



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Computación

**Psiquiatría computacional: diferencias individuales e identificación de
déficits en funciones ejecutivas a partir de patrones de trayectorias
visuales y manuales**

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires en
el área de Ciencias de la Computación

Gustavo Ezequiel Juantorena

Director de tesis: Dr. Juan Esteban Kamienkowski

Co-director de tesis: Dr. Agustín Petroni

Consejero de Estudios: Dr. Pablo Turjanski

Lugar de trabajo: Laboratorio de Inteligencia Artificial Aplicada, Instituto de Ciencias
de la Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos
Aires - CONICET

Buenos Aires, 2026

Psiquiatría computacional: diferencias individuales e identificación de déficits en funciones ejecutivas a partir de patrones de trayectorias visuales y manuales

La medición objetiva de las funciones ejecutivas es un desafío central de la neuropsicología y la salud mental. Las pruebas tradicionales, ancladas en el formato de lápiz y papel, resumen procesos cognitivos complejos en unas pocas medidas globales y dependen íntegramente de profesionales especializados, lo que limita tanto su precisión como su alcance, especialmente en regiones con recursos escasos.

En esta tesis desarrollamos y validamos un conjunto de herramientas computacionales que, a partir del registro de alta resolución de trayectorias manuales y oculares, buscan medir las funciones ejecutivas con mayor detalle y llevar la evaluación desde el laboratorio hacia el consultorio. La prueba neuropsicológica central en esta tesis fue una versión computarizada del Trail Making Test (cTMT) que registra de manera simultánea los movimientos de la mano y los ojos, lo que permitió descomponer una tarea aparentemente unitaria en sus procesos subyacentes, replicar sus efectos clásicos y derivar marcadores internos de memoria de trabajo y control inhibitorio.

Al llevar esta herramienta a pacientes con deterioro cognitivo leve, las métricas oculares discriminaron entre grupos allí donde las medidas clásicas no lo lograron. Sin embargo, la dificultad de adquirir datos oculares de calidad en población clínica motivó un giro hacia los movimientos de la mano, una señal con menores requerimientos de hardware. Modelos de aprendizaje automático entrenados únicamente con trayectorias manuales alcanzaron un valor diagnóstico cercano al de una batería neuropsicológica completa, sin necesidad de equipamiento especializado y bajo un esquema de validación riguroso y reproducible.

Para escalar este enfoque construimos una plataforma web de acceso abierto, diseñada para la comunidad hispanohablante, que sirvió de infraestructura para la recolección de datos multimodales a gran escala. Sobre ella, una versión online del cTMT mostró que las trayectorias del cursor predicen diferencias individuales en memoria de trabajo, control inhibitorio y edad. Por último, recuperamos el registro ocular de forma remota mediante un sistema de seguimiento ocular por cámara web que mejoramos y validamos con la tarea de antisacadas, sin requerir hardware especializado.

En conjunto, creamos y validamos herramientas digitales que aportan marcadores cognitivos objetivos, reproducibles y ecológicamente válidos, con potencial de implementación remota y a gran escala. Estas contribuciones se enmarcan en los campos emergentes de la neuropsicología digital y la psiquiatría computacional, con especial atención a la accesibilidad y a poblaciones históricamente subrepresentadas en la investigación cognitiva.

Palabras claves: Psiquiatría computacional, Funciones ejecutivas, Aprendizaje automático, Seguimiento ocular, Experimentos en línea, Evaluación psicológica remota.

Computational psychiatry: individual differences and identification of executive-function deficits from patterns of visual and manual trajectories

The objective measurement of executive functions is a central challenge in neuropsychology and mental health. Traditional tests, anchored in a pencil-and-paper format, reduce complex cognitive processes to a few global measures and rely entirely on specialized professionals, which constrains both their precision and their reach, particularly in regions with scarce resources.

In this thesis we developed and validated a set of computational tools that, based on high-resolution recordings of hand and eye trajectories, aim to measure executive functions in greater detail and to bring assessment from the laboratory to the clinic. The central neuropsychological instrument was a computerized version of the Trail Making Test (cTMT) that simultaneously records hand and eye movements, allowing us to decompose an apparently unitary task into its underlying processes, replicate its classic effects, and derive internal markers of working memory and inhibitory control.

When applied to patients with mild cognitive impairment, ocular metrics discriminated between groups where classical measures did not. However, the difficulty of acquiring high-quality eye-tracking data in clinical populations motivated a shift toward hand movements, a signal with lower hardware requirements. Machine-learning models trained solely on manual trajectories reached a diagnostic value close to that of a full neuropsychological battery, without specialized equipment and under a rigorous, reproducible validation scheme.

To scale this approach, we built an open-access web platform, designed for the Spanish-speaking community, that served as infrastructure for large-scale multimodal data collection. Building on it, an online version of the cTMT showed that cursor trajectories predict individual differences in working memory, inhibitory control and age. Finally, we brought eye recording back remotely through a webcam-based eye-tracking system that we improved and validated with the antisaccade task, requiring no specialized hardware.

Taken together, we created and validated digital tools that provide objective, reproducible and ecologically valid cognitive markers, with potential for remote, large-scale deployment. These contributions sit within the emerging fields of digital neuropsychology and computational psychiatry, with particular attention to accessibility and to populations historically underrepresented in cognitive research.

Keywords: computational psychiatry, executive functions, machine learning, eye tracking, online experiments, remote psychological assessment.

Agradecimientos

En primer lugar, quiero agradecer a mis directores: Agus y Juan.

Agus estuvo desde el primer momento y, a pesar de la distancia, siempre se hizo el tiempo para que tuviéramos reuniones semanales y para ayudarme en todo lo que necesité.

Juan es el principal responsable de que yo haya llegado a esta etapa. Estuvo absolutamente en todos los momentos críticos y, además de haber aprendido muchísimo con él, me alegra haber generado un vínculo en el que logramos entendernos muy bien.

A los *Tesinguistas* (Lara, Gastón y Damián), por la amistad y la compañía de siempre, pero muy especialmente en esta última etapa de escritura, donde su apoyo fue clave.

A los tesistas de licenciatura que codirigí (Fran, Agus y Gian), con quienes viví una gran experiencia enseñando y especialmente aprendiendo. Con ellos generamos parte de esta tesis.

A todo el LIAA, que objetivamente es el mejor laboratorio del mundo gracias a la gente que lo compone.

A Jessi, por apoyarme durante la mayor parte de este proceso.

A Pablo, por sus valiosos consejos durante nuestras caminatas volviendo del labo.

A mis viejos, que son los responsables de que yo me haya podido educar y siempre hicieron un esfuerzo enorme para que yo pudiera cumplir mis metas.

A CONICET, por haberme otorgado una beca sin la cual no habría podido hacer este trabajo.

A la UBA, a Exactas y en particular al ICC, donde me abrieron las puertas y me dieron todos los elementos para poder formarme e investigar.

A la educación pública nacional, no solo como un ente abstracto, sino por la gente que la compone. Al culminar esta etapa, habré completado todos los niveles educativos en instituciones públicas, algo que me ha abierto puertas que de otra manera jamás habría imaginado. Especialmente en el contexto actual, es más importante que nunca valorar su rol.

Índice

Introducción general.....	8
Capítulo 1: Desarrollo y validación de una versión computarizada del Trail Making Test.....	11
Introducción.....	11
Métodos.....	14
Participantes.....	14
TMT computarizado (cTMT).....	14
Procedimiento.....	15
Generación de estímulos.....	17
Registros conductuales.....	17
Medidas de validación externa.....	18
Análisis de datos.....	19
Definición de métricas de rendimiento.....	19
Criterios de corrección del ensayo.....	20
Análisis de los movimientos oculares.....	20
Descomposición de las fijaciones en etapas.....	21
Medida interna de la memoria de trabajo.....	22
Medida interna del control inhibitorio.....	23
Resultados.....	24
Movimientos de la mano: rendimiento global del cTMT.....	24
Movimientos oculares.....	25
Descomposición de la tarea basada en movimientos oculares y manuales.....	26
Memoria de trabajo visual.....	28
Control inhibitorio.....	29
Discusión.....	30
Conclusiones.....	33
Capítulo 2: Aplicación del cTMT en deterioro cognitivo leve: integración de seguimiento ocular y manual.....	34
Introducción.....	34
Consideraciones metodológicas: EyeLink en modo remoto.....	35
Métodos.....	35
Participantes.....	35
Protocolo experimental.....	35
Análisis.....	36
Resultados.....	36
Discusión.....	37
Conclusiones.....	38
Capítulo 3: Evaluación digital del deterioro cognitivo leve mediante aprendizaje automático y movimientos de la mano en el Trail Making Test.....	39
Introducción.....	39
Métodos.....	40
Participantes.....	40

TMT Computarizado (cTMT).....	41
Extracción de características de las mediciones de la mano.....	42
Esquema de aprendizaje automático.....	43
Importancia de las Características: Enfoque SHAP.....	45
Modelos de Referencia.....	46
Resultados.....	46
Análisis Exploratorio.....	46
Clasificación del grupo diagnóstico.....	47
Predicción de Puntuaciones de Pruebas Neuropsicológicas.....	50
Discusión.....	55
Conclusiones.....	58
Capítulo 4: DataPruebas: una plataforma para recolección de datos a escala.....	59
Introducción.....	59
Interfaz y Compatibilidad Experimental.....	60
Calidad de Datos Online.....	62
Captura de Trayectorias Motoras y Datos Multimodales.....	62
Limitaciones Actuales y Trabajo Futuro.....	63
Discusión.....	63
Conclusiones.....	64
Capítulo 5: Predicción de diferencias individuales en funciones ejecutivas a partir del TMT online.....	65
Introducción.....	65
Métodos.....	66
Participantes.....	66
Criterios de inclusión.....	66
Plataforma web y aspectos de implementación.....	67
Tareas Experimentales.....	68
Trail-Making Test.....	68
Stop-Signal Task.....	69
Change-Detection Task.....	69
Go/No-Go.....	69
Procesamiento de datos y extracción de características.....	70
Controles de calidad.....	70
Preprocesamiento de las trayectorias.....	70
Extracción de características y segmentación.....	71
Variables objetivo.....	72
Esquema de aprendizaje automático.....	73
Análisis estadístico.....	73
Importancia de características: enfoque SHAP.....	74
Resultados.....	74
Datos conductuales.....	74
Resultados de predicción.....	75
Análisis de importancia de características.....	77
Discusión.....	79

Implicaciones teóricas: el cTMT como herramienta de evaluación cognitiva continua.....	79
Limitaciones y direcciones futuras.....	80
Conclusiones.....	81
Capítulo 6: Seguimiento ocular basado en web.....	82
Introducción y motivación.....	82
Desarrollo e implementación del prototipo.....	85
Arquitectura base y selección de herramientas.....	85
Optimización y adaptación al contexto experimental.....	86
Detección de movimiento de cabeza.....	87
Detección de parpadeos.....	88
Estimación de distancia a pantalla.....	89
Validación técnica de los módulos de monitoreo.....	90
Experimento de detección de parpadeos.....	91
Experimento de estimación de distancia a la pantalla.....	93
Experimento de detección de rotaciones de cabeza.....	93
Resultados.....	93
Desempeño del sistema y de los módulos de monitoreo.....	93
Optimización del rendimiento (frame rate).....	93
Detección de Parpadeo: Comparación de Algoritmos.....	94
Detección de Rotación de Cabeza.....	96
Estimación de Distancia a la Pantalla.....	97
Estudio de caso: validación de la tarea de antisacadas.....	98
Participantes.....	98
Materiales.....	98
Protocolo Experimental.....	99
Preprocesamiento de datos.....	100
Análisis estadístico.....	101
Resultados de la tarea de antisacadas.....	101
Discusión y Conclusiones.....	102
Conclusiones generales.....	105
Bibliografía.....	107
Apéndice A: Repositorios de Código y Recursos Digitales.....	126
A.1. Plataforma DataPruebas (Caps. 4, 5 y 6).....	126
A.2. Implementación del Virtual Chinrest (Li et al., 2020) (Caps. 5 y 6).....	126
A.3. Repositorio de tareas cognitivas en jsPsych (Cap. 5).....	126
A.4. Biblioteca NeuroTask (Caps. 3 y 5).....	126
A.5. Biblioteca Pyxations (Cap. 6).....	127
A.6. Web Eye-Tracker: seguimiento ocular por cámara web (Cap. 6).....	127
Apéndice B: Repositorios de datos.....	128
B.1. Datos del cTMT en pacientes con DCL (Cap. 3).....	128
B.2. Datos del cTMT online (Cap. 5).....	128
B.3. Datos de la tarea de antisacadas (Cap. 6).....	128
Material suplementario.....	129
Figura C1.S1: Medidas de rendimiento para dos umbrales de objetivos diferentes.....	129

Figura C1.S2: Medidas de rendimiento para el primer tercio de los ensayos.....	130
Figura C1.S3: Distribución espacial de las trayectorias exploradas por la mano al fijar la vista en un nuevo ítem para un participante.....	131
Figura C5.S1: Valores medios SHAP - Ridge - Accuracy.....	132
Tabla C3.S1: Atributos extraídos del cursor (estudio presencial).....	133
Tabla C3.S2: Grilla de hiperparámetros de clasificación.....	134
Tabla C3.S3: Grilla de hiperparámetros de regresión (capítulos 3 y 5).....	135
Tabla C5.S1: Atributos extraídos del cursor (estudio online).....	136

Introducción general

La medición objetiva del comportamiento y la cognición es uno de los desafíos centrales de la salud mental. A diferencia de otras áreas de la medicina, la psiquiatría y la neuropsicología dependen en buena medida de la observación clínica y los autoinformes, susceptibles a sesgos y limitados a evaluaciones puntuales. En los últimos años, los avances en capacidad computacional y el acceso masivo a dispositivos digitales dieron lugar a campos como la psiquiatría computacional (Montague et al., 2012) y la neuropsicología digital (Germine et al., 2019), que buscan medir la cognición y la conducta de forma cuantitativa y reproducible.

En este marco, el fenotipado digital (*digital phenotyping*) propone caracterizar a las personas a partir de su “huella digital comportamental”, es decir, las señales que surgen de su interacción con dispositivos (Insel, 2017, 2018). De estas señales se derivan biomarcadores digitales, definidos como datos objetivos y cuantificables recolectados mediante dispositivos digitales (Piau et al., 2019). A diferencia de los biomarcadores tradicionales, como el análisis de líquido cefalorraquídeo o las neuroimágenes, permiten un seguimiento continuo, no invasivo y ecológicamente válido, con capacidad para detectar cambios sutiles en la cognición que pueden preceder a los síntomas clínicos por años (Kourtis et al., 2019; Song et al., 2025). Si bien buena parte de este campo se apoya en el monitoreo pasivo de sensores (actividad, geolocalización, uso del teléfono), una vertiente complementaria se basa en tareas cognitivas estructuradas, donde la forma en que una persona interactúa (cómo se mueve, dónde mira, cuánto demora) resulta tan informativa como su respuesta final (Insel, 2018). Para tal fin, el aprendizaje automático (*machine learning*) ocupa un lugar central: entrenado sobre datos conductuales, permite extraer patrones que escapan al análisis clínico tradicional y construir predictores objetivos de capacidades cognitivas específicas.

Algunas de las medidas de alta resolución temporal que permiten estudiar los procesos cognitivos que subyacen a una tarea son los movimientos oculares y manuales, permitiendo ir más allá de lo que permite una medida global como el tiempo de respuesta. Las trayectorias revelan cuándo y dónde se orienta la atención, cómo se planifica la acción siguiente, y en qué momento se inhibe una respuesta incorrecta. Por un lado, el seguimiento ocular (*eye tracking*, en idioma inglés) permite registrar métricas como fijaciones, sacadas y tamaño pupilar, las cuales reflejan correlatos directos de la atención visual, la carga cognitiva y los estados afectivos (Duchowski, 2017; Yarbus, 1967). Durante una fijación, la mirada permanece relativamente estable sobre una región del estímulo y se extrae la información visual, mientras que las sacadas son los movimientos rápidos y balísticos que reorientan la mirada de una fijación a la siguiente. Estas señales ofrecen un acceso privilegiado a los procesos que ocurren durante la planificación y ejecución de una tarea, antes de que se produzca cualquier respuesta observable. Por otro lado, los movimientos del mouse emergieron como herramienta para el estudio de procesos psicológicos en tiempo real a partir del trabajo pionero de Spivey y colaboradores (2005), quienes demostraron que los movimientos del cursor reflejan la competencia continua entre representaciones cognitivas durante la toma de decisiones. Estudios posteriores confirmaron que estas trayectorias (incluyendo desviaciones, submovimientos y tiempos de inicio) correlacionan con medidas estándar de carga cognitiva y carga de trabajo mental (Cha & Min, 2022; Thorpe et al., 2022), posicionándose como biomarcadores digitales accesibles sin hardware especializado (Wulff et al., 2026).

Un terreno natural para este enfoque son las pruebas neuropsicológicas clásicas: instrumentos ampliamente validados pero anclados históricamente en el formato de lápiz y papel. Digitalizarlas

permite registrar el proceso, y no solo el resultado final. Entre las más extendidas se encuentra el Test del Trazo (Trail Making Test, o TMT), herramienta central de esta tesis.

En 1945, el gobierno de Estados Unidos evaluó a 256 veteranos de la Segunda Guerra Mundial utilizando un test desarrollado por Partington y Leiter en 1938 (Partington & Leiter, 1949), el mismo mostró una fuerte correlación con medidas de inteligencia de la época (Brown et al., 1958). Durante su incorporación como subtest de la *Army Individual Test Battery*, el instrumento fue rebautizado como *Trail Making Test*, y comenzó a emplearse para la evaluación rápida del funcionamiento intelectual y la identificación de alteraciones asociadas a daño cerebral. Posteriormente, el Departamento de Guerra autorizó desclasificar la prueba solo para uso interno del ejército, permitiendo su reproducción y adopción en entornos clínicos y civiles.

El TMT es hoy uno de los instrumentos neuropsicológicos más utilizados en evaluaciones clínicas y de investigación a nivel mundial (Reitan, 1958; Tombaugh, 2004). La prueba se administra en lápiz y papel y consta de dos partes: en la Parte A el participante conecta 25 números en orden ascendente; en la Parte B debe alternar entre números (1 a 13) y letras (A a L) en orden numérico y alfabético. El TMT resulta sensible a alteraciones en funciones ejecutivas como la memoria de trabajo, el control inhibitorio y la atención, y muestra resultados consistentes en múltiples poblaciones clínicas. Sin embargo, tanto su estructura como las métricas que extrae no han cambiado desde su creación hace más de ochenta años. Su versión tradicional ofrece pocas medidas (esencialmente el tiempo total y la cantidad de errores), no distingue los procesos ejecutivos que intervienen en la tarea, y presenta alta variabilidad al medir un único ensayo por condición. Estas limitaciones revelan la necesidad de versiones computarizadas que midan con mayor precisión y que permitan descomponer la tarea en sus componentes subyacentes.

Uno de los principales desafíos de salud pública del siglo XXI lo constituyen las demencias, con algunas fuentes proyectando que la cantidad de personas que la padecen casi se triplicará para el año 2050 (Nichols et al., 2022). La enfermedad de Alzheimer representa aproximadamente el 70% de los casos (Rostagno, 2023), y el deterioro cognitivo leve (DCL) se reconoce como una etapa prodrómica de alto riesgo de progresión. Detectar déficits en funciones ejecutivas durante esta ventana temprana es clave para el desarrollo de intervenciones oportunas y biomarcadores diagnósticos.

