



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

EVALUACIÓN DE CARGA COGNITIVA Y ESTADO EMOCIONAL MEDIANTE
SENSORES PSICO-FISIOLÓGICOS EN TAREAS DE REDACCIÓN

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL ELÉCTRICA

ANTONIA LARRAÑAGA RECART

PROFESOR GUÍA:
ÁNGEL JIMÉNEZ MOLINA

MIEMBROS DE LA COMISIÓN:
ANDRÉS CABA RUTTE
FELIPE TOBAR HENRÍQUEZ

Este trabajo ha sido parcialmente financiado por proyecto Fondecyt 11130252

SANTIAGO DE CHILE
2018

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE: INGENIERA CIVIL ELÉCTRICA
POR: ANTONIA LARRAÑAGA RECART
FECHA: 2018
PROFESOR GUÍA: Dr. ÁNGEL JIMÉNEZ MOLINA

EVALUACIÓN DE CARGA COGNITIVA Y ESTADO EMOCIONAL MEDIANTE SENSORES PSICO-FISIOLÓGICOS EN TAREAS DE REDACCIÓN

El presente trabajo tiene como objetivo principal la identificación y medición de niveles de carga cognitiva y del estado emocional en personas que realizan una tarea de redacción, mediante la utilización de sensores psico-fisiológicos. Se utilizan 6 sensores psico-fisiológicos: *eye tracker*, electrocardiograma (ECG), electroencefalograma (EEG), sensor de respuesta electrodermal (GSR), sensor de temperatura y fotopleletismógrafo (PPG). Existe amplia evidencia de que las señales psico-fisiológicas, entregadas por los sensores anteriormente nombrados, pueden medir tanto la carga cognitiva (entendida como el esfuerzo mental al realizar una tarea) como el estado emocional (medido en términos de valencia y excitación) experimentado por un usuario al realizar una tarea en particular.

Esta memoria está enmarcada en el proyecto Fondecyt “*A Cognitive Resource-Aware Mobile Service Framework to Support Human-Computer-Interactions in Ubiquitous Computing Environments*”, liderado por el profesor Ángel Jiménez. El problema a estudiar es de los diferentes niveles de carga cognitiva y distintos estados emocionales que experimenta un usuario mientras escribe un ensayo.

Para esto, se desarrolla un diseño experimental que consiste en que cada usuario de la muestra escriba un ensayo sobre un tema previamente dado, igual para todos. El experimento se realiza a 70 participantes, a los cuales se les miden las señales psico-fisiológicas ya mencionadas. El principal resultado esperado es el de evaluar si los sensores pueden (o no) medir los diferentes niveles de carga cognitiva y estado emocional que experimenta un usuario dado y obtener un clasificador para estos niveles.

Para poder identificar los niveles existentes de carga cognitiva y de valencia y excitación, y poder etiquetarlos, se utilizan algoritmos de aprendizaje no supervisado. En particular, se utiliza *Self-Organizing Maps* (SOM) en conjunto con clustering aglomerativo. Para saber si las señales diferencian los niveles identificados, se realiza un análisis estadístico mediante tests de ANOVA-MR. Luego, para la clasificación de estos niveles, se utiliza k-vecinos más cercanos (KNN), máquinas de soporte vectorial (SVM) y redes neuronales (ANN). Para cada categoría que se busca medir, se encuentra el clasificador que tiene un mejor desempeño promedio para todos los sujetos.

Para carga cognitiva se obtiene un desempeño de 75,55 % de exactitud y 61,14 % de F1 con SVM, para la clasificación de dos niveles. Para excitación se obtiene un 80,5 % de exactitud y 60,49 % de F1 con ANN. Para valencia se obtiene un 74,96 % de exactitud y 49,31 % de F1 con KNN-3. Con esto, se concluye sobre la factibilidad de la identificación y clasificación de niveles de carga cognitiva, valencia y excitación, pero que aún falta investigación sobre las características a extraer de las señales, en particular del EEG, con el fin de aumentar la exactitud y F1 en el desempeño de los clasificadores.

Tabla de Contenido

1. Introducción	1
1.1. Motivación	1
1.2. Definición del problema	2
1.2.1. Detalles del proyecto Fondecyt	3
1.3. Objetivos	4
1.3.1. Objetivos generales	4
1.3.2. Objetivos específicos	4
1.4. Metodología	4
1.5. Resultados esperados y alcances	5
1.6. Organización de la memoria	6
2. Marco Teórico	7
2.1. Tareas de redacción	7
2.1.1. Etapas	7
2.1.2. Estructura de un ensayo	8
2.2. Teoría de la carga cognitiva	9
2.2.1. Carga cognitiva en el proceso de escritura	9
2.2.2. Medición de la carga	10
2.3. Estado emocional	11
2.4. Medidas psico-fisiológicas	12
2.4.1. Seguimiento ocular	12
2.4.2. Tamaño de la pupila	13
2.4.3. Electroencefalografía	14
2.4.4. Actividad electrodérmica	16
2.4.5. Actividad del sistema cardiovascular	17
2.4.6. Temperatura corporal	18
2.5. Procesamiento y clasificación de los datos	19
2.5.1. Selección de características	19
2.5.1.1. Eliminación recursiva de características (RFE)	20
2.5.2. Clustering	21
2.5.2.1. Clustering Aglomerativo	21
2.5.2.2. Self-organizing Maps (SOM)	23
2.5.3. Clasificación	24
2.5.3.1. Evaluación del clasificador	24
2.5.3.2. Clasificadores	25
2.5.3.3. K-vecinos más cercanos (KNN)	25

2.5.3.4.	Máquinas de soporte vectorial (SVM)	26
2.5.3.5.	Redes neuronales artificiales (RNA)	29
2.6.	Estado del arte	33
2.6.1.	Carga cognitiva en el proceso de escritura	33
2.6.2.	Medición del estado del usuario con sensores psico-fisiológicos	35
2.6.2.1.	Medición de carga Cognitiva	35
2.6.2.2.	Medición de emociones	37
3.	Ejecución experimental	40
3.1.	Instrumentación	40
3.1.1.	Unidad GSR+ de Shimmer	40
3.1.1.1.	Respuesta galvánica de la piel	41
3.1.1.2.	Sensor óptico de pulso de Shimmer	42
3.1.2.	Unidad <i>Bridge Amplifier</i> de Shimmer	43
3.1.3.	ECG de tarjeta biomédica todo-en-uno de BITalino	44
3.1.4.	EEG Emotiv EPOC	45
3.1.5.	<i>Eye tracker</i> Tobii T120	48
3.2.	Grupo experimental	49
3.3.	Diseño experimental	50
3.3.1.	Preparación del experimento	50
3.4.	Procedimiento experimental	51
3.4.1.	Configuración sensores	52
4.	Análisis de datos	54
4.1.	Definición de observaciones	54
4.1.1.	Etiquetado de subtareas	54
4.1.2.	Duración promedio de las etapas del ensayo	55
4.2.	Preprocesamiento de las señales	56
4.2.1.	Señal GSR	56
4.2.2.	PPG	56
4.2.3.	Temperatura Corporal	57
4.2.4.	ECG	58
4.2.5.	EEG	59
4.2.6.	Diámetro pupilar	59
4.3.	Extracción de características	60
4.3.1.	Eye Tracker	61
4.3.2.	GSR	61
4.3.3.	ECG	62
4.3.4.	PPG	62
4.3.5.	Temperatura corporal	62
4.3.6.	EEG	62
4.3.6.1.	Carga cognitiva	63
4.3.6.2.	Detección de emociones	63
4.4.	Etiquetado de las muestras mediante Clustering	64
4.5.	Clasificación	66
5.	Resultados y discusión	68

5.1.	Análisis estadístico	68
5.1.1.	Diferenciación entre etapas de un ensayo	68
5.1.2.	Diferenciación entre estados para cada etapa del ensayo	69
5.1.2.1.	Carga cognitiva	69
5.1.2.2.	Valencia	70
5.1.2.3.	Excitación	70
5.2.	Minería de datos	71
5.2.1.	Selección de características	71
5.2.2.	Clasificación de carga cognitiva	73
5.2.2.1.	Elección de tamaño de observaciones	73
5.2.2.2.	Porcentaje de clasificación según número de clases	74
5.2.2.3.	Clasificación por sensor	74
5.2.2.4.	Nuevas características EEG para carga cognitiva	75
5.2.2.5.	Clasificación de nivel alto y bajo de carga cognitiva	76
5.2.2.6.	Clasificación entre-sujetos	76
5.2.3.	Clasificación de estado emocional	77
5.2.3.1.	Tamaño de observaciones	77
5.2.3.2.	Promedio de porcentajes de clasificación según número de clases	79
5.2.3.3.	Clasificación por sensor	80
6.	Conclusiones y trabajo futuro	82
	Bibliografía	85
	Anexo A. Grupo experimental	93
	Anexo B. Cantidad de clusters por participante	95
B.1.	Carga Cognitiva	95
B.2.	Valencia	96
B.3.	Excitación	97
	Anexo C. Análisis estadístico	99
C.1.	Carga cognitiva	99
C.1.1.	Borrador	99
C.1.2.	Introducción	100
C.1.3.	Tesis	102
C.1.4.	Desarrollo	103
C.1.5.	Conclusión	104
C.2.	Valencia	105
C.2.1.	Borrador	105
C.2.2.	Introducción	106
C.2.3.	Tesis	107
C.2.4.	Desarrollo	108
C.2.5.	Conclusión	109
C.3.	Excitación	110
C.3.1.	Borrador	110
C.3.2.	Introducción	111
C.3.3.	Tesis	112

C.3.4. Desarrollo	113
C.3.5. Conclusión	114
Anexo D. Detalle de clasificación	116
D.1. Carga cognitiva	116
D.2. Carga cognitiva para dos clases	117
D.3. Valencia	117
D.4. Excitación	119

Índice de Tablas

2.1. Kernels más utilizados	28
4.1. Duración promedio de cada etapa estructural del ensayo	55
4.2. Duración promedio de cada etapa mental durante la elaboración del ensayo	55
4.3. Características extraídas para cada señal	61
4.4. Características extraídas del EEG	62
4.5. Promedio de niveles de carga obtenidos para los distintos usuarios.	66
4.6. Promedio de niveles de valencia obtenidos para los distintos usuarios.	66
4.7. Promedio de niveles de excitación obtenidos para los distintos usuarios.	66
4.8. Cantidad de datos promedio para todos los participantes	67
5.1. Resumen de señales que presentan un p-value significativo ($<0,05$) para diferenciar entre etapas del ensayo	69
5.2. Resumen señales que presentan al menos una característica con p-value significativo ($<0,05$) para los niveles de carga cognitiva	69
5.3. Resumen señales que presentan al menos una característica con p-value significativo ($<0,05$) para los niveles de valencia	70
5.4. Resumen señales que presentan al menos una característica con p-value significativo ($<0,05$) para los niveles de excitación	70
5.5. Resumen 5 características más seleccionadas para la clasificación de niveles de carga cognitiva	71
5.6. Resumen 5 características más seleccionadas para la clasificación de niveles de valencia.	72
5.7. Resumen 5 características más seleccionadas para la clasificación de niveles de excitación.	72
5.8. Promedio de porcentajes de clasificación para la carga cognitiva para los mejores clasificadores encontrados, maximizando F1 con observaciones de tamaño $t = 2$ [s]	73
5.9. Promedio de porcentajes de clasificación para la carga cognitiva para los mejores clasificadores encontrados, maximizando F1 con observaciones de tamaño $t = 5$ [s]	73
5.10. Promedio de porcentaje de clasificación por número de clase para carga cognitiva	74
5.11. Promedio de porcentaje de clasificación para niveles de carga cognitiva, para cada sensor por separado.	74
5.12. Desempeño promedio del mejor clasificador con nuevas características para el EEG	75

5.13. Desempeño promedio de clasificadores para discernir entre nivel alto y bajo de carga cognitiva	76
5.14. Desempeño de clasificación entre-sujetos para carga cognitiva	77
5.15. Promedio de porcentajes de clasificación para los niveles de valencia para los mejores clasificadores encontrados, maximizando F1 con observaciones de tamaño $t = 2$ [s]	77
5.16. Promedio de porcentajes de clasificación para los niveles de valencia para los mejores clasificadores encontrados, maximizando F1 con observaciones de tamaño $t = 5$ [s]	78
5.17. Promedio de porcentajes de clasificación para los niveles de excitación para los mejores clasificadores encontrados, maximizando F1 con observaciones de tamaño $t = 2$ [s]	78
5.18. Promedio de porcentajes de clasificación para los niveles de excitación para los mejores clasificadores encontrados, maximizando F1 con observaciones de tamaño $t = 5$ [s]	78
5.19. Promedio de porcentajes de clasificación por número de clase para nivel de valencia	79
5.20. Promedio de porcentajes de clasificación por número de clase para nivel de excitación	79
5.21. Promedio de porcentajes de clasificación para niveles de valencia, para cada sensor por separado.	80
5.22. Promedio de porcentajes de clasificación para niveles de excitación, para cada sensor por separado.	80
5.23. Índice de motivación promedio para todos los sujetos en cada etapa del ensayo	81
A.1. Grupo experimental	93
B.1. Niveles de carga cognitiva por usuario	95
B.2. Niveles de valencia por usuario	96
B.3. Niveles de excitación por usuario	97
C.1. Resultados ANOVA-MR para niveles de carga cognitiva en la etapa borrador del ensayo.	99
C.2. Resultados ANOVA-MR para niveles de carga cognitiva en la etapa introducción del ensayo.	100
C.3. Resultados ANOVA-MR para niveles de carga cognitiva en la etapa tesis del ensayo.	102
C.4. Resultados ANOVA-MR para niveles de carga cognitiva en la etapa desarrollo del ensayo.	103
C.5. Resultados ANOVA-MR para niveles de carga cognitiva en la etapa conclusión del ensayo.	104
C.6. Resultados ANOVA-MR para niveles de valencia en la etapa borrador del ensayo.	105
C.7. Resultados ANOVA-MR para niveles de valencia en la etapa introducción del ensayo.	106
C.8. Resultados ANOVA-MR para niveles de valencia en la etapa tesis del ensayo.	107
C.9. Resultados ANOVA-MR para niveles de valencia en la etapa desarrollo del ensayo.	108

C.10.Resultados ANOVA-MR para niveles de valencia en la etapa conclusión del ensayo.	109
C.11.Resultados ANOVA-MR para niveles de excitación en la etapa borrador del ensayo.	110
C.12.Resultados ANOVA-MR para niveles de excitación en la etapa introducción del ensayo.	111
C.13.Resultados ANOVA-MR para niveles de excitación en la etapa tesis del ensayo.	112
C.14.Resultados ANOVA-MR para niveles de excitación en la etapa desarrollo del ensayo.	113
C.15.Resultados ANOVA-MR para niveles de excitación en la etapa conclusión del ensayo.	114
D.1. Detalle por sujeto para mejor clasificador de carga cognitiva (ANN con alfa = 0,01)	116
D.2. Detalle por sujeto para clasificación de cargas cognitivas extremas (SVM, kernel rbf y $C = 10$)	117
D.3. Detalle por sujeto para mejor clasificador de valencia (KNN, $k = 3$)	117
D.4. Detalle para mejor clasificador excitación (ANN, alfa = 0.01)	119

Índice de Ilustraciones

2.1. Estados emocionales y su posición en el plano valencia-excitación	11
2.2. Reflexión de la córnea y pupila brillante, visto en la imagen de la cámara infrarroja	13
2.3. Sistema 10-20 de electrodos	14
2.4. Distribución de las bandas de frecuencia	14
2.5. Señal GSR, con sus componentes fásica (naranja) y tónica (amarilla) diferenciadas	17
2.6. Ciclo cardíaco ideal	17
2.7. El primer nivel de abstracción es obtenido al crear un set de vectores prototipos (por ejemplo con SOM). El segundo nivel se crea al clusterizar el SOM	23
2.8. Ilustración de una clasificación mediante KNN, para $K=3$. Se toman los 3 vecinos más cercanos y, de acuerdo a estos, se entrega una probabilidad de pertenencia a una clase de X_1 y X_2	25
2.9. Esquemático de SVM linealmente separable	26
2.10. Caso SVM no lineal: gracias a la función φ el espacio se transforma y los datos son linealmente separables	27
2.11. Modelo de perceptrón	29
2.12. Esquema RNA MLP: diagrama para red de dos capas. Las variables de entrada, ocultas y de salida son representadas por nodos y los pesos son representados como links entre los nodos, donde el bias se representa como un link que viene de variables adicionales. Las flechas muestran la dirección de la información durante la propagación hacia adelante.	30
2.13. Funciones de activación	31
3.1. Unidad GSR+ de Shimmer	41
3.2. Zona recomendada para medición de GSR en pies	41
3.3. Sensor óptico de pulso en oreja	42
3.4. Implementos necesarios para medición de temperatura en la piel	43
3.5. Posición electrodos ECG	44
3.6. Sensor para medir EEG	45
3.7. Posicionamiento del sensor	46
3.8. Sensores de referencia	46
3.9. Verificación de posición de sensores en <i>Xavier TestBench</i>	47
3.10. Eye Tracker Tobii T120	48
3.11. Posicionamiento del usuario para el eye tracker	49
3.13. Sala donde se llevan a cabo los experimentos	51

3.14. Participante con los sensores ya posicionados.	52
3.12. Página 2 y 3 de encuesta respondida por el usuario	53
4.1. Preprocesamiento de señal GSR	56
4.2. Preprocesamiento de señal PPG	57
4.3. Señal de PPG con forma menos clara	57
4.4. Preprocesamiento de la temperatura corporal	58
4.5. Preprocesamiento señal ECG	58
4.6. Preprocesamiento canal AF3 de EEG	59
4.7. Preprocesamiento señal de diámetro pupilar	60
4.8. Proceso para clusterizar las muestras	65

Capítulo 1

Introducción

1.1. Motivación

Hoy en Chile, periódicamente se aplican distintas pruebas estandarizadas: nacionales (SIMCE) e internacionales (TIMMS, IALS, UNESCO, PISA). Todas estas revelan una verdad ineludible: el nivel de comprensión lectora, tanto de los escolares como de los adultos chilenos, es muy insuficiente, comparado con el de países del mundo desarrollado. En particular, los niños y niñas chilenos leen y escriben poco y mal; no dominan las habilidades para decodificar y, lo que es más grave: no comprenden lo que leen [1].

Los resultados del último SIMCE (2015) son decisivos. Con respecto a los distintos ámbitos que mide el SIMCE, los resultados entregados revelan que [2]:

- Capacidad de transmitir un propósito en el texto: cerca de un 15 % de los niños de 6to básico no logran esta habilidad. Ésta debiese ser una capacidad desarrollada en los primeros años de la educación escolar.
- Organización de texto: cerca de la mitad de los estudiantes no escribe textos con una organización clara y completa.
- Coherencia interna del texto (su comprensibilidad): un tercio de los estudiantes alcanza el nivel más alto de esta competencia, mientras que otro tercio no logra escribir textos comprensibles
- Desarrollo de ideas: entre un 12 y 13 % es capaz de desarrollar sus ideas en profundidad y de forma clara, mientras que cerca de la mitad de los estudiantes no logra este objetivo

Pero este problema no es exclusivo en niños, en 2016 se realizó PIAAC (Evaluación Internacional de las Competencias de los Adultos), un estudio internacional sobre competencias de adultos. En particular, para las competencias de comprensión lectora se obtiene que [3]:

- Chile presenta un puntaje promedio de 48 puntos más bajo que el promedio de países OCDE en comprensión lectora.
- Al compararlo con la última prueba IALS realizada en Chile (1998), se obtiene que el bajo desempeño de Chile en comprensión lectora se ha mantenido constante en los

últimos 17 años (220 puntos) pese al aumento del PIB per cápita y de la población con educación superior.

- 53 % de los adultos chilenos se encuentra en nivel de desempeño bajo, mientras que en el promedio OCDE sólo el 19 % se encuentra en el nivel de desempeño bajo.

Con respecto a lo descrito anteriormente, la literatura tiene varios puntos clave:

- Hoy en día en Chile la mayoría de los niños y adultos no entienden lo que leen y no son capaces de comunicarse eficazmente por escrito [1][2][3][4][5].
- El fracaso en la adquisición de la lectura y la escritura se explicaría en mayor medida por determinadas prácticas pedagógicas, más que por las condiciones sociales o culturales familiares de los alumnos [1].
- La calidad de la alfabetización en la educación inicial tiende a determinar el “éxito” o “fracaso” escolar de los niños [1].
- Para solucionar este problema, es fundamental avanzar en el desarrollo de instrumentos de evaluación con validez individual, que ayuden a los profesores a detectar a los alumnos más rezagados y las áreas en las que requieren reforzamiento [2].

1.2. Definición del problema

Con respecto a los antecedentes mencionados en la Sección 1.1, resulta clave el último punto sobre el desarrollo de instrumentos de evaluación que sean personalizados para cada persona y que ayuden a los profesores a identificar alumnos que se van quedando atrás; esto quizás porque les cuesta la tarea que están realizando, porque tienen un déficit en el área de la tarea o por otra razón.

Con esto en mente, se busca aportar a la solución del problema de alfabetización de niños y adultos chilenos, mediante los avances tecnológicos y científicos de hoy. Así, desde el Laboratorio de *Web Science and Smart Technologies* (WeSSTLab) [6], a cargo del académico del Departamento de Industrias de la FCFM Ángel Jiménez Molina, quién dirige el Fondecyt 11130252: “*A Cognitive Resource-Aware Mobile Service Framework to Support Human-Computer-Interactions in Ubiquitous Computing Environments*”, nace la idea de realizar un proyecto que tiene como objetivo final el diseño de una herramienta educativa de redacción personalizada que apoye al docente, basada en la medición del estudiante con sensores psico-fisiológicos.

Existe amplia evidencia en la literatura de que los sensores psico-fisiológicos pueden medir tanto la carga cognitiva (esfuerzo mental) del usuario en tiempo real (por ejemplo [7][8][9][10][11]) como pueden medir las distintas emociones que experimenta el usuario al realizar una tarea (por ejemplo [12][13]) o recibe ciertos estímulos (por ejemplo [14][15][16]). Se ha probado para tareas aritméticas, de desarrollo de código, de lectura, juegos; pero existe poca literatura en cuanto a actividades de redacción.

Así, como primera aproximación al proyecto planteado nace este trabajo: se busca saber si es posible medir el esfuerzo mental y/o emociones de un usuario al realizar actividades de

redacción, tales como escribir un texto o expresar una idea vía escritura.

En particular, se busca medir la carga cognitiva experimentada por un usuario durante una actividad de redacción. También interesa medir el nivel de valencia/excitación (que se relaciona con el estado emocional del usuario) durante esta misma actividad. Se busca responder si, utilizando una combinación de sensores psico-fisiológicos no invasivos, se puede medir la carga cognitiva y/o estado emocional experimentado por el usuario durante la actividad. Para responder estas interrogantes, se debe diseñar un experimento, realizarlo en distintas personas (50-60) y analizar los datos obtenidos mediante métodos estadísticos y de minería de datos.

1.2.1. Detalles del proyecto Fondecyt

Este trabajo de título se desarrolla en el marco del proyecto Fondecyt otorgado en el año 2013 titulado: “*A Cognitive Resource-Aware Mobile Service Framework to Support Human-Computer-Interactions in Ubiquitous Computing Environments*” a cargo del profesor del Departamento de Ingeniería Industrial Ángel Jiménez Molina. En este proyecto se pretende desarrollar un mecanismo de ingeniería cognitiva que seleccione, componga y desarrolle funcionalidades durante el tiempo en que se ejecute, tomando en consideración el contexto situacional y los recursos cognitivos empleados por el usuario de acuerdo a las tareas de interacción humano-computador (en inglés Human-Computer Interactions, HCI) realizadas.

El objetivo general del proyecto Fondecyt es reducir el agotamiento de recursos cognitivos mientras se proveen servicios. Para esto los objetivos específicos que lo sustentan son: 1) Modelar relaciones entre recursos mentales, tareas HCI y actividades; 2) Definir métricas para evaluar agotamiento de recursos cognitivos; 3) Crear algoritmos de Binding (Algoritmo capaz de seleccionar el servicio idóneo a entregar según el contexto); 4) Diseñar un modelo computacional para analizar servicios según demandas cognitivas e interferencias; 5) Desarrollar un framework para dispositivos móviles que englobe todo lo anterior. Esto se hace considerando la ubicación, los intereses y las preferencias de los actores de los procesos, así como la actividad en la que se encuentran inmersos, los factores ambientales, sociales y temporales y la información interpretada de sensores, entre otras fuentes de información contextual, y aplicables a diferentes dominios como mobile marketing, sistemas de recomendación, e-health y gestión en tiempo real de la cadena de suministros, entre otros.

Además, existen dos alumnos desarrollando temas paralelos del mismo Fondecyt. Ambos ya terminaron su memoria: Hernán Lira (“Medición de cargas cognitivas durante actividades de interacción humano computador en ambiente móvil usando sensores psico-fisiológicos” [17]) y Cristian Retamal (“Estudio del comportamiento de la carga cognitiva de usuarios que navegan en un sitio web” [18]). Ambos estudian la carga cognitiva de usuarios en distintos escenarios.

1.3. Objetivos

1.3.1. Objetivos generales

Diseñar y ejecutar un experimento con el fin de evaluar si se puede - o no - medir la carga cognitiva y el estado emocional durante la elaboración de una actividad de redacción, mediante señales psico-fisiológicas, utilizando algoritmos de preprocesamiento de señales, análisis estadísticos y modelos supervisados y no supervisados de minería de datos.

1.3.2. Objetivos específicos

- Conocer sobre la teoría de la carga cognitiva, la medición de las emociones, las señales psico-fisiológicas, el proceso de escritura y la relación entre estos cuatro tópicos.
- Diseñar e implementar un experimento que permita la medición de la carga cognitiva y el estado emocional de usuarios que redactan un texto, utilizando sensores psico-fisiológicos.
- Aplicar o diseñar algoritmos de pre-procesamiento de señales psico-fisiológicas para la eliminación de ruido y artefactos.
- Determinar si los sensores psico-fisiológicos son capaces de discernir realmente los estados cognitivos y emocionales en cada etapa (introducción, tesis, desarrollo, conclusión) de la escritura del ensayo y si estas etapas se diferencian entre sí.
- Evaluar modelos no supervisados de minería de datos a las señales psico-fisiológicas, para identificar los niveles de carga cognitiva, valencia y excitación experimentados por el usuario mientras está realizando su ensayo.
- Evaluar modelos supervisados de minería de datos para la clasificación de los estados y niveles de carga cognitiva, valencia y excitación identificados anteriormente.
- Discutir resultados y entregar posibles vías de desarrollo para un trabajo futuro de dispositivos de captura de señales y herramienta computacional aplicable en el aula.

1.4. Metodología

A continuación, se presenta la metodología que se seguirá en este trabajo. Ésta se basa en trabajos anteriores de este Fondecyt [17][18].

1. Estudio del estado del arte con respecto a la carga cognitiva, estado emocional, proceso de redacción de un texto, señales psico-fisiológicas y procesamiento de datos. El fin de esto es generar el suficiente conocimiento para entender el problema planteado.
2. Desarrollo del diseño experimental.
3. Realización del experimento: medición de las señales fisiológicas del usuario mientras realice el experimento. Se utilizarán los siguientes sensores: *eye tracker* (Tobii), ECG

(BITalino), EEG (Emotiv EPOC), GSR, PPG y sensor de temperatura (Shimmer Sensing).

4. Procesamiento de las señales para eliminar ruidos y artefactos. Para esto, se utilizan filtros, normalización y otros métodos.
5. Análisis de datos utilizando Python. Se utilizan test estadísticos y técnicas de minería de datos para la identificación y clasificación de los datos en nivel de carga cognitiva y nivel de valencia/excitación.
6. Se evalúa el desempeño de la clasificación mediante análisis estadísticos y figuras de mérito como: exactitud (*accuracy*), precisión (*precision*), sensibilidad (*recall*) y F1 (promedio armónico entre precisión y sensibilidad).
7. Discusión en torno a los resultados obtenidos y a las tablas presentadas.
8. Conclusiones sobre el éxito (o fracaso) de las clasificaciones y del análisis realizado.
9. Conclusiones que ahonden sobre el trabajo realizado y como esto agrega valor por sobre otros métodos. Proponer vías para seguir desarrollando en torno a este tema y con miras en el objetivo del diseño de una herramienta educativa de redacción personalizada.

1.5. Resultados esperados y alcances

Para este trabajo, se esperan los siguientes resultados:

- Estado del arte de la medición de carga cognitiva y estado emocional mediante sensores psico-fisiológicos.
- Diseño experimental válido y replicable.
- Procesamiento de las señales para mejorar su calidad.
- Análisis estadístico que determine si las señales pueden discernir entre estados para todas las etapas del ensayo.
- Tabla comparativa de desempeño de distintos clasificadores al clasificar carga mental y estado emocional y análisis de estas clasificaciones.

Con respecto a los alcances, el foco del trabajo se encuentra en el resultado científico de la posibilidad de evaluar estados cognitivos y emocionales en la tarea de redacción. No está dentro del alcance de la memoria el diseño y construcción de dispositivos de captura de señales (sensores). Se utilizan dispositivos pre-existentes con los que cuenta el WeSSTLab. Tampoco está considerada la implementación de una herramienta computacional final reproducible en el aula.

1.6. Organización de la memoria

El presente documento consta de 6 capítulos, cada uno comenzando con una breve introducción a los temas abordados. A continuación, se presenta una síntesis de cada uno de ellos:

- El Capítulo 2 se compone de todos los antecedentes contingentes a esta investigación, comenzando por definir conceptos básicos para comprender esta investigación (tareas de redacción, carga cognitiva, estado emocional y señales psico-fisiológicas) y, luego, revisando el estado del arte en cuanto a medición de la carga cognitiva en tareas de redacción y a la medición de carga cognitiva y estado emocional mediante el uso de sensores psico-fisiológicos.
- El Capítulo 3 describe la metodología propuesta, en particular, describe el experimento diseñado y realizado. Se presentan los sensores a utilizar, algunas características técnicas y recomendaciones para un buen uso.
- En el Capítulo 4 se presenta el preprocesamiento (limpieza) de los datos y su análisis.
- En el Capítulo 5 se exponen y analizan los resultados más importantes en base a los datos analizados.
- Finalmente, el Capítulo 6 integra las conclusiones extraídas de la clasificación y el análisis de los resultados obtenidos. Complementariamente, se realizan comentarios sobre el trabajo futuro en vías del desarrollo de una plataforma basada en sensores psico-fisiológicos para monitorear a estudiantes en actividades de redacción.

Capítulo 2

Marco Teórico

A continuación, se detalla sobre distintos conceptos y definiciones necesarias para comprender los temas tratados en esta memoria. Primero, se ahonda sobre la tarea de redactar un texto: las etapas que se siguen y sobre la estructura de un ensayo. Luego, se profundiza sobre dos temas claves: la carga cognitiva y el estado emocional; se detalla sobre cómo se definen, sobre como se miden y las distintas opciones para hacerlo. Luego, se describen las señales que serán utilizadas para, en este caso, clasificar nivel de carga y nivel de valencia/excitación. Se presentan distintos algoritmos de pre-procesamiento y procesamiento de los datos. Por último, se revisa el estado del arte actual en estas áreas.

2.1. Tareas de redacción

En esta sección se ahonda sobre los procesos seguidos por un ser humano al desarrollar un texto escrito. Se describen las etapas mentales generales seguidas, y, luego, se detalla sobre la estructura de un ensayo, dado que este es el texto que se le pedirá a los usuarios reallizar mientras se les miden sus señales psico-fisiológicas.

2.1.1. Etapas

En general, al escribir un texto, las personas siguen cierta estructura mental. En [19] y [20] describen esta metodología, con las siguientes etapas:

1. Pre-escritura o planificación: dentro de esta etapa, se cuenta con distintos recursos.
 - Lectura de los materiales que han de proporcionar datos, ideas, juicios y pensamientos que se desean plasmar en el documento. Es importante entenderlos y realizar conexiones dinámicas con lo que se desea plantear.
 - Organización de las ideas.
 - Enumerar o jerarquizar información, según como quiere presentarse en el texto.

2. Puesta en texto: cada oración, párrafo o capítulo debe girar, dentro de lo posible, en torno a una idea. Comprende cierta estructura: presentación de una idea, explicaciones, pruebas, discusiones o ampliaciones, para terminar con un enunciado de la idea con un tono rotundo y definitivo.
3. Revisión o corrección: el texto debe ser evaluado en torno a los objetivos propuestos al inicio. Se recomienda revisarlo cíclicamente y que la revisión final no sea inmediata después de terminando el texto. En cuanto al orden de la revisión final, se recomienda revisar el cuerpo del texto y, luego, la introducción y conclusión.

Es importante notar que estas etapas no se conciben como sucesivas, sino como procesos iterativos que se van realizando mientras se escribe. Es decir, la puesta en texto puede verse interrumpida en cualquier minuto para volver a la planificación o al observarse una falencia, se puede agregar o quitar un tema o, en la revisión, darse cuenta que se debe volver a plantear la idea de un párrafo. Esta acción es permanente hasta que el escrito se da por concluido [19]. En [20] hablan de la importancia de estas etapas: se argumenta que la gran parte de los problemas de redacción de un texto, se deben fundamentalmente, a que el escritor no planea ni revisa sus escritos.

2.1.2. Estructura de un ensayo

Un ensayo es un texto argumentativo donde el autor busca exponer su punto de vista sobre un tema. Así, el ensayo se ve definido por el uso de la expresión personal y por la posición que toma el ensayista en relación al tema planteado; por su condición subjetiva. Dentro de las ventajas de este texto argumentativo se encuentra su sencillez productiva, su brevedad (aunque no existe regla en cuanto a su duración) y la comunicación ágil que existe entre un ensayista y su lector. Dentro de sus características se encuentra en que el autor debe utilizar distintos recursos literarios y argumentos tal que resulte “objetiva” la tesis que desea demostrar [20].

Un ensayo cuenta, principalmente, con tres secciones [20]:

- **Introducción:** incluye una breve presentación del tema, motivación de la escritura, la tesis planteada y un pequeño adelanto de qué se hablará a continuación.
- **Desarrollo:** se exponen los argumentos. Es un argumento (o idea principal) por párrafo.
- **Conclusión:** se reafirma la tesis, basándose en lo expuesto a lo largo del texto y se cierra el ensayo.

Además, se debe tener en cuenta que la tesis planteada en el documento argumentativo (en particular un ensayo) debe estar sustentado por un argumento. Este argumento está compuesto por una base (razón de por qué planteo esta tesis), una garantía (relaciona la base con la tesis) y un respaldo (fuente externa que valide lo planteado).

2.2. Teoría de la carga cognitiva

La teoría de la carga cognitiva se basa en la idea de que existe una memoria de trabajo que tiene una capacidad limitada de operación y almacenamiento cuando trabaja con información nueva, pero es ilimitada cuando trabaja con información proveniente de la memoria de largo plazo. Además, la teoría dice que la memoria de trabajo está organizada tal que existe una central ejecutiva que cuenta con dos sub-sistemas parcialmente independientes: el que procesa información visual y espacial (*sketchpad* visuoespacial) y el que procesa información auditiva (loop fonológico) [21].

Uno de los principales objetivos de esta teoría, es de desarrollar métodos de instrucciones para utilizar de manera eficiente esta capacidad limitada de procesamiento cognitivo. Para esto, surge el concepto de carga cognitiva que se define como la carga que la ejecución de cierta tarea impone en el sistema cognitivo de la persona en cuestión. La carga cognitiva es limitada y al ser sobrepasada, la persona puede caer en su rendimiento. Por esto, interesa medir la carga e implementar sistemas que se adapten al nivel de ésta [22].

2.2.1. Carga cognitiva en el proceso de escritura

La escritura es uno de los procesos cognitivos más complejos que los humanos realizan durante su vida. Requiere un largo número de componentes cognitivos que operan en diferentes niveles de representación. Se reconocen cuatro sets de procesos: [23]:

- La etapa de planificación, donde se construye un mensaje pre-verbal que corresponde a las ideas que el escritor desea comunicar. En esta etapa, las ideas son recuperadas de la memoria de largo plazo y (re-)organizadas si es necesario.
- El proceso de traducción, que involucra una representación a nivel lingüístico en donde el mensaje pre-verbal de la planificación se transforma en un mensaje verbal. Por lo tanto, la estructura conceptual elaborada durante la planificación es gramaticalmente codificado al recuperar en el léxico mental las propiedades sintácticas y morfológicas de las palabras.
- La comparación del texto escrito con la representación mental de la intención del escritor de su propio texto, para la mejora de éste.
- El componente a nivel motor de la representación. La ejecución o transcripción gráfica permite que el autor transforme el mensaje lingüístico en una secuencia de movimientos motores para producir la escritura.

En [24] abordan sobre los principales componentes de la teoría cognitiva que inciden en el proceso antes descrito:

- En el proceso de planificación del texto se utiliza principalmente la central ejecutiva y el sistema esclavo visuoespacial al elegir el tono del texto, crear ideas para el texto y organizando ideas. Cuando un escritor crea una visualización mental del escrito, se utiliza el sistema visuoespacial.

