias del vídeo Paranormal

relación entre sujetos.

- **Prinzenzeit:** Este es un vídeo de bajo arousal donde es difícil sacar conclusiones a los picos observados. Sólo en banda theta (quizá relacionado con ERPs) podemos ver picos que se corresponden con cambios de plano como en los segundos 10, 18, 34, 44, 48 o 59.
- **Paranormal:** En banda alfa vemos dos picos muy claros. El primero de ellos se corresponde con el segundo susto (segundo 30) y la pantalla en negro de después hasta el segundo 34. El segundo momento se corresponde con el último susto (segundo 54) y el logo. Observamos como el estado que se crea debido al susto se mantiene durante el visionado del logo de la película, lo cual interesa al creador ya que es una película de miedo y la asociación de sustos con la marca puede crear el efecto deseado en los usuarios. El primer susto también tiene un efecto (14 segundos) pero no es tan relevante. En banda delta podemos ver el primer susto y el tercero. En banda theta podemos apreciar también los tres sustos (14, 30 y 54) pero se capturan otras relaciones por lo que no es un resultado tan claro como en banda alfa.

5.5. Multimedia

En este apartado añadimos un resultado adicional que no está basado en componentes fisiológicas pero que puede ser una ayuda para formar por ejemplo sistemas combinados

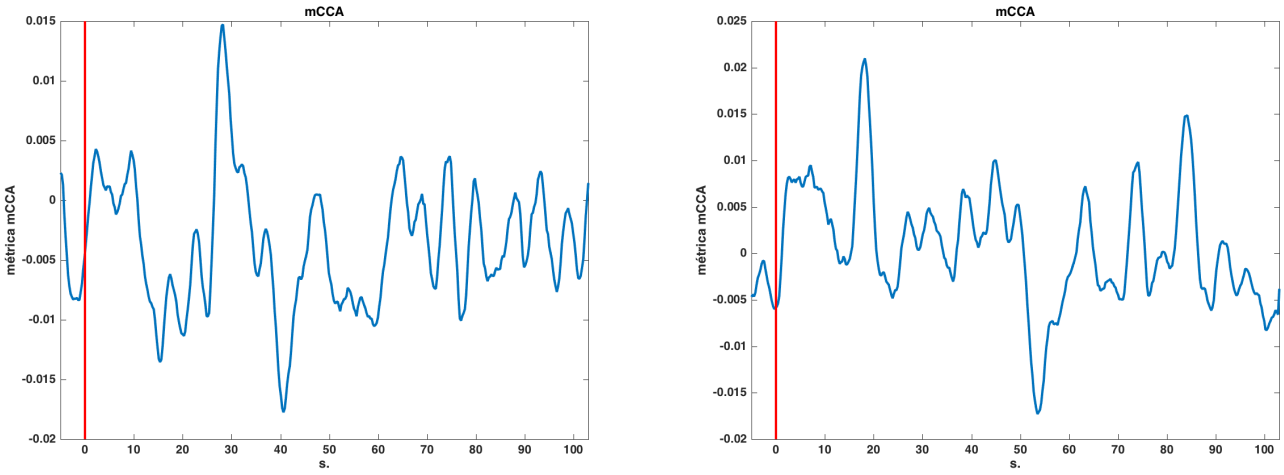


Figura 5.1: Resultado mCCA vídeo REC con PCA (izda.) y sin PCA (dcha.)

entre información multimedia del propio vídeo y la respuesta fisiológica común de muchos sujetos. De forma general podemos ver que, por ejemplo, en el vídeo de Paranormal se capturan muy bien los tres momentos que corresponden con los tres sustos del vídeo. En el vídeo de Esquí el momento que resulta es el grito. En el vídeo de CocaCola se puede observar una especie de estructura en el mismo que puede coincidir con el interés del diseñador. El vídeo empieza de una forma tranquila, después va creciendo en arousal, llega a un valle para finalizar en lo alto que es cuando aparece la imagen de la marca (figuras 5.61 - 5.65).

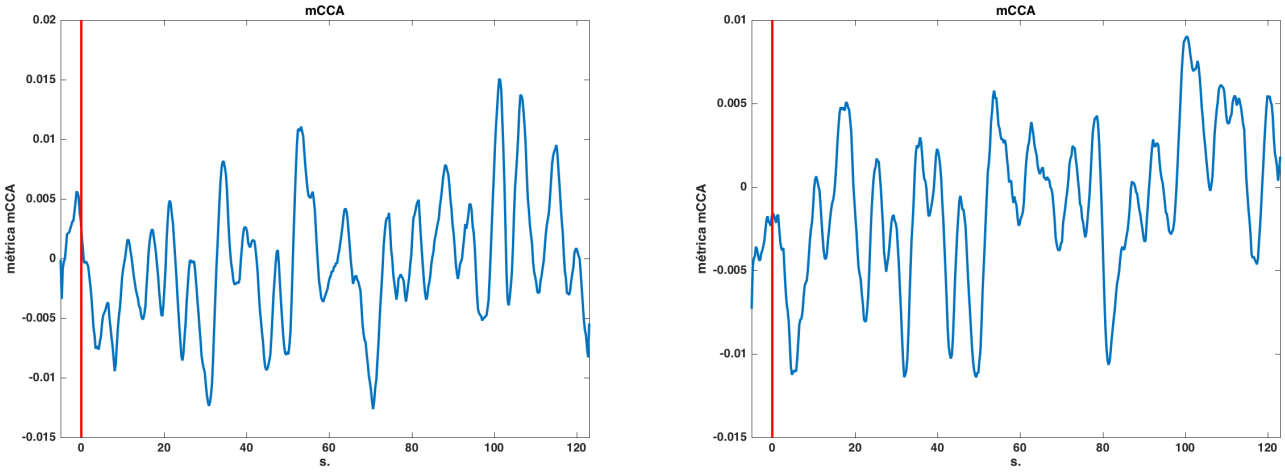


Figura 5.2: Resultado mCCA vídeo CocaCola con PCA (izda.) y sin PCA (dcha.)

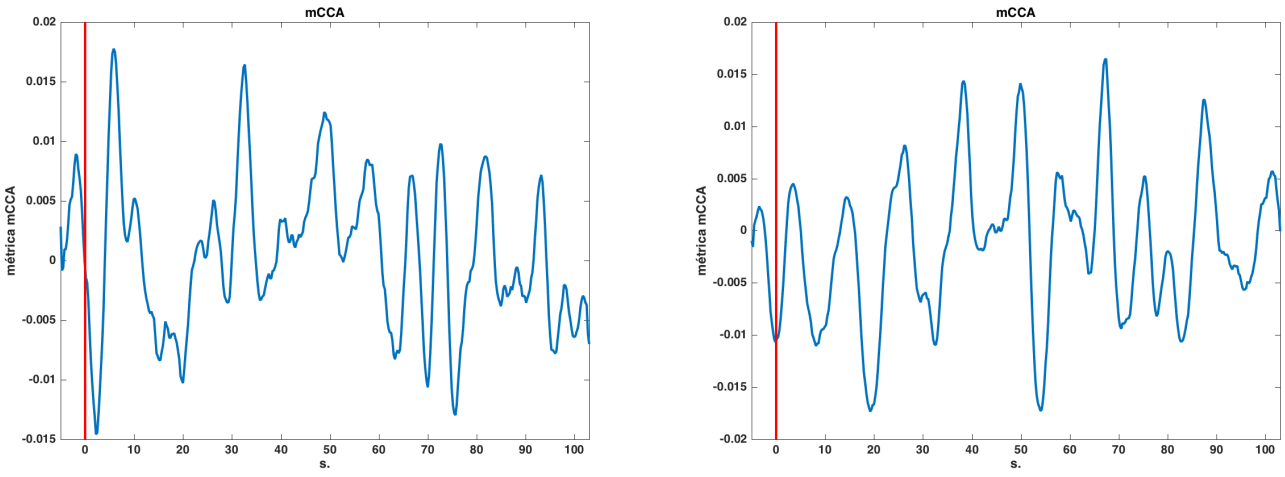


Figura 5.3: Resultado mCCA vídeo Esquí con PCA (izda.) y sin PCA (dcha.)

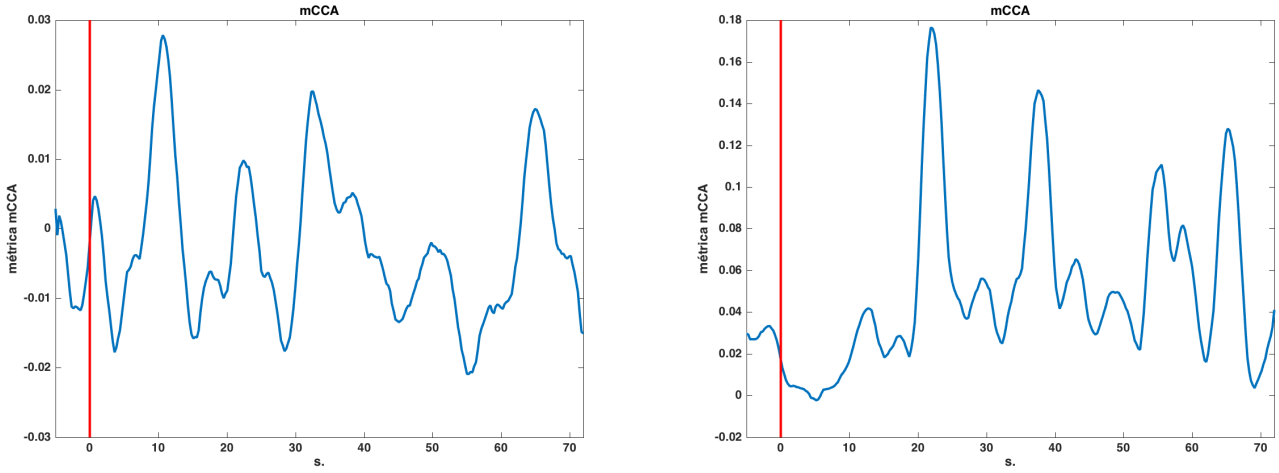


Figura 5.4: Resultado mCCA vídeo Prinzenzeit con PCA (izda.) y sin PCA (dcha.)

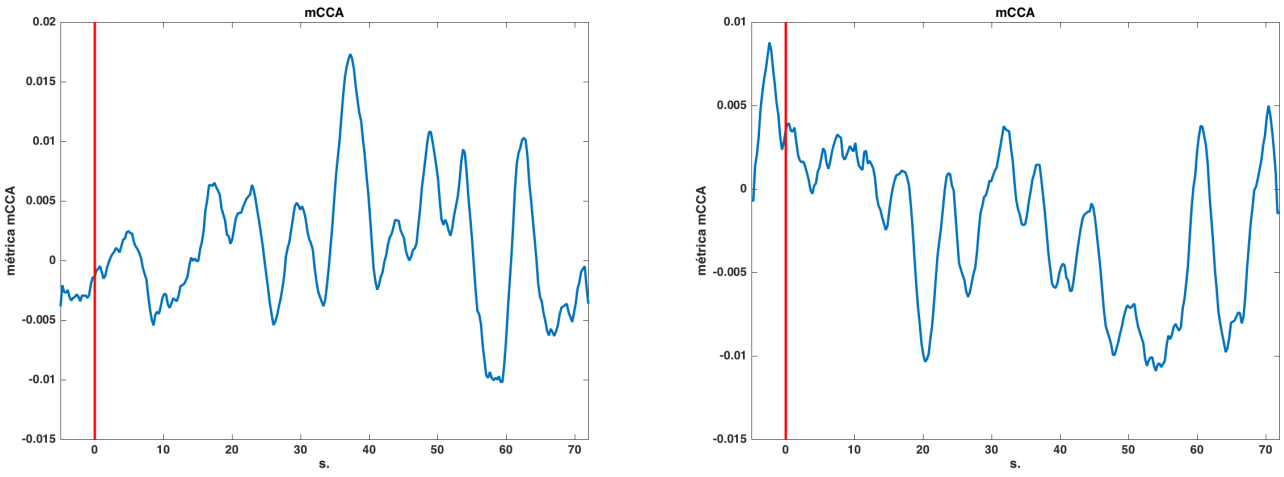


Figura 5.5: Resultado mCCA vídeo Paranormal con PCA (izda.) y sin PCA (dcha.)

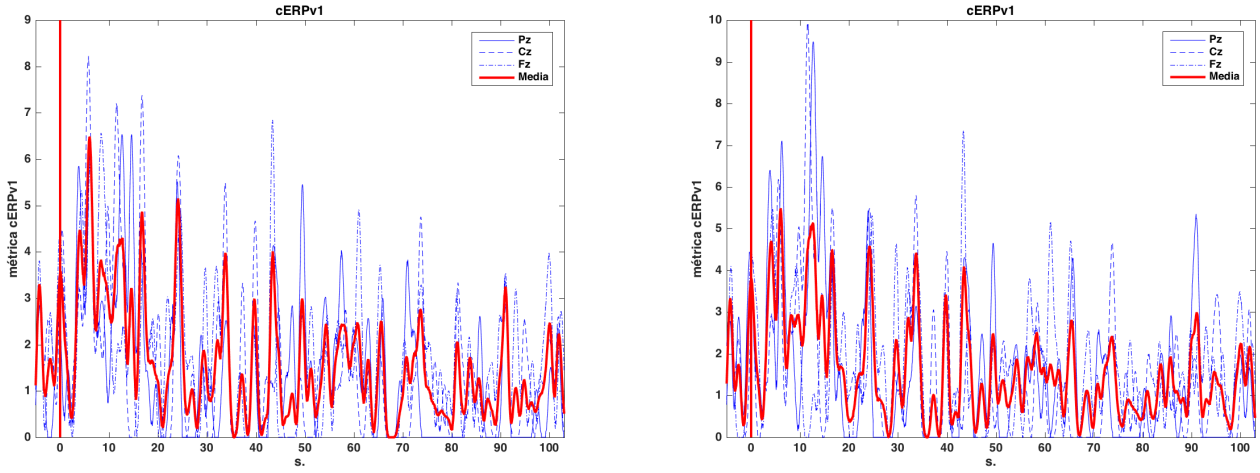


Figura 5.6: Resultado cERPv1 vídeo REC con PCA (izda.) y sin PCA (dcha.)

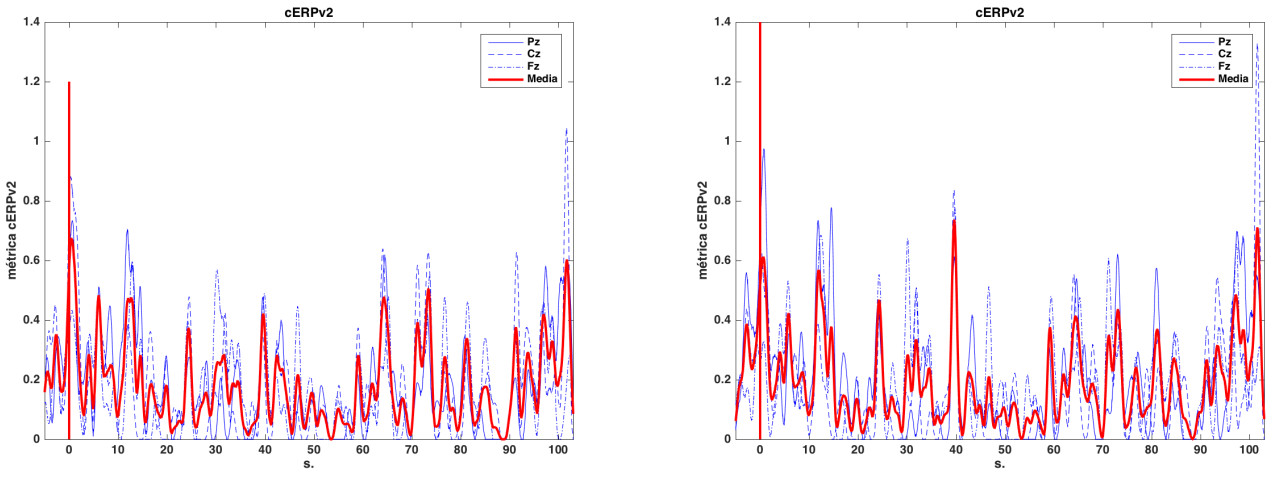


Figura 5.7: Resultado cERPv2 vídeo REC con PCA (izda.) y sin PCA (dcha.)