En este sentido, las funciones ejecutivas (FE) son sensibles al envejecimiento normal y se deterioran de forma pronunciada en el DCL y la enfermedad de Alzheimer, lo que las convierte en objetivos clave para la evaluación clínica temprana. Se trata de un grupo de funciones cognitivas de alto orden que permiten la planificación, el monitoreo y la regulación del comportamiento orientado a metas, con componentes centrales como la memoria de trabajo, el control inhibitorio y la flexibilidad cognitiva (Diamond, 2013). Su medición precisa, sin embargo, presenta desafíos inherentes: dado que ninguna tarea ejecutiva se desarrolla de forma aislada, separar la contribución de cada componente resulta complejo, un problema conocido como el *impurity problem* (Miyake et al., 2000). Medir estas funciones de forma precisa, accesible y a gran escala sigue siendo un desafío abierto.

La evaluación neuropsicológica tradicional requiere infraestructura clínica especializada, lo que limita su alcance en regiones con recursos sanitarios escasos, como América Latina, donde la prevalencia de demencia crece de forma pronunciada (Nichols et al., 2022). Los avances en conectividad y la proliferación de dispositivos personales abren la posibilidad de llevar estas evaluaciones fuera del laboratorio, reduciendo barreras geográficas y económicas. Este contexto

motiva el desarrollo de herramientas de evaluación cognitiva remota que sean válidas, accesibles y escalables a grandes poblaciones.

El objetivo general de esta tesis es construir herramientas computacionales que, a partir de las trayectorias visuales y manuales, permitan medir las funciones ejecutivas con mayor resolución que las pruebas tradicionales y, mediante aprendizaje automático, predecir diferencias individuales e identificar déficits ejecutivos de forma accesible y escalable.

Para alcanzarlo, se plantean los siguientes objetivos específicos:

1. Desarrollar y validar una versión computarizada del TMT (cTMT) con registro simultáneo de alta resolución de los movimientos de la mano y los ojos, que controle las configuraciones espaciales de los ítems y permita descomponer la tarea para extraer marcadores de funciones ejecutivas como la memoria de trabajo y el control inhibitorio.
2. Evaluar la capacidad de las métricas oculares y manuales del cTMT para discriminar a pacientes con deterioro cognitivo leve (DCL) de controles sanos.
3. Determinar el valor diagnóstico de los marcadores digitales derivados de las trayectorias manuales del cTMT para la detección de DCL mediante aprendizaje automático, en comparación con tests neuropsicológicos estándar.
4. Extender el cTMT a un contexto online y a gran escala, y evaluar mediante modelos de aprendizaje automático si las trayectorias del cursor permiten predecir diferencias individuales en funciones ejecutivas.
5. Adaptar y validar un sistema de seguimiento ocular remoto por cámara web, a partir de software de código abierto existente, incorporando mejoras técnicas que habiliten su uso en la evaluación neuropsicológica fuera del laboratorio.

Llevar varios de estos objetivos a gran escala requirió, además, construir una infraestructura propia de recolección de datos multimodales (*DataPruebas*, **Capítulo 4**).

La tesis se organiza en seis capítulos que abordan estos objetivos de forma progresiva. En el **Capítulo 1** desarrollamos y validamos una versión computarizada del TMT (cTMT) que registra simultáneamente, con alta resolución, los movimientos de la mano y los ojos, descomponiendo la tarea en fases de monitoreo, planificación y exploración. En el **Capítulo 2** aplicamos el cTMT con registro ocular y manual a pacientes con deterioro cognitivo leve (DCL), identificando métricas oculares que discriminaron entre grupos más allá de las medidas clásicas, y documentamos los desafíos de la adquisición de datos de *eye tracking* en población clínica. En el **Capítulo 3** aplicamos técnicas de aprendizaje automático sobre los datos manuales del cTMT para detectar DCL en adultos mayores, comparando el valor diagnóstico de los marcadores digitales con el de tests neuropsicológicos estándar. En el **Capítulo 4** presentamos *DataPruebas*, una plataforma web de acceso abierto que creamos para la recolección de datos multimodales a gran escala en la comunidad hispanohablante, que sirvió como infraestructura para los estudios siguientes. En el **Capítulo 5** extendemos el cTMT a un contexto online con 484 participantes, investigando si las trayectorias del cursor pueden predecir diferencias individuales en funciones ejecutivas. Finalmente, en el **Capítulo 6** desarrollamos y validamos un sistema de seguimiento ocular basado en cámara web, mejorando la biblioteca *WebGazer.js* para permitir evaluaciones neuropsicológicas remotas, validado con la tarea de antisacadas.

Capítulo 1: Desarrollo y validación de una versión computarizada del Trail Making Test

Este capítulo se basa en el trabajo publicado en la revista Scientific Reports de la editorial Nature:

Linari, I. *, Juantorena, G. E. *, Ibáñez, A., Petroni, A., & Kamienkowski, J. E. (2022). Unveiling Trail Making Test: Visual and manual trajectories indexing multiple executive processes. *Scientific Reports*, 12(1), 14265. <https://doi.org/10.1038/s41598-022-16431-9>.

* Estos autores contribuyeron equitativamente

Introducción

El TMT (Fig. 1), una de las pruebas neuropsicológicas más utilizadas para evaluar funciones ejecutivas, presenta importantes limitaciones, señaladas tanto por nuestro grupo como por otros (Bowie & Harvey, 2006; R. T. Lange et al., 2005; Linari et al., 2022). Primero, no distingue los procesos ejecutivos que intervienen en la tarea. A pesar de que su resolución requiere capacidades tales como preparación y ejecución motora, coordinación visomotora, memoria de trabajo e inhibición, las métricas relevadas no separan estos componentes y el rendimiento final funciona como un índice global. Esto se conoce como el *impurity problem* (Miyake et al., 2000), el cual plantea que dado que ninguna tarea ejecutiva puede desarrollarse sin involucrar, además, el uso de funciones no ejecutivas, su discriminación es más compleja. Un ejemplo es que el TMT exige una velocidad de ejecución motora mínima para completar los trazos. Estas demandas motoras no son parte del proceso ejecutivo que se desea medir, pero pueden influir en el rendimiento final y generar confusión entre funciones motoras y ejecutivas (Salthouse, 2011). Segundo, la prueba ofrece pocas medidas, en general solo el tiempo total de resolución de cada condición del test y en algunos casos la correspondiente cantidad de errores, lo que impide un análisis detallado de los procesos subyacentes. Tercero, presenta una alta variabilidad, dado que solo se mide un ensayo por condición, lo que no refleja la capacidad real del participante al ser susceptible a efectos de práctica y a un rendimiento desproporcionadamente más lento en la primera administración (McKeown et al., 2021). En cuarto lugar, las configuraciones espaciales de los ítems son fijas y sus efectos están en gran medida sin explorar; por lo que constituyen un factor de confusión a la hora de comparar la parte A y la parte B (Fossum et al., 1992; Gaudino et al., 1995). En conjunto, estas limitaciones revelan la necesidad de nuevas versiones de la tarea que ayuden a entender el proceso subyacente de las FE, en las que el tiempo y las trayectorias de las manos se midan con mayor precisión y se controlen las configuraciones espaciales.

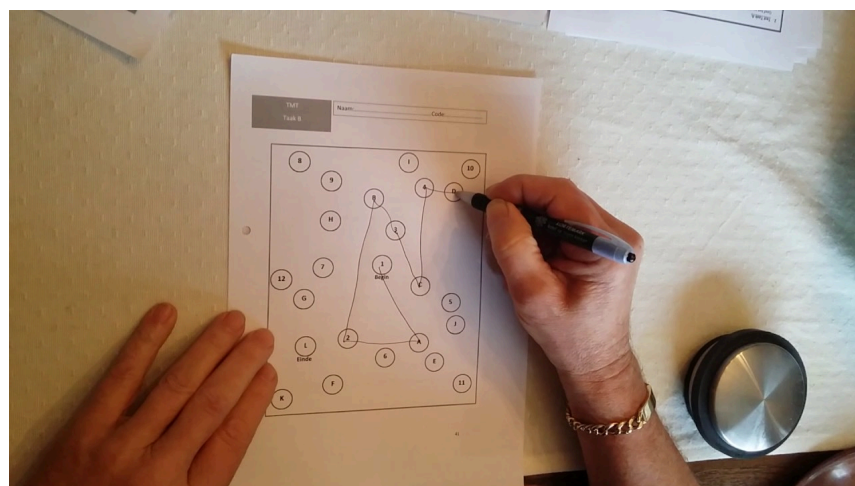


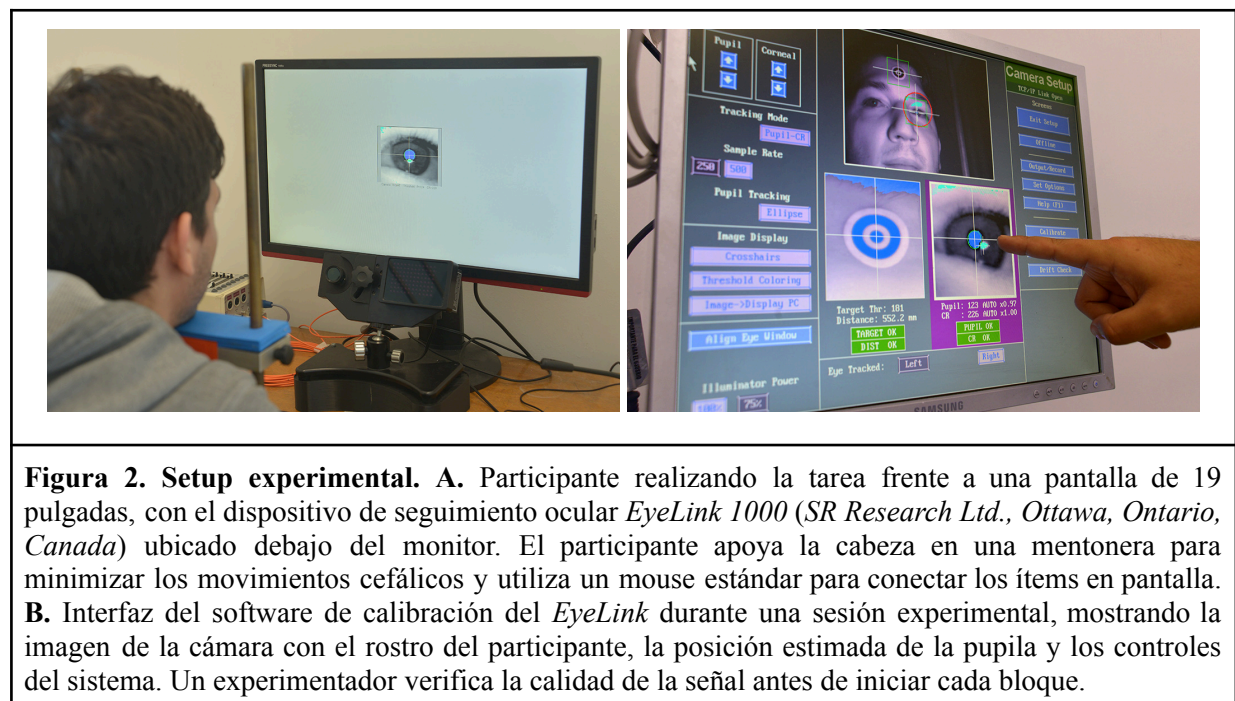
Figura 1. Parte B del Trail Making Test clásico. El participante recibe una hoja y cuando el evaluador la voltea debe comenzar a unir los puntos sin levantar el lápiz. Si comete un error el evaluador se lo indica y vuelve a la última posición correcta. Finalmente se registran el tiempo total en segundos (con un máximo de 300 segundos) y en algunos casos también la cantidad de errores.

Líneas de investigación previas han destacado la importancia de los movimientos de las manos y los ojos como herramientas valiosas para estudiar las FE. Por ejemplo, algunos estudios mostraron una relación estrecha entre la memoria de trabajo espacial y la planificación de los movimientos oculares en varios experimentos (Theeuwes et al., 2009; Woodman & Luck, 2007). Además, otras líneas de trabajo han mostrado una relación entre los marcadores de búsqueda visual y el control inhibitorio (supresión reactiva y proactiva, Huang et al., 2021; Moher & Egeth, 2012). Por otra parte, al menos dos estudios demostraron que los recursos centrales implicados en la selección de respuestas son compartidos por los movimientos de las manos o los ojos durante una tarea secuencial (Remington et al., 2011; Shalom & Sigman, 2015). En general, estos y otros estudios más recientes (por ejemplo, Eckstein et al., 2017) demostraron que muchos procesos cognitivos superiores intervienen en las tareas que requieren movimientos de las manos y los ojos.

En los últimos años, algunos estudios han intentado descomponer el TMT en subcomponentes más pequeños, con el fin de examinar con mayor detalle qué procesos son relevantes durante la tarea. Bajo esa premisa se han desarrollado versiones digitales del TMT que presentan una medida más refinada del tiempo, algunas de las cuales miden las trayectorias de las manos (Dahmen et al., 2017; Fellows et al., 2017; Salthouse & Fristoe, 1995; Woods et al., 2015). Al momento de publicar el trabajo en el que se basa este capítulo (Linari et al., 2022), solo un estudio de seguimiento ocular en el TMT había analizado la tarea en términos de monitoreo y planeamiento, midiendo la interacción entre los movimientos de las manos y los ojos (Wölwer & Gaebel, 2002). Siendo el monitoreo el momento en que las fijaciones oculares se encontraban cerca de la mano, mientras que la planificación involucraba aquellas en las cuales la mano estaba lejos (en base a un umbral definido). A pesar de su trabajo pionero, con importantes contribuciones teóricas y metodológicas, Wölwer y Gaebel midieron los movimientos oculares con un dispositivo de seguimiento ocular de baja resolución y baja frecuencia de muestreo, lo que supuso una fuerte limitación para medir las fijaciones oculares (Wölwer et al., 2003, 2012; Wölwer & Gaebel, 2002).

Uno de los objetivos de nuestro trabajo fue intentar abordar la mayoría de las limitaciones del TMT mediante el diseño de una versión computarizada (Computerized Trail Making Test o cTMT), implementada utilizando *MATLAB* y *Psychtoolbox* y que puede ser ejecutada en una computadora estándar. Esta herramienta incluye múltiples ensayos, control sobre las configuraciones espaciales de los ítems y un registro simultáneo de alta resolución de los movimientos manuales y oculares, utilizando un mouse y un dispositivo de seguimiento ocular (*eye tracker*, en idioma inglés) respectivamente (Fig. 2).

Nuestro diseño permite medir el rendimiento en múltiples ensayos, tiene una configuración espacial controlada y captura los movimientos de las manos y los ojos con alta resolución temporal y espacial. Aún más importante, esta implementación nos permitió revelar diferentes procesos subyacentes. Durante el análisis, segmentamos *a posteriori* la tarea en tres fases: monitoreo, planeamiento y una nueva fase que llamamos exploración. La exploración consistía en aquellos movimientos oculares que examinaban la escena antes del primer movimiento de la mano en un ensayo, recopilando información de la misma antes de comenzar la tarea real de concatenar los ítems en la pantalla (Ver sección *Descomposición de las fijaciones en etapas*).



Asimismo, abordamos la validación del cTMT verificando que logre replicar los patrones de rendimiento del TMT clásico y analizando su convergencia con una batería neuropsicológica estandarizada. Nuestro estudio no solo evaluó los componentes ejecutivos intrínsecos a la tarea, sino que profundizó en su validación externa, focalizándose en la identificación de marcadores de memoria de trabajo y control inhibitorio.

Con base en la literatura previa y el diseño experimental propuesto, establecimos un conjunto de hipótesis específicas. En primer lugar, proponemos que el cTMT replica los patrones de resultados del TMT clásico y muestra validez convergente con las medidas externas de funciones ejecutivas. En segundo lugar, dada la mayor carga cognitiva, proponemos que el cTMT-B exhibe una dinámica oculomotora distintiva respecto al cTMT-A, caracterizada por una mayor cantidad de fijaciones. En tercer lugar, postulamos que estas métricas oculares reflejan las exigencias ejecutivas de alto orden

inherentes al cTMT-B, las cuales se encuentran ausentes en la condición de búsqueda visual simple. Cuarto, hipotetizamos que un subconjunto de estas características es capaz de explicar la variabilidad individual en el rendimiento medido de manera clásica. Finalmente, planteamos la posibilidad de derivar marcadores oculares y manuales novedosos para memoria de trabajo e inhibición, validados mediante su correlación con la evaluación neuropsicológica externa.

Métodos

Participantes

La muestra original consistió en 61 participantes que fueron evaluados con la versión computarizada del Trail Making Test (cTMT). Los mismos fueron reclutados principalmente en la universidad (FCEyN, UBA, Argentina) y por redes sociales. Además reportaron no tener antecedentes neurológicos o psiquiátricos, no estar bajo tratamiento con psicofármacos al momento del reclutamiento, ni tampoco haber consumido alcohol y/o drogas recreativas en las veinticuatro horas previas al experimento. Excluimos a 12 participantes por adquisición deficiente de datos o por no seguir las instrucciones. La muestra final consistió en 49 participantes (24 mujeres; 18-42 años; media $\pm$ DE = 25.7 ± 5.4), excepto para mediciones online adicionales (ver sección *Change detection task*), donde la muestra consistió en 41 participantes (tasa de deserción del 16%, aproximadamente 1 año después). Esta tasa no afectó los resultados principales, dado que un análisis de potencia para una prueba de Wilcoxon sobre PC y RT (Ver sección *Definición de Métricas de Rendimiento*) mostró que solo 12 participantes resultaron suficientes para alcanzar potencia empírica ≥ 0.99 (simulación Monte Carlo, 10000 iteraciones; biblioteca *MKpower* en R). Para este cálculo, estimamos la distribución *a priori* aproximada de PC (RT) a partir de una muestra de cinco participantes con media y desviación estándar de aproximadamente -15 (1.5 ms) y 10 (1.0 ms), respectivamente. Todos los sujetos fueron *naïf* respecto de los objetivos del experimento y presentaron visión normal o corregida a normal. Todos los estudios contaron con revisión y aprobación del IRB del *CEMIC Medical Centre* y calificación del *Department of Health and Human Services (HHS, USA)*: IRb00001745–IORG 0001315. Todos los participantes firmaron consentimiento informado conforme a la Declaración de Helsinki.

TMT computarizado (cTMT)

Implementamos el cTMT en *MATLAB* utilizando el *Psychophysics Toolbox (Psychtoolbox-3)*, de modo que puede ejecutarse en una computadora estándar. La tarea conserva el formato del TMT clásico con varias modificaciones. Redujimos el número de ítems a 20 por ensayo (en lugar de los 25 habituales) para facilitar la distribución espacial en pantalla. El diseño incluye múltiples ensayos por condición, lo que permite obtener medidas estadísticamente estables del rendimiento individual. Además, controlamos la configuración espacial de los ítems igualándola entre las condiciones A y B: a diferencia de la versión original, donde las ubicaciones de los ítems difieren sistemáticamente entre partes, aquí las mismas distribuciones se presentan en ambas condiciones, eliminando ese factor de confusión.

Realizamos el registro conductual de forma simultánea mediante dos canales: los movimientos de la mano, a través de un mouse estándar, y los movimientos oculares, con un dispositivo *EyeLink 1000 (SR Research)* de alta resolución temporal y espacial. Esta combinación

permite extraer medidas internas de cada componente ejecutivo de la tarea. Los detalles del procedimiento y de cada sistema de registro se describen en las secciones subsiguientes.

