



Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales
Departamento de Física

Dinámica cerebral durante el proceso de reconocimiento de imágenes con EEG

Tesis de Licenciatura en Ciencias Físicas

Carolina Vlatko

Dirección: Luz Bavassi
Co-Dirección: Rodrigo S. Fernández

Colaboradores: Juan Cruz Berón

Abril 2024

TEMA: Dinámica cerebral durante el proceso de reconocimiento de imágenes con EEG

ALUMNA: Carolina Vlatko

LU: 215/18

LUGAR DE TRABAJO: Laboratorio de Neurociencias de la Memoria Instituto de Cálculo, Instituto de Fisiología, Biología Molecular y Neurociencias, CONICET/UBA.

DIRECCIÓN: Luz Bavassi (IFIBYNE-UBA/CONICET) y Rodrigo S. Fernández (IFIBYNE-UBA/CONICET)

FECHA DE INICIACIÓN: Marzo 2023

FECHA DE FINALIZACIÓN: Abril 2024

FECHA DE EXAMEN: 15/04/2024

INFORME APROBADO POR:

Autor/a Carolina Vlatko	Jurado Dra. Ana Amador
Director/a Dra. Luz Bavassi	Jurado Dr. Enzo Tagliazucchi
Profesor/a de la Tesis de Licenciatura Dra. Silvina Ponce Dawson	Jurado Dr. Juan E. Kamienkowski

Resumen

En la neurociencia y la psicología cognitiva se investiga ampliamente el problema de la organización y representación de la información en el cerebro. En otros aspectos, se estudia particularmente el proceso de reconocimiento básico de objetos, ya que implica entender las relación entre los estímulos del mundo real y la codificación que realiza nuestro cerebro. En esta tesis se llevó a cabo un experimento de electroencefalografía con 7 participantes, con el cual se investigó la representación neural de objetos pertenecientes a diferentes categorías semánticas: caras, animales, frutas y verduras, ropa, lugares, instrumentos musicales, vehículos y elementos de cocina. Se implementaron dos condiciones experimentales: la de estímulos de visualización, donde se presentaron imágenes de distintas categorías, y la condición de elaboración, en donde a partir del nombre de los objetos se indicó realizar un proceso de imaginación. Para investigar la señal registrada se utilizaron métodos de análisis multivariado como RSA y algoritmos de aprendizaje supervisado con enfoque en la dinámica temporal de la actividad cerebral.

En los estímulos visuales se vieron los típicos potenciales relacionados a eventos (ERPs) N100 y N170, este último predominante en los estímulos de caras humanas. Con el análisis de RSA, se identificaron patrones temporales en la actividad cerebral, en algunos sujetos semejantes a los ERPs pero con una latencia temporal de unos pocos milisegundos. Las curvas de similaridad presentaron variabilidad entre sujetos, y para cada sujeto presentaron semejanza entre las categorías. Se destacó la estabilidad en la representación de caras humanas y lugares, así como su consistencia entre los participantes.

Se aplicaron clasificadores de aprendizaje automático para decodificar las señales cerebrales y diferenciar las categorías de caras y lugares. Los resultados mostraron una precisión significativa en la clasificación con picos de exactitud de casi el 90 % entre 160 ms y 190 ms post-estímulo. Se vio que los clasificadores utilizan principalmente una señal temprana para discriminar las categorías, mientras que la distinción entre representaciones dada por RSA aparece mayormente en tiempos posteriores a los 200 ms.

Además, se analizaron los coeficientes del clasificador SVM para identificar las áreas cerebrales más relevantes en la clasificación, observando cualitativamente una predominancia de la región occipital y parietal posterior derecha. Los clasificadores entrenados con la señal en la condición de visualización se utilizaron para clasificar los ensayos de la condición de elaboración. Los resultados indican que la decodificación de reconocimiento de objetos podría estar relacionada con procesos de imaginación o elaboración. A pesar de los resultados favorables, se sugiere ampliar el estudio con más participantes y más ensayos en la condición de elaboración.

Adicionalmente se plantea seguir explorando las características utilizadas para la clasificación. Por ejemplo, utilizando información sobre el espacio de frecuencias de las señal electroencefalográfica, la cual está relacionada con procesos de memoria.

Palabras Clave: Electroencefalografía, Análisis de representación de similaridad, Clasificador

Índice general

Chapter	Page
1. Introducción	1
1.1. Contexto histórico	1
1.2. Representación neural y análisis de similaridad	3
1.3. Algoritmos de clasificación con aprendizaje supervisado	5
A. Formalización del problema	5
A.1. Regresión Logística	6
A.2. Support Vector Machines	8
B. Reconocimiento de objetos	10
1.4. Objetivos	10
2. Metodología experimental	12
2.1. Participantes	12
2.2. Protocolo experimental	12
2.3. Estímulos	14
2.4. Equipo	14
2.5. Preprocesado	15
A. Etapa inicial	15
B. Segmentación en épocas	16
B.1. ALE (Algoritmo de Limpieza de Épocas)	16
B.2. Comparación con la librería Autoreject	18
2.6. Análisis de similaridad representacional	20
A. Matriz de similaridad por categoría semántica	21
B. Similaridad intra y entre categorías	22
B.1. T-test no paramétrico de permutación de clusters	23
2.7. Algoritmos de clasificación de aprendizaje supervisado	24
A. Preprocesado y validación	24
B. Matriz de confusión	26
2.8. Clasificación de la señal de elaboración	27
3. Resultados	29
3.1. Potenciales relacionados a eventos	29
3.2. Análisis de similaridad representacional	30
A. Matriz de similaridad por condición	31
B. Similaridad intra y entre categorías	31
C. Comparación entre las categorías caras y lugares	34
3.3. Algoritmos de clasificación de aprendizaje supervisado	35

A.	Comparación entre los clasificadores	35
A.1.	Exactitud de clasificación	35
A.2.	Tiempo de cómputo	36
B.	Clasificación en la condición de visualización	36
B.1.	Matriz de confusión	37
B.2.	Exploración del modelo	38
3.4.	Clasificación de la señal de elaboración	38
4.	Discusión y conclusiones	41
	Bibliography	46
5.	Apéndices	50

Agradecimientos

A mi directora Luz por darme la oportunidad de trabajar en algo nuevo para mí. Por acompañarme en el sinuoso camino de la investigación, guiándome a cada paso. Por confiar siempre en mí y darme el empujón que necesitaba para terminar mi licenciatura. Por demostrarme que la ciencia es un trabajo de las personas, para las personas y con las personas. En este sentido extendiendo los agradecimientos al grupo del Laboratorio de Neurociencias de la Memoria, por cada día compartido.

A mi mamá, por la cual debería existir el título de co-licenciada. Todos mis logros son también suyos, así como todas mis alegrías y desafortunadamente mis penas. Por acompañarme desde siempre y para siempre. A mi hermana Camila, por ser mi hermana, por querernos incondicionalmente y por seguir compartiendo nuestras vidas. A Marce, por ser mi familia, por el amor y apoyo incondicional construido. A mi familia que me apoyó siempre.

A mis amigas del ILSE, por tenerme en su vida tantos años y tantos más. Por ser prácticamente hermanas.

A mis amigos de la facultad, por todo lo que los quiero y por hacerme muy feliz en esta etapa. Por hacer cada cursada única, cada proceso colectivo. Por trascender infinitamente la facultad, y ser parte indispensable de mi vida. Por toda la diversión, el amor, el afecto.

A la FIFA por las fiestas, los viajes y las luchas. Por entender que venir a la facultad es mucho más que estudiar. Por promover la organización estudiantil y el sentido de comunidad.

A Christian y mis compañeros del LIAF, a quienes admiro infinitamente. Por adoptarme primero como pequeña estudiante de Laboratorio 6, y abrirme de nuevo las puertas cuando quise volver por más. Por ayudarme, enseñarme y por ser (quizás sin querer) mis modelos a seguir. Por ser tremendamente compañeros dentro y fuera del laboratorio.

A todos mis profesores, JTP's y ayudantes de la facultad por su dedicación. Por inspirarme a ser parte del plantel docente. Al Departamento de Física por el inmenso trabajo para que todo esto funcione.

A mis compañeros del Equipo de Apoyo Escolar y Acompañamiento Educativo de la UBA, y a los estudiantes que asisten al espacio cada sábado, por expandir en territorio el derecho a la educación. Por enseñarme sobre compromiso y militancia. Por ser parte de mi trayectoria.

A la universidad pública, en este caso a la Universidad de Buenos Aires, por darme la oportunidad de ser segunda generación universitaria en mi familia. Por darme una educación de excelencia, pública y gratuita, a mí y a muchos miles. Por lo valioso de un proyecto de país de derechos humanos. Por el orgullo que eso me genera.

Por todo lo nombrado y mucho más, agradezco a mi patria y a los habitantes del suelo argentino.

Capítulo 1

Introducción

Reconocer el mate en el escritorio, el mismo escritorio o a la persona que pasó por enfrente de la oficina es una tarea sencilla. Al sistema visual le resulta sumamente fácil el proceso de detectar y clasificar objetos entre miles de posibilidades, y le insume no más que una fracción de segundo. Este proceso cognitivo se conoce como “reconocimiento básico de objetos”, y se define como la discriminación extremadamente rápida y precisa de categorías de objetos, independientemente de cambios en posición, tamaño, pose y contexto de fondo [1]. Esta habilidad para asignar etiquetas (por ejemplo, sustantivos) a objetos particulares, comprende desde titulaciones precisas de “identificación” (mesa, tomate, Micaela) hasta asignaciones generales de “categorización” (mueble, fruta, persona) [1]. El reconocimiento de objetos es una actividad esencial en nuestra vida diaria, la realizamos permanentemente y nos permite interactuar con nuestro entorno de manera efectiva. En los últimos tiempos, la neurociencia cognitiva ha estado investigando las representaciones cerebrales-visuales, con el objetivo de relacionar estos patrones de actividad con nuestro comportamiento. Para llevar a cabo esta investigación, se han utilizado herramientas computacionales, como análisis multivariado y aprendizaje automático, que han demostrado ser de gran utilidad en este campo.

1.1. Contexto histórico

En la década de los 80' comenzaron a reportarse casos de pacientes con trastornos en el reconocimiento y comprensión de objetos dentro de categorías específicas. Estos pacientes, generalmente debido a daños cerebrales, perdían la capacidad para identificar y entender objetos pertenecientes a una categoría concreta, como por ejemplo los animales, mientras mantenían su conocimiento sobre otros objetos intacto [2–4]. Esto dio inicio a un nuevo campo de investigación en la psicología cognitiva, ya que desde entonces se registró una gran cantidad de casos de deterioro cognitivo desproporcionado sobre una categoría semántica comparada a otra.

En estas investigaciones se llevaron a cabo diferentes experimentos comportamentales con pacientes para determinar las características de sus deterioros. En la Figura 1.1 se muestran evaluaciones realizadas sobre un paciente con deterioro semántico dispar para la categoría “animales” en comparación con otras categorías inanimadas como herramientas, muebles, elementos de cocina. El paciente no podía (a) identificar y asignar el nombre correcto a los animales, (b) identificar y decidir si los animales presentados eran reales o inventados o (c) completar correctamente las partes de los animales, entre otras evaluaciones.

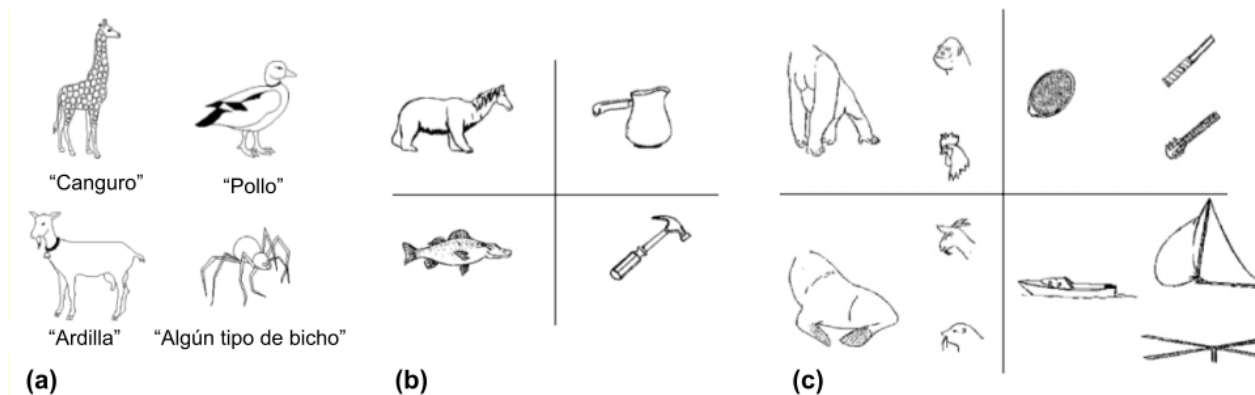


Figura 1.1: Algunos ejemplos de las evaluaciones que mostraron un deterioro semántico específico a la categoría de “animales” en un paciente. Se evaluó: **(a)** Identificación y nominación. **(b)** Decisión sobre la realidad del objeto. **(c)** Decisión de las partes del objeto o la cabeza del animal o animal [5,6].

Estos estudios indicaban que los seres vivos y los artefactos podrían procesarse de manera diferente en el cerebro [7, 8] y se desarrollaron varias teorías que intentaron explicar estos hechos en base a principios subyacentes sobre las representaciones de los conceptos [6]. Por ejemplo, se desarrolló la teoría “Sensorial/Funcional” que postulaba que el déficit de reconocimiento sobre cada categoría estaba asociado un déficit en la modalidad del conocimiento; modalidad sensorial (identificación y nombramiento de los objetos) o funcional (como incapacidad de decidir si una ballena nada o vuela, o cómo se usa una herramienta). Finalmente esta como otras hipótesis no se comprobaron [9], y la evidencia sugirió a los científicos inclinarse por teorías que asumen un sistema conceptual ampliamente distribuido que no está organizado por categorías [10]. Por ejemplo, la teoría de estructura conceptual propone que los conceptos de distintas categorías semánticas tienen estructuras diferentes. Por lo tanto, el daño al sistema afecta cada categoría de manera única, resultando en déficits selectivos. Añade que las características del entorno perceptual determinan patrones específicos de distinción y correlación en las representaciones internas, los cuales causan comportamientos deficitarios en categorías afectadas por daño [11]. Si bien hay varias teorías con coincidencias y diferencias en algunos postulados, el consenso general sostiene que las propiedades compartidas entre objetos se reflejan bajo una organización semántica, permitiendo generalizar conceptos que pertenecen a una categoría particular (como animales, herramientas o edificios) [8].

Un avance enorme en este área se debió a la implementación de técnicas de neuroimagen como fMRI (Resonancia Magnética Funcional) y EEG (Electroencefalograma) para estudiar la actividad cerebral. El fMRI proporciona imágenes detalladas del cerebro en tiempo real, mostrando áreas específicas que se activan durante diferentes tareas o estímulos. Su funcionamiento se basa en aplicar campos magnéticos para medir los cambios en el flujo sanguíneo relacionados con la actividad neural, y es un método que se caracteriza por su alta resolución espacial. Por su parte, el EEG registra la actividad eléctrica del cerebro mediante electrodos colocados en el cuero cabelludo. En este caso, el equipo se destaca por su alta resolución temporal típicamente en el rango de 100 Hz a 1000 Hz, y por su bajo costo económico en relación a fMRI. Los métodos son no invasivos y aptos para una gran variedad de sujetos de diferentes edades y condiciones. A partir de este instrumental y aplicando métodos de análisis multivariado se vieron diferencias en la actividad cerebral

relacionada al reconocimiento de objetos en sujetos sanos, entre la categoría de seres vivos y la de objetos fabricados por la humanidad. Adicionalmente, se diferenciaron categorías específicas de objetos, como caras, partes del cuerpo, animales, frutas/verduras, edificios, herramientas y muebles [8].

En los estudios de EEG, típicamente se estudian los ERPs (Event-related Potentials). Estos son picos de amplitud de la señal post-estímulo cuando se promedia entre todos los estímulos de una condición. En estudios con estímulos visuales, se vio que los rostros humanos pueden evocar un componente N170 ¹ más prominente y estable en el tiempo en comparación con objetos que no son rostros. Esta diferencia en la topografía espacial del N170 sugiere la existencia de un proceso neural específico para el reconocimiento facial [12, 13]. Paralelamente, mediante el análisis de las distribuciones de amplitudes, polaridades y topografías del ERP se puso el foco en analizar la evolución temporal de la selectividad de categorías. Estos estudios han demostrado consistentemente que la actividad negativa en el cuero cabelludo entre 120 y 200 ms está influenciada por la categoría del objeto presentado, con una especial relevancia en electrodos laterales como P7 y P8 [14, 15].

A principios de siglo, comenzaron a implementarse MVPA (Multivariate Pattern Analysis). A diferencia de los métodos univariados convencionales el análisis multivariado tiene en cuenta el patrón de la actividad espacial del cerebro completo. Gracias a esto se determinó que los objetos visuales están representados de manera ampliamente distribuida y superpuesta, lo que resulta en patrones espaciales distintos de activación cerebral en la corteza occipital y temporal ventral [8, 12]. De esta manera, se ha utilizado el fMRI y el EEG para identificar los correlatos neurales implicados en la clasificación de imágenes y palabras, destacando la importancia de esta técnica en el estudio de la percepción visual [16, 17].

En resumen, se puede decir que los resultados de clasificación con datos de neuroimágenes mediante MVPA confirmaron la existencia de patrones espaciales distribuidos y demostraron la posibilidad de discriminar objetos visuales utilizando patrones de activación cerebral. Un concepto ampliamente utilizado para describir estos fenómenos es el de “representación neural” que a su vez es estudiado con un análisis denominado *Representation Similarity Analysis* o RSA.

1.2. Representación neural y análisis de similaridad

En este contexto de la neurociencia cognitiva resulta clave el concepto de “representación neural” puesto que asume que la información está representada en patrones definidos de actividad neural [18]. La investigación sobre la representación neural abarca estudios sobre las características y propiedades de los patrones en la actividad neural que codifican y almacenan en el cerebro los estímulos externos e internos. Esto incluye la representación de información sensorial, como imágenes visuales o sonidos, así como información cognitiva y conceptual, como ideas y recuerdos. Por lo tanto, la representación neural es un concepto que vincula el mundo físico con la forma en que se experimenta mentalmente [18]. En la práctica, las representaciones neurales se miden como patrones de actividad neural, por ejemplo, durante la percepción de estímulos o la imaginación.

Si bien existen debates sobre las implicaciones filosóficas de este planteamiento, la neurociencia cognitiva se limita a entender la representación neural como una relación sistemática y de correlación entre las

¹La componente N170 se denota con N por ser un potencial negativo y 170 por los milisegundos posteriores al estímulo

características del mundo y la actividad neural observada. En este marco, estudiar la representación neural es de suma importancia para comprender cómo se representa y organiza la información en el cerebro y cómo la calidad de las representaciones neurales influye en la cognición.

En la práctica, se emplea el **análisis de similitud representacional** (RSA por sus siglas en inglés) para evaluar la **similitud** entre representaciones neurales utilizando técnicas de análisis multivariado. En RSA, los datos de actividad cerebral se comparan con medidas de similitud, para discernir los patrones de respuesta ante diferentes estímulos o condiciones experimentales. Este enfoque revela la estructura subyacente de la representación neural y contribuye al entendimiento sobre la organización y procesamiento de la información en el cerebro [19].

RSA tiene aplicaciones relevantes para entender cómo los patrones neurales se relacionan con diferentes procesos cognitivos, como el reconocimiento de objetos específico a las categorías semánticas. En este caso, se puede cuantificar la información específica a una categoría que está representada en los patrones de activación neural. Para ello se mide la diferencia entre la similitud dentro de una categoría comparada con la similitud entre categorías. Por ejemplo, esta aproximación se aplica como medida del envejecimiento neural, puesto que hay indicaciones que las representaciones neurales se vuelven más estables (o menos distintas) con la edad [20,21].

Un área clave en la implementación de RSA es el estudio de la memoria. La estabilidad temporal de la similitud de las representaciones y su grado de distintividad con respecto a otras representaciones están relacionados con el rendimiento de la memoria. Para evaluar la estabilidad de la representación a lo largo del tiempo se analizan los patrones de actividad neural inducidos por repeticiones de un estímulo. Estas medidas de similitud logran ser predictivas sobre la precisión y capacidad de evocación de recuerdos [22,23]. Mediante el análisis de RSA, se pueden explorar otros aspectos de la memoria, como la reinstalación o reactivación. En este caso, se evalúa la similitud entre los patrones de activación durante la codificación y la recuperación en tareas de reconocimiento o evocación, siendo crucial para el modelado de memorias episódicas. Además, se investigan los efectos de “memorias subsiguientes” analizando con RSA las diferencias en la actividad neural entre los ítems que fueron subsecuentemente recordados y los que no.

Cabe destacar que la similitud entre representaciones es, en principio, una noción subjetiva. En cada estudio se definen los atributos de la actividad cerebral que se van a comparar tales como los patrones espacio-temporales o del espacio de frecuencias, entre otros; y también se define la medida para cuantificar la distancia entre esos atributos. Una medida de similitud entre las representaciones neurales es la correlación entre los patrones de actividad en respuesta a distintos estímulos [24]. Por ejemplo, las representaciones de contenido similar, como los estímulos faciales, exhiben una correlación más alta entre sí.

Por otro lado, también es sumamente utilizado la idea de **disimilitud representacional**. En este sentido la inteligencia artificial es una herramienta que ha ganado relevancia en los últimos años. En esta línea se busca definir un modelo que explique la mejor manera de separar los datos observados, generalmente en términos de rendimiento predictivo, pero con el agregado de interpretar las estimaciones obtenidas para los parámetros del modelo. Concretamente, los parámetros del modelo pueden ser mapeados de nuevo al espacio original, encontrando la disposición espacial y los patrones de conectividad de las áreas cerebrales relevantes [8]. Para desarrollar estos modelos, se emplean algoritmos de clasificación y otras técnicas de aprendizaje supervisado. Estos modelos no solo son capaces de clasificar patrones neurales, sino que también pueden utilizarse para

el análisis de similitud representacional. La idea es que la clasificación realiza un proceso de decodificación, donde predice la condición experimental o el estímulo a partir de la actividad cerebral. Luego los pares de condiciones que tienen representaciones similares muestran una baja precisión de decodificación, y los pares de condiciones que tienen representaciones disímiles muestran una alta precisión de decodificación [25]. De esta forma la decodificación por clasificación de la señal de EEG puede interpretarse como medida de **disimilitud representacional**, indicando las relaciones de similitud entre los patrones de activación de sensores relacionados con condiciones experimentales.

1.3. Algoritmos de clasificación con aprendizaje supervisado

La mayoría de los algoritmos de aprendizaje automático pueden clasificarse como aprendizaje supervisado o aprendizaje no supervisado. El objetivo del aprendizaje supervisado es encontrar una función que mapea los valores de los atributos de los datos, hasta el valor de otro atributo, llamado atributo objetivo [26]. Es decir, se busca generalizar el comportamiento de los datos con un modelo de manera que cuando se tiene un dato nuevo, se pueda predecir su atributo objetivo.

Se denomina aprendizaje *supervisado* ya que en cada instancia de la búsqueda de la función que mejor ajusta a los datos, el algoritmo se guía con la respuesta del atributo objetivo. De esta forma el conjunto de datos actúa de supervisor retroalimentando al algoritmo con esta información en el proceso de aprendizaje. Para ello es necesario que los datos contengan esta información, lo que no es un problema trivial. De otra manera habría que utilizar los algoritmos de aprendizaje no supervisados en donde el objetivo ya no es tan específico y existen distintas métricas para encontrar patrones y regularidades en los datos [26]. De todas formas, en el aprendizaje supervisado la búsqueda del modelo que “mejor” ajuste a los datos depende del algoritmo utilizado y no es única; esto se llama “sesgo de aprendizaje” y parte del problema es elegir la mejor combinación entre los datos y el algoritmo elegido.

En este trabajo se utilizaron dos algoritmos de aprendizaje supervisado para la clasificar la señal electroencefalográfica en categorías semánticas, estos fueron Regresión Logística y Support Vector Machines.

A. Formalización del problema

En términos generales, el objetivo de una clasificación es tomar un vector de entrada \mathbf{x} , denominado *feature* o vector de características, y asignarlo a una de las K clases. Las clases C_k con $k = 1, 2, \dots, K$ generalmente son clases disjuntas, es decir, a cada entrada le corresponde una sola clase. El espacio de entrada queda entonces dividido en regiones cuyos límites se llaman superficies de decisión. En este trabajo se usaron modelos lineales para la clasificación, lo que significa que las superficies de decisión son funciones lineales del vector de entrada \mathbf{x} , es decir quedan definidas por hiperplanos de dimensión $D-1$ en el espacio de entrada D -dimensional. Si los datos pueden separarse exactamente en sus clases correspondientes mediante este tipo de superficies, se dice que los datos son linealmente separables. [27].

Entonces, en un modelo lineal las variables toman la forma

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0 \quad (1.1)$$

donde \mathbf{w} es el vector de pesos y la normal del hiperplano de decisión y la constante w_0 es su corrimiento del origen. Luego la superficie de decisión queda definida por los \mathbf{x} tales que $y(\mathbf{x}) = 0$.

Aunque los datos sean linealmente separables, puede haber ruido o datos atípicos, que valen la pena clasificar incorrectamente con el objetivo de no sobreajustar el modelo a este tipo de datos. Esta flexibilidad se conoce como “margen suave”. Para estos problemas, y para penalizar coeficientes que sobreajusten los datos, se introduce un parámetro de regularización cuyo valor se suele determinar por métodos de validación cruzada. A continuación se describen los dos algoritmos utilizados en este trabajo, que buscan por distintos métodos encontrar el hiperplano que separe correctamente la mayoría de los datos, particularmente en el caso de dos clases.