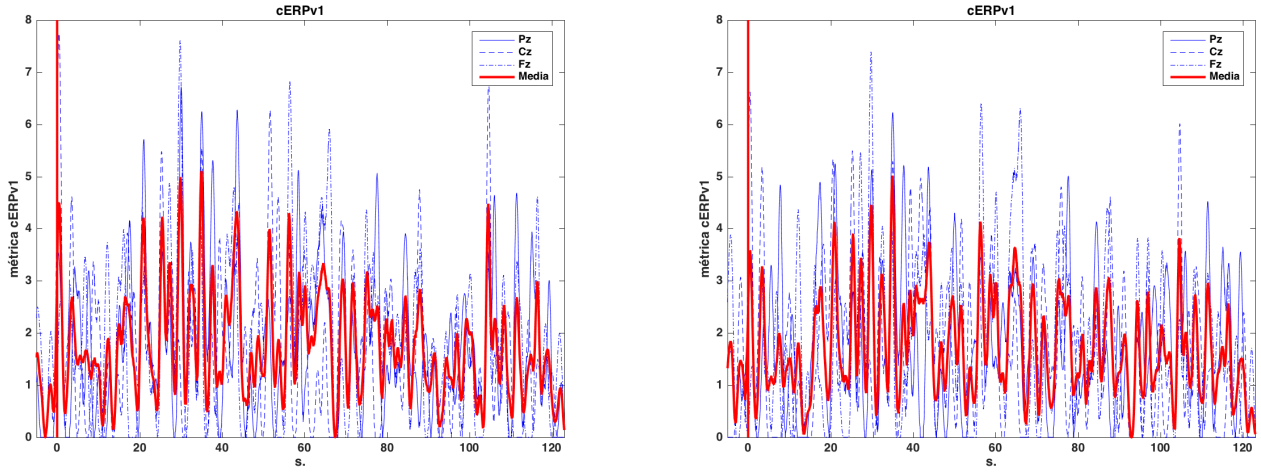


Figura 5.8: Resultado cERPv1 vídeo CocaCola con PCA (izda.) y sin PCA (dcha.)

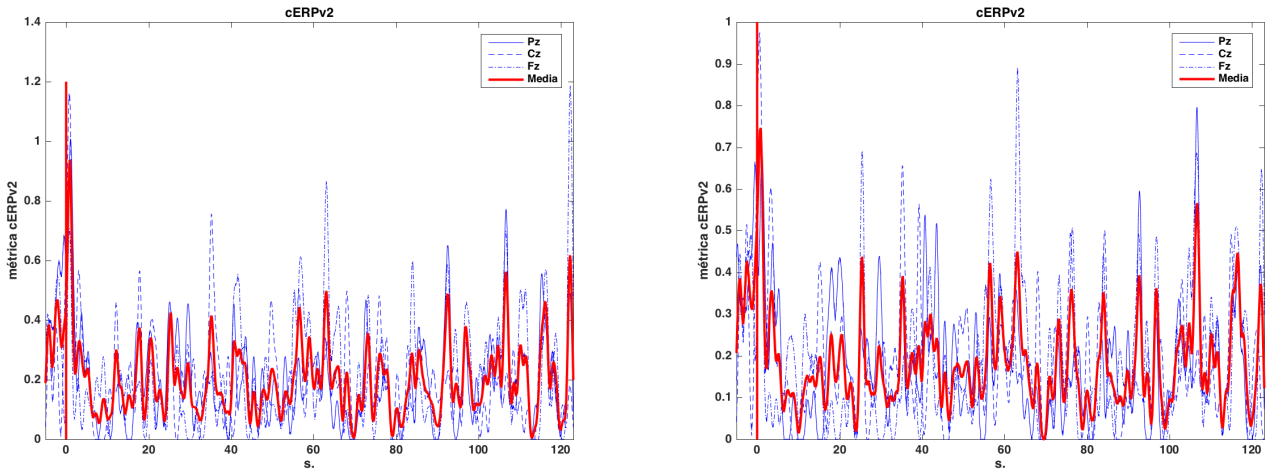


Figura 5.9: Resultado cERPv2 vídeo CocaCola con PCA (izda.) y sin PCA (dcha.)

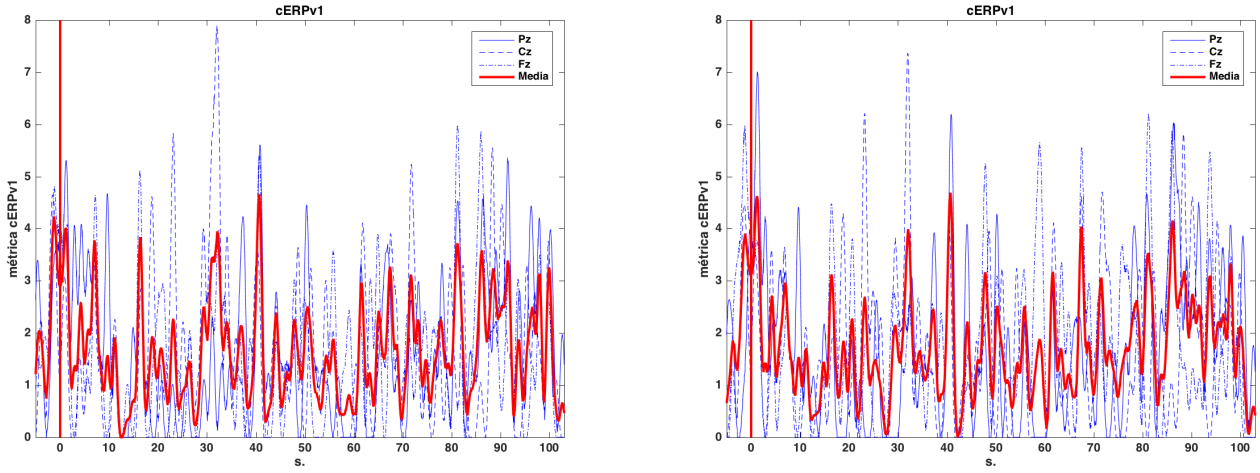


Figura 5.10: Resultado cERPv1 vídeo Esquí con PCA (izda.) y sin PCA (dcha.)

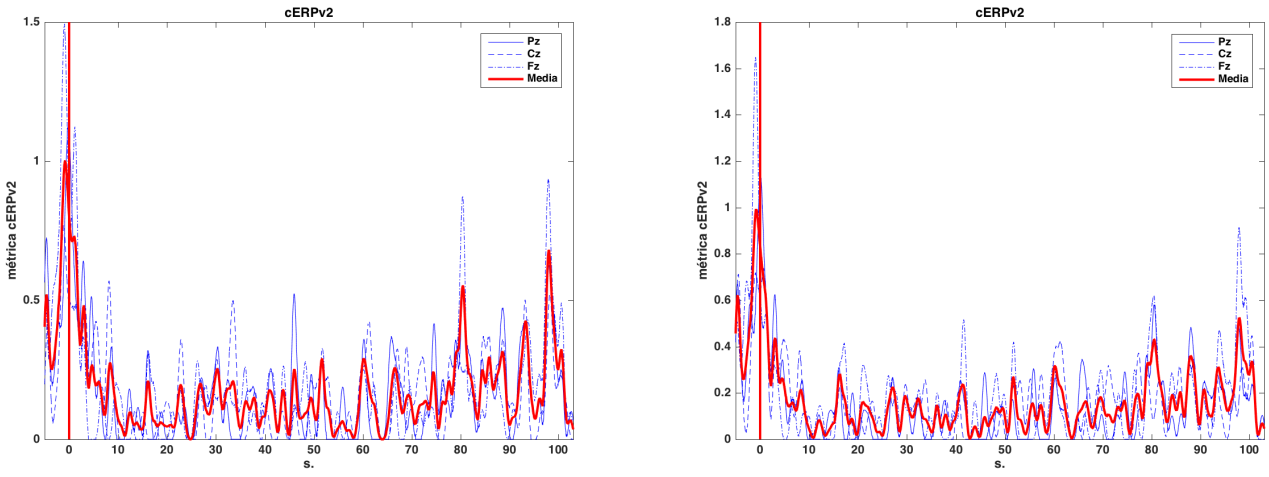


Figura 5.11: Resultado cERPv2 vídeo Esquí con PCA (izda.) y sin PCA (dcha.)

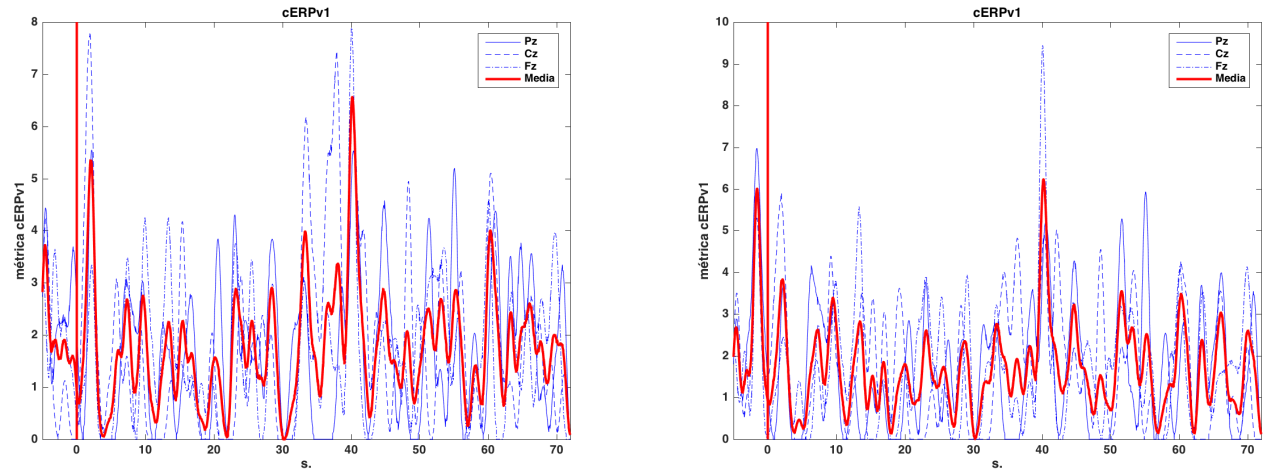


Figura 5.12: Resultado cERPv1 vídeo Prinzenzeit con PCA (izda.) y sin PCA (dcha.)

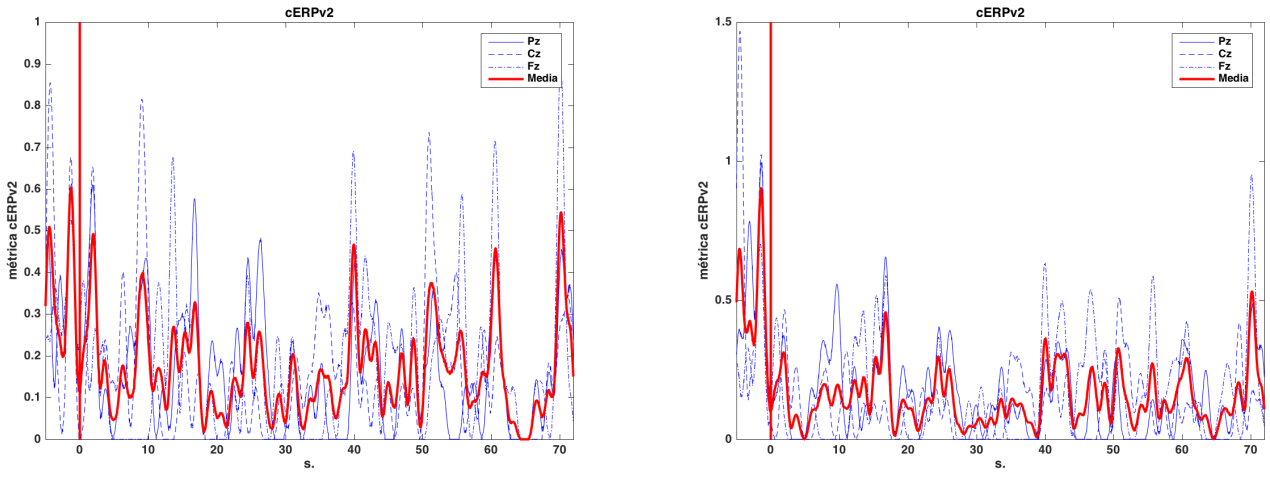


Figura 5.13: Resultado cERPv2 vídeo Prinzenzeit con PCA (izda.) y sin PCA (dcha.)

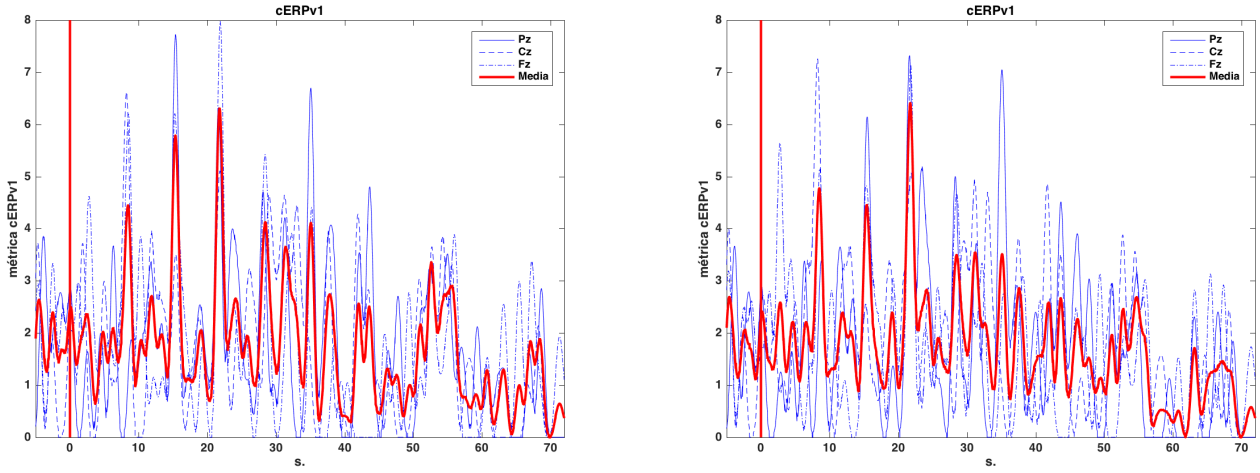


Figura 5.14: Resultado cERPv1 vídeo Paranormal con PCA (izda.) y sin PCA (dcha.)

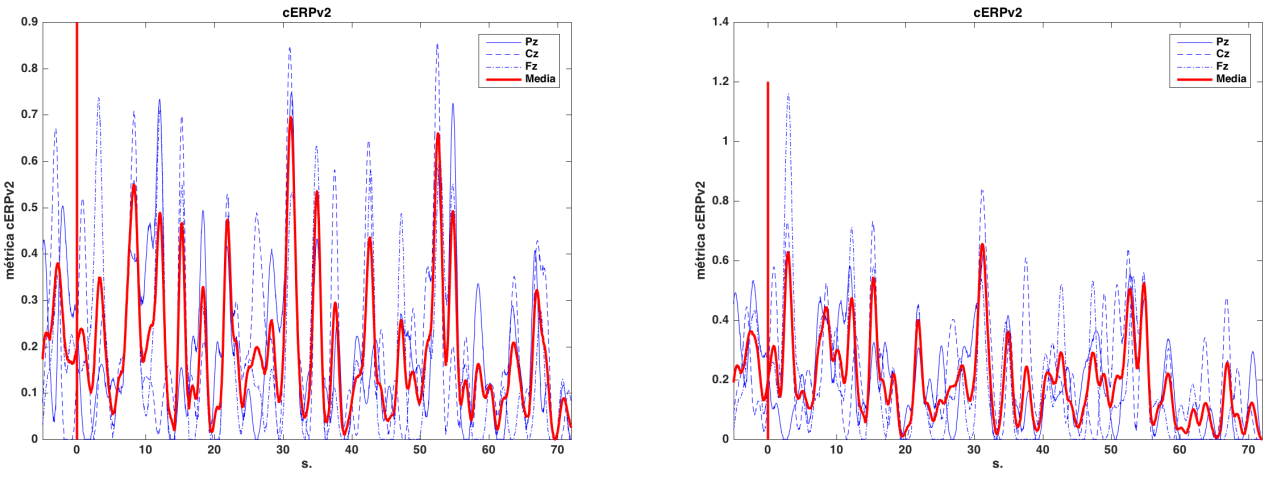


Figura 5.15: Resultado cERPv2 vídeo Paranormal con PCA (izda.) y sin PCA (dcha.)