Procedimiento

Seguimos el diseño original del TMT con algunas modificaciones detalladas a continuación (Fig. 3A). En cada ensayo los participantes debían conectar 20 ítems en orden consecutivo (en lugar de 25, por una cuestión de espacio en la pantalla al generar diferentes distribuciones, ver sección *Generación de estímulos*). En el TMT-A presentamos únicamente números (1 a 20). En TMT-B presentamos números (1 a 10) y letras mayúsculas (A a J) y solicitamos al participante que alterne iniciando en el número 1 (1, A, 2, B, etc.). Dividimos la tarea en 5 bloques de 20 ensayos, separados por cuatro pausas. Cada ensayo inició cuando los participantes presionaron el botón izquierdo del mouse. Al presionarlo, los estímulos aparecían en pantalla y los participantes debían pasar por cada ítem sin soltar el botón, buscando reproducir la ejecución motora del trazo continuo del lápiz sobre el papel en la tarea clásica. Al soltar el botón, los estímulos desaparecían para dar lugar a un punto de fijación en la pantalla. Establecimos un límite de 25 segundos por ensayo. Los ensayos finalizaban al soltar el botón o bien al alcanzar el límite temporal. Instruimos a los participantes para corregir su trayectoria si detectaban que habían pasado por un ítem incorrecto. Registramos los movimientos de los ojos y la mano durante todo el ensayo y no brindamos *feedback* en el momento. Cada participante completó 100 ensayos, 50 TMT-A y 50 TMT-B, alternados estrictamente. La sesión duró entre 40 y 60 minutos, incluyendo calibraciones y recalibraciones del *eye tracker*. Presentamos los estímulos con el software *Psychophysics Toolbox* Version 3 en *MATLAB*.

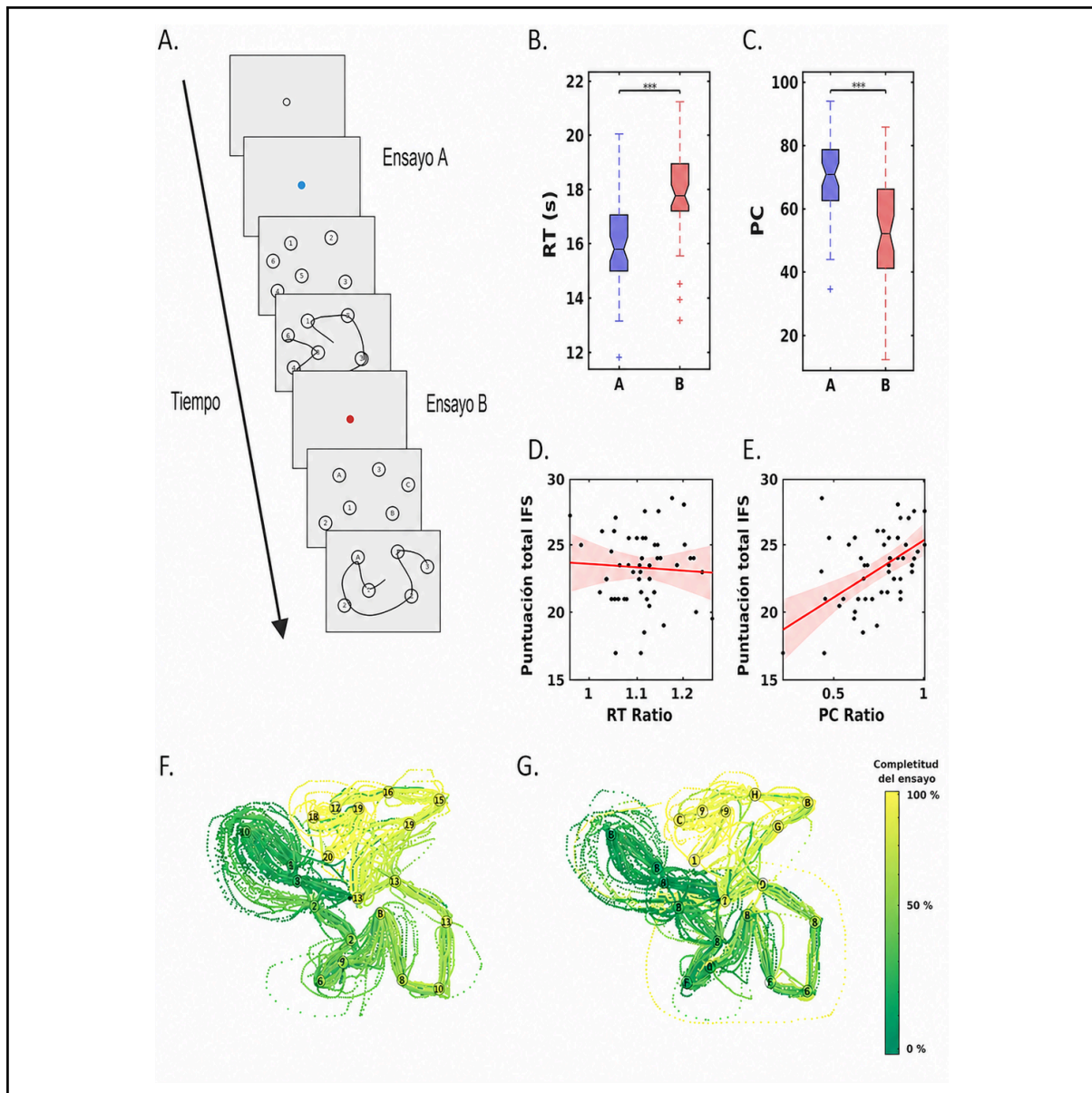


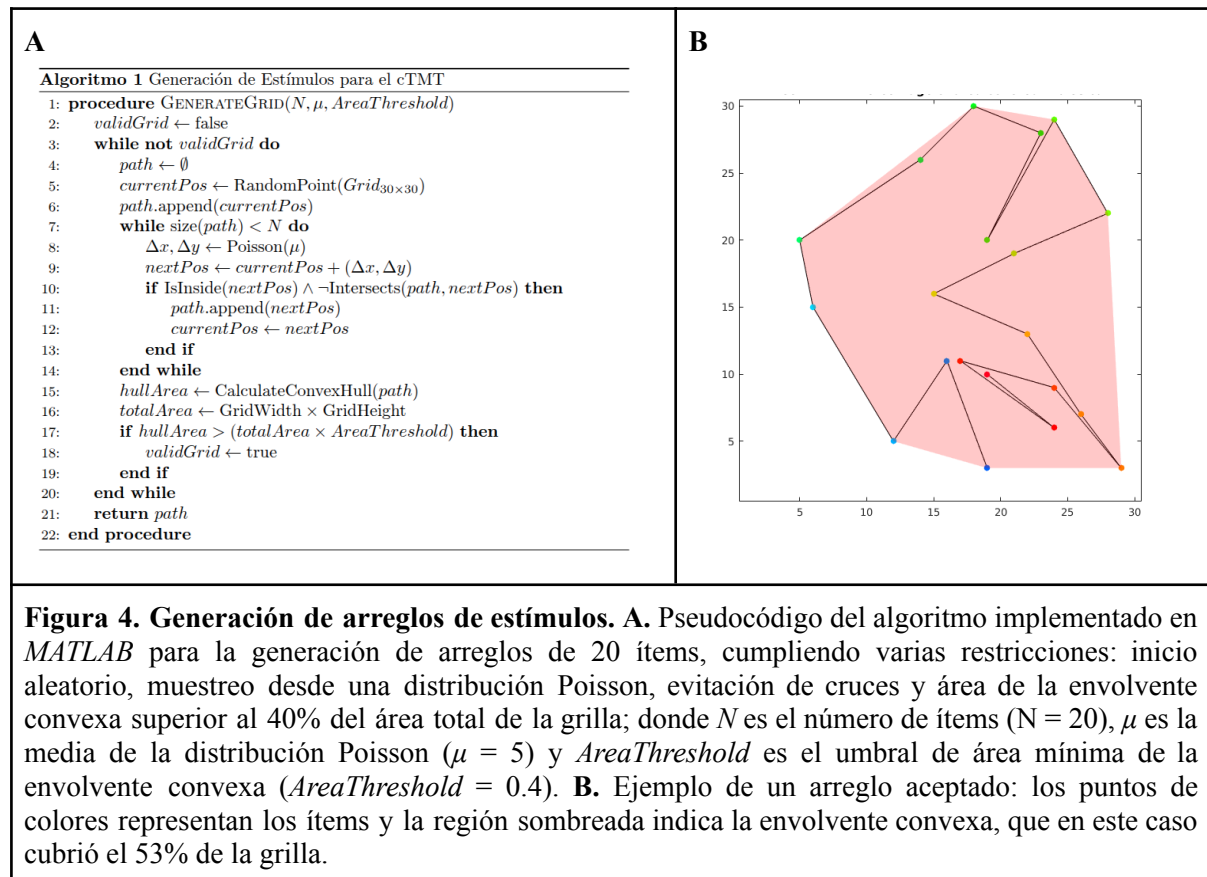
Figura 3. Diseño experimental del cTMT y validación de la tarea. **A.** Secuencia temporal de un ensayo: la prueba comienza al pulsar el botón del mouse, los ítems aparecen en pantalla y el participante debe conectarlos en orden sin soltar el botón, hasta completar la secuencia o alcanzar el límite de 25 segundos. **B.** Tiempo de respuesta (RT): tiempo en conectar 12 ítems en orden correcto (ver sección *Criterios de corrección del ensayo*). **C.** Porcentaje de ensayos completados (PC), para cTMT-A y cTMT-B. Los asteriscos indican diferencias estadísticamente significativas (test de Wilcoxon). **D.** Correlación entre la puntuación total del IFS y el ratio RT (RT-B/RT-A). **E.** Correlación entre la puntuación total del IFS y el ratio de finalización (PC-B/PC-A). **F.** Trayectoria de la mano en un ensayo de tipo A. **G.** Trayectoria de la mano en un ensayo de tipo B con la misma configuración espacial. La barra de colores representa la evolución temporal relativa de cada sujeto.

Cada bloque iniciaba con una corrección de deriva (*drift correction*) del *eye tracker*, en la cual el participante debía fijar su vista sobre un pequeño círculo en el centro de la pantalla (20 píxeles, Fig. 3A), con el objetivo de asegurar una buena calidad de datos. Tras esta corrección, un punto rojo o azul indicaba el tipo de ensayo siguiente (azul para A, rojo para B) y el nuevo ensayo comenzaba al hacer click. Recomendamos descansar entre bloques y reanudar cuando el participante lo considerara oportuno. Antes de continuar, realizábamos nuevamente la corrección de deriva con un punto central.

Si el sistema no detectaba el ojo o la deriva superaba los 2 grados de ángulo visual (valor por defecto del dispositivo), el experimento se detenía y solo se reanudaba luego de que el participante llamara al experimentador y este ejecutara la rutina de recalibración.

Generación de estímulos

Mantuvimos las distribuciones espaciales de ítems constantes entre participantes, pero aleatorizamos el orden y la asignación A/B de las configuraciones espaciales entre sujetos. Seleccionamos las posiciones una a una sobre una grilla 30×30. Elegimos aleatoriamente la posición inicial y extrajimos desplazamientos horizontales y verticales de una distribución Poisson con $\mu = 5$. Tras ubicar 20 ítems, calculamos el área de la envolvente convexa de la trayectoria resultante y aceptamos arreglos solo si presentaron un área $> 40\%$ del área total de la grilla. Las celdas se espaciaron a 20 píxeles (0.44° de ángulo visual). La grilla cubrió 600×600 píxeles. Cada ítem fue un dígito/carácter rodeado por un círculo de radio 10 píxeles, centrado en una celda. Generamos varias distribuciones y seleccionamos 100. El área final de la envolvente convexa fue $50 \pm 7\%$ del área total de la grilla.



Registros conductuales

Los participantes se sentaron frente a una pantalla de 19 pulgadas (*SyncMaster 997 MB*, resolución de 1024x768 píxeles, tasa de refresco de 100 Hz; *Samsung, Suwon, Corea*) a una distancia de 65 cm, con un ángulo de 29.3 grados en horizontal y 22.5 grados en vertical. Utilizamos una mentonera (*chin-rest*) alineada con el centro de la pantalla para limitar los movimientos de la cabeza.

Para el registro de seguimiento ocular, el dispositivo utilizado fue un *EyeLink 1000* (SR Research Ltd., Ottawa, Ontario, Canadá) para registrar la ubicación de la mirada de ambos ojos a una frecuencia de muestreo de 1 kHz. Según los datos proporcionados por el fabricante, la precisión nominal media es de 0.5 grados y la resolución espacial de 0.01 grados de error cuadrático medio. Calibramos la mirada del participante con una cuadrícula estándar de 9 puntos para ambos ojos. Realizamos una corrección de deriva (drift correction) integrada antes de cada bloque de 20 ensayos. Según los resultados de las medidas de deriva, el participante avanzaba en el experimento o tenía que llamar al experimentador para que realizara una nueva calibración.

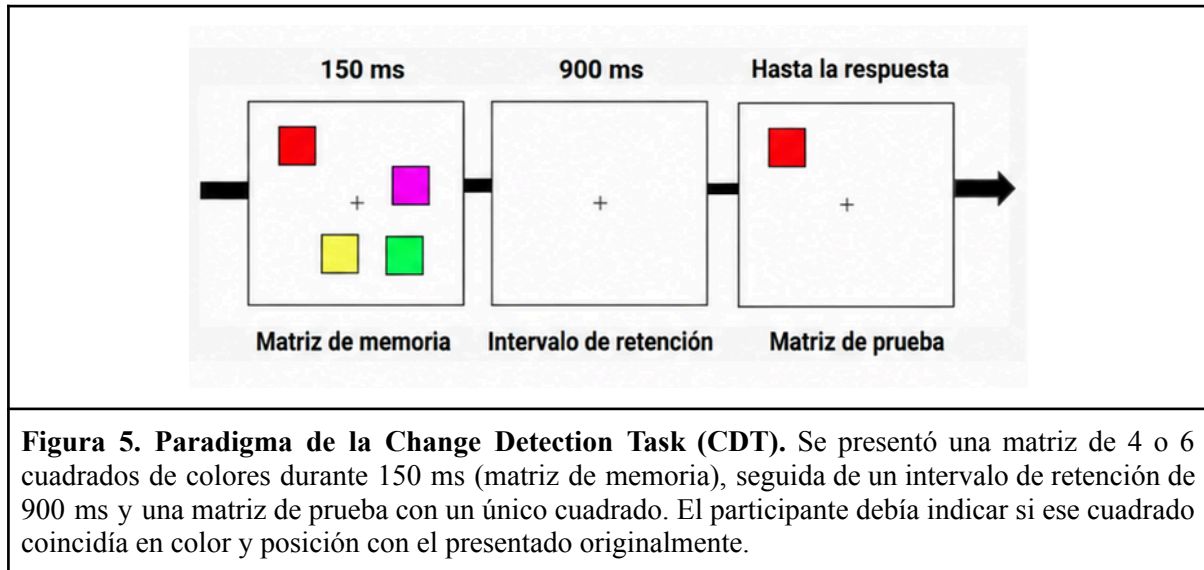
Seleccionamos el ojo mejor calibrado para cada participante basándonos en la exploración visual de cada prueba. Todos los movimientos oculares fueron etiquetados como fijaciones, sacadas y parpadeos por el software del rastreador ocular utilizando los umbrales predeterminados para los experimentos cognitivos (30°/s para velocidad, 8000°/s² para aceleración y 0.1° para movimiento; Cornelissen et al., 2002).

Por su parte, para registrar los movimientos de la mano utilizamos un mouse estándar. El muestreo alcanzó hasta 1000 Hz, no homogéneo, ya que guardamos posiciones en cada marca de tiempo (*timestamp*) solo cuando hubo movimiento. En este caso no afectó el objetivo porque nos enfocamos en eventos de llegada y partida de cada ítem. Extrajimos la secuencia de ítems seleccionados a partir de los datos de mano, con precisión espacial de un píxel.

Medidas de validación externa

Administramos la batería *INECO Frontal Screening* (IFS; Torralva et al., 2009) y una prueba canónica de memoria visual de trabajo, *Change Detection Task* (CDT; Luck & Vogel, 2013). Implementamos la CDT online utilizando la biblioteca *jsPsych* (de Leeuw, 2015) en el lenguaje *JavaScript* y se desplegó en la plataforma *Cognition* (www.cognition.run). Esto fue previo a la creación de nuestra propia plataforma de experimentos online (ver *DataPruebas* en el **Capítulo 4**). Recolectamos el IFS según lo recomendado por su estudio de validación (Torralva et al., 2009). El IFS evalúa funciones ejecutivas con alta sensibilidad para caracterizar déficits en poblaciones clínicas. Incluye una tarea de programación motora, tareas de interferencia y go/nogo basadas en secuencias motoras, una tarea de control inhibitorio verbal en la cual la persona debía completar la palabra final de una oración, tareas de memoria de trabajo verbal y espacial, un backward digit span, y una medida de abstracción con interpretación de proverbios. Cada tarea sumó puntos hasta un total que iba de 0 a 30. En su estudio de validación (Torralva et al., 2009), con un punto de corte de 25, el IFS mostró una sensibilidad de 96.2% y una especificidad de 91.5% para diferenciar controles de pacientes, y el puntaje total correlacionó con pruebas ejecutivas clásicas como el tiempo del TMT-B ($\rho = -0.75$; $p < 0.001$). El IFS posee una buena consistencia interna ($\alpha = 0.80$), sensibilidad para disfunción frontal-ejecutiva y se mantiene notablemente similar con la edad. Lo administramos mediante entrevista con el experimentador y demandó aproximadamente 10 minutos. La tarea de detección de cambios (*change detection task* o CDT) es una evaluación sencilla que permite estimar de manera fiable la capacidad de la memoria de trabajo visual (*VWM*, por sus siglas en inglés). A los participantes se les presentó una matriz de 4 o 6 cuadrados de colores durante 150 ms y, tras un intervalo de 900 ms sin estímulos, aparecía en pantalla un único cuadrado. A continuación debían indicar si ese cuadrado formaba parte de la matriz original, es decir, si tenía el mismo color que el presentado en esa posición específica. Las respuestas se registraron con dos teclas del teclado,

utilizando el dedo índice de cada mano. Hubo ensayos consistentes e inconsistentes. En nuestro diseño experimental *online* incluimos 120 ensayos y medimos los tiempos de respuesta (RT) y las respuestas en teclado (Fig. 5).



Para evaluar la capacidad de VWM, calculamos el número de ítems almacenados en memoria de trabajo para cada tipo de ensayo (K) (Ecuación 1) en los ensayos con matrices de 6 elementos (K_6), de 4 elementos (K_4) y el valor promedio entre ambos (K_{promedio}).

$$K = N_{\text{conjunto}} \left(2 \frac{\text{ensayos correctos}}{\text{todos los ensayos}} - 1 \right) \quad (1)$$

donde N_{conjunto} corresponde al número de cuadrados presentado en un arreglo para un ensayo específico (4 ó 6).