A.1. Regresión Logística

La Regresión Logística es un modelo probabilístico utilizado principalmente para problemas de clasificación binaria. En este enfoque se modelan las densidades de probabilidad condicional de clase $p(x|C_k)$, así como las probabilidades a priori de clase $p(C_k)$, y luego se utilizan para calcular las probabilidades posteriores $p(C_k|x)$ a través del teorema de Bayes [27]. En el problema de dos clases, la probabilidad posterior para la clase C_1 puede escribirse como

$$p(C_1|x) = \frac{p(x|C_1)p(C_1)}{p(x|C_1)p(C_1) + p(x|C_2)p(C_2)} = \frac{1}{1 + \exp(-a)} = \sigma(a) \quad (1.2)$$

donde se definió

$$a \equiv \ln \frac{p(x|C_1)p(C_1)}{p(x|C_2)p(C_2)}. \quad (1.3)$$

De esta forma se llega a la función logística

$$\sigma(a) = \frac{1}{1 + \exp(-a)}, \quad (1.4)$$

conocida como sigmoidea que es fundamental en este modelo. La función $\sigma(\cdot)$ toma cualquier valor real y lo transforma en un valor entre 0 y 1, que es interpretado como una probabilidad. Su inversa es la función logit y en el contexto de Regresión Logística se la modela como una función lineal del vector de características \mathbf{x} de la forma

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \mathbf{w}^T \mathbf{x} + w_0, \quad (1.5)$$

donde \mathbf{w} es el vector de pesos y w_0 es el término de sesgo. El objetivo es encontrar los parámetros \mathbf{w} que maximizan la verosimilitud del modelo para los datos de entrenamiento. La función de verosimilitud representa la probabilidad de obtener un conjunto de resultados dadas las características de entrada con ciertos parámetros del modelo. Para un conjunto de datos etiquetados $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\}$, donde $y_i \in \{0, 1\}$, la verosimilitud se define como

$$L(\mathbf{w}, w_0) = \prod_{i:y_i=1} \sigma(\mathbf{w}^T \mathbf{x}_i + w_0) \prod_{i:y_i=0} (1 - \sigma(\mathbf{w}^T \mathbf{x}_i + w_0)). \quad (1.6)$$

Para implementar la clasificación por Regresión Logística en este trabajo se utilizó la función `LogisticRegression` de la librería `sklearn`. Esta implementación se basa en la Ecuación 1.6 para el caso de clases binarias con término de regularización. Se toma el logaritmo para obtener una sumatoria de términos, y la función de costo que se minimiza es

$$\min_{\mathbf{w}} C \sum_{i=1}^N \mathbf{s}_i (-y_i \log(\sigma(\mathbf{w}^T \mathbf{x}_i + w_0)) - (1 - y_i) \log(1 - \sigma(\mathbf{w}^T \mathbf{x}_i + w_0))) + r(\mathbf{w}) \quad (1.7)$$

donde \mathbf{s}_i corresponde a los pesos asignados por el usuario a una muestra de entrenamiento específica, en este trabajo no se asignaron pesos puesto que se usaron datos balanceados entre clases. La función $r(w)$ depende del tipo de regularización que se quiera aplicar, en este caso se usó la norma ℓ_2 de \mathbf{w} . Por último, el grado de regularización se determina por el valor de la constante C , que como se mencionó suele elegirse por métodos de validación cruzada, optimizando la capacidad predictiva del clasificador.

Para ejemplificar el uso de este modelo, se puede plantear una pregunta como “¿Cuántas horas de estudio se necesitan para aprobar un examen?”. En este caso, se tiene un conjunto de datos con la cantidad de horas de estudio de las personas y el resultado de su examen (Aprobado/No aprobado).

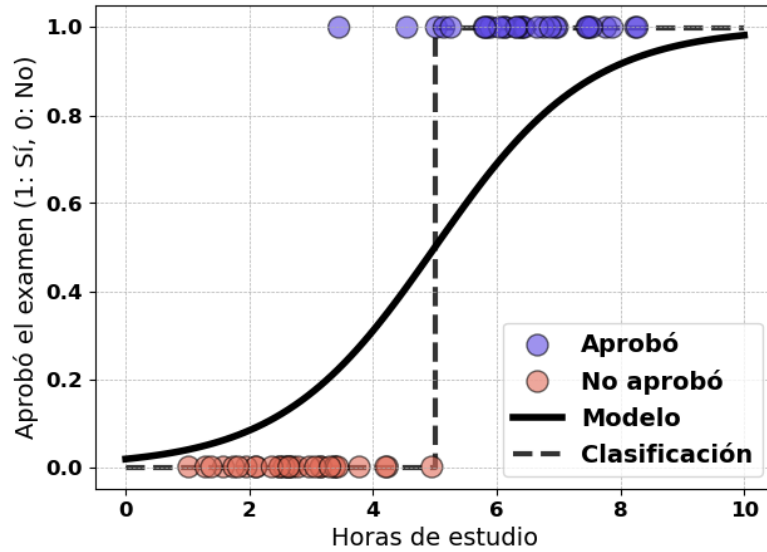


Figura 1.2: Clasificación con Regresión Logística en donde se predice el resultado de un examen (Aprobado/Desaprobado) según las horas de estudio empleadas. Durante el entrenamiento se busca la función sigmoidea que maximiza verosimilitud de los datos de entrenamiento (Ecuación 1.6).

El modelo entrenado con estos datos predice la probabilidad de aprobar un examen dadas las horas de estudio mediante la función sigmoidea como se muestra en la Figura 1.2. Usualmente se determina el umbral

en 50 %, y si la probabilidad es mayor o menor se predice que un examen fue aprobado o desaprobado respectivamente, como se muestra en línea punteada en la figura.

A.2. Support Vector Machines

Al igual que la regresión logística, este algoritmo de aprendizaje supervisado también busca el hiperplano óptimo que separa las clases en un espacio de alta dimensión. Sin embargo, SVM aborda este problema a través del concepto de margen. El margen se define como la distancia perpendicular mínima entre el hiperplano de decisión y cualquiera de las muestras. Así, SVM se centra en hallar el hiperplano que maximiza dicho margen entre las dos clases, y los puntos más cercanos a este hiperplano de cada clase reciben el nombre de vectores de soporte.

Para formular este problema, se considera la distancia perpendicular a un punto \mathbf{x} está dada por $|y(\mathbf{x})|/\|\mathbf{w}\|$. Por otro lado, se pide que los datos sean clasificados correctamente como

$$t_i \cdot y(\mathbf{x}_i) \geq 0 \quad \forall i \quad (1.8)$$

donde $t_i = \{-1, 1\}$ son los valores que puede tomar la etiqueta. De manera que se buscan los parámetros \mathbf{w} y w_0 que maximizan el margen de la forma

$$\max_{\mathbf{w}, w_0} \left\{ \frac{1}{\|\mathbf{w}\|} \min_i [t_i (\mathbf{w}^T \mathbf{x}_i + w_0)] \right\}. \quad (1.9)$$

Para simplificar el problema, se pueden reescalar \mathbf{w} y w_0 , y establecer la relación

$$t_n (\mathbf{w}^T \mathbf{x}_i + w_0) = 1 \quad (1.10)$$

para el punto más cercano a la superficie de decisión. Es decir, el margen queda a distancia 1 del hiperplano de decisión, y se sigue exigiendo la condición de la Ecuación 1.8 para todos los puntos. Con esta simplificación solo queda resolver

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad (1.11)$$

bajo esa misma restricción (Ecuación 1.8).

Como se mencionó anteriormente, los problemas generalmente no son siempre perfectamente separables mediante un hiperplano. Por ello, se quisiera permitir que algunas muestras estén del lado incorrecto, lo que se denomina una formulación de “margen suave”. En el caso de SVM lineal este problema se puede resolver implementando la función de pérdida llamada “hinge” o bisagra

$$\ell(\mathbf{x}_i) \max(0, 1 - t_i (\mathbf{w}^T \mathbf{x}_i + w_0)). \quad (1.12)$$

Esta función devuelve diferentes valores según la distancia del punto el \mathbf{x}_i al hiperplano de decisión y a los márgenes [28]. Entre los distintos casos, la función de pérdida devuelve un valor mayor a 1 para los datos clasificados incorrectamente. Teniendo esto en cuenta la función que se busca optimizar es

$$\min_{\mathbf{w}, w_0} \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \max(0, 1 - t_i(\mathbf{w}^T \mathbf{x}_i + w_0)) \right\}, \quad (1.13)$$

y es la expresión que usa la función `LinearSVC` de la librería `sklearn`. Intuitivamente, se trata de maximizar el margen (minimizando $\|\mathbf{w}\|$), mientras se incurre en una penalización cuando una muestra es clasificada incorrectamente o está dentro del límite del margen. El término de penalización C controla la fuerza de esta penalización, y como resultado, actúa como un parámetro de regularización inverso.

Para visualizar el concepto del margen, se puede ver el ejemplo en donde se clasifica entre dos especies de flores en base a sus características, largo y ancho de sus pétalos. En este caso el vector de características tiene dos dimensiones, por lo que el hiperplano es una recta que separa al plano en las dos clases. En la Figura 1.3 se muestran los datos de entrenamiento donde cada color representa una especie. La línea sólida es la recta de decisión y en línea punteada se tienen los márgenes, por último marcados con círculo negro se tienen los vectores de soporte. De esta manera, un nuevo dato consiste en un vector de características con el ancho y largo de sus pétalos, luego se predice la especie de la flor según la ubicación de este dato respecto de la recta de decisión.

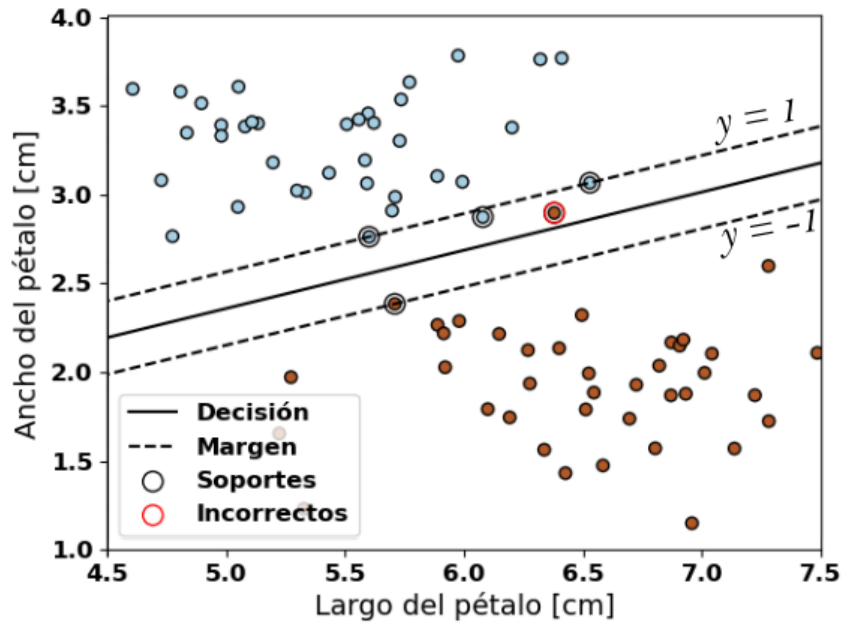


Figura 1.3: Clasificación con Support Vector Machines lineal. Se clasifica la especie de la flor (rojo/azul) según el largo y ancho de los pétalos. Se ve que es un clasificador de margen suave, puesto que se permitieron datos por dentro del margen o mal clasificados durante el entrenamiento.

En cuanto al entrenamiento, se puede ver que el clasificador es de “margen suave” puesto que se permitieron clasificaciones incorrectas y datos por dentro del margen. Según la Ecuación 1.12, los vectores de soporte y los datos por fuera del margen (bien clasificados) tienen valor de pérdida 0. Por otro lado, los datos mal clasificados aportan a la función de pérdida con su distancia a la superficie de decisión ($\ell > 1$). Los datos bien clasificados pero dentro del margen, también se penalizan con la distancia a la superficie de decisión que en este caso es $\ell < 1$.

B. Reconocimiento de objetos

Existen varios trabajos en los que se han utilizado clasificadores para identificar individualmente el reconocimiento de objetos específicos a categorías en participantes sanos usando EEG, MEG y fMRI [8, 12, 15, 25, 29–34]. En general, en estos estudios identificaban categorías semánticas a conjuntos de objetos agrupado por un significado más abstracto, con diferentes grados de jerarquización [15]. Por ejemplo, una categoría podía ser “ropa” (remeras, camisas, pantalones, etc.) y no sólo estímulos de diferentes “pantalones”. Las categorías más utilizadas fueron caras humanas, animales, objetos de la naturaleza como frutas y verduras, y objetos fabricados por el ser humano (por ej. herramientas o inmobiliaria). La duración de la presentación de estímulos en estos trabajos fue desde 200 ms a 500 ms. Por otra parte, existe una gran variabilidad en la cantidad de imágenes presentadas que se necesitan para construir el clasificador. Por ejemplo, en [30] registran 128 ensayos por participante, mientras que en [15] la cantidad es de 5184 separadas en dos sesiones experimentales. El efecto esta variación se estudió en [32], donde se reportó el desempeño de un clasificador de aprendizaje supervisado, en función de la cantidad de datos de entrenamiento y a la repetición de los mismos, cuando se discrimina entre caras y objetos inanimados. Allí se reportó un límite superior de 90 % de exactitud utilizando 120 imágenes repetidas 10 veces cada una. Sólo con el 16 % de esos ensayos (40 imágenes y 5 repeticiones) obtuvieron una exactitud de clasificación del 80 %. En la mayoría de los trabajos se usaron clasificadores lineales obteniendo buenos resultados [35], además permiten una interpretación de los modelos más directa que aquellos con transformaciones no-lineales.

1.4. Objetivos

En este proyecto se propuso realizar un experimento que permita explorar y caracterizar la dinámica de las representaciones neurales involucradas en el reconocimiento de objetos de diferentes categorías semánticas.

Para ello se plantearon los siguientes objetivos específicos:

- Diseñar y desarrollar un experimento de EEG de presentación de imágenes pertenecientes a 8 categorías semánticas (caras humanas, animales, lugares, frutas y verduras, instrumentos musicales, elementos de cocina, ropa, vehículos). A partir de la bibliografía sobre experimentos previos, determinar el número de imágenes pertenecientes a cada categoría, el tiempo de exposición de cada imagen y el número de repeticiones necesarias.
- Estudiar la estabilidad de las representaciones neurales para cada categoría. Generar una escala que ordene las categorías de imágenes según nivel de similaridad de los patrones de actividad cerebral utilizando RSA. Hallar el momento en el que la actividad cerebral presenta la máxima semejanza para cada categoría. Elegir dos categorías para usar en un experimento subsecuente del Laboratorio de Neurociencias. Para ello, identificar las dos categorías que presenten las similitudes más estables y consistentes entre participantes, y estudiar la disimilitud representacional entre ambas con RSA. Esperamos que la visualización de imágenes de caras presente mayor similaridad entre sus representaciones cerebrales que las otras categorías semánticas.

- Desarrollar un clasificador lineal de patrones espacio-temporales que permita distinguir entre dos categorías elegidas. Analizar la calidad de diferentes clasificadores lineales (Regresión Logística y Support Vector Machines) en base a su exactitud y tiempo de cómputo. Explorar el modelo dado por el clasificador identificando los parámetros más relevantes.
- Estudiar la semejanza que existe en la actividad neural involucrada en el proceso de visualización de un objeto de una categoría con el proceso de elaboración de dicho objeto para el uso de las herramientas de análisis en posteriores experimentos de memoria. Utilizar el modelo de clasificación obtenido para el reconocimiento de objetos (visualización) para identificar el proceso de imaginar objetos de una dada categoría (elaboración).

Capítulo 2

Metodología experimental

En este capítulo se detalla el sistema experimental en tanto sus participantes (Sección 2.1), diseño y protocolo (Sección 2.2), estímulos (Sección 2.3) y equipos utilizados (Sección 2.4).

Por otra parte, se describen los métodos utilizados para el preprocesado y limpieza de la señal, desde su registro hasta su forma final utilizada en los análisis subsiguientes (Sección 2.5).

A continuación del preprocesado y la limpieza, comienzan a detallarse los métodos de análisis utilizados. Con el objetivo de comparar y distinguir entre condiciones, se implementaron principalmente análisis de Similaridad Representacional (Sección 2.6) y algoritmos de clasificación de aprendizaje supervisado (Sección 2.7).

2.1. Participantes

Para este experimento, originalmente se registraron los datos de 12 sujetos (5 mujeres, 7 hombres), con edades comprendidas entre 23 y 27 años ($M=24,7$, $STD=1,3$). Sin embargo por un error del reloj interno del equipo de registro los datos debieron ser descartados. Los datos utilizados en este trabajo fueron tomados posteriormente en una nueva computadora de registro, sobre 7 nuevos sujetos (2 mujeres, 5 varones) de edades comprendidas entre 24 y 27 años ($M= 24,7$, $STD= 1,0$).

El proyecto contó con una aprobación del Comité de Ética del Instituto Médico Alfredo Lanari (Apéndice E).

2.2. Protocolo experimental

El experimento consistió principalmente en la presentación de estímulos correspondientes a 8 (ocho) categorías semánticas diferentes:

- Caras humanas (femeninas y masculinas)
- Animales (no humanos)
- Frutas y verduras
- Lugares (interiores y exteriores)

- Instrumentos musicales
- Elementos de cocina
- Ropa y elementos de indumentaria
- Vehículos.

Los estímulos se dieron en dos condiciones experimentales diferentes: *visualización* y *elaboración*. La condición de visualización consistió en la presentación de imágenes de objetos correspondientes a cada categoría. Con el fin de mantener la atención del sujeto en la pantalla, se le asignó una tarea de atención similar a [32]. Se incluyeron aleatoriamente imágenes de diferentes banderas y la tarea del sujeto fue contar la cantidad de apariciones de la bandera Argentina.

El experimento se dividió en bloques para el descanso del participante. El contenido de cada bloque de visualización fue el siguiente: se colocaron 40 imágenes únicas de cada categoría, presentadas en orden pseudo-aleatorio de modo que no haya imágenes consecutivas de la misma categoría. Se intercalaron aleatoriamente las imágenes de la tarea de atención, en una cantidad también aleatoria de entre 5 y 14 veces. La duración de cada imagen en pantalla fue de 0,7 s. Para separar un estímulo del otro, entre cada imagen hubo un intervalo con la pantalla en blanco, que tuvo una duración aleatoria de entre 0,5 y 1,0 segundos. De este modo la duración de un bloque de visualización fue de aproximadamente 8 minutos. En una única sesión experimental se realizaron 5 de estos bloques. Al final de cada bloque el sujeto debía responder cuántas banderas de Argentina habían aparecido en el bloque (tarea de atención) utilizando el mouse. Se solicitaba al sujeto que no realizara movimientos con el cuerpo a lo largo de cada bloque experimental. Durante el experimento el sujeto se encontraba solo, sentado frente a la computadora en el cuarto del laboratorio.

Por otro lado, el bloque de elaboración consistió en pedirle al sujeto que se imaginara y visualizara mentalmente el objeto que representa una palabra. Se presentaron 6 (seis) palabras por cada categoría mencionada, con una duración de 1,7 segundos cada una. El orden de las palabras fue pseudo-aleatorio respetando la misma restricción que las imágenes, que no haya dos ensayos consecutivos de la misma categoría. De la misma manera se tuvo un intervalo de pantalla en blanco entre palabras cuya duración fue de 0,5 a 1,0 segundos aleatoriamente. En este caso, la duración de este bloque fue de aproximadamente 2 minutos. Para corroborar la atención del sujeto, se interrumpió el bloque a la mitad preguntando si había aparecido la palabra “BANDERA”, siendo esta posibilidad aleatoria entre participantes.

En resumen, el esquema del experimento se muestra en la Figura 2.1. Se realizaron tres bloques de visualización (presentación de imágenes), con los descansos correspondientes, se continuó con el bloque de elaboración (presentación de palabras) y se finalizó con otros dos bloques de visualización. La duración total del experimento en la computadora fue de aproximadamente 50 minutos por sujeto. Se puede observar que en principio se contó, por sujeto, con 200 ensayos de visualización por categoría (1600 en total) y 48 ensayos de elaboración.

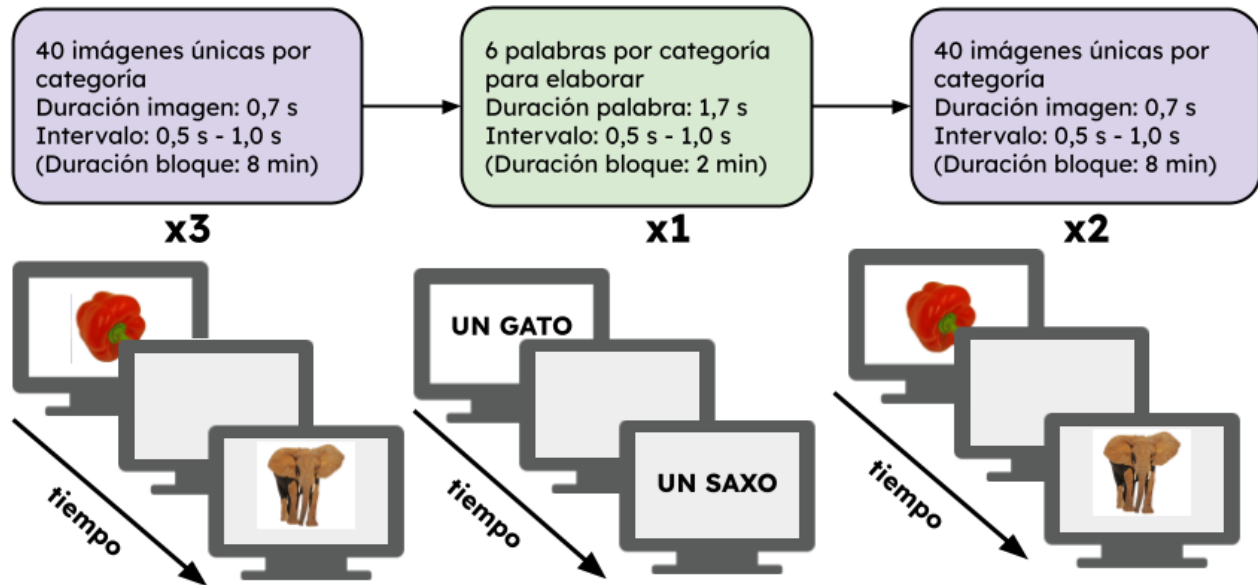


Figura 2.1: Esquema de bloques del protocolo experimental.

2.3. Estímulos

Las imágenes utilizadas fueron seleccionadas por el Licenciado Juan Cruz Berón de diferentes bases de datos. Las imágenes de los “lugares” se obtuvieron de la página *Places* [36] comúnmente utilizados para entrenar algoritmos de aprendizaje automático para clasificación de escenas. Se extrajeron 20 imágenes de escenas de interiores y 20 de exteriores. Las caras humanas fueron seleccionadas de la base de datos *FACES* [37], realizada especialmente para experimentos de neurociencia cognitiva. Se seleccionaron 40 imágenes de emotividad neutra, balanceadas entre hombres y mujeres y por rango etario (jóvenes, adultos, adultos mayores). El resto de las imágenes fueron obtenidas de BOSS (*Bank of standarized stimuli*) [38], también creado especialmente para estos experimentos, y fueron separadas por categoría manualmente por Juan Cruz Berón. En el APÉNDICE A se encuentra el listado de todas las imágenes utilizadas separadas por categoría.

El fondo de las imágenes, a excepción de los “lugares” fue removido para obtener el objeto en un fondo blanco sin contraste con el fondo blanco de la pantalla de estímulos. Todas las imágenes se ajustaron para equiparar el tamaño, y se corroboró que la distribución RGB por categoría fuese similar. Además, la resolución se limitó a 850×850 ya que de otro modo el desempeño de PsychoPy se veía perjudicado.

Las palabras utilizadas se basaron en las imágenes presentadas y se pueden consultar en el APÉNDICE B.

2.4. Equipo

Los registros neurofisiológicos se realizaron con un equipo de electroencefalografía Akonik (Bio-PC) de resolución temporal de 256 Hz, con 30 sensores en el montaje 10-20 (Figura 2.2) junto con la referencia de mastoides enlazada. El equipo fue conectado a una computadora de registro que contaba con el software EEG V12 de Akonik. El experimento se programó en el software PsychoPy (v2022.2.5) [39] y se llevó a cabo en una computadora de estímulos. Para la sincronización temporal estímulo-registro se utilizó un Arduino

NANO, enviando pulsos desde la computadora de estímulo hacia dos canales externos del equipo Akonik, obtentiéndose así marcas de on-set y off-set del estímulo.

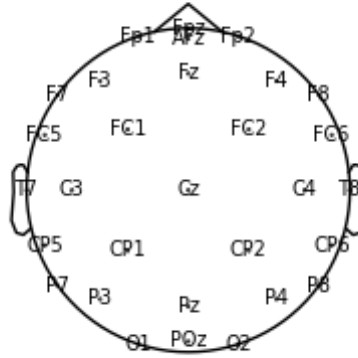


Figura 2.2: Topográfico con la posición de los 30 sensores utilizados en el experimento.

2.5. Preprocesado

A. Etapa inicial

La totalidad del preprocesado se ejecutó a través de la librería MNE-Python [40]. La señal de EEG registrada fue cargada en el objeto `Raw` de MNE, una estructura de datos que contiene la información completa de la misma: los valores de amplitud a lo largo del tiempo e información adicional como la frecuencia de muestreo y la ubicación espacial de los electrodos.

En primer lugar se aplicó un filtro Butterworth de 4to orden entre las frecuencias de 1 Hz y 40 Hz, con el método IIR (Infinite Impulse Response) y fase cero. A continuación se eliminaron los intervalos de descansos entre bloques experimentales. Se inspeccionó visualmente el espectro de frecuencias de los sensores y se interpolaron los sensores Fp2 (cuatro sujetos), F8 (un sujeto) y P7 (un sujeto).

Finalmente, se removieron los artefactos oculares. Estos surgen de los parpadeos y movimientos oculares del sujeto, ya que generan un patrón específico en la señal electroencefalográfica que no es del interés de este trabajo. Se aplicó una técnica denominada ICA (Independent Component Analysis) [41] que separa la señal en componentes que representan fuentes independientes cuya combinación lineal resulta en la señal registrada. Asumiendo que los artefactos oculares son una de estas fuentes independientes de la señal, se pueden identificar como componentes de ICA y reconstruir la señal después de remover estas componentes. Para este proceso se utilizó la clase `preprocessing.ICA` de MNE, la cual implementa una descomposición de componentes principales (PCA) que usa como input del algoritmo de ICA. La cantidad de componentes de PCA se determinó tal que se reúna el 99 % de la varianza explicada.

B. Segmentación en épocas

La señal continua preprocesada se dividió en segmentos temporales correspondientes a los ensayos, y fueron almacenados en la estructura de datos de MNE Epochs. Para la condición de visualización cada época se correspondió al segmento temporal de $-0,02$ s a $0,5$ s, y para la de elaboración la época fue de $-0,02$ s a $2,5$ s. El tiempo de 0 s corresponde al comienzo del estímulo (imagen o palabra según el caso), el cual fue registrado con la marca temporal del Arduino.