- En el proceso de traducción se utiliza el loop fonológico y la central ejecutiva, dado que durante la generación de texto se crea un discurso interno sobre la dicción y el orden de ésta. El discurso crea representaciones fonológicas de las palabras, lo que utiliza el loop fonológico. Además, el escritor debe elegir que texto generado por el discurso interno será almacenado y utilizado, lo que utiliza la central ejecutiva.
- En el proceso de revisión, donde el autor lee lo escrito para corregirlo, se exige a la central ejecutiva y al loop fonológico. Leer un texto escrito por uno mismo demanda más capacidad en la memoria de trabajo que leer textos de otros autores, dado que los errores son transparentes por la familiaridad que existe con el texto. Además, editar un texto incluye distintas tareas: descifrar errores gramaticales y cambiarlos, revisar la fluidez de las oraciones y párrafos, detectar errores de organización y arreglarlos. Cada una de estas tareas por sí solas no demandan gran capacidad a la central ejecutiva, pero en conjunto demandan bastante. Se cree que esta etapa es la más demandante cognitivamente [25].
- Al momento de escribir, se utiliza la central ejecutiva para preparar al sistema motor para tipear, escribir o dictar. Sin embargo, se ha demostrado que este proceso es el que menos consume capacidad de la memoria de trabajo.

2.2.2. Medición de la carga

Existen distintas maneras de medir la carga cognitiva, en [9] detallan 4 clases de mediciones:

1. Medidas analíticas: basan la estimación de la carga en información general del sistema y del usuario. Por ejemplo, información sobre el conocimiento previo del usuario con respecto al tema tratado o la dificultad intrínseca a la tarea a realizar.
2. Medidas subjetivas: se basan en pedirle al usuario que evalúe la carga cognitiva que está experimentando o experimentó durante la tarea. Tiene la desventaja de que no es posible medir en tiempo real la carga cognitiva con esta técnica dado que se hace *a posteriori*. Si se hiciese durante sería una interrupción.
3. Medidas por rendimiento: se infiere la carga cognitiva experimentada por el usuario basándose en una medida de rendimiento establecida, por ejemplo, el tiempo que se demora el usuario en completar cierta tarea. Como desventaja se observa que no refleja todas las variaciones en la carga cognitiva y es apropiada solo para ciertas actividades.
4. Medidas psico-fisiológicas: se ha observado que los cambios en varios procesos y estados del cuerpo humano covarían con cambios en la carga cognitiva. Así, el monitorear ciertas funciones del cuerpo permite inferir la carga experimentada en el momento. Como desventaja se observa que para obtener estas mediciones corporales se deben utilizar distintos sensores (medidas del ritmo cardíaco, del tamaño de la pupila..) que son invasivos o tienen limitaciones, lo que no permite un uso diario ni fuera de un ambiente de laboratorio.

2.3. Estado emocional

Las emociones se definen como un proceso inconsciente que emana desde una expresión de sentimientos y pensamientos [26]. Un problema mayor al tratar de entender las emociones, es la definición de una emoción. De hecho, los psicólogos no están de acuerdo en cuanto a qué es considerado una emoción ni a cuántos tipos de emociones existen [27]. Lo que si se consensúa es que las emociones son una combinación entre un factor físico (emociones primarias o corporales) y otro cognitivo (emociones mentales) [13].

Psicólogos y neurocientíficos han explicado diversas teorías de la emoción: las dos más utilizadas corresponden al modelo de emociones discretas, propuesto por Ekman y al modelo de dos dimensiones valencia-excitación, propuesta por Lang:

- El modelo de emociones discretas dice que existe un set de emociones básicas, universales a través de las culturas. Se han sugerido diferentes categorías de emociones, en las que se ha tenido más acuerdo son: felicidad, tristeza, sorpresa, enojo, disgusto y miedo [28].
- El modelo del plano valencia-excitación: la valencia define si la emoción es positiva o negativa y en que grado; la excitación define la intensidad de la emoción, yendo desde calmado (valor bajo) a emocionado (valor alto) [13]. En la Figura 2.1 (traducida de [29]) se muestra este plano y se ejemplifica dónde irían ciertas emociones.



Figura 2.1: Estados emocionales y su posición en el plano valencia-excitación

Se utilizan distintas maneras para medir y clasificar emociones, entre ellas: análisis del habla, análisis de gestos/expresión facial o análisis de cambios en los patrones psico-fisiológicos [15]. La ventaja de este último es que si una persona no expresa emociones visiblemente (mediante el habla o su expresión), los cambios psico-fisiológicos son inevitables y detectables. Esto, dado que los nervios del sistema nervioso simpático se activan cuando una persona está excitada positiva o negativamente. Esta activación aumenta el ritmo cardíaco, aumenta la tasa de respiración, aumenta la presión sanguínea y disminuye la variabilidad en el ritmo cardíaco [28].

2.4. Medidas psico-fisiológicas

La psico-fisiología es donde se relaciona la anatomía y la fisiología del ser humano con los fenómenos psicológicos que pueda experimentar, es decir, relaciona las experiencias y comportamientos del organismo en un ambiente físico y social [30].

A continuación, se describen ciertas medidas psico-fisiológicas. Éstas corresponden a las que pueden ser medidas utilizando el equipamiento disponible en el centro de investigación donde se desarrolla esta memoria. Para cada una se desarrolla sobre qué es, cómo se mide y algunas características útiles para la clasificación de emociones (nivel de valencia/excitación) y/o carga cognitiva descritas en la literatura.

2.4.1. Seguimiento ocular

El seguimiento ocular es una técnica en donde los movimientos de los ojos de una persona son medidos con el fin de que el investigador sepa adonde está mirando el usuario en todo momento y la secuencia en que los ojos de la persona cambian de una locación a otra.

Las principales medidas utilizadas en el seguimiento ocular son dos: fijaciones y sacadas. Las fijaciones son aquellos momentos en que los ojos están relativamente estacionarios mientras que las sacadas son los movimientos de ojos rápidos que suceden entre fijaciones.

Para obtener estas medidas, se utiliza un dispositivo llamado *eye tracker*. Estos funcionan midiendo el punto de vista con el método de la reflexión-de-córnea/centro-pupilar. Usualmente consisten de un computador de escritorio con una cámara infrarroja montada bajo o al lado de la pantalla, con un software de procesamiento de imágenes que ubica e identifica las características del ojo utilizadas para el seguimiento. Durante la operación del eye tracker, luz infrarroja de un LED integrado en la cámara es dirigido al ojo para crear una fuerte reflexión en los ojos para que sea más fácil seguirlos (se utiliza luz infrarroja para evitar encandilar al usuario con luz visible). La luz entra a la retina, y una gran proporción es reflejada de vuelta, haciendo que la pupila se vea brillante y como un disco bien definido (efecto pupila-brillante). La reflexión de la córnea también se genera por la luz infrarroja, mostrándose como un resplandor pequeño pero agudo (ver Figura 2.2, de [31]). Una vez que el software de procesamiento ha identificado el centro de la pupila y la ubicación del reflejo de la córnea, se calcula el vector entre ellos y mediante trigonometría, calcula la posición de la vista [31].

En general, las medidas obtenidas desde un eye tracker ayudan a entender el procesamiento de información visual y mostrada en una pantalla y, también, sobre cuán eficientemente está diseñada una interfaz [31]. En particular, se ha encontrado relación entre el promedio de la duración de la fijación ocular con la dificultad que está teniendo el usuario al extraer información de un monitor [8].

También, se puede obtener la tasa de parpadeo de un usuario, midiendo la cantidad de sacadas de éste. Se ha comprobado que ésta es inversamente proporcional con la atención o

carga mental. Es decir, mientras mayor sea la atención o carga mental demandada por una tarea menor será la tasa de parpadeo [8].

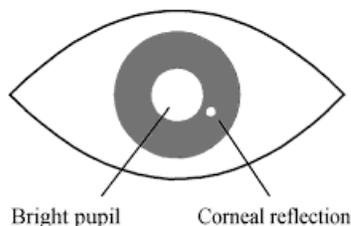


Figura 2.2: Reflexión de la córnea y pupila brillante, visto en la imagen de la cámara infrarroja

2.4.2. Tamaño de la pupila

La pupila es una abertura central en el iris que actúa a modo de diafragma; controla la cantidad de luz que entra en el ojo y es uno de los elementos oculares que mejora la calidad de la imagen que se forma en la retina. La pupila es diferente entre individuos y, además, entre los ojos del mismo sujeto, con asimetrías hasta de un 20 por ciento en individuos normales que, pocas veces, supera los 0.5 mm. El diámetro suele ser de 3 a 4 mm, con tendencia a ser más grande en la niñez y más pequeña de forma progresiva, con el aumento de la edad. Los cambios en los tamaños pupilares dependen también de las variaciones en el calibre de los vasos sanguíneos, inervados por la porción simpática del sistema nervioso autónomo [32].

Para obtener esta señal se puede utilizar un eye tracker o una cámara común y corriente que grabe la cara de la persona mientras realiza la tarea y luego realizar el procesamiento de esa imagen para recuperar la pupila.

Bajo condiciones de iluminación controlada, estudios demuestran que el tamaño de la pupila es una medida confiable y efectiva de la carga mental. Se ha encontrado que el aumento del tamaño de la pupila se relaciona con aumentos en la carga mental experimentada [33]. A pesar de esto, en [9] concluyen que no existen relación entre el tamaño de la pupila y la dificultad experimentada por un usuario al leer textos de distinta dificultad.

La medición del tamaño de la pupila tiene como ventajas que no se necesitan sensores invasivos y aparatosos, es altamente sensible y, además, se cree que es una medida directamente relacionada con la carga cognitiva experimentada por el usuario: se supone que ésta responde a cambios en la carga dentro de la escala de milisegundos [9].

El tamaño pupilar es altamente sensible a factores no relacionados con la carga cognitiva (por ejemplo a la luz ambiental). Por esto, cuando es utilizado para medir carga cognitiva, se utilizan distintas técnicas para controlar estas influencias externas. Entre ellas se cuentan: luz constante, evitar movimiento ocular, uso de estímulos no visuales, uso de tareas cortas-similares y evaluando sólo los valores medios promediados entre las tareas y los sujetos. [9]

2.4.3. Electroencefalografía

La electroencefalografía (EEG) mide la actividad eléctrica del cerebro, utilizando electrodos ubicados en el cuero cabelludo [8]. Se pueden utilizar distintas distribuciones de electrodos, en particular interesa la llamada 10-20 (Figura 2.3), donde se definen 21 posiciones de electrodos en la cabeza del usuario. Los nombres de los electrodos se componen de una letra, que indica la región del cerebro (F = lóbulo frontal, T = lóbulo temporal, C = centro, O = lóbulo occipital) y un número, que indica la posición (par = ubicación derecha de la cabeza, impar = ubicación izquierda de la cabeza).

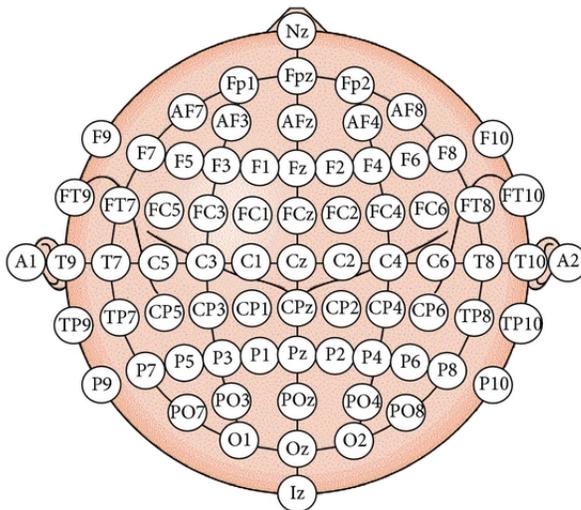


Figura 2.3: Sistema 10-20 de electrodos

La señal cruda del EEG corresponde a las oscilaciones neuronales que se experimentan en el cerebro, obteniéndose una señal que es una mezcla de varias frecuencias subyacentes, que se consideran un reflejo de cierto estado cognitivo, afectivo o atencional. Se definen 5 bandas, diferenciadas por su frecuencia: delta (1-4 Hz), theta (4-8 Hz), alfa (8-12 Hz), beta (13-25 Hz) y gamma (>25 Hz), que se muestran en la imagen 2.4 [34].

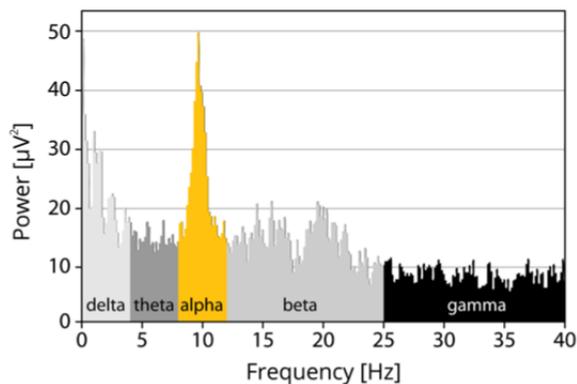


Figura 2.4: Distribución de las bandas de frecuencia

- Banda Delta: solo están presente en estados de sueño profundo no-REM. Se asocia con la profundidad del sueño; mientras más fuerte sea el ritmo delta, mayor es la profundidad del sueño. En general esta banda de frecuencias se encuentra más fuerte en el hemisferio derecho del cerebro, y su fuentes más típicas están localizadas en el hipotalamo.
- Banda Theta: se correlaciona con la dificultad de operaciones mentales, por ejemplo durante periodos de atención focalizada o de captación de información, procesamiento y aprendizaje y durante recuerdo de memorias. Se ha encontrado que la frecuencia de la banda theta se convierte en más prominente cuando aumenta la dificultad de la tarea. Esta banda puede ser obtenida desde toda la corteza, lo que indica que se genera mediante una amplia red que involucra las zonas prefrontal, central, parietal y temporal cortices.
- Banda Alfa: tienen varias correlaciones con funciones sensoriales, motoras y de memoria. Se observa un aumento en los niveles de la banda de poder alfa durante relajaciones mentales y físicas con los ojos cerrados. Al contrario, la banda alfa se ve reducida (o suprimida) durante actividades mentales o corporales con los ojos abiertos. La supresión de la banda alfa constituye una señal de estados de actividad mental y compromiso con la tarea. Se interpreta esta supresión como que el cerebro está recolectando información desde varios sentidos, coordinando recursos de atención y enfocándose en lo que interesa en el momento. La banda alfa es generada en sitio posteriores corticales, incluyendo la región occipital, parietal y temporal posterior del cerebro.
- Banda Beta: se correlaciona una alza en la banda beta con pensamientos ansiosos o activos y con estados de concentración activa. También, se observa un aumento en esta banda cuando se planea la ejecución de movimientos, en particular, cuando se requiere alguna acción que involucre movimientos finos de dedos y atención focalizada. Esta banda se genera en las regiones posterior y frontal del cerebro.
- Banda Gamma: sobre esta banda no se conoce mucho ni sobre dónde se genera ni sobre su función en el cerebro. Existe una discusión sobre si esta banda tiene un propósito dado en el proceso cognitivo o si es un subproducto de otros procesos neuronales como movimiento de ojo y micro sacadas.

Existe evidencia de que al aumentar la carga cognitiva, se suprime la banda alfa en el lóbulo parietal y que aumenta la banda theta en el lóbulo frontal [8] [35][36][37]. Sin embargo, en [38] el autor indica que, a pesar de que las oscilaciones de las bandas alfa y theta reflejan cambios en la carga cognitiva y desempeño de la memoria, es importante definir la banda alfa y theta para cada sujeto a partir del peak de la frecuencia de su banda alfa (IAF por sus siglas en inglés: *Individual Alpha Frequency*). Es decir, las frecuencias de corte no son las mismas para cada persona. A pesar de esto, existe literatura que utiliza las bandas “estándar” del EEG para clasificar carga cognitiva con buenos resultados [39][40][41][42].

Con respecto a la detección de emociones, las bandas alfa y beta son las más relevantes. En [43] se realiza un estudio sobre el reconocimiento del nivel de valencia y excitación en 3 escenarios distintos, utilizando la señal del EEG: viendo imágenes emotivas, escuchando música con los ojos cerrados y viendo imágenes al mismo tiempo que se escucha música (donde la imagen y el audio apuntan a la misma emoción). Se concluye sobre las características más relevantes en la detección de emociones para cada escenario y sobre el carácter personal de las emociones, es decir, de cómo para cada persona las señales del EEG se comportan de

manera diferente al demostrar las mismas emociones. A pesar de esta naturaleza personal, existen trabajos donde se clasifican emociones (como nivel de valencia/excitación o emociones discretas), con buenos resultados [29][44][45][46].

2.4.4. Actividad electrodérmica

La actividad electrodermal (EDA) o respuesta galvánica de la piel (GSR) mide la conductividad que tiene la piel en cierto momento. Ésta ha sido relacionada fuertemente con la excitación, atención, estado emocional, estrés y ansiedad . También, se ha demostrado que esta medida puede indicar los niveles cognitivos del usuario, la dificultad de la tarea y distinguir entre carga cognitiva y estrés [8].

En general, esta medida se toma con un sensor que mide la conductividad de la piel mediante uno o dos electrodos ubicados en las manos o en los pies. La conductividad de la piel varía con los cambios en la humedad (sudor) y puede revelar cambios en el sistema nervioso [47].

Las ventajas de esta medida es que es de bajo costo, fácil captura y robusta, pero se debe tener cuidado con el movimiento de los sensores ya que es muy fácil capturar solo ruido [47]. Se ha demostrado que para que esta medida sea efectiva, los sujetos deben relajarse entre tareas (ver un video relajante, escuchar música relajante..) [8].

La señal que se obtiene consta de dos partes: la de baja frecuencia (señal tónica) y la de alta frecuencia (señal fásica). Se muestra la señal cruda, separada en tónica y fásica en la Figura 2.5 (de [34]).

- Señal tónica o nivel conductivo de la piel (SCL por sus siglas en inglés: *Skin Conductance Level*): componente de la señal que varía lentamente en el tiempo, presentando leves cambios en la escala de 10-100 segundos. El alza y decaimiento de la señal cambia constantemente dentro de un mismo sujeto, dependiendo de su hidratación, sequedad de su piel o regulación autonómica. Esta componente puede diferir altamente entre sujetos. Estos antecedentes han llevado a concluir que, quizás, esta componente no tiene mucha información.
- Señal fásica o respuesta conductiva de la piel (SCR por sus siglas en inglés: *Skin Conductance Response*): esta componente está sobre la tónica y muestra alteraciones significativamente más rápidas. La señal es sensible a eventos de estímulo emocionales específicos y las variaciones (peaks) inducidos por los estímulos emocionales ocurren entre 1-5 segundos después del inicio del estímulo.

En cuanto a la detección de emociones, esta señal se asocia a la excitación del sistema nervioso autónomo simpático, lo que implica que no puede ser controlada de manera consciente. Por lo mismo, esta señal es un reflejo de la excitación emocional, dado que ofrece un avistamiento a los procesos fisiológicos y psicológicos de la persona [48]. En [49] el autor utiliza la señal GSR como indicador del nivel de excitación experimentado por el usuario.

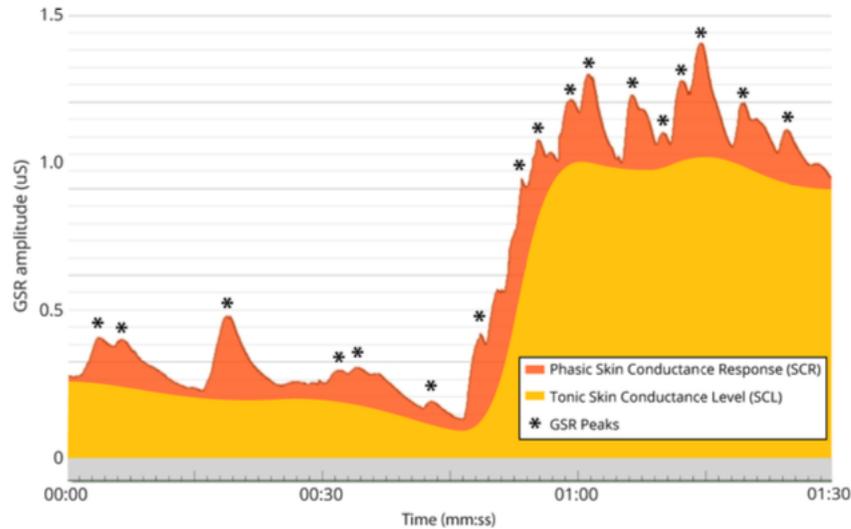


Figura 2.5: Señal GSR, con sus componentes fásica (naranja) y tónica (amarilla) diferenciadas

2.4.5. Actividad del sistema cardiovascular

El sistema cardiovascular consiste del corazón, una bomba, y la vasculatura, el sistema de distribución, que juntos se aseguran que la sangre alcance todos los tejidos del cuerpo. El corazón provee con un flujo constante de sangre oxigenada, enviando sangre a los pulmones (circulación pulmonar) y luego al resto del cuerpo (circulación sistémica) [30].

Para medirlo se utiliza un electrocardiograma (ECG), que consiste en un sistema que mide la actividad eléctrica del corazón. Para esto, se disponen 3 o más electrodos que se pegan en el dorso del sujeto y que recopilan los datos necesarios para obtener las ondas que describen el ciclo cardíaco.

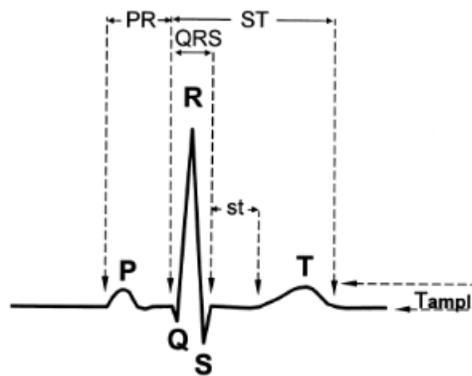


Figura 2.6: Ciclo cardíaco ideal

De esta señal (mostrada en la Figura 2.6, de [30]) se pueden sacar varias medidas:

- Periodo cardíaco: tiempo en milisegundos entre dos beats adyacentes. En general se mide el tiempo entre dos ondas R consecutivas, dado que es la de mayor amplitud.

Esta medida es recíproca al ritmo cardíaco (HR) [bpm], que es la cantidad de latidos por minuto [30].

- Variación del ritmo cardíaco (HRV): se refiere a la oscilación en el intervalo temporal entre dos beats consecutivos del corazón. Se refiere tanto a la variación del ritmo cardíaco instantáneo como de los intervalos RR. Se mide de distintas maneras, las que incluyen análisis en el dominio del tiempo y de la frecuencia [50].
- Arritmia del seno respiratorio (RSA): el ritmo cardíaco aumenta durante la inspiración y decrece durante la exhalación, por lo que esta medida se utiliza como un índice del control automático del sistema nervioso en el corazón [30]
- Intervalo entre beats (IBI).

Se ha demostrado que para tareas que involucran la manipulación de información, existe un aumento en el ritmo cardíaco [51].

En cuanto a la medición de emociones, diferencia entre emociones positivas y negativas. En particular, la variación del ritmo cardíaco (HRV), es útil para diferencias entre estrés y esfuerzo mental en adultos [28]. En [49] el autor utiliza el ritmo cardíaco como indicador del nivel de valencia experimentado por el usuario.

2.4.6. Temperatura corporal

Los cambios en la temperatura corporal reflejan cambios localizados en el flujo de sangre, causado por resistencia vascular o presión arterial. Ambas causas son mediadas por el sistema nervioso autónomo, por lo que se puede asociar a un indicador de cambios emocionales en el individuo donde se mide [16].

Para medirlo, se utilizan distintas técnicas: termografía (mide la temperatura mediante las ondas infrarrojas emitidas por el cuerpo, se presenta como una representación gráfica de la temperatura) o mediante un termistor (mide la temperatura por conducción).

En [52] se mide la temperatura del dedo índice en 3 experimentos. Concluyen que la temperatura del dedo disminuye durante situaciones de estrés y aumenta durante condiciones de relajación. Por otro lado, en [53] utilizan la medición de la temperatura facial mediante termografía. Se concluye que la temperatura de la nariz de un usuario varía con el nivel de carga mental experimentada.

2.5. Procesamiento y clasificación de los datos

Para el procesamiento de los datos utilizando algoritmos de aprendizaje de máquinas, existe un *pipeline*, tal que de muchos datos (en este caso señales) sea posible extraer información útil y entendible para humanos. En el caso de señales biológicas, éste sigue los siguientes pasos [28]:

1. Elaborar base de datos: independiente de lo que se quiera clasificar, se necesita el diseño de un experimento que permita la extracción de señales significativas. Por ejemplo, en el caso de la clasificación de emociones la emoción inducida o medida debe ser genuina, dado que se trata de reacciones del sistema nervioso autónomo.
2. Pre-procesamiento de los datos: dado que se trata de señales biológicas, las señales deben limpiarse de ruidos e interferencias externas (generadas por la electrostática de los aparatos utilizados o por el movimiento de los músculos). Para esto, se pueden utilizar filtros, segmentación de la señal, remuestreo de la señal, normalización de valores u otra técnica.
3. Extracción de características: puede ser información estadística de la señal o información en el dominio del tiempo, de la frecuencia, tiempo-frecuencia o de otro tipo.
4. Reducción de características: las características extraídas pueden o no tener correlación con la información que se busca extraer. Estas características no correlacionadas bajan el rendimiento del algoritmo. Para esto, existen distintos algoritmos. En particular, interesa la eliminación recursiva de características (RFE)[54].
5. Clasificación: en esta etapa las características seleccionadas son utilizadas para entrenar al clasificador. Existen distintos tipos de clasificadores, entre ellos, supervisados y no supervisados. En esta memoria se utilizará SOM (*Self Organizing Map*) y clustering aglomerativo [55] como algoritmos no supervisados, para la identificación de niveles de carga cognitiva y valencia/excitación y, luego, con las etiquetas obtenidas se utilizarán clasificadores supervisados como SVM (*Support Vector Machine*) [56], redes neuronales [57] o k vecinos más cercanos [58].

A continuación, se explica en detalle las dos últimas secciones: la selección de características y técnicas de clasificación.

2.5.1. Selección de características

En todos los problemas de clasificación, identificar las características que mejor describen los datos de entrada es crítico para minimizar el error de clasificación. Además, estas técnicas permiten reducir la dimensionalidad del problema, lo que implica menor complejidad computacional y menores requisitos de almacenamiento de datos. También, permiten un mejor conocimiento sobre los datos o, simplemente, una mejor visualización de estos [54]. Entregarle datos menos redundantes implica que el clasificador tendrá menos oportunidades de tomar decisiones basadas en ruido, por lo que se evita el sobreajuste y se aumenta la exactitud.

2.5.1.1. Eliminación recursiva de características (RFE)

La eliminación recursiva de características (RFE, Recursive Feature Elimination) es un procedimiento de eliminación hacia atrás, es decir, se comienza con todas las características y se eliminan progresivamente las menos útiles. Puede producir buenos resultados pero a expensas de conjuntos de características posiblemente más grandes. El método RFE es un proceso iterativo que elimina características según un criterio [18].

Consiste en 3 pasos:

1. Entrenar un clasificador.
2. Calcular el criterio de puntuación.
3. Eliminar las características con menor puntaje.

- **RFE-SVM** [54]: En este caso se usa SVM como clasificador (ver Sección 2.5.3.4). La idea es definir la importancia de una característica en términos de su contribución al SVM según una función de decisión. En cada paso del procedimiento RFE, el SVM es entrenado con el conjunto de datos dado, se calcula la función de decisión D (ver Ecuación 2.1), y la característica con menor contribución se descarta. SVM debe ser re-entrenado en de cada iteración. Por lo tanto, el método se reduce a la simple eliminación iterativa de la característica x_i con el menor peso en valor absoluto $|w_i|$. Los pesos w_i de la función de decisión se obtienen de un subconjunto de ejemplos de entrenamiento llamados vectores de soporte. La función de decisión se define como:

$$D(x_i) = w_i \cdot x_i + b \quad (2.1)$$

- **RFE-RF** En [59] utilizan RFE con Random Forest (RF) como clasificador para la selección de características para clasificar el agotamiento mental en EEG.

Random Forest consiste en la creación de muchos árboles de decisión al azar. Para clasificar un nuevo objeto desde un vector de entrada, el vector se evalúa en cada uno de los árboles del bosque. Cada árbol clasifica el vector y se dice que cada árbol “vota” por esa clase. El bosque elige la clase que tiene más votos entre todos los árboles. Para la construcción de los árboles, se utiliza un conjunto *bootstrap* (datos obtenidos con remuestreo estadístico) de, aproximadamente, dos tercios de los datos totales. El tercio restante se deja afuera como casos *out-of-bag* (OOB).

Cada árbol crece de la siguiente manera:

1. Si hay M variables de entrada, se especifica un número $m \ll M$ tal que para cada nodo, se seleccionan m variables al azar de las M y las mejores se utilizan para dividir el nodo. El valor de m es constante durante el crecimiento del bosque
2. Cada árbol crece lo máximo posible. Así, se mantiene un bajo sesgo pero alta varianza.

Una de las características de RF es que es un clasificador fuerte, con baja varianza y sesgo, construido por árboles débiles (con baja correlación, bajo sesgo y alta varianza).

Otra característica de RF es que puede entregar una medición de importancia de cada característica, según su contribución al desempeño de la clasificación. La contribución de una característica se calcula comparando el rendimiento de los árboles de decisión

en los casos OOB originales y el obtenido a partir de los casos OOB permutados con el contenido de esa característica enmascarada (mediante permutación aleatoria de los valores de esa característica). Se calcula de la siguiente manera:

$$\text{IMP}(m) = \frac{\sum_i (N_i - N_i^{m\text{-permutados}})}{N_{\text{tree}}}, \quad (2.2)$$

donde N_i y $N_i^{m\text{-permutados}}$ denotan el número de votos emitidos por la correcta clasificación del árbol i -ésimo de las clases OOB antes y después de la permutación de la característica m -ésima y N_{tree} corresponde al número total de árboles del bosque.

El coeficiente de importancia (ecuación 2.2) es muy inestable para ser usado como el criterio de elección de características en el algoritmo de RFE. Por esto, se define el siguiente criterio, que es el z-score de la medida $\text{IMP}(\cdot)$:

$$C_t(m) = \frac{\text{IMP}(m)}{\frac{\sigma_m}{\sqrt{N_{\text{tree}}}}}, \quad (2.3)$$

donde σ_m corresponde a la desviación estándar de las disminuciones en el número de votos para la clase correcta debido a la permutación de la característica m -ésima [59].

2.5.2. Clustering

En el *clustering* o análisis exploratorio de los datos, no existen etiquetas de clases disponibles. La meta del *clustering* es separar datos finitos, no etiquetados en un conjunto finito y discreto de estructuras de datos ocultas “naturales”. Así, una razón directa para la clasificación no supervisada, es explorar la naturaleza desconocida de los datos que se integran con poca o ninguna información previa [60].

2.5.2.1. Clustering Aglomerativo

Los algoritmos de *clustering* aglomerativo parten con N clusters, donde cada uno incluye exactamente un dato. Luego prosigue una serie de operaciones de unión que, eventualmente, lleva a todos los objetos a un *cluster*. El algoritmo general se resume así [60]:

1. Se empieza con N *clusters* con un dato. Se calcula la matrix de proximidad (basada en alguna función de distancia) para los N *clusters*.
2. Dentro de la matriz de proximidad, se busca la mínima distancia $D(C_i, C_j) = \min_{1 \leq m, l \leq N, m \neq l} D(C_m, C_l)$, con $D(\cdot, \cdot)$ la función de distancia, y se combina el *cluster* C_i con el *cluster* C_j para formar el nuevo *cluster* C_{ij} .
3. Se actualiza la matriz de distancias, calculando las distancias entre el *cluster* C_{ij} y el resto de los *clusters*.
4. Se repite el paso 2 y 3 hasta que solo quede un *cluster*.

Existen variantes del algoritmo donde el paso 4 se realiza hasta que quede un número k de *clusters*, esto dado que la manera de elegir el número de *clusters* para el algoritmo original,

es la inspección visual de un dendograma o gráfico de árbol que muestra la evolución de la unión de los *clusters* y resulta poco viable para un gran número de *clusters*.

Definiciones de distancias

La unión de un par de *clusters* para la formación de uno nuevo depende de la definición de la función de distancia entre dos *clusters*. Se resumen algunas a continuación [60]:

- El algoritmo de enlace simple: la distancia entre un par de *clusters* está determinada por los dos objetos más cercanos en los diferentes *clusters*. Por esto también es llamado el método del vecino más cercano. Este método tiende a generar *clusters* elongados, lo que da paso al “efecto cadena” donde dos *clusters* distintos pueden estar unidos debido a la existencia de ruido. Para datos limpios y separados entre sí, este método funciona bien.
- El algoritmo de enlace completo: este método utiliza la distancia entre los elementos más distantes entre sí en los distintos *clusters*. Es efectivo en descubrir *clusters* pequeños y compactos.
- El algoritmo de enlace de centroides: dos *clusters* son unidos en base a la distancia entre sus centroides (que se define como el promedio de todos sus elementos).
- El método de Ward o método de varianza mínima: tiene como objetivo minimizar el aumento de la suma del error cuadrático intra-clase,

$$E = \sum_{k=1}^K \sum_{x_j \in C_k} \|x_i - m_k\|^2, \quad (2.4)$$

donde $\|\cdot\|$ corresponde a la norma euclidiana, K al número de *clusters* y m_k al centroide del *cluster* C_k creado por la unión de dos *clusters*.

Los algoritmos de enlace simple y completo son llamados métodos de grafos, dado que consideran todos los puntos de un par de *clusters* cuando calculan su distancia intra-*cluster* mientras que los otros son considerados métodos geométricos dado que usan los centro geométricos de los *clusters* para representarlos y calcular su distancia.

Métodos para evaluar el número de clusters formados

Dado el algoritmo de los clusters aglomerativos, se debe tener algún tipo de parámetro para elegir el número K de clusters formados. Existen diversos índices para la evaluación de la calidad de los clusters formados: entre ellos se encuentra la inspección visual, la utilización de conocimiento experto o métodos más analíticos, como la utilización de índices. En particular, interesan dos:

- Índice de silueta [61]: se calcula según la ecuación 2.5.

$$S = \frac{b - a}{\max(a, b)} \quad (2.5)$$

con a el promedio de las distancias intra-cluster, b la distancia entre un punto y el cluster más cercano al cual no pertenece.

Para este índice su mejor valor es 1 y su peor valor es -1. Los valores cercanos a 0 indican que existen clusters que se traslapan. Valores negativos indican que una muestra ha sido asignada al cluster erróneo, dado que hay un cluster más similar.

- Índice de Calinski-Harabasz (CH) [18]: es una relación entre la varianza entre-cluster y la variación total intra-cluster. Se calcula según la ecuación 2.6.

$$\text{CH}(k) = \frac{\text{traza}(B/k - 1)}{\text{traza}(W/N - k)} \quad (2.6)$$

donde B es la matriz de la suma de los errores cuadráticos entre diferentes clusters (entre-clusters) y W es la diferencia al cuadrado de todos los objetos de un cluster con respecto a su centroide (intra-cluster).

Se busca maximizar el índice de $\text{CH}(k)$ a la hora de elegir un k óptimo en una clusterización.

2.5.2.2. Self-organizing Maps (SOM)

El algoritmo SOM crea un set de vectores de prototipo que representan los datos y lleva a cabo una proyección con preservación de topología de los prototipos desde el espacio de entrada multidimensional a una grilla de dimensión más baja (en general de dos dimensiones). Así, SOM se presenta como una herramienta de visualización de datos de alta dimensionalidad, revelando diferentes características de los datos como, por ejemplo, la existencia -o no- de clusters. Sin embargo, la visualización de los datos solo provee información cualitativa. Para producir información cuantitativa, es necesario seleccionar grupos de unidades del mapa.

Fue planteado por Kohonen y es un tipo de red neuronal artificial no supervisada, donde las neuronas son entrenadas a partir de aprendizaje competitivo, es decir, las neuronas compiten entre sí para ser activadas, resultando en la activación de solo una.

Clusterización del SOM

En [55] plantean la clusterización del SOM ya obtenido, mediante algoritmos de clustering aglomerativos o de partición (se parte con todo el espacio con un cluster y se va separando en distintos clusters, según algún criterio). La Figura 2.7 (traducida de [55]) resume este proceso.