Análisis de datos

Definición de métricas de rendimiento

Para el análisis computacional del rendimiento de la tarea, establecimos un límite temporal de 25 segundos para la finalización de cada ensayo. Dado que la mayoría de los participantes no alcanzaron a completar los 20 ítems del ensayo, decidimos utilizar un criterio de 12 ítems correctamente alcanzados, iniciando desde el primer elemento, para declarar la finalización del ensayo y definir el Tiempo de Respuesta (RT). Este tiempo se cuantificó como el período necesario para lograr la concatenación de los primeros 12 ítems. De manera análoga, para determinar el porcentaje de ensayos completados (PC), calculamos la proporción de ensayos que fueron completados exitosamente hasta el ítem número 12. La elección de este umbral resultó en una estimación eficiente del rendimiento, permitiéndonos mantener una cantidad de datos estadísticamente razonable. Confirmamos la robustez de esta decisión mediante una replicación de los resultados con dos umbrales alternativos, demostrando que los hallazgos principales no dependen de ella (véase la Fig. C1.S1 para una réplica de los resultados con otros dos umbrales). Además, estos criterios generaron resultados sólidos a lo largo de la tarea, dado que no hubo efectos de aprendizaje significativos, como

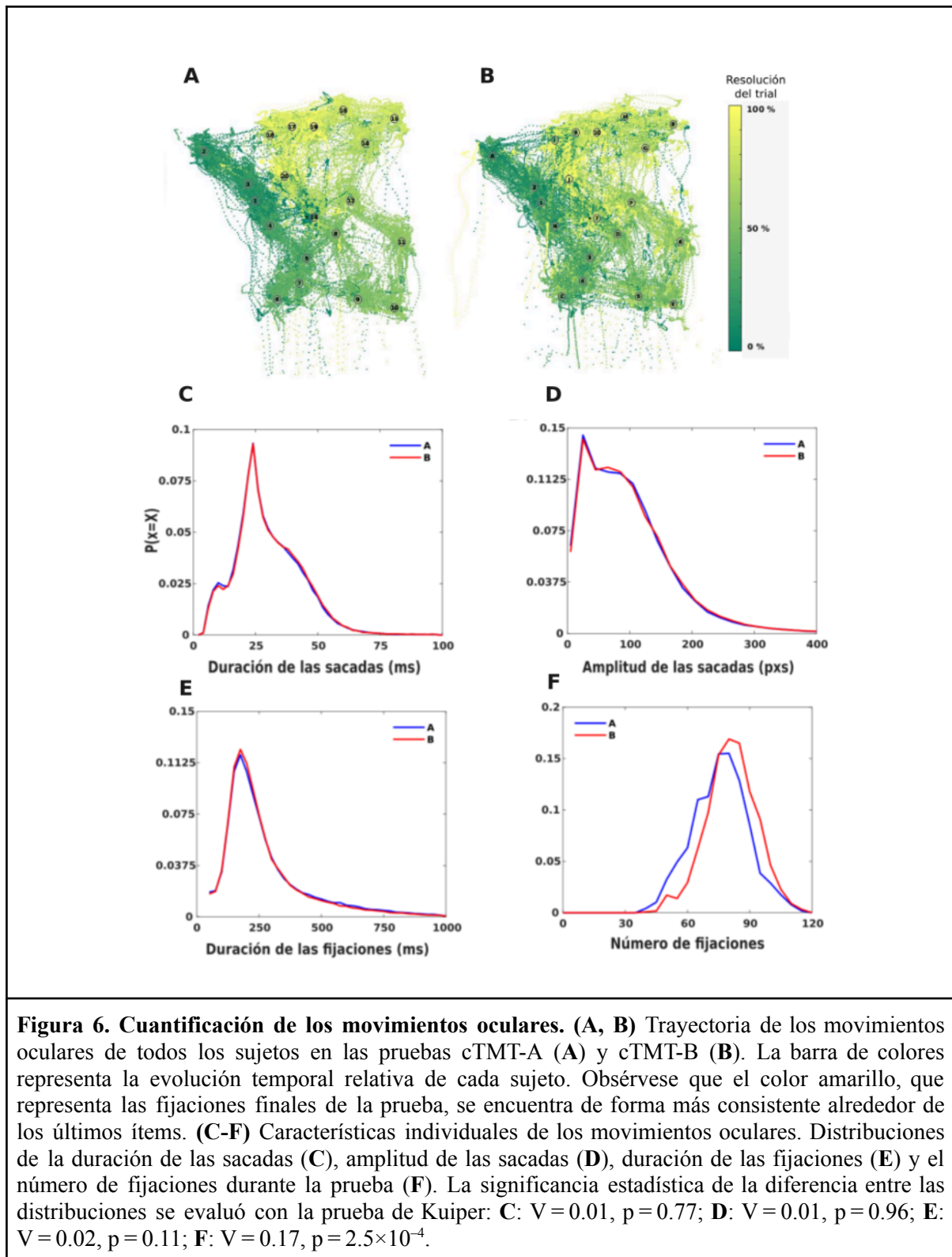
reveló la comparación de los primeros y últimos tercios de los ensayos en relación con la proporción (B/A) para PC y RT.

Criterios de corrección del ensayo

Los ensayos correctos eran aquellos que cumplían el criterio de completitud (definido en la sección *Definición de métricas de rendimiento*) y también presentaban una secuencia de ítems correctamente concatenada. Para ello, la trayectoria trazada por el mouse debía entrar en todos los ítems en el orden correcto (por ejemplo, 1-A-2-B, etc.). Para definir una trayectoria como correctamente concatenada, evaluamos la secuencia de ítems producida por el participante. Un ítem se daba por alcanzado cuando la trayectoria entraba en un umbral de 10 píxeles desde su centro. Un criterio adicional era que las trayectorias no debían cruzarse. En otras palabras, la curva de la trayectoria no debía tocarse a sí misma, como en el TMT original.

Análisis de los movimientos oculares

Las fijaciones son los períodos en que la mirada permanece estable sobre una región y las sacadas los movimientos rápidos que las conectan. Con el fin de comparar sus distribuciones entre cTMT-A y cTMT-B (véase la Figura 6C-F), realizamos un filtrado de los datos crudos conservando solo los ensayos correctos, además descartamos aquellas fijaciones de duración superior a 1000 ms y las sacadas superiores a 100 ms. Dada la disparidad en la duración y cantidad de eventos oculomotores entre las condiciones A y B (siendo la B más extensa), procedimos a equilibrar las muestras mediante un submuestreo (*downsampling*). Redujimos el volumen de datos por un factor de 10, lo que permitió comparar las densidades de probabilidad de fijaciones y sacadas sin el sesgo introducido por la diferencia en el número total de eventos. Debido a ello las distribuciones no cambiaron visualmente después del filtrado. Aplicamos la prueba de Kuiper (Kuiper, 1960) (biblioteca *twosamples* en el lenguaje de programación R; Dowd, 2020) para comprobar estadísticamente las diferencias entre cTMT-A y cTMT-B.

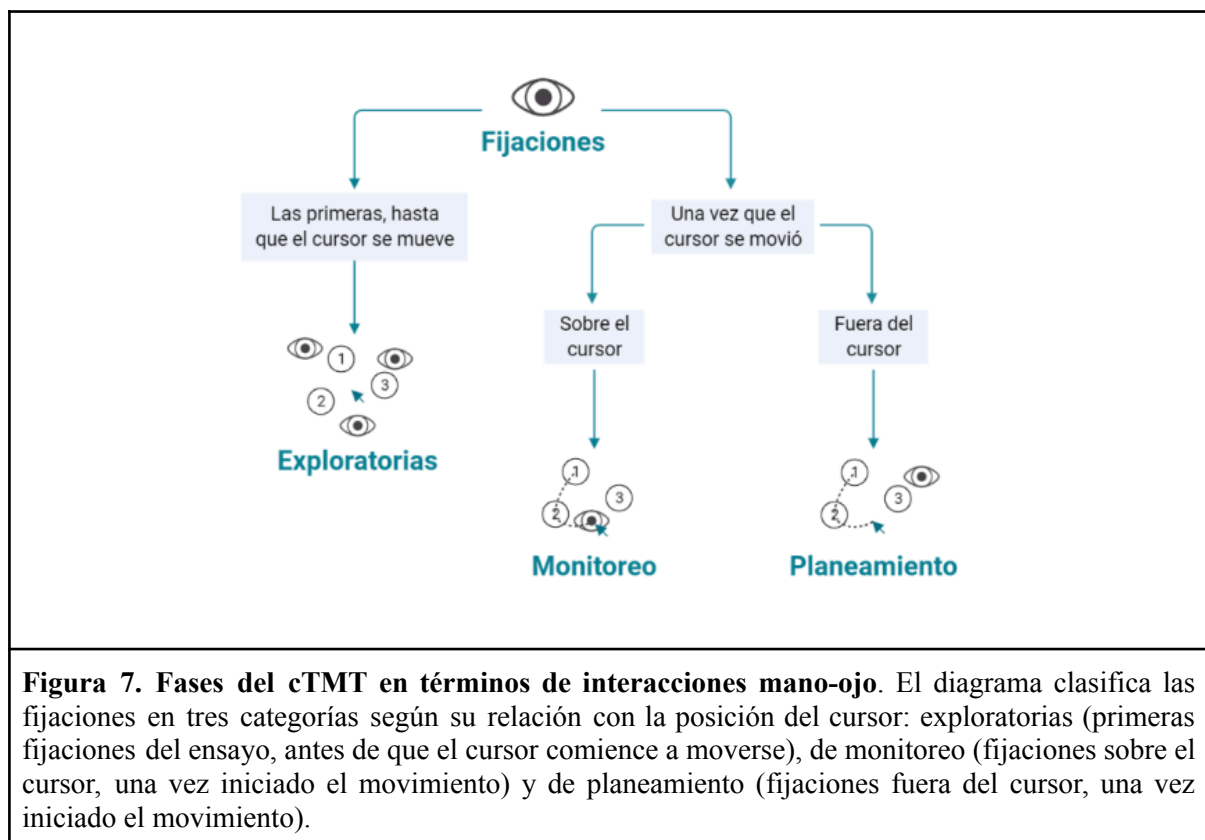


Descomposición de las fijaciones en etapas

Continuando con el trabajo pionero de Wölwer y Gaebel (Wölwer & Gaebel, 2002), clasificamos las fijaciones en función de su relación con la posición del mouse en tres etapas diferentes: fijaciones de monitoreo si se encontraban cerca del cursor (a menos de 25 píxeles) en

cualquier momento durante el intervalo de fijación; fijaciones se definieron como fijaciones de planeamiento si estaban lejos del cursor (a más de 25 píxeles) durante todos los intervalos de fijación (Wölwer et al., 2003, 2012). A partir de las fijaciones de planeamiento, también definimos un nuevo grupo separado denominado fijaciones de exploración, que corresponden a las fijaciones que se producen antes del primer movimiento de la mano.

Calculamos el número medio de fijaciones y la duración media de las fijaciones para cada participante y condición. Utilizamos el test de Wilcoxon signed-rank para comparar las condiciones cTMT-A y cTMT-B. Estimamos los tamaños del efecto (e.s. por sus siglas en inglés) para estas pruebas como $\frac{Z}{\sqrt{n}}$.



Medida interna de la memoria de trabajo

Adicionalmente a las métricas de desempeño global, incorporamos una medida interna diseñada para evaluar específicamente la memoria de trabajo visual: la tasa de ítems recordados. Definimos esta variable operacionalmente identificando aquellos ítems que fueron inspeccionados visualmente (fijados) con antelación, pero cuyo paso con el mouse posterior se ejecutó sin requerir una nueva fijación inmediata (o re-fijación). Es decir, se asume que la acción motora fue guiada por la información espacial almacenada durante el intervalo, sin depender de una validación visual directa al momento de pasar sobre el ítem con el mouse.

La Figura 8 ilustra esta distinción. El caso (i) ejemplifica un ítem recordado: el ítem "2" es observado durante la búsqueda del "1" y, posteriormente, es alcanzado por el cursor directamente, sin necesidad de volver a fijar la mirada sobre él. En contraposición, el caso (ii) representa un ítem no recordado: a pesar de que el "2" fue visto previamente, el participante requiere confirmar su posición con una nueva fijación inmediatamente antes de la acción manual, lo cual sugiere que no se encontraba disponible en la memoria de trabajo al momento de la ejecución.

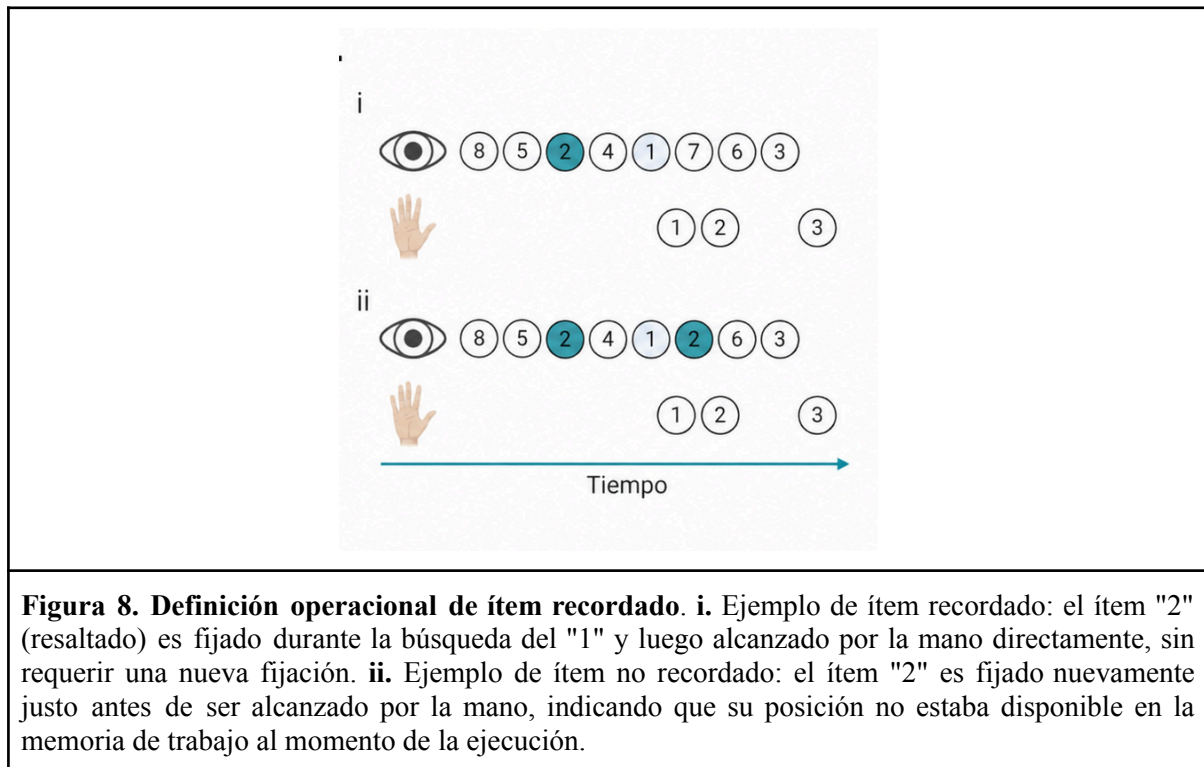


Figura 8. Definición operacional de ítem recordado. i. Ejemplo de ítem recordado: el ítem "2" (resaltado) es fijado durante la búsqueda del "1" y luego alcanzado por la mano directamente, sin requerir una nueva fijación. ii. Ejemplo de ítem no recordado: el ítem "2" es fijado nuevamente justo antes de ser alcanzado por la mano, indicando que su posición no estaba disponible en la memoria de trabajo al momento de la ejecución.

Medida interna del control inhibitorio

Las trayectorias de la mano tienden a dirigirse hacia los objetivos fijados (Bulloch et al., 2015; Fisk & Goodale, 1985) cuando estos son los siguientes en la secuencia (detecciones correctas). Por otro lado, la inhibición ocurre durante las fijaciones en ítems que no continúan la secuencia (detecciones falsas), momento en que la mano debe mantener su trayectoria sin orientarse hacia el ítem (Fig. 9A). La falta de inhibición, en este contexto, queda definida como la tendencia a orientar la trayectoria de la mano hacia las detecciones falsas.

Para agrupar todas las trayectorias manuales proyectadas en la dirección del ítem fijado, primero las segmentamos entre el inicio y el fin de las fijaciones en los ítems medidas con seguimiento ocular. Segundo, restamos la posición de la mano al momento del inicio de la fijación tanto en dirección vertical como horizontal. Por lo tanto, todas las trayectorias comienzan en el origen (0,0). Tercero, las proyectamos en la dirección del ítem fijado y las normalizamos según la distancia entre el punto inicial y el ítem. Así, el ítem fijado quedó ubicado en el punto (1,0). Finalmente, separamos las fijaciones al ítem que en ese momento correspondía alcanzar manualmente según la secuencia correcta de la tarea (detecciones correctas) de las fijaciones a otros ítems (detecciones falsas) (ver Fig. 9A).

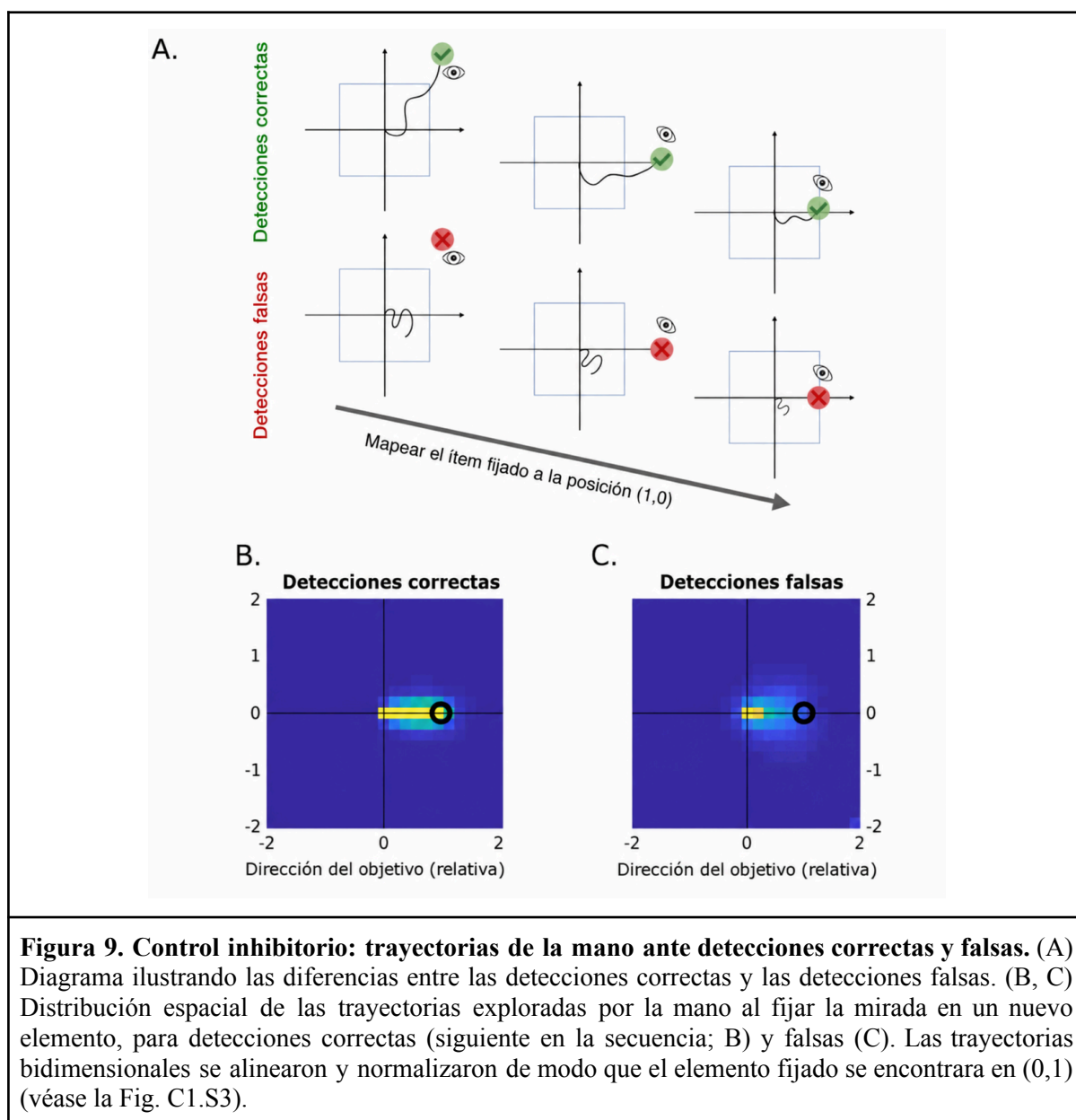


Figura 9. Control inhibitorio: trayectorias de la mano ante detecciones correctas y falsas. (A) Diagrama ilustrando las diferencias entre las detecciones correctas y las detecciones falsas. (B, C) Distribución espacial de las trayectorias exploradas por la mano al fijar la mirada en un nuevo elemento, para detecciones correctas (siguiente en la secuencia; B) y falsas (C). Las trayectorias bidimensionales se alinearon y normalizaron de modo que el elemento fijado se encontrara en (0,1) (véase la Fig. C1.S3).

Estimamos la distribución espacial de las trayectorias mediante el histograma 2D de las mismas (ver Fig. C1.S3). Estimamos el patrón temporal como la posición relativa al ítem fijado (y proyectada como se describió anteriormente) en función del tiempo.