Una vez segmentada la señal continua en épocas, se realizó la limpieza de épocas descripta a continuación. Por último, para disminuir la dimensión de los datos y reducir los tiempos de cómputo en los análisis posteriores, se redujo la frecuencia de muestreo de todas las épocas a 128 Hz. De esta manera las épocas de la condición de visualización tuvieron una cantidad de 68 muestras temporales y las de elaboración de 324 muestras temporales (Tabla 2.1).

Condición	Tiempos [s]	Frecuencia de muestreo [Hz]	Nº de épocas
Visualización	$-0,02$ a $0,5$	128	$40 \times 8 \times 5$
Elaboración	$-0,02$ a $2,5$	128	6×8

Tabla 2.1: Información sobre datos experimentales segmentados en épocas.

De esta manera la señal originalmente continua quedó segmentada en épocas relacionadas con los estímulos. Para cada participante la cantidad de épocas fue de 1600 en la condición de visualización (40 imágenes $\times 8$ categorías $\times 5$ repeticiones) y 48 épocas en la condición de elaboración (6 palabras $\times 8$ categorías).

Durante el experimento, se le indicó a cada participante no realizar movimientos con el cuerpo. Sin embargo, la señal electroencefalográfica es sensible a pequeños movimientos incluso faciales. Esto, y un débil contacto de los electrodos con el cuero cabelludo pueden introducir señales ruidosas de amplitud comparable o mayor a la de la actividad cerebral relacionada con el experimento. Por este motivo, en el preprocesado de la señal de EEG es necesario incluir una “limpieza” de épocas para corregir o mitigar estos ruidos y obtener una señal de calidad que permita un análisis preciso. Para eso se implementó ALE (Algoritmo de Limpieza de Épocas) el cual se describe a continuación.

B.1. ALE (Algoritmo de Limpieza de Épocas)

Para eliminar o corregir épocas que posean ruido o movimientos se realizó un algoritmo automático para descartar épocas e interpolar sensores. Este algoritmo se desarrolló con dos objetivos principales: **(1)** fomentar la capacidad de escalabilidad en el manejo de grandes volúmenes de datos, y **(2)** garantizar la replicabilidad de los resultados. En los trabajos de M/EEG, es una práctica común descartar épocas mediante inspección visual, así como también realizar la interpolación de sensores ruidosos intra-época para corregir la época y evitar la pérdida significativa de datos experimentales. Es inmediato ver que un criterio cualitativo como el de la inspección visual puede variar según quien lo ejecute, y por lo tanto llevar a variaciones no replicables en los datos. Paralelamente, en un trabajo con 1600 épocas por sujeto, el proceso sería sumamente ineficiente. Por esto cabe destacar que ALE, el método desarrollado y usado para los datos de este trabajo es automático y determinista cumpliendo con los dos objetivos mencionados.

Los objetivos operativos del algoritmo fueron primero identificar épocas consideradas “ruidosas” en comparación con el comportamiento general de la señal del sujeto; y luego decidir si la época podía utilizarse con una corrección por interpolación de los sensores ruidosos, o bien debía ser completamente descartada. La tensión entre estas dos opciones surge de evitar introducir señales ruidosas y sesgos por sobre-interpolación de sensores en el análisis, y simultáneamente intentar conservar la mayor cantidad posible de datos experimentales. El algoritmo se implementa por sujeto, y el procedimiento se puede observar esquemáticamente en la Figura 2.3. En primer lugar, para obtener una medida del comportamiento general de la señal, se obtienen todos los valores ptp (del inglés *peak to peak*) de todos los sensores y de todas las épocas, obteniéndose 30×1600 valores. El valor ptp se refiere a la diferencia de voltaje entre el mínimo voltaje y máximo voltaje de la época, en tiempos no necesariamente consecutivos. A partir de la distribución de los voltajes ptp, se consideraron valores atípicos a los superiores al percentil 95 %. Es decir, se tomó como umbral de “ruido” al valor ptp correspondiente al percentil 95 % de todos los ptp del sujeto. Luego, si en una época más del 10 % de los sensores (3) superaron el umbral ptp, se consideró una época predominantemente ruidosa y se descartó.

Por otro lado, si la época tuvo tres o menos sensores ruidosos (con ptp mayor al umbral) se buscó un segundo criterio para decidir sobre la interpolación de los mismos. Para decidir si interpolar o conservar la señal de estos sensores en una época, se estudió si el sensor presentó un comportamiento similar y coherente con el resto de sensores en la época, a pesar de haber superado el umbral ptp. La métrica que se utilizó para comparar el comportamiento temporal de los sensores en una época, fue la distancia euclídea calculada según la Ecuación 2.1. La distancia d_i correspondió a la distancia euclídea entre el sensor i y el valor medio de los otros sensores, es decir, excluyendo al sensor i . Entonces, dentro de cada época y para cada sensor i , se calculó la distancia d_i para evaluar el comportamiento de cada sensor en comparación con el conjunto restante.

$$d_i = \|\mathbf{x}_i - \bar{\mathbf{x}}_{-i}\|_2 \quad i = 1, 2, \dots, 30 \quad (2.1)$$

De modo que para considerar como comportamiento atípico de un sensor en la época, se tomó como umbral intra-época el percentil 95 % de las distancias calculadas. Entonces los sensores considerados ruidosos (por el umbral ptp) y de comportamiento atípico en la época (por el umbral de distancia euclídea) se interpolaron para obtener una señal corregida en estos casos particulares. En la Figura 2.4 se puede observar un ejemplo del resultado de este algoritmo. En rojo se muestran las épocas consideradas ruidosas porque más del 10 % de los canales superaron el umbral ptp, y por lo tanto fueron rechazadas. En azul, se muestra el caso en el que se consideraron ruidosos una minoría de sensores ($< 10\%$), por haber superado el umbral ptp y el de distancia euclídea, de modo que fueron interpolados y la época fue conservada. En verde, se ejemplifica el caso de sensores que superaron el umbral ptp pero permanecieron sin interpolarse por no haber superado el segundo umbral, sugiriendo que el comportamiento del sensor fue coherente con los otros en la época.

Como se mencionó, el algoritmo se diseñó para identificar ruido en la señal de un sujeto según su comportamiento general, por lo tanto se implementó por separado para cada sujeto. Pero además, se implementó por separado para las condiciones de visualización y elaboración, debido a que los ensayos tenían distinta duración y un tipo de actividad cognitiva diferente, por ende un comportamiento general diferente. Al

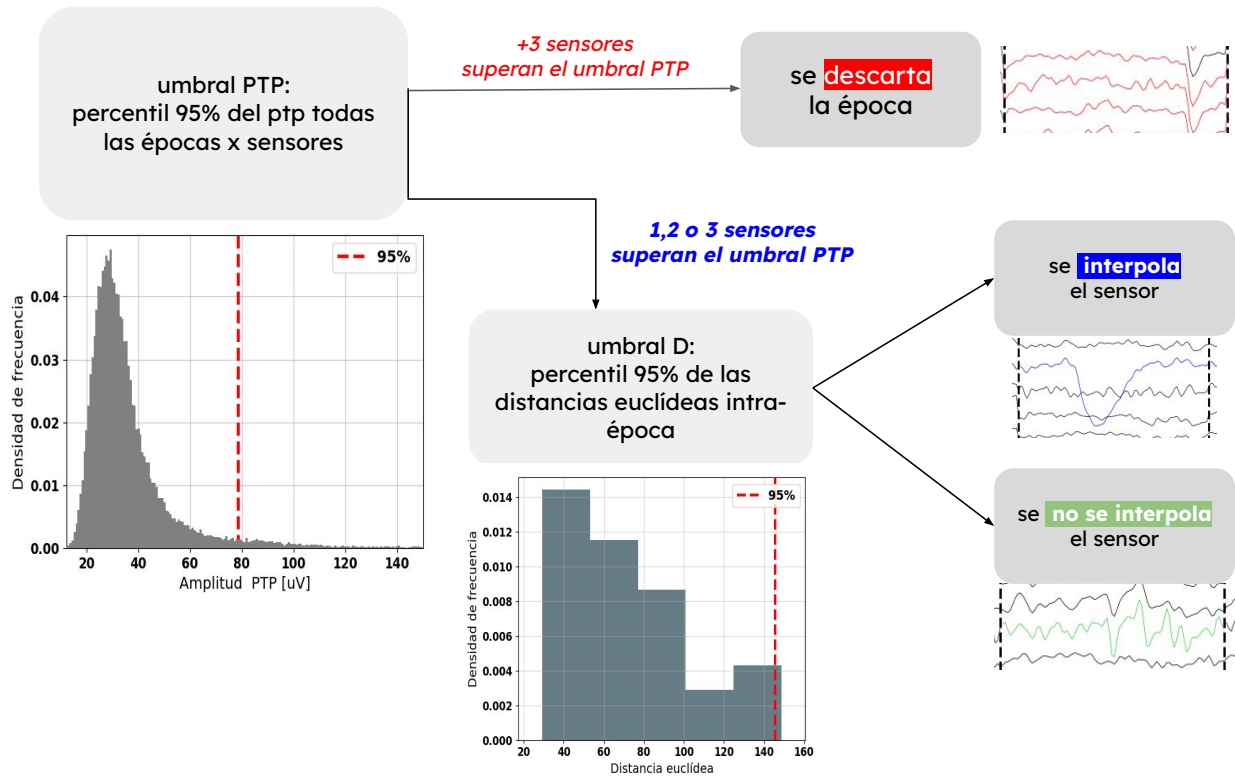


Figura 2.3: Esquema del algoritmo automático para descartar épocas e interpolar sensores. Primero se obtiene un umbral a partir del percentil 95 % de los voltajes pico a pico de todos los sensores y épocas. Una época se rechaza si más el 10 % de los canales superan dicho umbral. Sino, estos canales si interpolan si superan el umbral calculado a partir de las distancias euclídeas según la Ecuación 2.1.

implementar el ALE por separado, se computaron umbrales independientes para cada condición. De hecho, los umbrales ptp en la condición de elaboración (MEAN= 103 μ V, STD=36 μ V) fueron mayores que en la condición de visualización (MEAN= 55 μ V, STD=9 μ V) para todos los sujetos.

Se puede notar que el método tiene parámetros “libres”. Es decir, se puede implementar variando por ejemplo, el percentil de los umbrales o la métrica que mide la similitud del comportamiento de los sensores ya que se podría usar la correlación de Pearson o la distancia coseno, entre otras.

B.2. Comparación con la librería Autoreject

Se comparó el desempeño de ALE desarrollado en esta tesis con el algoritmo automático de limpieza de épocas de la librería Autoreject [42]. La librería fue desarrollada para rechazar y reparar sensores de MEG y EEG también de manera automática. El método también busca encontrar umbrales ptp pero con un enfoque diferente. Con ALE se obtuvo un umbral ptp por sujeto, teniendo en cuenta a todas los canales de todas las épocas, y el segundo umbral utilizado (Ecuación 2.1) fue un umbral intra-época. En el algoritmo

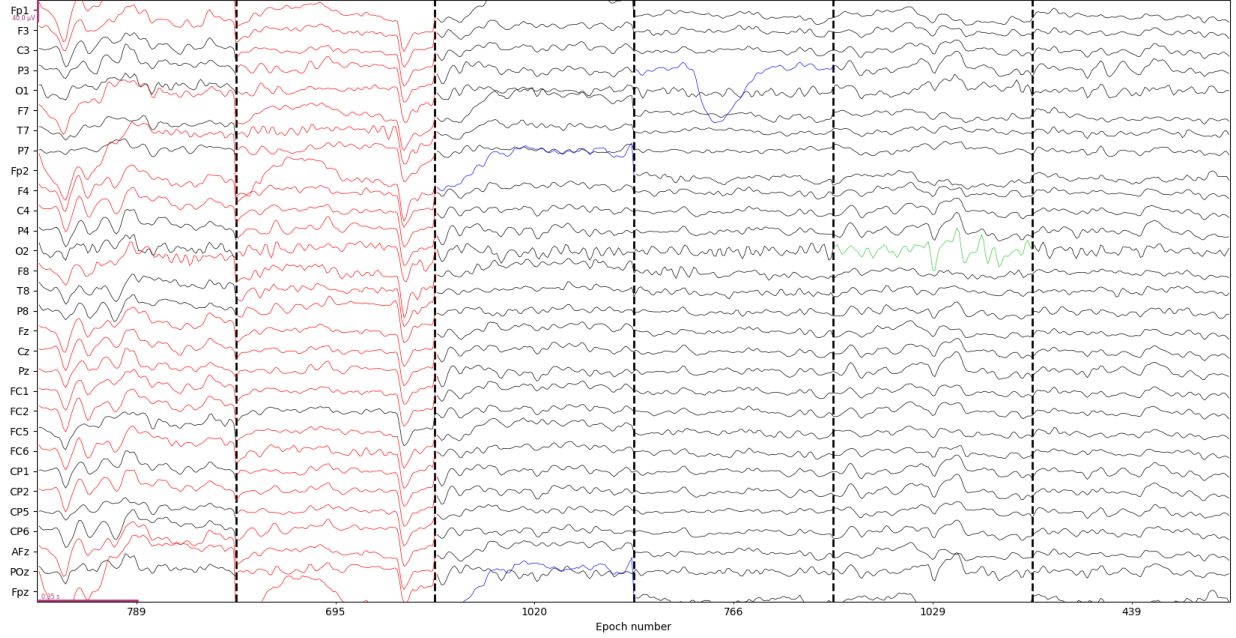


Figura 2.4: Ejemplo de algunas épocas limpiadas por el algoritmo. En rojo se encuentran los sensores rechazados, en azul los interpolados y en verde los que habían superado el umbral ptp pero quedaron intactos por no superar el umbral de la distancia euclídea.

de autoreject, se plantea que el umbral, además del sujeto, debe ser dependiente del sensor. Es decir, para cada sujeto se optimiza un umbral por sensor. Más aún, introducen dos parámetros. El parámetro κ indica la cantidad máxima de sensores que superan el umbral se pueden tolerar para no rechazar la época. En el caso de este trabajo, sería equivalente a la tolerancia de hasta 3 sensores “malos”(el 10%). El segundo, llamado ρ indica la cantidad de sensores máxima, ya catalogados como “malos-que pueden ser interpolados. Nótese que necesariamente $\rho < \kappa$. A diferencia del Autoreject, en este trabajo la decisión para interpolar los canales “malos”se tomó utilizando el umbral por época sobre la distancia euclídea (Ecuación 2.1). Otra diferencia fundamental entre ambos métodos, es que los umbrales en nuestro caso se tomaron en base al percentil sobre todos los datos, épocas y sujetos. En el Autoreject, los umbrales y parámetros se obtienen en base a una validación cruzada en *k-folds* que busca minimizar el error. Para cada umbral, se toma el valor medio de las épocas no-rechazadas por dicho umbral, y el error se define como la diferencia entre ese valor medio y la mediana de un nuevo set de datos (el nuevo set de datos correspondería al *fold* de la validación cruzada). Finalmente, para cada sujeto se obtienen los parámetros κ y ρ y además el umbral ptp depende de cada sensor. Para mayor detalle se puede consultar la bibliografía de los desarrolladores del método [42].

Para comparar ambos algoritmos se hizo la limpieza de épocas en la condición de visualización con cada uno, y se calculó el porcentaje de épocas -sobre las 1600- rechazadas coincidentemente por ambos métodos, el porcentaje de épocas rechazadas únicamente por el Autoreject, y el porcentaje rechazado solo por el ALE. Los resultados se muestran en la Figura 2.5.(a). Se vio que predominó la coincidencia en el rechazo de épocas en todos los sujetos. Por otra parte, en términos del método que rechazó más épocas, no hubo una tendencia consistente entre los sujetos. En promedio, los parámetros del algoritmo de autoreject obtenidos para los sujetos fueron $\kappa = 14$ y $\rho = 2$. Es decir, una época fue rechazada si tenía más 14 sensores ruidosos,

y mientras que en ALE fue con más de 3 sensores, aunque en cada caso según el umbral respectivo. Lo importante es que en ambos casos la cantidad máxima de sensores interpolados por época fue similar (2 en autoreject y 3 en ALE). Los umbrales ptp obtenidos en ambos métodos se pueden observar en la Figura 2.5 (b). La distribución corresponde a los umbrales ptp de los sensores de cada sujeto para el Autoreject, mientras que el punto de color representa el umbral ptp único por sujeto dado por ALE. Se observó entonces que los métodos fueron altamente consistentes en cuanto a los umbrales ptp obtenidos, a la cantidad de canales interpolados por época y en la coincidencia en las épocas que fueron finalmente rechazadas por los mismos.

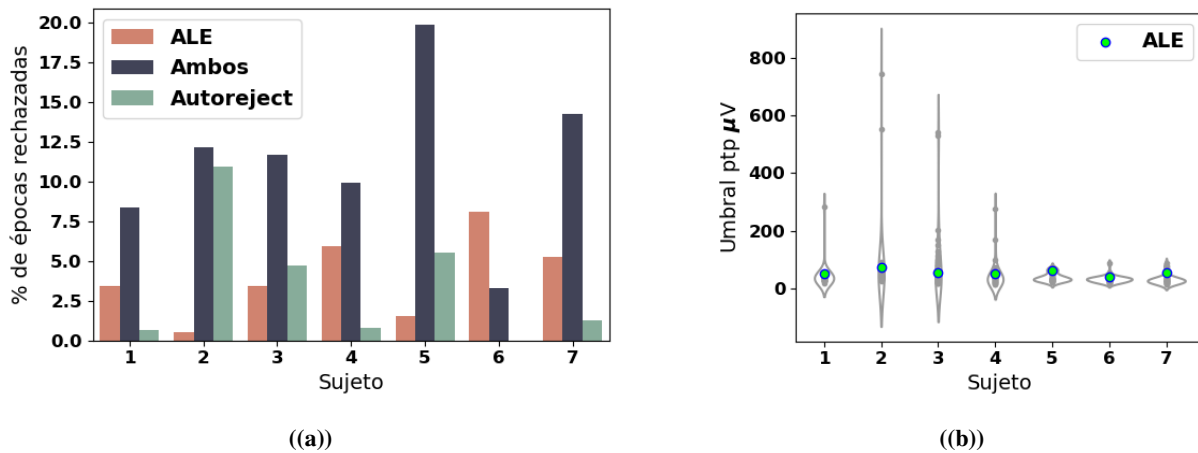


Figura 2.5: Comparación de la limpieza de épocas según el algoritmo automático implementado, ALE y Autoreject [42]. (a) Gráfico de barras del porcentaje de épocas rechazadas sobre el total de un sujeto. La etiqueta “ambos” refiere a épocas rechazadas coincidentemente por ambos métodos. (b) Distribución por sujeto de los umbrales ptp de los sensores dados por el Autoreject, y con un punto el umbral ptp único por sujeto dado por ALE.

2.6. Análisis de similitud representacional

Como se mencionó en la introducción, la estabilidad de una representación neural y la distintividad entre representaciones son propiedades clave para obtener precisión y exactitud en la evocación de memorias. En este capítulo se detalla la implementación de RSA (análisis de la similitud representacional) de las categorías semánticas a través del tiempo y entre ellas. Primero se explica el método para obtener las matrices de similitud y la correlación a través del tiempo. Se implementaron las matrices intra-categoría junto con un modelo nulo temporal. Luego, se compararon las matrices de similitud intra y entre categoría mediante un test estadístico de permutaciones no paramétrico. Finalmente, en base a estos resultados se seleccionaron las dos categorías más estables para continuar con el análisis de clasificación con algoritmos de aprendizaje supervisado.

A. Matriz de similitud por categoría semántica

Para estudiar la similitud en un conjunto de ensayos a través del tiempo se implementó una adaptación del método dado por [43], con lo que se estudió la condición de visualización. Para un sujeto, se toma un conjunto de N ensayos, por ejemplo, ensayos pertenecientes a la misma categoría. Cada ensayo se puede pensar como una matriz de $ch \times M$ siendo $ch = 30$ la cantidad de sensores y $M = 68$ la cantidad de muestras temporales dadas por la duración de las épocas (Ver Tabla 2.1). En la Figura 2.6 se esquematiza la forma matricial de un ensayo, donde cada columna corresponde al patrón espacial en un instante de tiempo t_i , es decir, la distribución de los voltajes de los ch sensores en ese tiempo. Entonces, para obtener la matriz de similitud espacial en un instante de tiempo particular se calculan los coeficientes de Pearson (en valor absoluto) entre todos los pares de vectores espaciales en ese tiempo. Por ejemplo entre el ensayo A y C , el coeficiente $R(AC)$ en el tiempo t_i se calculó como

$$R(AC) = \frac{|\sum_{k=1}^{ch} (A_k - \bar{A})(C_k - \bar{C})|}{\sqrt{\sum_{k=1}^{ch} (A_k - \bar{A})^2} \sqrt{\sum_{k=1}^{ch} (C_k - \bar{C})^2}} \quad (2.2)$$

donde A y C representan los vectores espaciales en el tiempo t_i de los ensayos A y C , respectivamente. En la Figura 2.6 se muestra un esquema del procedimiento, y el elemento de la matriz de similitud representa la correlación $R(AC)$. Al calcular las correlaciones espaciales entre todos los pares de vectores espaciales de un tiempo en particular se completa la matriz de similitud de ese tiempo. Se ve entonces que la matriz de similitud a cierto tiempo para un conjunto de N ensayos es una matriz cuadrada simétrica de tamaño $N \times N$ donde los valores de la diagonal son siempre 1. Finalmente, se promediaron los $\frac{N(N-1)}{2}$ elementos fuera de la diagonal, obteniendo entonces un sólo R para cada tiempo por conjunto de N ensayos.

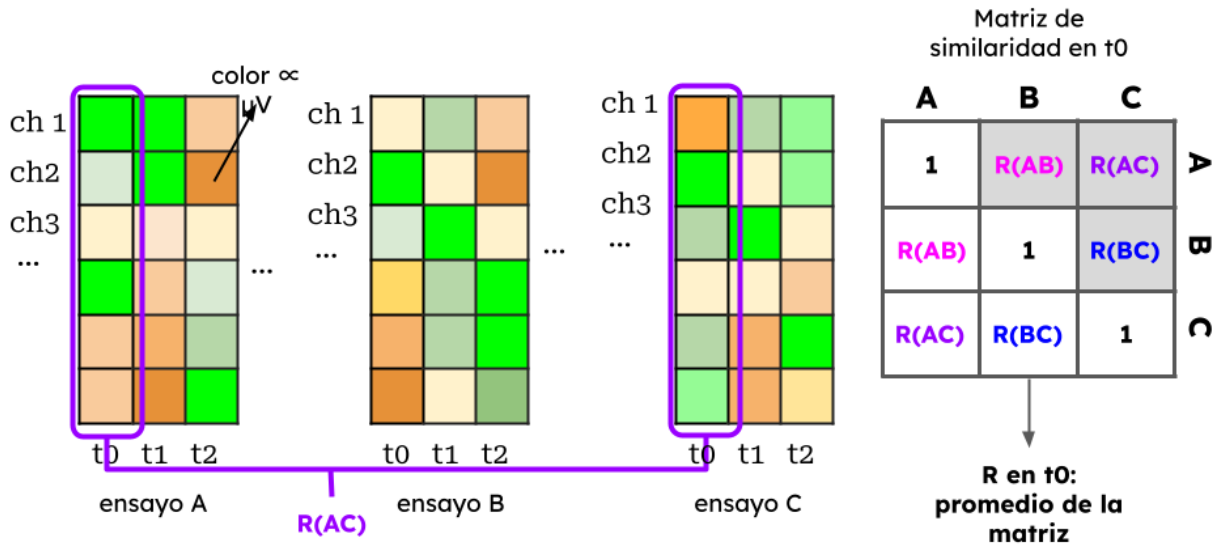


Figura 2.6: Esquema delineando la generación de la matriz de similitud a un momento específico t_0 , concebida como la correlación espacial entre ensayos (Ecuación 2.2). En el último paso se promediaron los elementos superiores de la diagonal para obtener el $R(t_0)$ [43].

Este valor R representa entonces la similaridad entendida como la correlación en la distribución espacial de los voltajes. Para estudiar la dinámica temporal de la similaridad en una categoría, se aplicó este procedimiento tomando como conjunto a los ensayos de cada categoría por separado. Aplicando el procedimiento a todos los tiempos se obtuvo la curva de similaridad por categoría en el tiempo. Para cada categoría el tamaño de las matrices de similaridad fue de aproximadamente 200×200 ensayos (40 imágenes únicas por categoría $\times 5$ repeticiones, descontando los ensayos descartados en la limpieza).

Para evaluar si el comportamiento de la curva de similaridad en el tiempo corresponde a una dependencia temporal separable del azar, se creó un modelo nulo bajo la hipótesis nula de que no existe relación entre la correlación espacial con el tiempo. Para ello, se mezclaron las columnas de la matriz de datos de $ch \times M$ mencionada anteriormente (siendo $M = 68$ la cantidad de muestras temporales). De esta forma, cuando se calcularon las matrices de similaridad, lo que se obtuvo es la correlación entre el vector espacial A a un tiempo t_i contra el vector espacial del ensayo C a un tiempo t_j . Por lo tanto la pregunta planteada es si la correlación de los sensores de los ensayos en un cierto tiempo es mayor que la correlación entre señales correspondientes a diferentes tiempo del estímulo, es decir, si se cumple $R(A_i C_i) > R(A_i C_j)$ con $i \neq j$, en promedio de los pares de vectores y para 1000 permutaciones de tiempo para cada categoría de las que se tomó como umbral de este modelo nulo el percentil 95 %.

B. Similaridad intra y entre categorías

Por el momento, se propusieron métodos para evaluar la similaridad de los patrones espaciales de las categorías a través del tiempo. Pero esto no implica que sea diferente la similaridad entre una categoría y otra. Para determinar si las categorías son similares entre sí y disímiles entre ellas se calcularon matrices de similaridad de tipo intra-categoría y entre-categoría como se muestran esquemáticamente en la Figura 2.7. Para que el resultado entre los dos tipos de matrices sea comparable se computaron matrices del mismo tamaño, y el procedimiento fue también intra-sujeto y de la siguiente manera.