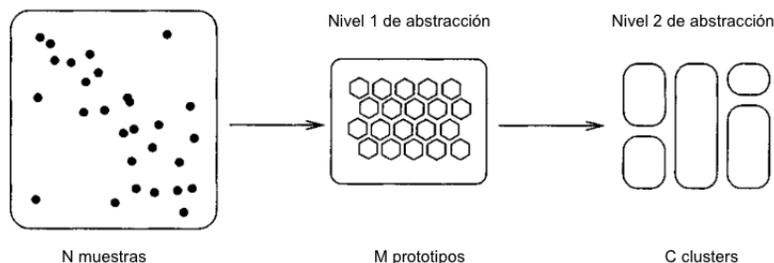


Figura 2.7: El primer nivel de abstracción es obtenido al crear un set de vectores prototipos (por ejemplo con SOM). El segundo nivel se crea al clusterizar el SOM

Primero, se forma un número grande de prototipos - mucho más grande que el número esperado de clusers - utilizando SOM u otro algoritmo de cuantización vectorial. Los prototipos

son interpretados con como “proto-clusters”, que luego, en un siguiente paso, son combinados para formar los clusters finales. Cada dato de los datos originales pertenece al mismo cluster que su prototipo más cercano [55].

En [55] se evalúan los clusters obtenidos con el índice de David-Bouldin y, luego, se comparan con los obtenidos solo por clusterizar. Se concluye sobre la utilidad de esta técnica propuesta como método de clusterización, comparando los clusters formados por SOM+clustering con los formados solamente por clustering, y reducción del tiempo computacional dedicado a esto.

2.5.3. Clasificación

En la clasificación se realiza un mapeo desde un conjunto de vectores de entrada, denotados como $x \in R^d$, con d la dimensionalidad del espacio de entrada, a un set finito de etiquetas de clases discretas, representadas como $y \in 1, \dots, C$, con C es el número total de clases, que es modelado en términos de alguna función matemática $y = y(x, w)$ con w un vector de parámetros ajustables. Los valores de este parámetro son determinados (optimizados) por un algoritmo de aprendizaje inductivo (también llamado inductor), cuyo objetivo es la minimización de una función de costo en un set finito de datos de entrada-salida (x_i, y_i) , $i = 1 \dots N$, donde N es la cardinalidad finita del conjunto de datos representativos disponibles. Cuando el inductor converge o termina, se genera un clasificador inductivo [60].

2.5.3.1. Evaluación del clasificador

Para evaluar el desempeño del clasificador existen distintas métricas, en particular, interesan:

- Exactitud (*accuracy*): calcula la fracción de predicciones correctas. Se calcula según la ecuación 2.7

$$\text{exactitud}(y_i, \hat{y}_i) = \frac{1}{N} \sum_{i=0}^{N-1} I(y_i = \hat{y}_i), \quad (2.7)$$

con N = número de muestras, y_i = etiqueta asignada para la muestra por el clasificador, \hat{y}_i = etiqueta real de la muestra. Esta métrica es la más usada, pero tiene la desventaja de que cuando las clases no están balanceadas (en cuanto a número de muestras), al optimizar el modelo según esta muestra éste priorizará clasificar las muestras como las de la clase más grande. Así, siempre tendrá predicciones correctas en la clase de mayor tamaño y siempre se equivocará en las más pequeñas; pero dado el desbalance existente, la cantidad de predicciones correctas hará que la métrica tenga un valor cercano a 1.

- Precisión (*precision*): intuitivamente es la habilidad de un clasificador de no etiquetar una muestra como positiva cuando es negativa. Es decir, la habilidad de disminuir los falsos positivos de una clase. Se calcula según la ecuación 2.8

$$\text{precisión} = \frac{VP}{VP + FP}, \quad (2.8)$$

con VP = verdaderos positivos y FP = falsos positivos.

- Sensibilidad (*recall*): intuitivamente es la habilidad del clasificador de encontrar todas las muestras de una clase, es decir, es la habilidad de disminuir los falsos negativos de una clase. Se calcula según la ecuación 2.9.

$$\text{sensibilidad} = \frac{VP}{VP + FN}, \quad (2.9)$$

con FN = falsos negativos.

- F1: es la media armónica entre precisión y sensibilidad, considerando ambos valores al entregar un puntaje. Se calcula según la ecuación 2.10

$$F1 = \frac{(1 + \beta^2)\text{precisión} \cdot \text{sensibilidad}}{(\beta^2 \cdot \text{precisión}) + \text{sensibilidad}}, \quad (2.10)$$

donde β ajusta el peso que se le da a cada término. Al elegirla como 1 se les da igual importancia. Esta métrica, al considerar tanto los FP como FN, se vuelve útil a la hora de enfrentar un desbalance de clases, puesto que impide que todo se clasifique como la clase de mayores muestras, puesto que eso disminuiría su precisión y, por ende, F1.

Para obtener el desempeño del clasificador (dado que estas métricas son para cada clase clasificada), se calcula la métrica para cada clase y luego se promedian todas, sin considerar el desbalance de clases que pueda existir.

2.5.3.2. Clasificadores

2.5.3.3. K-vecinos más cercanos (KNN)

Este clasificador “mira” los k puntos más cercanos del set de entrenamiento con la entrada de prueba x , cuenta cuántos miembros de cada clase están en este sub-set, y retorna la fracción empírica como el estimado, como se muestra en la Figura 2.8, donde ese ejemplifica el funcionamiento de KNN para 2d y $K = 3$ [58].

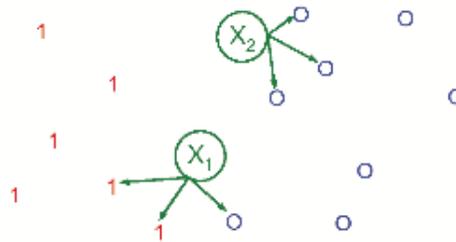


Figura 2.8: Ilustración de una clasificación mediante KNN, para $K=3$. Se toman los 3 vecinos más cercanos y, de acuerdo a estos, se entrega una probabilidad de pertenencia a una clase de X_1 y X_2

Formalmente:

$$P(y = c | x, D, K) = \frac{1}{K} \sum_{i \in N_k(x, D)} \mathbb{I}(y_i = c), \quad (2.11)$$

con $N_k(x, D)$ son los (índices de los) k puntos más cercanos a x en D e $\mathbb{I}(t)$ es la función definida como

$$\mathbb{I}(t) = \begin{cases} 1 & \text{si } t \text{ es verdadero} \\ 0 & \text{si } t \text{ es falso} \end{cases} \quad (2.12)$$

En la Figura 2.8, los 3 vecinos más cercanos del punto X_1 tienen etiqueta 1, 1, y 0 por lo que se predecirá $P(y = 1 | X_1, D, K = 3) = 2/3$. Para el punto X_2 los 3 vecinos más cercanos tienen etiquetas 0, 0, 0 por lo tanto se predice $P(y = 1 | X_2, D, K = 3) = 0/3$

2.5.3.4. Máquinas de soporte vectorial (SVM)

El algoritmo de máquinas de soporte vectorial (SVM) se utiliza para clasificar, basándose en la distancia mínima que poseen los elementos más cercanos al hiperplano separador (llamada margen). Se busca aquel hiperplano que otorga el máximo margen de separación entre clases (hiperplano óptimo).

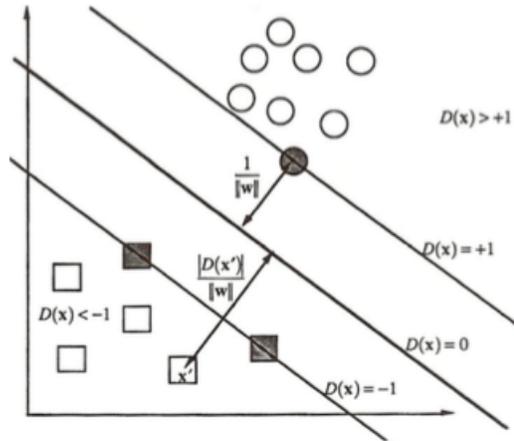


Figura 2.9: Esquemático de SVM linealmente separable

Para el caso en que **los datos son linealmente separables** (Figura 2.9), se puede observar como el hiperplano separador es aquel para el cual $D(x)=0$, y el margen está definido por los elementos de ambas clases que están más cercanos de éste. A estos elementos se les llama vectores de soporte, y se ven caracterizados por $D(x) = \pm 1$, con $D(x) = (w \cdot x) + b$. De esta manera, todos los puntos con $y_i = 1$ quedan en un lado del plano (y tienen $D(x) > 0$), mientras que todos los puntos con $y_i = -1$ quedan al otro lado (y tienen $D(x) < 0$). Por lo tanto, se busca un w (un vector normal al hiperplano) y b (un offset). En general existe más de un hiperplano capaz de realizar la separación, por lo que se busca aquel que: (1) Minimiza el error en la separación de las muestras (error de clasificación); (2) Maximiza el margen de separación (generalización del clasificador).

Este problema puede escribirse como uno de optimización:

$$\begin{aligned} \min \quad & \frac{1}{2} w \cdot w \\ \text{s.a} \quad & y_i(w \cdot x_i + b) \geq 1, \quad i = 1, \dots, m \end{aligned} \quad (2.13)$$

donde w corresponde al máximo vector normal al hiperplano, x_i corresponde a un vector de características que describe un punto de los datos e y_i corresponde a su etiqueta.

Para el caso en que **los datos no son linealmente separables** (no existe un plano que los separe), existe el llamado SVM de margen relajado. Se introduce la variable ξ_i para cada punto x_i . Si el punto es aquel que puede ser separado, entonces $\xi_i = 0$. Si no, $\xi_i > 0$ es el valor de la discrepancia, que se puede expresar con la siguiente inecuación

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i \quad (2.14)$$

Se busca un trade-off entre hacer los ξ_i pequeños y hacer la distancia entre márgenes lo más grande posible. Luego, el problema de optimización a resolver es:

$$\begin{aligned} \min \quad & \frac{1}{2} w \cdot w + C \sum \xi_i \\ \text{s.a} \quad & \xi_i \geq 0, \\ & y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, m \end{aligned} \quad (2.15)$$

Para elegir el valor de C ($0 < C < \infty$), se varía desde una solución que obtiene una distancia entre márgenes muy ancha (sin importar cuántos puntos hay adentro o en el lado incorrecto) a una solución que se conforma con un margen inexistente. El primer caso es menos preciso para datos de entrenamiento pero más robusto en datos nuevos; el segundo es lo más preciso posible en datos de entrenamiento pero frágil para datos nuevos.

El último caso, es cuando **los datos no son lineales**. Para estos problemas (Figura 2.10, de [56]), se transforma el espacio mediante la utilización de un kernel (función no lineal), resultando en que el problema es linealmente separable dentro del espacio del Kernel. La idea básica (mostrada en la Figura 2.10) es que una superficie de separación no lineal en n -dimensiones puede mapearse (o aproximarse) por un hiperplano lineal en N -dimensiones [56]. Se muestran los kernels más utilizados en la Tabla 2.1.

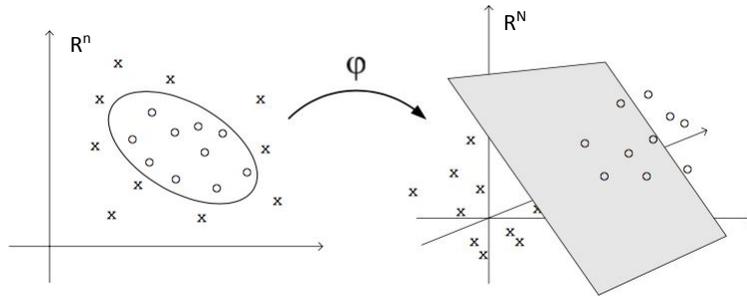


Figura 2.10: Caso SVM no lineal: gracias a la función φ el espacio se transforma y los datos son linealmente separables

Tabla 2.1: Kernels más utilizados

Lineal	$K(x_i, x_j) = x_i \cdot x_j$
Polinomial	$K(x_i, x_j) = (ax_i \cdot x_j + b)^d$
Sigmoideo	$K(x_i, x_j) = \tanh(ax_i \cdot x_j + b)$
Función gaussiana de base radial (RBF)	$K(x_i, x_j) = \exp(-\frac{1}{2} x_i - x_j ^2/\sigma^2)$

SVM se define como un clasificador binario, es decir, para clasificar entre dos clases. Existen dos aproximaciones para una **clasificación multiclase**: la primera se basa en construir y combinar varios clasificadores binarios mientras que la segunda se basa en considerar todos los datos en una fórmula de optimización [62]:

- **Uno contra todos**: se construyen k modelos de SVM, con k el número de clases. El modelo m es entrenado con todos los ejemplos en la clase m con etiqueta positiva y todo el resto de las clases con etiqueta negativa. Así dado l datos de entrenamiento $(x_1, y_1), \dots, (x_l, y_l)$, con $x_i \in R^n, i = 1..l$ y $y_i \in 1, \dots, k$ es la clase de x_i , el modelo m de SVM resuelve el siguiente problema:

$$\begin{aligned} \min_{w^m, b^m, \xi^m} \quad & \frac{1}{2}(w^m)^T w^m + C \sum_{i=1}^l \xi_i^m \\ & (w^m)^T \phi(x_i) + b^m \geq 1 - \xi_i^m, \text{ si } y_i = m, \\ & (w^m)^T \phi(x_i) + b^m \leq -1 + \xi_i^m, \text{ si } y_i \neq m, \\ & \xi_i^m \geq 0, i = 1, \dots, l, \end{aligned} \quad (2.16)$$

donde los datos de entrenamiento x_i están mapeados a un espacio dimensional mayor por la función ϕ y C es el parámetro de castigo.

Minimizar $\frac{1}{2}(w^m)^T w^m$ significa que se quiere maximizar $2/\|w^m\|$, que corresponde al margen entre dos grupos de datos. Cuando los datos no son linealmente separables, existe un castigo $C \sum_{i=1}^l \xi_i^m$ que reduce el número de errores de entrenamiento. El concepto básico de SVM es buscar el balance entre el término de regularización $\frac{1}{2}(w^m)^T w^m$ y los errores de entrenamiento. Después de resolver (2.16), existen k funciones de decisión:

$$\begin{aligned} & (w^1)^T \phi(x) + b^1, \\ & \vdots \\ & (w^k)^T \phi(x) + b^k. \end{aligned} \quad (2.17)$$

Se dice que x está en la clase que tiene el mayor valor de la función de decisión

$$\text{clase de } x \equiv \operatorname{argmax}_{m=1, \dots, k} ((w^m)^T \phi(x) + b^m). \quad (2.18)$$

Prácticamente, se resuelve el problema dual de 2.16, donde el número de variables es el mismo que el número de datos en 2.16. Así, se resuelven k problemas de programación cuadrática de l -variables

- **Uno contra uno**: este método construye $k(k-1)/2$ clasificadores donde cada uno es entrenado con datos de dos clases. Para datos de entrenamiento de las clases i y j , se resuelve el siguiente problema:

$$\begin{aligned} \min_{w^{ij}, b^{ij}, \xi^{ij}} \quad & \frac{1}{2}(w^{ij})^T w^{ij} + C \sum_t \xi_t^{ij} \\ & (w^{ij})^T \phi(x_t) + b^{ij} \geq 1 - \xi_t^{ij}, \text{ si } y_t = i, \\ & (w^{ij})^T \phi(x_t) + b^{ij} \leq -1 + \xi_t^{ij}, \text{ si } y_t = j, \\ & \xi_t^{ij} \geq 0. \end{aligned} \quad (2.19)$$

Existen distintos métodos para el testeo de los $k(k-1)/2$ clasificadores. En [62] utilizan una estrategia de votos: si $\text{sign}((w^{ij})^T \phi(x) + b^{ij})$ dice que x está en la clase i , entonces se le suma un punto a la clase i . Sino, se le suma un voto a la clase j . Luego, se predice que x está en la clase con el mayor número de votos. En el caso que dos clases tengan la misma cantidad de votos, se elige aquella con un índice menor.

Prácticamente se resuelve el problema dual de 2.19, donde su cantidad de variables es la misma que el número de datos en ambas clases. Así, si en promedio cada clase tiene l/k puntos de datos, se deben resolver $k(k-1)/2$ problemas de programación cuadrática, donde cada uno tiene aproximadamente $2l/k$ variables.

2.5.3.5. Redes neuronales artificiales (RNA)

Un perceptrón o neurona (Figura 2.11, de [63]) es la unidad básica de una red neuronal artificial. Cuenta con tres elementos básicos :

- Pesos sinápticos y bias que ponderan las entradas
- Sumador o combinador lineal
- Función de activación no-lineal, que limita la amplitud de la salida.

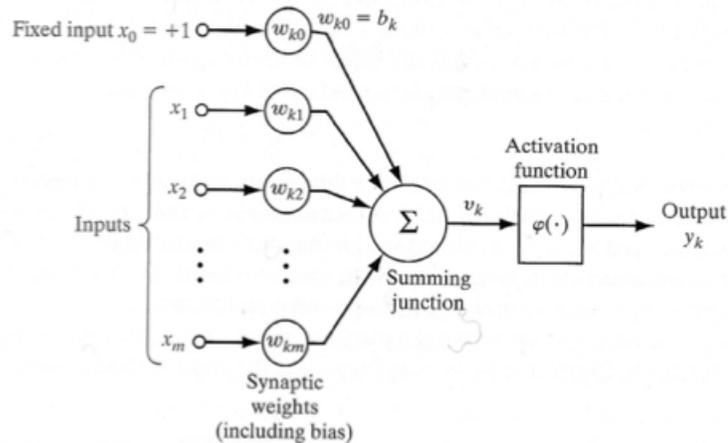


Figura 2.11: Modelo de perceptrón

La ecuación de una neurona es

$$\begin{aligned} S_k &= \sum_{j=0}^n w_{kj} x_j \\ y_k &= \varphi(s_k), \end{aligned} \quad (2.20)$$

donde x_j es la entrada j -ésima, w_{kj} es el peso sináptico de la neurona k asociada a la entrada j , s_k es la salida del combinador lineal e y_k es la salida no-lineal de la neurona. Además, se define x_0 como un input fijo en 1 y $w_{k0} = b_k$ es el bias de la neurona.

Las redes MLP consisten de un sistema de neuronas o nodos interconectados (Figura 2.12, de [63]). Este modelo representa un mapeo no lineal entre un vector de entrada (x_i) y uno

de salida (y_i). Los nodos están conectados por pesos (w_{ij}^k) y funciones de salida, que son una función de la suma de las entradas del nodo modificadas por una función de activación. Es la superposición de muchas funciones de activación no lineales que permite que el MLP aproxime funciones no lineales. La salida de un nodo es escalada por su peso asociado y es la entrada del nodo de la siguiente capa. Esto implica una dirección del procesamiento de la información (por lo que son conocidos como redes neuronales *feed-forward*) [57]. El número de entradas depende de la cantidad de elementos que tenga el vector que se desea clasificar, el número de salidas depende de las clases a clasificar y el número de neuronas en la capa oculta es un parámetro de diseño.

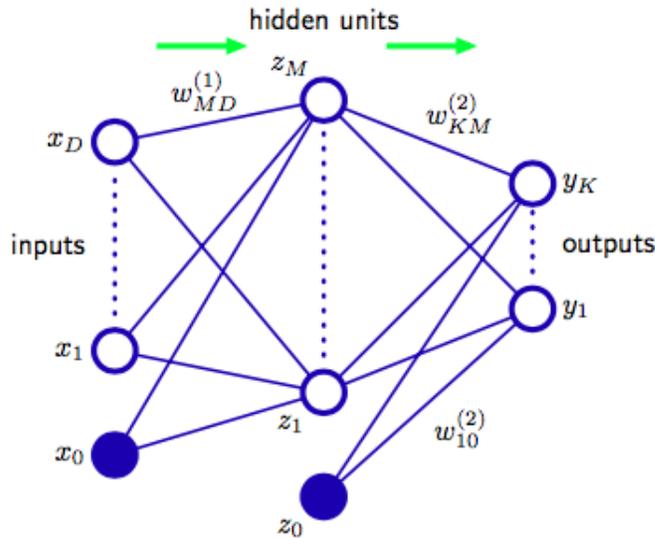


Figura 2.12: Esquema RNA MLP: diagrama para red de dos capas. Las variables de entrada, ocultas y de salida son representadas por nodos y los pesos son representados como links entre los nodos, donde el bias se representa como un link que viene de variables adicionales. Las flechas muestran la dirección de la información durante la propagación hacia adelante.

El modelo básico de una red neuronal se describe como una serie de transformaciones funcionales. Primero, se contruyen M combinaciones lineales de las variables de entrada:

$$a_j = \sum_{i=0}^D w_{ji}^{(1)} x_i \quad (2.21)$$

donde $j = 1, \dots, M$ y el superíndice (1) indican que los parámetros están en la primera capa de la red. Luego, cada uno de ellos es transformado utilizando una función de activación $h(\cdot)$, que es diferenciable y no lineal

$$z_j = h(a_j) \quad (2.22)$$

Esto corresponde a la salida de las unidades ocultas. Luego, estos valores son combinados linealmente para dar la salida de la unidad de activación

$$a_k = \sum_{j=0}^M w_{kj}^{(2)} z_j \quad (2.23)$$

donde $k = 1, \dots, K$ y K es el número total de salidas. Esta transformación corresponde a la segunda capa de la red. Finalmente, la salida de la unidad de activación son transformadas

utilizando una función de activación apropiada ($\sigma(\cdot)$), para dar un set de variables de salida tal que $y_k = \sigma(a_k)$.

Estas etapas pueden combinarse para obtener la función de la red

$$y_k(x, w) = \sigma \left(\sum_{j=0}^M w_{kj}^{(2)} h \left(\sum_{i=0}^D w_{ji}^{(1)} x_i \right) \right) \quad (2.24)$$

Así, se puede observar que una red neuronal es una función no lineal desde un set de variables de entrada x_i hacia un set de variables de salida y_k controlado por un vector w de parámetros ajustables.

La arquitectura recién descrita (y mostrada en la Figura 2.12) es la que más se usa en la práctica. Sin embargo, es fácilmente generalizable, por ejemplo al considerar capas adicionales de procesamiento, cada una que consiste en una combinación lineal de la ecuación 2.23 seguidas de una transformación por elemento utilizando una función de activación no lineal.

Las redes MLP tienen la habilidad de aprender a través del entrenamiento. Para esto se requiere un set de datos de entrenamiento, que consiste de una serie de vectores de entrada con sus salidas asociadas (entrenamiento supervisado). Durante el entrenamiento, se le presenta el set de entrenamiento a la red y ésta va ajustando sus pesos hasta que se encuentra el mapeo salida-entrada deseado (dado un margen de error). Una vez obtenido esto, la red puede generalizar para un set de datos nuevos de entrada [57].

Algunos parámetros para tener en consideración:

- **Funciones de activación:** su elección es determinada por la naturaleza de los datos y la distribución asumida de las variables etiquetadas.

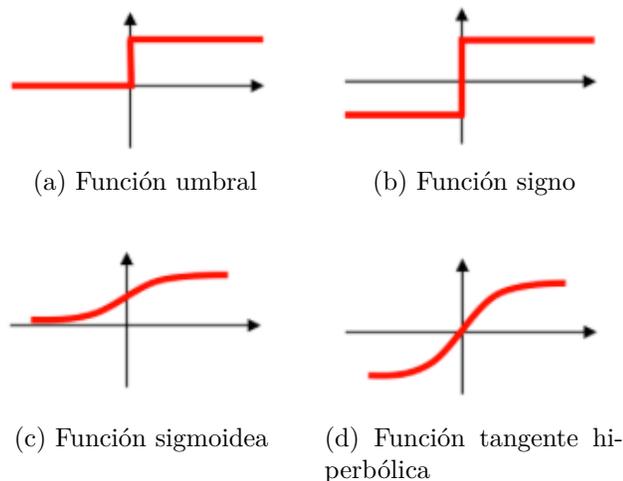


Figura 2.13: Funciones de activación

– Función umbral o heaviside (Figura 2.13(a)): $\phi(u) = \begin{cases} 1 & \text{si } u \geq 0 \\ 0 & \sim \end{cases}$

– Función signo (Figura 2.13(b)): $\phi(u) = \begin{cases} 1 & \text{si } u \geq 0 \\ -1 & \sim \end{cases}$

– Función sigmoidea (Figura 2.13(c)): $\phi(u) = \frac{1}{1+e^u}$

– Función tangente hiperbólica (Figura 2.13(d)): $\phi(u) = \frac{e^u-1}{e^u+1}$

- **Cantidad de neuronas en la capa oculta** [18] : surge la cuestión de cuál es el número óptimo de neuronas en esta capa. Si bien la cantidad óptima de neuronas en la capa oculta se obtiene de manera heurística y dependerá de cada caso, siempre se puede partir con un número “óptimo”.

Un número elevado de neuronas en la capa oculta puede generar sobreentrenamiento, es decir, no generaliza la respuesta del clasificador, y muy pocas producirán subentrenamiento, es decir, no se alcanza a resolver el problema.

Se han encontrado dos criterios para definir el número de neuronas en la capa oculta (para casos de una capa oculta), los cuales son :

– Regla de la pirámide geométrica: $h = \sqrt{n \cdot m}$, con m la cantidad de entradas a la red y n la cantidad de salidas.

– En función del número de muestras al entrenar: $h = \log_2(T)$, con T la cantidad de muestras en el conjunto de entrenamiento.

- **Sobreajuste del modelo** [18]: sucede cuando la red neuronal logra reproducir muy bien el comportamiento de los datos de entrenamiento pero no el de datos nuevos, es decir, la red no logra generalizar la respuesta. En general, para evitar el sobreentrenamiento se utiliza un conjunto de validación luego de cada iteración en el proceso de entrenamiento para comprobar si existe sobreentrenamiento y corregirlo variando los parámetros. Durante el proceso iterativo de optimización de los parámetros de la red, se comparan los errores obtenidos con los datos de entrenamiento y con los datos de validación. En el caso de que durante sucesivas iteraciones, el error con los datos de entrenamientos disminuya, mientras que el error con los datos de validación aumente, se detiene el proceso de ajuste, como un criterio de parada adicional.

2.6. Estado del arte

A continuación, se presenta el estado del arte del tema en cuestión. Se encuentran dos grandes aristas: sobre la medición de carga cognitiva en la escritura y sobre la medición de carga cognitiva y estado emocional utilizando sensores psico-fisiológicos.

2.6.1. Carga cognitiva en el proceso de escritura

Existe literatura con respecto a la investigación de cómo afecta la carga cognitiva y los componentes de la memoria de trabajo (central ejecutiva, loop fonológico y *sketchpad* visuoespacial) al desarrollo del proceso de escritura. A continuación, se detalla sobre algunos trabajos en el tema. Para cada investigación, se indica el objetivo de ésta, el experimento realizado, la manera de medir la carga cognitiva y los resultados obtenidos.

- R. Vanderberg and H. L. Swanson [24]

Objetivo(s)	Busca investigar la relación entre los componentes de la memoria de trabajo y la macroestructura (planificación, traducción y revisión) y microestructura (gramática, puntuación..) de la escritura.
Experimento	116 estudiantes de educación media deben escribir una historia y un ensayo en el colegio. Además, se realizan pruebas que miden la memoria de corto plazo y la memoria de trabajo.
Medida de carga cognitiva	Se revisa cualitativamente ambos escritos (se asigna puntaje por distintos aspectos: estructura, puntuación, gramática, borrador..) y se realizan correlaciones entre estas pruebas y las de memoria de corto plazo y de trabajo.
Resultados	Se valida la importancia de la memoria de trabajo en el proceso de escritura. Se demuestra que el proceso de escritura está más ligado al componente de atención de la memoria de trabajo que al de almacenamiento de información.

- K. Yu, J. Epps, and F. Chen [64]

Objetivo(s)	Busca clasificar el nivel de carga mental durante el proceso de escritura, mediante la examinación de características de la letra manuscrita.
Experimento	Se muestran una o más palabras en la interface del experimento, las que desaparecen después de un tiempo especificado. Los participantes deben escribir una oración con las palabras. Se varía el número de palabras y el tiempo de exposición de éstas.
Medida de carga cognitiva	Se mide la curvatura y la velocidad de la letra manuscrita. Se obtienen características y se aplica un clasificador de ventana de Parzen. Para la clasificación, los usuarios auto-reportan la dificultad experimentada.
Resultados	La mayor precisión de clasificación obtenida es de 65.1%.

- R. A. Alves, S. L. Castro, and T. Olive [25]

Objetivo(s)	Se busca estimar la distribución de los procesos de la escritura, y sus demandas cognitivas, a través de los procesos de pausa y tipeo. Además, se busca saber como estos procesos se ven afectados por las habilidades de tipeo de los usuarios.
Experimento	Los participantes deben escribir un texto, basados en una imagen que se muestra. Mientras escriben deben reaccionar lo más rápido posible a un pito que suena cada 15-45 segs. Inmediatamente deben verbalizar la etiqueta que se ajusta mejor a la actividad que fue interrumpida.
Medida de carga cognitiva	Se mide la ocurrencia de los procesos de escritura mediante verbalización y sus demandas cognitivas se miden a través de la interferencia en el tiempo de reacción (IRT).
Resultados	El proceso de traducción fue el más reportado durante tipeo. No se puede caracterizar ninguno de los procesos de escritura como típicos del proceso de pausa. Con respecto a las demandas cognitivas, se concluye que la revisión es el proceso que más demanda. Por último, se observa que las habilidades de tipeo tienen un impacto perjudicial sobre la calidad del texto.

- L. Bourke and A.-M. Adams [65]

Objetivo(s)	Examinar el efecto de ciertos factores cognitivos (memoria de trabajo, conocimiento de vocabulario, habilidad cognitiva y habilidades de lectura) en la diferencia en el progreso en el desarrollo de la escritura en niños.
Experimento	Los niños deben escribir una historia, según una imagen que se les muestra. También hacen pruebas que miden los componentes fonológicos, visuoespaciales y de la central ejecutiva.
Medida de carga cognitiva	Se le asigna puntaje a la escritura del niño, según ciertos parámetros. Se correlacionan con los puntajes obtenidos en las pruebas cognitivas.
Resultados	Los resultados indican que el desempeño en lectura y habilidades visuoespaciales son cruciales para determinar si un niño de 4-5 años está en un nivel deseado de escritura, según los estándares del estudio.

En general, para la literatura del área de carga cognitiva y escritura, se busca entender cómo afecta la carga cognitiva el proceso de escritura. Qué componentes de la memoria de trabajo se ven activados en las distintas etapas. Se buscan correlaciones entre calidad del texto (según diversos indicadores) y desempeños asociados a tareas cognitivas específicas o asociar cierta etapa al nivel cognitivo demandado. Esto resulta bastante engorroso dado que se debe contar con expertos que lean los textos escritos y que evalúen de manera pareja y equivalente. El único trabajo presentado que busca clasificar carga cognitiva según escritura,

se basa en el análisis de la letra manuscrita de la persona: la curvatura de ésta, la rapidez de la escritura o cuánto carga el lápiz.

2.6.2. Medición del estado del usuario con sensores psico-fisiológicos

2.6.2.1. Medición de carga Cognitiva

En el último tiempo, se ha investigado bastante sobre la medición de carga cognitiva utilizando medidas psico-fisiológicas. A continuación, se presenta una revisión de distintos trabajos que tratan este tema. Ésta está basada en la revisión realizada por Hernán Lira [17] en su trabajo de memoria, pero tiene algunas modificaciones. Para cada trabajo se indica su objetivo, que sensores usa, el experimento realizado y los resultados obtenidos.

- C. Ikehara and M. Crosby [66]

Objetivo(s)	Evaluar carga cognitiva
Sensor(es) psico-fisiológicos	Eyetracker, presión del mouse, GSR, pulso, oxímetro.
Experimento	Seleccionar en la pantalla las fracciones cuyo valor es mayor a 1/3. Dos grados de dificultad (varía la velocidad de aparición de las fracciones, la dificultad de resolución de las fracciones, cantidad de fracciones en la pantalla y cantidad de tiempo disponible).
Resultados	GSR y eyetracker tienen la mayor significancia para detectar la dificultad de una tarea

- H. Schultheis and A. Jameson [9]

Objetivo(s)	Buscan evaluar si el tamaño de la pupila es un buen indicador para la carga cognitiva.
Sensor(es) psico-fisiológicos	Eyetracker y EEG
Experimento	Medir nivel de carga cognitiva mientras el usuario lee textos de distinta dificultad.
Resultados	Concluyen que no existe relación entre el tamaño de la pupila y la carga cognitiva experimentada por el usuario.

- G. Wilson and C. Russell [11]

Objetivo(s)	Determinar el nivel de precisión alcanzado por una red neuronal cuando se desea medir en tiempo real el nivel de carga mental de los usuarios al realizar una tarea compleja.
Sensor(es) psico-fisiológicos	EEG, pestañeo del ojo, ritmo cardíaco y respiración
Experimento	Los participantes deben realizar la prueba MATB de la Nasa con dos niveles de dificultad. Los participantes son entrenados por 3 días, para evitar el efecto del aprendizaje en la medición.
Resultados	Se alcanza una precisión de 85 %, 82 % y 86 % para la clasificación de tarea nivel base, fácil y difícil respectivamente.

- E. Haapalainen, S. Kim, J. Forlizzi and A. Dey [7]

Objetivo(s)	Medición de cargas cognitivas en tareas cognitivas básicas
Sensor(es) psico-fisiológicos	Eyetracker, EEG, ECG, flujo de calor y pulso
Experimento	Problemas de percepción visual y velocidad cognitiva
Resultados	ECG y flujo de calor detectan con un 80 % de precisión tareas de alta y baja demanda cognitiva

- T. Fritz, A. Begel, S. Müller, S. Yigit-Elliott and M. Züger [8]

Objetivo(s)	Detectar cuando un desarrollador de software está experimentando dificultad mientras programa.
Sensor(es) psico-fisiológicos	Eyetracker, GSR y EEG
Experimento	Medir nivel de carga cognitiva mientras los programadores corrigen códigos de distinta dificultad.
Resultados	Predicen dificultad de una tarea con 85 % de precisión

- S. Iqbal, P. Adamczyk, X. Zheng and B. Bailey [33]

Objetivo(s)	Demostrar como la carga mental de un usuario cambia durante la ejecución de una tarea
Sensor(es) psico-fisiológicos	Eyetracker
Experimento	Se realizan dos tareas: una de plan de rutas y la otra de corrección de documentos escritos
Resultados	Demuestran que diferentes tipos de subtareas imponen distinta carga mental en el usuario y que la carga mental decrece en el límite entre subtareas.

En general, en todos estos trabajos, el usuario no se encuentra realizando actividades de redacción (excepto en [33], donde no es el foco), sino, tareas de distintos ámbitos (lectura, tareas cognitivas específicas, programación, resolución de problemas matemáticos..). De estos, se observa que para medir la carga cognitiva cuando el usuario está realizando cierta tarea, luego de finalizada ésta se pide un auto-reporte. Éste puede ser libre [9][11] o más formal, como el test NASA TLX (*Task Load Index*) [10], o ambos [8]. Además, se utilizan varios sensores de interés para este trabajo: eye tracker, ECG, EEG y actividad electrodermal. Para la validación del experimento en general se utiliza un test ANOVA [66][33] y/o los resultados del test NASA TLX en correlación con algún parámetro (tiempo de completitud de la tarea, reporte subjetivo..) [10][8][7] y, cuando se utilizan algoritmos de aprendizaje de máquina, se utiliza KNN, redes neuronales [11] y SVM. Además, desde estas investigaciones previas se pueden sacar procedimientos útiles para este trabajo: realizar un pequeño grupo piloto, previo al experimento, para validar que el experimento realmente cumpla para lo que se diseñó [7][8] y normalizar las señales para eliminar las diferencias individuales entre individuos (por ejemplo, utilizando una señal de base) [8][66][47][67].

Con respecto a las otras memorias de este Fondecyt, se ha publicado un paper [68] y se está trabajando en otro [69]. En ambos se busca medir la carga cognitiva a partir de sensores psico-fisiológicos. El trabajo basado en la memoria de Hernán Lira [68], busca medir

la carga cognitiva en usuarios que hacen *multitasking* en *smartphones* (por ejemplo, mandan mensajes mientras caminan o esperan una micro), utilizando dos mediciones no invasivas: el oxígeno en la sangre y el ritmo cardiaco (ambas medidas con un pulsoxímetro). En el caso del trabajo basado en la memoria de Cristian Retamal [69], se desea medir la carga cuando el usuario está navegando por un sitio web. Para esto, se utilizan los mismos sensores que en su memoria (y que en esta): ECG, EGG, GSR, temperatura corporal, Eye tracker y pulso cardíaco. Luego, se clusteriza según el diámetro de la pupila, para obtener los distintos niveles de carga cognitiva experimentadas y, desde ahí, se clasifica. Se realizan análisis de redes neuronales con combinaciones de los distintos sensores utilizados, obteniendo que el EEG es el que mejor resultado obtiene por sí solo (88.46 % de precisión promedio).

2.6.2.2. Medición de emociones

Para lograr la clasificación de emociones en un sujeto, se han utilizado distintas medidas psico-fisiológicas. Por ejemplo, en [14] utilizan mediciones electroencefalográficas (EEG) mientras que en [16] utilizan actividad electrodermal, variación de la temperatura corporal y ritmo cardíaco. En [28] se detallan las medidas psico-fisiológicas más comunes utilizados en estudios para clasificar emociones:

1. Sistema cardiovascular: variación del ritmo cardiaco (HRV), arritmia seno respiratoria (ASR), output cardíaco (volumen bombeado por el corazón), intervalo entre beats (IBI) y presión sanguínea.
2. Actividad electrodermal: conducción de la piel, respuesta galvánica de la piel (GSR)
3. Sistema respiratorio: respiraciones por minuto, volumen de la respiración.
4. Electromiograma (actividad muscular)
5. Actividad cerebral: electroencefalograma (EEG), imágenes cerebrales.