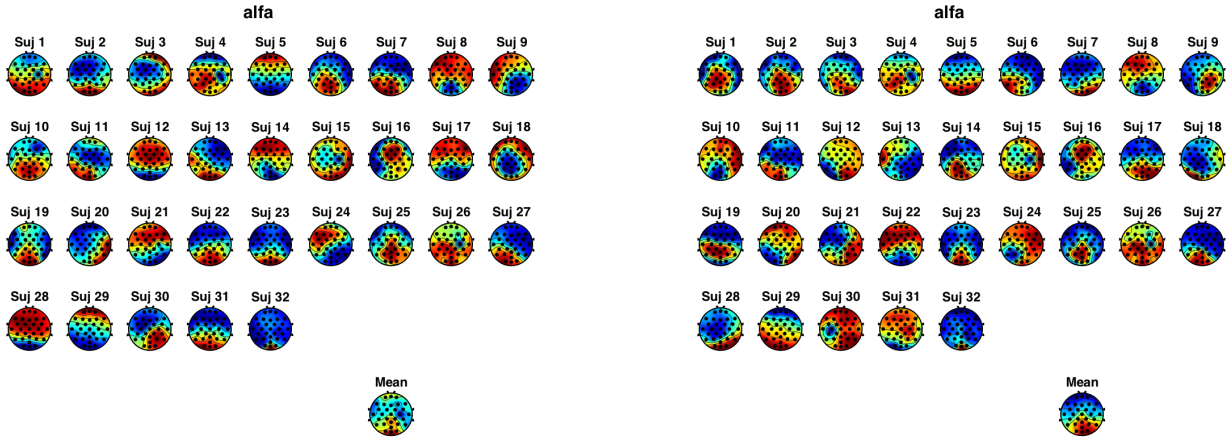


Figura 5.16: Activaciones cerebrales cSPoC vídeo REC con PCA (izda.) y sin PCA (dcha.) en banda alfa

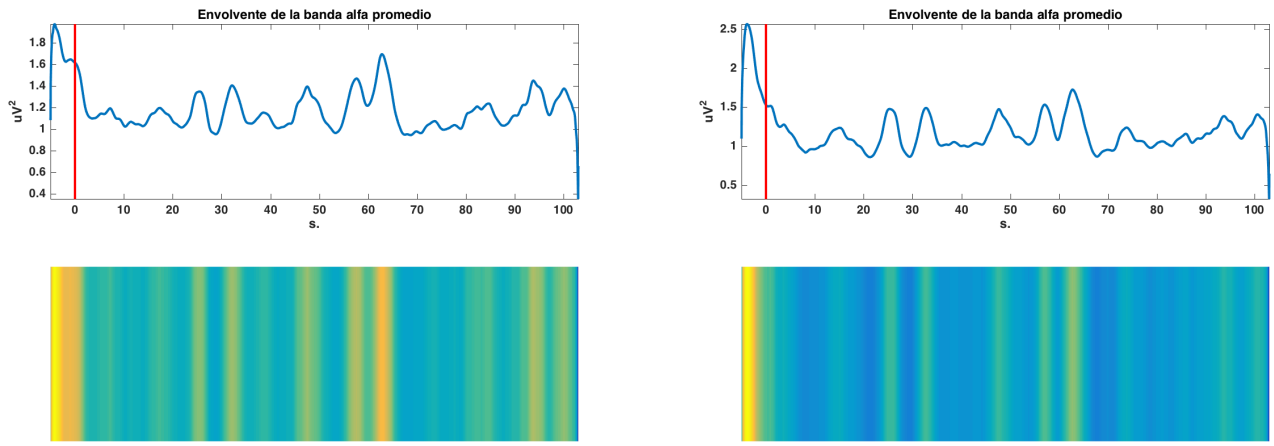


Figura 5.17: Envolvente banda alfa vídeo REC con PCA (izda.) y sin PCA (dcha.)

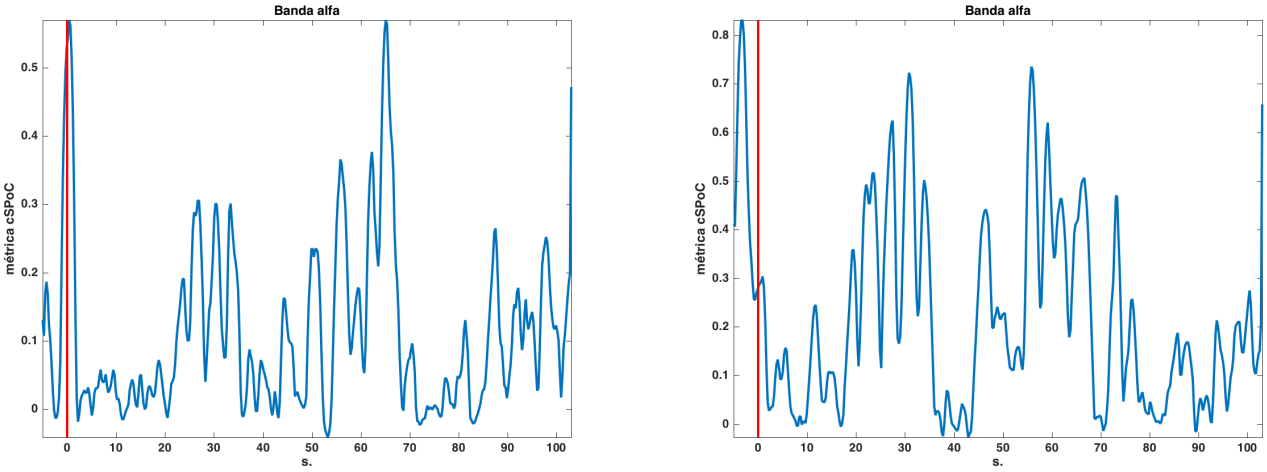


Figura 5.18: Resultados cSPoC vídeo REC con PCA (izda.) y sin PCA (dcha.) en banda alfa

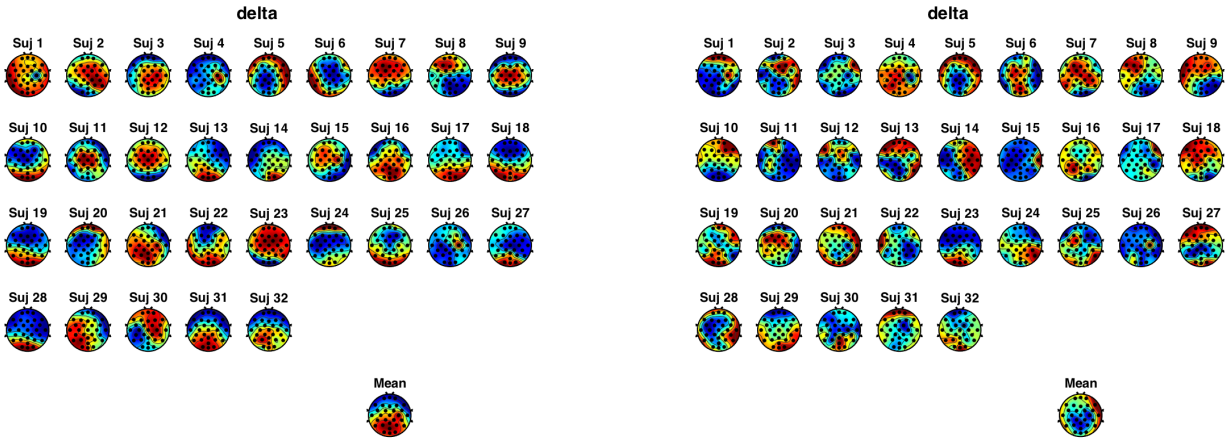


Figura 5.19: Activaciones cerebrales cSPoC vídeo REC con PCA (izda.) y sin PCA (dcha.) en banda delta

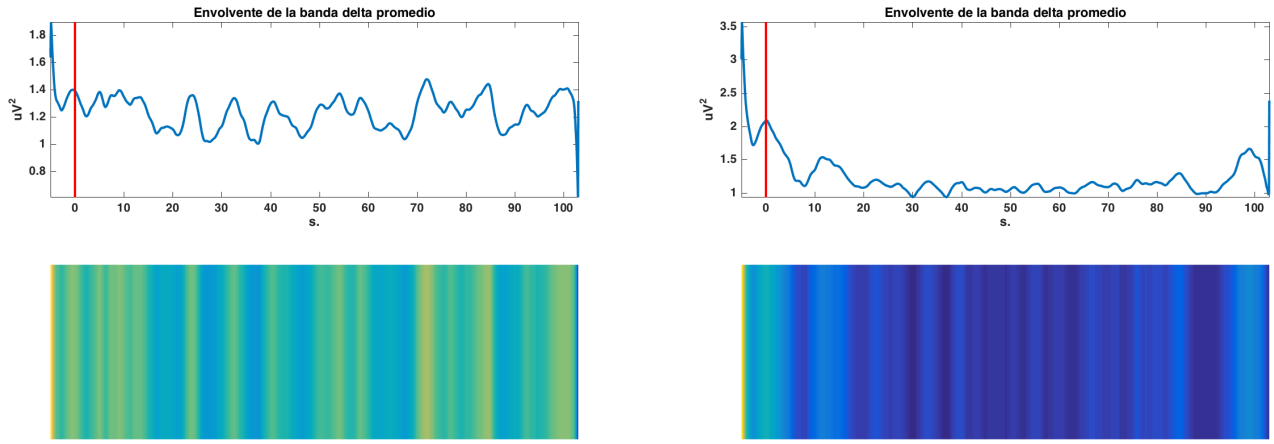


Figura 5.20: Envolvente banda delta vídeo REC con PCA (izda.) y sin PCA (dcha.)

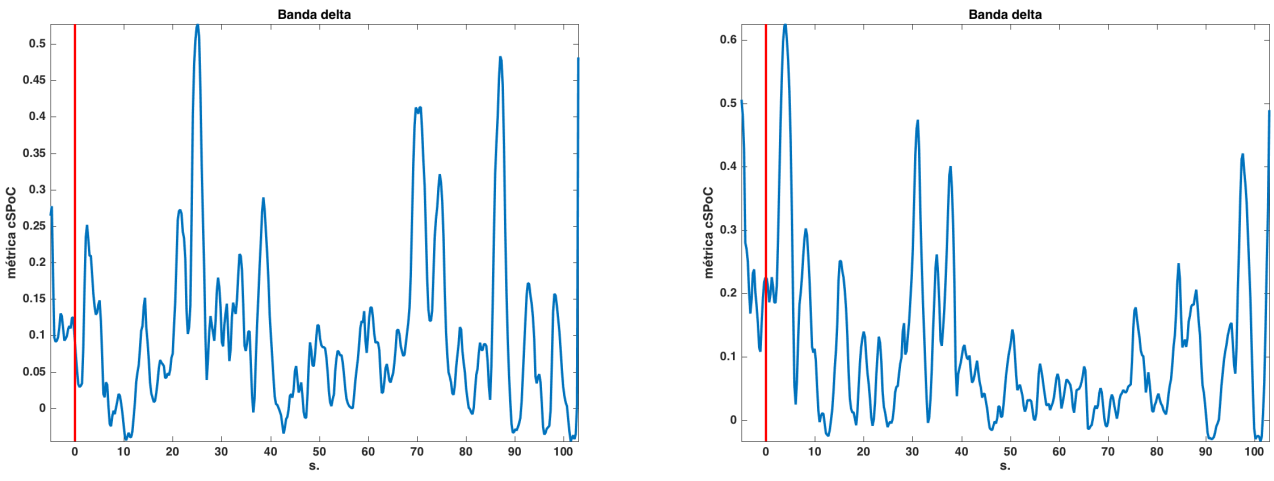


Figura 5.21: Resultados cSPoC vídeo REC con PCA (izda.) y sin PCA (dcha.) en banda delta

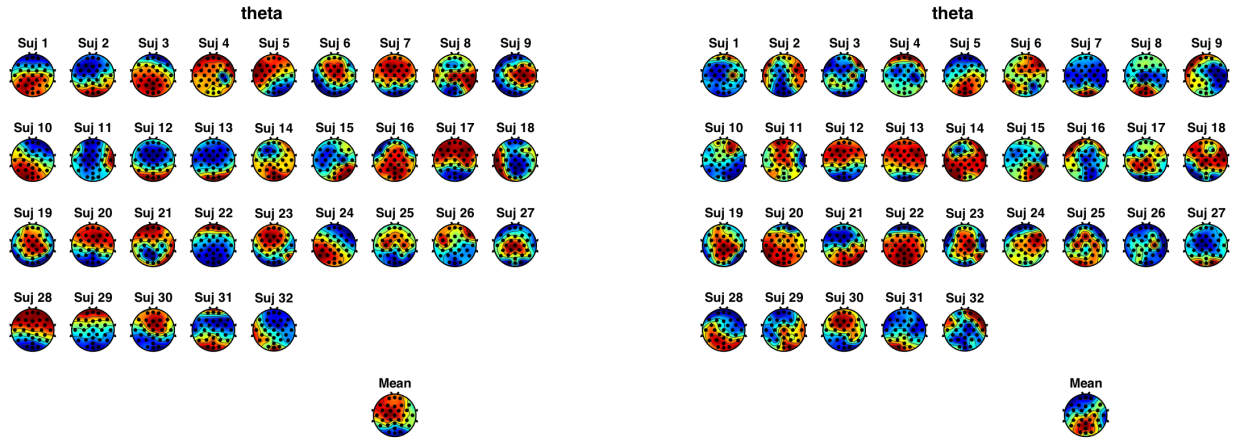


Figura 5.22: Activaciones cerebrales cSPoC vídeo REC con PCA (izda.) y sin PCA (dcha.) en banda theta

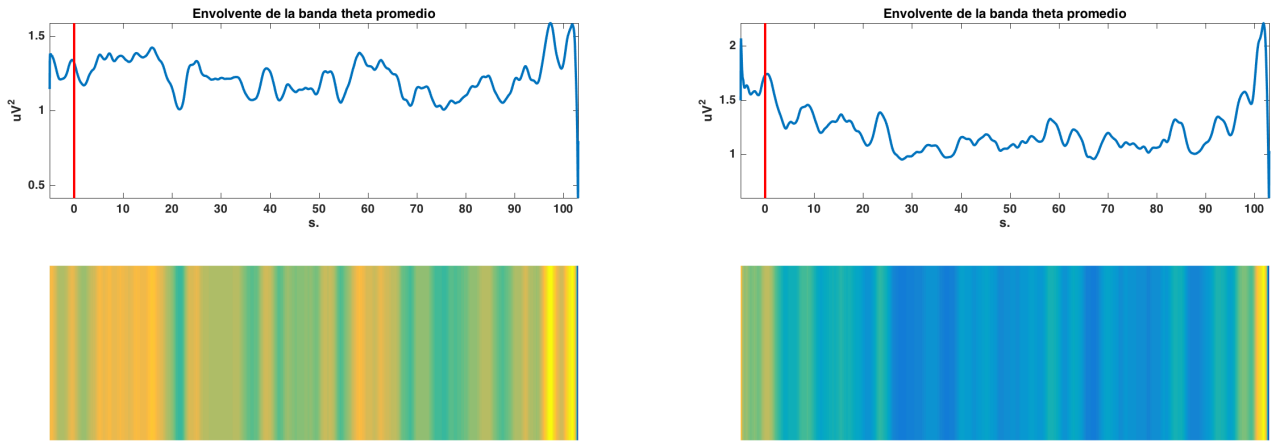


Figura 5.23: Envolvente banda theta vídeo REC con PCA (izda.) y sin PCA (dcha.)