Resultados

Movimientos de la mano: rendimiento global del cTMT

El análisis del desempeño global reveló un aumento significativo en el RT en la parte B comparado con la A (*Wilcoxon signed-rank test*: $p = 1.1 \times 10^{-9}$, $z = -6$, e.s. = 0.86; Fig. 3B). Asimismo, el PC fue menor en B comparado con A (*Wilcoxon signed-rank test*: $p = 1.1 \times 10^{-9}$, $z = 6$, e.s. = 0.86; Fig. 3C). Es importante notar que estos resultados se mantienen incluso considerando solo los

primeros 30 ensayos, incluyendo ambas condiciones y excluyendo los primeros para descartar posibles errores (primeros 2 ensayos tanto de TMT-A como de TMT-B) (ver Fig. C1.S2). Esto es consistente con resultados previos para la tarea TMT de lápiz y papel (Margulis et al., 2018; Tombaugh, 2004).

A continuación, se evaluó la asociación entre el rendimiento basado en movimientos de la mano y las FE. Observamos una correlación significativa entre el porcentaje de completitud (PC-B/PC-A) y el puntaje total del IFS (Fig. 3E; Correlación de Spearman: $\rho = 0.437$, $p = 1.7 \times 10^{-3}$), pero no con el ratio de RT (Fig. 3D; Correlación de Spearman: $\rho = 0.006$, $p = 0.97$). Estos resultados sirvieron como validación de la versión propuesta del TMT. Cabe mencionar que el hecho de que la correlación de la puntuación del IFS con el PC fuera significativa, pero no con el RT, podría ser una consecuencia directa de la presión temporal debido al límite de tiempo de nuestro diseño, ausente en el TMT original.

En resumen, los resultados de nuestro cTMT se asemejan al TMT clásico, incluso tomando solo los primeros 30 ensayos. Estos resultados se relacionan exclusivamente con el cambio de listas (solo números o números y letras) ya que las mismas configuraciones espaciales se presentaron en ambos tipos de ensayos. Finalmente, el rendimiento mostró correlación con la puntuación del IFS, una prueba externa de *screening* de funciones ejecutivas.

Movimientos oculares

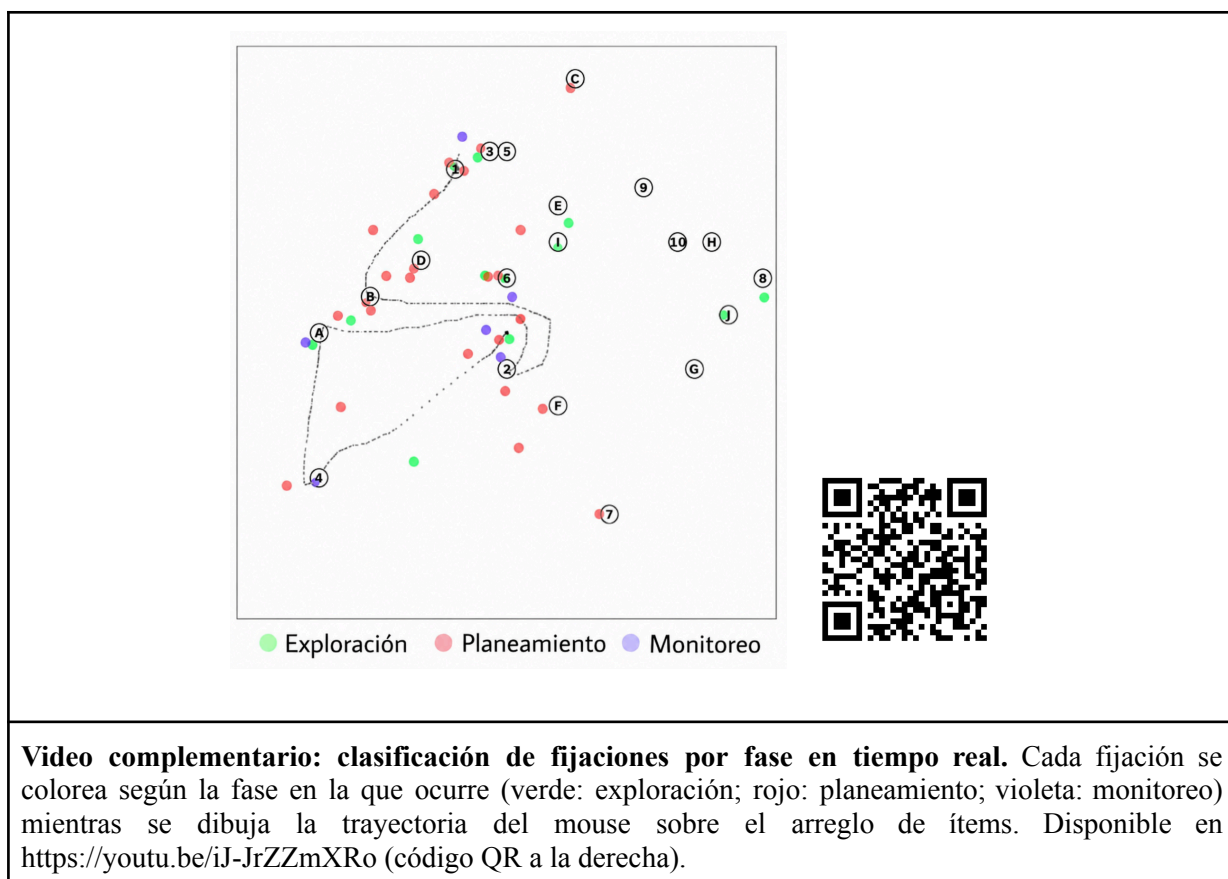
Observamos una estructura similar en las trayectorias de exploración visual (*scanpaths*, en idioma inglés), excepto que el cTMT-A presenta, cualitativamente, mayor consistencia de color a lo largo de la trayectoria, revelando que en el cTMT-A casi todos los ensayos alcanzaron el último ítem. La Figura 6A,B ilustra los patrones de exploración ocular de dos ensayos representativos (cTMT-A y cTMT-B, respectivamente) con idéntica configuración espacial. El cTMT-B, por otro lado, presentó mayor variabilidad en el número de ítems alcanzados, lo cual también se refleja en las barras de error más grandes observadas en el gráfico de barras de PC en la Fig. 3C.

En cuanto a las fijaciones y sacadas, cTMT-A y cTMT-B fueron indistinguibles en varias medidas clásicas de seguimiento ocular, incluyendo la duración de las sacadas y las fijaciones, así como la amplitud de las sacadas (Fig. 6C-E; prueba de Kuiper: $V = 0.01$, $p = 0.77$; $V = 0.01$, $p = 0.96$; $V = 0.02$, $p = 0.11$ respectivamente). Sin embargo, el número de fijaciones mostró una clara diferencia entre ambas condiciones, con un número mayor en cTMT-B (prueba de Kuiper: $V = 0.17$, $p = 2.5 \times 10^{-4}$; mediana (TMT-A) = 76, mediana (TMT-B) = 81; Fig. 6F). La ausencia de diferencias en la duración y amplitud de las sacadas, junto con la duración de las fijaciones sugieren un mecanismo visual similar entre las partes A y B. La diferencia observada específicamente para el número de fijaciones podría tener su origen en un procesamiento más complejo de la tarea, más relacionado con procesos cognitivos de orden superior que con mecanismos visuales. En otras palabras, para resolver ambas partes de la tarea, los sujetos parecen utilizar su maquinaria visual de una manera muy similar, excepto en los ensayos de cTMT-B donde requeriría un escaneo más intensivo de la escena visual. Para explorar con más detalle los posibles mecanismos implicados en el rendimiento diferencial de A y B, dividimos la tarea en tres fases y analizamos los movimientos oculares en cada una de ellas.

Descomposición de la tarea basada en movimientos oculares y manuales

Como se ha observado en la sección anterior, la diferencia en el tiempo necesario para completar la tarea en ambas condiciones se explica, principalmente, por el número de fijaciones realizadas durante la prueba, y no por la duración de la fijación o la duración de la sacada. En esta sección, nos centraremos en el número de fijaciones en nuestro análisis. Nuestro objetivo fue comprender qué aspectos de la resolución de la tarea cambian entre las partes A y B, revelados por el tipo de fijación.

En el trabajo anterior que ha analizado trayectorias visuales con cierto detalle durante la ejecución del TMT, las fijaciones durante el TMT se han clasificado en dos fases: planeamiento y monitoreo (Wölwer & Gaebel, 2002). Aquí utilizamos una clasificación similar, con el agregado de una nueva fase de exploración inicial. Esta última corresponde a aquellas fijaciones que se producen al inicio de cada ensayo antes del movimiento de la mano y tiene en cuenta no solo la búsqueda del primer elemento, sino también la exploración inicial de la escena (Fig. 7). Por otro parte, la fase de monitoreo consistía en fijaciones que se producían cerca del cursor y estaban más relacionadas con la ejecución motora de la tarea, mientras que la fase de planeamiento consistía en fijaciones que se producían más lejos del cursor y las asociamos con aspectos más ejecutivos de la tarea. Para ilustrar la clasificación de las fases, creamos un vídeo en el que las fijaciones se colorean según la fase en la que se producen en tiempo real (véase el vídeo complementario a continuación).



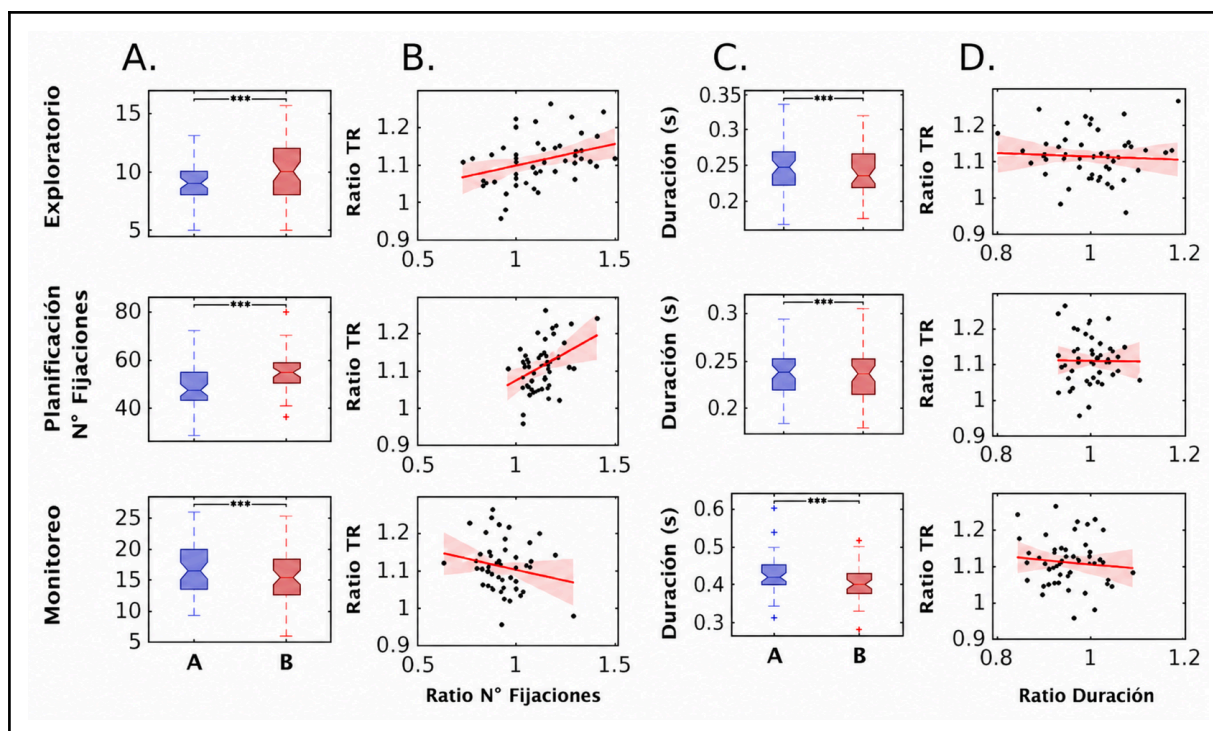


Figura 10. Descomposición de la tarea en fases de movimientos oculares. (B) Boxplots de la mediana del número de fijaciones en cada una de las tres fases (exploración, planeamiento y monitoreo) para ambas partes (cTMT-A y cTMT-B). (C) Correlación entre la relación RT (RT-B/RT-A) y la relación del número de fijaciones (B/A). (D) Boxplot de la duración media de las fijaciones en cada fase (exploratoria, de planeamiento y de monitoreo) para ambas partes (cTMT-A y cTMT-B). (E) Correlaciones entre la relación RT (RT-B/RT-A) y la relación de duración de las fijaciones en cada fase.

En primer lugar, comparamos el número de fijaciones y la duración de las fijaciones entre el cTMT-A y el cTMT-B en cada fase. Observamos un mayor número de fijaciones exploratorias (Wilcoxon signed-rank test: $p = 9 \times 10^{-4}$, $z = -3.3$, e.s. = 0.47) y de planeamiento en la parte B (Wilcoxon signed-rank test: $p = 1.9 \times 10^{-9}$, $z = -6$, e.s. = 0.86), siguiendo la tendencia de la tarea general. Por el contrario, hubo un menor número de fijaciones de monitoreo en la parte B (Wilcoxon signed-rank test: $p = 1.1 \times 10^{-4}$, $z = 3.8$, e.s. = 0.54; Fig. 10B). En cuanto a la duración de las fijaciones, no hubo diferencias significativas entre cTMT-A y cTMT-B en la exploración (Wilcoxon signed-rank test $p = 0.98$, $z = -0.01$) y el planeamiento (Wilcoxon signed-rank test: $p = 0.7$, $z = 0.38$). Sin embargo, en la fase de monitoreo, la duración de las fijaciones fue mayor en el cTMT-A (Wilcoxon signed-rank test: $p = 2.4 \times 10^{-6}$, $z = 4.7$, e.s. = 0.67; Fig. 10D). Las figuras 10B y 10D muestran que el número de fijaciones era más informativo que la duración, lo que explica las diferencias entre cTMT-B y cTMT-A, y a su vez es coherente con la distribución de los movimientos oculares representada en la Figura 6.

Para explorar la asociación entre los movimientos oculares y el rendimiento en las tres fases de la tarea, calculamos la relación (B/A) entre el número de fijaciones y su duración correspondiente, y fueron correlacionadas con una medida del rendimiento (RT-B/RT-A). Encontramos correlaciones significativas entre la relación RT y la relación del número de fijaciones en las fases de exploración (correlación de Spearman: $\rho = 0.39$, $p = 5.2 \times 10^{-3}$) y planeamiento (correlación de Spearman: $\rho = 0.29$, $p = 4.3 \times 10^{-2}$; Fig. 10B,C), pero no en el caso de las de monitoreo (correlación de Spearman: ρ

= -0.22, $p = 0.13$). Esto vuelve a ser coherente con las diferencias en la distribución del número de fijaciones (Fig. 6F). Como se observa en las distribuciones de la Fig. 6E, la duración de las fijaciones no varió entre las condiciones, por lo que era de esperar que la relación de duración de las fijaciones no afectara a la relación de RT (correlación de Spearman en exploración: $\rho = -0.05$, $p = 0.71$; planeamiento: $\rho = -0.02$, $p = 0.91$; monitoreo: $\rho = -0.05$, $p = 0.74$; Fig. 10E).

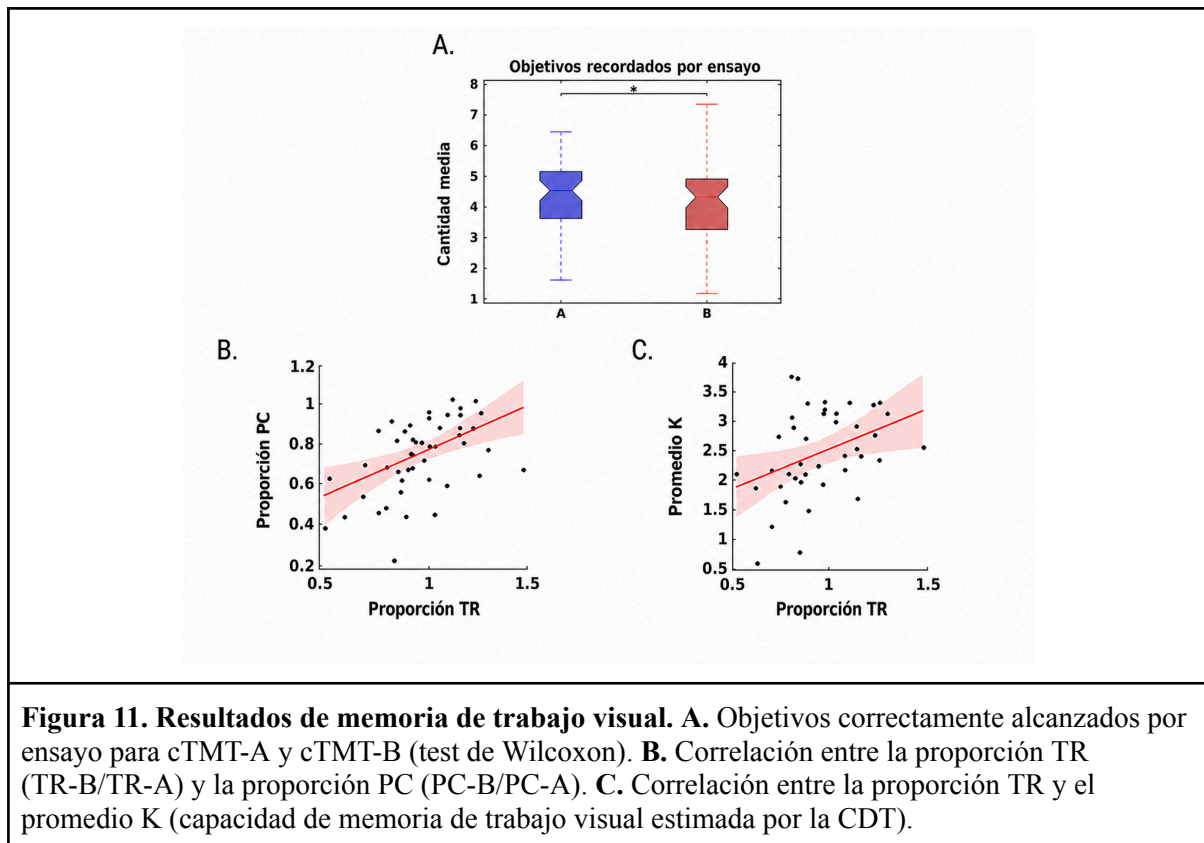
En resumen, el número de fijaciones, pero no la duración/amplitud de las fijaciones o sacadas, varió entre las partes A y B y proporcionó medidas adecuadas del rendimiento de la tarea. La división de la tarea en tres fases reveló los diferentes aspectos del proceso ejecutivo (exploración, planeamiento y monitoreo). El aumento del número de fijaciones en B frente a A, tanto en la exploración como en el planeamiento, así como la disminución del monitoreo, caracterizaron las diferentes etapas. Además, se observó un aumento pequeño pero significativo del tiempo de fijación en la parte A frente a la B solo en la fase de monitoreo. Por último, el aumento del número de fijaciones observado en B/A para las fases de exploración y planeamiento se correlacionó positivamente con el rendimiento relativo B/A (ratio RT).

Memoria de trabajo visual

En esta sección, derivamos una medida interna de la memoria de trabajo visual utilizando los movimientos oculares y manuales en el cTMT. Posteriormente inspeccionamos de qué manera esta medida interna de la memoria de trabajo visual afectaba al rendimiento en el cTMT-B con respecto al cTMT-A. Por último, comparamos la medida derivada con el rendimiento individual en una tarea validada de memoria de trabajo visual. Para cuantificar nuestra medida, estimamos el número de ítems recordados junto con la búsqueda (TR), es decir, el número de ítems que no tenían fijaciones justo antes de que la mano los alcanzara (véase la sección Medida interna de la memoria de trabajo), incluyendo solo los ensayos correctos.