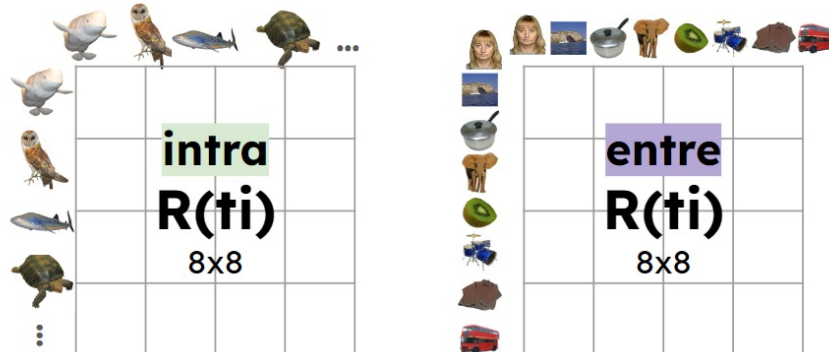


Figura 2.7: Esquema que representa la diferencia entre las matrices de similaridad intra-categoría (los ensayos pertenecen a imágenes únicas de la misma categoría), y las de entre-categorías. Esta última se conforma con ensayos de una imagen por categoría. Se construyeron del mismo tamaño para que el cálculo del R (promedio) fuese comparable.

Para construir las matrices **intra-categoría** se tomaron subconjuntos de $N = 8$ ensayos de una misma categoría pero con el cuidado de no tomar dos ensayos del mismo estímulo pues según la bibliografía [24] esto corresponde a una relación de auto-similaridad (en lugar de intra-similaridad). Con estos ensayos se obtuvo una curva de similaridad en el tiempo con el método descrito en la Subsección A. Es decir, el valor medio sobre la matriz de un determinado tiempo corresponde a un punto en la curva temporal. Luego, se tomó otro subconjunto de 8 ensayos diferentes a los primeros y se obtuvo nuevamente una curva de similaridad en el tiempo. Esto se repitió hasta agotar los ensayos disponibles de cada categoría de cada sujeto. Así se obtuvieron aproximadamente 25 curvas (200 ensayos divididos en subconjuntos de 8). Para tener un análisis que abarque las correlaciones entre todos los estímulos correspondientes a una categoría, este procedimiento se repitió pero alternando aleatoriamente la división de los subconjuntos de 8 ensayos, hasta alcanzar una cantidad de aproximadamente 1000 curvas de similaridad en el tiempo. Es decir, esta repetición con permutación en los subconjuntos se realizó para (a) obtener comparaciones estadísticamente robustas y (b) cruzar lo más ampliamente posible los estímulos correspondientes a una categoría. Por ejemplo, si en la primera separación de subconjuntos el estímulo de la imagen de la tortuga y del elefante quedaron en conjuntos diferentes, se da la posibilidad de que se crucen en otra repetición. Y así con todos los estímulos para que la comparación sea abarcativa.

Como se mencionó, la similaridad intra-categoría se quiere contrastar contra una posible similaridad en estímulos visuales de distintas categorías, llamada entre-categoría. Para construir las matrices entre-categoría, se tomaron subconjuntos con un ensayo por categoría, es decir, subconjuntos de 8 ensayos hasta agotar los datos disponibles, obteniendo aproximadamente 200 matrices de (8×8) . De nuevo este proceso se repitió hasta obtener alrededor 1000 matrices de ensayos entre-categorías.

Nótese que como se tuvieron 40 imágenes únicas por categoría, nuevamente se quiere abarcar en lo posible la comparación entre todos los estímulos de distintas categorías. Se propuso un límite de 1000 repeticiones dado que el espacio de permutaciones para construir estas matrices $\ggg 40^8$ y el tiempo de cómputo por matriz es de aproximadamente 1,5 s.

B.1. T-test no paramétrico de permutación de clusters

Para realizar la comparación entre estas dos clases de curvas de similaridad (intra y entre categoría) se implementó un test estadístico no paramétrico de clusters [44], el cual indica si las curvas son significativamente diferentes en ventanas temporales a las que el test denomina clusters. Por ejemplo, en fMRI un cluster podría referirse a un conjunto de voxels vecinos.

El test se implementó con MNE-Python, y el método se describe a continuación. Para cada punto temporal se tienen n valores de similaridad de cada clase, supongamos A y B.

1. Se calcula el típico t-test con los datos de cada punto temporal. Si hay 2 o más puntos consecutivos en el tiempo para los que del test se obtuvo que la similaridad de A fue mayor que la similaridad de B con un $p\text{-value} < 0,05$ se dice que se encontró un cluster. El mismo se caracteriza por el valor T, que es la suma de los t valores de los puntos del cluster.
2. Se permutan aleatoriamente las etiquetas de los datos de las dos clases. De modo que se obtienen n datos por punto temporal de cada clase siendo esta etiqueta aleatoriamente correcta (o incorrecta).

3. Se repite el punto 1 en los datos con las etiquetas alteradas. De todos los clusters encontrados con sus respectivos T , solo se almacena el T de mayor valor, que se puede llamar T_{max} .
4. Se repiten los puntos 2 y 3 una cierta cantidad de veces que por defecto es 10000.
5. Se obtiene la distribución de los T_{max} , cada uno obtenido en cada repetición. Esta es la distribución de los máximos valores de T sobre un modelo nulo.
6. De esta distribución se obtiene el valor del percentil 95 % (que corresponde a $p\text{-value} < 0,05$). Se considera a este valor como umbral dado por el modelo nulo T_0 .
7. De los clusters sobre los datos originales obtenidos en el punto 1, se consideran significativos únicamente los que tuvieron $T > T_0$.

Es decir, la idea general del test es encontrar la distribución empírica de los máximos valores de los t-tests obtenidos por la permutación en la etiqueta de los datos, como el modelo nulo. Luego, se pide que los t-test obtenidos sobre los datos originales superen al del modelo nulo considerando que eso pueda pasar por azar con una probabilidad menor al 5 %.

Cabe destacar, como se mencionó en el punto 1, que se pidió que la similaridad intra-categoría fuese no solo distinta sino que mayor a la de entre-categorías.

2.7. Algoritmos de clasificación de aprendizaje supervisado

En esta sección se describe el procedimiento sobre la clasificación de los ensayos usando algoritmos de aprendizaje supervisado. Los algoritmos utilizados fueron el de Support Vector Machines con kernel lineal y la Regresión Logística. La clasificación se realizó siempre entre dos categorías. Se aplicó un enfoque de resolución temporal en la clasificación, esto es que se entrenó y evaluó un clasificador con los patrones espaciales de los ensayos para cada tiempo t_i . En el marco de RSA, la exactitud del clasificador a través del tiempo se interpretó como la disimilaridad de los patrones de activación para las distintas categorías [25]. Es decir, la hipótesis de un buen desempeño en la clasificación implica que existe distintividad entre las categorías, y en estos tiempos los sensores poseen información acerca del procesado y reconocimiento del objetos. Además, se analizó el modelo obtenido por el algoritmo, siendo que los pesos de cada vector de características indican la relevancia de cada sensor en la clasificación. Para cada clasificador, se optimizó el parámetro de regularización C aplicando métodos de validación cruzada.

A. Preprocesado y validación

Como se mencionó, la entrada del clasificador fue el vector de patrón espacial en un determinado tiempo, de modo que se entrenó y evaluó un clasificador para cada tiempo. La evaluación del clasificador se realizó mediante la validación cruzada LOO (Leave-One-Out). Como se muestra en la Figura 2.8, esto significa que se entrena el modelo con los $N-1$ ensayos de los N disponibles y se hace una evaluación sobre el dato excluido. Es decir, a este nuevo vector de entrada se aplican los pesos dados por el entrenamiento y esto da una predicción de la etiqueta. Si el ensayo fue clasificado correctamente el resultado es positivo, y en caso

contrario es negativo. Este procedimiento se repite N veces, habiendo usando todos los datos una vez para evaluación. Finalmente, el porcentaje o proporción de aciertos en la clasificación es lo que se denomina la **exactitud** del clasificador. Entonces para evaluar el desempeño del clasificador se utilizó esta métrica de exactitud que depende del tiempo.

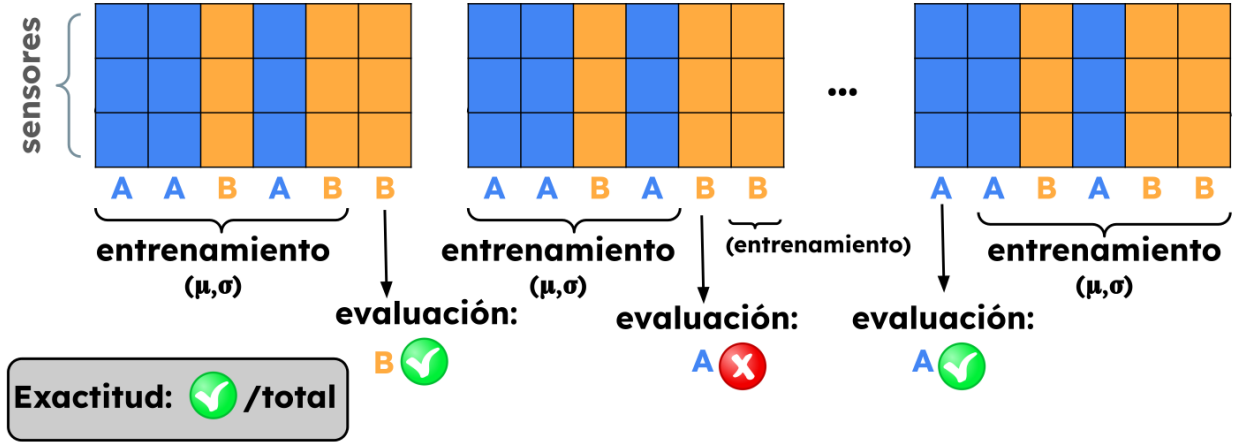


Figura 2.8: Esquema del método de validación cruzada LOO (Leave-One-Out). Se tienen los ensayos de las condiciones A y B en un cierto tiempo, donde los volajes de los sensores son el vector de características de tamaño $ch = 30$. Los ensayos se dividen en *entrenamiento* y *evaluación*. Del conjunto de entrenamiento se obtienen los parámetros para el re-escaleo (μ, σ) , el cual se aplica también al ensayo de evaluación. Luego, se entrena el clasificador y se realiza la evaluación. Luego de iterar este procedimiento por todos los ensayos se obtiene la exactitud como la proporción de aciertos en el test.

En general, el desempeño del clasificador puede considerarse satisfactoria si la exactitud supera el umbral de chance, que en el caso de dos clases es del 50 %. Sin embargo, para descartar una exactitud aparente por encima del azar, se contrastaron los datos contra un modelo nulo. El procedimiento para obtener el modelo nulo se muestra en la Figura 2.9. El mismo consiste en intercambiar aleatoriamente las etiquetas de los ensayos, realizar el entrenamiento y validación de LOO y obtener la exactitud de este set de datos etiquetado incorrectamente de manera aleatoria. Este procedimiento se realizó una cierta cantidad de veces (discutida posteriormente), hasta obtener una distribución de exactitudes del modelo nulo de la cual se toma el percentil 95 % como umbral de significancia.

El preprocesado utilizado sobre los datos para realizar la clasificación consistió en aplicar el re-escaleo estándar `Standard Scaler` de la librería `sklearn.preprocessing`. Esto es importante para disminuir la sensibilidad a la escala en los algoritmos de aprendizaje automático, puesto que algunas características pueden dominar sobre otras en el proceso de optimización y afectar el desempeño del clasificador. Además el re-escaleo puede ayudar a la convergencia de los métodos iterativos que implementan los algoritmos de optimización. Este re-escaleo consiste en restarle a cada dato la media del conjunto, y dividirlo por la desviación estándar del conjunto, es decir, a partir de una variable x_i se obtiene una nueva variable z_i como

$$z_i = \frac{x_i - \mu}{\sigma}. \quad (2.3)$$

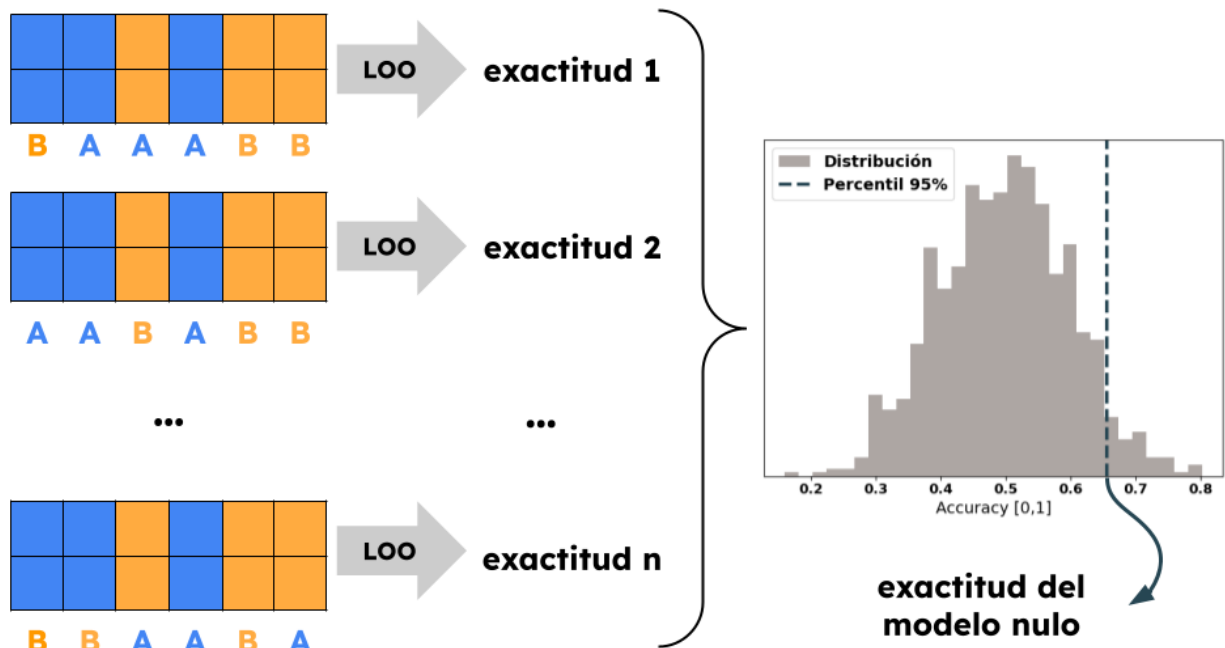


Figura 2.9: Esquema para calcular el umbral de significancia en base a la exactitud del modelo nulo. Las etiquetas de los ensayos se intercalan aleatoriamente y se obtiene la exactitud mediante LOO-CV. (Figura 2.8). Repitiendo el procedimiento se obtiene una distribución de exactitudes de la cual se calcula el percentil 95 %.

Lo importante a destacar sobre este paso, es que la media μ y la desviación estándar σ se obtuvieron sobre el conjunto de entrenamiento, y se aplican a todos los datos, es decir que no se usaron los datos de evaluación para obtener los parámetros de re-escalo.

En algunos conjuntos de datos, el entrenamiento del modelo puede depender fuertemente de los datos de inicialización del entrenamiento. Para caracterizar esta variabilidad se hicieron múltiples repeticiones para el cálculo de la exactitud para un sujeto. También se estudió la cantidad de permutaciones del modelo nulo suficiente para obtener un umbral de significancia robusto.

B. Matriz de confusión

Un métrica de evaluación común en el contexto de problemas de clasificación, es la de la matriz de confusión. Esta matriz se construye según los aciertos y errores en la etapa de evaluación, comparando las verdaderas etiquetas vs. las predichas pero diferenciado por categoría, como se ve en el esquema de la Figura 2.10. Este análisis suele ser sumamente útil cuando se quiere predecir, por ejemplo, el resultado de una enfermedad donde tiene mayor gravedad diagnosticar falsos negativos que falsos positivos. En esta tesis se estudió la exactitud específica a cada categoría A y B como las proporciones de “Verdadero A” y “Verdadero B” respectivamente.

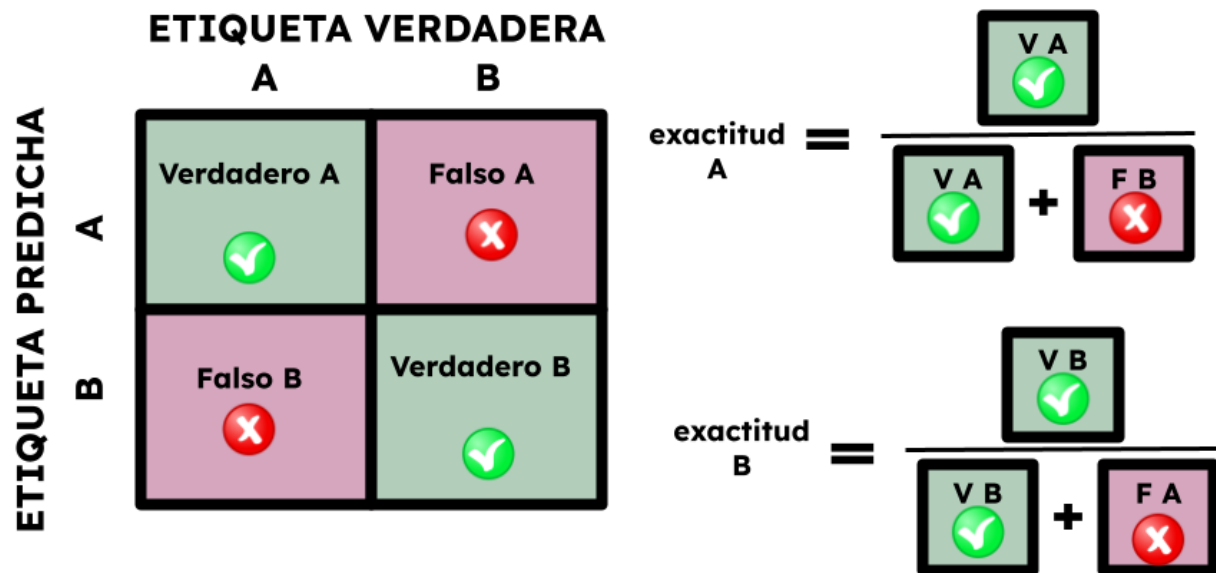


Figura 2.10: Esquema de la matriz de confusión sobre el desempeño de un clasificador. A la derecha se muestra cómo se obtuvo lo que se llamó la exactitud específica a cada categoría.

2.8. Clasificación de la señal de elaboración

El objetivo de esta sección fue estudiar si pueden agrupar por categoría los ensayos de elaboración a partir de información dada por los ensayos de visualización. Para recordar, la señal de elaboración corresponde a la generada durante el bloque experimental donde se le pidió al participante que se imagine mentalmente los objetos dados por las palabras en pantalla. Dado que los ensayos de visualización y elaboración comprenden distintas duraciones (0,5 s y 1,7 s respectivamente) y que no se conoce el momento exacto en el cual podría comenzar el proceso de imaginación, se realizó el siguiente análisis (Ver Figura 2.11).

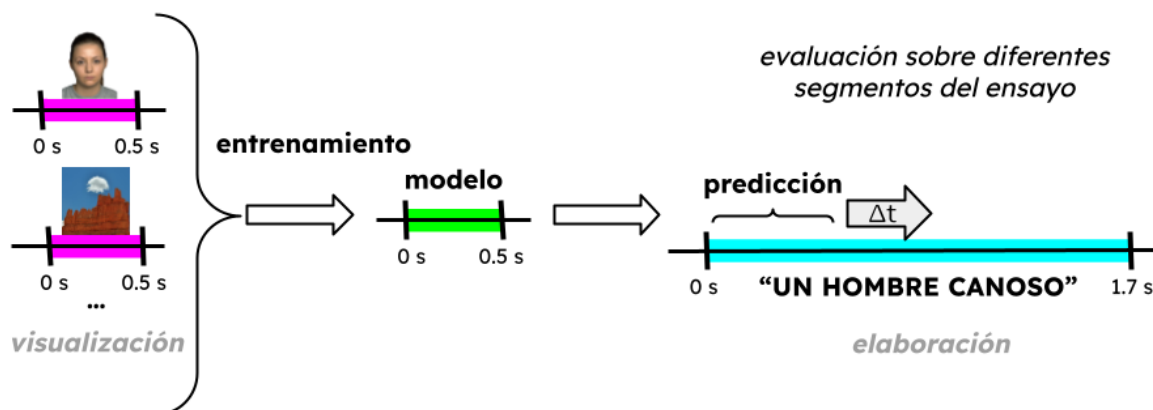


Figura 2.11: Esquema utilizado para la clasificación de los ensayos de elaboración a partir del entrenamiento con los ensayos de visualización. Se realizan las predicciones de la etiqueta aplicando los coeficientes por ventanas temporales móviles a la señal de elaboración. El objetivo es determinar las ventanas temporales con información relacionada a la categoría.

Primero se entrenó el clasificador SVM con los ensayos de las imágenes de dos categorías. Con este entrenamiento se obtuvo un modelo, es decir, valores para cada coeficiente a través del tiempo de 0 s a 0,5 s. Luego se usaron los coeficientes de este modelo para clasificar la señal de elaboración en ventanas temporales de 0,5 segundos, comenzando desde el inicio y corriéndose 1/128 s (una muestra temporal) hasta recorrer el total del ensayo de elaboración. Para cada clasificación se realizó además la prueba de significancia, con la permutación aleatoria de etiquetas (Subsección A).

Capítulo 3

Resultados

3.1. Potenciales relacionados a eventos

En un primer momento se realizó una exploración visual de las señales de electroencefalográficas. A partir de la señal obtenida con el preprocesado, se obtuvieron los potenciales relacionados a eventos (ERPs, por sus siglas en inglés de Event-related potential), es decir, potenciales relacionados temporalmente con un estímulo. Los ERPs se obtienen promediando los ensayos sobre alguna condición o clase, en este caso sobre cada categoría semánticas en la condición de visualización (Figura 3.1) y por otro lado en la condición de elaboración (Figura 3.2). Luego se promedió sobre todos los sujetos y se tomó la desviación estándar entre sujetos. Para que la amplitud entre sujetos sea comparable, se normalizó la señal por sujeto dividiendo por su valor máximo en términos absolutos.

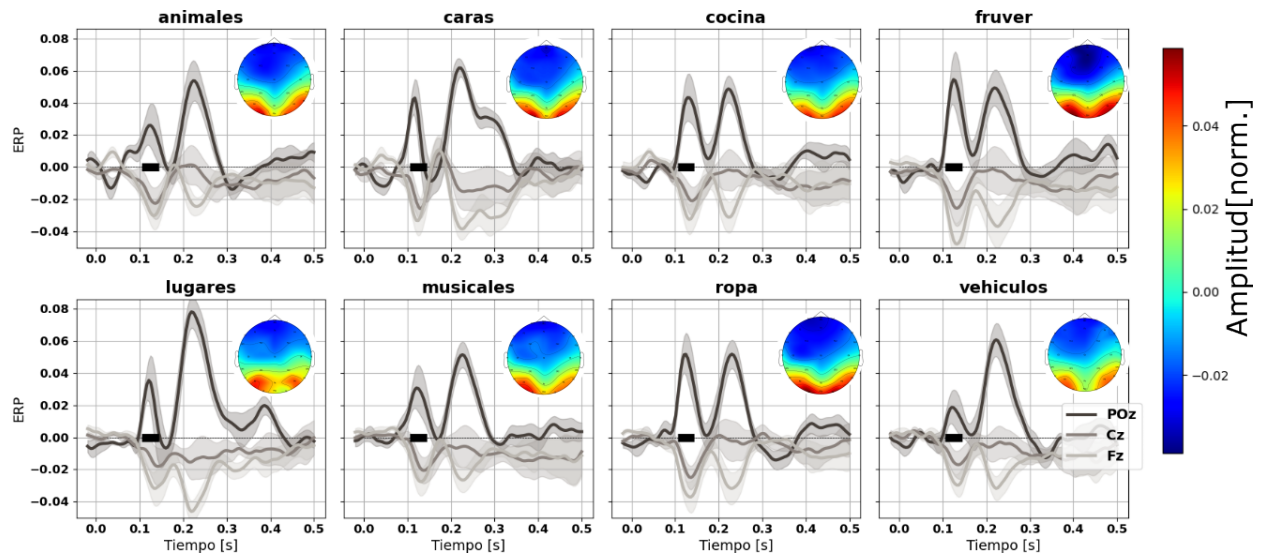


Figura 3.1: ERPs (Evoked-related potential) de tres sensores centrales - Fz, Cz y POz - promediado entre participantes para cada condición. Primero se normalizó por el valor máximo absoluto de cada sujeto, luego se promedió por categoría dentro de cada sujeto y por último se promedió sobre los sujetos. En cada gráfico se muestra el topográfico del promedio temporal señalado en línea gruesa.

En la condición de visualización, en los tres canales centrales, y para todas las categorías se observa un potencial aproximadamente a partir de los 100 ms y un segundo pico luego de los 200 ms, siendo estos potenciales polarizados entre la parte frontal y posterior. La categoría de caras humanas presenta un pico de potencial de mayor duración que en el resto de las categorías. Algunas categorías presenta un tercer pico a tiempos más largos (350 ms) pero con menos de 50 % menos de amplitud. En términos generales, la actividad es similar en todos los casos, involucrando un procesamiento rápido de bajo nivel hasta los 200 ms después del estímulo. Por estas características lo más probable es que estos potenciales estén relacionados a la tarea de mirar la pantalla, y como reacción a la observación de un estímulo visual [13]. En principio las topografías en los primeros 100 ms se ven similares y existen diferencias sutiles entre categorías que suceden después de los 200 ms.