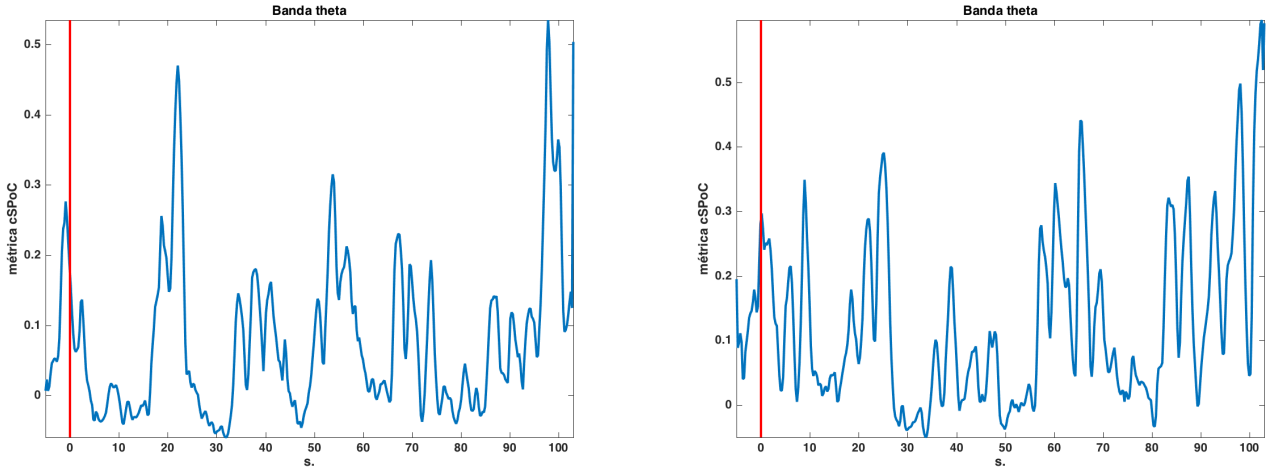


Figura 5.24: Resultados cSPoC vídeo REC con PCA (izda.) y sin PCA (dcha.) en banda theta

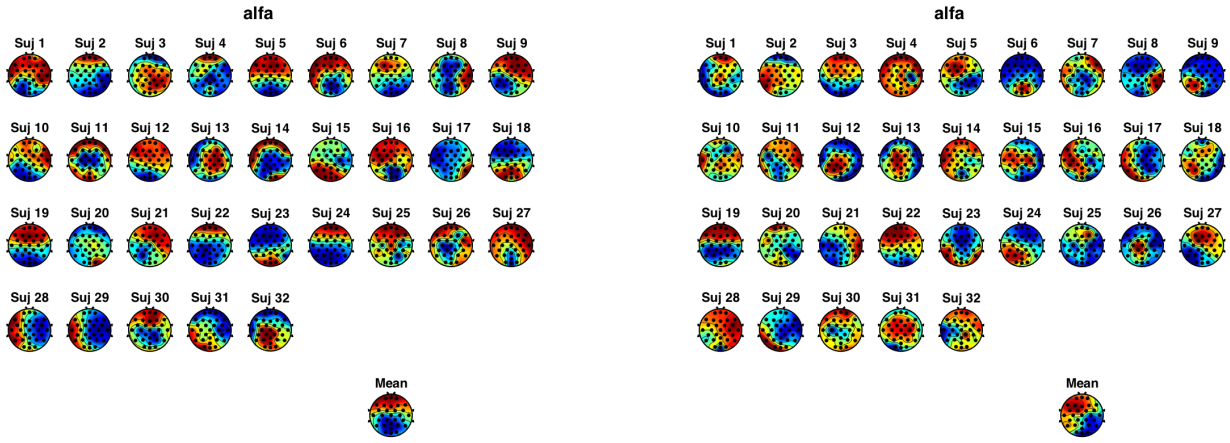


Figura 5.25: Activaciones cerebrales cSPoC vídeo CocaCola con PCA (izda.) y sin PCA (dcha.) en banda alfa

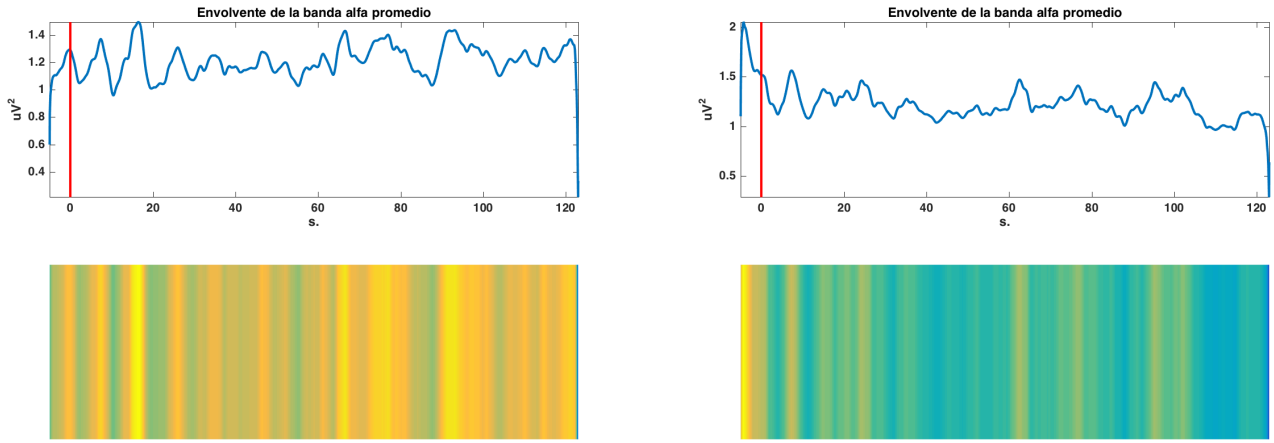


Figura 5.26: Envolvente banda alfa vídeo CocaCola con PCA (izda.) y sin PCA (dcha.)

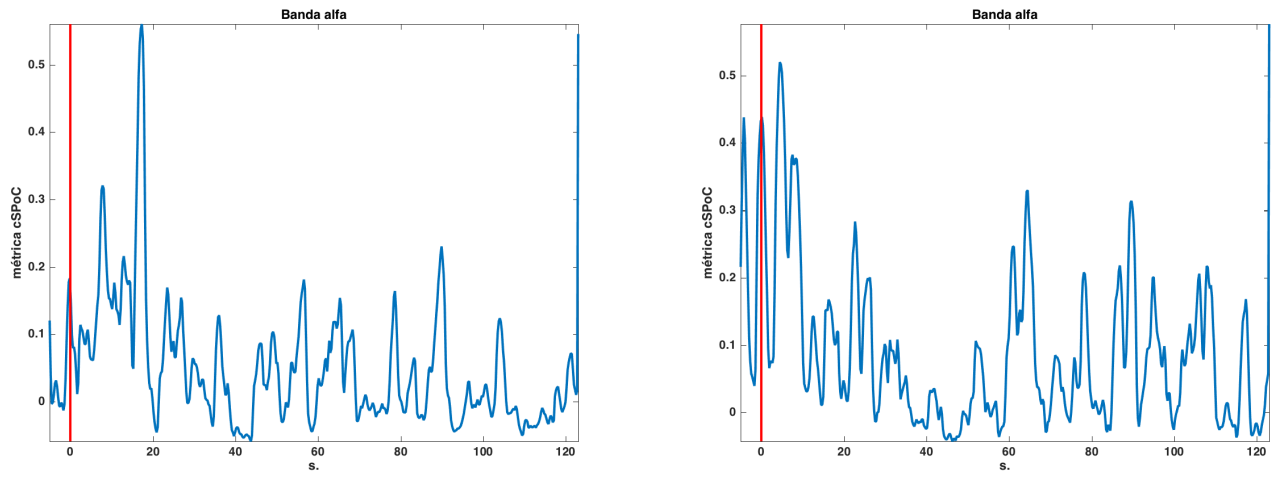


Figura 5.27: Resultados cSPoC vídeo CocaCola con PCA (izda.) y sin PCA (dcha.) en banda alfa

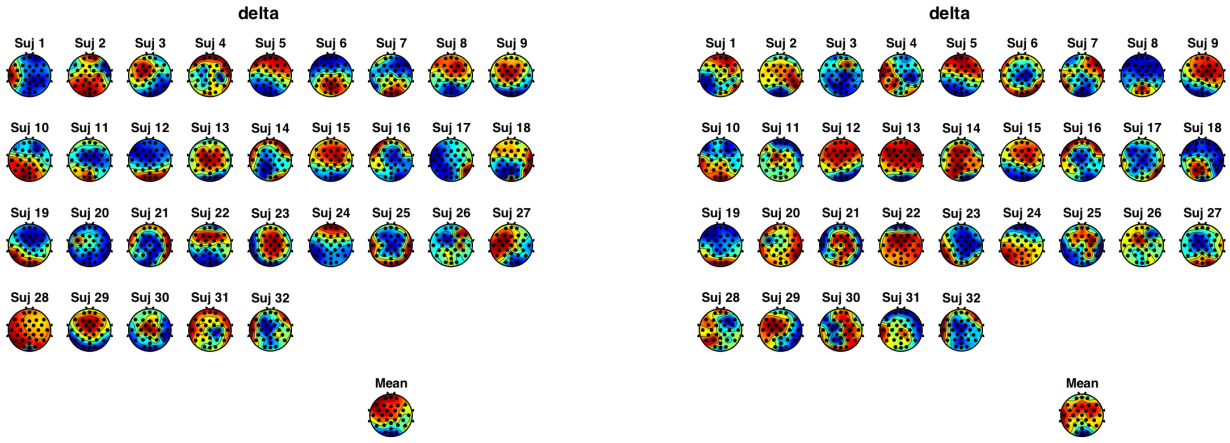


Figura 5.28: Activaciones cerebrales cSPoC vídeo CocaCola con PCA (izda.) y sin PCA (dcha.) en banda delta

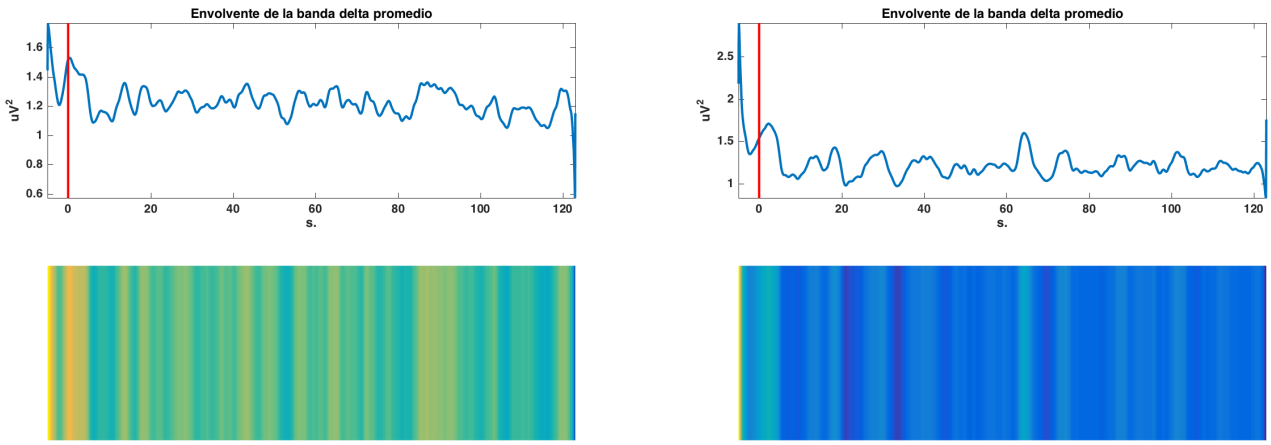


Figura 5.29: Envolvente banda delta vídeo CocaCola con PCA (izda.) y sin PCA (dcha.)

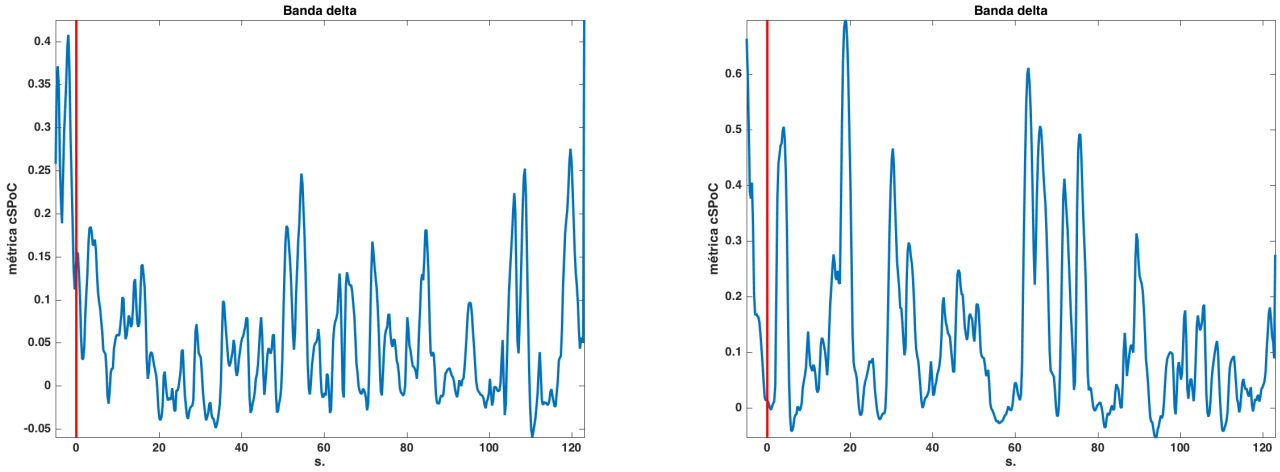


Figura 5.30: Resultados cSPoC vídeo CocaCola con PCA (izda.) y sin PCA (dcha.) en banda delta

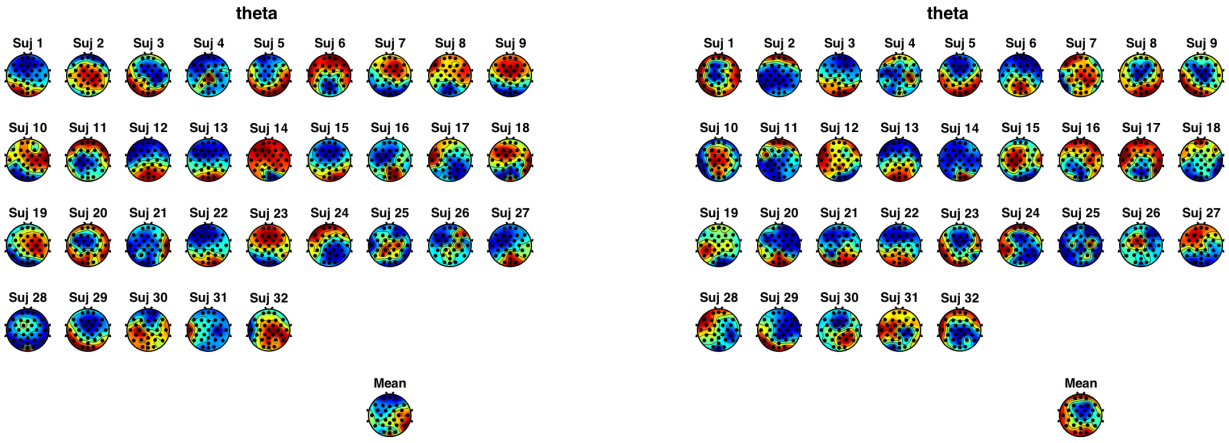


Figura 5.31: Activaciones cerebrales cSPoC vídeo CocaCola con PCA (izda.) y sin PCA (dcha.) en banda theta

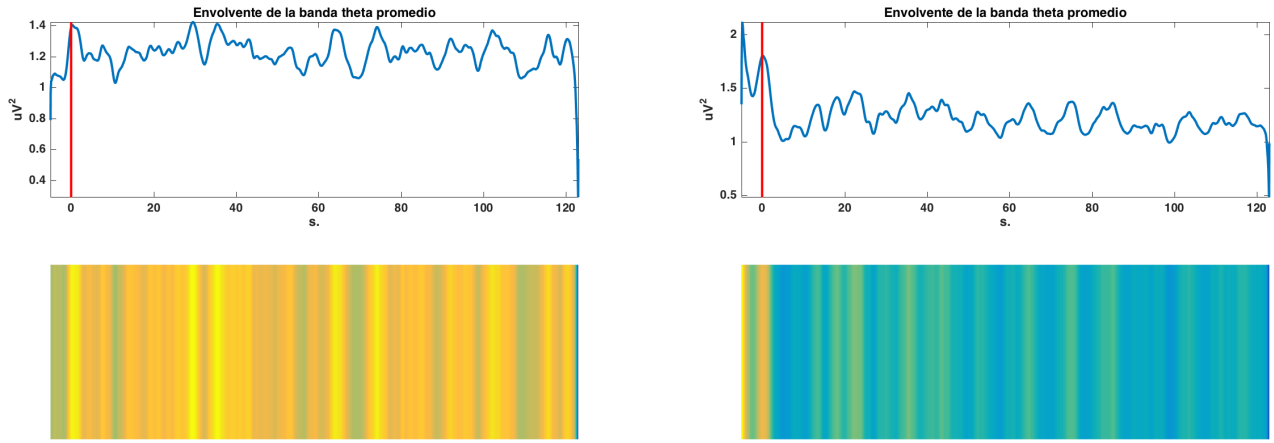


Figura 5.32: Envolvente banda theta vídeo CocaCola con PCA (izda.) y sin PCA (dcha.)