En promedio, los participantes recordaron más objetivos en el cTMT-A que en el cTMT-B (Fig. 11A; $A = 4.60 \pm 0.87$; $B = 4.38 \pm 0.98$; Wilcoxon signed-rank test: $p = 0.028$; $z = 2.2$, e.s. = 0.31). Este resultado concuerda con un menor número de fijaciones generales en el cTMT-A, dado que una mayor memoria de la ubicación de los ítems implica una menor búsqueda de los mismos en la escena (véase la figura 6). También concuerda con un menor número de fijaciones de planeamiento (véase la figura 10). Esto sugiere que los participantes memorizaron la ubicación de más ítems por adelantado y tuvieron que volver a mirar al mismo menos veces para completar correctamente la prueba en el cTMT-A.

La relación TR (TR en B/TR en A) se correlacionó con la relación PC (Fig. 11B; correlación de Spearman: $\rho = 0.48$, $p = 5.3 \times 10^{-4}$), lo que indica que la mejora relativa en el recuerdo de los ítems en B se asoció con el rendimiento general de la tarea. Además, se evaluó la asociación de la relación TR con una medida externa de la memoria de trabajo, la capacidad de la memoria de trabajo visual (K_{promedio}) estimada a partir de una tarea de detección de cambios (CDT, Fig. 5) (véase la sección Medida interna de la memoria de trabajo), que mostró una correlación moderada (Fig. 11C; correlación de Spearman: $\rho = 0.43$, $p = 5.3 \times 10^{-3}$, $N = 41$).



En resumen, extrajimos una nueva medida interna de la memoria de trabajo visual en el cTMT que se correlacionaba tanto con el rendimiento (ratio PC) como con una medida externa canónica de la memoria de trabajo visual (CDT), lo que sugiere que es posible aislar componentes individuales de la FE dentro del cTMT.

Control inhibitorio

A partir de los datos sobre los movimientos oculares y manuales, nos propusimos derivar una segunda medida interna asociada a una función ejecutiva como el control inhibitorio. Cuando la mirada se fija en un nuevo ítem, este puede ser el siguiente de la secuencia o no, es decir, puede tratarse de una detección correcta o incorrecta del objetivo. En este último caso, la trayectoria manual debe evitar seguir a los ojos y esperar hasta que se encuentre el ítem correcto. Este comportamiento es evidente cuando se alinean todas las trayectorias exploradas por la mano después de fijar la vista en un nuevo ítem (Fig. 9B, C, Fig. C1.S3). La distribución espacial de estas trayectorias muestra que, cuando se identificaba un ítem correcto, la mano se movía directamente hacia el objetivo (Fig. 9B), mientras que cuando se producía una detección falsa, la mano permanecía inmóvil (Fig. 9C) o se movía en otras direcciones (Fig. C1.S3), lo que muestra una inhibición de las acciones motoras tempranas. Para cuantificar este comportamiento, estimamos el desplazamiento en la dirección del nuevo ítem, para las detecciones correctas y falsas, y tanto para el cTMT-A como para el cTMT-B.

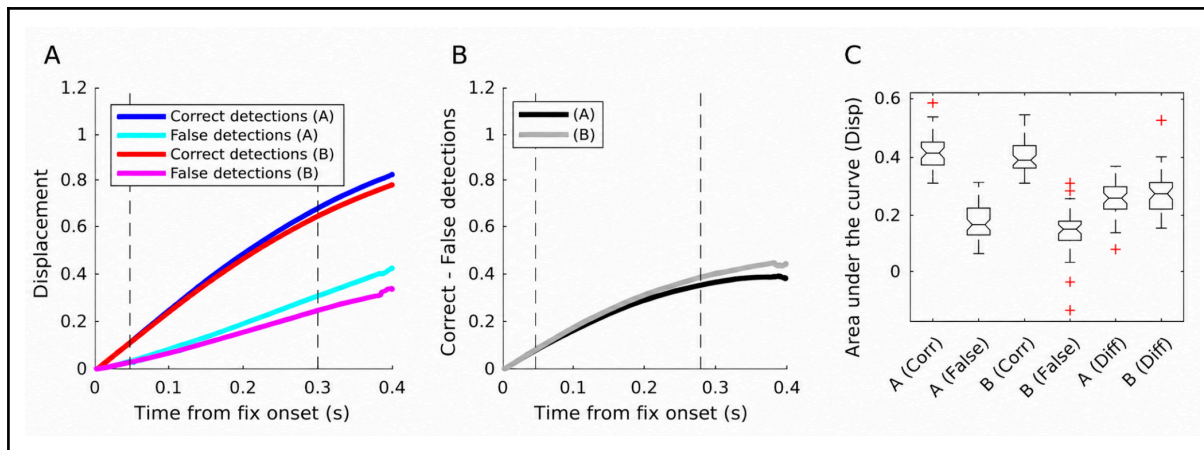


Figura 12. Control inhibitorio: desplazamiento de la mano ante detecciones correctas y falsas. **A.** Desplazamiento de la mano en la dirección del nuevo ítem en función del tiempo desde el inicio de la fijación, para detecciones correctas y falsas en cTMT-A y cTMT-B. **B.** Curvas de diferencia entre detecciones correctas y falsas para cTMT-A y cTMT-B. **C.** Área bajo las curvas de diferencia para cTMT-A y cTMT-B (test de Wilcoxon).

En consonancia con las distribuciones espaciales, el desplazamiento de la mano para las detecciones correctas fue mayor que para las detecciones falsas, alcanzando la posición del objetivo (desplazamiento = 1, Fig. 12A) y revelando un proceso motor inhibitorio. Curiosamente, las curvas en cTMT-A y cTMT-B fueron similares, como también se evidenció en las curvas de diferencia entre detecciones correctas y falsas (Fig. 12B). El área bajo las curvas de diferencia fue significativamente diferente de cero tanto para cTMT-A como para cTMT-B (Fig. 12C; Wilcoxon signed rank test: TMT-A: $p < 10^{-8}$, TMT-B: $p < 10^{-8}$, $z = 6.1$, e.s. = 0.87), pero no hubo diferencias significativas entre cTMT-A y cTMT-B (Wilcoxon signed rank test, $p = 0.29$, $z = 1.1$).

En resumen, extrajimos una nueva medida interna de inhibición en el cTMT que podría captar la dinámica de los procesos de control inhibitorio dentro de la tarea, pero no mostró correlación significativa con las medidas externas de CI del IFS. Una posible explicación es la multidimensionalidad del control inhibitorio y la naturaleza de las pruebas inhibitorias del IFS: su subtest Go/NoGo, que evalúa inhibición motora simple, suele estar saturado en participantes neurotípicos, mientras que su subtest de Hayling evalúa inhibición verbal. Ambas dimensiones difieren de la inhibición visoespacial que captura el cTMT.

Discusión

En el presente capítulo, nuestro objetivo fue diseñar una versión computarizada del TMT (cTMT) que pudiera mitigar sus principales limitaciones de la versión en lápiz y papel. En concreto, derivamos nuevas medidas dentro de la tarea para reflejar procesos de FE individuales, basándonos en el registro preciso de los movimientos de la mano y los ojos. En primer lugar, validamos el cTMT demostrando que los tiempos de reacción y los perfiles de rendimiento son coherentes con el TMT clásico. Además, observamos una correlación significativa entre la tasa de finalización (PC) y una batería independiente de funciones ejecutivas (IFS). En segundo lugar, demostramos que las características de los movimientos oculares eran muy similares en el cTMT-A y el cTMT-B, y solo diferían en el número de fijaciones, lo que implica que los mecanismos visuales son similares entre las

condiciones, pero difieren en los procesos de nivel superior. En tercer lugar, cuando la tarea se dividió en tres etapas diferentes (exploración, planeamiento y monitoreo), encontramos un mayor número de fijaciones de exploración y planificación en el cTMT-B, y un menor número de fijaciones de monitoreo. Esto podría interpretarse como mayores costes de planificación y ejecutivos (de alto nivel) en B, y menos recursos dedicados a procesos de menor nivel (supervisión de los movimientos de las manos). En cuarto lugar, la cantidad media de objetivos recordados fue mayor en el cTMT-A, y la proporción de objetivos recordados entre cTMT-B y cTMT-A se correlaciona con la tasa de finalización de toda la tarea.

Según nuestro conocimiento, al momento de la publicación del artículo en que se basa este capítulo (2022), este fue el primer estudio que utilizó movimientos oculares y manuales de alta resolución durante el TMT. Incluso resultando una cita frecuente entre trabajos del área (~70 citas al momento de la publicación de esta tesis) considerándolo como un antecedente de relevancia (Del Gatto et al., 2025; Goset et al., 2024; Recker & Poth, 2023).

Un aspecto importante de nuestro trabajo es que pudimos diseccionar la tarea y extraer marcadores individuales de las FE, abordando una de las principales limitaciones del TMT tradicional, lo que convierte al cTMT en una herramienta prometedora para la investigación y el uso clínico.

Partiendo desde nuestra primera hipótesis, replicamos los resultados generales del TMT clásico: la resolución de las pruebas de tipo B llevó más tiempo, mientras que el porcentaje de finalización fue mayor en las pruebas de tipo A. Además, encontramos una correlación entre la puntuación total del IFS y la relación de finalización (PC-B/PC-A), pero no con la relación de tiempos de resolución (RT), probablemente debido al factor limitante del tiempo (los participantes tenían la presión de resolver cada ensayo en un máximo de 25 segundos, a diferencia del TMT clásico). Las implementaciones digitales anteriores del TMT ampliaron el análisis de la versión clásica al extraer más características (Dahmen et al., 2017; Fellows et al., 2017; Salthouse & Fristoe, 1995; Woods et al., 2015), pero en este trabajo también nos centramos en extraer medidas internas como marcadores de las FE.

Al momento de publicar nuestro estudio, la literatura sobre movimientos oculares en el TMT presentaba limitaciones técnicas: se analizaron fijaciones globales con alta resolución (Hicks et al., 2013) o se intentó descomponer la tarea empleando dispositivos de baja frecuencia (Jyotsna et al., 2020; Wölwer et al., 2003, 2012; Wölwer & Gaebel, 2002). Al emplear un *eye tracker* de alta resolución, observamos que las dinámicas oculomotoras fueron similares entre condiciones, con la excepción de un mayor número de fijaciones en el cTMT-B. Esto sugiere que ante una mayor carga cognitiva, se intensifica la exploración visual sin alterar mecánicamente las sacadas, un correlato de la asignación de recursos atencionales consistente con la evidencia previa (Hicks et al., 2013; Hyönä, 2010).

Expandiendo la taxonomía en fases propuesta por Wölwer (planeamiento y monitoreo), introdujimos las fijaciones exploratorias y evaluamos no solo los tiempos totales (Wölwer et al., 2003, 2012; Wölwer & Gaebel, 2002), sino la frecuencia y duración de cada clase. Encontramos que el cTMT-B demandó más fijaciones exploratorias y de planeamiento (asociadas a procesos ejecutivos superiores (Gouveia et al., 2007)), en concordancia con patrones observados en poblaciones clínicas. Inversamente, la duración de las fijaciones de monitoreo (ligadas al control motor) fue mayor en el cTMT-A, lo que indica que en contextos de alta demanda cognitiva, los participantes compensan

sacrificando el monitoreo visomotor continuo para priorizar la planificación. La duración de los movimientos sacádicos se mantuvo sin cambios significativos en ambas condiciones.

Respecto a la métrica de memoria, cuantificamos los ítems recordados (TR) observando un mayor registro en el cTMT-A, lo que es coherente con una menor necesidad de búsqueda visual. La asociación de TR tanto con el rendimiento general de la tarea como con la prueba externa CDT (Luck & Vogel, 2013) aporta validez a nuestra métrica. Cabe señalar que trabajos anteriores relacionaban el rendimiento en el TMT con la memoria de trabajo, pero los resultados dependían de las pruebas que se administraban (pruebas canónicas y complejas, como la Escala de Memoria de Wechsler (WMS) y el Test de Clasificación de Cartas de Wisconsin (WCST), entre otras) (Kortte et al., 2002). A diferencia de esos abordajes generales, nuestro enfoque es pionero al correlacionar directamente marcadores internos del TMT con evaluaciones específicas de la memoria de trabajo visual.

Adicionalmente, buscamos evaluar el control inhibitorio (CI) a partir de las trayectorias visomotoras. El rol del CI en el TMT ha sido objeto de debate (Sánchez-Cubillo et al., 2009): mientras algunos autores reportan correlaciones con pruebas de interferencia como Stroop (Arbuthnott & Frank, 2000; Chaytor et al., 2006; Spikman et al., 2001), otras medidas específicas arrojan resultados contradictorios (Langenecker et al., 2007; Miner & Ferraro, 1998), reflejando la naturaleza multidimensional del CI (Diamond, 2013; H. C. Meyer & Bucci, 2016; Sperl et al., 2021). Al cuantificar el desplazamiento manual ante detecciones visuales correctas y falsas, observamos diferencias significativas que evidencian un proceso inhibitorio interno. Sin embargo, no hallamos correlación con las métricas externas del IFS, posiblemente porque este evalúa inhibición motora simple o verbal (Diamond, 2013; Torralva et al., 2009), mientras que el cTMT demanda inhibición visoespacial. Futuras investigaciones deberán emplear paradigmas específicos (ej., Flanker o Spatial Stroop) para desenredar los subcomponentes del CI involucrados en el TMT.

Resulta relevante destacar que este estudio se realizó en población neurotípica, donde la menor variabilidad reduce la probabilidad de hallar grandes diferencias individuales. A pesar de ello, logramos identificar correlaciones significativas con evaluaciones externas, lo cual nos resulta alentador en pos de su aplicación en cuadros neurodegenerativos (como Alzheimer o demencia frontotemporal) donde el TMT es fundamental y el seguimiento ocular cobra cada vez más fuerza diagnóstica como lo hemos mostrado en un reciente trabajo de revisión sistemática que publicamos recientemente (Verón et al., 2025). Dado que la duración de las evaluaciones es crítica en contextos clínicos, demostramos que nuestro protocolo captura efectos robustos analizando únicamente los ensayos iniciales.

Finalmente, el diseño de nuestro cTMT permitió mitigar limitaciones intrínsecas de la versión clásica. Al igualar la configuración espacial de los ítems en ambas condiciones, controlamos un sesgo histórico que explicaba gran parte de las diferencias temporales entre TMT-A y TMT-B (Fossum et al., 1992; Gaudino et al., 1995), asegurando que las variaciones observadas respondan al costo del cambio de reglas cognitivas. Además, al incluir múltiples ensayos, logramos estimaciones más estables que estudios previos (Fellows et al., 2017; Poreh et al., 2012; Salthouse, 2011; Wölwer et al., 2003, 2012; Wölwer & Gaebel, 2002; Woods et al., 2015), confirmando estadísticamente la ausencia de efectos de aprendizaje significativos durante la sesión, tal como reveló la comparación entre el primer y el último tercio de los ensayos en cuanto a la relación (B/A) para la tasa de completitud (PC) y el tiempo de respuesta (RT).

Conclusiones

Nuestra contribución con el cTMT demuestra que es posible transformar una prueba neuropsicológica clínica estándar en una herramienta capaz de diseccionar estos componentes con alta resolución. Al extraer marcadores continuos y detallados del comportamiento visomotor, no solo logramos una caracterización más robusta de los mecanismos cognitivos implicados durante una única tarea compleja, sino que generamos un conjunto de características (*features*) fundamentales para el entrenamiento de modelos predictivos.

Este avance metodológico supera sustancialmente las limitaciones del TMT de lápiz y papel, sentando las bases para el desarrollo de herramientas diagnósticas digitales más precisas para trastornos como la enfermedad de Alzheimer o demencias frontotemporales. Asimismo, allana el camino hacia evaluaciones más ecológicas basadas en el comportamiento espontáneo en entornos naturales.

Capítulo 2: Aplicación del cTMT en deterioro cognitivo leve: integración de seguimiento ocular y manual

Este capítulo se basa en el trabajo publicado en la conferencia ETRA (Eye Tracking Research and Applications):

Juantorena, G. E., Berrios, W., Fernández, M. C., Ibáñez, A., Petroni, A., & Kamienkowski, J. E. (2024). Enhancing Cognitive Assessment: Integrating Hand and Eye Tracking in the Digital Trail-Making Test for Mild Cognitive Impairment. In 2024 Symposium on Eye Tracking Research and Applications (ETRA '24). ACM.

Introducción

En el **Capítulo 1** demostramos que el cTMT captura procesos ejecutivos con mayor detalle que la versión clásica en población neurotípica. El paso siguiente fue evaluar su utilidad clínica en una población con deterioro cognitivo leve (DCL), un estadio frecuentemente considerado prodrómico de la enfermedad de Alzheimer y otras demencias (Rostagno, 2023). En esta etapa, los déficits cognitivos son sutiles y a menudo escapan a las métricas globales del TMT tradicional, lo que hace especialmente relevante contar con herramientas de evaluación digital de mayor resolución (Germine et al., 2019; Smith et al., 2023).

La Organización Mundial de la Salud (2017) proyecta que la carga global de la demencia casi se triplicará en los próximos treinta años. Este síndrome neurológico abarca diversos trastornos que deterioran progresivamente la memoria, la cognición y el funcionamiento diario, con una heterogeneidad considerable en sus causas subyacentes y manifestaciones clínicas (Mukadam et al., 2024). Sin embargo, estas cifras ocultan importantes disparidades regionales: los países de ingresos bajos y medios, particularmente en América Latina, presentan las tasas de prevalencia más altas y los aumentos más pronunciados, lo que genera desafíos significativos en regiones donde los recursos sanitarios son escasos (Aranda et al., 2021; Baez et al., 2023; Custodio et al., 2017; Kalaria et al., 2008).

La enfermedad de Alzheimer (EA), que representa aproximadamente el 70% de los casos de demencia, se define actualmente como un continuo biológico donde los cambios neuropatológicos preceden a los síntomas clínicos por años (Jack Jr. et al., 2024). Este marco enfatiza la importancia de la detección temprana. En este contexto, el Deterioro Cognitivo Leve (DCL), anteriormente considerado meramente prodrómico, se reconoce ahora como parte de este continuo de la EA. Aunque no todos los individuos con DCL progresan a demencia, el riesgo es considerable: estudios clínicos reportan tasas de conversión de alrededor del 40% en cinco años, con variabilidad según el subtipo y el contexto de evaluación (Salemme et al., 2025).

Nuestro objetivo fue aplicar el cTMT con registro simultáneo de movimientos oculares y manuales a un grupo de pacientes con DCL y controles sanos apareados por edad. Buscamos identificar marcadores digitales continuos que permitieran discriminar entre ambos grupos de manera más sensible que las medidas clásicas limitadas al tiempo total y el porcentaje de completitud (Piau et al., 2019).

Consideraciones metodológicas: EyeLink en modo remoto

En el Capítulo anterior, participantes neurotípicos realizaron la tarea con el *eye tracker* *EyeLink 1000 Plus* en modo estabilizado, apoyando la cabeza en una mentonera que minimiza los movimientos de la cabeza, permitiendo adquirir datos con mejor calidad (Fig. 2A). Sin embargo, durante una experiencia piloto con adultos mayores neurotípicos (N=9), varios participantes manifestaron incomodidad con este dispositivo, lo que dificultaba su permanencia durante toda la sesión experimental.