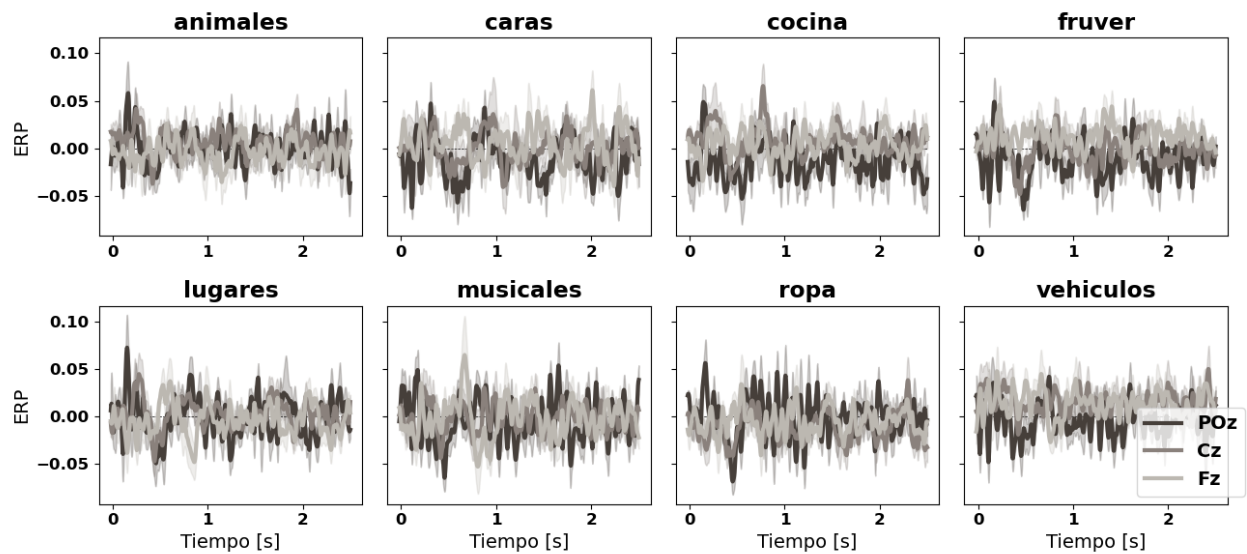


Figura 3.2: ERPs de tres sensores centrales - Fz, Cz y POz - en la condición de elaboración separados por categoría. Primero se normalizó por el valor máximo absoluto de cada sujeto, luego se promedió por categoría dentro de cada sujeto y por último se promedió sobre los sujetos.

Como era de esperarse, en la condición de elaboración no se vieron potenciales claros, indistintamente de la categoría. La respuesta a la “presentación de una palabra” parece no tener una estructura consistente en el tiempo como en el caso de la presentación de una imagen.

3.2. Análisis de similitud representacional

Con el objetivo de identificar patrones en la actividad neural se implementó el análisis de RSA. Estudiar la existencia de estos patrones podría indicar la posibilidad de identificar el procesamiento de la visualización de diferentes categorías. El análisis de RSA se implementó en modalidad intra-sujeto, es decir, de forma independiente entre participantes.

A. Matriz de similitud por condición

Para cada participante, de forma individual, se calculó la similitud a través del tiempo para condición según la Sección 2.6. A cada condición se le restó el umbral de lo que se llamó modelo nulo temporal. Es decir, cuando se calculó la correlación de los sensores entre ensayos de la misma condición pero en diferentes tiempos. Luego se graficó la diferencia entre la correlación y el umbral del modelo nulo de cada categoría. Esto se ve en la Figura 3.3 separados por sujeto, y distinguiendo entre categorías por el color.

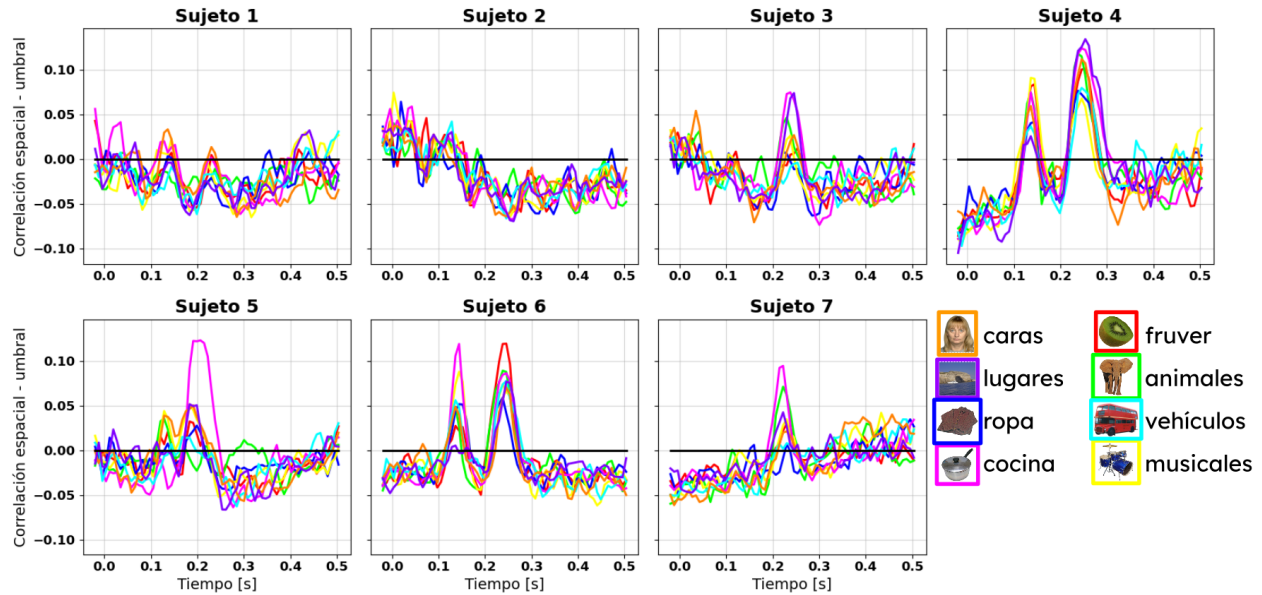


Figura 3.3: Gráfico por sujeto de la similitud intra-categoría a través del tiempo, cada color corresponde a una categoría. La curva es la diferencia entre la correlación espacial y el umbral del modelo nulo. La hipótesis nula fue que la correlación no depende del tiempo, por lo que el modelo nulo se hizo calculando la correlación entre vectores espaciales de distintos tiempos.

Se observó que la estructura de estas curvas es variable entre los sujetos. Dentro de cada sujeto, las categorías se comportan de manera parecida. De este modo, las ventanas de similitud significativa coinciden, por sujeto, entre categorías. En la mayoría de los participantes esta ventana se observa alrededor de los 200 ms. En tres de los sujetos, la curva de similitud se comporta como los ERPs de la Figura 3.1.

B. Similitud intra y entre categorías

Se realizó un análisis para determinar si las curvas de similitud intra-categorías se diferencian de las curvas entre-categorías, y más aún, si las primeras son mayores que las segundas. Para cada participante, se obtuvieron curvas de intra y entre similitud, permutando la combinación de ensayos tomados para calcular la matriz de similitud y luego la curva. Luego, para cada participante se buscaron ventanas de tiempo donde exista diferencias significativas entre la similitud intra y entre categorías. Para ello, se realizó un análisis estadístico (ver Subsubsección B.1). Un ejemplo, para el participante 5, puede verse en la Figura 3.4. La curva en color negro corresponde al valor medio de las correlaciones de las matrices entre-categoría, junto

con su error en sombreado. Lo mismo sucede para las curvas en color (distinto de negro), correspondiendo a cada categoría de las matrices intra-categoría.

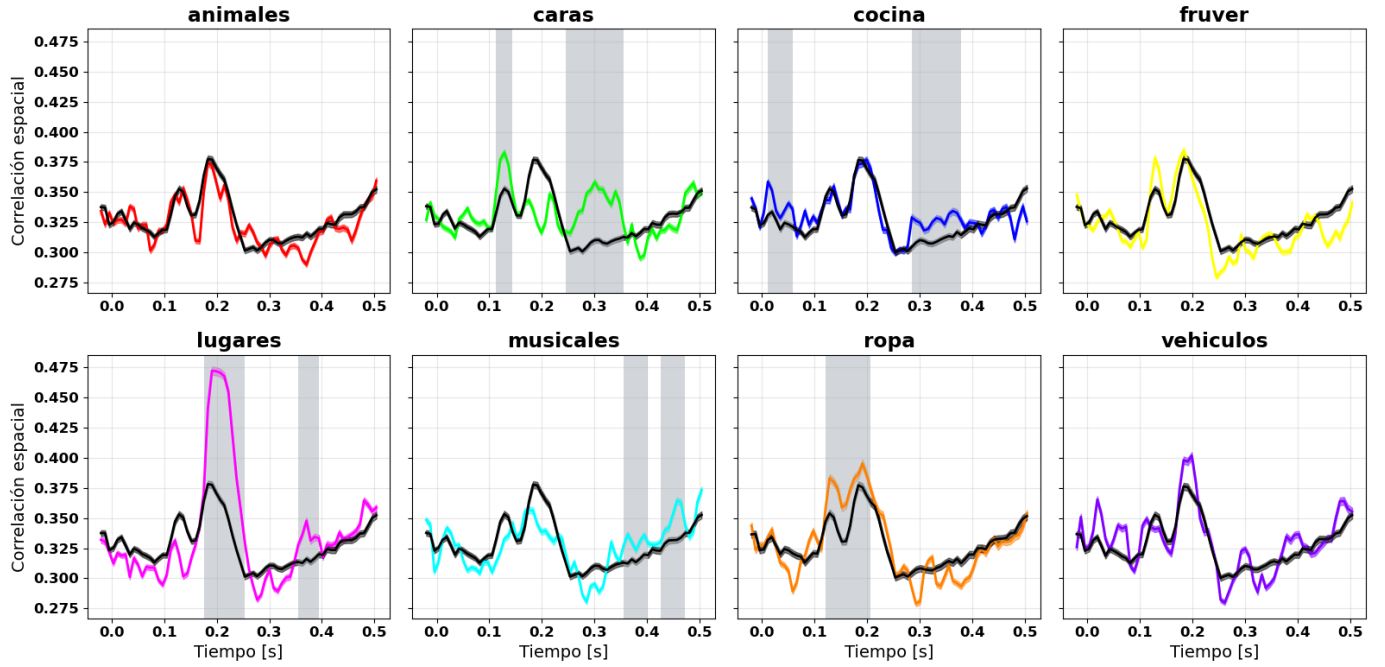


Figura 3.4: Curvas de la similaridad intra-categoría (color) y entre-categoría (negro) para las distintas categorías con los datos del sujeto 5. Se graficó el valor medio de las permutaciones y el error del promedio en sombreado del mismo color. El sombreado en gris representa las ventanas temporales en donde las clases fueron significativamente diferentes según el test de permutaciones [44].

Luego, se buscó cuantificar las ventanas significativas por categoría a través de todos los sujetos. Para ello se realizó el gráfico de barras de la Figura 3.5, donde se muestra la proporción de sujetos que cuentan con al menos una ventana significativa en cada categoría. Se encontró al menos una ventana temporal significativa en todos los sujetos en las categorías de caras y lugares. Las categorías de animales y frutas y verduras fueron la de menos consistencia entre sujetos.

Luego, se analizó la ubicación y extensión temporal de las ventanas significativas por categoría a través de los sujetos. Las mismas se muestran en la Figura 3.6, diferenciadas en color por sujeto. En gris se marcaron los límites de extensión mínima de las ventanas significativas por categoría, es decir, la extensión desde el primer tiempo en donde algún sujeto presenta una ventana significativa para esa categoría hasta el último. Para cuantificar la consistencia de las ventanas temporales significativas a través de los sujetos para las distintas categorías se calculó lo que se denominó el porcentaje de “área efectiva” abreviado como AE. El AE para cada categoría, se obtuvo como la proporción de área que ocupan las ventanas significativas con respecto a su área gris correspondiente. Esta métrica representa la consistencia en la localización y confinación de las ventanas significativas a través de los sujetos.

En la Tabla 3.1 se puede ver que el mayor AE lo tuvieron las caras y los lugares con 38,3 % y 32,6 %, respectivamente. En ambas categorías se tuvieron sujetos con ventanas significativas extensas y con consistencia entre todos los sujetos. En la categoría ropa, si bien la mayoría de los sujetos cuentan con una ventana

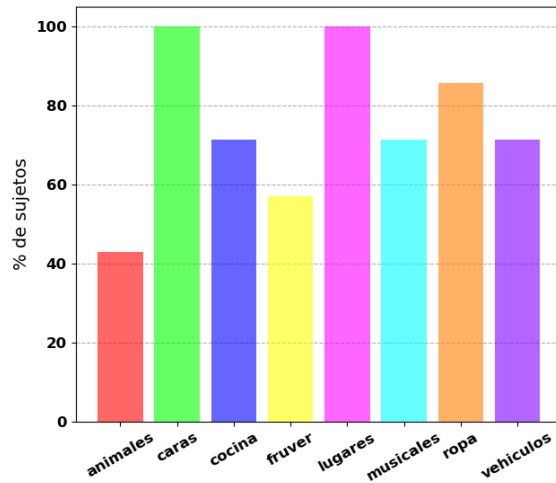


Figura 3.5: Gráfico de barras sobre el porcentaje de sujetos -sobre el total de 7 sujetos- que tuvieron al menos una ventana temporal significativa. Esto es los tiempos en donde las curvas de similitud intra y entre-categoría se diferenciaron significativamente.

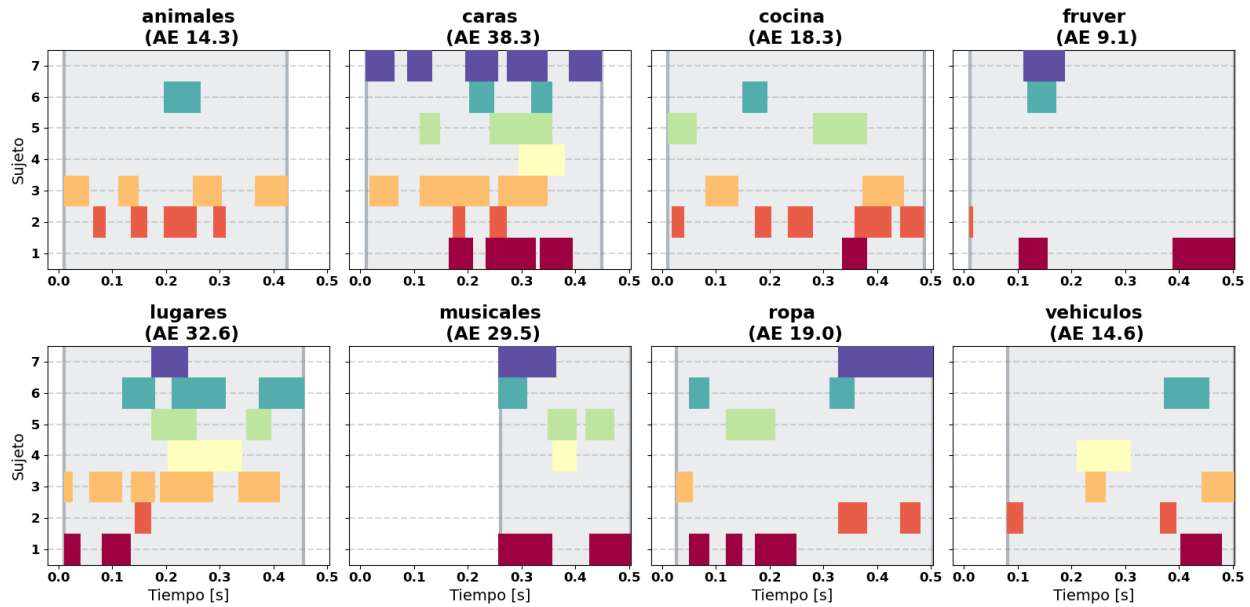


Figura 3.6: Representación de las ventanas temporales, por categoría, en donde las curvas de similitud intra y entre categoría fueron significativamente diferentes. En este caso cada color corresponde a un sujeto para facilitar la visualización. Luego, se calculó el AE (área efectiva) por categoría como el porcentaje del área significativa sobre el área de la extensión mínima de los tiempos significativos. Esta métrica indica la confinación y consistencia de ventanas significativas por categoría a través de los sujetos.

significativa, no predomina que sean extensas en el tiempo, ni confinadas en una locación temporal del ensayo. El menor AE lo tuvo la categoría frutas y verduras (9,1 %) ya que se tienen ventanas significativas en pocos sujetos y no se encontraron confinadas temporalmente, por ejemplo en comparación a la categoría

Animales	Caras	Cocina	Fruver
14,3 %	38,3 %	18,3 %	9,1 %
Lugares	Musicales	Ropa	Vehículos
32,6 %	29,5 %	19,0 %	14,6 %

Tabla 3.1: Tabla con las áreas efectivas (AE) por categoría.

musicales. En la categoría animales solo tres sujetos obtuvieron ventanas significativas, y en términos de AE quedó en ante-último lugar.

C. Comparación entre las categorías caras y lugares

En base a los resultados expresados antes, el análisis siguiente se centró en la discriminación entre las categorías de caras y lugares, con el objetivo de utilizarse en un experimento subsecuente en el Laboratorio de Neurociencias de la Memoria. Se realizó el mismo análisis estadístico no paramétrico (Subsubsección B.1) pero entre la similitud intra-categoría de caras e intra-categoría de lugares. Se obtuvieron ventanas significativas para todos los sujetos, y se muestran en la Figura 3.7. Si bien no se encontró una ventana temporal significativa que común a todos los sujetos, la ventana de [280-380] ms es significativa para 6 de los 7 sujetos.

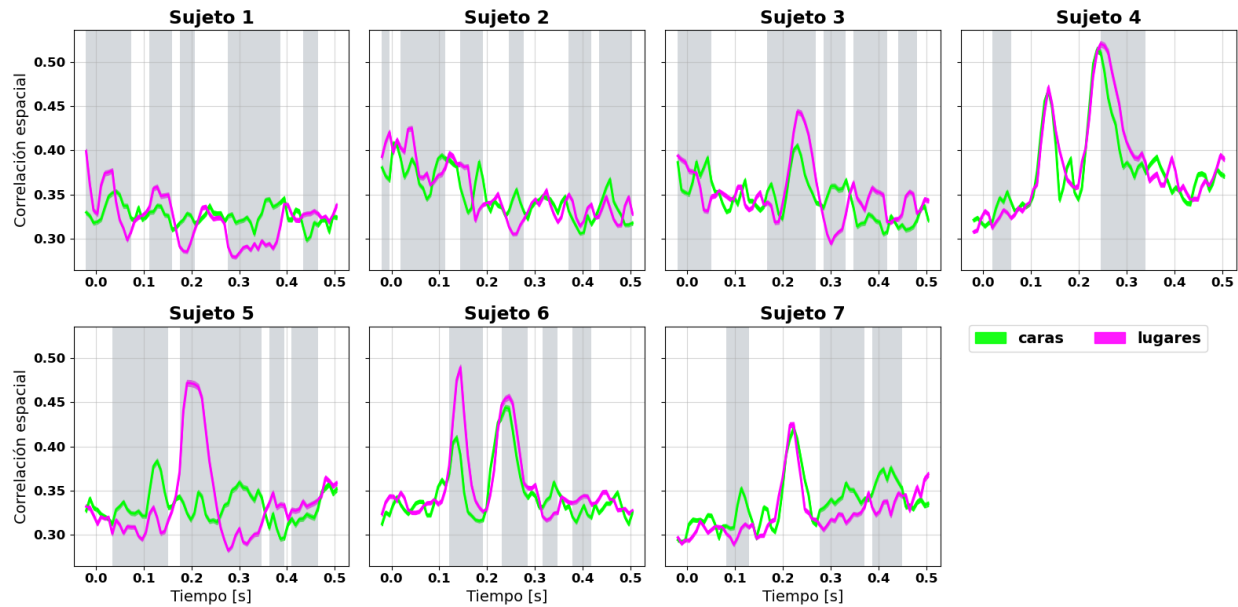


Figura 3.7: Comparación entre las curvas de similitud para las categorías “caras”(verde) y “lugares”(rosa). La curva es el valor medio y el error está en sombreado del mismo color. Se realizó el test de permutaciones para determinar las ventanas temporales (gris) en donde las curvas de las dos clases fueron significativamente diferentes. (Subsubsección B.1)

3.3. Algoritmos de clasificación de aprendizaje supervisado

Se utilizaron los algoritmos de clasificación con el objetivo de encontrar diferencias en la actividad neural de las categorías caras y lugares. Los clasificadores utilizados fueron Support Vector Machines (Lineal) y Regresión Logística. Previo su implementación para los siguientes análisis, se estudió la validación necesaria y se optimizó el parámetro de regularización de cada uno (APÉNDICE D). Luego, para ambos clasificadores se utilizó $C = 0,1$.

A. Comparación entre los clasificadores

A.1. Exactitud de clasificación

Se entrenó y evaluó un clasificador para cada sujeto. Más aún, para estudiar la dinámica temporal del reconocimiento de objetos se implementó un clasificador para cada muestra temporal. Para cada tiempo se calculó la exactitud con LOO-CV y el umbral de significancia dado por el modelo nulo.

En la Figura 3.8 se tiene la exactitud de la clasificación para cada sujeto, y se ve el comportamiento similar a lo largo del tiempo. Comparando ambos clasificadores se vio que el desempeño fue equivalente.

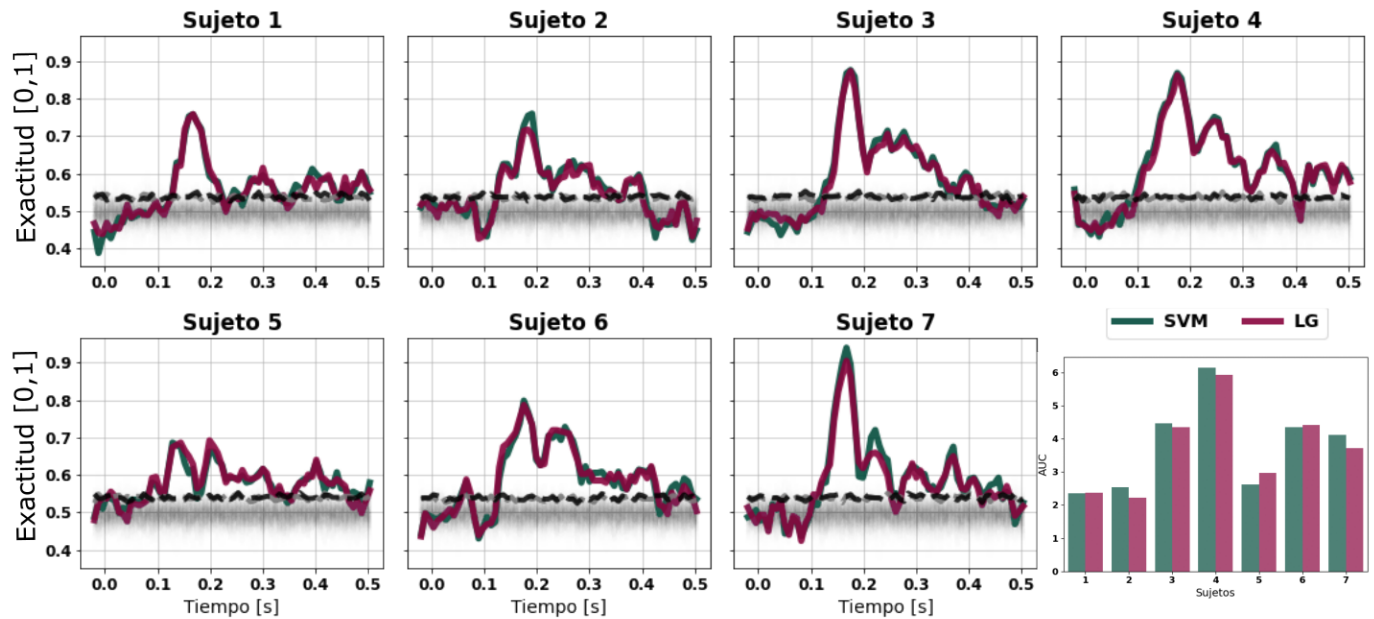


Figura 3.8: Curvas de exactitud a través del tiempo en contraste con el modelo nulo en gris sombreado, y el umbral de significancia en línea negra punteada. Se vio que el desempeño de los clasificadores fue equivalente. El gráfico de barras indica el área bajo la curva significativa por sujeto. (LR: Logistic Regression, SVM: Support Vector Machines)

Para determinar en detalle esta observación, en la Figura 3.8 se muestra un gráfico de barras realizado a partir del AUC (área bajo la curva significativa) de cada clasificador por sujeto.

Ningún clasificador predomina sobre el otro a través de los sujetos, sino que la dependencia en el desempeño fue según el sujeto, y no según el clasificador.

A.2. Tiempo de cómputo

Se almacenaron las duraciones del tiempo de cómputo para cada permutación de LOO-CV de las 5000 permutaciones que se realizaron para un sujeto en la etapa de validación (APÉNDICE D). En la Figura 3.9 se evidenció que SVM es significativamente más rápido que la regresión logística (LR). La iteración de menor duración de LR es el 144 % de la de mayor duración de SVM.

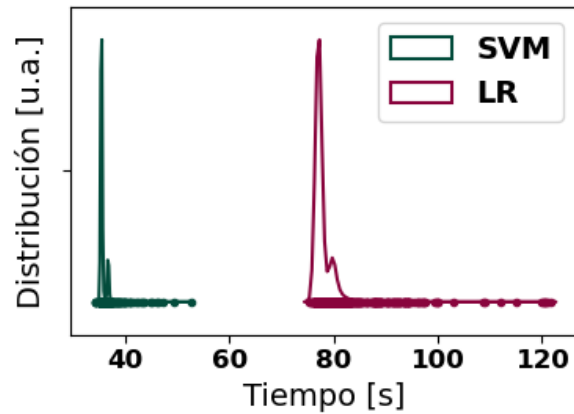


Figura 3.9: Distribución de los tiempos de cómputo de los algoritmos en las diferentes iteraciones de LOO-CV. (LR: Logistic Regression, SVM: Support Vector Machines)

También se observó que la duración en el cómputo de la regresión logística tuvo mayor variabilidad, lo que puede significar que la convergencia dependió fuertemente (o al menos en mayor medida) del conjunto de datos de entrenamiento. (STD SVM: 0,7 s, STD LR: 2,9 s).

Debido a esta gran diferencia en el tiempo de cómputo y considerando que el desempeño entre ambos clasificadores fue equivalente, se utilizó SVM para los análisis subsecuentes.

B. Clasificación en la condición de visualización

Utilizando los resultados de la clasificación de SVM, para facilitar la visualización el desempeño del clasificador a través del tiempo y de los participantes, se realizó un mapa de color que se muestra en la Figura 3.10. El color depende del AUC significativa, es decir, el área entre la exactitud y el umbral de significancia.

Se encontraron coincidencias entre sujetos, siendo la ventana temporal de entre 150 ms y 250 ms la de mejor desempeño del clasificador. En línea negra se trazaron las ventanas temporales en donde fue significativa la diferencia entre las similitudes intra-categorías para caras y lugares (Figura 3.7). En general, se observa que las ventanas de diferencia entre similitudes no es la misma que en la mayor exactitud del clasificador.