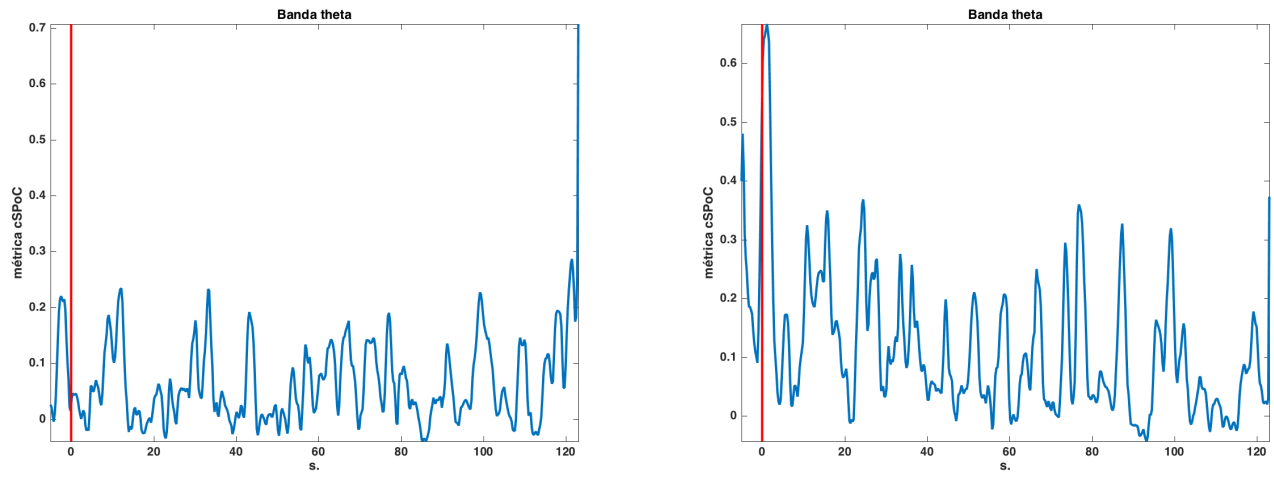


Figura 5.33: Resultados cSPoC vídeo CocaCola con PCA (izda.) y sin PCA (dcha.) en banda theta

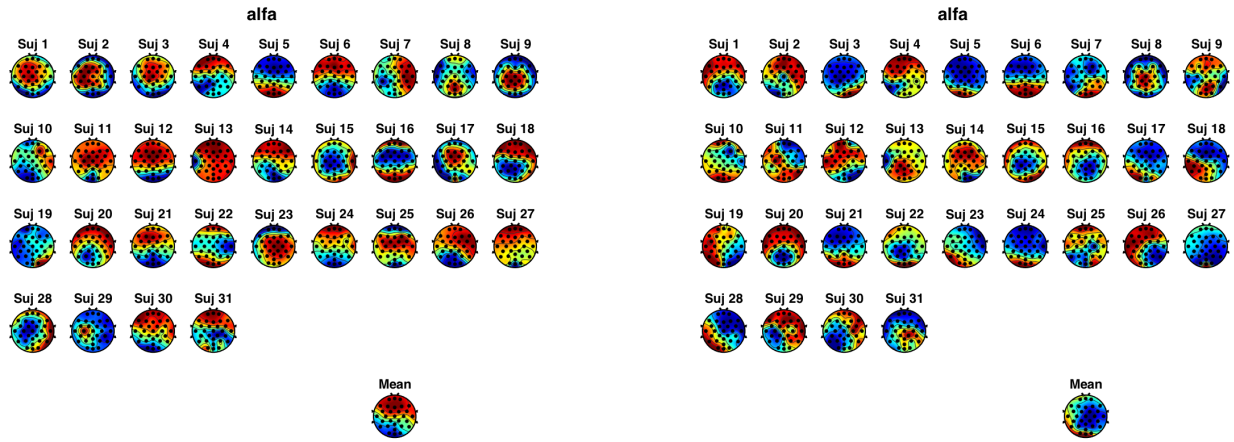


Figura 5.34: Activaciones cerebrales cSPoC vídeo Esquí con PCA (izda.) y sin PCA (dcha.) en banda alfa

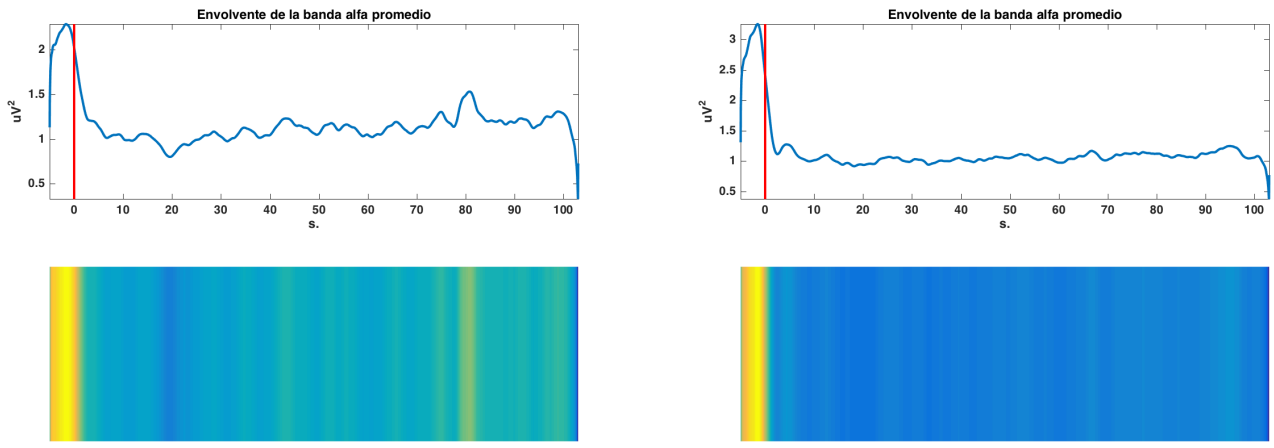


Figura 5.35: Envolvente banda alfa vídeo Esquí con PCA (izda.) y sin PCA (dcha.)

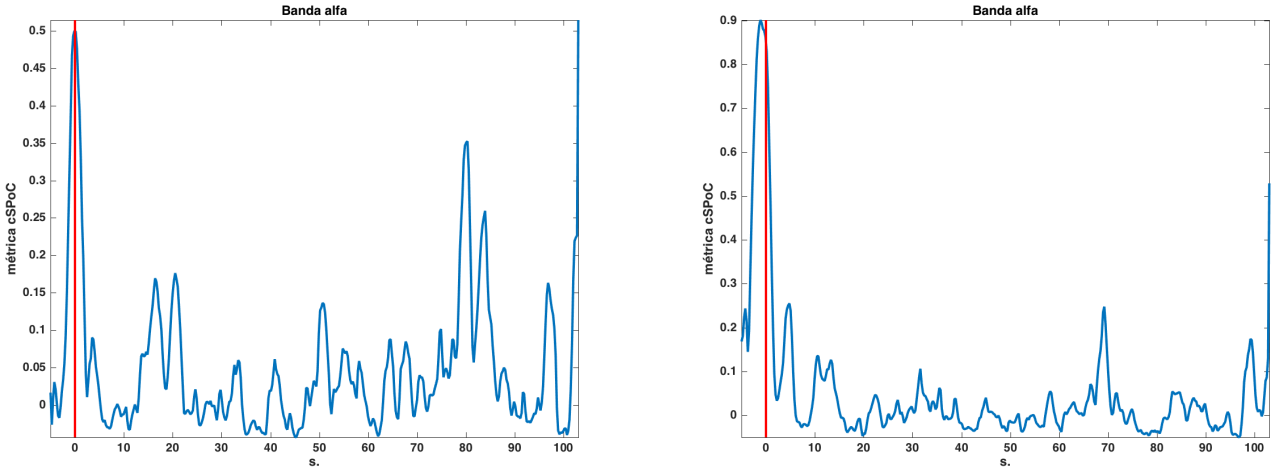


Figura 5.36: Resultados cSPoC vídeo Esquí con PCA (izda.) y sin PCA (dcha.) en banda alfa

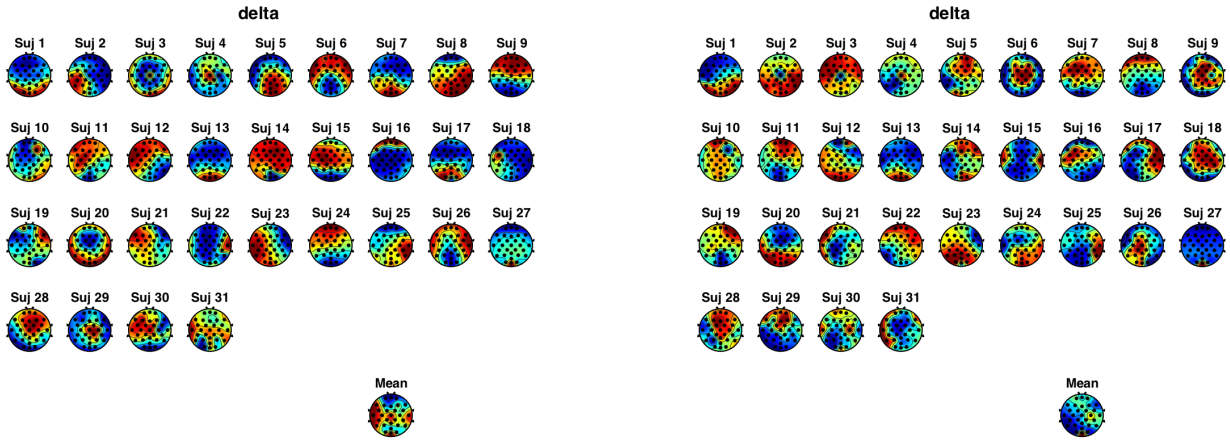


Figura 5.37: Activaciones cerebrales cSPoC vídeo Esquí con PCA (izda.) y sin PCA (dcha.) en banda delta

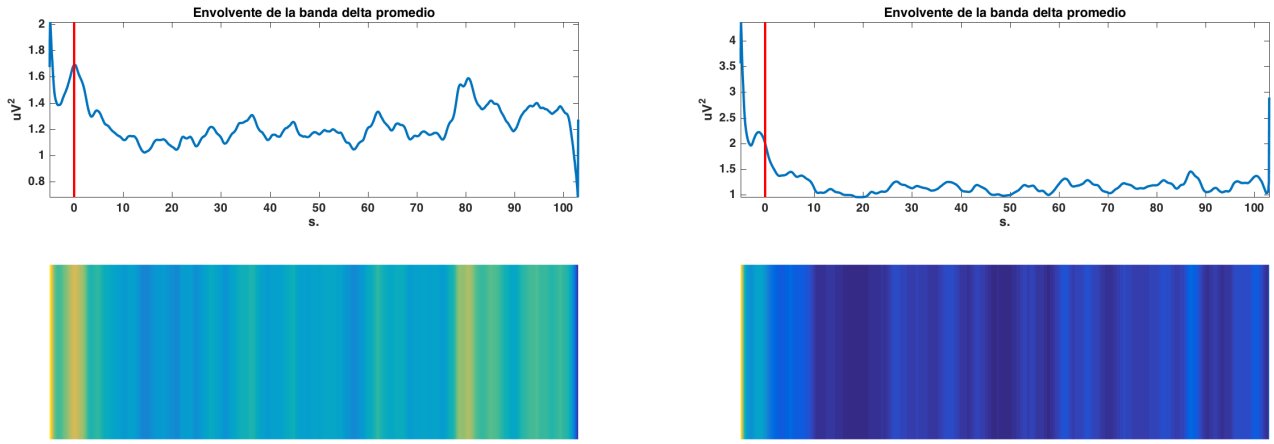


Figura 5.38: Envolvente banda delta vídeo Esquí con PCA (izda.) y sin PCA (dcha.)

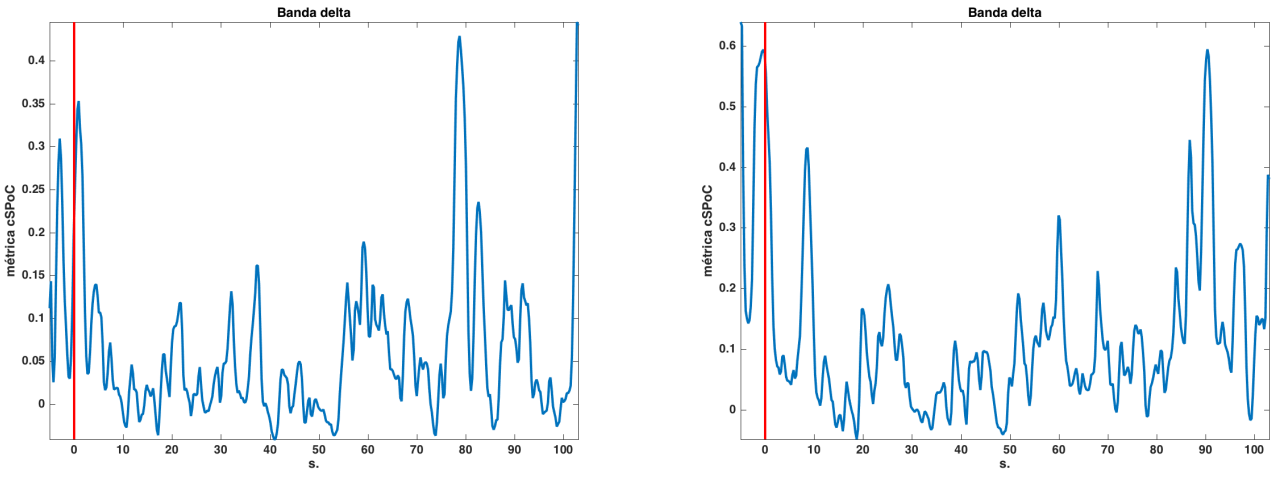


Figura 5.39: Resultados cSPoC vídeo Esquí con PCA (izda.) y sin PCA (dcha.) en banda delta

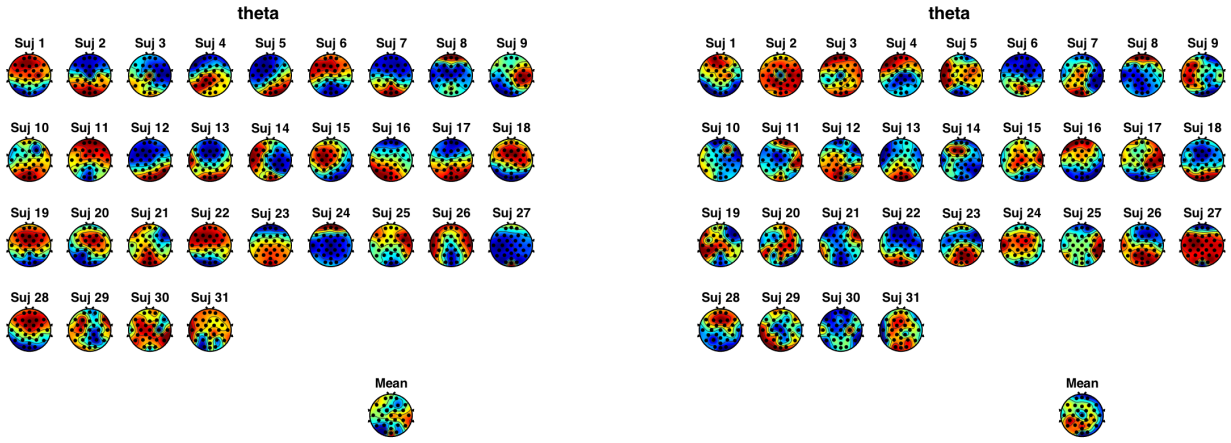


Figura 5.40: Activaciones cerebrales cSPoC vídeo Esquí con PCA (izda.) y sin PCA (dcha.) en banda theta

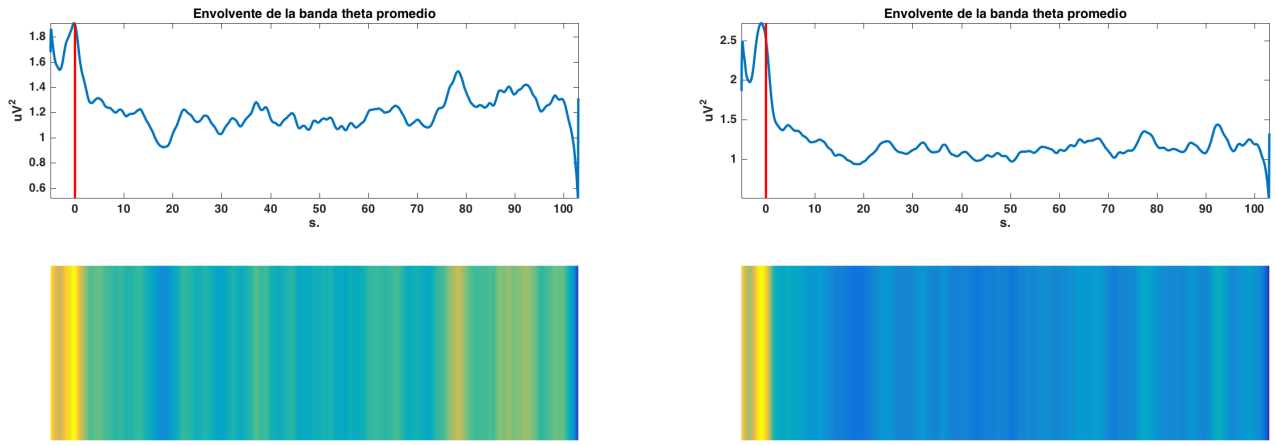


Figura 5.41: Envolvente banda theta vídeo Esquí con PCA (izda.) y sin PCA (dcha.)