A continuación, se presentan ciertos trabajos de interés para esta memoria (ya sea por metodología o señales usadas):

- P. Rani, C. Liu, N. Sarkar, y E. Vanman [12].

Objetivo(s)	Comparación de distintas técnicas de machine learning para clasificación de emociones (ansiedad, aburrimiento, compromiso, frustración, enojo).
Sensor(es) psico-fisiológicos	Características del sistema cardiovascular, actividad electrodermal y electromiograma.
Experimento	Participantes deben resolver anagramas y jugar pong. Se varía la dificultad de cada tarea.
Resultados	SVM da la mejor clasificación, aunque todos los métodos utilizados entregan resultados competitivos. El árbol de regresión da el segundo mejor resultado.

- A. Haag, S. Goronzy, P. Schaich, and J. Williams [15]

Objetivo(s)	Reconocer emociones mediante la utilización de bio-sensores. Buscan medir valencia y excitación.
Sensor(es) psico-fisiológicos	Electromiograma, ECG, actividad electrodermal, temperatura de la piel, pulso de volumen sanguíneo (BVP) y respiración.
Experimento	Se les muestran fotos del set IAPS (vienen etiquetadas con nivel de valencia y excitación). Se muestran 6 sets de 5 fotos que inducen la misma emoción. Se utiliza una red neuronal para la clasificación.
Resultados	Se obtienen precisión de 96.6 % y 89.9 % para el reconocimiento de excitación y valencia, respectivamente.

- C. Lee, S. Yoo, Y. Park, N. Kim, K. Jeong, and B. Lee [13]

Objetivo(s)	Estimar emociones mediante el uso de redes neuronales y los cambios en el sistema nervioso autónomo. Buscan clasificar tristeza, placer calmado, placer interesante y miedo.
Sensor(es) psico-fisiológicos	Variabilidad del ritmo cardíaco y actividad electrodermal.
Experimento	Se les muestra a los usuarios distintos videoclips (análogo a cambiar el canal de la televisión). Usuario auto-reportan sus emociones al terminar cada vídeo.
Resultados	Se alcanza una tasa total de precisión de 80.2 %, siendo la emoción “miedo” la con una tasa de clasificación más alta.

- K. Kim, S. Bang and S. Kim [16]

Objetivo(s)	Buscan clasificar el estado emocional entre 4 emociones: tristeza, enojo, estrés y sorpresa; utilizando sensores poco invasivos para el sujeto.
Sensor(es) psico-fisiológicos	Actividad electrodermal, variación de la temperatura corporal y actividad cardiovascular
Experimento	Crean 4 escenarios distintos que inducen la emoción buscada. Para esto, utilizan estímulos visuales, auditivos y cognitivos.
Resultados	Para 50 sujetos, encuentran un clasificador con 78.4 % de clasificaciones correctas para 3 emociones y 61.8 % para 4 emociones.

En la mayoría de los trabajos realizados sobre la clasificación de emociones, el experimento se basa en inducir emociones claras (ya sea con material de audio, imágenes u otra cosa) [14][15][27][16]. Sin embargo, el propósito de esta memoria es ver si dada una actividad cotidiana, es posible clasificar cómo se siente el usuario en ese instante, parecido a lo que se hace en [12] donde se miden emociones mientras el usuario juega pong o hace anagramas o en [13] donde se busca clasificar las emociones de ver vídeos al azar. Para lograr esta clasificación, es necesario tener una idea de lo que siente el sujeto mientras está realizando la tarea encomendada. En [13] buscan que el usuario auto-reporte las emociones que siente luego de ver cada videoclip y, además, utilizan un test SAM (*Self-Assessment Manikini*) para que el usuario reporte sobre el nivel de valencia/excitación experimentado. Es importante notar que solo utilizan la data de aquellos usuarios que reportaran emociones claras en SAM. En [12] utilizan solo el auto-reporte luego de cada actividad.

Por otro lado, en [28] se presenta un revisión de trabajos anteriores que desean reconocer emociones utilizando sensores. De ésta, se concluye que dado que las emociones son altamente subjetivas, un sistema general para clasificar las emociones básicas no ha sido logrado aún. La mayoría de los sistemas desarrollados es dependiente del usuario y aquellos que son independientes, les falta precisión. Esto, dado que dado un contexto, diferentes individuos expresan la misma emoción con un patrón de respuesta distinto [43][12].

Con respecto al estado del arte recién presentado se observa que existe un vacío en la literatura: no se ha investigado sobre las tareas de redacción, el proceso de la escritura y la carga cognitiva experimentada por el usuario en el mismo instante cuándo escribe ni se ha buscado medir la carga cognitiva o nivel de valencia/excitación experimentada por el usuario que escribe utilizando sensores psico-fisiológicos, que es el propósito bajo el cual nace este trabajo de título. Así, la idea es combinar lo que se conoce de la interacción de la memoria de trabajo con el proceso de escritura (aprendido gracias a la Sección 2.6.1) con la metodología, procedimiento y procesamiento de las investigaciones presentadas en la Sección 2.6.2.

Capítulo 3

Ejecución experimental

En esta sección se detalla sobre los sensores utilizados, sobre el grupo que fue sometido al experimento, al diseño de éste y al procedimiento realizado para asegurar que todos los participantes tuvieran el mismo conocimiento y escenario para el experimento.

3.1. Instrumentación

Los sensores a utilizar son:

- Unidad GSR+ de Shimmer
- Sensor óptico de pulso de Shimmer
- Unidad *Bridge Amplifier+* de Shimmer
- ECG de tarjeta biomédica todo-en-uno de BITalino
- EEG Emotiv Epoc
- *Eye Tracker* Tobii T120

A continuación, se presenta una breve descripción técnica del sensor y recomendaciones de uso.

3.1.1. Unidad GSR+ de Shimmer

La unidad GSR+ de Shimmer puede medir la característica eléctrica o conductancia de la piel (GSR) como también puede capturar el fotopleletismograma (PPG), utilizando el sensor óptico de pulso Shimmer. Para la medición de la señal GSR se utilizan las entradas rojas y negras que se observan juntas en la Figura 3.1. Para la medición de PPG se utiliza la entrada negra más alejada.



Figura 3.1: Unidad GSR+ de Shimmer

3.1.1.1. Respuesta galvánica de la piel

La resistencia de la piel es proporcional a la humedad (sudor) que se encuentra en la piel. Para medirla, no existe un lugar estándar, aunque en general se mide en las falanges de la mano (y es también donde lo recomienda el manual de GSR de Shimmer [70]). Dado que para este experimento se necesita utilizar ambas manos para escribir sobre el teclado, la señal GSR se mide en la planta del pie, lugar que se menciona en [71] y [48] y que está validado en [72] y [73]. En [71] se indica que en casos donde ambas manos sean necesarias para la ejecución del experimento, el siguiente mejor lugar para la obtención de esta señal son las plantas de los pies. Dado que éstas están sujetas a la presión del apoyo del pie, se recomienda el sitio mostrado en la Figura 3.2, de [71], con las letras A y B.

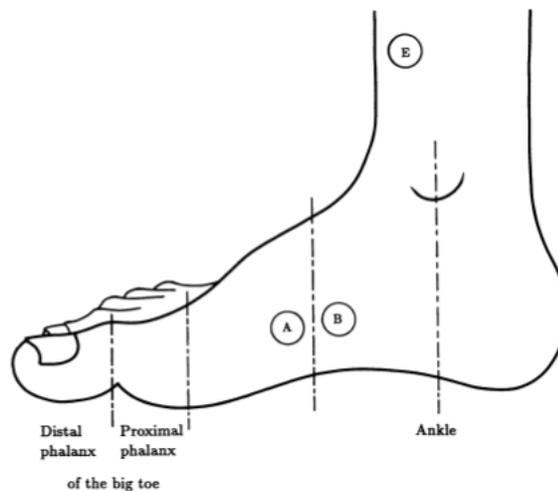


Figura 3.2: Zona recomendada para medición de GSR en pies

Recomendaciones para una buena medición [48][70]

- Para garantizar medidas precisas, el sensor debe estar conectado a la piel en todo momento. Es importante asegurarse que los sensores estén sujetos a la palma de la mano o pie, sin estar demasiado apretados, dado que esto deteriora los datos.

- Normalmente no se necesita mucha preparación del sitio donde se medirá. Sin embargo, si un sujeto tiene la piel demasiado grasa, se debe limpiar con alcohol antes de poner los electrodos. Por otro lado, si el sujeto tiene la piel muy seca (en ancianos por ejemplo), se puede aplicar crema previo al experimento.
- Se le debe instruir al usuario que respire de manera normal, evite el movimiento de extremidades, evite hablar y se siente cómodamente durante el experimento. Todo esto con el fin de evitar artefactos en la señal.

3.1.1.2. Sensor óptico de pulso de Shimmer

El sensor óptico de pulso permite medir el fotopleletismograma (PPG), también llamado volumen de pulso sanguíneo (BVP por sus siglas en inglés) de un dedo, lóbulo de la oreja o cualquier parte del cuerpo. Un fotopleletismograma es la medición volumétrica de un órgano, obtenida ópticamente. Se obtiene iluminando la piel y midiendo los cambios en la absorción de la luz. En particular, en este caso, se medirá en la oreja (como se muestra en la Figura 3.3)

El PPG mide la dinámica cardiovascular al detectar cambios en la translucidez arterial. Cuando el corazón bombea sangre, las arterias se vuelven más opacas, permitiendo que menos luz pase desde el emisor de luz en el sensor hacia el receptor. Dada esta relación, se puede derivar el ritmo cardiaco al detectar los peaks de la señal PPG.



Figura 3.3: Sensor óptico de pulso en oreja

Dado que el sensor para medir la señal GSR y PPG es el mismo, se debe construir un

alargador para el sensor de PPG. Esto, ya que el que se tiene mide 1 mt y no alcanza a llegar desde el pie (donde se mide el GSR) hacia la oreja (donde se mide el PPG). Para construir el alargador se utiliza un manos libres y entrada de manos libres de un celular viejo. Lo más crítico es que el alargador no le meta más ruido a la señal. Se mide la señal PPG en un sujeto con y sin alargador, y se comprueba que éste no distorsiona la señal.

Recomendaciones para una buena medición [74]

- Se recomienda una frecuencia de muestreo sobre los 100 Hz.
- A pesar de que se puede medir en distintas partes del cuerpo, se recomienda el lóbulo de la oreja dado que los artefactos por movimiento tienden a ser mínimos y, además, no existe actividad muscular que cause interferencia, lo que reduce el ruido de la señal.

Conversión de señal PPG a ritmo cardiaco: para convertir la señal del PPG a ritmo cardiaco, se deben detectar los peaks de la señal. Para esto no basta un valor de threshold, dado que la amplitud puede variar bastante en una misma medición (debido a la respiración, actividad muscular, movimiento del sensor y calidad de agarre del sensor al lugar de medición). Una vez identificado los peaks de la señal, se calcula el tiempo entre peaks y se deriva el ritmo cardiaco.

3.1.2. Unidad *Bridge Amplifier* de Shimmer

La unidad *Bridge Amplifier+* de Shimmer permite obtener distintas mediciones, entre ellas se puede obtener la temperatura corporal y ambiental mediante una resistencia altamente sensible a la cuál se le puede conectar un sensor de temperatura. En la Figura 3.4(a) se observa el dispositivo y en la Figura 3.4(b) se muestra el sensor que se le conecta [18].



(a) Unidad *Bridge Amplifier* de Shimmer



(b) Sonda de temperatura superficial de la piel

Figura 3.4: Implementos necesarios para medición de temperatura en la piel

Recomendaciones para una buena medición

- El sensor puede ponerse en el dedo pulgar o en la axilar. Se recomienda la axila dado que existe menos movimiento, lo que reduce el ruido en la señal.
- Se recomienda una frecuencia de muestreo baja, por ejemplo, 10 Hz.

3.1.3. ECG de tarjeta biomédica todo-en-uno de BITalino

La tarjeta todo-en-uno de BITalino viene integrada con distintos sensores: electromiograma (EMG), ECG, GSR, EEG y acelerómetro. En particular, se utiliza para obtener la señal del ECG, que refleja la actividad eléctrica del corazón.

Posición de los electrodos

Se utiliza el estándar Lead II para el posicionamiento de los electrodos. La configuración utilizada se muestra en la Figura 3.5.

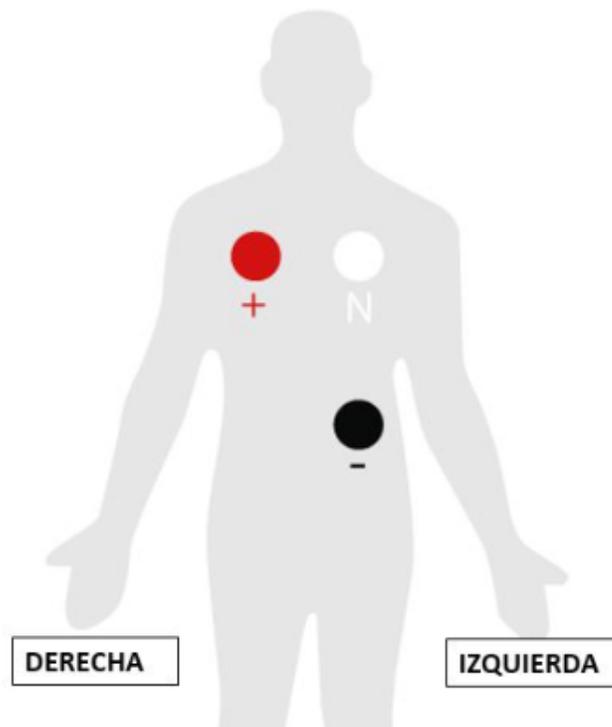


Figura 3.5: Posición electrodos ECG

Es importante que cada electrodo quede lejos del corazón y en la dirección de la extremidad correspondiente. Por ejemplo, el electrodo positivo puede ponerse en cualquier lugar del pecho, siempre y cuando se coloque lejos del corazón en la dirección del hombro derecho [18].

Recomendaciones de uso

- El dispositivo puede muestrear a 10, 100 o 1000 Hz. Se recomienda una frecuencia de muestreo >100 Hz para ECG.
- Se recomienda un filtro pasabanda con frecuencias de corte 2-40 [Hz], dado que se elimina la mayoría del ruido ocasionado por el movimiento de músculos y se elimina el ruido de la frecuencia de la red (60 [Hz] en este caso) [75].
- Para una correcta medición, se recomienda la limpieza de la piel con alcohol antes de poner los electrodos. También, para ahorrar tiempo, se recomienda utilizar los electrodos que ya vienen con gel conductor incorporado.

Conversión de señal ECG a ritmo cardiaco: se realiza de manera análoga a la descrita en el PPG (detección de peaks).

3.1.4. EEG Emotiv EPOC

Para la medición de EEG se utiliza el casco EEG Emotiv EPOC, mostrado en la Figura 3.6a. Éste cuenta con 16 canales, de los cuales 14 son efectivos (AF3, AF4, F3, F4, F7, F8, FC5, FC6, P7, P8, T7, T8, O1 y O2) y 2 son de posicionamiento. La disposición de estos sigue la configuración 10-20 (ver Figura 2.3), se muestran en la Figura 3.6b.

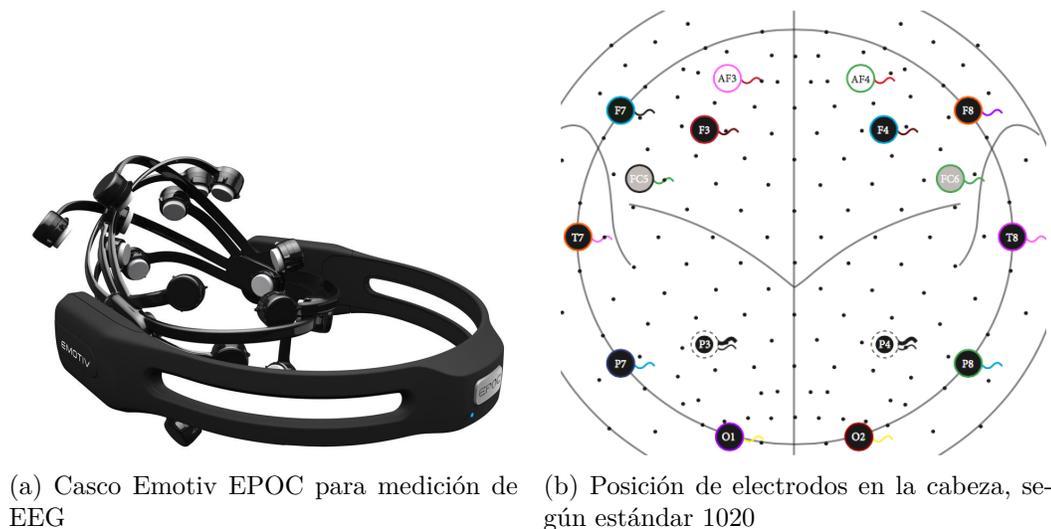


Figura 3.6: Sensor para medir EEG

Posición del sensor [76] Utilizando ambas manos, deslizar el casco desde la parte superior de la cabeza del usuario hacia abajo. Poner el sensor como se indica en la Figura 3.7, teniendo especial atención en el posicionamiento de los sensores de goma en el hueso atrás de la oreja (flecha negra en la segunda cabeza de la Figura 3.7a). Además, se deben ajustar los sensores frontales para que queden, aproximadamente, con una distancia de 3 dedos desde la frente hasta la ceja (Figura 3.7b).

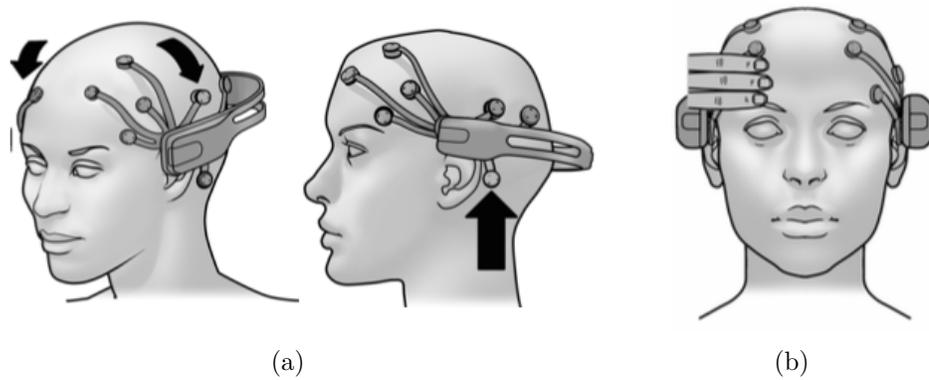


Figura 3.7: Posicionamiento del sensor

Una vez posicionado el sensor, se deben mantener apretados los sensores de referencia, que se ubican justo arriba de las orejas (con flecha en Figura 3.8). Un buen contacto en los electrodos de referencia es clave para una buena señal.

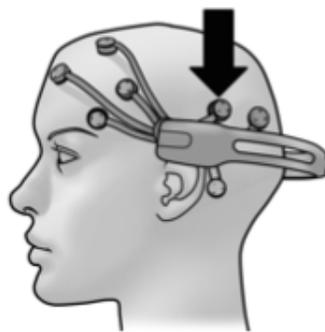


Figura 3.8: Sensores de referencia

Una vez obtenida una buena señal en los electrodos de referencia, apretar suavemente cada sensor contra el cuero cabelludo hasta que todas las luces correspondientes a los sensores estén en verde en el software *Xavier Testbench* (Figura 3.9).

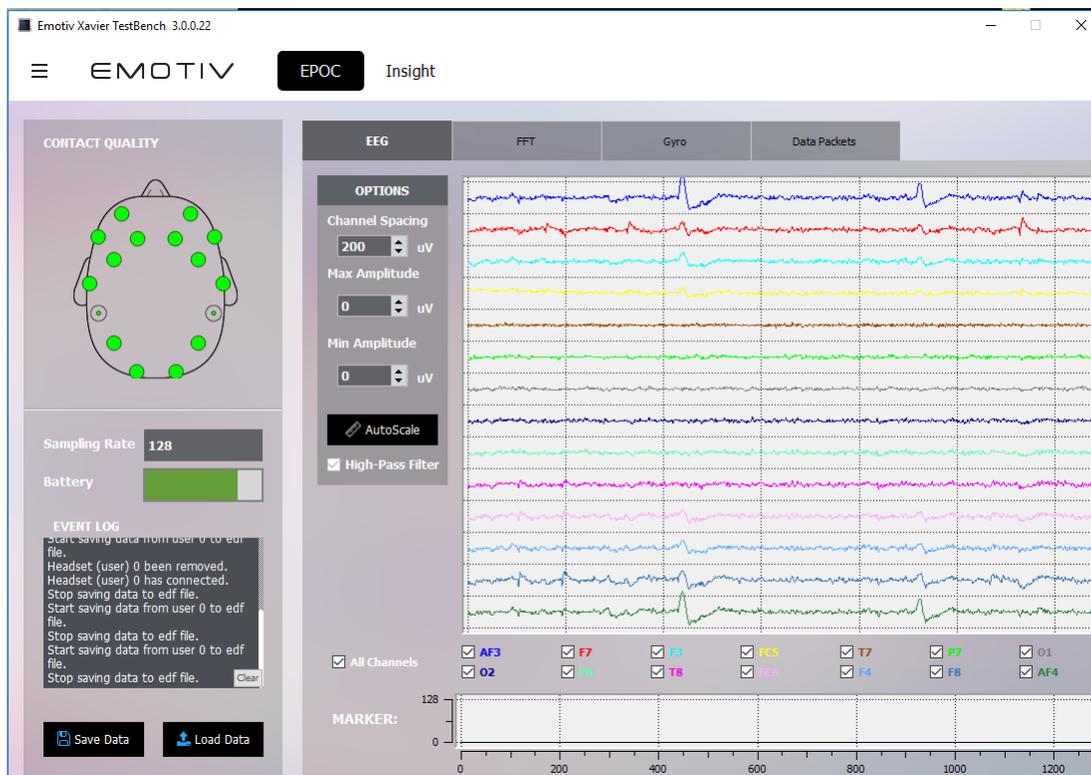


Figura 3.9: Verificación de posición de sensores en *Xavier TestBench*

Recomendaciones para una buena medición [76]

- Antes de posicionar el sensor, hidratar los electrodos con solución salina, para mejorar su conductividad.
- Revisar que cada uno de los 16 electrodos esté bien puesto en su soporte y que no estén sueltos.
- Asegurar que cada electrodo quede en contacto con el cuero cabelludo, retirando el cabello de la zona.
- Para cada sensor que tenga una calidad de contacto baja (negro, naranja o rojo en software), revisar que el sensor no esté suelto, que éste esté firmemente presionando el cuero cabelludo y revisar la hidratación (re-hidratar si es necesario).
- Si se tiene problemas con los sensores de referencia, se pueden probar posiciones alternativas (por ejemplo cambiando los sensores de goma por los de referencia, y dejando la posición atrás de las orejas como referencia).

Recomendaciones de uso [76]

- No utilizar el sensor en la cabeza mientras esté conectado para cargarse.
- Decirle explícitamente al usuario que evite mover la cabeza, apretar la mandíbula, mover la lengua o cualquier movimiento que vaya a agregar ruido a la medición.

3.1.5. *Eye tracker* Tobii T120

El *eye tracker* Tobii T120 es un instrumento de medición que permite seguir la mirada de un sujeto. Este modelo en particular viene integrado con una pantalla de 17 pulgadas, se muestra en la Figura 3.10.



Figura 3.10: Eye Tracker Tobii T120

Recomendación para posicionar al participante [77]

- Antes de la calibración, asegurarse que el participante esté en una postura cómoda que puedan mantener durante la duración de la prueba.
- Los ojos del participante debiesen quedar a, aproximadamente, 65-70 cms del eye tracker (Figura 3.11a).
- La cara del participante debiese estar paralela a la pantalla (Figura 3.11).
- Asegurarse que la ventana de status del eye tracker muestre dos ojos estables, con las barras de los lados con color verde y distancia de 65 cms (Figura 3.11b).
- Asegurarse que el participante pueda leer texto y alcanzar el mouse y teclado desde la posición en la que está.

Recomendaciones para una correcta medición

- Evitar que el usuario se mueva después de la calibración: usar una silla fija (sin ruedas), ajustar el alto de la pantalla para que el usuario no se incline hacia adelante y dar instrucciones explícitas de no moverse.
- Controlar la iluminación variable, para evitar que afecte en la medición del diámetro pupilar.

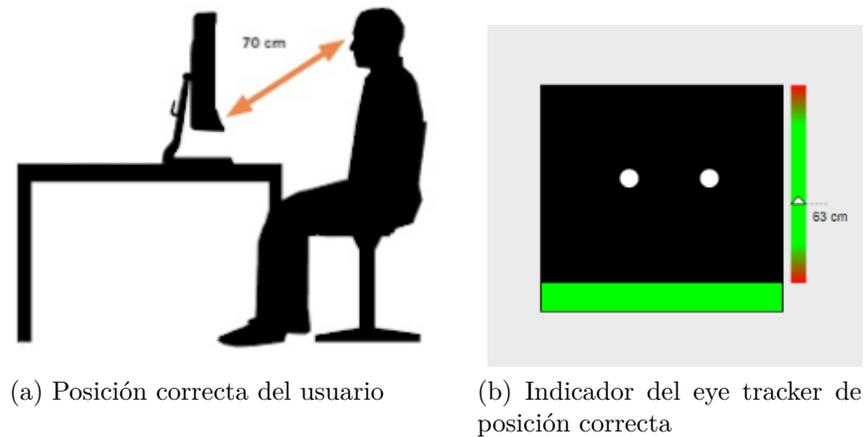


Figura 3.11: Posicionamiento del usuario para el eye tracker

3.2. Grupo experimental

Para realizar el experimento, se utiliza una muestra por conveniencia. Ésta consiste en 70 participantes (48 hombres y 22 mujeres) con una edad promedio de 24.7 años (± 1.8 años), en su mayoría estudiantes de ingeniería. Entre aquellos que no eran estudiantes de ingeniería, se cuenta una economista, una profesora de lenguaje, un estudiante de literatura, una estudiante de pedagogía, un ingeniero comercial y algunos ingenieros graduados hace poco. Todos los profesionales mencionados se encontraban graduados (de postgrado o pregrado) hace 3 años o menos. Así, todos estaban relacionados de alguna u otra manera al ambiente universitario.

De todos los participantes, se eligen solo aquellos que sus mediciones no presentan problemas (todos los electrodos del EEG en verde o -uno o dos- en naranja durante toda la medición, electrodos se mantienen bien pegados durante toda la medición, etc.). Además, se deben descartar cuatro por razones externas (no siguen las instrucciones, existen interrupciones durante el experimento..). Así, se obtiene un total de 55 participantes válidos (39 hombres y 16 mujeres con una edad promedio de 24.6 años (± 2.18 años)). Se detalla sobre género, edad y ocupación del grupo experimental en el anexo A (Tabla A.1).

Todos los participantes deben leer y firmar un consentimiento informado para el uso de sus señales. Además, en éste declaran no haber consumido drogas recreativas (alcohol, tabaco u otras) en un periodo de 24 hrs. Todos los participantes cuentan con visión o visión corregida normal y no presentan enfermedades que puedan alterar los resultados.

3.3. Diseño experimental

El experimento consiste en que el usuario debe escribir un ensayo sobre una pregunta dada (la misma para todos). La pregunta es: "Estás de acuerdo o en desacuerdo con la siguiente afirmación: estudiar en grupo es más efectivo que estudiar solo". Se elige esta pregunta desde ensayos disponibles online del TOEFL, dada su temática cercana con el ambiente de la mayoría de los participantes (estudiantes universitarios o egresados hace poco).

Se utiliza la plataforma entregada por Google para realizar encuestas y se realiza una encuesta con 3 páginas: en la primera debe llenar sus datos personales (nombre, rut, mail, sexo, edad, si sabe qué es un ensayo y si se siente preparado para escribir un ensayo sobre un tema cualquiera), en la segunda se le explica la estructura de un ensayo y sus principales componentes, según lo establecido en las Sección 2.1.2, y en la tercera se les presenta la pregunta y se les da espacio para escribir el ensayo. Se muestran las dos últimas partes de la encuesta en la Figura 3.12. La encuesta y pregunta del ensayo es la misma para todos los participantes.

La tercera hoja está dividida según las partes de un ensayo (introducción, tesis, argumento n1, argumento n2 y conclusión). Además, se les da un espacio para que escriban un borrador, si es que quieren, y se les recomienda revisar su ensayo antes de entregar. Cada participante tiene 30 minutos para elaborar su ensayo, desde que leen la pregunta. Dado que no se les puede interrumpir durante el experimento, se les facilita un reloj en la pantalla (se puede observar en la Figura 3.14b) y un recordatorio de que tienen 30 minutos en la tercera página de la encuesta. Si el participante se pasa del tiempo establecido, se deja que siga escribiendo hasta que envíe el ensayo.

3.3.1. Preparación del experimento

Se adquieren los insumos necesarios: electrodos desechables para el ECG y GSR, cinta micropore para pegar el sensor de temperatura a la axila y suero fisiológico para hidratar los electrodos del EEG.

Todo el experimento se realiza en la misma sala, donde se cuenta con regulación de la temperatura ambiental (siempre es la misma). Además, para evitar contaminar las medidas del eye tracker, se ponen cortinas *black-out* en las ventanas y puerta de la sala (Figura 3.13). Así, se realiza el experimento completamente a oscuras.

Por último, para evitar las interrupciones durante el experimento (que abran o toquen la puerta), se pone un cartel afuera de la sala donde se indica que se está en experimento, la fecha del día y los horarios dónde se estará en experimento.

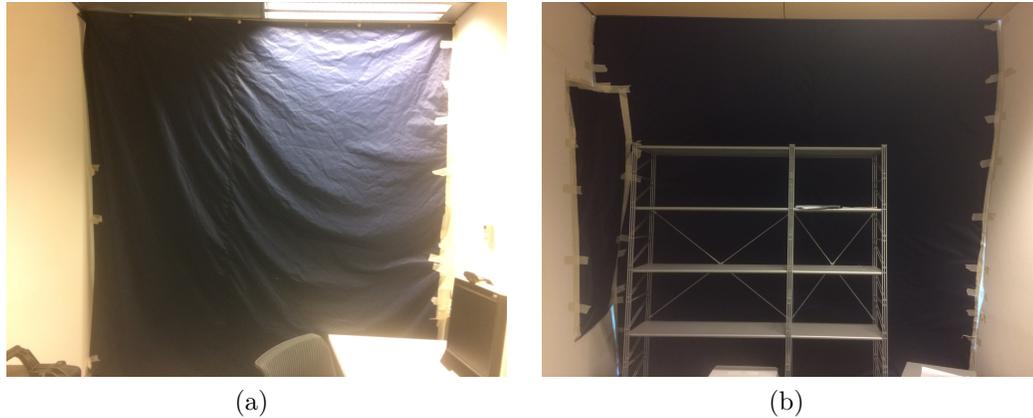


Figura 3.13: Sala donde se llevan a cabo los experimentos

3.4. Procedimiento experimental

Antes de la llegada del usuario, dado que se utilizan dos computadores para el experimento, se sincroniza la hora de ambos. Luego, se preparan los sensores (se conectan los electrodos necesarios, se hidrata el EEG, se configura la frecuencia de muestreo...) y se verifica la conexión a internet.

Una vez que llega el usuario, se le explica el experimento y se le pide que lea y firme el consentimiento informado. Se le pide que se saque aros de la oreja derecha, si es que tiene, y la zapatilla y calcetín de su pie no dominante. Luego, se comienzan a poner los sensores: primero, se pone el EEG, luego el PPG y GSR, luego el sensor de temperatura, el ECG y, por último, se calibra el eye tracker. Se le indica al usuario durante todo este proceso que durante el experimento evite moverse para evitar contaminar las señales con ruido. También se le da indicaciones sobre el experimento: no se puede conversar durante el experimento, no se pueden hacer preguntas ni interrumpir de ninguna manera y que escriba un ensayo como siempre lo ha hecho, de ninguna manera especial.

Antes de comenzar el experimento, el usuario se ve sometido a una relajación de 5 minutos. Ésta consiste en cerrar los ojos (la sala ya está oscura) y escuchar una canción durante 5 minutos (disponible en ¹) y se le indica que se intente relajar lo más posible, sin quedarse dormido. Una vez se acaba la canción, el usuario abre los ojos (se le instruye desde antes) y comienza el experimento. Si el usuario se demora en abrir los ojos, se deja hasta que los abra.

Una vez realizado el experimento, se desenchufa al usuario de los sensores y se le pide que no le cuente a nadie la pregunta del ensayo ni el procedimiento del experimento. Luego, se pasan las señales al computador, guardándolas en una carpeta con el nombre del usuario.

¹<https://www.youtube.com/watch?v=UfkT9m9NQpg>

3.4.1. Configuración sensores

Para medir las señales de GSR y PPG se utiliza el sensor *Shimmer GSR+ Unit*, con una frecuencia de muestreo de 120 Hz. Para la medición de GSR, se posicionan los electrodos en el pie no dominante del participante. El sensor de PPG se ubica en el lóbulo de la oreja derecha.

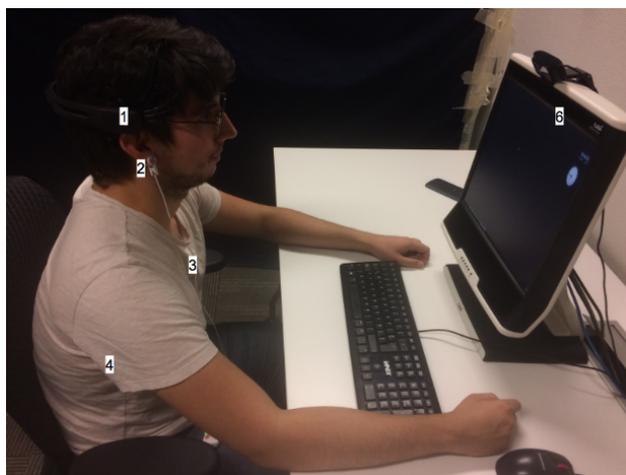
Para medir la temperatura corporal, se utiliza el sensor *Shimmer Bridge Amplifier+ Unit* con una frecuencia de muestreo de 120 Hz. El sensor se posiciona bajo la axila derecha. Para sincronizar ambos sensores Shimmer se utiliza el software *Consensus Pro*.

Para medir el ECG se utiliza el dispositivo *BITalino BioMedical Development All-in-One Board* con una frecuencia de muestreo de 100 Hz. Se posicionan los electrodos según la configuración Lead II. Para la visualización de la señal se utiliza el software *OpenSignals (r)evolution*.

Para medir el EEG se utiliza el sensor *Emotiv EPOC EEG* con una frecuencia de muestreo de 128 Hz. Antes de poner el sensor, se hidratan los electrodos con una solución salina. Para capturar datos y verificar el posicionamiento de los electrodos se utiliza el software *Emotiv Xavier Testbench*.

Para medir la dilatación pupilar y seguimiento de la mirada se utiliza un dispositivo *eye tracker Tobii T120* con una frecuencia de muestreo de 120 Hz. Para la calibración, toma de datos y etiquetado de muestras se utiliza el software *Tobii Studio*

En la Figura 3.14 se muestra un participante con todos los sensores puestos, listo para ejecutar el experimento ya descrito.



(a) Sujeto con sensores (1)EEG, (2)PPG, (3)ECG -bajo la polera-, (4)Temperatura -bajo la polera- y (6)Eye Tracker



(b) Sujeto con sensor (2)PPG y (6)GSR

Figura 3.14: Participante con los sensores ya posicionados.

Recuerda que el ensayo es un texto argumentativo, donde un autor busca exponer su punto de vista sobre un tema. Por lo tanto, está definido por el uso de la expresión personal y por la posición que toma el ensayista en relación al tema planteado.

A continuación se explica la estructura general de un ensayo.

Estructura de un ensayo

- 1. Introducción:** incluye una breve presentación del tema, motivación de la escritura, la tesis planteada y un pequeño adelanto de qué se hablará a continuación. Recuerda que una buena tesis solo tiene una idea principal y puede ser expresada en una oración corta.
 - Frase introductoria y general que relacione el tópico del ensayo con lo que vas a defender/refutar
 - Especificar sobre el tema del ensayo
 - Opcional: mostrar la postura contraria de la discusión (que tu no apoyas)
 - Da tu opinión
- 2. Cuerpo o desarrollo:** se exponen los argumentos, estos deben ser del tipo lógico racional. Evita utilizar argumentos basados en afectividad o subjetividad, como apelar a la crueldad, el sentido común o los sentimientos de las personas.

Es un argumento (o idea principal) por párrafo.