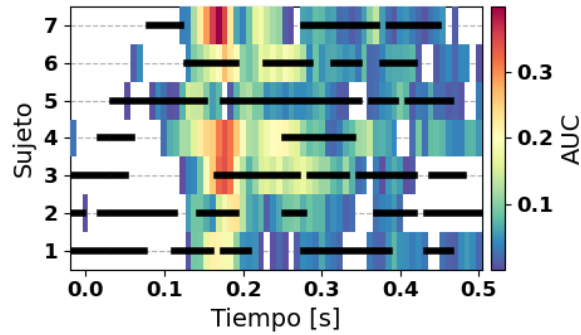


Figura 3.10: Mapa de color del AUC (área bajo la curva significativa) a través de tiempo para todos los sujetos. La ventana temporal de 150 ms a 250 ms fue la de mejor desempeño para todos los sujetos. En línea negra se graficaron las ventanas de temporales en donde caras y lugares se distinguen según RSA.

B.1. Matriz de confusión

Para estudiar si alguna de las categorías predomina en la clasificación sobre la otra se analizó la matriz de confusión. En la Figura 3.11 se muestran las curvas de la exactitud de cada categoría a lo largo del tiempo y para los distintos sujetos. En línea gruesa se marcaron los tiempos en los que el acierto en la clasificación fue significativo. En este caso se tuvo en cuenta el nivel de significancia a partir de cada elemento de la matriz de confusión. Es decir que por un lado se pidió que el porcentaje de caras correctamente etiquetadas sea mayor al del modelo nulo (en el percentil 95 %) y que el porcentaje de caras clasificadas incorrectamente sea menor a la del modelo nulo (menor al 5 %). En otras palabras se pidió que clasificador acierte más que el modelo nulo y se equivoque menos. Análogamente para la categoría de lugares.

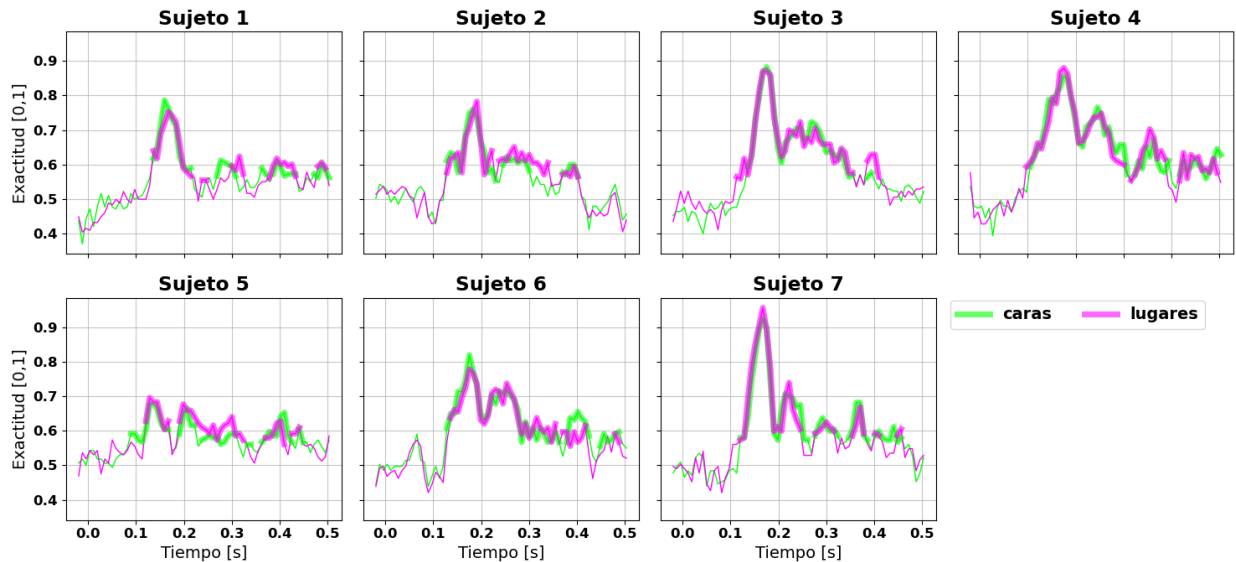


Figura 3.11: Curvas de la exactitud a través del tiempo (curva fina). Las líneas gruesas simbolizan los momentos de exactitud significativa de la matriz de confusión.

De estas curvas, y teniendo en cuenta la AUC significativa, no se encontró diferencia entre las categorías. Otra forma de decir esto es que al clasificar un dato nuevo, no utilizado en el entrenamiento, no se cometieron más o menos errores dependiendo de la categoría de este dato nuevo. Este resultado es positivo para el experimento subsiguiente en el Laboratorio de Neurociencias de la Memoria puesto que indica que en principio la exactitud específica de clasificación entre las categorías está balanceada.

B.2. Exploración del modelo

El modelo dado por SVM permite explorar los coeficientes del hiperplano optimizado para separar el espacio de atributos de las dos clases. Estos coeficientes indican la importancia o influencia de cada característica en la separación de las clases, por lo que se pueden identificar los sensores más relevantes para diferenciar los patrones de actividad neural de las categorías. Para estudiar el modelo obtenido por el clasificador, se utilizaron solamente las ventanas temporales en donde la exactitud de la clasificación fue significativamente superior al modelo nulo consistentemente entre todos los sujetos (Ver Figura 3.10). Bajo esta condición se eligieron dos ventanas temporales, la primera comprendida entre los 128 ms y 324 ms, y la segunda entre los 337 ms y 394 ms; habiendo una separación de 13 ms entre ambas. Luego, se promedió el valor de cada coeficiente dentro de cada ventana temporal, para cada sujeto. De forma que se obtuvo un valor del coeficiente para cada sensor por sujeto.

Para ver los sensores con mayor peso en el modelo, se graficaron las distribuciones de los coeficientes de todos los sensores por sujeto. En la Figura 3.12 se ve que la mayoría de los sujetos tuvieron algunos sensores con peso predominante. Se marcó con línea punteada el percentil 95 %, con lo que se pueden observar los sensores encuentran entre el 5 % de los más altos. Además, se ve que estos valores se alejan de la distribución considerándose atípicos según el método del rango intercuartil. Se ve que estos sensores coinciden en la región de la corteza occipital posterior y derecha, y parietal derecha. En la segunda ventana significativa (según el clasificador y para todos los sujetos) también tuvieron peso significativo los sensores frontales en los sujetos 1 y 2.

[width=0.8]figures/ch05/ch05_{svm}coef_s/bitmapflechas_v0.png

Figura 3.11: (a)

[width=0.8]figures/ch05/ch05_{svm}coef_s/bitmapflechas_v1.png

Figura 3.11: (b)

Figura 3.12: Distribución de los valores de los coeficientes de los sensores separados por sujeto. Cada coeficiente es el valor medio en la ventana temporal (a) [128 ms-324 ms] y (b) [337 ms-394 ms]. En estas ventanas la clasificación fue significativamente acertada para todos los sujetos. Además se señalaron los sensores con los coeficientes de valor superior al percentil 95 %.

3.4. Clasificación de la señal de elaboración

Para estudiar si se podía relacionar la información de la señal de visualización con la de elaboración se entrenaron clasificadores de Support Vector Machines lineales utilizando los ensayos en la condición

de visualización, pero la clasificación se implementó sobre los ensayos de elaboración. Es decir, se buscó clasificar la señal de los procesos de elaboración de los objetos, usando la información de los patrones neurales en el proceso de visualización y reconocimiento (Sección 2.8).

Para encontrar el momento en el que podrían encontrarse patrones del proceso de elaboración, la clasificación se realizó sobre ventanas temporales móviles de forma de recorrer por completo el ensayo de elaboración. En la Figura 3.13 se muestran los resultados obtenidos.

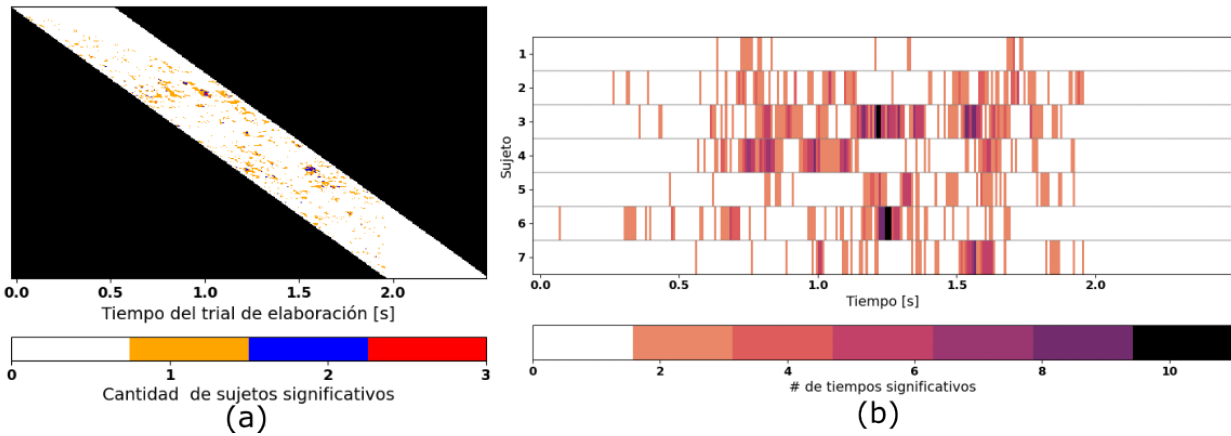


Figura 3.13: Resultados de la clasificación de los ensayos de elaboración. Se entrenó en modelo con los ensayos de visualización y se evaluaron segmentos del ensayo de elaboración comenzando en diferentes tiempos. **(a)** En color se muestra la cantidad de sujetos para las cuales, en esa ventana de tiempo del ensayo de elaboración, fue clasificado correctamente de forma estadísticamente significativa. **(b)** La escala de color representa la cantidad de tiempos del ensayo de visualización que sirvieron para clasificar correctamente ese tiempo de elaboración.

En la Figura 3.13 (a), el eje horizontal indica el tiempo del ensayo de elaboración. La diagonal blanca representa los segmentos en donde se evaluó la señal de elaboración. Para comprender el armado de este gráfico, se puede comenzar por la esquina superior izquierda, donde la diagonal blanca se ubica sobre el segmento de 0 – 0,5 s. Aquí se usaron los coeficientes del modelo (que varían en el tiempo durante medio segundo), se aplicaron al primer medio segundo del ensayo de elaboración. Es decir, los coeficientes en el tiempo t_0 se usaron para clasificar la señal en el tiempo t_0 de elaboración, y los del t_1 para clasificar la señal en el tiempo t_1 y así sucesivamente. La fila se encuentra en blanco lo que quiere decir que de esta forma no fue correcta la clasificación. En la segunda fila, el on-set es de 1/128 s, es decir, una muestra temporal. En caso entonces se usaron los coeficientes del tiempo t_0 para clasificar la señal de elaboración en el tiempo t_1 , etc. Los colores representan la cantidad de sujetos que tuvieron una clasificación correcta significativa en ese segmento. Esta figura es complementada por la Figura 3.13 (b) en donde se marcaron los tiempos en donde la clasificación fue significativa. En este caso, a escala de color representa la cantidad de tiempos del ensayo de visualización que sirvieron para clasificar correctamente ese tiempo de elaboración.

La aplicación de este defasaje permitió mostrar que la señal de elaboración pudo ser clasificada correctamente después de los 500 ms. También existió una segunda clasificación exitosa en aproximadamente los 1500 ms. De modo que se pudo observar que los tiempos de elaboración son mucho más largos que los de visualización. Por otra parte, se observaron muchos tiempos con significancia pero aislados (mayormente

en los sujetos 1, 2 y 5) que indicarían una significancia poco confiable, a diferencia de casos como los sujetos 3, 4 y 6 en donde hubo estabilidad de la discriminación. Para discernir entre estos casos con mayor confiabilidad haría falta tener más ensayos de elaboración para tener una estadística más robusta.

Capítulo 4

Discusión y conclusiones

Desde el campo de la psicología cognitiva hasta la neurociencia, se estudió ampliamente el problema de la codificación de la información en la actividad cerebral. Como en una gran variedad de temas, los déficits y trastornos cognitivos fueron un puntapié para generar preguntas en la neurociencia. Así es el caso del procesamiento de reconocimiento de objetos y la organización de la información en el cerebro diferenciada por categorías semánticas. La evidencia de este siglo identificó varias categorías que pueden discriminarse entre las señales de neuroimagen utilizando análisis de patrones multivariados. Las categorías más estudiadas fueron las caras humanas, animales, objetos y lugares. En esta tesis se llevó a cabo un experimento con registros electroencefalográficos sobre el reconocimiento visual de objetos con estímulos de imágenes de diferentes categorías. Adicionalmente se implementaron estímulos de palabras para los cuales la tarea fue imaginarse el objeto que nombra la palabra, a este tipo de ensayos se los denominó condición de elaboración. El objetivo, de manera general, fue estudiar la señal electroencefalográfica con distintos métodos de análisis multivariado para investigar la estabilidad de las representaciones neurales. Luego, se buscó decodificar esta señal, es decir, discriminar la categorías semánticas correspondientes a los estímulos en esta ocasión con algoritmos de aprendizaje supervisado. Seguidamente, se buscó relacionar la decodificación con los procesos fisiológicos, explorando las regiones cerebrales relevantes durante la decodificación. A lo largo de la tesis, se trabajó con un enfoque temporal. Es decir, valiéndose de la alta resolución temporal del equipo de EEG de 256 Hz, todos los análisis se realizaron en función del tiempo para determinar la dinámica temporal de estos procesos.

En primer lugar, se diseñó un experimento acorde a los trabajos previos. Se eligió un protocolo que no durara más de una hora y de una sola sesión. Para poder realizar un análisis intra-sujeto se utilizó una cantidad de 1600 estímulos por participante, que pueden subdividirse en 8 categorías semánticas de 40 imágenes únicas de cada una, y cada imagen repetida 5 veces. En un experimento similar [32] se vio que aumentar la cantidad de presentaciones de un mismo estímulo contribuye a aumentar la exactitud de un clasificador. Además, se vio que este aumento en la exactitud proporcional a la cantidad de repeticiones, sino que la diferencia fue disminuyendo. En este trabajo el volumen de datos fue suficiente para los análisis intra-sujeto, realizándose análisis cualitativos y cuantitativos de los datos. Además, la exactitud de nuestro clasificador fue comparable con el estimado por el trabajo [32]. Queda pendiente analizar la variación de nuestros resultados disminuyendo la cantidad de estímulos, por ejemplo a 3 repeticiones de las imágenes para que el experimento dure tan solo media hora. Es decir, ver si esto es posible sin perder el desempeño como indicaría aquel trabajo.

Una vez realizado el experimento, se exploró a primer orden de la señal observando los ERPs, en este caso usando tres canales sobre la línea central Fz, Cz y POz. En la condición de visualización (Figura 3.1) se vieron potenciales provenientes de la región occipital con dos componentes que comienzan en los 100 ms y 200 ms como se esperaba para un estímulo visual [12, 45]. Entre las categorías, las caras humanas se destacan por tener más estabilidad en la segunda componente. Esta diferencia en la topografía espacial es consistente con la bibliografía, que muestra que la componente N170 es más prominente en los estímulos de rostros en comparación con objetos que no son rostros [13]. Observado los topográficos queda claro que a partir de los 100 ms post-estímulo la activación es predominantemente occipital indistintamente de la categoría, indicando un procesamiento de bajo nivel en respuesta a la visualización de una imagen. Por otro lado, los potenciales relacionados a la condición de elaboración no presentaron ninguna estructura, como se muestra en la Figura 3.2 para mismos tres canales sobre la línea central. Esto también era esperado, dado que en pensamientos creativos, como la elaboración mental de una imagen a partir de una palabra, es un proceso menos controlado que la visualización de una imagen. El momento en el que se desencadena el proceso de elaboración no es certero, por lo que no se puede obtener un potencial vinculado al estímulo de la misma forma que en visualización [46]. En consecuencia, para obtener información sobre la condición de elaboración se utilizaron los resultados de los algoritmos de clasificación para la condición de visualización, buscando una relación entre estas representaciones.

En cuanto a los ensayos en la condición de visualización, se prosiguió con estudiar la similaridad espacial a través del tiempo (RSA). Para validar la dependencia de la similaridad con el tiempo se usó un modelo nulo que consistió en calcular la correlación entre los sensores de los ensayos en distintos tiempos. Se observó que la estructura de estas curvas (Figura 3.3) depende principalmente del participante, y que para un mismo participante las categorías presentan curvas similares entre sí. En este mismo sentido, se vio que las ventanas significativas coincidían entre las categorías para el mismo sujeto, igualmente entre sujetos en general coincidieron alrededor de los 200 ms. De esta manera se vio que para los sujetos 4 y 6 las curvas de similaridad se asemejaron a los ERPs, aunque con una activación casi 50 ms posterior a estos. En los casos de los sujetos 3 y 7, la similaridad conserva un solo pico semejante a la componente de 200 ms de los ERPs, de nuevo con una postergación de algunos milisegundos. Por consiguiente, el aumento de la similaridad no parece responder únicamente al procesamiento de bajo orden visual ya que tiene una diferencia temporal con la típica componente N100 [47].

Para estudiar la estabilidad de la representación neural específica a cada categoría, se compararon las curvas de similaridad intra-categoría frente a entre-categorías para cada categoría y separadas por sujeto (Ver APÉNDICE C). Se observó una gran variabilidad de estas representaciones entre los sujetos. No obstante, es importante destacar que cada sujeto mostró patrones de similaridad específicos para cada categoría, es decir, que se diferenciaron de las similaridad entre-categoría en distintos segmentos temporales (Figura 3.5). Estas diferencias que se presentaron para las diferentes categorías, se estudiaron a través de los sujetos ubicando estos segmentos temporales en la Figura 3.6. Se vio que gran parte de estas diferencias se presentaron después de los 250 ms, más del doble de tiempo desde la primera reacción al estímulo según los ERPs. En particular fue interesante el caso de los instrumentos musicales, en 5 sujetos las ventanas significativas se ubicaron exclusivamente después de este tiempo de 250 ms. También se vieron ventanas significativa para todos los sujetos en la categoría de ropa pero de poca duración y deslocalizadas temporalmente entre suje-

tos. Las categorías que más se destacaron fueron las caras humanas y los lugares con representaciones más estables para todos los sujetos. Con respecto a las caras, esto era esperable de acuerdo a la bibliografía en donde la respuesta a este estímulo es de mayor intensidad en términos de amplitud y tiene una temporalidad característica de pico de potencial en los 170 ms [13]. En cuanto a los lugares, asumimos que la diferenciación se debe, principalmente, a que los estímulos son esencialmente diferente al resto. El resto de las categorías presentan un objeto sin fondo, mientras que los lugares son una composición de diferentes objetos, adicionalmente la imagen de un lugar mantiene una forma cuadrada. Para corroborar esta hipótesis se podría hacer un clasificador de imágenes. Para terminar de definir las diferencias entre las categorías de caras y lugares se realizó el test de diferenciación estadístico entre estas dos clases que se vio en la Figura 3.7, cuyo resultado fue coherente con los anteriores y se encontró coincidencia entre los sujetos en la diferenciación de las clases alrededor de los 300 ms. Según la literatura [15], la categoría animales también podría haberse distinguido por su similaridad. Una posible explicación de no haber encontrado este resultado es que en el trabajo de referencia [15] la mayoría de los animales usados fueron mamíferos (75 %), mientras que en este trabajo los mamíferos fueron un 50 %.

En resumen, se vio que la métrica de similaridad a partir de la correlación espacial entre ensayos tiempo a tiempo como implementación de RSA sirvió para ver que las categorías que muestran más similaridad, estable en el tiempo y consistente entre sujetos, fueron las caras y lugares. De modo que se eligieron para ser utilizadas en un experimento subsecuente en el Laboratorio de Neurociencias de la Memoria.

El problema de distinguir la información sobre las categorías semánticas en las señales se continuó estudiando mediante el uso de clasificadores con aprendizaje supervisado. Se utilizaron dos clasificadores lineales, es decir, que aplican una transformación lineal al vector de características. En línea con el enfoque de RSA, el vector de características fue el vector de amplitudes de todos los sensores en una muestra temporal. Se buscó clasificar la señal para cada tiempo a partir de esta información. Cabe destacar que la clasificación entre los ensayos de caras y lugares fue acertada por encima del umbral de significancia para todos los sujetos. En todos los casos, este umbral se encontró en un valor de exactitud alrededor del 55 %, levemente por encima del azar ideal de 50 %. Además, las ventanas temporales de exactitud significativa coinciden con la bibliografía [15, 25]. Ambos algoritmos de clasificación implementados tuvieron un desempeño equivalente con la diferencia de que SVM fue superior en términos de los tiempos de cómputo. Haciendo uso de la matriz de confusión, se determinó que ninguna categoría predominó sobre otra (Ver Figura 3.11).

En la Figura 3.8, se puede ver que en tres participantes se alcanzó casi un 90 % de exactitud en tiempos alrededor de los 150 ms post-estímulo. Es notable que este grado de exactitud corresponde a una clasificación que utiliza únicamente los voltajes de los sensores en un solo tiempo, y realiza una separación lineal de las clases. Además, ese tiempo de máxima exactitud es consistente entre los sujetos. La máxima exactitud en relación al umbral para todos los sujetos se dio en la ventana de entre 160 ms y 190 ms, a excepción por un sujeto en los 130 ms. Esta localización los picos de exactitud se ve claramente en la Figura 3.10. En esta figura también se muestran en línea negra los tiempos en los que el análisis de RSA indicó distintividad entre las categorías. Se vio que las ventanas significativas de estos dos métodos en su mayoría no coincidieron. En donde la clasificación fue máxima, no hubo diferencias de similaridad. Esto puede indicar que durante la ventana de 150-200 ms en donde predomina la actividad occipital, el clasificador puede discernir las

categorías mientras que RSA no distingue las representaciones. Por lo que podría ser que cuando termina la activación de bajo nivel por la visualización de una imagen, RSA comienza a distinguir las representaciones.

Es interesante notar que en los sujetos 1, 2 y 5 las curvas de similaridad no tuvieron una estructura marcada (Figura 3.3) relacionada con los ERPs de visualización, y por otro lado estos sujetos fueron los que presentaron menor AUC en la clasificación (Figura 3.8). Adicionalmente, la mayoría de los sujetos mostraron estructuras de similaridad semejantes a las de exactitud, observando la cantidad de picos y su localización temporal. En cuanto a esto último, cabe notar que RSA tuvo una pequeña postergación temporal respecto a la clasificación, que ya se había señalado también respecto a los ERPs.

Se realizó una breve exploración de los coeficientes para los sensores obtenidos por el clasificador SVM. Como se describió en la introducción, estos coeficientes definen la normal al hiperplano que separa las clases. Observando el ejemplo de la Figura 1.3, se puede ver que una menor componente el eje horizontal en la normal del hiperplano, implica un plano que tiende a la dirección horizontal. De esta manera el valor de la característica en ese eje, pierde relevancia a la hora de separar las clases. Esto mismo se puede pensar con los coeficientes que acompañan a cada sensor, en un espacio de 30 dimensiones. Como se realizó una clasificación para cada tiempo, se obtuvieron coeficientes para cada tiempo. Por lo que para explorar estos coeficientes se tomó el promedio en una ventana de tiempo particular. Esta ventana de tiempo se eligió tal que la clasificación haya sido significativamente correcta para los 7 participantes. Se encontraron dos ventanas cumpliendo esta condición; (a) [128 ms-324 ms] y (b) [337 ms-394 ms], donde la primera es mucho más larga que la segunda; y por la separación entre las mismas (13 ms) podría considerarse directamente como la misma ventana. Sin embargo se conservó la separación para observar si se presentan diferencias en los coeficientes. Cuando se vio la distribución de los valores de estos coeficientes, separados por sujeto, se vio que varios sensores se destacaban. Se pueden tomar varias formas de determinar cuáles, como los que tuvieron un valor más alto que el 95 % de los datos, o aquellos que se consideraron estadísticamente atípicos (según el gráfico de cajas que usa el rango intercuartil). O simplemente identificándolos cualitativamente en la Figura 3.12. Lo que se destaca de esta figura es que aunque los sensores destacados no coincidieron exactamente entre todos los sujetos, se pudo apreciar que corresponden a la misma región del cerebro, principalmente occipital. La predominancia que se vio sobre el hemisferio derecho coincide con experimentos previos [48–50] aunque existen otros resultados en diferentes direcciones dejando inconclusa esta cuestión (Revisión [10]).

Teniendo en cuenta los resultados, se pone en discusión la propuesta de algunos trabajos sobre la clasificación como medida de disimilaridad (complemento de la similaridad) [25], o al menos para esta tesis. Ante todo, existen diferencias en la metodología, RSA tiene en cuenta todos los sensores en todo momento con igual peso, mientras que los modelos de los clasificadores les asignan pesos a los sensores. De esta manera, al momento de la predicción el clasificador puede anular parcialmente información que estima que no es relevante para discernir las categorías. Dado que el clasificador priorizó la región occipital y el momento donde se da el reconocimiento básico de bajo nivel, esto podría indicar que se basó en características como las formas de las imágenes y no tanto en un procesamiento neural de reconocimiento semántico. Por lo tanto, estos resultados se consideran preliminares y se proponen diversos estudios. Por ejemplo, queda pendiente estudiar si utilizando los sensores occipitales exclusivamente, la clasificación pierde o mantiene exactitud, lo que indicaría el nivel de irrelevancia de otros sensores. También queda evaluar el clasificador con las otras categorías para determinar si los lugares fueron efectivamente discriminados por cuestiones de forma de la

imagen. Para ello también se propone clasificar las categorías a partir de los estímulos, para estudiar el grado de diferencia del estímulo en sí mismo con respecto al de su representación neural. Principalmente sería de utilidad incorporar registros de una cantidad mayor de participantes para estudiar la extrapolación de estos resultados preliminares.