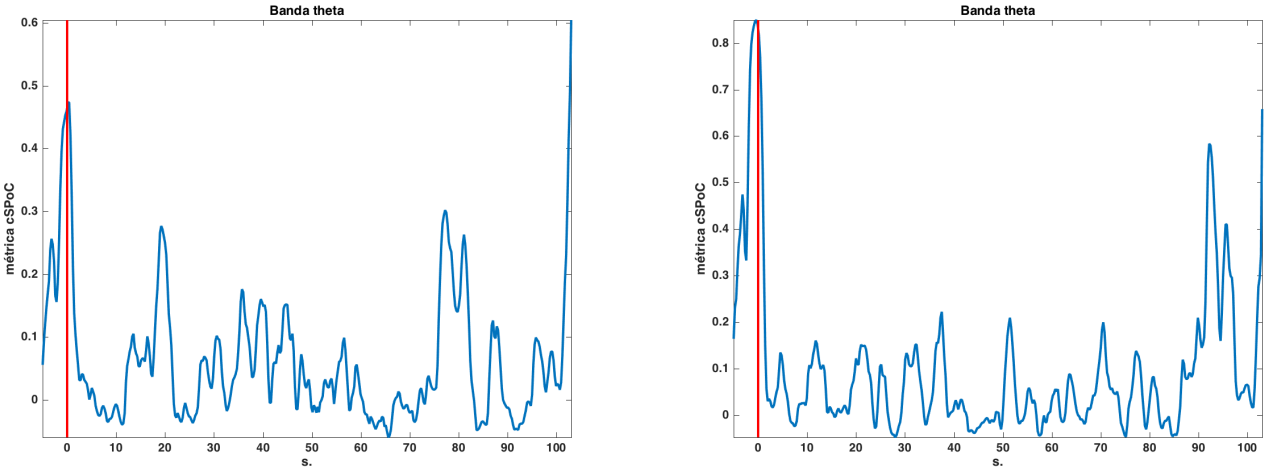


Figura 5.42: Resultados cSPoC vídeo Esquí con PCA (izda.) y sin PCA (dcha.) en banda theta

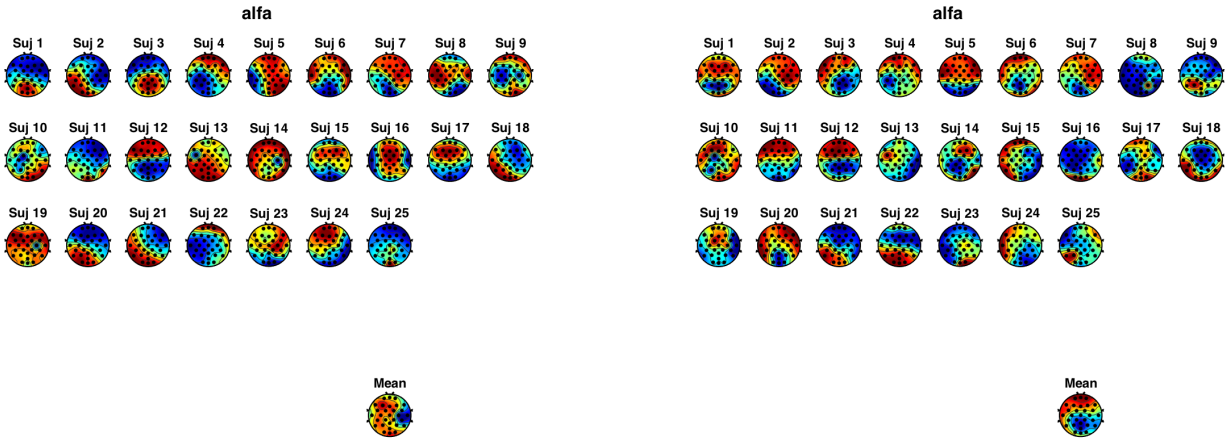


Figura 5.43: Activaciones cerebrales cSPoC vídeo Prinzenzeit con PCA (izda.) y sin PCA (dcha.) en banda alfa

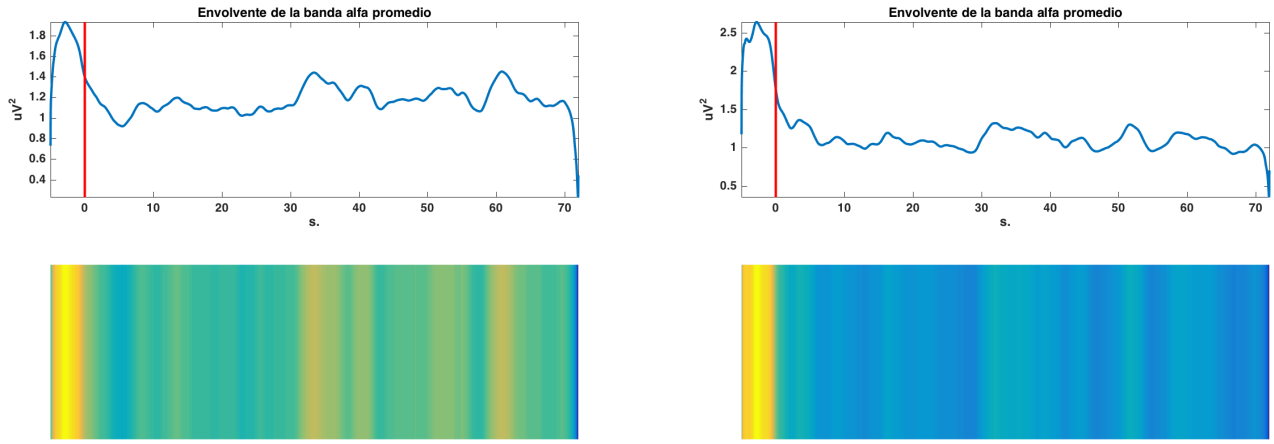


Figura 5.44: Envolvente banda alfa vídeo Prinzenzeit con PCA (izda.) y sin PCA (dcha.)

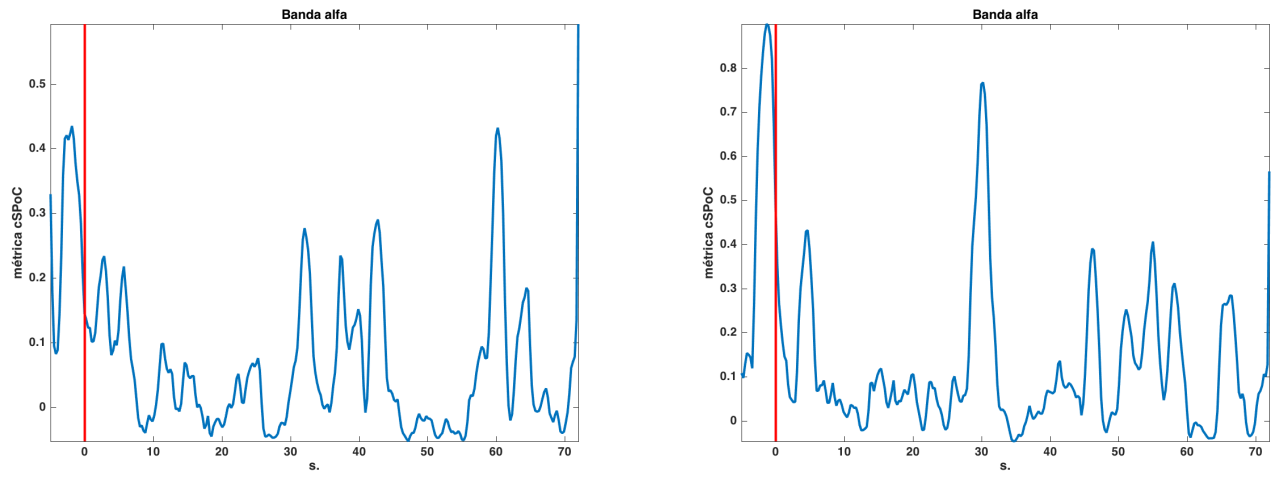


Figura 5.45: Resultados cSPoC vídeo Prinzenzeit con PCA (izda.) y sin PCA (dcha.) en banda alfa

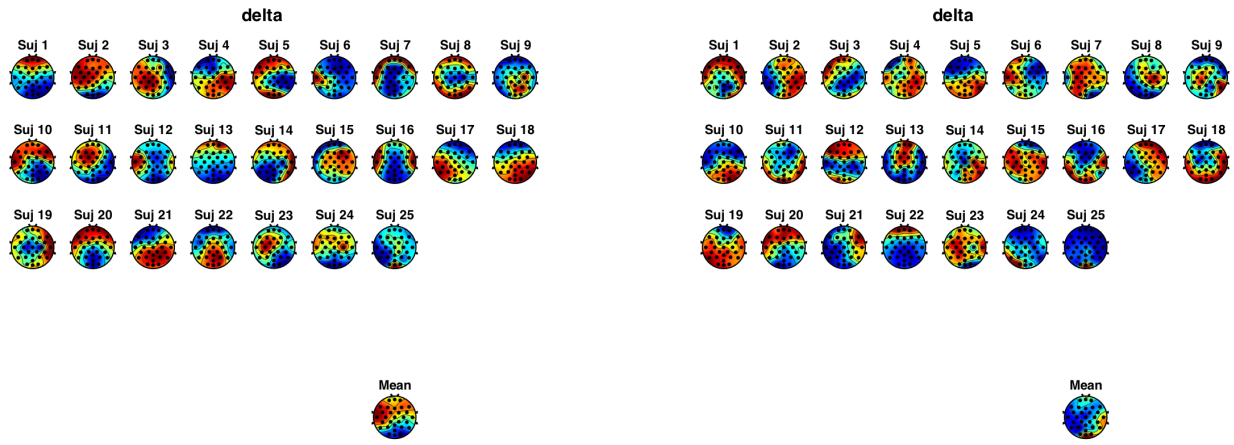


Figura 5.46: Activaciones cerebrales cSPoC vídeo Prinzenzeit con PCA (izda.) y sin PCA (dcha.) en banda delta

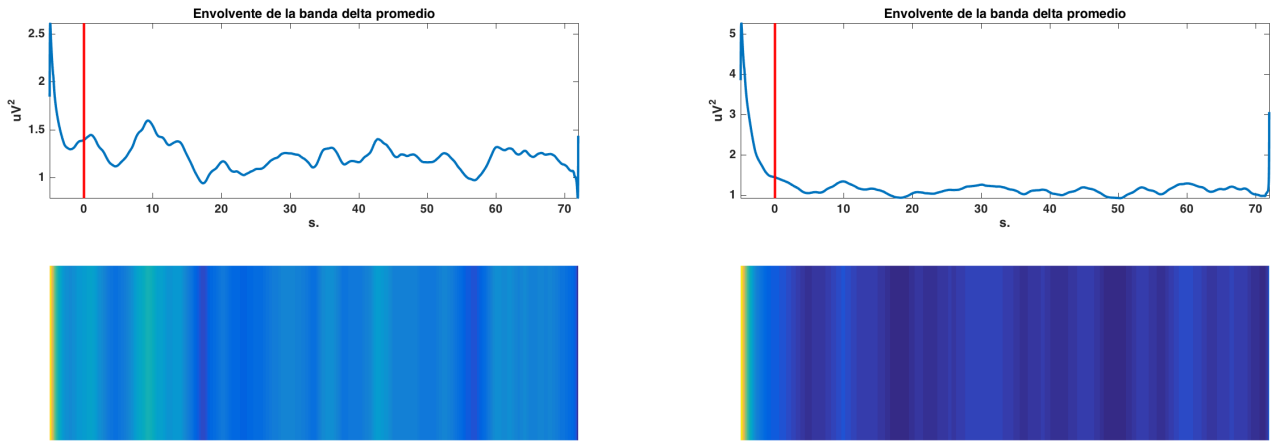


Figura 5.47: Envolvente banda delta vídeo Prinzenzeit con PCA (izda.) y sin PCA (dcha.)

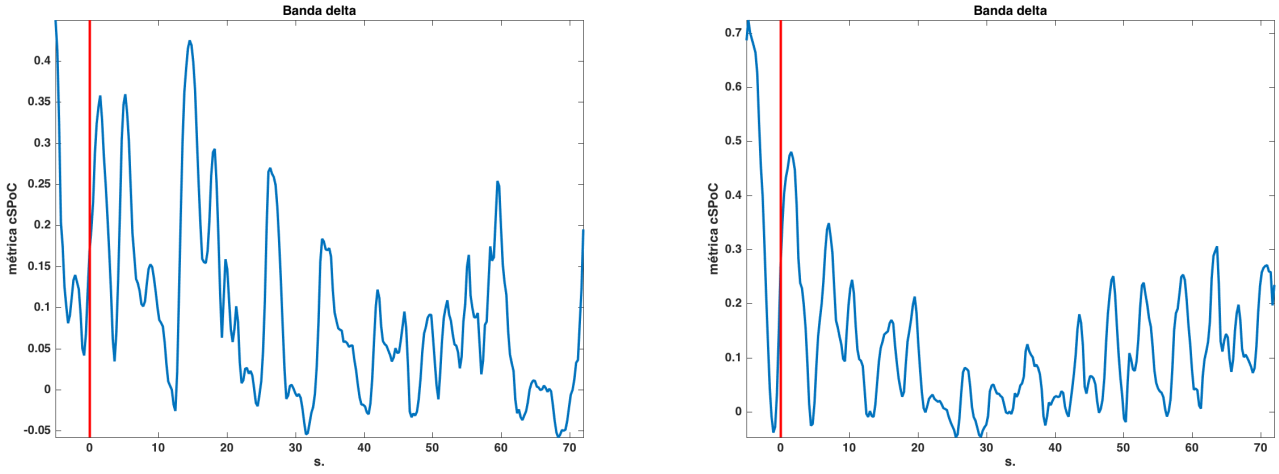


Figura 5.48: Resultados cSPoC vídeo Prinzenzeit con PCA (izda.) y sin PCA (dcha.) en banda delta

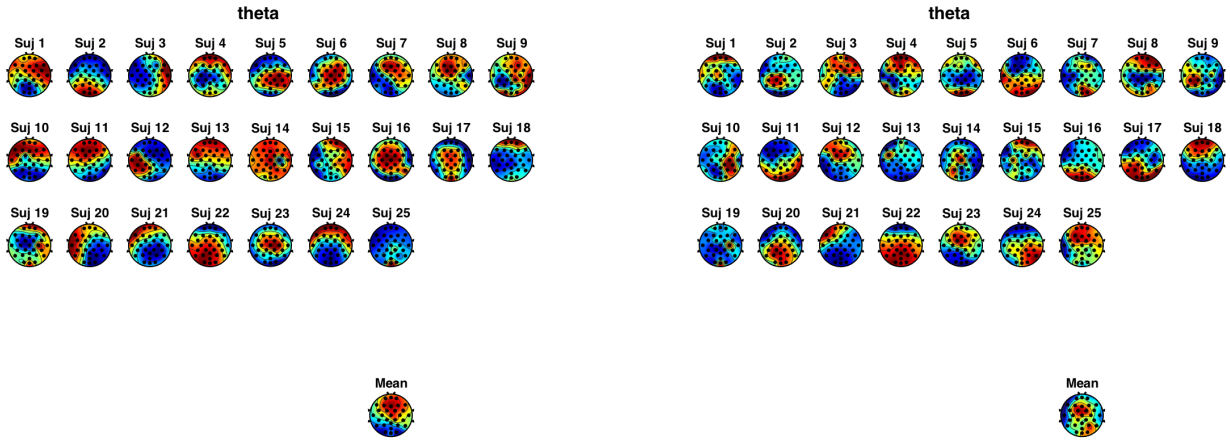


Figura 5.49: Activaciones cerebrales cSPoC vídeo Prinzenzeit con PCA (izda.) y sin PCA (dcha.) en banda theta

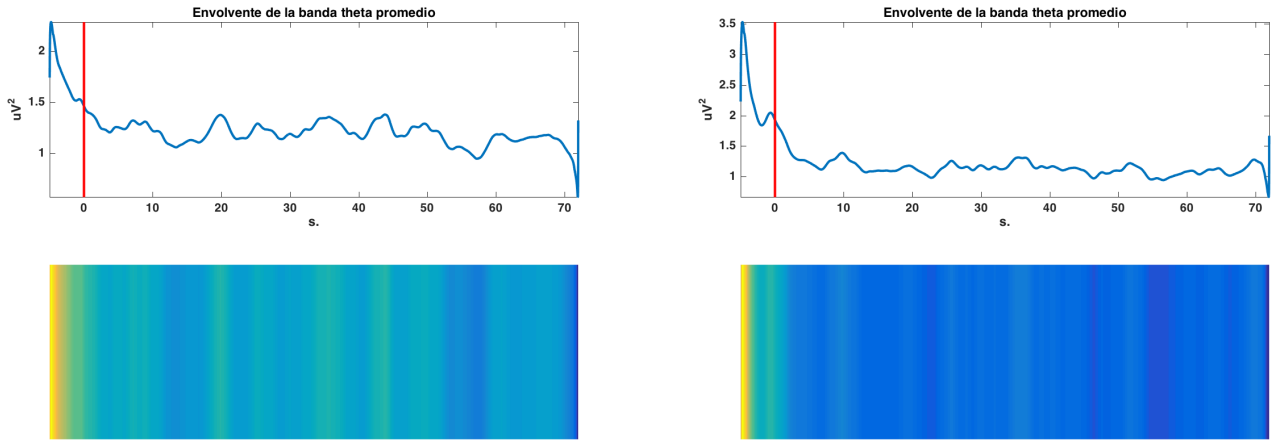


Figura 5.50: Envolvente banda theta vídeo Prinzenzeit con PCA (izda.) y sin PCA (dcha.)