Se sugiere la siguiente estructura al presentar cada argumento que explica tu tesis:

 - Explica el argumento si es necesario
 - Da un ejemplo concreto o más detalle sobre tu argumento
 - Opcional: segundo ejemplo o detalle
 - Explica por qué el (o los) ejemplo(s) respalda(n) tu opinión
- 3. Conclusiones:** se reafirma la tesis, basándose en lo expuesto a lo largo del texto y se cierra el ensayo.
 - Volver a exponer la tesis
 - Opcional: habla sobre la postura contraria de la discusión (opiniones o razones opuestas) para darle más fuerza a tu tesis
 - Explicita las consecuencias o implicancias de tu tesis y plantea alguna reflexión final para cerrar el tema o vincularlo con otras problemáticas.

Resumen de la estructura sugerida

Introducción

- Explicar el tema
- Presentar la tesis

Desarrollo

- **Párrafo 1**
 - Defensa de tu tesis (razón de por qué)
 - Ideas que complementen la defensa
- **Párrafo 2**
 - Defensa de tu tesis
 - Ideas que complementen la defensa

Conclusión

- Reafirmar la tesis

(a) Se explica la estructura de un ensayo

Estás de acuerdo o en desacuerdo con la siguiente declaración: es más efectivo estudiar en grupo que estudiar solo

Tienes 30 minutos para escribir un ensayo sobre el tema mencionado arriba (en morado). Te recomiendo que uses los primeros minutos para organizar tu ensayo (en el espacio llamado borrador) y los últimos minutos para revisar lo escrito.

Recuerda la estructura sugerida:

Introducción

- Explicar el tema
- Presentar la tesis

Desarrollo

- **Párrafo 1**
 - Defensa de tu tesis (razón de por qué)
 - Ideas que complementen la defensa
- **Párrafo 2**
 - Defensa de tu tesis
 - Ideas que complementen la defensa

Conclusión

- Reafirmar la tesis
- Resumen de lo presentado

Borrador

Tu respuesta

Introducción *

Tu respuesta

Presenta tu tesis (oración corta) *

Tu respuesta

Desarrollo: argumento n1 *

Tu respuesta

Desarrollo: argumento n2 *

Tu respuesta

Conclusión *

Tu respuesta

Una vez terminado y revisado tu ensayo, aprieta enviar

ATRÁS
ENVIAR

(b) Se presenta la pregunta, se recuerda que tiene 30 minutos y se da el espacio para escribir

Figura 3.12: Página 2 y 3 de encuesta respondida por el usuario

Capítulo 4

Análisis de datos

4.1. Definición de observaciones

Durante el proceso de escribir un ensayo, se pasa por diferentes etapas, como se explica en la secciones 2.1.1 y 2.1.2. Por un lado están las etapas mentales: planificar, escribir y revisar (no necesariamente secuenciales) y por el otro las secciones del ensayo en sí: introducción, desarrollo y conclusión. Se busca separar las mediciones realizadas según estas etapas. Cada observación está caracterizada según qué sección del ensayo el usuario está escribiendo (introducción, tesis, desarrollo1, desarrollo2 o conclusión) y sobre qué proceso está realizando en esa etapa, por ejemplo, `introducción_escribiendo` o `conclusión_revisando`. Además, se agrega la sección de borrador, donde el usuario planifica lo que va a escribir. Así, las etiquetas de la actividad del usuario son de la forma `x_y`, con $x = \{\text{borrador, introducción, desarrollo1, desarrollo2, conclusión}\}$ e $y = \{\text{escribiendo, no escribiendo, revisando}\}$. Además se agrega la etiqueta `nn`, cuando el usuario no está haciendo nada definido (pasando rápidamente por cada sección, por ejemplo) y las etiquetas `enviar_i`, `enviar_f` que indican cuando el usuario empieza a enviar hasta que aprieta el botón de enviar la encuesta. Una vez separadas las muestras, se segmentan en ventanas de duración uniforme (5 y 2 segundos), sin solapamiento. Cuando existe un residuo dentro de una etapa, debe ser de duración mínima de 500 [ms] para ser considerada válida.

4.1.1. Etiquetado de subtareas

Dado que el usuario no siempre mira a la pantalla al escribir, no se puede etiquetar de manera automática la señal del eye tracker. Por lo tanto, se realiza manualmente desde el *software Tobii Studio*, mirando el vídeo de grabación de la pantalla disponible para cada usuario y agregando eventos que indiquen la actividad realizada por el usuario. Luego, se sincroniza el resto de las señales con los tiempos del eye tracker, y se segmentan según las anotaciones realizadas.

4.1.2. Duración promedio de las etapas del ensayo

Una vez etiquetadas y segmentadas las muestras, se obtiene las Tablas 4.1 y 4.2, donde se resumen la duración promedio (y su desviación estándar)¹ de cada etapa para los participantes. Para las etapas que no eran obligatorias (borrador y revisar), se añade entre paréntesis la cantidad de personas que las realizan. Es importante notar que solo se etiquetaba como `borrador_` cuando el usuario estaba sobre el área del borrador en la encuesta, por lo que si, por ejemplo, escribe sus ideas principales en cada sección, no se etiqueta como `borrador_`, tampoco si la persona escribe todo el ensayo en la sección borrador y luego copia y pega en las secciones correspondientes. Con respecto a revisar, solo se etiqueta como `_revisando` cuando la persona termina de escribir una sección y vuelve atrás para leerla, no se etiqueta como `_revisando` si, por ejemplo, la persona queda en la mitad de una idea y vuelve atrás para leer lo que lleva (dentro de una misma sección).

Tabla 4.1: Duración promedio de cada etapa estructural del ensayo

Etapa	Duración [min]
Borrador (38/55)	3,36±3,79
Introducción	5,36±2,69
Tesis	1,34±0,95
Desarrollo 1	6,78±2,68
Desarrollo 2	6,18±3,16
Conclusión	3,78±1,81

Tabla 4.2: Duración promedio de cada etapa mental durante la elaboración del ensayo

Etapa	Duración [min]
Escribiendo	14,46±3,99
No escribiendo	8,73±4,57
Revisando (46/55)	3,12± 2,03

De la Tabla 4.1 se observa que la variabilidad de la distribución de los tiempos es relativamente alta entre sujetos (especialmente para el borrador). De la Tabla 4.2 se observa que, en promedio, las personas pasan algo más de la mitad del tiempo escribiendo el ensayo y el resto del tiempo pensando o revisando.

¹Se utilizará la notación $x\pm y$, con x = valor promedio e y = desviación estándar, para todas las tablas que denoten un promedio en este documento.

4.2. Preprocesamiento de las señales

A continuación se detalla el proceso de limpieza y remoción de artefactos de cada señal obtenida por la toma de datos.

4.2.1. Señal GSR

Se obtiene la señal cruda desde el *software Consensys*. Lo primero que se hace es reducirle la tasa de muestreo, para aliviar la complejidad computacional. Desde 120 [Hz] se baja a 10 [Hz], sin perder resolución [48]. Para eliminar ruidos e impurezas no deseadas, se utiliza un filtro pasabanda de tipo *Butterworth* con frecuencias de corte 1 [Hz] y 5 [Hz]. Luego, se aplica una ventana de Hamming para suavizar la señal. Por último, se separa la parte fásica de la señal con la parte tónica. Todo esto se realiza con el toolbox Ledalab² [78] para Matlab. La Figura 4.1 ilustra los resultados de este proceso, para un individuo tipo³.

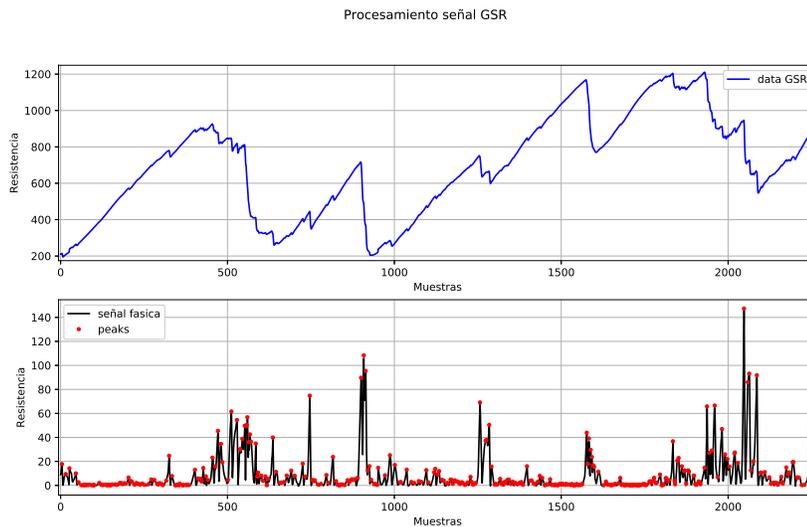


Figura 4.1: Preprocesamiento de señal GSR

4.2.2. PPG

La señal de PPG se obtiene del *software Consensys*. Desde esta señal se puede derivar el ritmo cardiaco. Para esto, se necesita detectar los *peaks* de la señal. Para esto, primero, se filtra la señal con un filtro FIR pasabandas de 16 Hz. Se utiliza un filtro FIR dado que mantienen la fase de la señal. Luego, para suavizarla se utiliza una ventana de Blackman. Una vez obtenida la señal limpia, se detectan sus *peaks*. Los resultados de este proceso se ilustran en la imagen 4.2.

²<http://www.ledalab.de>

³Todas las figuras de esta sección (Figura 4.1 - Figura 4.7) son ilustrativas de un sujeto para el preprocesamiento al cuál fueron sometidas todas las señales de todos los sujetos.

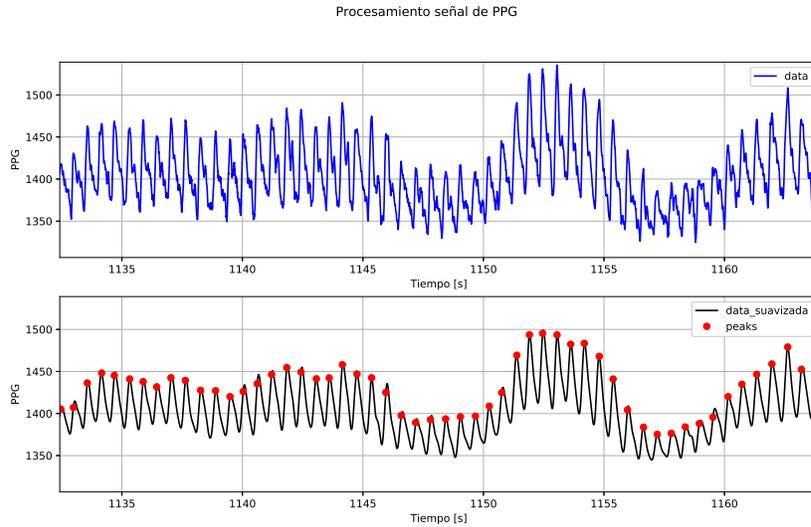


Figura 4.2: Preprocesamiento de señal PPG

Al analizar más detenidamente la señal PPG para distintos sujetos, se observa que no siempre es tan fácil distinguir los peaks en la señal (Figura 4.3), sea por ruidos en la señal u otros motivos, por lo que se decide derivar el ritmo cardíaco de la señal del ECG y utilizar esta señal de manera cruda.

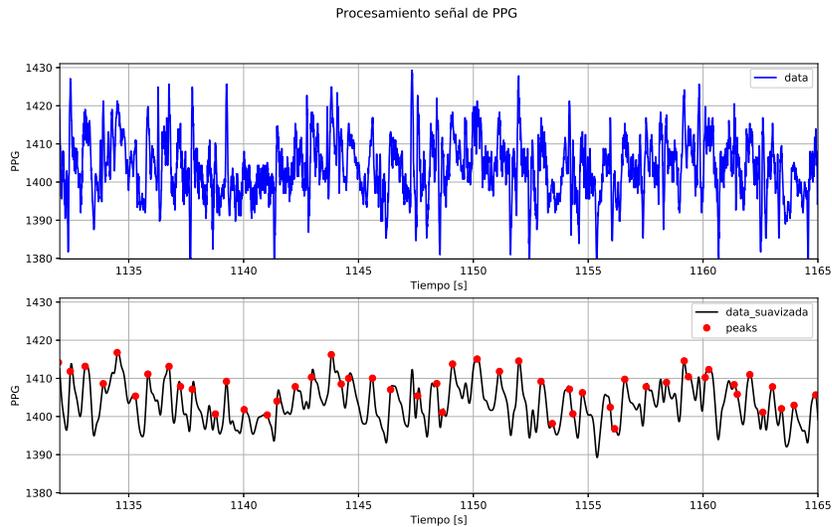


Figura 4.3: Señal de PPG con forma menos clara

4.2.3. Temperatura Corporal

La señal de temperatura corporal también se obtiene desde el *software Consensys*. Dado que se recomienda una frecuencia de muestreo baja para esta señal, lo primero que se hace es bajarle las muestras por segundo. De 120 [Hz] se baja a 50 [Hz]. Luego, se filtra la señal

con un filtro pasabajo con frecuencia de corte 10 [Hz]. Se ilustran los resultados del proceso en la Figura 4.4

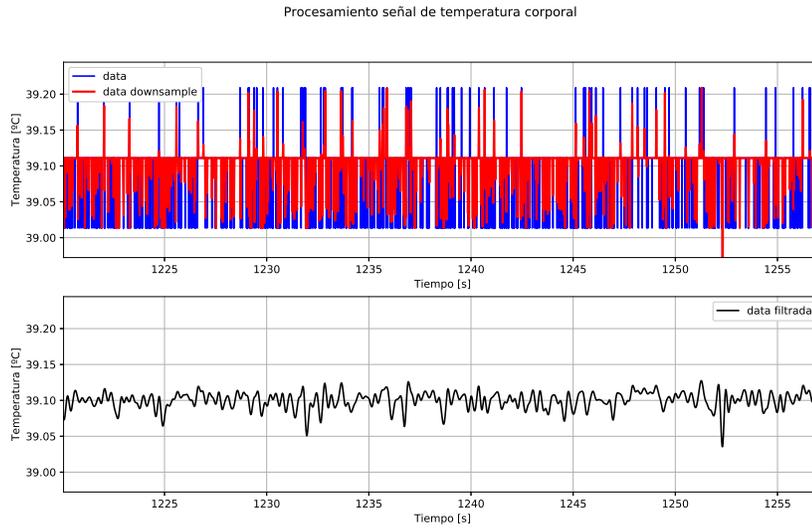


Figura 4.4: Preprocesamiento de la temperatura corporal

4.2.4. ECG

La señal ECG se obtiene desde el *software Open Signals (r)evolution*. Primero, se filtra la señal con un filtro pasabanda con frecuencias de corte 3 [Hz] y 45 [Hz]. Para encontrar los *peaks* de la señal se utiliza la librería de python BiossPy ⁴. En particular, se utiliza un segmentador de Hamilton [79]. Se ilustran los resultados del proceso en la Figura 4.5.

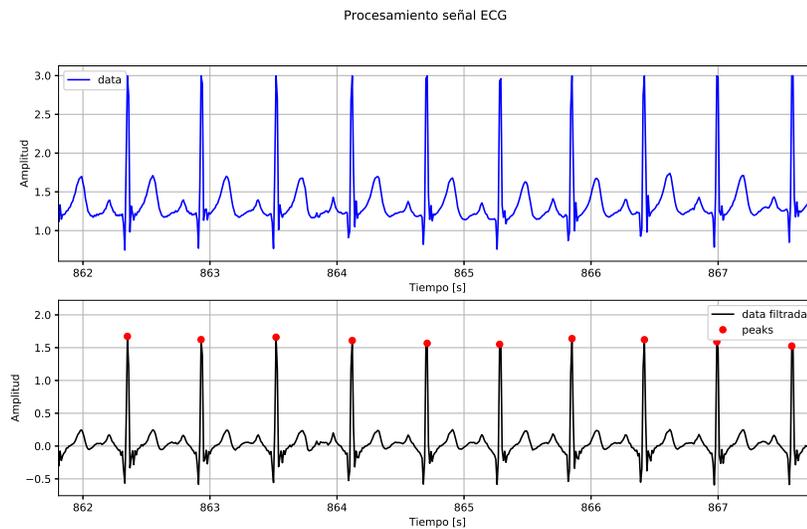


Figura 4.5: Preprocesamiento señal ECG

⁴<http://biosppy.readthedocs.io>

Se observan *peaks* mucho más claros que en la señal PPG, por lo que la derivación del ritmo cardíaco es más confiable, pero tiene la desventaja de que es un método mucho más invasivo para el usuario (3 electrodos en el pecho vs un clip en la oreja).

4.2.5. EEG

La señal de EEG está contaminada por muchos ruidos proveniente de movimientos inevitables para el sujeto, por ejemplo, movimiento de la lengua, de los ojos, la mandíbula... Existen distintas aproximaciones para intentar eliminar estos artefactos sin eliminar información útil para el análisis. En particular, se sigue el método basado en un filtro de Hampel de [80].

Lo primero que se hace es filtrar la señal del EEG con un filtro pasabanda para las frecuencias que interesan. Así, se utiliza un filtro *butterworth* pasabanda, con frecuencias de corte 4 - 40 Hz. Luego, se aplica un filtro de Hampel, para eliminar *outliers*. Se ilustra el preprocesamiento del canal AF3 en la Figura 4.6.

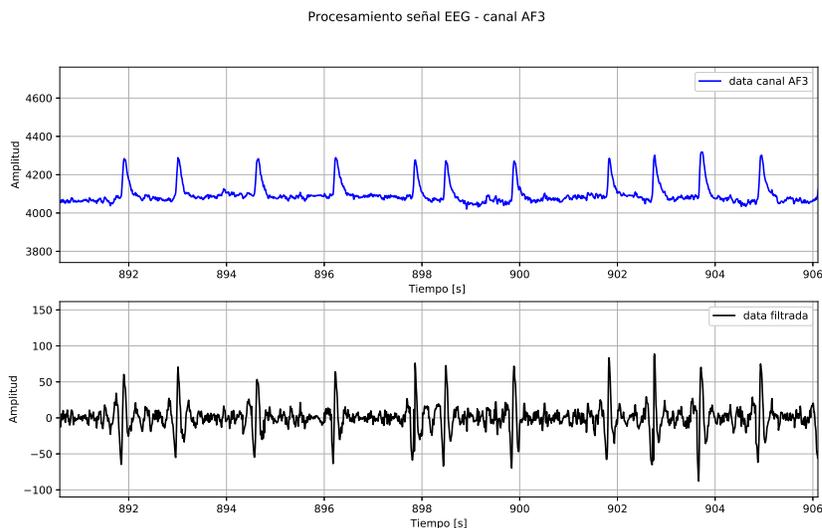


Figura 4.6: Preprocesamiento canal AF3 de EEG

4.2.6. Diámetro pupilar

La información sobre el diámetro pupilar se obtiene desde el *software TobiiPro Studio*. Éste entrega, para cada muestra, el diámetro de la pupila detectado por el *eye tracker* en [mm] y el grado de confiabilidad de este dato. Así, se tiene el diámetro de cada pupila (izquierda y derecha) y su grado de confiabilidad asociado (confiabilidad izquierda, confiabilidad derecha). El grado de confiabilidad va desde el 0 (muy confiable) hasta el 4 (no es confiable). También, se obtiene información sobre las muestras que son sacadas (incluye pestañeos) y fijaciones [77].

Para obtener una medida del diámetro pupilar de la persona, se busca aquella pupila con mayor grado de confiabilidad. Luego, se eliminan las muestras etiquetadas como sacadas por el *software* y se interpola linealmente ese dato. Por último, se pasa la señal obtenida por un filtro pasa bajo de orden 2 Hz, para eliminar el ruido de la señal. Se muestra el proceso en la Figura 4.7.

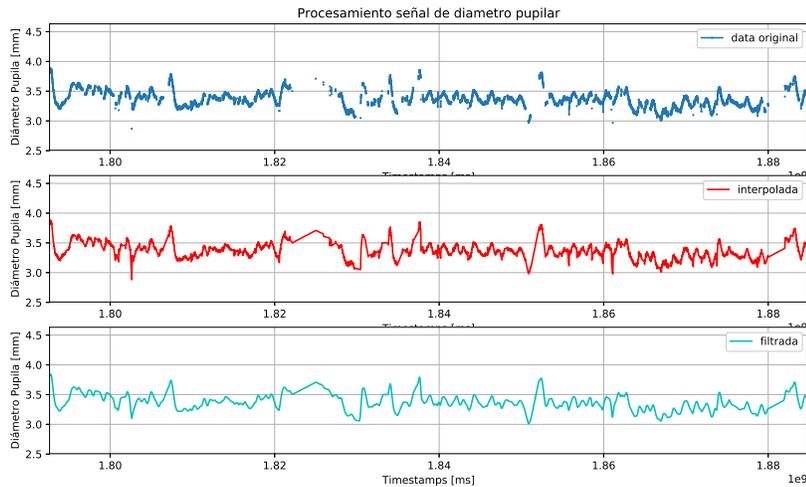


Figura 4.7: Preprocesamiento señal de diámetro pupilar

Se observa que hay secciones de la señal donde casi no existen datos del diámetro pupilar. Esto se debe a que la mayoría de los sujetos no mira a la pantalla al escribir, si no, al teclado, por lo que se pierde esa información. Por lo mismo, para cada observación se evalúa si se puede o no obtener la información del diámetro de pupila. Para esto, se determina que si el número de mediciones válidas es mayor al 50 % del número total de datos, se interpola la señal y se obtiene un valor de diámetro pupilar para cada instante. Una vez que se calcula esto para todas las observaciones de un usuario, se evalúa si el número de observaciones con valor de pupila es mayor al 80 % del número total de observaciones. Si esto se cumple, se mantiene esta medida. Si no, se elimina la medida para ese usuario. Así, existen usuarios que tienen la medida de diámetro pupilar (16/55) y otros que no. Se destaca en **negrita** en la Tabla A.1 (Anexo A), aquellos que tienen la medida válida.

4.3. Extracción de características

Se eligen distintas características para cada señal. Éstas deben poder describir de manera certera la señal. Para esto, se utiliza como base la extracción realizada en [18]. A continuación, se describen las características extraídas para cada observación de la señal. Se resumen en la Tabla 4.3.

Tabla 4.3: Características extraídas para cada señal

Señal	Características extraídas
Eye Tracker	promedio pupila, varianza pupila, cantidad de fijaciones y cantidad de sacadas.
GSR	GSR acumulada, promedio y potencia espectral
Componente Fásica	número peaks, máximo, promedio
ECG	promedio, mediana ECG y ECG MAD
HR (desde ECG)	promedio, desviación estándar, rms, AVNN, SDNN, rMSDD
PPG	promedio, desviación estándar, mediana, máximo y mínimo
Temperatura	pendiente, promedio y mediana
EEG (detalle en Tabla 4.4)	Separación en bandas de frecuencia y características de ellas

4.3.1. Eye Tracker

Se extrae el promedio y la varianza del tamaño de la pupila. También el número de fijaciones y sacadas [8][66].

4.3.2. GSR

En [67] se demuestra que la señal del GSR normalizada tiene mejores resultados con respecto a la señal completa. Se define la señal de GSR normalizado como:

$$GSR_{normalizado}(i, k, t) = \frac{GSR(i, k, t)}{\frac{1}{m} \sum_{j=1}^m \sum_{t=1}^{T_{ij}} GSR(i, j, t)} \quad (4.1)$$

donde $GSR(i,k,t)$ corresponde a cada punto de la señal para el sujeto i , tiempo t y tarea k ; m corresponde a la cantidad de tareas (en este caso, a la cantidad de observaciones para cada sujeto).

Con base a esto y a los resultados obtenidos por [18] se extraen las siguientes características:

- GSR acumulado:

$$GSR_{acumulado}(i, k) = \sum_t GSR_{normalizado}(i, k, t) \quad (4.2)$$

- GSR promedio:

$$GSR_{promedio}(i, k) = \frac{\sum_t GSR_{normalizado}(i, k, t)}{T} \quad (4.3)$$

donde T es el tiempo total de todas las observaciones.

Además, de la señal fásica se obtiene el número de *peaks*, el valor máximo y el promedio [8].

4.3.3. ECG

Se extrae el promedio y mediana de la señal. También se extrae la varianza de la ECG MAD (*median absolute desviation*), que es una medida de la dispersión estadística [7].

$$\text{ECG MAD} = |ECG_i - \text{mediana}(ECG)| \quad (4.4)$$

Ritmo Cardíaco

Se deriva el ritmo cardíaco desde los *peaks* del ECG y se le calcula el promedio, la desviación estándar y el valor cuadrático medio (rms por sus siglas en inglés: *root mean square*). Luego, se calcula el intervalo entre beats (IBI por sus siglas en inglés: *Inter Beat Interval*). Se calcula el promedio (AVNN), la desviación estándar (SDNN) y el rms (rMSDD), a partir de éste. [7][81][82].

4.3.4. PPG

Se toma el promedio, desviación estándar, mediana y valores máximo y mínimo de la señal.

4.3.5. Temperatura corporal

Se extrae el promedio y la pendiente de la señal (utilizando una regresión lineal) [15]. Además, se extrae la mediana.

4.3.6. EEG

Dada la complejidad de la señal del EEG, se sacan diferentes características para carga cognitiva, excitación y valencia. Se resumen en la Tabla 4.4 y se detallan a continuación

Tabla 4.4: Características extraídas del EEG

	Características extraídas del EEG
Carga Cognitiva	e_totalF3_theta, e_totalF4_theta, e_totalF7_theta, e_totalF8_theta, entropiaNorm_F3_theta, entropiaNorm_F4_theta, entropiaNorm_F7_theta, entropiaNorm_F8_theta, stdF3_theta, stdF4_theta, stdF7_theta, stdF8_theta, e_totalP7_alfa, e_totalP8_alfa, entropiaNormP7_alfa, entropiaNormP8_alfa, stdP7_alfa, stdP8_alfa
Valencia	beta-alfaF3, beta-alfaF4, beta-alfaF7, beta-alfaF8, e_totalF3_beta, e_totalF4_beta, e_totalF7_beta, e_totalF8_beta, e_totalP7_beta, e_totalP8_beta, cF7F8, asimetria_a/b_F4F3, asimetria_a/b_F8/F7
Excitación	e_totalP7_beta, e_totalP8_beta, cP7O2, cP8O1, cP7P8, cO1O2, b/a_AF3, b/a_AF4, b/a_F3, b/a_F4

4.3.6.1. Carga cognitiva

Existe evidencia de que las bandas más relevantes a la hora de distinguir carga cognitiva son las bandas alfa y theta en los lóbulos parietales y frontales, respectivamente [8][35][36][38]. Por lo tanto, se utiliza la banda de la señal que corresponde a las bandas alfa y theta y se limitan los canales del EEG a aquellos correspondientes a lóbulos frontales y parietales (F3, F4, F7, F8, P7 y P8). Se resumen las características en la Tabla 4.4

Dado los resultados obtenidos en [39][41] se decide separar las bandas de frecuencia del EEG utilizando transformada de Wavelet. La transformada de Wavelet separa la señal mediante una cascada de filtros pasa altos y pasa bajos. Se seleccionan los coeficientes de aproximación y detalle, según las bandas de frecuencias requeridas (alfa y theta).

Luego, se extraen las siguientes características [39]:

- Entropía normalizada:

$$\text{EntropiaNorm}_j = - \sum_{k=1}^{M_j} \bar{a}_j[k] \log_2(\bar{a}_j[k]), \quad (4.5)$$

con M_j la cantidad de coeficientes y \bar{a}_j los coeficientes wavelet normalizados según la ecuación 4.6.

$$\bar{a}_j = \frac{|a_j[k]|}{\sum_{k=1}^{M_j} |a_j[k]|} \quad (4.6)$$

- Energía total del canal: representa la energía del segmento de EEG bajo análisis para una banda de frecuencia particular.

$$\text{Energía}_j = \sum_{k=1}^{M_j} |a_j[k]|^2 \quad (4.7)$$

- Desviación estándar de la energía total

4.3.6.2. Detección de emociones

Se seleccionan las características que más se repiten en el análisis realizado en [43] y se agregan las características utilizadas en [29], dado sus resultados obtenidos (obtienen >75% de exactitud para la clasificación de valencia y excitación al escuchar música). En ambos casos se utilizan los canales y bandas que los autores utilizan. Para la separación en bandas de frecuencia se utiliza la transformada de Fourier. Se describen las características a continuación.

Valencia

- Potencia espectral de un canal [43]: se calcula utilizando el método de Welch [83]. Se calcula para los canales frontales y parietales en la banda β
- Diferencia $\beta - \alpha$ [43]: expresa el cambio relativo de actividad entre dos canales o entre dos bandas de frecuencias. Se calcula para los canales frontales.

- Coherencia entre canales [43]: indica el nivel de concordancia entre la actividad de dos canales cerebrales, en una banda de frecuencia dada. Se calcula según la ecuación 4.8.

$$C_{xy} = \frac{|P_{xy}|^2}{P_{xx} \cdot P_{yy}} \quad (4.8)$$

donde P_{xx} y P_{yy} son las potencias espectrales estimadas de X e Y, y P_{xy} es la potencia espectral cruzada estimada de X e Y. Se calcula para los canales F_7 y F_8

- Asimetría frontal [29]: se compara la potencia de alfa y beta en canales frontales correspondientes al hemisferio derecho (F_{par}) e izquierdo (F_{impar}). Esto, dado que se ha demostrado que la inactivación del lóbulo frontal izquierdo se asocia a emociones negativas. Por el otro lado, la inactivación del lóbulo frontal derecho está asociada a una emoción positiva.

$$\text{Asimetría Frontal} = \frac{\alpha F_{par}}{\beta F_{par}} - \frac{\alpha F_{impar}}{\beta F_{impar}} \quad (4.9)$$

donde αF_{par} corresponde a la banda α del canal F_4 o F_8 y αF_{impar} corresponde a la banda α del canal F_3 o F_7

Excitación

- Potencia espectral de un canal [43]: se calcula para los canales parietales en la banda β
- Coherencia entre canales [43]: se calcula para los canales parietales y occipitales y para combinaciones entre ellos.
- Coeficiente β/α [29]: se calcula para los canales frontales.

4.4. Etiquetado de las muestras mediante Clustering

Se busca etiquetar cada observación según su nivel de carga cognitiva, valencia y excitación. Para cada una de estas categorías se utiliza un distinto *gold-standard*.

- Nivel de carga cognitiva: se ha comprobado la relación que existe entre el tamaño de la pupila y el nivel de carga cognitiva [18][33]. Por lo tanto, para identificar los niveles de carga cognitiva experimentados por el usuario al escribir un ensayo, se clusteriza el promedio del diámetro de pupila con mayor grado de confiabilidad y la varianza de éste. Dado que se utiliza la medida pupilar como *gold-standard*, solo 16 de los 55 sujetos son utilizados como muestra en la clasificación de carga cognitiva.
- Nivel de valencia: en [84], se encuentra una alta correlación entre el nivel de valencia experimentado y el ritmo cardiaco. Así, se utilizan las características extraídas para HR (Tabla 4.3) para clusterizar el nivel de valencia experimentado por el usuario.
- Nivel de excitación: en [84], se encuentra una alta correlación entre el nivel de excitación experimentado y la señal GSR. Así, se utilizan las características extraídas para GSR y componente fásica (Tabla 4.3) para clusterizar el nivel de excitación experimentado por el usuario.

Para cada una de estas categorías, se siguen los siguientes pasos:

- Mediante SOM se construye el mapa de proto-núcleos
- Luego, con un algoritmo de clustering aglomerativo, se itera sobre el número K de clusters finales. La regla de decisión para elegir K es cuando el índice de Calinski-Harabasz indica el mismo número de clusters que el índice de Silueta.

Dado que SOM tiene una inicialización aleatoria, se realizan los pasos anteriores 50 veces y se elige el número de clusters K tal que sea el que más se repite, dentro de los cuales los índices estaban de acuerdo.

Se ilustra el proceso de clusterización para carga cognitiva en la Figura 4.8. En la Figura 4.8(a) se muestra el mapa de SOM ya clusterizado (se indica por los puntos del mismo color un mismo cluster) y en la Figura 4.8(b) se muestran los datos clusterizados (también se indica por color) en el espacio original. Los colores de una imagen y la otra no necesariamente se relacionan entre sí.

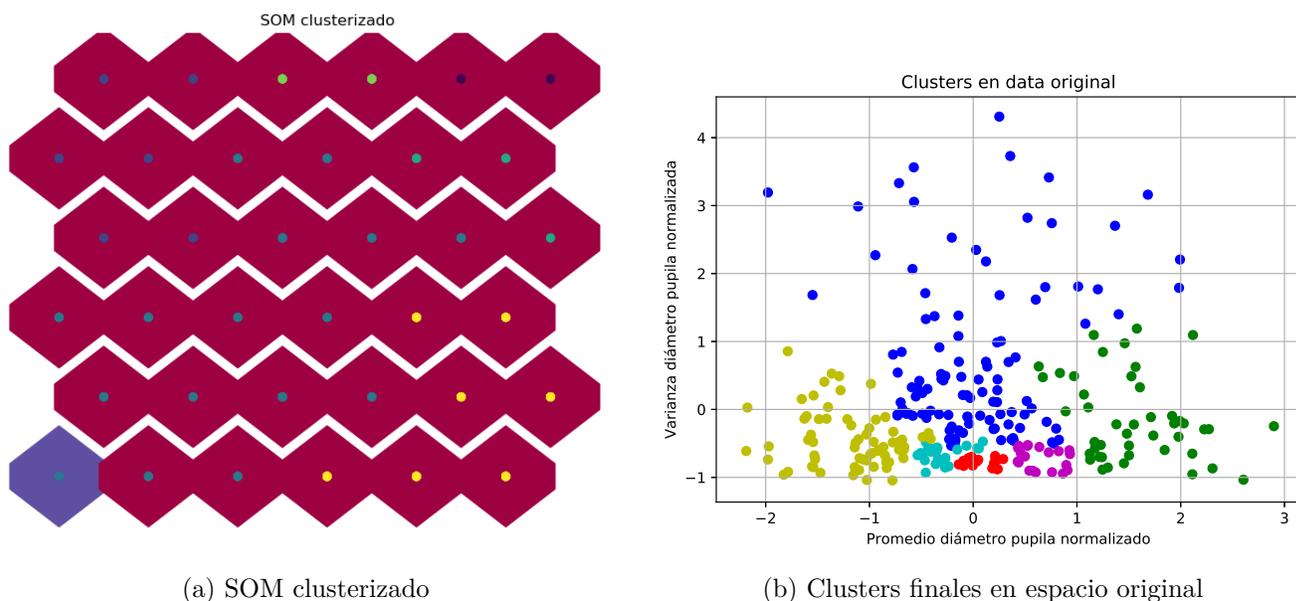


Figura 4.8: Proceso para clusterizar las muestras

Así, se obtienen niveles de carga cognitiva, valencia y excitación para cada usuario. En pos de la clasificación, se eliminan aquellos clusters que tengan menos de cinco muestras. Las Tablas 4.5, 4.6 y 4.7 resumen los niveles promedio de carga cognitiva, valencia y excitación obtenidos para todos los usuarios. Cada promedio se muestra en conjunto a su desviación estándar. Se muestran los niveles obtenidos para ventanas de tiempo de 5 segundos y de 2 segundos. Se muestra el detalle para cada participante en el Anexo B

Tabla 4.5: Promedio de niveles de carga obtenidos para los distintos usuarios.

t = 5		t = 2	
N clusters original	N clusters final	N clusters original	N clusters final
4,56±1,67	4,5±1,63	4,13±1,63	4,13±1,63

Tabla 4.6: Promedio de niveles de valencia obtenidos para los distintos usuarios.

t = 5		t = 2	
N clusters original	N clusters final	N clusters original	N clusters final
2,56±1,08	2,44±1,20	3,60±1,57	3,60±1,57

Tabla 4.7: Promedio de niveles de excitación obtenidos para los distintos usuarios.

t = 5		t = 2	
N clusters original	N clusters final	N clusters original	N clusters final
2,27±0,73	2,24±0,77	2,98±1,53	2,98±1,53

4.5. Clasificación

Se prueba con distintos algoritmos de aprendizaje supervisado y sus distintos parámetros:

- K-vecinos más cercanos: se prueba con $k = [1, 3, 5, 10]$
- SVM: se prueba con distintos kernels (lineal, rbf, polinomial -grado 2, 3 o 5- o sigmoideo), y $C = [1, 10, 100]$
- Redes neuronales: se prueba con una capa de neuronas ocultas, con distinto tamaño (teniendo como máximo el dictado por la regla de la pirámide, detallada en la Sección 2.5.3.5) y valores de alfa = $[0.001, 0.01, 0.1]$

Consideraciones importantes a tener en cuenta son:

- **Preparación de conjuntos de entrenamiento y prueba:** al clasificar de manera supervisada, se le entrega una porción de los datos al algoritmo para que aprenda (conjunto de entrenamiento) y otro para probar lo aprendido (conjunto de prueba). Es importante que el algoritmo sea probado con datos que nunca haya visto antes, para comprobar su capacidad de generalizar. Dado que se cuenta con un desbalance de clases (algunas con muchos datos y otras con pocos), se repartan los datos de manera que se mantenga la proporción de datos de cada clases, para cada subconjunto. Se separa la muestra con un 70 % entrenamiento y 30 % prueba (se muestra la cantidad de datos promedio en la Tabla 4.8). Cada clasificador se corre 5 veces. Se entrega el promedio de sus métricas y la desviación estándar de éste.