En cuanto a la condición de elaboración, se vio desde el comienzo del trabajo que la representación neural no fue tan fuerte y marcada como en un proceso de visualización de una imagen. A pesar de ello, nos pareció interesante estudiar la relación entre estas dos condiciones, particularmente, estudiar si la decodificación del reconocimiento de un objeto podía ayudar a decodificar las categorías de los estímulos en un proceso de imaginación. Esta idea se implementó utilizando el modelo dado por SVM entrenado con los ensayos de visualización, para predecir la categoría de la señal de elaboración. Es decir, determinar si de alguna forma, la representación de una categoría al momento de reconocer el objeto, guarda codificaciones que se relacionen con este otro proceso, el de imaginación. En la Figura 3.13 para los sujetos 3, 4 y 6 se vio estabilidad temporal en la exactitud significativa. En los sujetos 1 y 2 el desempeño en la clasificación de elaboración fue inferior con respecto a los otros participantes. De manera que los resultados en la clasificación de elaboración por sujeto se relacionaron con el desempeño del clasificador en la condición de la visualización. Para este análisis se contó con pocos ensayos por sujeto y a su vez pocos sujetos, lo que dificultó la validación de las predicciones. A pesar de ello se obtuvieron resultados prometedores por la consistencia entre los resultados con la visualización. Continuar con esta línea de investigación resulta de interés puesto que puede ayudar a determinar si los clasificadores logran captar efectivamente procesos neurales de mayor nivel. En este caso se implementaron sobre procesos de imaginación pero se podrían aplicar en experimentos de memoria, en donde además se podría usar información sobre el espacio de frecuencias de la actividad neural [30].

Finalmente, se puede decir que en esta tesis se desarrolló e implementó un experimento de electroencefalografía. Los análisis sobre estos registros arrojaron resultados preliminares, y se propusieron otra serie de estudios consecuentes con los ya desarrollados. Las técnicas utilizadas para estos análisis pueden implementarse en experimentos futuros del laboratorio, para vincular los estudios comportamentales de memoria con la actividad neural involucrada.

Bibliography

- [1] J. J. DiCarlo, D. Zoccolan, and N. C. Rust, “How does the brain solve visual object recognition?” *Neuron*, vol. 73, no. 3, pp. 415–434, 2012.
- [2] E. K. Warrington and R. McCarthy, “Category specific access dysphasia,” *Brain*, vol. 106, no. 4, pp. 859–878, 1983.
- [3] E. K. Warrington and T. Shallice, “Category specific semantic impairments,” *Brain*, vol. 107, no. 3, pp. 829–853, 1984.
- [4] E. K. Warrington and R. A. McCarthy, “Categories of knowledge: Further fractionations and an attempted integration,” *Brain*, vol. 110, no. 5, pp. 1273–1296, 1987.
- [5] A. Caramazza and J. R. Shelton, “Domain-specific knowledge systems in the brain: The animate-inanimate distinction,” *Journal of cognitive neuroscience*, vol. 10, no. 1, pp. 1–34, 1998.
- [6] A. Caramazza and B. Z. Mahon, “The organization of conceptual knowledge: the evidence from category-specific semantic deficits,” *Trends in cognitive sciences*, vol. 7, no. 8, pp. 354–361, 2003.
- [7] B. Z. Mahon and A. Caramazza, “Concepts and categories: A cognitive neuropsychological perspective,” *Annual review of psychology*, vol. 60, pp. 27–51, 2009.
- [8] I. Simanova, M. Van Gerven, R. Oostenveld, and P. Hagoort, “Identifying object categories from event-related eeg: toward decoding of conceptual representations,” *PloS one*, vol. 5, no. 12, p. e14465, 2010.
- [9] E. Capitani, M. Laiacona, B. Mahon, and A. Caramazza, “What are the facts of semantic category-specific deficits? a critical review of the clinical evidence,” *Cognitive Neuropsychology*, vol. 20, no. 3-6, pp. 213–261, 2003.
- [10] C. Gerlach, “A review of functional imaging studies on category specificity,” *Journal of Cognitive Neuroscience*, vol. 19, no. 2, pp. 296–314, 2007.
- [11] L. Tyler, H. Moss, M. Durrant-Peatfield, and J. Levy, “Conceptual structure and the structure of concepts: A distributed account of category-specific deficits,” *Brain and Language*, vol. 75, no. 2, pp. 195–231, 2000. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0093934X00923538>

- [12] C. Wang, X. Hu, L. Yao, S. Xiong, and J. Zhang, "Spatio-temporal pattern analysis of single-trial eeg signals recorded during visual object recognition," *Science China Information Sciences*, vol. 54, pp. 2499–2507, 2011.
- [13] S. J. Luck, "Event-related potentials." 2012.
- [14] T. Carlson, D. A. Tovar, A. Alink, and N. Kriegeskorte, "Representational dynamics of object vision: the first 1000 ms," *Journal of vision*, vol. 13, no. 10, pp. 1–1, 2013.
- [15] B. Kaneshiro, M. Perreau Guimaraes, H.-S. Kim, A. M. Norcia, and P. Suppes, "A representational similarity analysis of the dynamics of object processing using single-trial eeg classification," *Plos one*, vol. 10, no. 8, p. e0135697, 2015.
- [16] F. Pulvermüller, W. Lutzenberger, and H. Preissl, "Nouns and verbs in the intact brain: Evidence from event-related potentials and high-frequency cortical responses," *Cerebral cortex*, vol. 9, no. 5, pp. 497–506, 1999.
- [17] M. Kiefer, "Perceptual and semantic sources of category-specific effects: Event-related potentials during picture and word categorization," *Memory & Cognition*, vol. 29, pp. 100–116, 2001.
- [18] V. R. Sommer, L. Mount, S. Weigelt, M. Werkle-Bergner, and M. C. Sander, "Spectral pattern similarity analysis: Tutorial and application in developmental cognitive neuroscience," *Developmental cognitive neuroscience*, vol. 54, p. 101071, 2022.
- [19] N. Kriegeskorte, M. Mur, and P. A. Bandettini, "Representational similarity analysis-connecting the branches of systems neuroscience," *Frontiers in systems neuroscience*, vol. 2, p. 249, 2008.
- [20] J. D. Koen and M. D. Rugg, "Neural dedifferentiation in the aging brain," *Trends in cognitive sciences*, vol. 23, no. 7, pp. 547–559, 2019.
- [21] J. D. Koen, S. Srokova, and M. D. Rugg, "Age-related neural dedifferentiation and cognition," *Current Opinion in Behavioral Sciences*, vol. 32, pp. 7–14, 2020.
- [22] B. A. Kuhl, J. Rissman, and A. D. Wagner, "Multi-voxel patterns of visual category representation during episodic encoding are predictive of subsequent memory," *Neuropsychologia*, vol. 50, no. 4, pp. 458–469, 2012, multivoxel pattern analysis and cognitive theories. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0028393211004088>
- [23] G. Xue, "The neural representations underlying human episodic memory," *Trends in cognitive sciences*, vol. 22, no. 6, pp. 544–561, 2018.
- [24] V. R. Sommer, Y. Fandakova, T. H. Grandy, Y. L. Shing, M. Werkle-Bergner, and M. C. Sander, "Neural pattern similarity differentially relates to memory performance in younger and older adults," *Journal of Neuroscience*, vol. 39, no. 41, pp. 8089–8099, 2019.
- [25] R. M. Cichy and D. Pantazis, "Multivariate pattern analysis of meg and eeg: A comparison of representational structure in time and space," *NeuroImage*, vol. 158, pp. 441–454, 2017.

- [26] B. T. John D. Kelleher, *Data Science*, ser. The MIT Press Essential Knowledge. The MIT Press, 2018. [Online]. Available: <http://gen.lib.rus.ec/book/index.php?md5=eeb7eae8b32aec6286f683951f259e1>
- [27] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed., ser. Information science and statistics. Springer, 2006. [Online]. Available: <http://gen.lib.rus.ec/book/index.php?md5=6b552b24cae380bb656f7aaef7f81b46>
- [28] L. Rosasco, E. De Vito, A. Caponnetto, M. Piana, and A. Verri, “Are loss functions all the same?” *Neural computation*, vol. 16, no. 5, pp. 1063–1076, 2004.
- [29] B. Murphy, M. Poesio, F. Bovolo, L. Bruzzone, M. Dalponte, and H. Lakany, “Eeg decoding of semantic category reveals distributed representations for single concepts,” *Brain and language*, vol. 117, no. 1, pp. 12–22, 2011.
- [30] C. Kerrén, J. Linde-Domingo, S. Hanslmayr, and M. Wimber, “An optimal oscillatory phase for pattern reactivation during memory retrieval,” *Current Biology*, vol. 28, no. 21, pp. 3383–3392, 2018.
- [31] C. Wang, S. Xiong, X. Hu, L. Yao, and J. Zhang, “Combining features from erp components in single-trial eeg for discriminating four-category visual objects,” *Journal of neural engineering*, vol. 9, no. 5, p. 056013, 2012.
- [32] P. Shenoy and D. S. Tan, “Human-aided computing: utilizing implicit human processing to classify images,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2008, pp. 845–854.
- [33] A. Tzovara, M. M. Murray, G. Plomp, M. H. Herzog, C. M. Michel, and M. De Lucia, “Decoding stimulus-related information from single-trial eeg responses based on voltage topographies,” *Pattern Recognition*, vol. 45, no. 6, pp. 2109–2122, 2012.
- [34] G. A. Rousselet, J. S. Husk, P. J. Bennett, and A. B. Sekuler, “Single-trial eeg dynamics of object and face visual processing,” *Neuroimage*, vol. 36, no. 3, pp. 843–862, 2007.
- [35] Pooja, S. Pahuja, and K. Veer, “Recent approaches on classification and feature extraction of eeg signal: A review,” *Robotica*, vol. 40, no. 1, p. 77–101, 2022.
- [36] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, “Places: A 10 million image database for scene recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.
- [37] N. C. Ebner, M. Riediger, and U. Lindenberger, “Faces—a database of facial expressions in young, middle-aged, and older women and men: Development and validation,” *Behavior research methods*, vol. 42, pp. 351–362, 2010.
- [38] M. B. Brodeur, K. Guérard, and M. Bouras, “Bank of standardized stimuli (boss) phase ii: 930 new normative photos,” *PLOS ONE*, vol. 9, no. 9, pp. 1–10, 09 2014. [Online]. Available: <https://doi.org/10.1371/journal.pone.0106953>

- [39] J. Peirce, J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. K. Lindeløv, “Psychopy2: Experiments in behavior made easy,” *Behavior research methods*, vol. 51, pp. 195–203, 2019.
- [40] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. S. Hämäläinen, “MEG and EEG data analysis with MNE-Python,” *Frontiers in Neuroscience*, vol. 7, no. 267, pp. 1–13, 2013.
- [41] F. Artoni, A. Delorme, and S. Makeig, “Applying dimension reduction to eeg data by principal component analysis reduces the quality of its subsequent independent component decomposition,” *NeuroImage*, vol. 175, pp. 176–187, 2018.
- [42] M. Jas, D. A. Engemann, Y. Bekhti, F. Raimondo, and A. Gramfort, “Autoreject: Automated artifact rejection for meg and eeg data,” *NeuroImage*, vol. 159, pp. 417–429, 2017.
- [43] L. Wang, E. Wlotko, E. Alexander, L. Schoot, M. Kim, L. Warnke, and G. R. Kuperberg, “Neural evidence for the prediction of animacy features during language comprehension: Evidence from meg and eeg representational similarity analysis,” *Journal of Neuroscience*, vol. 40, no. 16, pp. 3278–3291, 2020.
- [44] E. Maris and R. Oostenveld, “Nonparametric statistical testing of eeg- and meg-data,” *Journal of Neuroscience Methods*, vol. 164, no. 1, pp. 177–190, 2007. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0165027007001707>
- [45] N. Kriegeskorte and R. A. Kievit, “Representational geometry: integrating cognition, computation, and the brain,” *Trends in cognitive sciences*, vol. 17, no. 8, pp. 401–412, 2013.
- [46] A. Dietrich and R. Kanso, “A review of eeg, erp, and neuroimaging studies of creativity and insight,” *Psychological bulletin*, vol. 136, no. 5, p. 822, 2010.
- [47] S. Makeig, M. Westerfield, T.-P. Jung, S. Enghoff, J. Townsend, E. Courchesne, and T. J. Sejnowski, “Dynamic brain sources of visual evoked responses,” *Science*, vol. 295, no. 5555, pp. 690–694, 2002.
- [48] C. J. Moore and C. J. Price, “A functional neuroimaging study of the variables that generate category-specific object processing differences,” *Brain*, vol. 122, no. 5, pp. 943–962, 1999.
- [49] H. E. Moss, J. M. Rodd, E. A. Stamatakis, P. Bright, and L. K. Tyler, “Anteromedial temporal cortex supports fine-grained differentiation among objects,” *Cerebral cortex*, vol. 15, no. 5, pp. 616–627, 2005.
- [50] H. Damasio, D. Tranel, T. Grabowski, R. Adolphs, and A. Damasio, “Neural systems behind word and concept retrieval,” *Cognition*, vol. 92, no. 1-2, pp. 179–229, 2004.
- [51] J. Stelzer, Y. Chen, and R. Turner, “Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (mvpa): random permutations and cluster size control,” *Neuroimage*, vol. 65, pp. 69–82, 2013.

Capítulo 5

Apéndices

APÉNDICE A.

Caras	Animales	Fruver	Musicales
MUJER 1	PERRO 1	FRUTOS ROJOS	FLAUTA
MUJER 2	PERRO 2	PERA	TROMPETA
MUJER 3	PÁJARO 1	BANANA	GITARRA 1
MUJER 4	PÁJARO 2	ZANAHORIA	GITARRA 2
MUJER 5	PAVO REAL	PALTA	GITARRA 3
MUJER 6	TUCÁN	BRÓCOLI	GITARRA 4
MUJER 7	FLAMENCO	CEBOLLA	MICRÓFONO 1
MUJER 8	HORMIGA	MORRÓN 1	MICRÓFONO 2
MUJER 9	LECHUZA	MORRÓN 2	VIOLONCHELO
MUJER 10	DELFÍN	PEPINILLO	CLARINETE
MUJER 11	OSO 1	SANDÍA	CLAVES DE MADERA
MUJER 12	OSO 2	PAPA	CONGA
MUJER 13	MAPACHE	KIWI	SAXO
MUJER 14	ELEFANTE	LECHUGA	FUNDA DE GITARRA
MUJER 15	CEBRA	UVA	VIOLÍN
MUJER 16	LANGOSTA	NARANJA	XILÓFONO
MUJER 17	GALLINA	LIMÓN	BALALAICA
MUJER 18	JIRafa	ANANÁ	BANJO
MUJER 19	CONEJO	AJÍ	BAQUETAS
MUJER 20	TARÁNTULA	TOMATE	BATERIA 1
HOMBRE 1	VAQUITA DE SAN ANTONIO	COLIFLOR	BATERIA 2
HOMBRE 2	MARIPOSA	REMOLACHA	GONG
HOMBRE 3	GATO	CIRUELA	PIANO 1
HOMBRE 4	CAMALEÓN	REMOLACHA	PIANO 2

Caras	Animales	Fruver	Musicales
HOMBRE 5	ARDILLA	PAPAYA	ARMÓNICA
HOMBRE 6	SERPIENTE	GRANADA	TROMBÓN
HOMBRE 7	TIBURÓN	BERENJENA	TECLADO
HOMBRE 8	VACA	CALABAZA	MARACAS
HOMBRE 9	PEZ	AJO	PARTITURA
HOMBRE 10	SAPO	FRUTILLA	VINILO
HOMBRE 12	TIGRE	LIMA	PANDERETA 1
HOMBRE 13	TORTUGA	MANGO	PANDERETA 2
HOMBRE 14	CERDO	ARÁNDANOS	TRIANGULO
HOMBRE 15	HIPOPÓTAMO	RABANITOS	PLATILLO 1
HOMBRE 16	CABRA	MANZANA 1	PLATILLO 2
HOMBRE 17	COCODRILO	MANZANA 2	ACORDEÓN
HOMBRE 18	CAMELLO	PUERRO	MEZCLADOR DE AUDIO
HOMBRE 19	ZORRINO	CEREZA	CONTRABAJO
HOMBRE 20	RINOCERONTE	CHOCLO	TAMBOR

Cocina	Lugares	Vehículos	Ropa
CUCHILLO 1	PRADERA	MONOPATÍN	CINTURÓN
CUCHILLO 2	RÍO 1	AUTO 1	BOTAS 1
CUCHILLO 3	RÍO 2	AUTO 2	BOTAS 2
CAFETERA	CASAS	TANQUE DE GUERRA	BOTAS 3
PAVA ELÉCTRICA	MONTAÑA NEVADA	AUTO 4	VISERA 1
DESTAPADOR	DUNAS	AUTO 5	VISERA 2
JARRA	DIQUE	APLANADORA	CARTERA
SECAPLATOS	LAGO 1	CAMIÓN 1	OJOTA
DETERGENTE	LAGO 2	CAMIÓN 2	GORRO 1
CACEROLA	ACANTILADO	CAMIÓN 3	GORRO 2
PALO DE AMASAR	PUERTO	CAMIÓN 4	GORRO 3
SECADOR DE VERDURAS	JARDÍN	CAMIÓN 5	GORRO 4
SALERO	EXTERIOR VERDE 1	CAMIÓN 6	SOMBRERO 1
PIMENTERO	EXTERIOR VERDE 2	ACOPLADO	SOMBRERO 2
COCINA	ARBOLADO	MONTACARGA	ZAPATILLAS 1
COLADOR 1	GLACIARES	MICRO 1	ZAPATILLAS 2
COLADOR 2	FORMACIÓN ROCOSA 1	MICRO 2	ZAPATILLAS 3
ESPÁTULA	FORMACIÓN ROCOSA 2	MICRO 3	TACO
TENEDOR	PLAYA	CAMIÓN 1	SANDALIA
SOPORTE DE CONDIMENTOS	DESIERTO	CAMIÓN 2	SOQUETE 1
ESPONJA	SUPERMERCADO	CAMIÓN 3	SOQUETE 2

Cocina	Lugares	Vehículos	Ropa
TAZA	COCINA 1	CAMIÓN 4	LENTES
LICUADORA	COCINA 2	CAMIÓN 5	REMERA 1
HELADERA	COCINA 3	CAMIÓN 6	REMERA 2
VASO TÉRMICO	OFICINA	TRICICLO	REMERA 3
TUPPER	TEATRO	KAYAK	BUFANDA
COPA	BIBLIOTECA	HELICÓPTERO 1	JUMPER
HORNALLA	CINE	HELICÓPTERO 2	CAMISA 1
PINZA PARA FIDEOS	SALA DE ESTAR 1	CARRITO DE GOLF	CAMISA 2
TOSTADORA 1	SALA DE ESTAR 2	TREN	CORBATA 1
TOSTADORA 2	PELUQUERÍA 1	TOPADORA	CORBATA 2
BATIDOR 1	PELUQUERÍA 2	BICICLETA	PANTALÓN 1
BATIDOR 2	RESTAURANTE 1	MOTO 1	PANTALÓN 2
BOWL	RESTAURANTE 2	MOTO 2	BOTINES
CUBETERA	CUARTO 1	TRICICLO DE GASOLINA	BERMUDAS 1
SARTÉN	CUARTO 2	JEEP	BERMUDAS 2
RALLADOR	COMEDOR 1	MOTO DE AGUA	BERMUDAS 3
PELA PAPA	COMEDOR 2	BARCO 1	BERMUDAS 4
PLATO 1	COMEDOR 3	BARCO 2	SOMBRERO 1
PLATO 2	GIMNASIO	BARCO 3	SOMBRERO 2

Tabla 5.1: Tablas de la descripción de las imágenes utilizadas como estímulos para bloque experimental de visualización.

APÉNDICE B.

Caras	Animales	Fruver	Musicales
UNA MUJER PELIRROJA	UN PERRO	UNA MANZANA	UNA FLAUTA
UNA JOVEN RUBIA	UN GATO	UNA BANANA	UNA GUITARRA
UNA SEÑORA MAYOR	UN ELEFANTE	UN BRÓCOLI	UN PIANO
UN SEÑOR CANOSO	UNA ARAÑA	UNA CEBOLLA	UN TAMBOR
UN JOVEN MOROCHO	UN PÁJARO	UNA LECHUGA	UN MICRÓFONO
UN HOMBRE RUBIO	UNA HORMIGA	UN PEPINO	UNA TROMPETA
Cocina	Lugares	Vehículos	Ropa
UNA CUCHARA	UNA PLAYA	UN AUTO	UNA REMERA
UNA SARTÉN	UN BOSQUE	UN COLECTIVO	UN PANTALÓN
UN TENEDOR	UNA MONTAÑA	UNA BICICLETA	UN JEAN
UN COLADOR	UN AULA	UN CAMIÓN	UNA BUFANDA
UN HORNO	UNA OFICINA	UN BARCO	UNA CAMISA
UN MICROONDAS	UN TEATRO	UN TRACTOR	UN BUZO

Tabla 5.2: Tabla de las palabras utilizadas como estímulos para bloque experimental de elaboración.

APÉNDICE C.

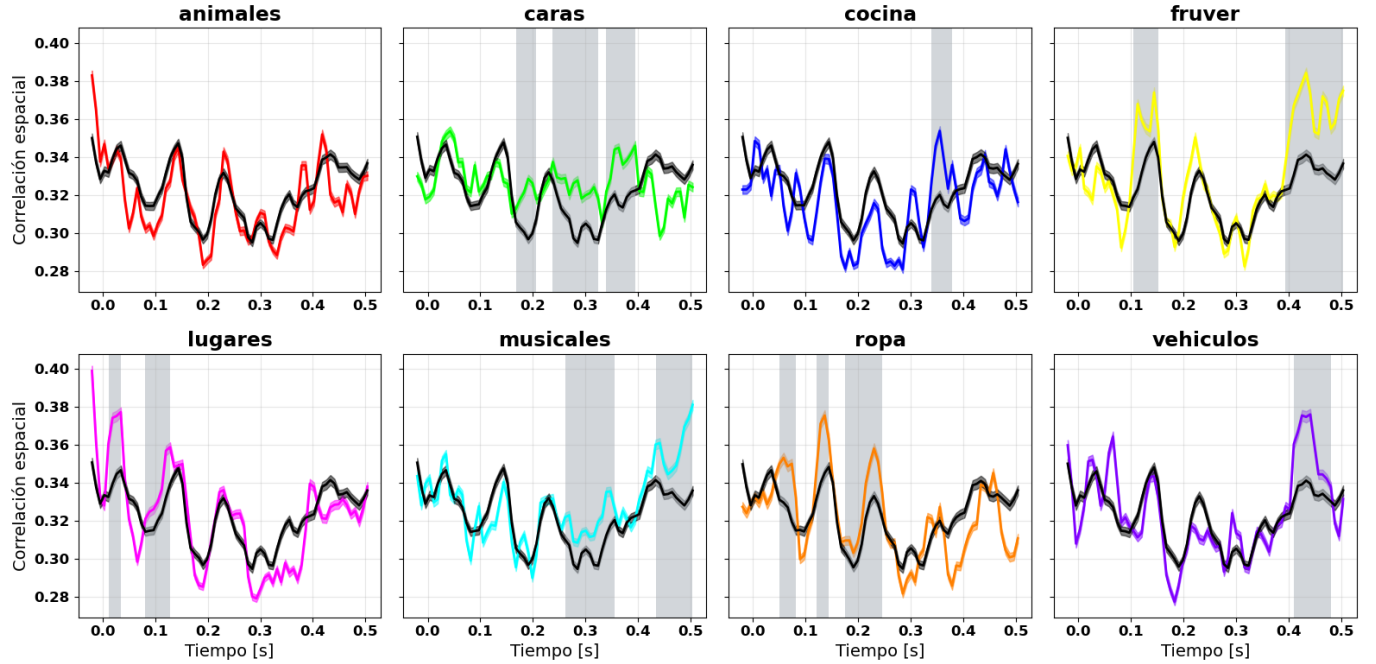


Figura 5.1: Curvas de similaridad intra y entre categorías para el **sujeto 1**. Se marcan en gris las ventanas temporales en donde se encontraron diferencias significativas entre estas dos clases.

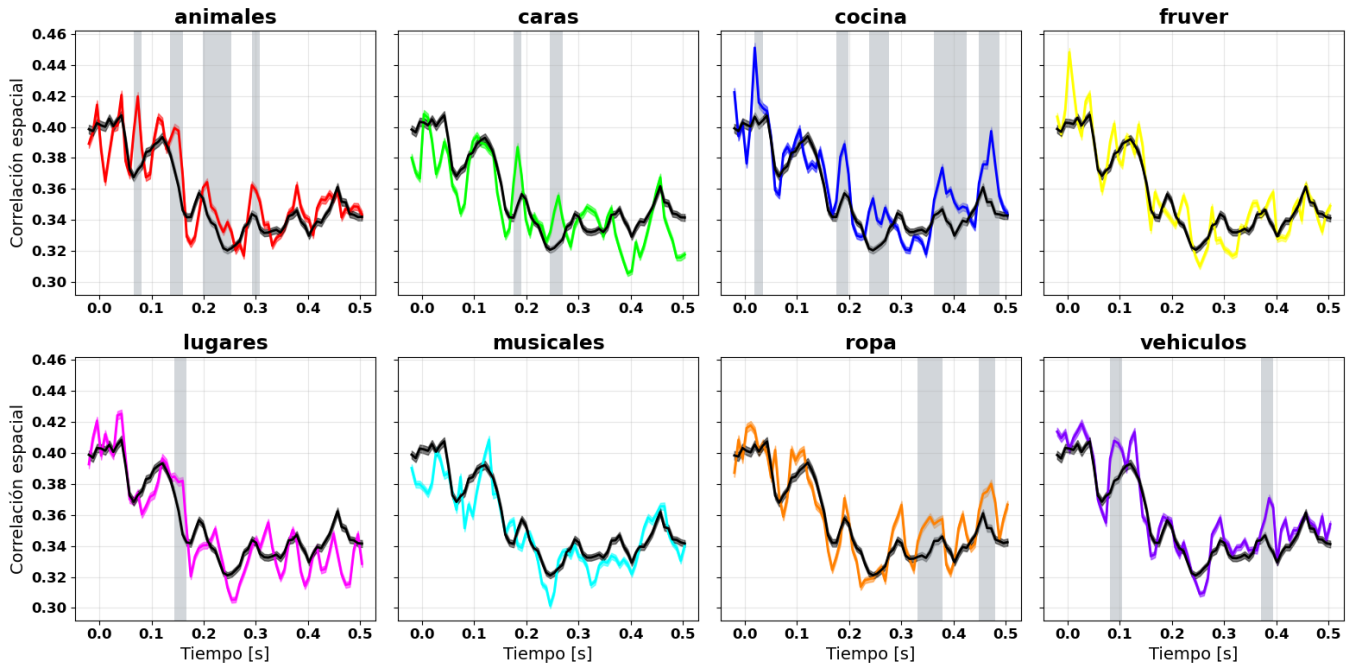


Figura 5.2: Curvas de similaridad intra y entre categorías para el **sujeto 2**. Se marcan en gris las ventanas temporales en donde se encontraron diferencias significativas entre estas dos clases.