Para abordar esta limitación, decidimos utilizar el *EyeLink 1000 Plus* en su modo remoto. En esta configuración, se coloca un pequeño sticker adhesivo con forma de blanco en la frente del participante (la figura 2 muestra un *setup* que incluye tanto la mentonera (2.A) como el blanco (2.B)), que la cámara del *eye tracker* rastrea junto con los reflejos corneales. Este seguimiento dual permite al sistema compensar los movimientos de la cabeza en tiempo real, eliminando la necesidad de la mentonera y ofreciendo una experiencia más confortable para los participantes. El modo remoto tolera desplazamientos de cabeza dentro de un volumen de captura (*headbox*) especificado por el fabricante, lo que nos pareció más adecuado para poblaciones clínicas y adultos mayores que pueden tener dificultades para mantener una postura rígida durante períodos prolongados.

No obstante, estudios previos han documentado que la calidad de los datos en modo remoto se degrada cuando los participantes adoptan posiciones no óptimas o realizan movimientos frecuentes fuera de la zona de captura, con incrementos en la pérdida de datos y *offsets* sistemáticos en la posición de la mirada (Niehorster et al., 2018). Esta limitación resultó particularmente relevante en nuestro estudio, como discutiremos más adelante.

Métodos

Participantes

Reclutamos 15 pacientes con DCL (edad: 75.8 ± 7.0 años) y 17 controles sanos pareados por edad (72.5 ± 8.9 años) en el Hospital Italiano de Buenos Aires. Todos los participantes tenían visión normal o corregida y proporcionaron consentimiento informado por escrito. El protocolo fue aprobado por el Comité de Ética del Hospital Italiano de Buenos Aires (ID: 11543). Los pacientes fueron diagnosticados con DCL basándose en una entrevista clínica, un reporte de un informante y una evaluación neuropsicológica que evidenció un rendimiento de -1.5 desvíos estándar en al menos dos pruebas de un mismo dominio cognitivo (atención, memoria, lenguaje, funciones ejecutivas o habilidades visoespaciales), con actividades de la vida diaria preservadas (Jak et al., 2009).

Protocolo experimental

Para este estudio reimplementamos el cTMT en *Python* utilizando *PsychoPy* (J. W. Peirce, 2007), un software de código abierto que ofrece mayor flexibilidad para integrar distintos dispositivos de registro, facilita la adaptación del protocolo experimental y resulta más accesible para su uso en entornos clínicos. La tarea fue administrada en una computadora de escritorio estándar (Intel Core i3, 8GB de RAM, Windows 11) con un mouse óptico USB Logitech M90 sobre una superficie dura y

uniforme. Los estímulos se presentaron en un monitor Samsung SyncMaster 997MB de 19 pulgadas a una resolución de 1024x768 píxeles y una frecuencia de actualización de 100 Hz, a una distancia de visualización de aproximadamente 65 cm. Los movimientos del mouse se registraron a una tasa de muestreo de ~60 Hz. Registramos la mirada del ojo derecho a 500 Hz utilizando un EyeLink 1000 Plus en modo remoto, con una calibración estándar de 9 puntos.

Basándonos en nuestro trabajo previo (Linari et al., 2022), adaptamos el protocolo a las características de la población de adultos mayores. Respecto a esa versión, redujimos el número de ítems por ensayo de 20 a 15 y aumentamos el radio de los círculos de los ítems de 10 a 30 píxeles para facilitar la adquisición de los ítems. La selección de ítems se realizó pasando el cursor sobre cada objetivo, sin necesidad de mantener presionado el botón del mouse, y se proporcionó retroalimentación visual continua resaltando los ítems cuando fueran alcanzados correctamente. En el cTMT-A, los ítems consistieron en números del 1 al 15 en orden ascendente; en el cTMT-B, en números del 1 al 8 y letras de la A a la G, que los participantes debían conectar en secuencia alternada (1-A-2-B-...). La distribución espacial de los ítems fue idéntica para todos los participantes y se presentó en el mismo orden; los arreglos utilizados fueron un subconjunto de los generados en Linari et al. (Linari et al., 2022).

Los participantes iniciaban cada ensayo haciendo clic en una cruz de fijación central, tras lo cual los estímulos aparecían en pantalla. La tarea completa incluyó 22 ensayos: 2 de práctica (uno de tipo A y uno de tipo B, no incluidos en el análisis) seguidos de 20 ensayos experimentales alternando cTMT-A y cTMT-B. Previo a los ensayos experimentales, el experimentador brindó instrucciones verbales mientras se mostraban instrucciones escritas simultáneamente en pantalla. Cada ensayo finalizaba al alcanzar el último ítem o al cumplirse un límite de 25 segundos. La duración total de la tarea no superó los 12 minutos, incluyendo la calibración.

Adicionalmente, como parte del proceso diagnóstico, los participantes completaron una batería de pruebas neuropsicológicas estándar no digitales: el Mini-Mental State Examination (MMSE; Tombaugh & McIntyre, 1992), el Trail Making Test en papel y lápiz (partes A y B) (Bowie & Harvey, 2006), el Digit Symbol Substitution Test y el Span de dígitos directo e inverso (subtests del WAIS-IV) (Hartman, 2009), y el Test del Reloj (Shulman, 2000).

Análisis

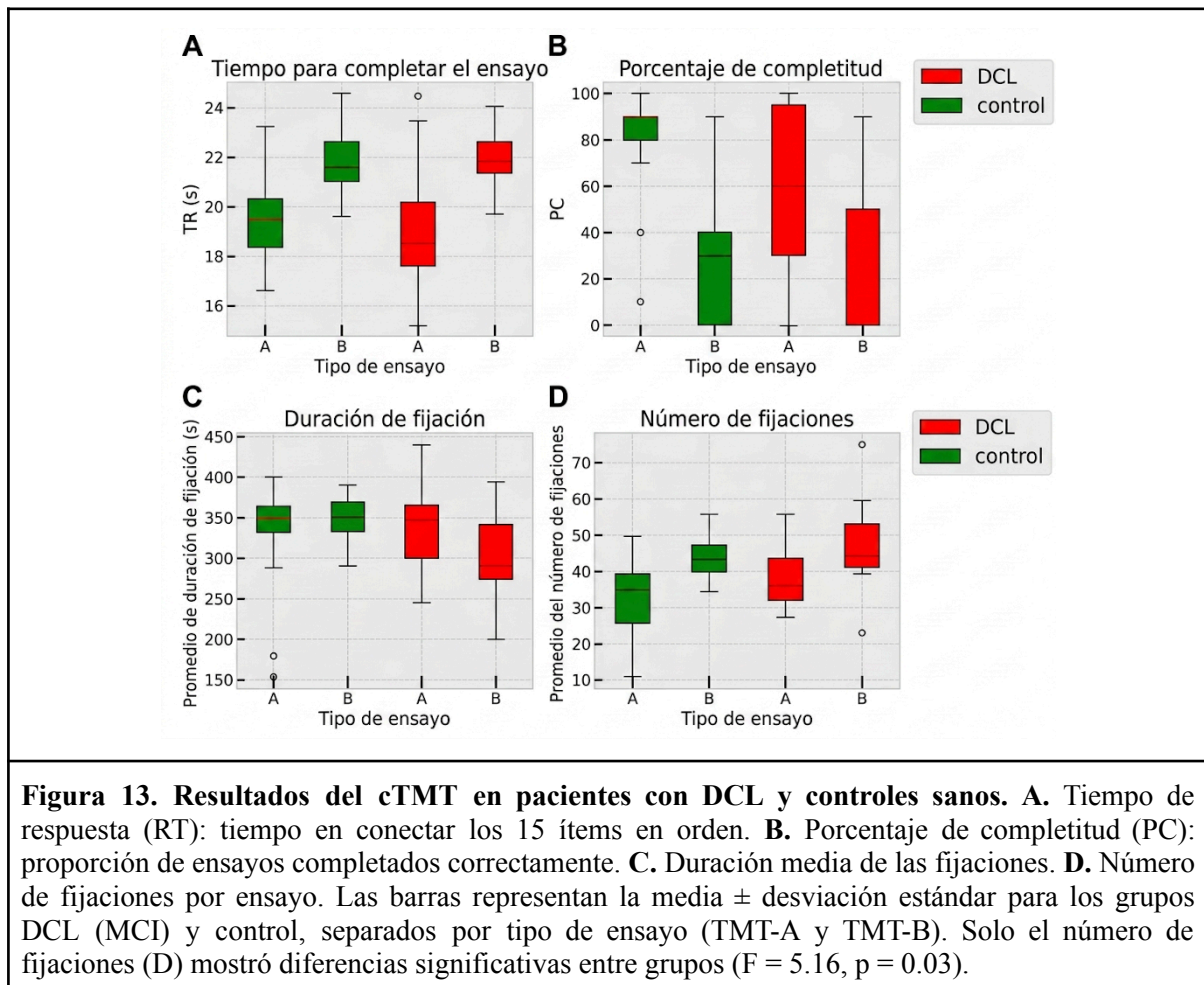
Para los movimientos oculares, descartamos fijaciones menores a 50 ms o mayores a 1000 ms, y calculamos la duración de las fijaciones y el número de fijaciones por ensayo, separando por tipo de ensayo (A o B) y grupo (DCL o control). Para los movimientos manuales, calculamos el porcentaje de ensayos completados correctamente (PC) y el tiempo de respuesta (RT). Utilizamos ANOVAs de dos factores (grupo $\times$ tipo de ensayo) para estimar los efectos principales y su interacción.

Resultados

Las trayectorias del mouse y su variabilidad fueron visualmente similares entre grupos. El análisis confirmó que tanto el RT (Figura 13.A) como el PC (Figura 13.B) mostraron diferencias significativas entre tipos de ensayo (RT: $F = 44.22$, $p = 9.9 \times 10^{-9}$; PC: $F = 32.70$, $p = 3.6 \times 10^{-7}$), pero no

entre grupos (RT: $F = 0.26$, $p = 0.61$; PC: $F = 2.16$, $p = 0.15$), ni interacciones significativas. Estas medidas clásicas no lograron discriminar entre pacientes con DCL y controles.

En cambio, al explorar las métricas oculares, encontramos diferencias significativas en el largo del *scanpath* (número de fijaciones por ensayo, Figura 13.D) tanto para el tipo de ensayo ($F = 8.53$, $p = 4.91 \times 10^{-3}$) como para el grupo ($F = 5.16$, $p = 0.03$), sin interacción significativa. Los pacientes con DCL realizaron más fijaciones que los controles en ambas condiciones, sugiriendo una estrategia de búsqueda visual menos eficiente. La duración de las fijaciones (Figura 13.C), en contraste, no mostró diferencias por grupo ni por tipo de ensayo.



Discusión

A pesar de los resultados prometedores del *scanpath length*, la recolección de datos de *eye tracking* en esta población presentó desafíos considerables. Si bien el modo remoto del EyeLink mejoró la comodidad de los participantes respecto a la mentonera, los pacientes con DCL tendían a seguir menos la consigna de permanecer relativamente quietos frente a la pantalla. Movimientos frecuentes de cabeza fuera del volumen de captura, dificultades para mantener la atención durante la calibración y problemas particulares como el uso de lentes multifocales produjeron una mayor proporción de datos perdidos o de baja calidad en el grupo clínico.

Estos inconvenientes son fuertemente consistentes con la evidencia previa sobre la degradación de la calidad del *eye tracking* remoto en participantes sin restricción de movimiento (Niehorster et al., 2018). De hecho, Niehorster y colaboradores demostraron empíricamente que el EyeLink 1000 Plus en su configuración remota es particularmente susceptible a generar errores sistemáticos (*offsets*) en la estimación de la mirada cuando los participantes adoptan posturas no óptimas.

Adicionalmente, a diferencia de la versión original del **Capítulo 1**, la reimplementación en *PsychoPy* no contó con una rutina de corrección de deriva (*drift correction*) entre bloques, lo que impidió mitigar estos desplazamientos progresivos generados por los cambios posturales durante la sesión.

La muestra reducida (15 DCL, 17 controles) y la proporción de datos oculares descartados limitaron el poder estadístico para explorar métricas oculares más sofisticadas, como las fases de exploración y planeamiento definidas en el **Capítulo 1**, o para entrenar modelos predictivos basados en estas señales.

Conclusiones

Este estudio piloto demostró que las métricas oculares del cTMT, en particular el número de fijaciones, pueden capturar diferencias entre pacientes con DCL y controles que escapan a las medidas conductuales clásicas. Este resultado refuerza el valor del seguimiento ocular como herramienta complementaria en la evaluación cognitiva digital, en línea con la creciente evidencia de su potencial diagnóstico en demencias (Verón et al., 2025).

Sin embargo, las dificultades prácticas en la adquisición de datos de movimientos oculares de calidad en población clínica nos llevaron a replantear la estrategia para el siguiente estudio. Con una muestra ampliada de participantes (tanto DCL como controles) y los datos del mouse disponibles con calidad consistente, decidimos adoptar un enfoque predictivo basado exclusivamente en los movimientos de la mano. Este enfoque, que presentamos en el **Capítulo 3**, tiene la ventaja adicional de no requerir hardware especializado, lo que facilita su escalabilidad y eventual implementación clínica.

Capítulo 3: Evaluación digital del deterioro cognitivo leve mediante aprendizaje automático y movimientos de la mano en el Trail Making Test

Este trabajo se encuentra publicado en la colección “Machine Learning in Psychology” de la revista Scientific Reports de la editorial Nature:

Juantorena, G.E., Capelo, G., Leon Vallejos, B.D. et al. Machine learning based digital assessment of mild cognitive impairment using mouse trajectories during the trail making test. Sci Rep (2026). <https://doi.org/10.1038/s41598-026-62955-9>

Este capítulo también incluye trabajo de la siguiente tesis de grado, de la cual fui co-director:

Gianluca Capelo “Predicción de funciones ejecutivas con aprendizaje automático: Estudios de caso en cohortes online y contexto clínico presencial” (2026)

Introducción

Los primeros esfuerzos por digitalizar el Trail Making Test (TMT) sentaron las bases para aprovechar los métodos computacionales en la evaluación cognitiva. Dahmen y colaboradores (2017) introdujeron una versión digital del TMT y aplicaron técnicas de aprendizaje automático para evaluar su validez predictiva en resultados clínicos y cognitivos. Los avances recientes en las adaptaciones digitales se han mostrado prometedores para mejorar la evaluación cognitiva en diversas poblaciones y modalidades. Por ejemplo, Park y Schott (2022) compararon los TMT tradicionales en papel y los informatizados en adultos jóvenes y también en mayores, y descubrieron que las versiones digitales pueden detectar con precisión las diferencias relacionadas con la edad. Varios estudios han empleado herramientas digitales para la detección precoz de enfermedades neurodegenerativas: Kobayashi y colaboradores (2022) utilizaron una batería de tareas digitales donde las personas debían trazar líneas o dibujar (entre las que se incluía el TMT) para detectar enfermedad de Alzheimer, mientras que Chudzik et al. (2024) analizaron la dinámica durante una versión de TMT *online* para diferenciar los patrones de la enfermedad de Parkinson. Otro grupo (Du et al., 2022) aplicó modelos ocultos de Markov a secuencias digitalmente generadas del TMT, extrayendo marcadores cognitivos más allá del tiempo total. Por su parte, Verma y Dhyan (2025) validaron una aplicación del TMT en Android con grupos de personas sanas o que transitaban un cuadro de depresión, respaldando el uso de evaluaciones móviles en distintas poblaciones. Basándose en esta tendencia, Lara-Garduno y colaboradores (2022) presentaron *SmartStrokes*, una aplicación de TMT para tablets que aplica modelos de clasificación a nivel de segmentos del trazo para detectar deterioro cognitivo leve (DCL). Por su parte Hanczár y colaboradores (2022) clasificaron DCL a partir de segmentos cortos de movimiento del mouse durante el TMT utilizando árboles de decisión y software propietario. Finalmente, Del Gatto y colaboradores (2025) desarrollaron el eTMT, una versión digital con la que demostraron que el número de distractores en pantalla afecta el rendimiento en la tarea, aportando evidencia sobre el rol de la búsqueda visual activa en el TMT más allá del cambio de tarea.

En paralelo, desarrollos recientes han utilizado herramientas de aprendizaje automático para analizar datos conductuales multimodales (incluyendo movimientos oculares, comportamiento motor e interacciones digitales) en la detección temprana de diferentes afecciones como el DCL o la enfermedad de Alzheimer. Por ejemplo, Revankar y colaboradores (2025) demostraron que la evaluación del comportamiento utilizando *smartphones*, incorporando datos motores de alta

resolución al interactuar con los dispositivos, puede distinguir entre la enfermedad de Alzheimer y la demencia de cuerpos de Lewy. Estos resultados son consistentes con esfuerzos más amplios en neurociencia y psiquiatría que aplican inteligencia artificial para extraer marcadores cognitivos latentes a partir de datos de alta dimensionalidad (Badrulhisham et al., 2024; Chekroud et al., 2021). Utilizando marcadores digitales basados en movimientos de mano y ojo durante una tarea de realidad virtual, Kim et al. (2023) entrenaron modelos de aprendizaje automático para la detección temprana del deterioro cognitivo, evidenciando las ventajas de las nuevas tecnologías en potenciales aplicaciones clínicas. Asimismo, Dörr et al. (2025) extrajeron características del habla durante una tarea de fluidez verbal semántica en participantes con DCL o envejecimiento saludable y entrenaron un clasificador de aprendizaje automático para examinar efectos de práctica.

Este desplazamiento desde puntuaciones estáticas hacia métricas dinámicas e interpretables sugiere que el aprendizaje automático aplicado a modalidades ecológicamente válidas (como el seguimiento ocular, la voz y los movimientos de la mano) puede mejorar significativamente el diagnóstico temprano y el monitoreo individualizado del deterioro cognitivo.

Nuestro principal objetivo en el trabajo que se expone en este capítulo fue investigar si los marcadores digitales derivados de nuestro Trail Making Test computarizado (cTMT) pueden apoyar la detección temprana del DCL (Juantorena et al., 2026). Buscamos capturar diferencias cognitivas sutiles combinando las medidas clásicas de tiempo de finalización con características detalladas del movimiento de la mano, como desviaciones de trayectoria, dinámicas de movimiento y cinemática del cursor. Evaluamos el valor diagnóstico de estas características digitales comparando su desempeño predictivo con el de la información demográfica y las pruebas neuropsicológicas estándar. Adicionalmente, examinamos si las características del cTMT pueden predecir las puntuaciones de los participantes en tareas neuropsicológicas clásicas (no digitales) para evaluar en qué medida los biomarcadores digitales pueden aproximarse a las evaluaciones clínicas establecidas. En última instancia, buscamos demostrar que los marcadores digitales obtenidos con hardware de uso masivo (una computadora estándar y un *mouse*) pueden complementar o reemplazar parcialmente las evaluaciones tradicionales realizadas por profesionales de la salud, ofreciendo una herramienta escalable y de bajo costo para el *screening* cognitivo. A diferencia de las implementaciones basadas en tablets (Lara-Garduno et al., 2022) o en software propietario (Hanczár et al., 2022), nuestro enfoque utiliza hardware estándar sin necesidad de equipamiento especializado ni licencias. Además, durante este trabajo desarrollamos *NeuroTask* (ver **Apéndice A.4**), una librería de *Python* de código abierto para la extracción de un conjunto amplio de características basadas en trayectorias del mouse. Hasta donde sabemos, esta combinación de hardware accesible, extracción sistemática de características multidimensionales y herramientas abiertas y transparentes no había sido aplicada previamente a la evaluación digital del TMT basada en *mouse*.

Métodos

Participantes

Esta muestra es una versión ampliada de la utilizada en el **Capítulo 2** e incluye a los mismos participantes junto con casos adicionales reclutados posteriormente. Se reclutó un total de 85 participantes, 48 con Deterioro Cognitivo Leve (DCL) y 37 Controles Sanos (CS). Tras aplicar los criterios de exclusión, 74 participantes (41 DCL) permanecieron en el análisis final. De los

participantes excluidos, seis no completaron al menos un ensayo de cada tipo (A o B), y cinco reportaron dificultades para usar el mouse y no pudieron conectar al menos cuatro ítems durante los ensayos de práctica; en todos los casos estas dificultades se debieron a falta de experiencia con la computadora y no a discapacidades motoras.