Tabla 4.8: Cantidad de datos promedio para todos los participantes

	t = 2	t = 5
Cantidad datos	824,6±219,57	348,36±94,21

- **Algoritmos para seleccionar y estandarizar características** Una vez separado el conjunto de entrenamiento y prueba, se estandarizan las características según la ecuación 4.10 para que sean comparables.

$$\bar{X} = \frac{X - \mu}{\sigma} \quad (4.10)$$

donde X corresponde a la característica, μ a su promedio y σ a su desviación estándar [54].

Luego, se aplica RFE-RF (Sección 2.5.1.1) para la selección de características. La importancia de la selección de características es que permite que solo las características más atingentes para cada sujeto entren a su clasificador, permitiendo reducir la dimensionalidad y mejorar la clasificación.

- **Métrica a maximizar:** se elige aquel clasificador que maximiza F1 (promedio armónico entre precisión y sensibilidad); si la diferencia entre los tres F1 de los clasificadores es poca, se elige aquel que tenga la mayor exactitud. Esto, dado que para el desbalance de clases se necesita un clasificador que clasifique lo más correctamente las clases pequeñas, lo que no se lograría maximizando solamente la exactitud.

Capítulo 5

Resultados y discusión

5.1. Análisis estadístico

Se realiza un análisis estadístico para responder principalmente a dos preguntas: ¿Son las etapas del ensayo (borrador, introducción, tesis, desarrollo, conclusión) diferenciables entre sí por las señales? ¿Pueden las señales diferenciar los distintos estados de carga cognitiva, valencia y excitación para cada una de estas etapas?

Para responder estas preguntas se realiza un test de normalidad y luego un test ANOVA-MR (de medidas repetidas) para todos los participantes, estandarizados por z-score. Se elige un intervalo de confianza de 95 %, es decir, que la probabilidad de equivocarse al rechazar la hipótesis nula sea menor a 0,05. A continuación, se presentan los resultados obtenidos y su análisis.

5.1.1. Diferenciación entre etapas de un ensayo

Se busca saber si las señales psico-fisiológicas pueden diferenciar entre las distintas etapas del ensayo. Para esto, se comparan todas las características de una señal en una etapa con las mismas características en otra. La hipótesis nula en este caso corresponde a que las medias de todas las características de una señal son iguales con respecto a las etapas comparadas. En la Tabla 5.1 se muestran los p-values de algunas características. De ésta, se puede observar que siempre existe al menos un sensor que puede diferenciar entre las etapas (p-value <0,05).

Tabla 5.1: Resumen de señales que presentan un p-value significativo ($<0,05$) para diferenciar entre etapas del ensayo

	Borrador	Introducción	Tesis	Desarrollo	Conclusión
Borrador	-	-	-	-	-
Introducción	8,13E-11	-	-	-	-
Tesis	3,02E-10	4,54E-06	-	-	-
Desarrollo	$<2E-16$	$<2E-16$	1,86E-10	-	-
Conclusión	$<2E-16$	$<2E-16$	5,56E-14	1,11E-08	-

5.1.2. Diferenciación entre estados para cada etapa del ensayo

Se busca responder a la pregunta de si las señales diferencian los estados de carga cognitiva, valencia y excitación para cada una de las etapas del ensayo. En este caso, la hipótesis nula corresponde a que la media de los grupos que se comparan son iguales. Para cada señal se toman todas sus características (listadas en la Sección 4.3) y se analiza si existe una diferencia significativa entre cada estado. A continuación (en las Tablas 5.2, 5.3, 5.4) se muestra un resumen de los resultados obtenidos.

5.1.2.1. Carga cognitiva

Se muestra en la Tabla 5.2 en amarillo y con el p-value de alguna de las características de la señal, aquellas señales que diferencian de manera significativa los estados de carga cognitiva para cada etapa del ensayo. Así, se observa que la única que no diferencia es la temperatura durante la tesis y la conclusión, todo el resto de las señales presenta al menos una característica con p-value significativo dentro del intervalo de confianza de 95 %. Se detalla sobre estos resultados en el Anexo C.1, donde en las Tablas C.1, C.2, C.3, C.4 y C.5 se muestran los p-values de cada característica para cada etapa del ensayo (borrador, introducción, tesis, desarrollo y conclusión, respectivamente) y se destacan con amarillo aquellos significativos.

Tabla 5.2: Resumen señales que presentan al menos una característica con p-value significativo ($<0,05$) para los niveles de carga cognitiva

	Borrador	Introducción	Tesis	Desarrollo	Conclusión
EyeTracker	2,84E-05	1,49E-06	3,00E-05	2,53E-08	0,01849
ECG	9,43E-13	2,00E-16	$<2,00E-15$	$<2,00E-15$	8,47E-15
HR	$<2,00E-16$	$<2,00E-16$	0,000344	0,000328	$<2,00E-15$
Temperatura	1,56E-10	2,47E-15		4,37E-05	
GSR	0,000252	1,42E-07	0,000363	1,54E-05	0,00438
PPG	6,28E-06	0,000891	0,00387	0,00296	x
EEG	0,000548	0,001445	0,0303	4,51E-11	0,01382

5.1.2.2. Valencia

En la tabla 5.3 se muestra en amarillo y con el p-value de alguna de sus características, aquellas señales que diferencian de manera significativa los estados de valencia, para cada etapa del ensayo. Se observa que la temperatura es la que menos distingue entre etapas (solo distingue para el borrador y la introducción). Además, el eye tracker no distingue para el borrador y el ritmo cardiaco (HR) no distingue para la conclusión. Así, las señales que logran diferenciar de manera significativa dentro del intervalo de confianza de 95 % para los niveles de valencia son ECG, GSR, PPG y EEG. En el Anexo C.2, en las tablas C.6, C.7, C.8, C.9 y C.10, se detallan las características de cada señal para cada etapa y el p-value asociado al test ANOVA-MR realizado. Se destacan en amarillo aquellos p-values menores a 0,05.

Tabla 5.3: Resumen señales que presentan al menos una característica con p-value significativo (<0,05) para los niveles de valencia

	Borrador	Introducción	Tesis	Desarrollo	Conclusión
EyeTracker		0,00203	0,0151	1,44E-05	0,0309
ECG	1,39E-05	8,20E-13	4,02E-05	6,51E-11	7,48E-05
HR	0,00252	0,02497	0,01591	1,34E-05	
Temperatura	0,000803	0,0263			
GSR	0,02503	0,02893	0,0121	0,0280	1,48E-05
PPG	1,80E-07	<2,00E-15	0,00035	1,28E-05	1,37E-08
EEG	0,0331	0,00207	0,0174	0,0046	0,0206

5.1.2.3. Excitación

En la Tabla 5.4 se muestran en amarillo y con el p-value de alguna de sus características, las señales que diferencian de manera significativa los estados de excitación. Se observa que durante la tesis es donde menos señales diferencian, lográndolo solo 4 de 7. También, que durante el desarrollo el ECG no diferencia. Así, las señales que efectivamente logran diferenciar los niveles de excitación durante el ensayo son PPG, EEG y HR.

Tabla 5.4: Resumen señales que presentan al menos una característica con p-value significativo (<0,05) para los niveles de excitación

	Borrador	Introducción	Tesis	Desarrollo	Conclusión
EyeTracker	0,019	1,49E-05		0,0005	4,89E-05
ECG	1,52E-05	1,47E-06	0,0117		0,0126
HR	7,40E-07	1,90E-15	1,31E-12	4,04E-09	0,007
Temperatura	0,0015	0,000826		7,11E-09	5,10E-05
GSR	7,36E-06	0,00046		4,05E-06	4,20E-06
PPG	3,29E-03	0,00346	0,0226	0,013879	0,0412
EEG	0,0226	1,76E-07	0,0373	8,08E-06	0,0010

5.2. Minería de datos

5.2.1. Selección de características

Interesa saber cuáles son las características más seleccionadas para todos los usuarios, utilizando RFE-RF (con las características estandarizadas). Dado que RF tiene un componente aleatorio, se aplica RFE-RF 5 veces para cada usuario, donde se cuenta cuántas veces salió cada característica (del 0 al 5). Una vez realizado este proceso para cada usuario, se observa cuáles son las características más relevantes para el conjunto. A continuación, se muestran las 5 características más seleccionadas en todo este proceso, para todo el grupo experimental (Tablas 5.5, 5.6 y 5.7). Al lado de cada característica, se anota la cantidad de veces promedio que fue elegida (5 siendo el máximo).

Para cada categoría (carga cognitiva, valencia y excitación) se tienen distintas características de las descritas en la Sección 4.3, dado que se deben eliminar las características que se utilizan como *gold-standard* para la obtención de etiquetas (Sección 4.4).

- **Carga cognitiva:** se tienen disponibles todas las características listadas en la Tabla 4.3, excepto diámetro y varianza de diámetro de pupila. Dentro de las características más seleccionadas (mostradas en la Tabla 5.5) se observa que, para ambos intervalos de tiempo, la señal fásica del GSR, la temperatura y el HR parecieran jugar un rol importante.

Tabla 5.5: Resumen 5 características más seleccionadas para la clasificación de niveles de carga cognitiva

t = 5	t = 2
Características más seleccionadas	Características más seleccionadas
numPeaksFasica 4,81	numPeaksFasica 4,37
numFijaciones 4,56	rMSDD 4,31
numSacadas 4,31	SDNN 4,25
rMSDD 4,12	promTemp 4,12
promTemp 3,81	medianaTemp 4,12

- **Valencia:** se eliminan todas las señales que se relacionan con el sistema cardiovascular (ECG, HR y PPG). Por lo tanto, se tienen disponibles las características listadas en la Tabla 4.3 correspondiente al *eye tracker*, GSR, temperatura y EEG. Además, dado que algunos usuarios tienen información del diámetro de la pupila (16) y otros no, se hace la diferenciación en el análisis. Es importante notar que, a pesar de que se hace la diferencia en esta sección (dado que los sujetos con características de diámetro pupilar tienen dos características más), en el posterior análisis de la clasificación no se diferencia entre sujetos con/sin pupila.

De los resultados obtenidos (Tabla 5.6), se puede observar que, nuevamente, la señal fásica y la señal del GSR parecieran jugar un rol importante en este estado mental. También, y más inesperado, resulta la presencia del número de fijaciones y sacadas en 3/4 escenarios (dado que son movimiento voluntarios y que, quizás, podrían asociarse a

valencia cuándo se observa una imagen o algo agradable pero resulta menos intuitivo en el caso de escribir algo). Para las observaciones de menor tiempo ($t=2$) la temperatura aparece como relevante.

Tabla 5.6: Resumen 5 características más seleccionadas para la clasificación de niveles de valencia.

t = 5		t = 2	
Sujetos sin pupila	Sujetos con pupila	Sujetos sin pupila	Sujetos con pupila
numPeaksFasica 3,41	numPeaksFasica 4,75	numPeaksFasica 3,18	numPeaksFasica 4,43
numFijaciones 3,13	promGSR 4,62	numFijaciones 2,65	numFijaciones 4,12
numSacadas 3,03	medianaTemp 4,37	numSacadas 2,56	numSacadas 4,12
promGSR 2,81	beta-alfaF8 4,37	promTemp 2,37	promGSR 4,06
powerGSR 2,79	numSacadas 4,31	medianaTemp 2,28	medianaTemp 4

- **Excitación:** en este caso se eliminan las características correspondientes a la señal del GSR y su componente fásica. Por lo tanto, quedan disponibles las características correspondientes a *eye tracker*, ECG, HR, PPG, temperatura y EEG (Tabla 4.3).

De la Tabla 5.7 se puede observar que las características extraídas del HR se repiten en todos los escenarios. También, la temperatura pareciera ser relevante. Para las observaciones más largas ($t=5$), el número de fijaciones aparece como importante también.

Tabla 5.7: Resumen 5 características más seleccionadas para la clasificación de niveles de excitación.

t = 5		t = 2	
Sujetos sin pupila	Sujetos con pupila	Sujetos sin pupila	Sujetos con pupila
numFijaciones 3,3	medianaTemp 4,43	rMSDD 3,29	SDNN 4,75
numSacadas 3,07	promTemp 4,25	SDNN 3,25	promTemp 4,68
AVNN 3,07	numFijaciones 4	medianaTemp 3,01	rMSDD 4,68
promHR 2,98	rmsHR 3,68	promTemp 2,96	numSacadas 4,62
rmsHR 2,98	promPupila 3,62	stdHR 2,8	stdHR 4,62

Con respecto a las características seleccionadas para carga cognitiva, valencia y excitación, resulta curioso que el EEG no sea protagonista, dado que existe amplia evidencia en la literatura de buenos porcentajes de clasificación solo utilizando el EEG, como por ejemplo en [39], [41] o [85].

Por otro lado, pareciera ser que el GSR y HR tendrían relevancia a la hora de clasificar valencia y excitación, respectivamente. Dado que ambas son parte del sistema nervioso autónomo simpático, que está encargado de preparar al cuerpo para posibles amenazas o estímulos externos, la aparición de las características de estas señales tiene sentido, dado que las emociones son el reflejo de una reacción del cuerpo ante un estímulo exterior y, dependiendo de cuál sea, el cuerpo debe prepararse (por ejemplo, al sentir miedo se debería querer huir, etc...).

5.2.2. Clasificación de carga cognitiva

5.2.2.1. Elección de tamaño de observaciones

Se prueba clasificar con observaciones de tamaño $t = 5$ segundos y $t = 2$ segundos, con el fin de observar la influencia del tamaño de la observación en la clasificación de carga cognitiva. Se prueban todos los clasificadores, y sus parámetros, mencionados en la Sección 4.5. Se presentan los mejores resultados obtenidos para cada clasificador (KNN, SVM y ANN) en las Tablas 5.8 y 5.9.

Tabla 5.8: Promedio de porcentajes de clasificación para la carga cognitiva para los mejores clasificadores encontrados, maximizando F1 con observaciones de tamaño $t = 2$ [s]

$t = 2$	Parámetros	Exactitud [%]	Precisión [%]	Sensibilidad [%]	F1 [%]
KNN	$k = 5$	$58,89 \pm 23,04$	$35,20 \pm 15,25$	$34,88 \pm 14,01$	$34,39 \pm 14,45$
SVM	kernel rbf $C = 10$	$59,90 \pm 22,06$	$36,29 \pm 14,95$	$35,83 \pm 13,83$	$35,57 \pm 14,33$
ANN	$\text{alfa} = 0,01$	$56,33 \pm 23,25$	$36,60 \pm 15,76$	$36,57 \pm 15,53$	$36,38 \pm 15,59$

Tabla 5.9: Promedio de porcentajes de clasificación para la carga cognitiva para los mejores clasificadores encontrados, maximizando F1 con observaciones de tamaño $t = 5$ [s]

$t = 5$	Parámetros	Exactitud [%]	Precisión [%]	Sensibilidad [%]	F1 [%]
KNN	$k = 10$	$56,35 \pm 20,69$	$35,70 \pm 17,94$	$35,64 \pm 15,54$	$34,27 \pm 16,82$
SVM	kernel rbf $C = 10$	$54,43 \pm 21,52$	$37,86 \pm 16,47$	$36,83 \pm 16,20$	$36,44 \pm 16,54$
ANN	$\text{alfa} = 0,1$	$51,88 \pm 22,13$	$36,70 \pm 18,16$	$35,82 \pm 16,69$	$35,63 \pm 17,15$

De las Tablas 5.8 y 5.9 se observa que, para la clasificación de niveles de carga cognitiva, se obtienen desempeños similares al optimizar F1. Se eligen observaciones de tamaños $t=2$ dado que maximizan la exactitud. También, que el mejor desempeño (en cuanto a la maximización de F1) lo obtiene el clasificador ANN con $\text{alfa} = 0,01$ (se puede ver el detalle de este clasificador para cada sujeto en la Tabla D.1, dentro del Anexo D). Este será el clasificador utilizado para las siguientes pruebas de carga cognitiva, a menos que se especifique lo contrario.

Con respecto a los resultados obtenidos, se observa que la exactitud obtenida es muy baja comparada con la literatura revisada, a modo de ejemplo en [8] obtienen un 80% de exactitud al clasificar carga cognitiva experimentada por programadores al corregir códigos con distintos niveles de dificultad. Aún así, dado que el promedio de clases es 4,13 (Tabla 4.5), el resultado de la clasificación es mejor que uno al azar (que sería una exactitud de 25% aprox).

Al observar los porcentajes de F1, se aprecia que a los clasificadores les cuesta distinguir entre los distintos niveles identificados por la clusterización, lo que podría ser señal de que: 1) Las características utilizadas no son las correctas para esta tarea, y/o, 2) Los niveles experimentados de carga cognitiva no son tan distinguibles en la tarea propuesta.

Con respecto al primer punto, existe evidencia en la literatura revisada de que las características utilizadas son útiles al distinguir los niveles buscados, pero en otro tipo de actividades [18], lo que podría indicar que no existe una diferencia entre carga cognitiva tan marcada al escribir un ensayo. En [85] se obtienen bajos porcentajes de clasificación (40 % de exactitud) al intentar clasificar 4 niveles de carga cognitiva muy similares entre sí, por lo que simplifican el problema clasificando los dos niveles más extremos de carga cognitiva presentes en su experimento. Pensando en el objetivo final de esta memoria, tendría sentido querer diferenciar cuando el alumno está con extremos de carga cognitiva más que cada nivel que experiencia.

5.2.2.2. Porcentaje de clasificación según número de clases

Dado que el número de clases no es homogéneo para todos los sujetos, interesa saber cómo se comporta el clasificador obtenido para los distintos números de clase. Se presenta el resultado en la Tabla 5.10, donde se observa un decrecimiento tanto en exactitud como en F1, al aumentar el número de clases.

Tabla 5.10: Promedio de porcentaje de clasificación por número de clase para carga cognitiva

Número de clases	Exactitud [%]	Precisión [%]	Sensibilidad [%]	F1 [%]
2	88,27±8,82	57,93±6,64	57,66±3,08	57,51±6,45
4	41,64±5,03	31,76±2,01	31,68±1,45	31,54±1,69
5	46,80±3,52	28,80±1,65	28,83±1,59	28,60±1,65
6	36,93 %± %5,51	21,28±2,43	21,60±2,73	21,38±2,55

5.2.2.3. Clasificación por sensor

Para conocer el aporte real de cada sensor a la clasificación (a diferencia del análisis exploratorio realizado en la Sección 5.2.1) se clasifican los mismos niveles de carga cognitiva pero utilizando cada sensor por separado.

Tabla 5.11: Promedio de porcentaje de clasificación para niveles de carga cognitiva, para cada sensor por separado.

Sensor	Exactitud [%]	Precisión [%]	Sensibilidad [%]	F1 [%]
Temperatura	57,19±23,30	35,49±17,06	33,97±15,47	33,30±15,88
EEG	49,31±26,22	30,08±15,45	29,97±15,03	28,85±15,14
GSR	55,71±24,00	29,68±18,46	31,66±16,86	29,59±18,20
PPG	51,34±25,23	30,24±14,63	30,12±14,52	29,86±14,16
ECG	53,68±23,41	28,89±15,21	30,41±14,41	28,64±15,43
Número fijaciones/ sacadas	62,14±21,38	30,31±16,19	33,27±12,78	30,46±13,84

Con respecto a los resultados obtenidos (Tabla 5.11), se observa que todos los sensores aportan de manera similar, siendo el con mayor F1 la temperatura y el número de fijaciones y sacadas. Esto indicaría que no existe un sensor que sea mejor para todas las personas en la medida de carga cognitiva. Este hecho resulta sorprendente, dado que, dentro de la literatura revisada, se demuestra que solo con EEG pueden llegar a altos porcentajes de clasificación, y en este caso, el EEG no sobresale por sobre el resto de los sensores (de hecho tiene un desempeño bastante bajo). Esto lleva a la pregunta ¿Serán las características extraídas las idóneas para esta tarea?

De este análisis, surgen dos nuevas pruebas, que se describen a continuación: probar con nuevas características para el EEG y probar utilizando solo dos clusters que representen los extremos de las cargas cognitivas vividas por los usuarios durante la escritura.

5.2.2.4. Nuevas características EEG para carga cognitiva

Con respecto a la carga cognitiva, existen muchas características de la señal EEG en la literatura. Por lo mismo, se prueba extrayendo las características utilizadas en [39], quiénes, utilizando las mismas características descritas en la Sección 4.3.6.1, pero para la banda δ en canales frontales, logran distinguir entre 7 niveles de carga cognitiva al realizar operaciones matemáticas con distintos niveles de dificultad, con una exactitud de 98 % para 12 sujetos.

Para estas nuevas características, se realiza nuevamente en análisis exploratorio de clasificadores (descrito en la Sección 4.5). Se muestra el desempeño del mejor clasificador obtenido en la Tabla 5.12.

Tabla 5.12: Desempeño promedio del mejor clasificador con nuevas características para el EEG

$t = 2$	Parámetros	Exactitud [%]	Precisión [%]	Sensibilidad [%]	F1 [%]
ANN	alfa = 0,01	57,62±22,95	37,31±15,26	36,25±14,59	36,28±14,69

De la Tabla 5.12 se observa de que el desempeño del clasificador cambia de manera poco significativa al clasificar con estas nuevas características de EEG (la exactitud aumenta en un 2,29 %, F1 disminuye en un 0,27 % y las desviaciones estándar también disminuyen en un 1,29 % y un 2,47 % respectivamente). Se teoriza sobre que, quizás, aún no se conozca bien el mecanismo de la escritura en el cerebro y que, por lo tanto, no se sabe cuáles características serían las más relevantes. Esto se puede extraer del mismo análisis que realizan en [39], dado que, dentro de la literatura revisada, la banda δ no es mencionada como una señal decidora sobre la variabilidad en el nivel de carga cognitiva de un usuario (a diferencia de las bandas α y θ que son ampliamente mencionadas). Quizás se podría encontrar algo interesante al realizar un experimento previo a este, donde se deban escribir textos breves con distintos niveles de dificultad (dado que los niveles de dificultad se relacionan con la carga cognitiva) y desde ahí, explorar el comportamiento de la señal del EEG.

5.2.2.5. Clasificación de nivel alto y bajo de carga cognitiva

Para cada participante, se seleccionan los dos clusters que más difieran en el valor promedio del tamaño de pupila y que tengan un número razonable de muestras. Luego, se prueba con todas las configuraciones de clasificadores nombrados en la Sección 4.5. Se presentan los mejores clasificadores de cada tipo en la Tabla 5.13, donde se puede ver que el mejor clasificador corresponde a SVM con kernel rbf y $C = 10$. Se detalla el desempeño de este clasificador para cada sujeto en la Tabla D.2, Anexo D.

Al comparar este desempeño (SVM Tabla 5.13) con el obtenido en la Tabla 5.12, se obtiene un aumento de 31,12% en exactitud promedio y de 68,53% en F1 promedio y una disminución total de la desviación estándar. Al compararlo con la clasificación para dos clases en la Tabla 5.10 se tiene una baja en exactitud promedio de 14,41% pero un aumento de F1 promedio en 6,32%. A pesar de esta mejora, los resultados siguen siendo más bajos que en la literatura revisada, a modo de ejemplo en [18] se alcanza una exactitud promedio de 89,58% al clasificar para dos niveles de carga cognitiva.

Tabla 5.13: Desempeño promedio de clasificadores para discernir entre nivel alto y bajo de carga cognitiva

$t = 2$	Parámetros	Exactitud [%]	Precisión [%]	Sensibilidad [%]	F1 [%]
KNN	$k = 3$	$74,89 \pm 12,26$	$61,39 \pm 7,29$	$60,05 \pm 6,80$	$60,21 \pm 6,99$
SVM	kernel rbf $C = 10$	$75,55 \pm 12,10$	$62,36 \pm 8,40$	$60,94 \pm 7,81$	$61,14 \pm 8,07$
ANN	alfa = 0,1	$74,38 \pm 12,13$	$61,69 \pm 7,87$	$60,96 \pm 7,16$	$61,10 \pm 7,40$

5.2.2.6. Clasificación entre-sujetos

Hasta ahora, todos los análisis realizados son para cada sujeto por separado (intra-sujeto). Interesa saber si se puede realizar un clasificador para todos los sujetos en conjunto. Para esto, se utiliza una línea base para normalizar las señales entre sujetos y hacerlas comparables. Como línea base se utilizan los 500 ms a partir del fin de la relajación (desde que el usuario abre los ojos) [86]. Una vez obtenida la línea base, se le resta a las señales y éstas se normalizan por z-score. Luego, se extraen las mismas características que para el caso intra-sujetos (Sección 4.3). Se clusteriza para diámetro y varianza de la pupila, obteniéndose un total de 5 niveles de carga cognitiva. Por último, se prueban distintos clasificadores (Sección 4.5). En este caso se cuenta con 11.510 datos y el procedimiento es el mismo del anterior (5 repeticiones, separación 70-30% en entrenamiento-prueba).

Se muestran los resultados de los mejores clasificadores obtenidos en la Tabla 5.14. Es importante notar que en este caso la desviación estándar corresponde a aquella obtenida por las 5 repeticiones realizadas para cada clasificador y no, como en el resto de las tablas, a la desviación estándar resultante del promedio de todos los sujetos.

Tabla 5.14: Desempeño de clasificación entre-sujetos para carga cognitiva

t=2	Parámetros	Exactitud [%]	Precisión [%]	Sensibilidad [%]	F1 [%]
KNN	k = 1	55,05±0,50	38,53±0,99	35,69±0,69	36,38±0,82
SVM	kernel Sigmoideo C = 10	53,84±1,04	38,36±1,01	37,21±1,13	37,63±1,07
ANN	alfa = 0.001	53,22±1,65	37,43±2,05	36,87±1,63	37,06±1,78

De la Tabla 5.14 se observa que aquel clasificador que maximiza F1 corresponde a SVM. Si se compara el desempeño obtenido con el desempeño para 5 clases intra-sujetos (Tabla 5.10), se observa una mejora en 15,04 % para la exactitud y en 31,57 % para F1. Además, al comparar el desempeño de F1 para estos 5 niveles vs el desempeño obtenido para los 4,14 niveles intra-sujeto (Tabla 5.8, clasificador ANN) se obtiene un aumento de 3,44 % en F1.

5.2.3. Clasificación de estado emocional

5.2.3.1. Tamaño de observaciones

Al igual que para el análisis de carga cognitiva, interesa saber con qué tamaño de observación se logra un mejor desempeño en los clasificadores y, además, encontrar el clasificador que obtiene un mejor desempeño promedio para todos los usuarios. Se presenta por separado los resultados obtenidos para niveles de valencia (Tablas 5.15 y 5.16) y para niveles de excitación (Tablas 5.17 y 5.18).

- **Valencia:** de las Tablas 5.15 y 5.16 se observa que para la clasificación de niveles de valencia, las observaciones con tamaño $t = 5[s]$ dan mejores resultados de clasificación, obteniéndose un alza de 66,87 % en exactitud y de 34,10 % en F1, al comparar ambos mejores clasificadores (ANN en Tabla 5.15 y KNN en Tabla 5.16).

Con respecto al clasificador, el con mejor desempeño es KNN con $k = 3$. Se elige este porque, dada la similitud entre F1s, se opta con el con mayor exactitud. Se puede ver el detalle para cada sujeto para este clasificador en la Tabla D.3 dentro del Anexo D. Este clasificador será el utilizado para las siguientes pruebas de niveles de valencia, a menos que se especifique lo contrario

Tabla 5.15: Promedio de porcentajes de clasificación para los niveles de valencia para los mejores clasificadores encontrados, maximizando F1 con observaciones de tamaño $t = 2 [s]$

t = 2	Parámetros	Exactitud [%]	Precisión [%]	Sensibilidad [%]	F1 [%]
KNN	k = 1	43,87±18,40	36,49±14,07	36,5±14,12	36,37±14,14
SVM	kernel rbf C = 10	48,34±18,71	37,68±14,36	37,01±13,68	36,15±13,80
ANN	alfa = 0,01	44,92±18,21	36,98±13,98	37,01±13,85	36,77±13,87

Tabla 5.16: Promedio de porcentajes de clasificación para los niveles de valencia para los mejores clasificadores encontrados, maximizando F1 con observaciones de tamaño $t = 5$ [s]

$t = 5$	Parámetros	Exactitud [%]	Precisión [%]	Sensibilidad [%]	F1 [%]
KNN	$k = 3$	$74,96 \pm 18,71$	$52,15 \pm 16,11$	$49,17 \pm 13,13$	$49,31 \pm 13,92$
SVM	kernel rbf $C = 100$	$72,12 \pm 18,50$	$51,16 \pm 14,07$	$49,96 \pm 14,41$	$49,86 \pm 13,91$
ANN	$\text{alfa} = 0,01$	$72,74 \pm 17,89$	$50,53 \pm 12,80$	$50,12 \pm 12,73$	$49,87 \pm 12,70$

- **Excitación:** de las Tablas 5.17 y 5.18 se puede ver que para niveles de excitación, al igual que para niveles de valencia, el tamaño más efectivo de observación es de $t = 5$ [s]. Al comparar ambos mejores clasificadores (ANN en Tablas 5.17 y 5.18), se puede ver un alza de 15,84% en exactitud y de 24,69% en F1.

El clasificador que presenta un mejor desempeño es una red neuronal con $\text{alfa} = 0.01$. Se puede ver el detalle para cada sujeto en la Tabla D.4, dentro del Anexo D. Este clasificador será el utilizado para las siguientes pruebas de niveles de excitación, a menos que se especifique lo contrario.

Tabla 5.17: Promedio de porcentajes de clasificación para los niveles de excitación para los mejores clasificadores encontrados, maximizando F1 con observaciones de tamaño $t = 2$ [s]

$t = 2$	Parámetros	Exactitud [%]	Precisión [%]	Sensibilidad [%]	F1 [%]
KNN	$k = 1$	$66,90 \pm 23,41$	$46,86 \pm 14,61$	$46,58 \pm 14,75$	$46,47 \pm 14,70$
SVM	kernel rbf $C = 10$	$71,60 \pm 22,60$	$50,10 \pm 15,90$	$47,37 \pm 14,74$	$47,37 \pm 15,14$
ANN	$\text{alfa} = 0,1$	$69,49 \pm 23,25$	$49,45 \pm 16,16$	$48,27 \pm 15,62$	$48,44 \pm 15,78$

Tabla 5.18: Promedio de porcentajes de clasificación para los niveles de excitación para los mejores clasificadores encontrados, maximizando F1 con observaciones de tamaño $t = 5$ [s]

$t = 5$	Parámetros	Exactitud [%]	Precisión [%]	Sensibilidad [%]	F1 [%]
KNN	$k = 3$	$82,14 \pm 13,99$	$65,85 \pm 16,88$	$57,66 \pm 12,19$	$59,11 \pm 13,67$
SVM	kernel rbf $C = 10$	$82,45 \pm 13,30$	$62,51 \pm 15,16$	$58,06 \pm 12,01$	$59,01 \pm 13,15$
ANN	$\text{alfa} = 0.01$	$80,50 \pm 14,00$	$61,86 \pm 14,62$	$60,06 \pm 13,10$	$60,46 \pm 13,61$

Sobre ambos resultados obtenidos, tanto para valencia (Tabla 5.16) como para excitación (Tabla 5.18) se repite la situación obtenida para carga cognitiva, donde el clasificador le cuesta distinguir entre una clase y otra (% de F1 más bajo que de exactitud).

5.2.3.2. Promedio de porcentajes de clasificación según número de clases

- **Valencia:** en la Tabla 5.19 se ve que existe un decaimiento importante para F1 entre 2 y 3 clases y para la exactitud entre 3 y 4 clases. Además, al igual que para carga cognitiva, se observa una relación inversamente proporcional entre número de clases y desempeño del clasificador.

Tabla 5.19: Promedio de porcentajes de clasificación por número de clase para nivel de valencia

Número de clases	Exactitud [%]	Precisión [%]	Sensibilidad [%]	F1 [%]
2	84,28±10,86	60,04±12,18	56,43±7,58	56,79±8,78
3	68,78±14,21	39,04±6,17	38,11±4,15	37,39±4,88
4	49,52±3,48	36,25±3,43	34,55±3,66	34,16±4,28
5	44,05±9,45	27,59±12,21	26,53±9,06	26,32±10,23
6	37,08±4,90	30,05±6,16	25,74±2,31	26,48±3,16

- **Excitación:** en la Tabla 5.20 se mantiene la tendencia de una baja en el desempeño de clasificación al aumentar el número de clases. También, al comparar las métricas F1 con las otras categorías a clasificar (Tabla 5.10 para carga cognitiva y 5.19 para valencia) se ve que los niveles de excitación, para 2 y 3 clases, pareciera ser para la que mejor se logra diferenciar entre clases, lo que podría indicar que existen niveles más marcados que para carga cognitiva y valencia o características que logran diferenciar de mejor manera.

Tabla 5.20: Promedio de porcentajes de clasificación por número de clase para nivel de excitación

Número de clases	Exactitud [%]	Precisión [%]	Sensibilidad [%]	F1 [%]
2	84,46±8,77	64,84±12,13	63,05±10,46	63,46±10,99
3	72,05±11,71	57,22±11,99	54,19±10,22	54,97±10,75
5	39,36±0,95	27,46±1,20	27,95±1,29	27,42±1,35

De los resultados recién analizados (Tablas 5.19 y 5.20) y de los resultados analizados para carga cognitiva (Tabla 5.10), se sigue que el desempeño del clasificador baja a medida que aumenta el número de clases. Esto podría indicar que al aumentar la cantidad de niveles estos se hacen menos distinguibles entre sí. Dado que el promedio de clases para valencia es 2,44 (Tabla 4.6) y para excitación es 2,24 (Tabla 4.7) más que sean muchos niveles poco distinguibles como es el caso de la carga cognitiva, donde al utilizar solo los niveles extremos mejora el desempeño total del clasificador (Tabla 5.13), pareciera ser que los dos niveles obtenidos tanto para valencia y excitación no son tan distinguibles entre sí.

5.2.3.3. Clasificación por sensor

- **Valencia:** De la Tabla 5.21 se observa que aquel sensor que maximiza F1 es el de GSR, lo que está en línea con las características más seleccionadas para esta categoría (Tabla 5.6). Nuevamente se repite que no existe un sensor que sea mejor para todos los sujetos, si no, todos pareciesen tener un aporte similar.

Tabla 5.21: Promedio de porcentajes de clasificación para niveles de valencia, para cada sensor por separado.

Sensor	Exactitud [%]	Precisión [%]	Sensibilidad [%]	F1 [%]
GSR	73,64±19,86	48,01±13,89	47,46±12,71	46,48±12,91
Temperatura	74,03±19,27	47,34±13,99	46,15±12,07	45,66±12,55
EEG	72,31±20,93	47,84±15,51	45,82±12,83	45,56±13,51
Eye Tracker	72,20±20,96	42,78±13,72	43,78±12,27	42,32±12,89

- **Excitación:** Desde la Tabla 5.22 se evidencia que los sensores que más distinguen entre niveles de excitación son el EEG y el ECG. El ECG se preveía, dado que sus características aparecen repetidamente en el análisis realizado en la Sección 5.2.1 (Tabla 5.7). Con respecto al aporte del EEG, quizás la presencia de éste sea la razón porqué la clasificación de excitación entrega mejor desempeño que la clasificación de carga cognitiva o valencia.

Tabla 5.22: Promedio de porcentajes de clasificación para niveles de excitación, para cada sensor por separado.

Sensor	Exactitud [%]	Precisión [%]	Sensibilidad [%]	F1 [%]
Temperatura	82,87±11,86	56,24±18,09	50,65±8,96	49,30±11,07
EEG	82,99±14,26	60,21±18,66	56,73±14,59	56,52±16,12
Eye Tracker	82,14±13,05	45,25±12,57	47,78±8,20	44,87±8,96
PPG	82,08±13,06	52,10±15,06	49,21±9,06	47,35±10,05
ECG	82,47±13,18	58,22±17,18	53,05±11,36	52,70±12,99

Con respecto al análisis recién realizado sobre los resultados obtenidos en la clasificación de emociones, se observa de que, a pesar de utilizar las características mencionadas en la literatura con buenos resultados, en este caso el clasificador pareciera tener dificultad al diferenciar entre niveles de emoción experimentados por el usuario, especialmente con los de valencia. La principal diferencia entre la literatura y este experimento es que en la mayoría de la literatura revisada se clasifican señales de emociones claras inducidas en sujetos (con imágenes o sonidos especiales para ello)[14][15][16] e incluso se eligen las imágenes y/o sonidos que estén en los extremos del plano valencia-excitación (Figura 2.1)[29], mientras que en este caso se intentan medir emociones en un contexto más cotidiano, lo que lleva a nuevas preguntas: ¿Se pueden medir emociones en el día a día? ¿Se pueden medir emociones al escribir? ¿Son éstas lo bastante claras y marcadas? ¿Se comporta el cerebro de la misma manera al expresar una emoción proveniente de una imagen que al escribir algo?