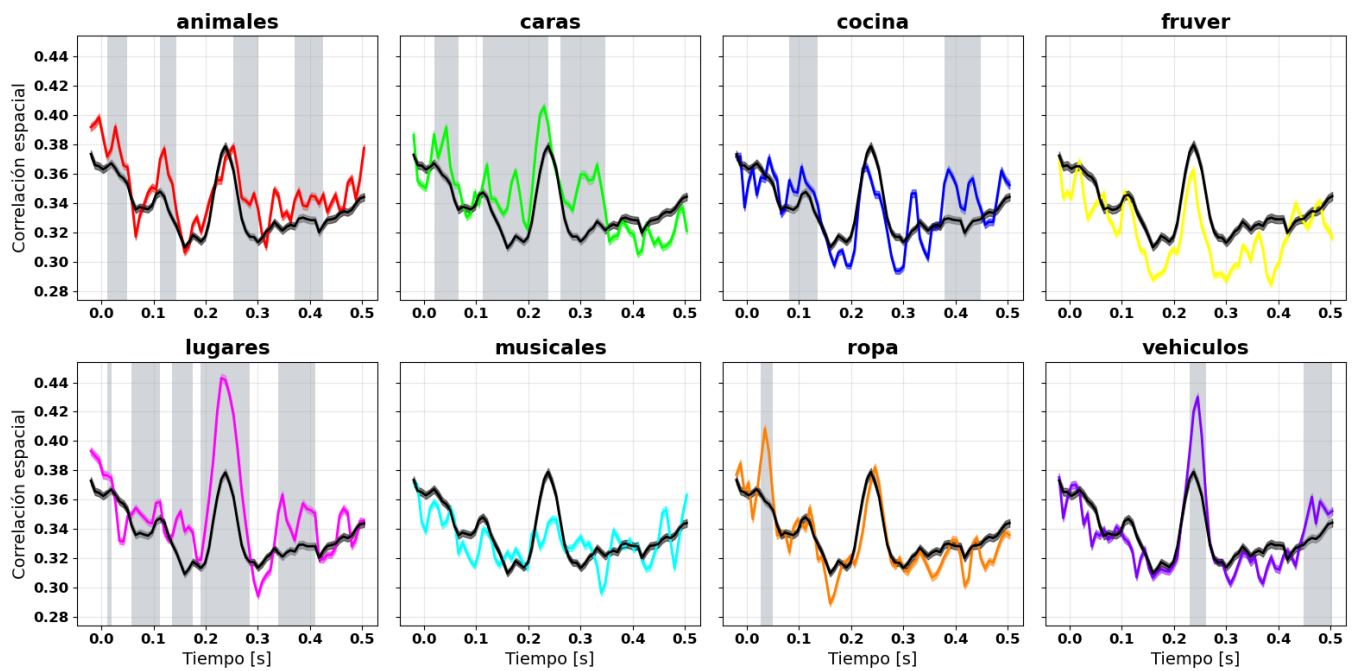


Figura 5.3: Curvas de similitud intra y entre categorías para el **sujeto 3**. Se marcan en gris las ventanas temporales en donde se encontraron diferencias significativas entre estas dos clases.

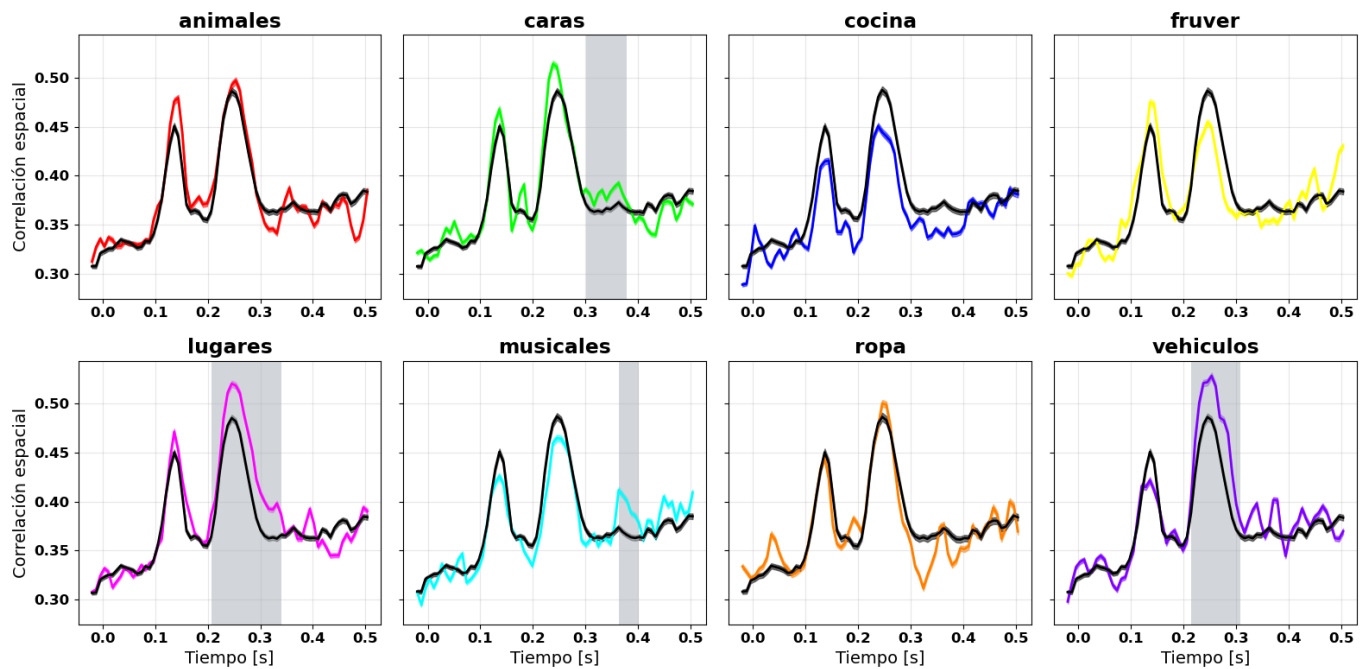


Figura 5.4: Curvas de similitud intra y entre categorías para el **sujeto 4**. Se marcan en gris las ventanas temporales en donde se encontraron diferencias significativas entre estas dos clases.

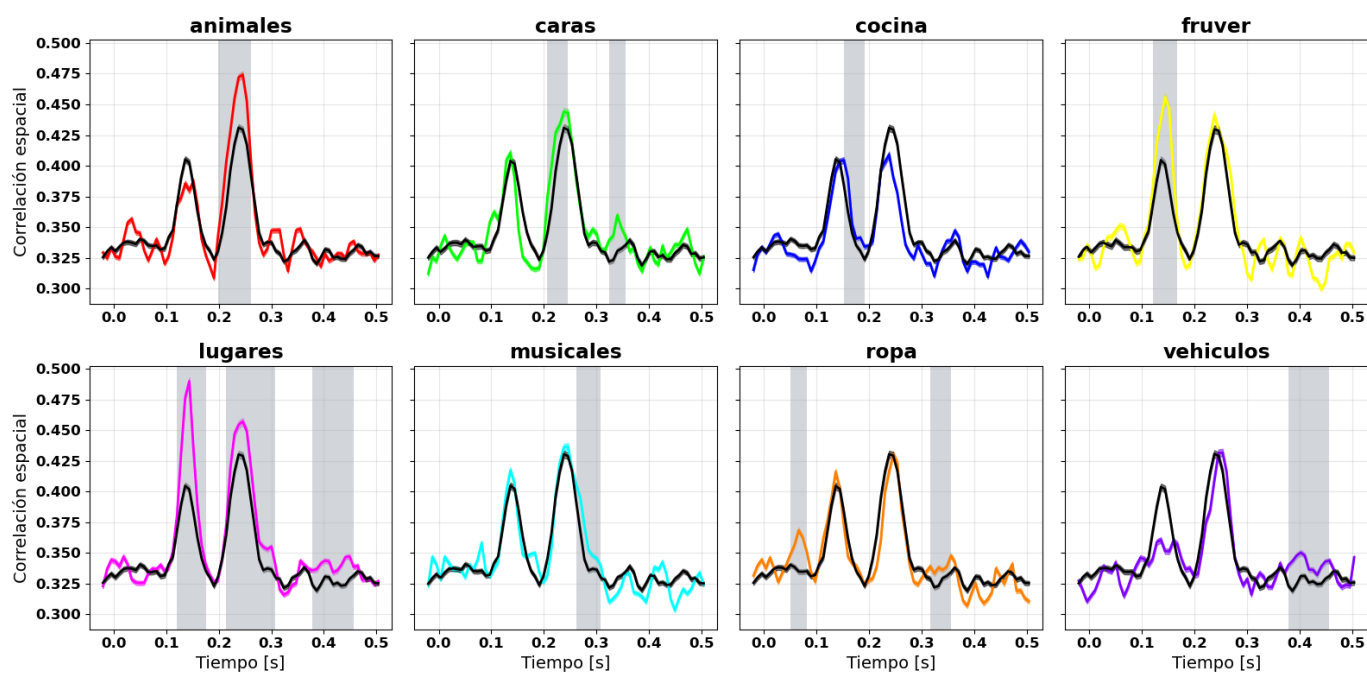


Figura 5.5: Curvas de similaridad intra y entre categorías para el **sujeto 6**. Se marcan en gris las ventanas temporales en donde se encontraron diferencias significativas entre estas dos clases.

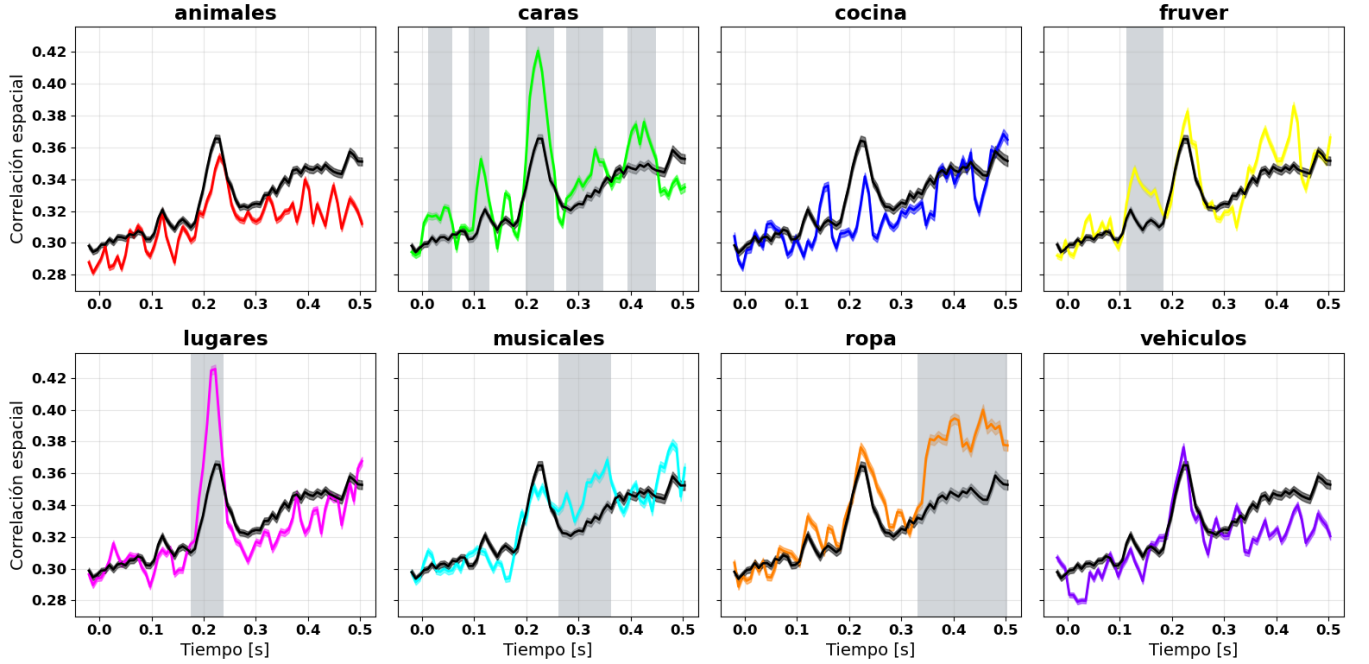


Figura 5.6: Curvas de similitud intra y entre categorías para el **sujeto 7**. Se marcan en gris las ventanas temporales en donde se encontraron diferencias significativas entre estas dos clases.

APÉNDICE D.

Validación de la clasificación

Se hicieron análisis de validación cruzada para determinar la confiabilidad al realizar la clasificación sobre los datos.

Como se describió en la Subsección A la exactitud se propuso calcular la exactitud mediante el LOO-CV (Leave-One-Out Cross Validation). La obtención del modelo podría depender la inicialización de la optimización. Para ver esto se calculó la exactitud mediante LOO-CV repetidas veces sobre un mismo conjunto de datos, particularmente se usaron los ensayos del sujeto 1. Se realizaron 100 repeticiones del LOO-CV¹, donde en cada repetición se obtiene un valor de exactitud de para tiempo del ensayo. Para determinar la variabilidad de los resultados frente a la reiteración, se calculó la desviación estándar para cada conjunto de exactitudes por tiempo. Luego se realizó un histograma con las desviaciones de todos los tiempos que se muestra en la Figura 5.7. Se vio para ambos clasificadores que la variabilidad de la exactitud se encuentra acotada en un rango del orden de 10^{-15} , es decir, a un orden de diferencia con el épsilon de la máquina. En resumen, se vio que con ambos clasificadores fue suficiente realizar una sola iteración de LOO-CV para obtener la exactitud sobre un conjunto de datos.

¹El orden de magnitud de la cantidad de repeticiones, en este caso 100, se eligió arbitrariamente para iniciar el análisis y por los resultados se vio que fue suficiente.

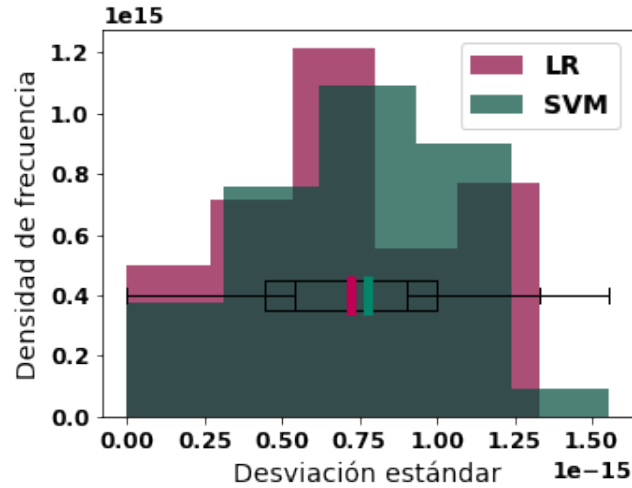


Figura 5.7: Histograma de desviaciones estándar de la exactitud en distintos momentos temporales. Para analizar la variabilidad en la exactitud de los clasificadores a lo largo de múltiples iteraciones del LOO-CV, se calcularon las curvas temporales de exactitud en 100 repeticiones. Luego, se determinó la desviación estándar de la exactitud en cada momento temporal y se representó gráficamente mediante un histograma. (LR: Logistic Regresion, SVM: Support Vector Machines).

Por otra parte, para la implementación del modelo nulo descrito en la Subsección A se introdujo una inspección sobre la cantidad de permutaciones mínima necesaria para obtener una convergencia del modelo nulo [51]. En otras palabras, para garantizar un modelo nulo robusto y optimizar el tiempo de cómputo. Para ver esta convergencia, se realizaron clasificaciones sobre los datos con permutación de etiquetas. En cada permutación se realizó una iteración de LOO-CV, obteniéndose un valor de exactitud para cada iteración. Luego, en la ?? se graficaron las distribuciones de estos valores de exactitud variando la cantidad de permutaciones. Con un punto se muestra el percentil 95 % de las distribuciones, que se quiere usar como umbral de significancia. y se observó que a partir de 100 permutaciones esta métrica converge.

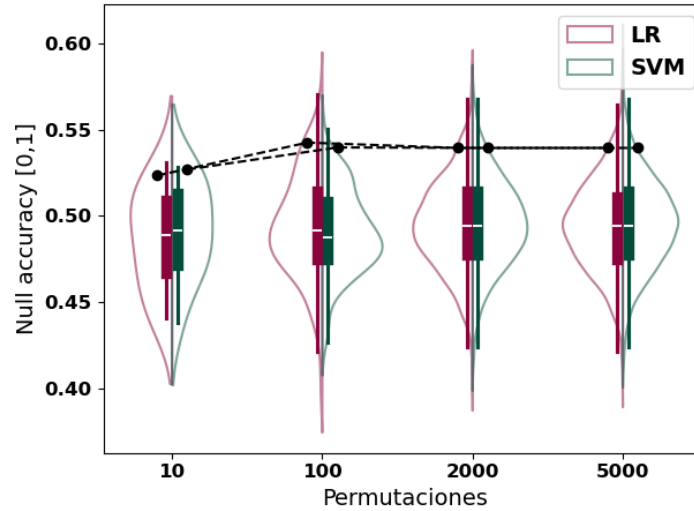


Figura 5.8: Distribuciones de la exactitud del modelo nulo según la cantidad de permutaciones realizadas. Los puntos negros denotan el percentil 95 % que se utilizó como umbral de significancia. (LR: Logistic Regression, SVM: Support Vector Machines).

Por lo tanto se consideró, para ambos clasificadores, que para obtener el modelo nulo fue suficiente realizar 100 permutaciones de etiquetas, implementando un LOO-CV en cada permutación y finalmente obtener el percentil 95 % como umbral de significancia.

Optimización de los parámetros de la regularización

Como se mencionó en la introducción, en ambos algoritmos de aprendizaje supervisado es importante controlar el sobreajuste del modelo a los datos de entrenamiento. En la regresión logística se añade un término de penalización proporcional a la norma L2 (cuadrados de los coeficientes) de los pesos del modelo. En el caso de SVM, el parámetro de regularización C también influye en la tolerancia hacia errores de clasificación en el conjunto de entrenamiento, es decir, influye en la sensibilidad a los outliers. Un exceso en la magnitud de este parámetro puede conducir a una sobre simplificación del modelo, volviéndose menos capaz de clasificar correctamente. De modo que el valor óptimo de este parámetro se estimó en base al desempeño del clasificador. Para ello se realizó un barrido de C en entre 10^{-4} y 10^4 en nueve pasos equiespaciados en escala logarítmica. Para cada constante se entrenó el clasificador y se validó según el procedimiento detallado en la Sección A, con la excepción de que se redujo el número de permutaciones a 10. En la Figura 5.9 se muestra dicha curva para cada sujeto y constante C , donde en línea punteada (negra) se graficó el máximo umbral de significancia de todas las constantes. De estos datos también se calcularon el pico de exactitud por sujeto y el área bajo la curva significativa, es decir, el área entre la curva de exactitud y su umbral de significancia. Estas métricas se muestran como mapa de color en la Figura 5.10. El mismo análisis se realizó para el algoritmo de Regresión Logística y los resultados fueron similares.

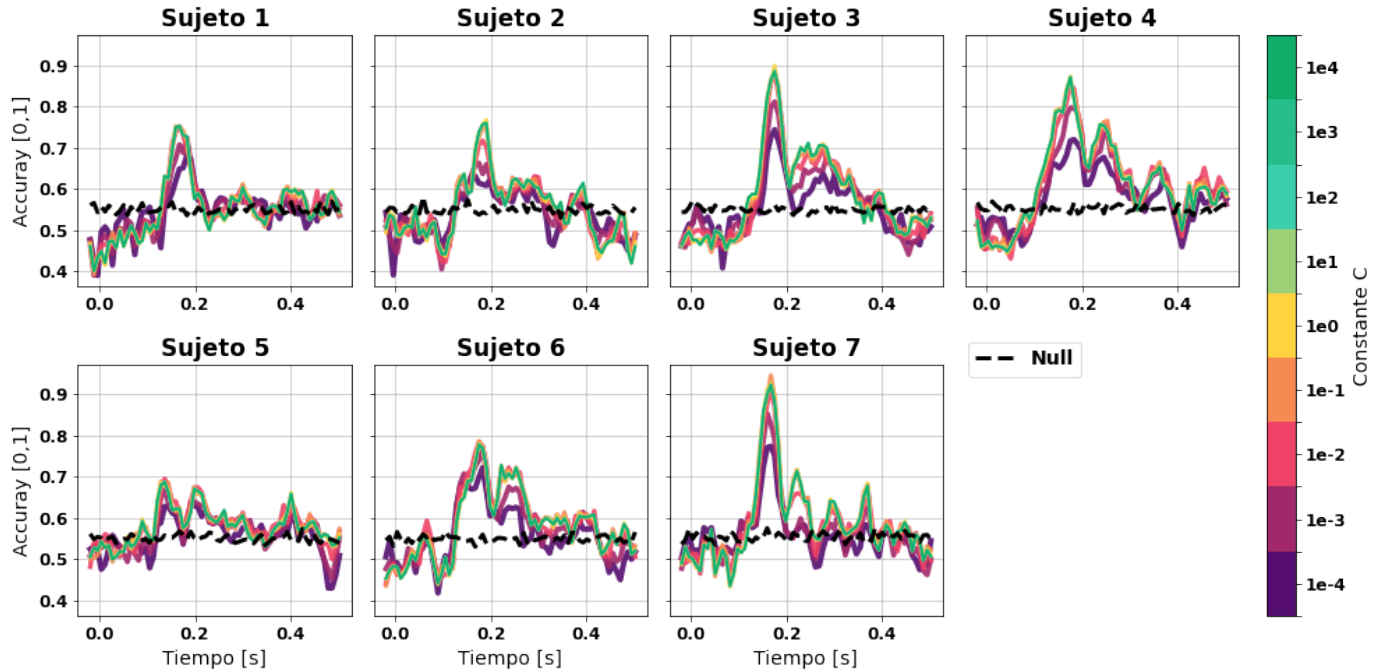


Figura 5.9: (SVM) Curvas de exactitud en función de la constante de regularización C. En línea punteada se tiene el umbral del significancia (obtenido considerando todas las constantes).

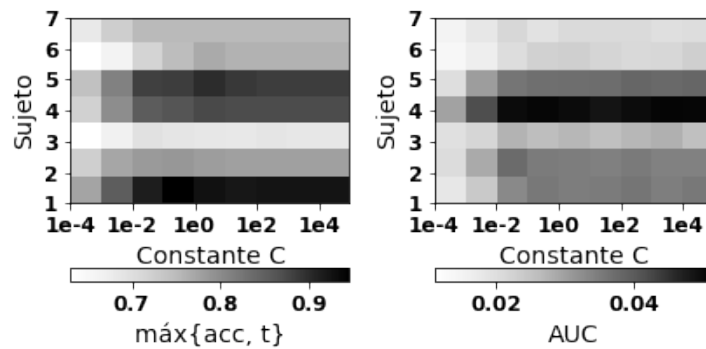


Figura 5.10: (SVM) Máximo valor de exactitud (izquierda) y área significativa bajo la curva (derecha) por sujeto para cada constante de regularización.

Teniendo en cuenta ambas figuras, se observó que el desempeño del clasificador SVM mejoró al aumentar la constante de regularización hasta alcanzar $C = 0,1$ donde se estabiliza. De modo que para los análisis posteriores se tomó este valor en donde la exactitud cambia de comportamiento. Se realizó el mismo análisis para el algoritmo de Regresión Logística y los resultados fueron similares, por lo que también se determinó $C = 0,1$ con el mismo criterio.

APÉNDICE E. Infome comité de ética



Comité de Ética IDIM Alfredo Lanari

Presidente:

Dr. Claudio Capuano (Médico, Master en Bioética y Derecho)

Vicepresidente:

Dra. Graciela Gómez (Médica)

Secretaria de Actas:

Dra. Mónica Costas (Bióloga)

Vocales:

Dra. Patricia Indaburu (Bioquímica)

Dra. María Isabel Zuleta (Antropóloga, Especialista en Bioética)

Dra. Mariana Puche (Abogada)

Dr. Gabriel Palacín (Médico)

Dr. Rodolfo López (Médico)

Dr. Eduardo Del Cerro (Médico)

Asesores externos:

Dra. Marina Khoury (Médica, especialista en Metodología y Epidemiología Clínica)

Dr. Constantino Touloupas (Médico, farmacólogo)

Asistente:

Sra. Ivana Rossetto

Protocolo: Nro 315

Título de la investigación: Fortalecimiento de la memoria por generalización episódica debida al proceso de reconsolidación y sus correlatos neuronales

Documentación presentada: Protocolo, documentos de consentimiento informado, CVs de equipo responsable

Investigadores responsables:

Dra. María Eugenia Pedreira, Dr. Rodrigo S. Fernández, Dra. Luz Bavassi

Lugar: Nuevo edificio del Instituto de Fisiología, Biología Molecular y Neurociencias (IFIByNE), Ciudad Universitaria. Teléfono: 4576-3348 interno 431

Declaración de la decisión tomada:

1.- Protocolo aprobado **X**

2.- Protocolo rechazado

3.- Solicitud de enmienda (no se aprueba hasta una modificación completa del protocolo o el ci)

4.- Protocolo aprobado con recomendación, (alguna recomendación sobre el proceso)

5.- Aprobación con condicionamiento (si falta algún documento que será entregado o se refiere a alguna conducta específica del investigador)

Justificación de la decisión: El protocolo fue evaluado por el CEI. Alcanzó los estándares de aprobación éticos.

Sugerencias del CEI (si corresponde): No corresponde

Responsabilidad	Corresponde	No Corresponde
Confirmar la aceptación de los requisitos del CEI	X	
Presentar oportunamente los informes periódicos y final de la investigación	X	
Presentar futuras enmiendas al protocolo	X	
Presentar futuras enmiendas al material de reclutamiento	X	
Presentar futuras enmiendas a la información para los potenciales participantes en la investigación	X	
Presentar futuras enmiendas al consentimiento informado	X	
Presentar reportes de eventos adversos y de eventos adversos serios e inesperados	X	
Reportar circunstancias no esperadas	X	
Reportar la terminación (programada o anticipada) del estudio	X	
Reportar decisiones significativas tomadas por otros CEIs	X	

Tesis disponible bajo Licencia:

CC BY-NC-SA 4.0 DEED

Creative Commons, Atribución - No Comercial - Compartir Igual 4.0 Internacional

Argentina, Buenos Aires, 2024