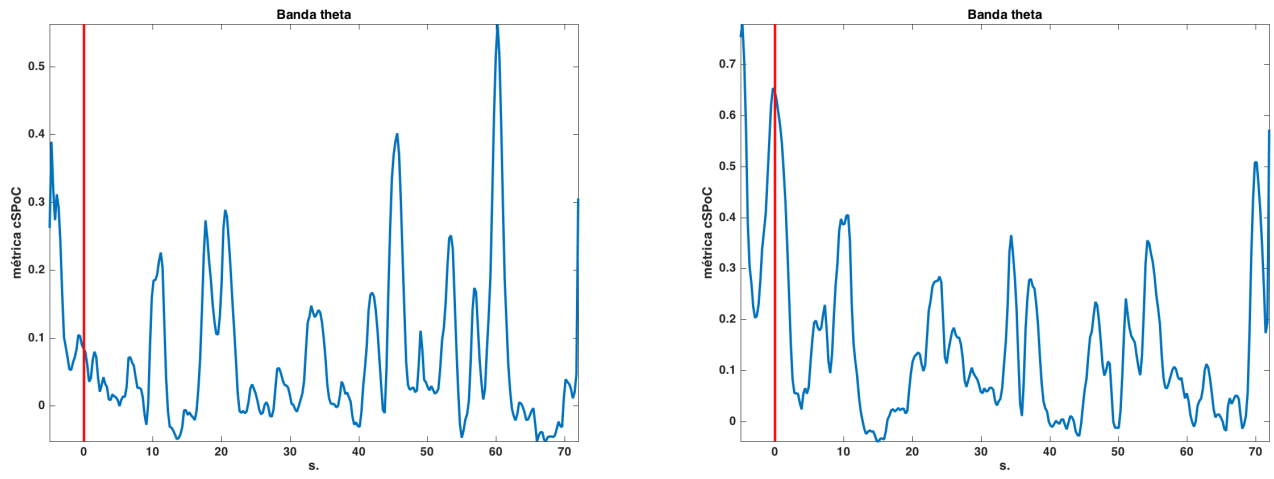


Figura 5.51: Resultados cSPoC vídeo Prinzenzeit con PCA (izda.) y sin PCA (dcha.) en banda theta

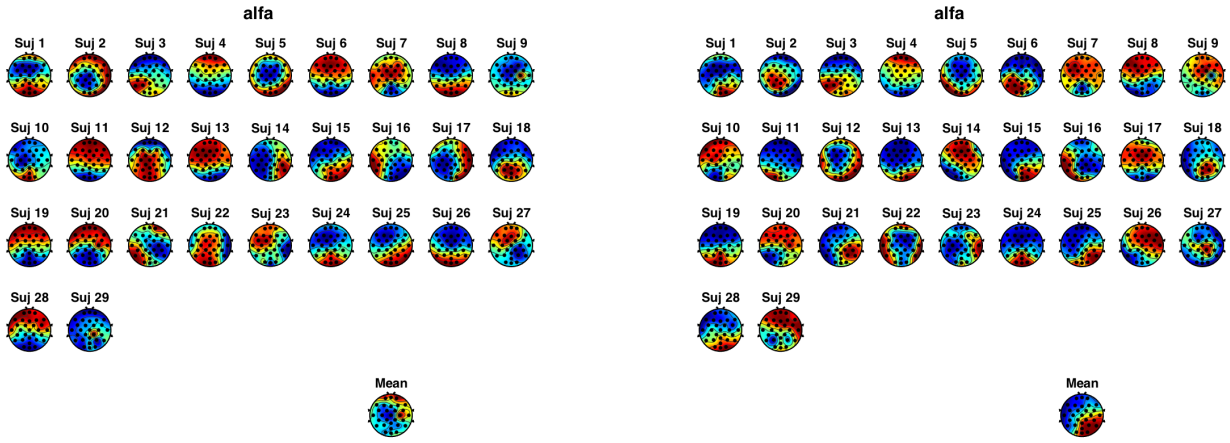


Figura 5.52: Activaciones cerebrales cSPoC vídeo Paranormal con PCA (izda.) y sin PCA (dcha.) en banda alfa

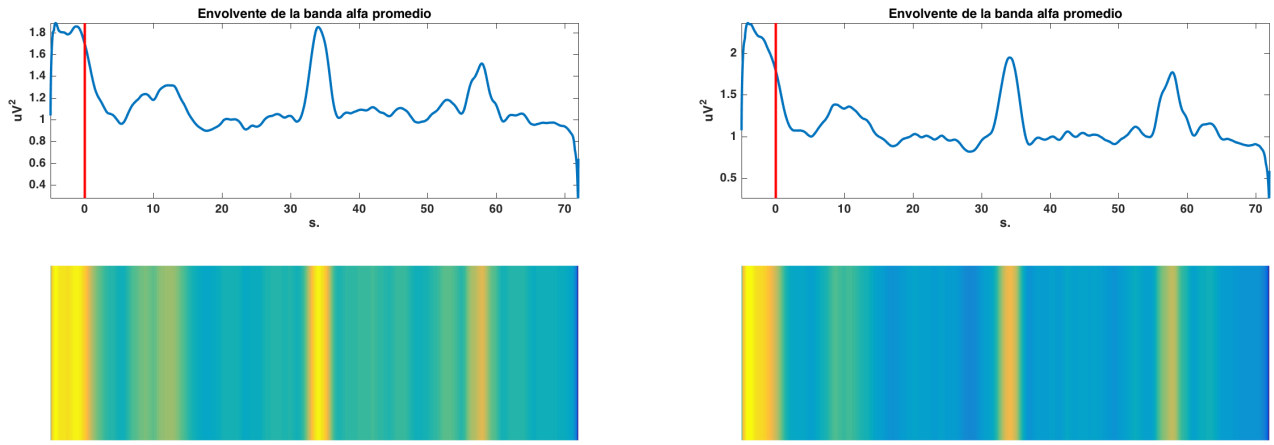


Figura 5.53: Envolvente banda alfa vídeo Paranormal con PCA (izda.) y sin PCA (dcha.)

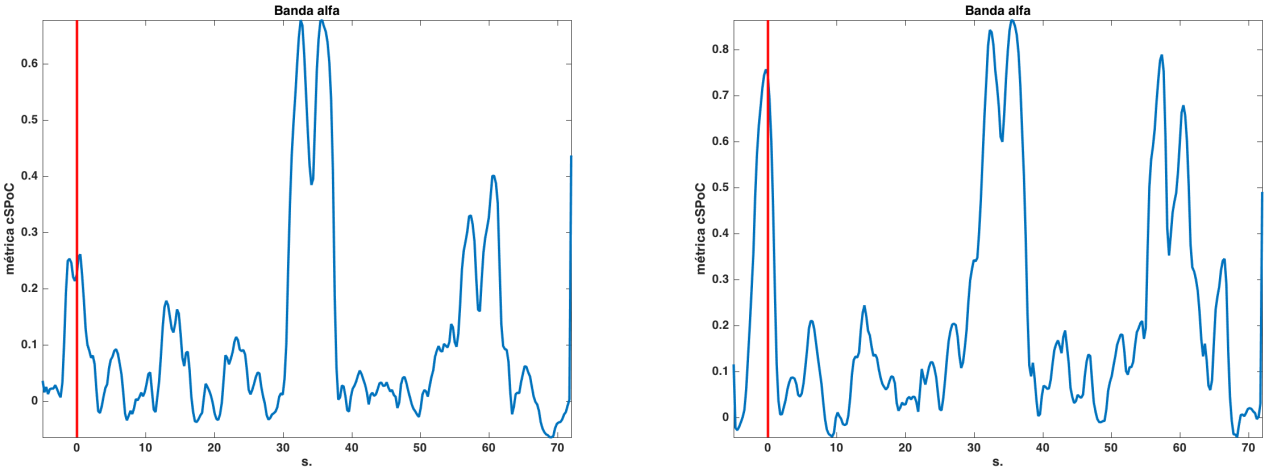


Figura 5.54: Resultados cSPoC vídeo Paranormal con PCA (izda.) y sin PCA (dcha.) en banda alfa

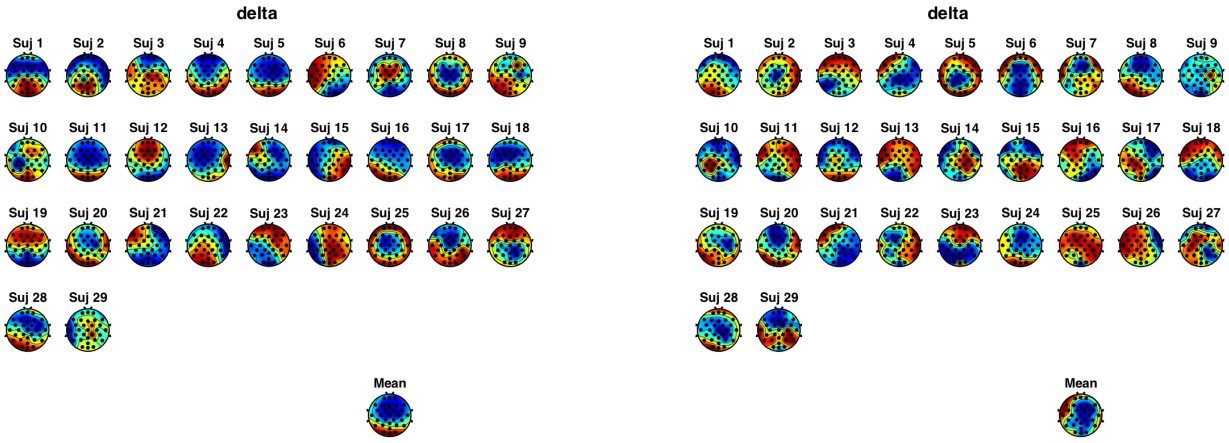


Figura 5.55: Activaciones cerebrales cSPoC vídeo Paranormal con PCA (izda.) y sin PCA (dcha.) en banda delta

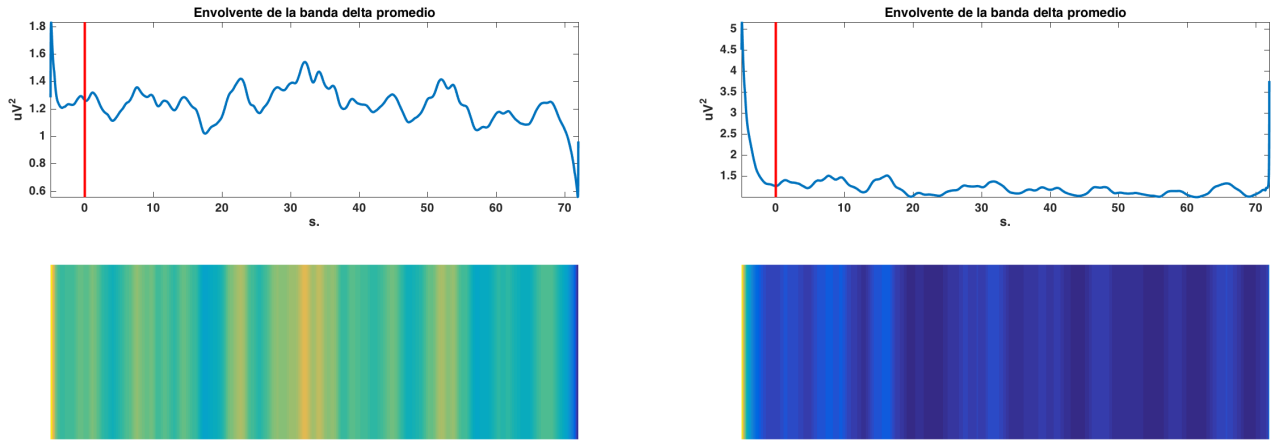


Figura 5.56: Envolvente banda delta vídeo Paranormal con PCA (izda.) y sin PCA (dcha.)

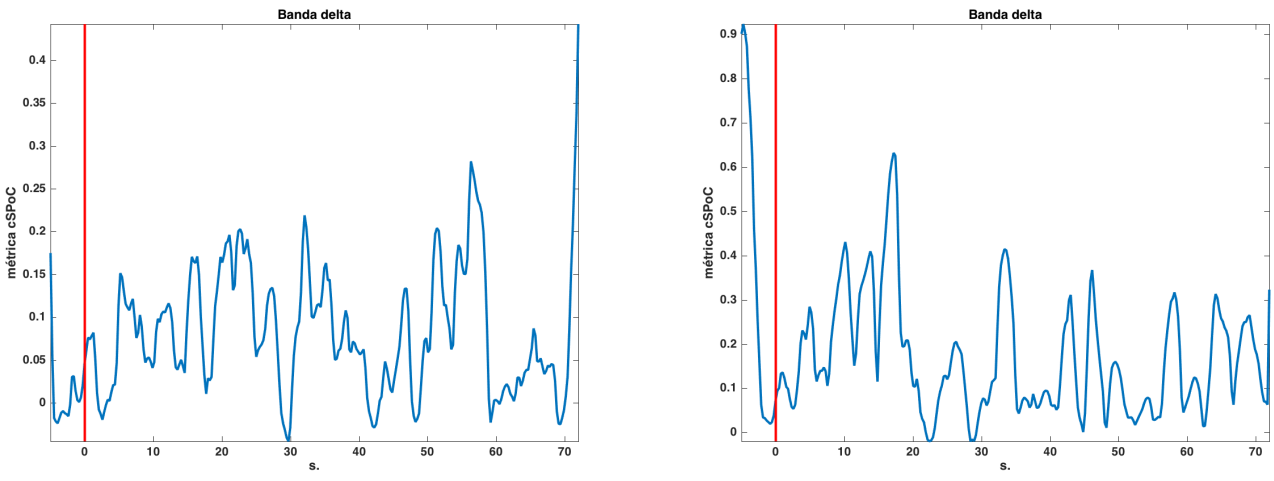


Figura 5.57: Resultados cSPoC vídeo Paranormal con PCA (izda.) y sin PCA (dcha.) en banda delta

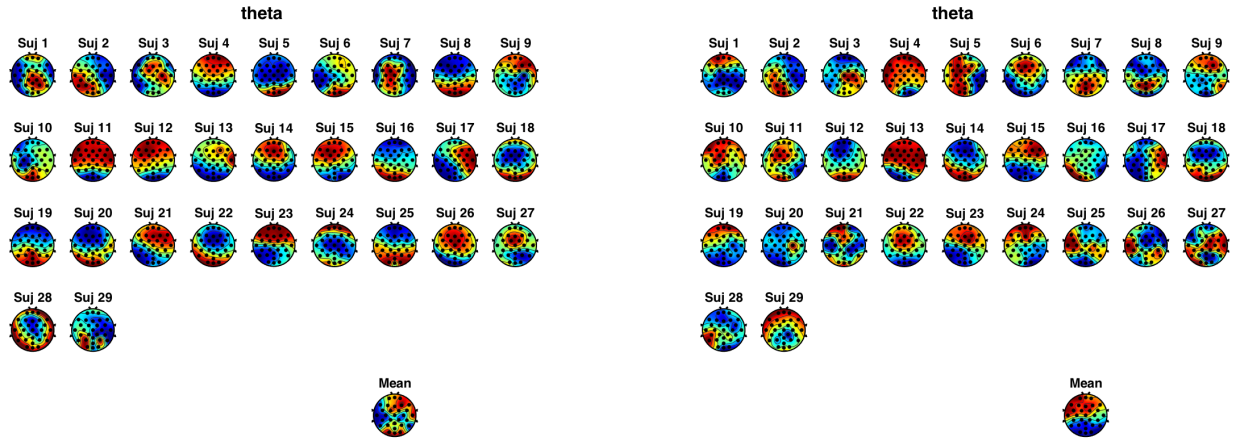


Figura 5.58: Activaciones cerebrales cSPoC vídeo Paranormal con PCA (izda.) y sin PCA (dcha.) en banda theta

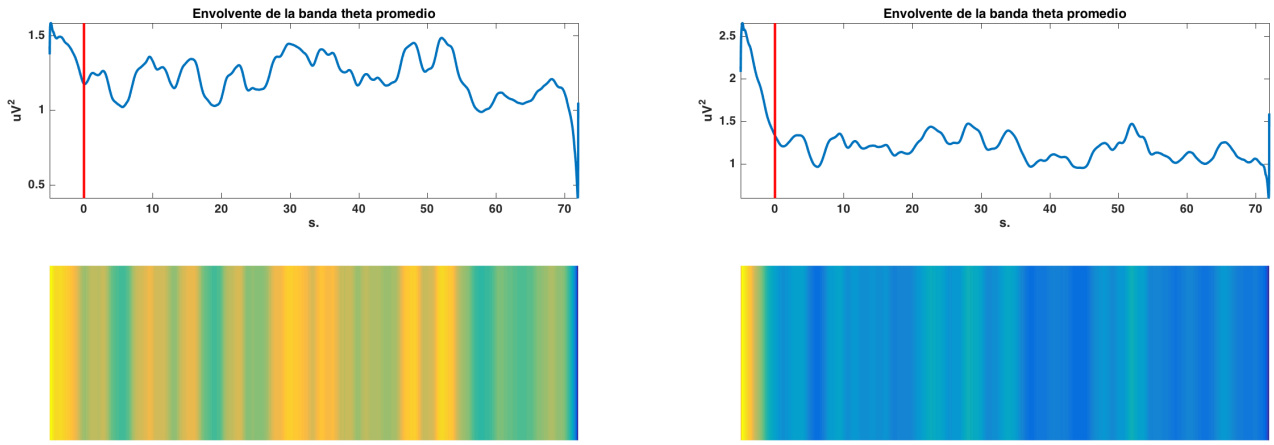


Figura 5.59: Envolvente banda theta vídeo Paranormal con PCA (izda.) y sin PCA (dcha.)