Con respecto a la última pregunta planteada, en [43] se realiza un análisis sobre las características más relevantes de EEG al clasificar emociones. Se obtienen características diferentes

para usuarios viendo imágenes y para usuarios viendo imágenes al mismo tiempo que se escucha música (ambos estímulos para inducir un mismo nivel de valencia/excitación). De esto, se podría extrapolar que, quizás, falta un paso previo para lograr medir valencia/excitación en el contexto de interés y es el de primero ver cómo se comporta el cerebro en los distintos usuarios, al escribir textos que generen emociones. Para esto, se propone realizar un nuevo experimento donde los usuarios escriban textos con temas ya etiquetados con ciertos niveles de valencia/excitación, para poder estudiar cómo se comportan las señales y cuáles serían las características más adecuadas para un escenario más cotidiano (como es el de esta memoria).

Por otro lado, con respecto a las emociones, la pregunta elegida para este ensayo no fue pensada en alguna que generara algún tipo de reacción en el usuario, por lo que, quizás, los grupos de emociones mostrados en los clusters sean demasiado similares las unas con las otras como para poder diferenciarlas claramente.

A partir de esto, y pensando en el objetivo final de esta memoria, es que se recomienda explorar la idea de tomar otro camino e intentar medir el índice de motivación de la persona [87], en vez de las emociones que pueda o no estar sintiendo. En [87] se desarrolla un índice a base del EEG que mide la motivación (*engagement*) del usuario frente a una tarea. Se define según la ecuación 5.1:

$$\text{índice motivación} = \frac{\beta}{\alpha + \theta}, \quad (5.1)$$

con β , α y θ la potencia combinada (suma de la potencia de la banda en todos los canales medidos) de las bandas de frecuencia del EEG. En [87] utilizan los electrodos Cz, T₅, P₃, Pz, P₄, P₁ y O₂ del sistema 10-20 y en [49] utilizan los electrodos P₃, C₃, Pz, Fz del sistema 10-20. Por otro lado en [88] se utilizan los canales occipitales del EEG Emotiv y se desarrolla un sistema para el monitoreo de la motivación de niños mientras leen lecciones escolares. Cuando el niño tiene un índice de motivación bajo, se desencadena un estímulo interactivo. Se demuestra que se mejora el aprendizaje y rendimiento del niño bajo este sistema.

Así, como primera aproximación, se calcula el índice de motivación de los canales occipitales [88], para todos los usuarios en las distintas etapas del ensayo, en ventanas de 5 segundos y se calcula su promedio. Se muestran los resultados obtenidos en la Tabla 5.23, donde se puede observar que, en promedio, la motivación de los usuarios aumenta al ir avanzando en el ensayo.

Tabla 5.23: Índice de motivación promedio para todos los sujetos en cada etapa del ensayo

Borrador	Introducción	Tesis	Desarrollo 1	Desarrollo 2	Conclusión
0,659 ± 0,202	0,670 ± 0,197	0,676 ± 0,202	0,677 ± 0,210	0,699 ± 0,230	0,702 ± 0,240

A partir de esto, sería interesante ver si se puede medir este mismo índice de motivación, pero con sensores que sean menos invasivos y más cómodos para el usuario. También se podría fijar un umbral de baja/alta motivación y reaccionar en torno a eso (como en [88]).

Capítulo 6

Conclusiones y trabajo futuro

Este trabajo se centra en la investigación sobre la medición a nivel individual de la carga cognitiva y valencia/excitación experimentado por un conjunto de usuarios mientras escriben un ensayo. Esto, como primera aproximación en el desarrollo de una plataforma basada en sensores psico-fisiológicos para monitorear, y ayudar, a estudiantes durante actividades de redacción.

Sobre el análisis estadístico realizado, se concluye que las señales diferencian entre las diferentes etapas del ensayo dentro de un intervalo de confianza de 95 %. También que para cada etapa del ensayo, las señales del eye tracker, ECG, HR, GSR y EEG pueden diferenciar los niveles de carga cognitiva; las señales de ECG, GSR, PPG y EEG pueden diferenciar los niveles de valencia y que las señales de HR, PPG y EEG pueden diferenciar los niveles de excitación (dentro de un intervalo de confianza de 95 %).

Sobre la clasificación, se puede concluir que para el contexto medido, las señales psico-fisiológicas logran diferenciar entre niveles de carga cognitiva, valencia y excitación, pero no logran hacerlo de manera tan efectiva en comparación a lo reportado para otras actividades en la literatura.

En particular:

- Con respecto a la clasificación de carga cognitiva, se obtienen desempeños promedios de clasificación relativamente bajos (exactitud 56,33 % y F1 36,38 %) al intentar clasificar todas los niveles detectados por la clusterización (4,14 en promedio). Se obtienen mejores resultados al seleccionar dos niveles de carga cognitiva (exactitud 75,55 % y F1 de 61,14 %). Esto resultaría útil para detectar cuando el estudiante está con una carga alta (le está costando la actividad) o baja (es muy fácil y se aburre). Además, se realiza un análisis entre-sujetos donde, al comparar los resultados de la clasificación de 5 niveles obtenidos para la clasificación entre-sujetos con la clasificación de 4,14 niveles obtenidos en el análisis intra-sujetos se obtiene un aumento en el desempeño de F1 en 3,44 %.
- Sobre la clasificación de valencia y excitación, se obtiene un mejor desempeño al clasificar excitación (80,50 % de exactitud y 60,49 % de F1) que valencia (74,96 % de exactitud

y 49,31 % de F1).

Esta diferencia entre los desempeños reportados por la literatura y los obtenidos se pueden deber a diversos factores. Se postula que los principales son: las características extraídas y la actividad en sí (la escritura de este ensayo en particular).

Sobre las características, al analizar las clasificaciones para cada sensor por separado, se obtiene que ninguno sobresale particularmente sobre los otros. En particular, esto indicaría que el EEG, que ha demostrado buenos resultados por sí solo en otros contextos para la clasificación de carga cognitiva y valencia o excitación, está siendo subutilizado. Esto, en el sentido de que es una señal muy compleja y que las características extraídas en este trabajo no reflejarían todo lo que puede entregar.

Con respecto a la actividad en sí, al analizar el desempeño del clasificador para el distinto número de niveles de carga cognitiva y excitación/valencia, se tiene una baja de desempeño al aumentar el nivel de estados analizados. Lo que implicaría que no se diferencia bien entre niveles, ni de carga cognitiva ni de valencia o excitación. Con respecto a la carga cognitiva, se comprueba esta hipótesis al realizar el análisis para solo dos niveles y obtener un aumento en el desempeño del clasificador. Con respecto a las emociones, dentro de la literatura la clasificación siempre se realiza para emociones que se inducen de manera clara y fuerte, no para tareas cotidianas como ésta, por lo que, quizás, el bajo desempeño de F1 se deba a que los niveles identificados no son tan diferenciables entre sí.

En base a estos resultados, se propone como trabajo futuro:

- Profundizar en el análisis de características para EEG (tanto en carga cognitiva como estado emocional), dentro del contexto de interés. Para esto, se podrían hacer experimentos de varios textos cortos que sean de distintos niveles de dificultad (induciendo distintos niveles de carga cognitiva) y/o que sean de temas polémicos o que generen una reacción en el usuario (para obtener emociones marcadas) y etiquetar de acuerdo a la percepción del usuario o de acuerdo al experimento en sí. Esto, para poder investigar a fondo el comportamiento de las bandas del EEG y aprovechar su potencialidad al máximo.
- En este trabajo solo se utilizaron las personas con medidas de pupila válida para la clasificación de carga cognitiva. Dado que la clasificación entre-sujetos presenta una leve mejora con un mayor número de clases, se podrían utilizar todos los datos válidos de todos los participantes (no solo de aquellos 16 con pupila válida), con el fin de tener mayor cantidad de datos de entrenamiento, y probar la clasificación de carga cognitiva con *deep learning*.
- En paralelo a intentar medir valencia y excitación, se podría profundizar en el índice de motivación (ecuación 5.1). Esto, dado que sería útil poder saber cómo enfrenta el estudiante la tarea que debe realizar y, en el caso de que esté desmotivado o se desmotive durante la actividad, poder tomar medidas contra esto y así garantizar su aprendizaje. También sería interesante saber si se puede medir de manera menos invasiva que con el EEG.

Bibliografía

- [1] A. Medina, “Enseñar a leer y a escribir: ¿en qué conceptos fundamentar las prácticas docentes?,” *Psykhé (Santiago)*, vol. 15, nov 2006.
- [2] S. Eyzaguirre, “Lecciones del simce de escritura.” <https://www.cepchile.cl/lecciones-del-simce-de-escritura/cep/2016-10-04/130520.html>. Visitado: 15/06/2017.
- [3] F. Donoso, A. Gallardo, and A. Matus, “La experiencia de Chile en PIAAC.” <http://www.mideuc.cl/wp-content/uploads/2016/08/La-experiencia-de-Chile-en-Piaac.pdf>, aug 2016. Visitado: 17/01/2018.
- [4] H. Beyer, “La escritura en escolares chilenos.” <https://www.cepchile.cl/la-escritura-en-escolares-chilenos/cep/2016-03-04/095035.html>. Visitado: 15/06/2017.
- [5] C. Sotomayor, P. Bedwell, G. Gómez, A. M. Domínguez, and E. Jéldrez, “Estudio sobre la calidad de la escritura en la educación básica,” in *Tercer congreso interdisciplinario de investigación en educación*, Centro de Investigación Avanzada en Educación Universidad de Chile and Centro de Estudios de Políticas y Prácticas en Educación Universidad Católica de Chile, aug 2014.
- [6] “Web Science and Smart Technologies Lab.” <http://wesstlab.dii.uchile.cl>. Visitado: 15/06/2017.
- [7] E. Haapalainen, S. Kim, J. F. Forlizzi, and A. K. Dey, “Psycho-physiological measures for assessing cognitive load,” in *Proceedings of the 12th ACM international conference on Ubiquitous computing - Ubicomp 10*, ACM Press, 2010.
- [8] T. Fritz, A. Begel, S. C. Müller, S. Yigit-Elliott, and M. Züger, “Using psycho-physiological measures to assess task difficulty in software development,” in *Proceedings of the 36th International Conference on Software Engineering - ICSE 2014*, ACM Press, 2014.
- [9] H. Schultheis and A. Jameson, “Assessing cognitive load in adaptive hypermedia systems: Physiological and behavioral methods,” in *Lecture Notes in Computer Science*, pp. 225–234, Springer Berlin Heidelberg, 2004.

- [10] E. Ferreira, D. Ferreira, S. Kim, P. Siirtola, J. Roning, J. F. Forlizzi, and A. K. Dey, “Assessing real-time cognitive load based on psycho-physiological measures for younger and older adults,” in *2014 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, IEEE, dec 2014.
- [11] G. F. Wilson and C. A. Russell, “Real-time assessment of mental workload using psychophysiological measures and artificial neural networks,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 45, no. 4, pp. 635–643, 2003.
- [12] P. Rani, C. Liu, N. Sarkar, and E. Vanman, “An empirical study of machine learning techniques for affect recognition in human–robot interaction,” *Pattern Analysis and Applications*, vol. 9, pp. 58–69, apr 2006.
- [13] C. Lee, S. Yoo, Y. Park, N. Kim, K. Jeong, and B. Lee, “Using neural network to recognize human emotions from heart rate variability and skin resistance,” in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, IEEE, 2005.
- [14] K. Schaaff and T. Schultz, “Towards emotion recognition from electroencephalographic signals,” in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, IEEE, sep 2009.
- [15] A. Haag, S. Goronzy, P. Schaich, and J. Williams, “Emotion recognition using biosensors: First steps towards an automatic system,” in *Lecture Notes in Computer Science*, pp. 36–48, Springer Berlin Heidelberg, 2004.
- [16] K. H. Kim, S. W. Bang, and S. R. Kim, “Emotion recognition system using short-term monitoring of physiological signals,” *Medical & Biological Engineering & Computing*, vol. 42, pp. 419–427, may 2004.
- [17] H. Lira, “Medición de cargas cognitivas durante actividades de interacción humano computador en ambiente móvil usando sensores psico-fisiológicos,” Tesis para optar al grado de Ingeniero Civil Industrial, Universidad de Chile. 2015.
- [18] C. Retamal, “Estudio del comportamiento de la carga cognitiva de usuarios que navegan en un sitio web,” Tesis para optar al grado de Ingeniero Civil Eléctrico, Universidad de Chile. 2017.
- [19] A. Torresi, “Escritura: Un estudio de las habilidades de los estudiantes de américa latina y el caribe,” UNESCO, dec 2010.
- [20] V. M. M. Martínez and S. S. J. Ríos, “Guía para la elaboración de ensayos de investigación (ensayo de un ensayo),” *Revista del Centro de Investigación. Universidad La Salle*, vol. 7, no. 26, pp. 63–79, 2006.
- [21] J. J. G. van Merriënboer and J. Sweller, “Cognitive load theory and complex learning: Recent developments and future directions,” *Educational Psychology Review*, vol. 17, pp. 147–177, jun 2005.
- [22] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. V. Gerven, “Cognitive load measu-

- rement as a means to advance cognitive load theory,” *Educational Psychologist*, vol. 38, pp. 63–71, mar 2003.
- [23] T. Olive, “Working memory in writing: Empirical evidence from the dual-task technique,” *European Psychologist*, vol. 9, pp. 32–42, jan 2004.
- [24] R. Vanderberg and H. L. Swanson, “Which components of working memory are important in the writing process?,” *Reading and Writing*, vol. 20, pp. 721–752, dec 2006.
- [25] R. A. Alves, S. L. Castro, and T. Olive, “Execution and pauses in writing narratives: Processing time, cognitive effort and typing skill,” *International Journal of Psychology*, vol. 43, pp. 969–979, dec 2008.
- [26] C. E. Valderrama and G. Ulloa, “Spectral analysis of physiological parameters for emotion detection,” in *2012 XVII Symposium of Image, Signal Processing, and Artificial Vision (STSIVA)*, IEEE, sep 2012.
- [27] S. A. Hosseini, M. A. Khalilzadeh, and S. Changiz, “Emotional stress recognition system for affective computing based on bio-signals,” *Journal of Biological Systems*, vol. 18, pp. 101–114, oct 2010.
- [28] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, “Physiological signals based human emotion recognition: a review,” in *2011 IEEE 7th International Colloquium on Signal Processing and its Applications*, IEEE, mar 2011.
- [29] R. Ramirez and Z. Vamvakousis, “Detecting emotion from EEG signals using the emotive epoc device,” in *Brain Informatics*, pp. 175–184, Springer Berlin Heidelberg, 2012.
- [30] J. T. Cacioppo, L. G. Tassinary, and G. Berntson, eds., *Handbook of Psychophysiology*. Cambridge University Press, 2007.
- [31] A. Poole and L. J. Ball, “Eye tracking in hci and usability research,” in *Encyclopedia of Human Computer Interaction*, pp. 211–214, Idea Group Reference, 2006.
- [32] F. E. L. Sarmiento, D. G. Prada, and C. Gutierrez, “Pupila, pupilometría y pupilografía,” *Acta Neurol Colomb*, vol. 24, pp. 188–197, oct 2008.
- [33] S. T. Iqbal, P. D. Adamczyk, X. S. Zheng, and B. P. Bailey, “Towards an index of opportunity: Understanding changes in mental workload during task execution,” in *In Proc of CHI 2005*, ACM Press, pp. 311–320, 2005.
- [34] I. B. R. Platform, “EEG Pocket Guide,” 2016.
- [35] P. Antonenko, F. Paas, R. Grabner, and T. van Gog, “Using electroencephalography to measure cognitive load,” *Educational Psychology Review*, vol. 22, pp. 425–438, apr 2010.
- [36] R. N. Roy, S. Bonnet, S. Charbonnier, and A. Campagne, “Mental fatigue and working memory load estimation: Interaction and implications for EEG-based passive BCI,” in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and*

Biology Society (EMBC), IEEE, jul 2013.

- [37] W. Klimesch, B. Schack, and P. Sauseng, “The functional significance of theta and upper alpha oscillations,” *Experimental Psychology*, vol. 52, pp. 99–108, jan 2005.
- [38] W. Klimesch, “EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis,” *Brain Research Reviews*, vol. 29, pp. 169–195, apr 1999.
- [39] P. Zarjam, J. Epps, and N. H. Lovell, “Beyond subjective self-rating: EEG signal classification of cognitive workload,” *IEEE Transactions on Autonomous Mental Development*, vol. 7, pp. 301–310, dec 2015.
- [40] H. U. Amin, A. S. Malik, R. F. Ahmad, N. Badruddin, N. Kamel, M. Hussain, and W.-T. Chooi, “Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques,” *Australasian Physical & Engineering Sciences in Medicine*, vol. 38, pp. 139–149, feb 2015.
- [41] L. Zhiwei and S. Minfen, “Classification of mental task EEG signals using wavelet packet entropy and SVM,” in *2007 8th International Conference on Electronic Measurement and Instruments*, IEEE, aug 2007.
- [42] R. F. Ahmad, A. S. Malik, H. U. Amin, N. Kamel, and F. Reza, “Classification of cognitive and resting states of the brain using EEG features,” in *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, IEEE, may 2016.
- [43] A. Choppin, “EEG-Based Human Interface for Disabled Individual: Emotion Expression with Neural Networks,” Thesis submitted for the Master Degree. Department of Information Processing, Interdisciplinary Graduate School of Science and Engineering. Tokyo Institute of Technology, 2000.
- [44] H. Xu and K. N. Plataniotis, “Subject independent affective states classification using EEG signals,” in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, IEEE, dec 2015.
- [45] A. T. Sohaib, S. Qureshi, J. Hagelbäck, O. Hilborn, and P. Jerčić, “Evaluating classifiers for emotion recognition using EEG,” in *Foundations of Augmented Cognition*, pp. 492–501, Springer Berlin Heidelberg, 2013.
- [46] Y. Liu, O. Sourina, and M. K. Nguyen, “Real-time EEG-based emotion recognition and its applications,” in *Lecture Notes in Computer Science*, pp. 256–277, Springer Berlin Heidelberg, 2011.
- [47] N. Nourbakhsh, Y. Wang, and F. Chen, “GSR and blink features for cognitive load classification,” in *Human-Computer Interaction – INTERACT 2013*, pp. 159–166, Springer Berlin Heidelberg, 2013.
- [48] I. B. R. Platform, “GSR Pocket Guide,” 2016.
- [49] M. Chaouachi and C. Frasson, “Exploring the relationship between learner EEG mental

- engagement and affect,” in *Intelligent Tutoring Systems*, pp. 291–293, Springer Berlin Heidelberg, 2010.
- [50] M. Malik, “Heart rate variability,” *Annals of Noninvasive Electrocardiology*, vol. 1, pp. 151–181, apr 1996.
- [51] J. T. Cacioppo and C. A. Sandman, “Physiological differentiation of sensory and cognitive tasks as a function of warning, processing demands, and reported unpleasantness,” *Biological Psychology*, vol. 6, pp. 181–192, apr 1978.
- [52] P. A. Boudewyns, “A comparison of the effects of stress vs. relaxation instruction on the finger temperature response,” *Behavior Therapy*, vol. 7, pp. 54–67, jan 1976.
- [53] E. Gomez, E. Salazar-López, E. Dominguez, and M. J. D. C. Serrano, “Neurotermografía y termografía psicósomática,” Fundación Internacional Artecittà, apr 2015.
- [54] I. Guyon, *Feature extraction : foundations and applications*. Berlin: Springer-Verlag, 2006.
- [55] J. Vesanto and E. Alhoniemi, “Clustering of the self-organizing map,” *IEEE Transactions on Neural Networks*, vol. 11, pp. 586–600, may 2000.
- [56] W. H. Press, A. S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes The Art of Scientific Computing*. Cambridge, 3rd ed., sept 2007.
- [57] M. Gardner and S. Dorling, “Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences,” *Atmospheric Environment*, vol. 32, pp. 2627–2636, aug 1998.
- [58] K. Murphy, *Machine learning: a probabilistic perspective*. Cambridge, Mass: MIT Press, 2012.
- [59] K.-Q. Shen, C.-J. Ong, X.-P. Li, Z. Hui, and E. Wilder-Smith, “A feature selection method for multilevel mental fatigue EEG classification,” *IEEE Transactions on Biomedical Engineering*, vol. 54, pp. 1231–1237, jul 2007.
- [60] D. C. W. I. Rui Xu, *Clustering*. Oxford: Wiley, 2009.
- [61] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, nov 1987.
- [62] C.-W. Hsu and C.-J. Lin, “A comparison of methods for multiclass support vector machines,” *IEEE Transactions on Neural Networks*, vol. 13, pp. 415–425, mar 2002.
- [63] C. Bishop, *Pattern recognition and machine learning*. New York: Springer, 2006.
- [64] K. Yu, J. Epps, and F. Chen, “Mental workload classification via online writing features,” in *2013 12th International Conference on Document Analysis and Recognition*, IEEE,

aug 2013.

- [65] L. Bourke and A.-M. Adams, “Cognitive constraints and the early learning goals in writing,” *Journal of Research in Reading*, vol. 33, pp. 94–110, feb 2010.
- [66] C. Ikehara and M. Crosby, “Assessing cognitive load with physiological sensors,” in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, IEEE.
- [67] N. Nourbakhsh, Y. Wang, F. Chen, and R. A. Calvo, “Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks,” in *Proceedings of the 24th Australian Computer-Human Interaction Conference on - OzCHI 12*, ACM Press, 2012.
- [68] A. Jiménez-Molina, C. Retamal, and H. Lira, “Using psycho-physiological sensors to asses mental workload in web browsing.” Working Paper.
- [69] A. Jiménez-Molina and H. Lira, “Towards a continuous assessment of cognitive workload for smartphone multitasking users,” in *International Symposium on Human Mental Workload: Models and Applications*, (Dublin, Ireland), 2017.
- [70] Shimmer, *GSR+ User Guide*. Real Time Technologies Ltd, revision 1.12 ed., 2017.
- [71] W. Boucsein, *Electrodermal Activity*. Springer US, second ed., 2012.
- [72] C. Kappeler-Setz, F. G. J. Schumm, B. Arnrich, and G. Tröster, “Towards long term monitoring of electrodermal activity in daily life,” *Personal and Ubiquitous Computing*, vol. 17, pp. 261–271, feb 2013.
- [73] E. Prasad, “Wearable electrodermal activity sensing - A validation study,” In partial fulfillment of the requirement for the degree Master of Science. Department of Computer Science, University of Houston. Dec 2013.
- [74] Shimmer, *Optical Pulse Sensor User Guide*. Real Time Technologies Ltd, revision 1.6 ed., 2016.
- [75] M. Patel, S. Lal, D. Kavanagh, and P. Rossiter, “Applying neural network analysis on heart rate variability data to assess driver fatigue,” *Expert Systems with Applications*, vol. 38, pp. 7235–7242, jun 2011.
- [76] Emotiv, *Emotiv Epoc User Manual*, 2014.
- [77] “Cómo posicionar al participante y el eye tracker.” <https://www.tobiipro.com/learn-and-support/learn/steps-in-an-eye-tracking-study/run/how-to-position-the-participant-and-the-eye-tracker/>. Último acceso: 13-10-2017.
- [78] M. Benedek and C. Kaernbach, “A continuous measure of phasic electrodermal activity,” *Journal of Neuroscience Methods*, vol. 190, pp. 80–91, jun 2010.

- [79] P. S. Hamilton and W. J. Tompkins, “Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database,” *IEEE Transactions on Biomedical Engineering*, vol. BME-33, pp. 1157–1165, dec 1986.
- [80] A. Quintero, M. Risk, and S. Liberczuk, “Eeg preprocessing with hampel filters,” pp.–, 2012.
- [81] T. N. A. S. o. P. T. F. of The European Society of Cardiology and Electrophysiology, “Heart rate variability : Standards of measurement, physiological interpretation, and clinical use,” *Circulation*, vol. 93, pp. 1043–1065, mar 1996.
- [82] J. M. Medeiros, “Development of a Heart Rate Variability analysis tool,” Thesis submitted for the Master Degree in Biomedical Engineering. Departamento de Física. Facultad de Ciências e Tecnologia, Universidade de Coimbra, 2010.
- [83] P. Welch, “The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms,” *IEEE Transactions on Audio and Electroacoustics*, vol. 15, pp. 70–73, jun 1967.
- [84] P. J. Lang, “The emotion probe: Studies of motivation and attention.,” *American Psychologist*, vol. 50, no. 5, pp. 372–385, 1995.
- [85] M. Honal and T. Schultz, “Determine task demand from brain activity,” 12 2017.
- [86] P. Van Gerven, F. Paas, J. J. G. Van Merrienboer, and H. G Schmidt, “Memory load and the cognitive pupillary response,” vol. 41, pp. 167–74, 04 2004.
- [87] A. T. Pope, E. H. Bogart, and D. S. Bartolome, “Biocybernetic system evaluates indices of operator engagement in automated task,” *Biological Psychology*, vol. 40, pp. 187–195, may 1995.
- [88] J. Huang, C. Yu, Y. Wang, Y. Zhao, S. Liu, C. Mo, J. Liu, L. Zhang, and Y. Shi, “Focus: Enhancing children’s engagement in reading by using contextual bci training sessions,” in *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI ’14, (New York, NY, USA), pp. 1905–1908, ACM, 2014.

Apéndice A

Grupo experimental

Se detalla sobre el grupo experimental, todos aquellos estudiantes no especificados se refieren a estudiantes de Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. Se destacan aquellos que tienen la medida de pupila válida (según lo explicado en la sección 4.2.6).

Tabla A.1: Grupo experimental

N	Género	Edad	Ocupación
1	Masculino	24	Estudiante
2	Masculino	25	Ingeniero Comercial
3	Masculino	23	Estudiante
4	Femenino	24	Estudiante
5	Masculino	33	Estudiante Doctorado
6	Femenino	24	Estudiante
7	Femenino	24	Estudiante
8	Masculino	24	Ingeniero
9	Masculino	23	Estudiante
10	Femenino	23	Estudiante
11	Masculino	24	Estudiante
12	Masculino	25	Estudiante Magister
13	Masculino	24	Estudiante
14	Masculino	26	Estudiante Literatura
15	Masculino	25	Estudiante Magister
16	Masculino	26	Estudiante
17	Femenino	22	Estudiante
18	Femenino	23	Estudiante
19	Femenino	24	Estudiante
20	Masculino	25	Estudiante
21	Masculino	26	Estudiante
22	Masculino	24	Estudiante Magister
23	Masculino	26	Estudiante Magister

N	Género	Edad	Ocupación
24	Masculino	29	Estudiante
25	Masculino	23	Estudiante
26	Masculino	23	Estudiante
27	Femenino	21	Estudiante
28	Masculino	26	Estudiante Magister
29	Masculino	27	Ingeniero
30	Femenino	23	Estudiante
31	Masculino	25	Estudiante
32	Femenino	26	Ingeniera Comercial
33	Femenino	32	Estudiante
34	Masculino	22	Estudiante
35	Masculino	27	Estudiante
36	Masculino	24	Estudiante Magister
37	Femenino	25	Profesora de Lenguaje
38	Femenino	24	Estudiante Pedagogía
39	Masculino	22	Estudiante
40	Masculino	26	Estudiante Magister
41	Masculino	24	Estudiante Magister
42	Femenino	24	Estudiante
43	Femenino	22	Estudiante
44	Masculino	24	Estudiante Magister
45	Masculino	24	Estudiante
46	Masculino	24	Licenciado
47	Masculino	24	Estudiante
48	Masculino	25	Estudiante
49	Femenino	24	Estudiante Magister
50	Masculino	22	Estudiante
51	Masculino	25	Estudiante Magister
52	Masculino	28	Ingeniero
53	Masculino	23	Estudiante Magister
54	Masculino	25	Estudiante
55	Masculino	24	Estudiante Magister

Apéndice B

Cantidad de clusters por participante

Se detalla para cada sujeto, la cantidad de niveles de carga cognitiva (Tabla B.1), valencia (Tabla B.2) y excitación (Tabla B.3). Para cada sujeto, se muestra el número de clusters original y final (luego de eliminados aquellos con menos de 5 muestras). Se destaca en **negrita** cuando un sujeto queda con una sola clase, lo que implica que no se puede realizar clasificación sobre sus datos.

B.1. Carga Cognitiva

Tabla B.1: Niveles de carga cognitiva por usuario

Sujeto	t = 5		t = 2	
	N inicial de clases	N final de clases	N inicial de clases	N final de clases
1	6	6	6	6
4	4	4	5	5
6	2	2	2	2
14	5	5	2	2
16	6	6	6	6
18	6	6	4	4
25	5	5	6	6
26	5	5	5	5
28	2	2	2	2
35	6	5	2	2
38	6	6	4	4
43	4	4	5	5
44	2	2	6	6
50	6	6	4	4
51	6	6	5	5
52	2	2	2	2

B.2. Valencia

Tabla B.2: Niveles de valencia por usuario

Sujeto	t = 5		t = 2	
	N inicial de clases	N final de clases	N inicial de clases	N final de clases
1	2	2	3	3
2	3	3	4	4
3	2	1	6	6
4	5	5	6	6
5	2	2	3	3
6	2	1	5	5
7	2	2	4	4
8	2	2	6	6
9	2	2	3	3
10	3	3	6	6
11	2	2	2	2
12	2	2	3	3
13	2	2	3	3
14	2	2	6	6
15	2	2	2	2
16	3	3	6	6
17	2	2	2	2
18	2	2	3	3
19	3	3	3	3
20	2	2	2	2
21	2	2	3	3
22	2	2	2	2
23	5	5	3	3
24	2	2	2	2
25	2	2	6	6
26	6	6	2	2
27	4	4	3	3
28	2	2	2	2
29	2	2	6	6
30	2	2	3	3
31	2	2	4	4
32	2	2	3	3
33	2	2	5	5
34	2	2	6	6
35	2	1	2	2
36	2	1	3	3
37	2	2	2	2
38	2	1	5	5
39	2	2	5	5
40	2	1	6	6
41	2	2	3	3
42	2	2	2	2

Sujeto	t = 5		t = 2	
	N inicial de clases	N final de clases	N inicial de clases	N final de clases
43	2	2	2	2
44	2	2	2	2
45	2	2	6	6
46	4	4	6	6
47	2	1	2	2
48	5	5	5	5
49	4	4	2	2
50	4	4	3	3
51	6	6	2	2
52	2	2	5	5
53	3	3	3	3
54	2	2	2	2
55	3	3	2	2

B.3. Excitación

Tabla B.3: Niveles de excitación por usuario

Sujeto	t = 5		t = 2	
	N inicial de clases	N final de clases	N inicial de clases	N final de clases
1	2	2	2	2
2	2	2	3	3
3	2	2	2	2
4	2	2	2	2
5	2	2	2	2
6	2	2	5	5
7	2	2	2	2
8	2	2	2	2
9	2	2	6	6
10	5	5	2	2
11	2	1	2	2
12	2	2	2	2
13	3	3	6	6
14	2	2	2	2
15	2	2	2	2
16	2	2	2	2
17	2	2	2	2
18	2	2	4	4
19	2	2	2	2
20	2	2	4	4
21	2	2	6	6
22	3	3	6	6

Sujeto	t = 5		t = 2	
	N inicial de clases	N final de clases	N inicial de clases	N final de clases
23	2	2	2	2
24	2	2	2	2
25	2	2	4	4
26	2	2	6	6
27	5	5	2	2
28	2	2	2	2
29	3	3	3	3
30	2	1	2	2
31	2	2	2	2
32	2	2	2	2
33	2	2	6	6
34	2	2	2	2
35	2	2	2	2
36	2	2	2	2
37	2	2	2	2
38	2	2	2	2
39	2	2	2	2
40	2	2	6	6
41	2	2	2	2
42	2	2	2	2
43	5	5	4	4
44	2	2	2	2
45	3	3	3	3
46	3	3	2	2
47	2	2	6	6
48	2	2	2	2
49	2	2	2	2
50	2	2	2	2
51	2	2	2	2
52	2	2	4	4
53	2	2	2	2
54	3	3	6	6
55	2	2	4	4

Apéndice C

Análisis estadístico

A continuación, se presenta el detalle del análisis con ANOVA con medidas repetidas para responder a la pregunta de si las señales pueden distinguir los distintos niveles de carga cognitiva, valencia y excitación en las distintas etapas del ensayo. Se destacan en amarillo aquellas características con un nivel de significancia menor a 0,05.

C.1. Carga cognitiva

C.1.1. Borrador

Tabla C.1: Resultados ANOVA-MR para niveles de carga cognitiva en la etapa borrador del ensayo.

Señal	Característica	F	Significancia (p-value)
EyeTracker	numFijaciones	17,761	2,84E-05
	numSacadas	1,815	0,1784
	promPupila	327,8	2,00E-16
	varPupila	6,676	0,00998
ECG	promECG	0,01	0,92
	mediana ECG	52,97	9,43E-13
	ecgMAD	20,37	7,54E-06
HR	stdHR	1,071	0,3012
	promHR	74,172	2,00E-16
	AVNN	68,547	6,66E-16
	SDNN	5,421	0,0202
	rmsHR	0,164	0,6854
	rMSDD	1,484	0,2237
Temperatura	promTemp	42,24	1,56E-10
	mediana Temp	0,853	0,356

Señal	Característica	F	Significancia (p-value)
GSR	numPeaksFasica	13,54	0,000252
	maxFasica	8,17	0,00439
	promFasica	3,22	0,073207
	gsrAcum	29,44	8,06E-08
PPG	ppgProm	0,821	0,365119
	ppgStd	20,728	6,28E-06
	ppgMediana	8,589	0,003496
	ppgMax	15,168	0,000108
	ppgMin	12,056	0,000549
EEG	etotalF3_theta	0,37	0,543175
	etotalF4_theta	3,511	0,061411
	etotalF7_theta	3,834	0,050631
	etotalF8_theta	6,724	0,009721
	entropiaF3_theta	0,875	0,349809
	entropiaF4_theta	0,571	0,450234
	entropiaF7_theta	2,219	0,136839
	entropiaF8_theta	0,045	0,832039
	stdF3_theta	12,064	0,000548
	stdF4_theta	15,46	9,31E-05
	stdF7_theta	2,531	0,112091
	stdF8_theta	0,639	4,24E-01
	etotalP7_alfa	0,102	0,74972
	etotalP8_alfa	8,025	0,004753
	entropiaP7_alfa	0,031	0,859879
entropiaP8_alfa	0,281	0,595965	
stdP7_alfa	3,729	0,053912	

C.1.2. Introducción

Tabla C.2: Resultados ANOVA-MR para niveles de carga cognitiva en la etapa introducción del ensayo.

Señal	Característica	F	Significancia (p-value)
EyeTracker	numFijaciones	1,503	0,2204
	numSacadas	23,279	1,49E-06
	promPupila	7,037	0,00804
	varPupila	12,078	0,00052
ECG	promECG	0,197	0,657
	mediana ECG	102,002	2,00E-16
	ecgMAD	48,188	5,02E-12

Señal	Característica	F	Significancia (p-value)
HR	stdHR	1,846	0,174433
	promHR	134,269	2,00E-16
	AVNN	14,721	0,000128
	SDNN	4,859	0,027598
	rmsHR	1,342	0,246723
	rMSDD	2,09	0,148365
Temperatura	promTemp	63,524	2,47E-15
	mediana Temp	0,984	0,321
GSR	numPeaksFasica	0,201	0,6543
	maxFasica	3,29	0,0699
	promFasica	0,004	0,9514
	gsrAcum	27,856	1,42E-07
PPG	ppgProm	0,365	0,545856
	ppgStd	11,971	0,000891
	ppgMediana	7,115	0,007697
	ppgMax	13,517	0,000242
	ppgMin	4,585	0,03236
EEG	etotalF3_theta	0,412	0,5212
	etotalF4_theta	0,043	0,836583
	etotalF7_theta	1,787	0,181429
	etotalF8_theta	2,957	0,085664
	entropiaF3_theta	10,172	0,001445
	entropiaF4_theta	0,189	0,663986
	entropiaF7_theta	1,791	0,180969
	entropiaF8_theta	0,002	0,967677
	stdF3_theta	1,581	0,208689
	stdF4_theta	11,48	0,000715
	stdF7_theta	52,997	4,58E-13
	stdF8_theta	2,509	0,113330
	etotalP7_alfa	0,32	0,571421
	etotalP8_alfa	1,347	0,245929
	entropiaP7_alfa	0,253	0,614828
	entropiaP8_alfa	2,492	0,114542
stdP7_alfa	0,099	0,752623	

C.1.3. Tesis

Tabla C.3: Resultados ANOVA-MR para niveles de carga cognitiva en la etapa tesis del ensayo.

Señal	Característica	F	Significancia (p-value)
EyeTracker	numFijaciones	1,081	0,29894
	numSacadas	17,674	3,00E-05
	promPupila	16,559	5,32E-05
	varPupila	10,729	0,00111
ECG	promECG	0,123	0,726
	mediana ECG	75,886	2,00E-16
	ecgMAD	0,047	0,828
HR	stdHR	12,96	0,000344
	promHR	81,097	2,00E-16
	AVNN	12,191	0,000514
	SDNN	0,274	0,600608
	rmsHR	0,115	0,734589
	rMSDD	0,055	0,814854
Temperatura	promTemp	2,187	0,13972
	mediana Temp	7,482	0,00641
GSR	numPeaksFasica	0,032	0,857952
	maxFasica	12,852	0,000363
	promFasica	0,004	0,948011
	gsrAcum	19,14	1,42E-05
PPG	ppgProm	0,217	0,64127
	ppgStd	8,407	0,00387
	ppgMediana	1,186	0,27661
	ppgMax	2,57	0,10939
	ppgMin	0,471	0,49275
EEG	etotalF3_theta	0,409	0,5226
	etotalF4_theta	0,882	0,2481
	etotalF7_theta	0,484	0,4867
	etotalF8_theta	0,998	0,3181
	entropiaF3_theta	0,02	0,8865
	entropiaF4_theta	2,977	0,085
	entropiaF7_theta	0,182	0,6701
	entropiaF8_theta	0,401	0,5268
	stdF3_theta	0,543	0,4615
	stdF4_theta	3,85	0,0502
	stdF7_theta	4,717	0,0303
	stdF8_theta	4,961	0,0263
	etotalP7_alfa	0,845	0,3584
	etotalP8_alfa	0,068	0,7942
	entropiaP7_alfa	0,558	0,4554
	entropiaP8_alfa	0,089	0,7655
stdP7_alfa	4.890	0,0274	

C.1.4. Desarrollo

Tabla C.4: Resultados ANOVA-MR para niveles de carga cognitiva en la etapa desarrollo del ensayo.

Señal	Característica	F	Significancia (p-value)
EyeTracker	numFijaciones	31,13	2,53E-08
	numSacadas	12,1	0,000508
	promPupila	12,76	0,000357
	varPupila	20,49	6,12E-06
ECG	promECG	0	0,995
	mediana ECG	265,196	2,00E-16
	ecgMAD	2,153	0,142
HR	stdHR	12,92	0,000328
	promHR	528,605	2,00E-16
	AVNN	2,295	0,129838
	SDNN	1,954	0,16216
	rmsHR	0,518	0,471549
	rMSDD	5,315	0,021178
Temperatura	promTemp	6,86	0,00884
	mediana Temp	16,73	4,37E-05
GSR	numPeaksFasica	3,872	0,0491
	maxFasica	18,72	1,54E-05
	promFasica	0,251	0,6164
	gsrAcum	2,914	0,0879
	powerGSR	4,839	0,0279
PPG	ppgProm	0,014	0,90694
	ppgStd	8,829	0,00296
	ppgMediana	10,202	0,00141
	ppgMax	33,71	6,75E-09
	ppgMin	16,856	4,09E-05
EEG	etotalF3_theta	0,007	0,934643
	etotalF4_theta	15,398	8,81E-05
	etotalF7_theta	0,103	0,747715
	etotalF8_theta	0,179	0,672089
	entropiaF3_theta	0,432	0,510994
	entropiaF4_theta	0,4	0,527057
	entropiaF7_theta	0,003	0,954407
	entropiaF8_theta	2,622	0,105435
	stdF3_theta	0,022	0,880864
	stdF4_theta	43,556	4,51E-11
	stdF7_theta	18,21	2,01E-11
	stdF8_theta	10,958	0,000938
	etotalP7_alfa	9,67	0,001883
	etotalP8_alfa	3,968	0,04612
	entropiaP7_alfa	16,413	5,16E-05
entropiaP8_alfa	1,169	0,279682	
stdP7_alfa	31,38	2,22E-08	

C.1.5. Conclusión

Tabla C.5: Resultados ANOVA-MR para niveles de carga cognitiva en la etapa conclusión del ensayo.

Señal	Característica	F	Significancia (p-value)
EyeTracker	numFijaciones	1,508	0,21958
	numSacadas	0,332	0,56462
	promPupila	5,558	0,01849
	varPupila	10,789	0,00103
ECG	promECG	0,396	0,529192
	mediana ECG	61,18	8,47E-15
	ecgMAD	14,156	0,000173
HR	stdHR	1,539	0,215
	promHR	136,916	2,00E-16
	AVNN	2,939	0,0866
	SDNN	0,751	0,3863
	rmsHR	3,873	0,0492
	rMSDD	3,45	0,0634
Temperatura	promTemp	3,123	0,0773
	mediana Temp	0,336	0,5625
GSR	numPeaksFasica	0,26	0,60986
	maxFasica	8,139	0,00438
	promFasica	0,958	0,32777
	gsrAcum	5,775	0,01634
PPG	ppgProm	0,38	0,53756
	ppgStd	0,315	0,57487
	ppgMediana	3,828	0,05056
	ppgMax	6,072	0,01382
	ppgMin	7,513	0,00618

Señal	Característica	F	Significancia (p-value)
EEG	etotalF3_theta	3,201	0,073747
	etotalF4_theta	10,158	0,00146
	etotalF7_theta	0,043	0,835568
	etotalF8_theta	0,119	0,730443
	entropiaF3_theta	0,325	0,568649
	entropiaF4_theta	0,119	0,730443
	entropiaF7_theta	0,325	0,568649
	entropiaF8_theta	0,119	0,730443
	stdF3_theta	0,999	0,317578
	stdF4_theta	32,544	1,34E-08
	stdF7_theta	8,882	0,002915
	stdF8_theta	7,728	0,00549
	etotalP7_alfa	0,228	0,633345
	etotalP8_alfa	0,78	0,377305
	entropiaP7_alfa	2,422	0,119844
	entropiaP8_alfa	0,257	0,612392
stdP7_alfa	13,14	0,000257	

C.2. Valencia

C.2.1. Borrador

Tabla C.6: Resultados ANOVA-MR para niveles de valencia en la etapa borrador del ensayo.

Señal	Característica	F	Significancia (p-value)
Eye Tracker	numFijaciones	0,012	0,912
	numSacadas	1,585	0,208
ECG	promECG	0,311	0,577
	mediana ECG	19,027	1,39E-05
	ecgMAD	90,199	2,00E-16
HR	stdHR	7,839	0,00519
	promHR	2,011	0,08295
	AVNN	9,163	0,00252
	SDNN	15,446	8,92E-05
	rmsHR	2,746	9,77E-02
	rMSDD	2,456	0,11729
Temperatura	promTemp	11,286	0,000803
	mediana Temp	0,001	0,9715
GSR	numPeaksFasica	2,418	0,12015
	maxFasica	0,039	0,8431
	promFasica	5,033	0,02503
	gsrAcum	10,026	0,00158
	powerGSR	5,36	0,02075

Señal	Característica	F	Significancia (p-value)
PPG	ppgProm	0,179	0,6721
	ppgStd	27,525	1,80E-07
	ppgMediana	3,701	0,0546
	ppgMax	0	0,948
	ppgMin	2,987	0,0842
EEG	betaAlfa_F3	0,077	0,7813
	betaAlfa_F4	0,809	0,3685
	betaAlfa_F7	1,543	0,2144
	betaAlfa_F8	0,11	0,74
	e_total_F3Beta	0,217	0,6412
	e_total_F4Beta	1,136	0,2867
	e_total_F7Beta	0,921	0,3374
	e_total_F8Beta	4,551	0,0331
	e_total_P7Beta	471	0,4928
	e_total_P8Beta	1,908	0,1675
	cF7F8	1,333	0,2485
	asimetría_alfaBeta_F4F3	40,629	2,53E-10
	asimetría_alfaBeta_F8F7	1,145	0,2847

C.2.2. Introducción

Tabla C.7: Resultados ANOVA-MR para niveles de valencia en la etapa introducción del ensayo.

Señal	Característica	F	Significancia (p-value)
EyeTracker	numFijaciones	2,796	0,0946
	numSacadas	9,539	0,00203
ECG	promECG	0,041	0,84
	mediana ECG	51,695	8,20E-13
	ecgMAD	41,562	1,33E-10
HR	stdHR	5,031	0,02497
	promHR	35,704	2,58E-09
	AVNN	9,284	0,00233
	SDNN	10,658	0,00111
	rmsHR	0,005	0,94624
	rMSDD	0,739	0,38997
Temperatura	promTemp	4,944	0,0263
	mediana Temp	1,222	0,2691
GSR	numPeaksFasica	4,777	0,028927
	maxFasica	0,487	0,485103
	promFasica	0,375	0,54015
	gsrAcum	2,948	0,04703
	powerGSR	13,806	0,000206
	ppgProm	0,045	0,832

Señal	Característica	F	Significancia (p-value)
PPG	ppgStd	79,811	2,00E-16
	ppgMediana	0,886	0,347
	ppgMax	0,864	0,353
	ppgMin	0,145	0,703
EEG	betaAlfa_F3	0,965	0,32607
	betaAlfa_F4	0,053	0,81795
	betaAlfa_F7	0,286	0,59289
	betaAlfa_F8	9,505	0,00207
	e_total_F3Beta	1,586	0,20794
	e_total_F4Beta	1,586	0,20794
	e_total_F7Beta	0,385	0,53478
	e_total_F8Beta	4,519	0,03361
	e_total_P7Beta	7,076	0,00785
	e_total_P8Beta	5,611	0,01791
	cF7F8	1,569	0,21046
	asimetría_alfaBeta_F4F3	26,667	2,58E-07
	asimetría_alfaBeta_F8F7	0,733	0,39204

C.2.3. Tesis

Tabla C.8: Resultados ANOVA-MR para niveles de valencia en la etapa tesis del ensayo.

Señal	Característica	F	Significancia (p-value)
EyeTracker	numFijaciones	5,936	0,0151
	numSacadas	0,477	0,4899
ECG	promECG	2,032	0,154
	mediana ECG	17,071	4,02E-05
	ecgMAD	21,684	3,83E-06
HR	stdHR	0,423	0,51565
	promHR	5,84	0,01591
	AVNN	1,945	0,16357
	SDNN	8,873	0,00299
	rmsHR	0,111	0,73957
	rMSDD	0,258	0,61143
Temperatura	promTemp	2,671	0,103
	mediana Temp	0,946	0,331
GSR	numPeaksFasica	2,273	0,132
	maxFasica	6,335	0,0121
	promFasica	0,248	0,6188
	gsrAcum	4,314	0,0381
	powerGSR	5,456	0,0198

Señal	Característica	F	Significancia (p-value)
PPG	ppgProm	1,419	0,234022
	ppgStd	12,89	0,000353
	ppgMediana	2,613	0,106419
	ppgMax	0,982	0,32205
	ppgMin	0,297	0,586192
EEG	betaAlfa_F3	0,657	0,4179
	betaAlfa_F4	0,028	0,8664
	betaAlfa_F7	0,216	0,6421
	betaAlfa_F8	0,474	0,4915
	e_total_F3Beta	0,066	0,7975
	e_total_F4Beta	0,462	0,4971
	e_total_F7Beta	1,852	0,174
	e_total_F8Beta	0,5	0,4799
	e_total_P7Beta	5,684	0,0174
	e_total_P8Beta	0,083	0,7733
	cF7F8	0	0,9975
	asimetría_alfaBeta_F4F3	6,501	0,011
	asimetría_alfaBeta_F8F7	0,881	0,3482

C.2.4. Desarrollo

Tabla C.9: Resultados ANOVA-MR para niveles de valencia en la etapa desarrollo del ensayo.

Señal	Característica	F	Significancia (p-value)
EyeTracker	numFijaciones	18,84	1,44E-05
	numSacadas	14,16	0,000169
ECG	promECG	1,381	0,24
	mediana ECG	42,769	6,51E-11
	ecgMAD	154,41	2,00E-16
HR	stdHR	18,947	1,34E-05
	promHR	14,21	0,000165
	AVNN	23,94	1,02E-06
	SDNN	5,122	0,023653
	rmsHR	0,666	0,414639
	rMSDD	6,098	0,013554
Temperatura	promTemp	1,234	0,267
	mediana Temp	0,216	0,642
GSR	numPeaksFasica	4,827	0,02805
	maxFasica	10,365	0,00129
	promFasica	2,543	0,1108
	gsrAcum	9,629	0,00192
	powerGSR	4,631	0,03143
	ppgProm	0	0,985569
	ppgStd	131,93	2,00E-16

Señal	Característica	F	Significancia (p-value)
PPG	ppgMediana	19,07	1,28E-05
	ppgMax	0,45	0,502394
	ppgMin	12,15	0,000494
EEG	betaAlfa_F3	0,127	0,72155
	betaAlfa_F4	0,87	0,35102
	betaAlfa_F7	0	0,97257
	betaAlfa_F8	0,17	0,68012
	e_total_F3Beta	0,348	0,55524
	e_total_F4Beta	0,033	0,85558
	e_total_F7Beta	0,193	0,6607
	e_total_F8Beta	0,093	0,76037
	e_total_P7Beta	8	0,00469
	e_total_P8Beta	1,315	0,25152
	cF7F8	32,689	1,13E-08
	asimetría_alfaBeta_F4F3	7,604	0,00584
	asimetría_alfaBeta_F8F7	0,661	0,4161

C.2.5. Conclusión

Tabla C.10: Resultados ANOVA-MR para niveles de valencia en la etapa conclusión del ensayo.

Señal	Característica	F	Significancia (p-value)
EyeTracker	numFijaciones	4,666	0,0309
	numSacadas	0,605	0,437
ECG	promECG	1,044	0,307
	mediana ECG	15,667	7,84E-05
	ecgMAD	68,88	2,00E-16
HR	stdHR	0,008	0,927
	promHR	1,571	0,21
	AVNN	0,025	0,874
	SDNN	1,635	0,201
	rmsHR	0,518	0,472
	rMSDD	0,518	0,472
Temperatura	promTemp	0,663	0,416
	mediana Temp	3,36	0,067
GSR	numPeaksFasica	0,002	0,966
	maxFasica	18,864	1,48E-05
	promFasica	21,851	3,16E-06
	gsrAcum	1,503	0,22
	powerGSR	2,076	0,15

Señal	Característica	F	Significancia (p-value)
PPG	ppgProm	0,491	0,4834
	ppgStd	32,513	1,37E-08
	ppgMediana	0,339	0,5605
	ppgMax	0,266	0,2607
	ppgMin	4,323	0,0377
EEG	betaAlfa_F3	0,324	0,5694
	betaAlfa_F4	0,008	0,9284
	betaAlfa_F7	0,273	0,6015
	betaAlfa_F8	0,072	0,7899
	e_total_F3Beta	1,838	0,1754
	e_total_F4Beta	0,192	0,6612
	e_total_F7Beta	0,477	0,4898
	e_total_F8Beta	0,965	0,3261
	e_total_P7Beta	1,262	0,2614
	e_total_P8Beta	0,913	0,2295
	cF7F8	1,355	0,2446
	asimetría_alfaBeta_F4F3	1,999	0,1576
	asimetría_alfaBeta_F8F7	5,367	0,0206

C.3. Excitación

C.3.1. Borrador

Tabla C.11: Resultados ANOVA-MR para niveles de excitación en la etapa borrador del ensayo.

Señal	Característica	F	Significancia (p-value)
EyeTracker	numFijaciones	5,518	0,019
	numSacadas	55,605	1,58E-13
ECG	promECG	0,152	0,697
	mediana ECG	18,854	1,52E-05
	ecgMAD	20,325	7,10E-06
HR	stdHR	24,742	7,40E-07
	promHR	24,977	6,57E-07
	AVNN	29,207	7,96E-08
	SDNN	4,114	0,0427
	rmsHR	0,668	0,414
	rMSDD	0,128	0,721
Temperatura	promTemp	10,125	0,0015
	mediana Temp	0,459	0,4984

Señal	Característica	F	Significancia (p-value)
GSR	numPeaksFasica	20,256	7,36E-06
	maxFasica	1,668	0,1967
	promFasica	0,909	0,3404
	gsrAcum	24,172	9,89E-07
	powerGSR	3,588	0,0584
PPG	ppgProm	0,568	0,45121
	ppgStd	8,671	3,29E-03
	ppgMediana	2,832	0,09261
	ppgMax	0,343	0,55827
	ppgMin	0,407	0,52369
EEG	eTotalP7_beta	0,519	0,4712
	eTotalP8_beta	5,212	0,0226
	cP7O2	0,405	0,5244
	cP8O1	35,877	2,70E-09
	cP7P8	17,02	3,93E-05
	cO1O2	41,66	1,51E-10
	b/a_AF3	20,629	6,08E-06
	b/a_AF4	0,14	0,7081
	b/a_F3	27,643	1,70E-07
	b/a_F4	5,366	0,0207

C.3.2. Introducción

Tabla C.12: Resultados ANOVA-MR para niveles de excitación en la etapa introducción del ensayo.

Señal	Característica	F	Significancia (p-value)
EyeTracker	numFijaciones	3,787	0,0517
	numSacadas	18,808	1,49E-05
ECG	promECG	0,272	0,602
	mediana ECG	23,278	1,47E-06
	ecgMAD	34,667	4,36E-09
HR	stdHR	2,192	0,13882
	promHR	63,871	1,90E-15
	AVNN	142,567	2,00E-16
	SDNN	32,944	1,05E-08
	rmsHR	8,276	0,00405
	rMSDD	0,06	0,80614
Temperatura	promTemp	11,206	0,000826
	mediana Temp	3,042	0,081232

Señal	Característica	F	Significancia (p-value)
GSR	numPeaksFasica	12,3	0,00046
	maxFasica	11,33	0,000773
	promFasica	4,024	0,044947
	gsrAcum	16,221	5,78E-05
	powerGSR	31,976	1,71E-08
PPG	ppgProm	0,273	0,60168
	ppgStd	0,923	0,33667
	ppgMediana	0,025	0,87471
	ppgMax	0,665	0,41469
	ppgMin	8,562	0,00346
EEG	eTotalP7_beta	2,681	0,10169
	eTotalP8_beta	0,005	0,94376
	cP7O2	0,323	0,56995
	cP8O1	27,411	1,76E-07
	cP7P8	25,639	4,37E-07
	cO1O2	26,136	3,39E-07
	b/a_AF3	5,498	0,0191
	b/a_AF4	0,294	0,58802
	b/a_F3	10,863	0,00101
	b/a_F4	0,876	0,34941

C.3.3. Tesis

Tabla C.13: Resultados ANOVA-MR para niveles de excitación en la etapa tesis del ensayo.

Señal	Característica	F	Significancia (p-value)
EyeTracker	numFijaciones	0,921	0,338
	numSacadas	0,539	0,463
ECG	promECG	1,072	0,3007
	mediana ECG	6,387	0,0117
	ecgMAD	3,645	0,0566
HR	stdHR	0,056	0,8126
	promHR	2,874	0,0904
	AVNN	52,159	1,31E-12
	SDNN	0,819	0,3657
	rmsHR	1,487	0,2231
	rMSDD	0,49	0,4843
Temperatura	promTemp	1,61	0,205
	mediana Temp	0,091	0,763

Señal	Característica	F	Significancia (p-value)
GSR	numPeaksFasica	3,342	0,0679
	maxFasica	0,071	0,7906
	promFasica	0,073	0,7876
	gsrAcum	0,177	0,6745
	powerGSR	1,479	0,2243
PPG	ppgProm	2,102	0,1476
	ppgStd	5,221	0,0226
	ppgMediana	2,945	0,0866
	ppgMax	0,14	0,7085
	ppgMin	2,14	0,1439
EEG	eTotalP7_beta	1,014	0,3144
	eTotalP8_beta	1,736	0,18812
	cP7O2	4,35	0,03735
	cP8O1	2,757	0,09725
	cP7P8	11,563	0,00071
	cO1O2	2,03	0,15469
	b/a_AF3	0,139	0,70923
	b/a_AF4	5,191	0,023
	b/a_F3	1,738	0,18784
	b/a_F4	0,266	0,6063

C.3.4. Desarrollo

Tabla C.14: Resultados ANOVA-MR para niveles de excitación en la etapa desarrollo del ensayo.

Señal	Característica	F	Significancia (p-value)
EyeTracker	numFijaciones	12,093	0,000509
	numSacadas	8,135	0,004356
ECG	promECG	0,399	0,5274
	mediana ECG	1,916	0,1664
	ecgMAD	2,859	0,0909
HR	stdHR	34,696	4,04E-09
	promHR	26,016	3,48E-07
	AVNN	436,133	2,00E-16
	SDNN	6,594	0,0103
	rmsHR	20,482	6,12E-06
	rMSDD	3,367	0,0666
Temperatura	promTemp	33,589	7,11E-09
	mediana Temp	2,038	0,154

Señal	Característica	F	Significancia (p-value)
GSR	numPeaksFasica	1,158	0,2818
	maxFasica	21,275	4,05E-06
	promFasica	5,831	0,0158
	gsrAcum	4,064	0,0438
	powerGSR	30,415	3,61E-08
PPG	ppgProm	6,057	0,013879
	ppgStd	8,306	0,003963
	ppgMediana	0,816	0,36625
	ppgMax	2,436	0,118615
	ppgMin	14,37	0,000151
EEG	eTotalP7_beta	0,591	0,44201
	eTotalP8_beta	19,948	8,08E-06
	cP7O2	8,743	0,00312
	cP8O1	86,253	2,00E-16
	cP7P8	79,254	2,00E-16
	cO1O2	95,509	2,00E-16
	b/a_AF3	3,104	0,07816
	b/a_AF4	0,715	0,39797
	b/a_F3	0,317	0,58368
	b/a_F4	4,509	0,03375

C.3.5. Conclusión

Tabla C.15: Resultados ANOVA-MR para niveles de excitación en la etapa conclusión del ensayo.

Señal	Característica	F	Significancia (p-value)
EyeTracker	numFijaciones	0,653	0,419
	numSacadas	16,532	4,98E-05
ECG	promECG	0,006	0,9373
	mediana ECG	1,954	0,1623
	ecgMAD	6,241	0,0126
HR	stdHR	0,043	0,834838
	promHR	7,287	0,007008
	AVNN	169,18	2,00E-16
	SDNN	13,879	0,000201
	rmsHR	0,081	0,775515
	rMSDD	14,153	0,000174
Temperatura	promTemp	16,487	5,10E-05
	mediana Temp	2,898	0,0889

Señal	Característica	F	Significancia (p-value)
GSR	numPeaksFasica	3,302	0,0693
	maxFasica	21,297	4,20E-06
	promFasica	1,844	0,1746
	gsrAcum	3,821	0,0508
	powerGSR	0,201	0,6543
PPG	ppgProm	0,037	0,847
	ppgStd	4,172	0,0412
	ppgMediana	2,13	0,1446
	ppgMax	5,737	0,0167
	ppgMin	15,475	8,67E-05
EEG	eTotalP7_beta	2,285	0,130804
	eTotalP8_beta	0,257	0,612523
	cP7O2	3,355	0,067172
	cP8O1	3,182	0,074551
	cP7P8	10847	0,001008
	cO1O2	2,768	0,096359
	b/a_AF3	14,608	0,000137
	b/a_AF4	0,177	0,673647
	b/a_F3	0,199	0,664314
	b/a_F4	0,015	0,90146

Apéndice D

Detalle de clasificación

Se detalla para cada participante, el desempeño del mejor clasificador para carga cognitiva (Tabla D.1), valencia (Tabla D.3) y excitación (Tabla D.4). Además, en la Tabla D.2 se detalla el desempeño del mejor clasificador para cada participante al clasificar carga cognitiva para los dos clusters extremos.

D.1. Carga cognitiva

Tabla D.1: Detalle por sujeto para mejor clasificador de carga cognitiva (ANN con alfa = 0,01)

Sujeto	NClases	Exact. [%]	Exact. std	Prec. [%]	Prec. std	Sens. [%]	Sens. std	F1 [%]	F1 std
1	6	32,25	2,53	18,94	2,12	19,17	2,13	19,00	2,10
4	5	41,56	2,59	28,88	0,82	29,15	0,72	28,60	0,54
6	2	93,56	1,24	60,79	7,35	59,95	6,73	59,75	6,51
14	2	90,69	1,75	50,79	5,14	49,74	2,64	49,80	3,15
16	6	39,81	2,89	21,17	2,72	20,77	1,54	20,81	1,87
18	4	37,38	5,85	29,47	5,45	30,12	5,64	29,64	5,43
25	6	32,39	1,99	20,67	0,74	20,94	1,26	20,71	1,01
26	5	49,03	3,72	27,99	3,03	27,58	2,13	27,53	2,56
28	2	74,44	1,62	58,25	2,14	58,14	2,58	58,09	2,31
35	2	97,14	1,04	52,52	6,84	54,36	9,88	53,27	8,03
38	4	47,19	5,45	32,57	6,48	31,96	5,52	32,08	5,90
43	5	48,69	3,27	27,28	1,35	27,63	1,22	27,34	1,12
44	6	43,28	2,94	24,73	3,90	25,52	3,83	25,00	3,83
50	4	40,36	1,22	33,24	1,81	32,97	2,17	32,89	2,00
51	5	47,93	2,15	31,07	1,67	30,96	1,58	30,92	1,59
52	2	85,52	1,90	67,31	4,56	66,14	4,60	66,66	4,58
Promedio	4,13	56,33	2,63	36,60	3,51	36,57	3,39	36,38	3,28
Std	1,63	23,25	1,40	15,76	2,20	15,53	2,50	15,59	2,21

D.2. Carga cognitiva para dos clases

Tabla D.2: Detalle por sujeto para clasificación de cargas cognitivas extremas (SVM, kernel rbf y $C = 10$)

Sujeto	NClases	Exact. [%]	Exact. std	Prec. [%]	Prec. std	Sens. [%]	Sens. std	F1 [%]	F1 std
1	2	61,65	2,03	57,15	2,11	56,54	2,15	56,48	2,30
4	2	72,41	5,74	66,79	7,57	64,72	7,96	65,11	8,15
6	2	94,36	1,24	59,88	12,65	54,94	5,45	56,21	7,61
14	2	93,24	1,09	47,03	0,03	49,53	0,58	48,25	0,29
16	2	69,54	2,21	64,88	2,80	63,79	2,77	64,11	2,88
18	2	71,28	2,20	69,45	2,41	69,13	2,10	69,22	2,19
25	2	62,32	1,46	61,72	1,40	61,58	1,54	61,48	1,57
26	2	74,01	2,16	70,95	2,59	70,18	2,97	70,46	2,85
28	2	79,37	0,42	61,22	2,09	55,80	2,51	56,01	3,55
35	2	98,04	0,36	49,11	0,00	49,91	0,18	49,50	0,09
38	2	66,53	3,62	56,51	5,83	55,57	4,66	55,59	5,22
43	2	74,30	1,29	72,89	1,41	72,13	1,34	72,37	1,32
44	2	76,38	1,79	75,02	2,02	74,59	1,26	74,55	1,45
50	2	57,49	3,67	56,97	3,69	56,97	3,74	56,91	3,76
51	2	70,49	1,72	56,29	2,70	54,96	2,18	54,97	2,67
52	2	87,41	1,58	71,92	5,18	64,70	2,81	67,06	3,36
Promedio	2,00	75,55	2,04	62,36	3,41	60,94	2,76	61,14	3,08
Std	0,00	12,11	1,35	8,40	3,17	7,81	1,95	8,08	2,28

D.3. Valencia

Tabla D.3: Detalle por sujeto para mejor clasificador de valencia (KNN, $k = 3$)

Sujeto	NClases	Exact. [%]	Exact. std	Prec. [%]	Prec. std	Sens. [%]	Sens. std	F1 [%]	F1 std
1	2	76,92	1,62	54,61	5,79	52,47	3,05	51,67	4,23
2	3	77,74	2,96	34,88	10,50	34,66	4,08	33,53	5,80
4	5	54,94	5,04	41,67	6,68	36,95	4,92	38,12	5,36
5	2	88,77	1,03	45,12	0,05	49,09	0,57	47,02	0,29
7	2	79,48	0,84	60,02	1,80	55,57	1,96	55,93	2,63
8	2	80,65	2,52	56,92	6,18	54,01	3,48	54,49	4,28
9	2	61,13	8,93	49,92	11,00	52,05	6,54	50,35	8,35
10	3	73,04	4,82	39,86	10,66	36,54	4,09	35,83	5,39

Sujeto	NClases	Exact. [%]	Exact. std	Prec. [%]	Prec. std	Sens. [%]	Sens. std	F1 [%]	F1 std
11	2	87,74	1,58	63,89	11,45	55,28	4,47	56,37	6,19
12	2	96,24	0,82	86,63	11,19	66,27	5,16	70,96	5,56
13	2	94,68	0,67	50,45	5,19	51,83	4,51	51,13	4,85
14	2	96,61	0,00	48,31	0,00	50,00	0,00	49,14	0,00
15	2	83,55	2,18	65,99	5,92	63,03	5,27	64,09	5,55
16	3	71,76	2,23	41,63	7,87	39,38	4,43	39,23	5,22
17	2	76,75	1,37	54,37	4,96	52,85	3,32	52,56	4,18
18	2	85,75	1,56	47,21	5,24	48,80	2,53	47,69	3,32
19	3	45,26	2,39	43,25	2,42	42,55	2,35	41,96	2,39
20	2	80,73	4,26	61,22	5,38	55,98	1,89	56,82	2,49
21	2	83,23	2,67	68,82	8,00	59,20	4,62	60,92	5,83
22	2	65,20	5,84	56,37	8,56	55,90	7,63	55,81	8,16
23	5	39,30	3,25	21,12	2,36	22,13	1,74	20,99	1,92
24	2	87,42	3,09	44,72	0,17	48,74	1,72	46,63	0,90
25	2	92,05	0,51	64,92	9,93	55,97	3,00	57,63	4,86
26	6	40,55	4,93	34,41	4,80	27,37	3,60	28,71	3,96
27	4	54,41	1,46	35,54	10,52	32,33	4,14	32,75	5,85
28	2	82,33	2,55	58,52	6,46	54,35	3,41	54,89	4,07
29	2	76,59	4,40	71,08	4,92	70,84	3,64	70,72	4,38
30	2	59,66	3,75	58,86	4,50	58,92	4,61	58,67	4,65
31	2	98,92	0,96	84,57	20,35	79,89	18,74	81,39	18,78
32	2	96,07	1,75	88,18	20,46	66,48	10,84	71,98	13,34
33	2	88,94	1,03	47,93	5,14	49,67	1,72	48,32	2,57
34	2	91,62	0,40	49,56	6,76	50,17	1,32	48,92	2,26
37	2	61,40	2,32	55,94	3,08	55,06	2,65	54,90	2,89
39	2	76,41	5,91	57,21	13,60	51,73	6,42	51,37	7,80
41	2	82,80	3,27	53,13	9,61	51,88	6,03	51,70	7,30
42	2	84,44	0,95	77,18	1,32	73,12	5,53	74,10	4,24
43	2	98,37	0,00	49,19	0,00	50,00	0,00	49,59	0,00
44	2	90,00	1,44	64,77	11,12	57,71	5,15	59,00	7,15
45	2	98,25	0,00	49,12	0,00	50,00	0,00	49,56	0,00
46	4	49,05	5,13	32,85	10,16	30,89	4,28	29,29	4,13
48	5	37,90	3,68	19,98	1,97	20,50	2,77	19,86	2,35
49	4	46,23	2,60	41,02	4,96	39,00	2,62	39,50	3,28
50	4	48,39	4,35	35,60	3,92	35,98	3,46	35,08	3,31
51	6	33,62	3,49	25,70	4,85	24,11	4,50	24,25	4,61
52	2	93,98	0,98	80,83	14,01	61,89	3,85	65,86	4,64
53	3	85,14	1,55	28,88	0,08	32,75	0,60	30,69	0,26
54	2	84,57	1,52	55,78	2,96	53,55	1,89	53,92	2,27
55	3	59,75	4,19	45,74	7,41	42,79	3,92	43,08	5,10
Promedio	2,65	74,97	2,56	52,16	6,55	49,17	3,90	49,31	4,52
Std	1,14	18,71	1,86	16,11	4,80	13,13	3,00	13,92	3,27

D.4. Excitación

Tabla D.4: Detalle para mejor clasificador excitación (ANN, alfa = 0.01)

Sujeto	NClases	Exact. [%]	Exact. std	Prec. [%]	Prec. std	Sens. [%]	Sens. std	F1 [%]	F1 std
1	2	94,53	1,67	87,20	4,22	84,13	6,21	85,37	4,99
2	2	67,55	2,64	56,41	3,47	56,04	3,20	56,12	3,28
3	2	73,74	5,57	60,48	5,39	58,53	2,33	58,76	3,41
4	2	94,18	1,89	84,30	5,56	76,17	9,45	78,83	8,02
5	2	82,74	2,23	48,38	5,29	48,65	3,76	48,32	4,29
6	2	86,67	2,65	49,27	6,21	49,74	6,97	49,48	6,54
7	2	75,17	4,48	59,94	5,81	59,99	4,95	59,63	5,26
8	2	90,00	1,60	80,53	2,92	82,38	3,04	81,28	2,69
9	2	81,13	2,76	70,60	5,88	67,77	3,49	68,63	3,69
10	5	38,26	5,47	27,08	6,41	27,31	5,77	26,65	5,76
12	2	68,12	5,07	51,39	6,06	51,52	5,38	51,10	5,82
13	3	70,85	1,97	66,51	4,14	59,31	6,99	61,10	4,28
14	2	74,92	5,07	60,48	6,19	60,06	5,93	59,67	5,63
15	2	89,35	1,63	52,07	5,18	52,88	3,91	52,42	4,51
16	2	94,82	0,94	56,76	10,82	55,57	8,32	56,00	9,27
17	2	96,24	0,87	48,70	0,01	49,39	0,45	49,04	0,23
18	2	82,30	3,68	70,72	6,65	67,85	4,88	68,88	5,51
19	2	96,84	1,15	49,11	0,01	49,29	0,58	49,20	0,30
20	2	82,39	1,47	67,55	3,62	67,17	7,89	66,59	5,59
21	2	80,00	2,47	69,19	4,12	66,15	4,65	67,15	4,55
22	3	61,60	1,85	50,69	3,74	48,91	1,59	49,22	1,50
23	2	81,69	1,99	51,00	4,16	51,04	4,68	50,99	4,41
24	2	83,71	3,22	59,83	6,31	59,10	5,26	59,21	5,65
25	2	77,69	2,76	56,73	8,76	57,69	7,57	56,89	8,18
26	2	87,64	0,93	77,76	1,65	75,63	4,00	76,39	2,80
27	5	39,82	3,84	26,50	3,56	27,10	3,26	26,63	3,32
28	2	84,17	1,18	62,33	3,69	63,76	5,45	62,88	4,45
29	3	77,67	3,58	60,47	3,98	59,02	6,77	59,36	5,27
31	2	93,98	1,46	48,34	0,03	48,56	0,75	48,44	0,39
32	2	78,57	2,53	59,07	3,67	60,42	4,42	59,44	3,80
33	2	87,58	1,03	75,25	2,29	73,48	3,73	73,92	2,52
34	2	61,01	2,00	51,08	2,08	51,08	1,98	50,99	2,01
35	2	76,80	4,12	54,39	8,12	54,82	5,83	54,43	6,95
36	2	89,44	2,39	81,20	6,38	73,15	4,87	76,13	5,43
37	2	85,87	2,05	69,44	3,40	66,45	3,31	67,34	2,81
38	2	79,61	5,05	61,03	7,84	57,38	5,37	58,26	5,75
39	2	86,50	1,66	73,32	5,55	67,69	2,49	69,64	2,85
40	2	70,97	3,92	59,26	4,22	58,53	3,51	58,74	3,85
41	2	79,25	2,54	65,46	4,27	64,14	3,66	64,59	3,77
42	2	94,13	1,63	83,11	7,46	79,52	7,59	80,51	6,11
43	5	40,00	4,31	28,81	5,12	29,44	4,82	28,98	4,97
44	2	90,85	2,82	73,96	15,11	62,12	5,14	64,84	6,88

Sujeto	NClases	Exact. [%]	Exact. std	Prec. [%]	Prec. std	Sens. [%]	Sens. std	F1 [%]	F1 std
45	3	83,33	2,29	71,07	4,25	66,93	7,30	68,02	5,05
46	3	55,24	6,63	37,52	7,11	37,38	7,84	37,13	7,52
47	2	95,10	0,62	56,54	10,77	54,29	6,09	54,93	7,73
48	2	84,95	3,38	65,41	7,94	63,01	5,12	63,92	6,17
49	2	96,23	0,84	79,53	4,87	80,91	9,52	79,28	5,12
50	2	95,91	1,25	93,86	4,08	85,88	3,10	89,27	3,25
51	2	94,47	1,04	52,43	5,16	53,47	5,91	52,74	5,30
52	2	76,42	3,17	67,07	4,49	66,67	3,69	66,74	3,95
53	2	84,76	2,33	74,19	5,83	69,66	1,28	71,21	2,48
54	3	83,62	1,94	57,09	3,36	53,62	3,79	55,00	3,41
55	2	88,35	3,79	78,31	11,57	72,66	8,63	73,88	7,96
Promedio	2,28	80,50	2,63	61,86	5,26	60,06	4,84	60,46	4,63
Std	0,74	14,00	1,42	14,62	2,79	13,10	2,21	13,61	2,02