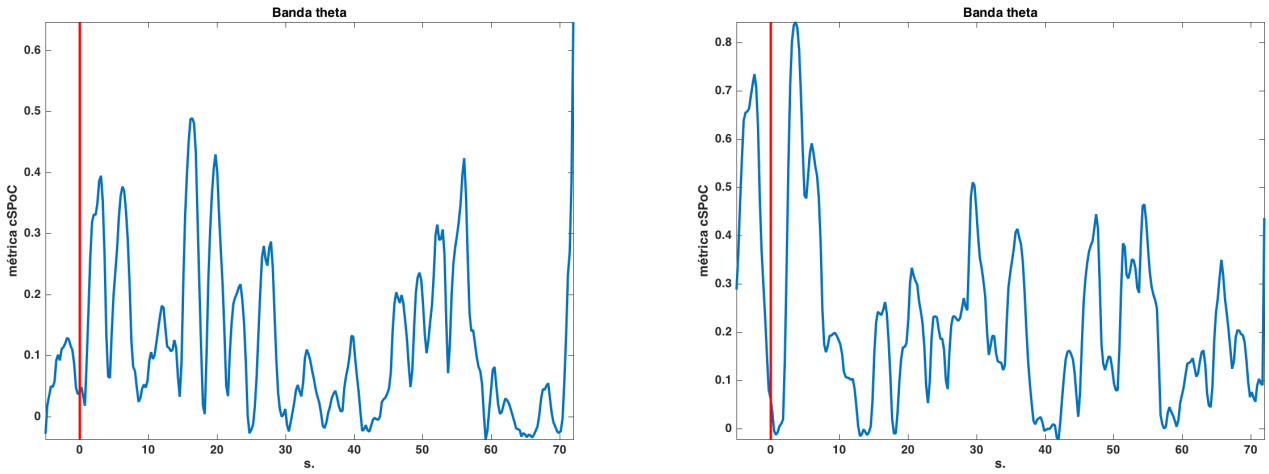


Figura 5.60: Resultados cSPoC vídeo Paranormal con PCA (izda.) y sin PCA (dcha.) en banda theta

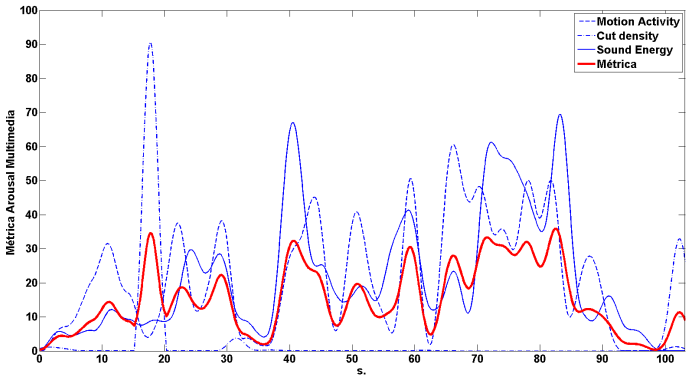


Figura 5.61: Resultado análisis multimedia REC

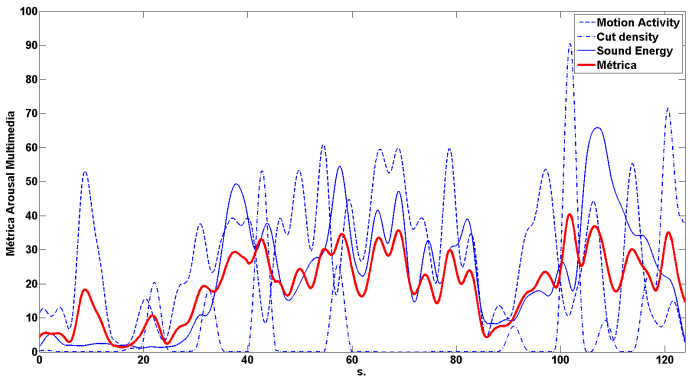


Figura 5.62: Resultado análisis multimedia CocaCola

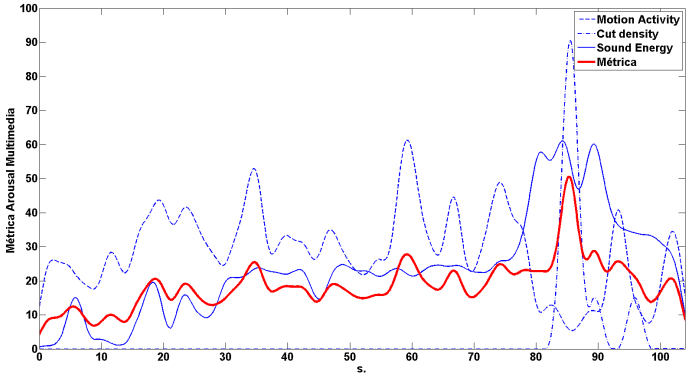


Figura 5.63: Resultado análisis multimedia Esquí

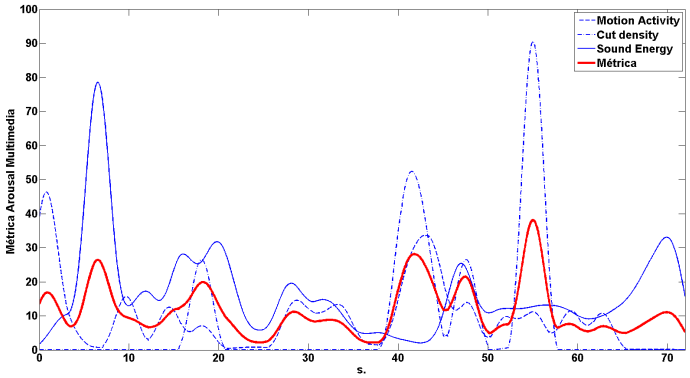


Figura 5.64: Resultado análisis multimedia Prinzenzeit

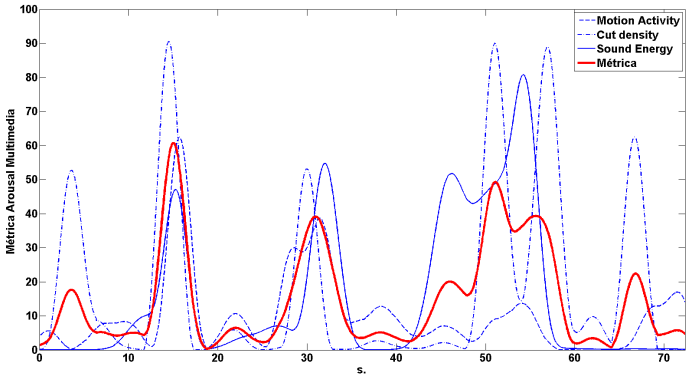


Figura 5.65: Resultado análisis multimedia Paranormal

6. Conclusiones y líneas futuras

En este proyecto fin de máster se han propuesto distintos métodos para desarrollar un sistema basado en la decodificación emocional a través del procesado de la señal encefalográfica (EEG) y la búsqueda de patrones temporales y frecuenciales comunes a lo largo del visionado de contenido multimedia. No existe un estado del arte avanzado en esta materia y el estudio realizado era exploratorio para conocer la viabilidad de los métodos por lo que los resultados presentados en este trabajo son esperanzadores y animan a continuar el desarrollo del sistema.

El objetivo de crear un sistema completamente automatizado se basaba en la consecución de dos objetivos parciales que eran el desarrollo de un filtrado automático de la señal de EEG y el uso de herramientas de relación entre señales multivariantes para encontrar patrones que nos den una idea de la evolución temporal de variables emocionales estimuladas por vídeos. En el capítulo 3 se expone un sistema formado por una serie de filtrados y procesados en serie que permiten la obtención de una señal limpia y libre de artefactos de una forma totalmente automática lo que permitiría la automatización completa del sistema. Hay que mencionar que a día de hoy todavía el preprocesado de EEG es una tarea bastante manual que involucra el conocimiento de numerosas técnicas. Con esta propuesta, el usuario final no tendría que preocuparse de ello. Por otra parte hemos presentado la opción de emplear diversas técnicas de procesado y filtros lineales que permiten maximizar las relaciones entre las señales fisiológicas (en este caso EEG) entre las respuestas de diferentes individuos. Esta es una aproximación novedosa al problema de la detección de emociones ya que parte de la premisa de que existe un sustrato neuronal común entre todos los individuos y éste podría ser decodificado usando técnicas no invasivas como puede ser el EEG. Este estudio tiene un carácter exploratorio ya que no existe apenas literatura sobre la materia por lo que los resultados de este estudio permiten abrir nuevas posibilidades para la creación de este tipo de sistemas.

Teniendo en cuenta todo lo anterior, los resultados comentados en el capítulo previo son esperanzadores y demuestran las oportunidades que este tipo de sistemas ofrecen para la posible detección de momentos clave en vídeos. Los sistemas basados en señales fisiológicas son una tecnología muy reciente, casi exclusivamente en fase de investigación en la actualidad pero con un futuro muy prometedor debido a sus amplísimas posibilidades. Por todo ello se abre un abanico de acciones futuras a realizar como las que se indican a continuación:

6. Conclusiones y líneas futuras

- Ampliación del número de vídeos para la comprobación de la efectividad del sistema desarrollado.
- Realización de un estudio con un protocolo de evaluación más controlado y medible para tener datos numéricos y no sólo subjetivos de la adaptación de sistema y emociones. Existe un problema común en la literatura sobre detección de emociones: es complicado definir "la verdad" para poder establecer métodos cuantitativos de evaluación. Podemos basar "la verdad" en encuestas subjetivas a una población muy grande sobre los momentos clave y de este modo poder incluso probar correlaciones entre los momentos detectados y la cantidad de gente que los selecciona como importantes.
- Aparte de la dimensión del arousal podemos pensar en buscar correlaciones con otras variables emocionales como puede ser la valencia.
- Posibilidad de combinar los sistemas expuestos en esta memoria para detectar con mayor precisión aquellos momentos más relevantes.
- La realización del mismo problema pero online, es decir, un sistema que decodifique en tiempo real lo que le va provocando el vídeo de manera que se pueda pensar, por ejemplo, en películas con líneas argumentales adaptativas.

Bibliografía

[1] Dan Ariely and Gregory S Berns. Neuromarketing: the hope and hype of neuroimaging in business. *Nat Rev Neurosci*, 11(4):284–292, apr 2010.

[2] Uri Hasson, Orit Furman, Dav Clark, Yadin Dudai, and Lila Davachi. Enhanced Intersubject Correlations during Movie Viewing Correlate with Successful Episodic Encoding. *Neuron*, 57(3):452–462, 2008.

[3] L. Nummenmaa, E. Glerean, M. Viinikainen, I. P. Jaaskelainen, R. Hari, and M. Sams. Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences*, 109(24):9599–9604, 2012.

[4] Jacek P. Dmochowski, Paul Sajda, Joao Dias, and Lucas C. Parra. Correlated Components of Ongoing EEG Point to Emotionally Laden Attention A Possible Marker of Engagement? *Frontiers in Human Neuroscience*, 6(May):112, 2012.

[5] Leif Sörnmo and Pablo Laguna. Bioelectrical signal processing in cardiac and neurological applications. 2005.

[6] Bruce H. Friedman. Feelings and the body: The Jamesian perspective on autonomic specificity of emotion. *Biological Psychology*, 84(3):383–393, 2010.

[7] Sylvia D. Kreibig. Autonomic nervous system activity in emotion: A review. *Biological Psychology*, 84(3):394–421, 2010.

[8] Tim Dalgleish. The emotional brain. *Nature reviews. Neuroscience*, 5(7):583–589, 2004.

[9] Charles Darwin. *The expression of the emotions in man and animals / by Charles Darwin.* New York ;D. Appleton and Co.,, 1916. http://www.biodiversitylibrary.org/bibliography/4820 — Includes index.

[10] W. James. What is an Emotion? *Mind*, 9, 1884.

[11] C.G. Lange and W. James. *The Emotions*. Hafner Pub. Co., 1922.

[12] Walter Cannon. The James-Lange-theory of emotions: A critical examination and an alternative theory. *The American Journal of Psychology*, 39(1/4):106–124, 1927.

[13] Paul Ekman, Robert W Levenson, and Wallace V Friesen. Autonomic Nervous System Activity Distinguishes among Emotions. *Science, New Series*, 221(4616):1208–1210, 2008.

[14] Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.

[15] J. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, (39):1161–1178, 1980.

[16] Linda Camras. *The American Journal of Psychology*, 94(2):370–372, 1981.

[17] T Mullen, C Kothe, and Ym Chi. Real-time estimation and 3D visualization of source dynamics and connectivity using wearable EEG. *Proceedings of the Fifth International Brain-Computer Interface Meeting*, pages 2–3, 2013.

[18] H. Shen, S. Jegelka, and A. Gretton. Fast kernel-based independent component analysis. *IEEE Transactions on Signal Processing*, 57(9):3498–3511, 2009.

[19] Aapo Hyvarinen. Fast and Robust Fixed-Point Algorithm for Independent Component Analysis. *IEEE Trans. Neur. Net.*, 10(3):626–634, 1999.

[20] Irene Winkler, Stefan Haufe, and Michael Tangermann. Automatic Classification of Artifactual ICA-Components for Artifact Removal in EEG Signals. *Behavioral and Brain Functions*, 7(1):30, 2011.

[21] Felix Biessmann, Michael Gaebler, Jan Peter Lamke, Ui Jong Ju, Stefan Hetzer, Christian Wallraven, and Klaus Robert Muller. Data-driven multisubject neuroimaging analyses for naturalistic stimuli. *Proceedings - 2014 International Workshop on Pattern Recognition in Neuroimaging, PRNI 2014*, (JUNE), 2014.

[22] Felix Biessmann. Machine Learning for Multimodal Neuroimaging, 2014.

[23] Wolfgang Klimesch. EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, 29(2-3):169–195, 1999.

[24] Vadim V. Nikulin, Guido Nolte, and Gabriel Curio. A novel method for reliable and fast extraction of neuronal EEG/MEG oscillations on the basis of spatio-spectral decomposition. *NeuroImage*, 55(4):1528–1535, 2011.

[25] Sven Dähne, Vadim V. Nikulin, David Ramírez, Peter J. Schreier, Klaus Robert Müller, and Stefan Haufe. Finding brain oscillations with power dependencies in neuroimaging data. *NeuroImage*, 96:334–348, 2014.

[26] Alan Hanjalic. Affective Video Content Representation. *IEEE Transactions on Multimedia*, 7(1):143–154, 2013.