Todos los participantes poseían visión normal o corregida a normal y realizaron la tarea utilizando un mouse con su mano dominante. El grupo DCL fue reclutado entre los pacientes de los profesionales involucrados en el estudio. El grupo CS fue reclutado entre los familiares, acompañantes y trabajadores del Hospital Italiano de Buenos Aires. Se excluyeron individuos con diagnósticos previos de condiciones neurológicas (ej. enfermedad de Parkinson) o vasculares. Se obtuvo el consentimiento informado por escrito de todos los participantes y el protocolo del estudio fue aprobado por el Comité de Ética del Hospital Italiano de Buenos Aires (ID: 11543) de conformidad con la Declaración de Helsinki.

Los pacientes fueron diagnosticados con DCL siguiendo los criterios descritos en el **Capítulo 2** y basados en Jak y colaboradores (2009): entrevista clínica, reporte de un familiar o cuidador y evaluación neuropsicológica con rendimiento de -1.5 desvíos estándar en al menos dos pruebas de un dominio cognitivo (atención, memoria, lenguaje, funciones ejecutivas o habilidades visuoespaciales), con actividades de la vida diaria preservadas. Se utilizó un criterio conservador para determinar el compromiso de cada dominio cognitivo, identificando el patrón neuropsicológico como uni- o multi-dominio.

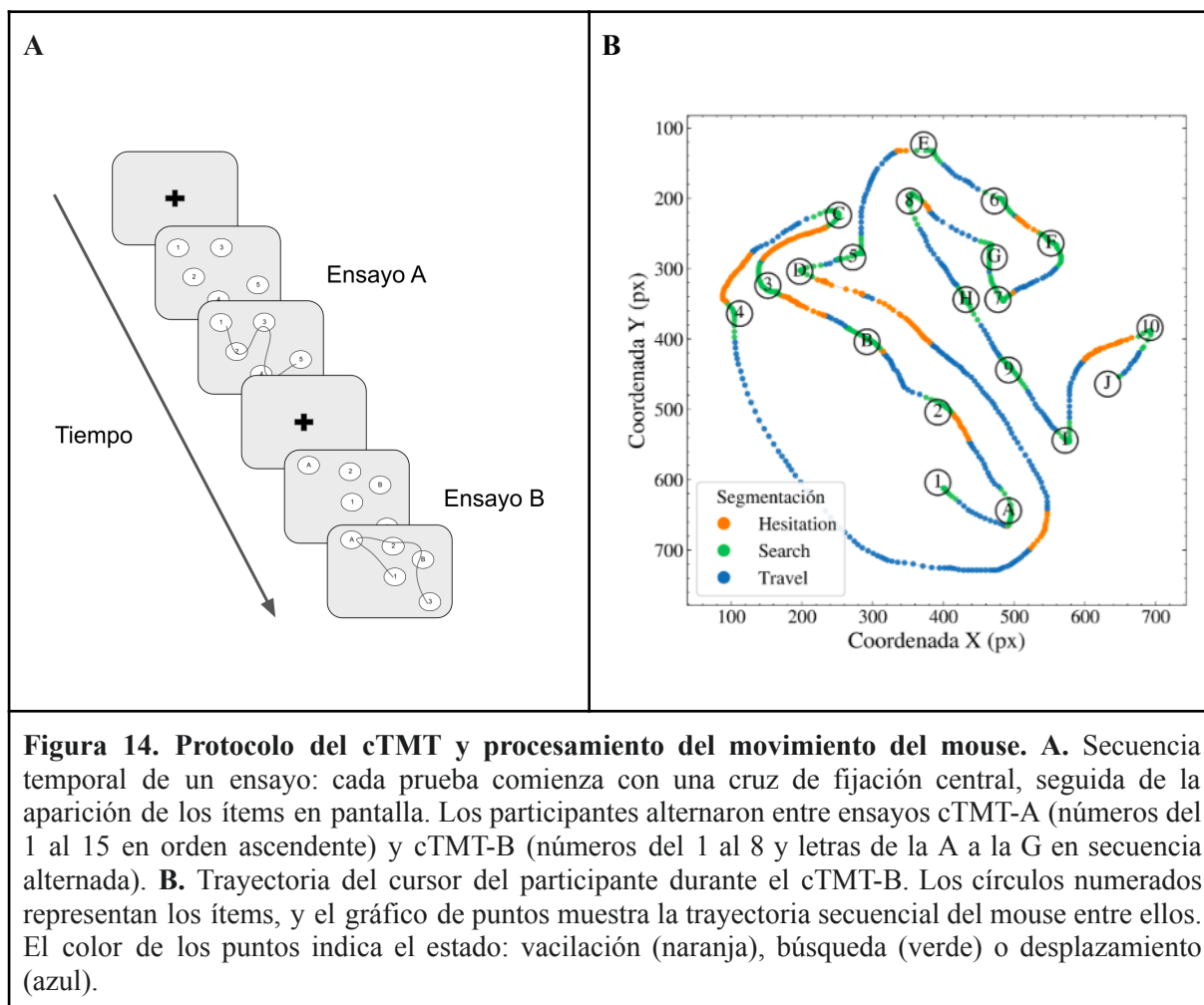
Además de la prueba digital, los participantes completaron una serie de tareas utilizadas en su evaluación diagnóstica, administradas dentro de las dos semanas previas o posteriores a la evaluación con el cTMT. Estas fueron: (1) Mini-Mental State Examination (MMSE), (2) Trail Making Test parte A (TMT-A) (3) Trail Making Test parte B (TMT-B), (4) Test de Símbolos y Dígitos, (5) Span de Dígitos Directo (6) Span de dígitos Inverso y (7) Test del Dibujo del Reloj. No se observaron diferencias demográficas significativas entre los grupos (Tabla 1).

Tabla 1. Datos demográficos. La prueba exacta de Fisher no indicó asociación entre grupo y sexo. Las pruebas *t* para muestras independientes no revelaron diferencias significativas en edad o años de educación.

	Controles Sanos (CS)	Deterioro Cognitivo Leve (DCL)	p-valor
Edad	73 ± 8	76 ± 7	0.078
Hombres / Mujeres	8 / 25	11 / 30	1.000
Años de educación	14 ± 3	14 ± 4	0.683

TMT Computarizado (cTMT)

Utilizamos el mismo protocolo cTMT descrito en el **Capítulo 2**, implementado en *Python* (v3.10.16) y *PsychoPy* (v2022.1.1; Peirce, 2007), sin registro de movimientos oculares. Los movimientos del mouse se registraron continuamente como coordenadas *x* e *y*, junto con sus marcas de tiempo (*t*), permitiendo la reconstrucción de la trayectoria completa (Fig. 14). El conjunto de datos resultante está disponible en el **Apéndice B.1**.



Extracción de características de las mediciones de la mano

Extrajimos mediciones de la mano a partir de series temporales del movimiento del mouse utilizando *NeuroTask* (<https://github.com/NeuroLIAA/neurotask>), una librería de *Python* que desarrollamos para este propósito, que permite extraer múltiples métricas a partir de las series temporales del cursor, favoreciendo la replicabilidad de los análisis.

Primero, obtuvimos medidas clásicas del TMT como el tiempo total, número de ítems alcanzados correcta e incorrectamente (respecto a la secuencia esperada) y cruces sobre trazos previos. A continuación, estimamos el grado de desviación respecto a la trayectoria ideal (definida como la conexión de los centros de cada ítem con líneas rectas). De este punto de referencia derivamos métricas como la diferencia acumulada y el área entre la trayectoria real e ideal. Además, incorporamos métricas cinemáticas, incluyendo velocidad promedio, velocidad máxima y características derivadas de la aceleración (por ejemplo, amplitud de zigzag, definida como la diferencia de tiempo media entre transiciones consecutivas de números y letras en la parte B). También incluimos otras medidas presentes en la literatura previa, como el tiempo dentro de cada ítem (intra-ítem o dwell time) y entre ítems (inter-ítem) (Thornton & Horowitz, 2020).

Inspirándonos en el trabajo de Lara-Garduno y colaboradores (2022), segmentamos cada prueba en diferentes estados para representar las acciones del sujeto (Fig. 14B). La asignación de estado se realizó punto a punto a lo largo de la trayectoria mediante un algoritmo basado en reglas con persistencia de estado: una transición entre estados se confirmaba solo si al menos cinco puntos consecutivos satisfacían la condición del nuevo estado, evitando así que fluctuaciones breves de velocidad generaran transiciones espurias. El estado de búsqueda se refiere a los periodos en los que el participante busca el siguiente ítem. Esto incluye puntos dentro de un ítem o inmediatamente después de abandonarlo. Sin embargo, la velocidad no puede superar un umbral de velocidad específico para cada sujeto, que se calculó como la mediana de las velocidades promedio registradas en el segmento del primer al segundo ítem en todos los ensayos válidos de la Parte A para ese participante. Usar la mediana entre ensayos hace al umbral robusto ante valores atípicos. Esta condición evita clasificar puntos fuera del ítem debido a movimientos involuntarios o proximidad. El estado de desplazamiento describe cuando el sujeto se mueve de forma decidida en una dirección determinada, incluyendo puntos que no cumplen la condición de búsqueda y superan el umbral de velocidad. La vacilación representa momentos de duda, corrección de la trayectoria o pérdida temporal del ítem, incluidos los segmentos inmediatamente posteriores a un estado de desplazamiento en los que el sujeto reduce la velocidad por debajo del umbral. Para cada estado, extrajimos métricas específicas, como la distancia recorrida, el tiempo transcurrido y la velocidad media, lo que permitió una caracterización más detallada de la resolución del sujeto.

Por último, consideramos válido un ensayo aunque no se alcanzara el objetivo final. Para garantizar la comparabilidad entre los participantes, establecimos un umbral de ocho ítems (más del cincuenta por ciento de la secuencia) y utilizamos solo la trayectoria hasta el último ítem alcanzado correctamente para calcularlo. Esto dio lugar a dos conjuntos de características: uno restringido a la trayectoria hasta el último objetivo alcanzado correctamente y otro sin esta limitación (variables etiquetadas con el prefijo “Completo” en la Tabla C3.S1).

Para obtener un único vector de características por participante apto para el entrenamiento de los modelos, las mismas se promediaron entre todos los ensayos válidos del mismo tipo (Parte A o Parte B) por separado, usando la media como función de agregación. Adicionalmente, se calcularon características del ratio (Parte B / Parte A) a partir de estos valores promediados, capturando el incremento relativo en la demanda cognitiva entre ambas partes. Esto resultó en una fila por participante, alineada con la información clínica y demográfica a nivel de sujeto.

Esquema de aprendizaje automático

Implementamos un esquema de validación cruzada anidada (nested cross-validation) con el fin de obtener estimaciones del desempeño casi insesgadas para todos los modelos (Varma & Simon, 2006) (Fig. 15). En el bucle externo, utilizamos una validación cruzada tipo Leave-One-Out (*LOOCV*), reservando en cada iteración a un sujeto como conjunto de prueba. Es importante señalar que todas las etapas de preprocesamiento se realizaron exclusivamente sobre los datos de entrenamiento, para prevenir la fuga de información o *data leakage* (Sasse et al., 2025). Estas etapas incluyeron la estandarización y la selección univariada de características mediante *SelectKBest* (scikit-learn; Varoquaux et al., 2015), reteniendo las mejores $k \leq 20$ características. La función de puntuación se eligió según el tipo de variable objetivo: el test F de ANOVA (*f_classif*) para la clasificación del grupo diagnóstico y el test F de regresión (*f_regression*). Cuando había menos de 20 características disponibles, se incluyeron todas. Los pasos se implementaron como un único objeto

Pipeline de scikit-learn, garantizando que ninguna información del sujeto de prueba influyera en ninguna decisión de preprocesamiento.

Dentro del bucle interno, llevamos a cabo una validación cruzada estratificada de 10 *folds* (10-fold CV) para el ajuste de hiperparámetros y la selección de modelos. La optimización de los mismos se ejecutó mediante grid search, utilizando el área bajo la curva ROC (ROC-AUC) para las tareas de clasificación y el error absoluto medio (MAE, por sus siglas en inglés) para las de regresión (Poldrack et al., 2020). Una vez seleccionados los parámetros óptimos, el modelo se re-entrenó con la partición de entrenamiento externa completa y se evaluó en el sujeto reservado; finalmente, las predicciones se agregaron a través de los *folds* para calcular la estimación del desempeño global (Fig. 15). Esta metodología es fundamental para obtener estimaciones de desempeño no sesgadas, ya que separa estrictamente los datos utilizados para la selección del modelo de los empleados para la evaluación del desempeño (Raschka, 2020).

Dado que los teoremas de no-free-lunch (Wolpert, 1996) establecen que ningún algoritmo es consistentemente superior a los demás en todas las situaciones, evaluamos varios modelos alternativos, incluyendo *Random Forest*, *Support Vector Machines* (SVM), Regresión Logística y *XGBoost*. Las grillas de hiperparámetros completas para todos los modelos están provistas en la Tabla C3.S2 de los materiales suplementarios.

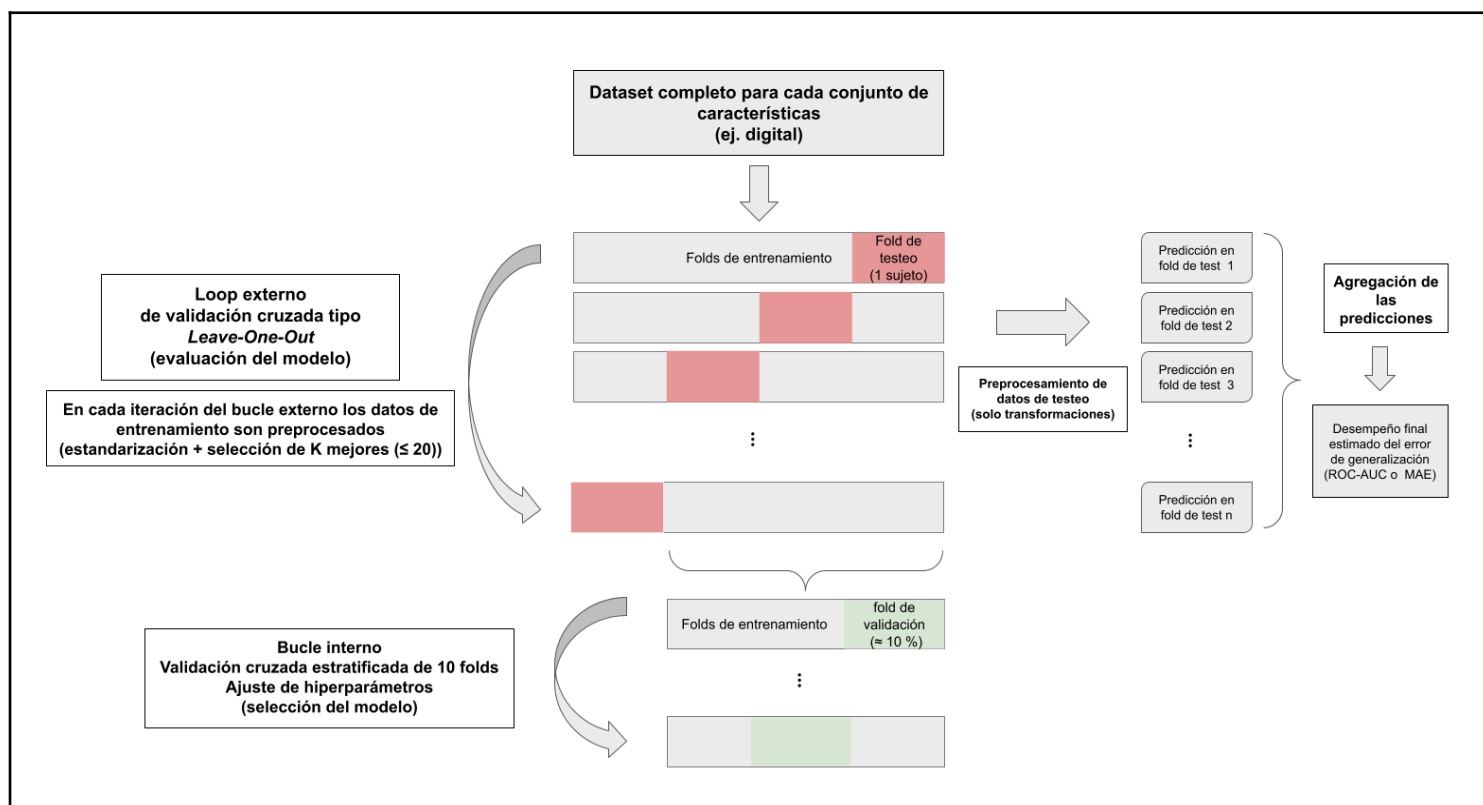


Figura 15. Esquema de aprendizaje automático. Representación del procedimiento de validación cruzada anidada utilizado. El bucle externo (Leave-One-Out CV) se utilizó para estimar el desempeño de generalización del modelo. En cada iteración, se excluyó a un sujeto como fold de prueba, mientras que los sujetos restantes formaron los *folds* de entrenamiento. El preprocesamiento, incluida la estandarización y la selección de características univariantes con *SelectKBest* (≤ 20 características), se realizó exclusivamente en los datos de entrenamiento y luego se aplicó al fold de prueba para prevenir la fuga de información o data leakage. Dentro de cada conjunto de entrenamiento externo, se realizó una validación cruzada estratificada de 10 *folds* (bucle interno) para el ajuste de hiperparámetros y la selección de modelos. Las predicciones para los sujetos excluidos se agregaron a través de los *folds* externos para calcular las estimaciones de desempeño finales, utilizando ROC-AUC para la clasificación y MAE para la regresión.

Importancia de las Características: Enfoque SHAP

Para interpretar los modelos, utilizamos SHAP (*SHapley Additive exPlanations*; Lundberg & Lee, 2017). Esta técnica, fundamentada en la teoría de juegos cooperativos, permite asignar a cada característica un valor que representa su contribución marginal a la predicción final, ofreciendo así una explicación tanto local como global del comportamiento del modelo. Los valores de Shapley se calcularon para el sujeto excluido en cada fold de *LOOCV*. En cada fold, el modelo seleccionado con sus hiperparámetros optimizados fue reajustado sobre el conjunto de entrenamiento, y los valores SHAP se calcularon para el sujeto excluido utilizando dicho conjunto como fondo (*background*). En las tareas de clasificación, las explicaciones se obtuvieron a partir de la probabilidad predicha de la clase positiva; en las de regresión, a partir de la predicción cruda. Dado que la selección de características (*SelectKBest*) podía dar lugar a la inclusión de diferentes subconjuntos de características en diferentes *folds*, agregamos los valores promediando la contribución absoluta de cada característica en todos los *folds* en los que aparecía. Este procedimiento arrojó puntuaciones de

importancia global que reflejan tanto la magnitud de la contribución de una característica como su estabilidad en los *folds*.

Modelos de Referencia

Establecimos dos líneas base para contextualizar el desempeño. Por un lado, los modelos demográficos funcionaron como límite inferior al ser entrenados exclusivamente con las variables sexo, edad y años de educación. Estas variables suelen estar asociadas al rendimiento cognitivo y normalmente se incluyen en las correcciones normativas de las puntuaciones de las pruebas neuropsicológicas. Aunque no son marcadores específicos de deterioro cognitivo, proporcionan un punto de referencia útil para evaluar el valor añadido de las características digitales derivadas de la tarea. Comparar el desempeño de este modelo con el de los modelos entrenados con características digitales permite cuantificar el valor discriminativo adicional del cTMT más allá de la información demográfica disponible de forma rutinaria.

Por otro lado, desarrollamos modelos con las puntuaciones de los tests neuropsicológicos que sirvieron como límite superior conservador. Para establecer esta referencia, entrenamos modelos utilizando las mismas evaluaciones estandarizadas empleadas para el diagnóstico clínico original del DCL, tal como se detalla en la sección *Participantes*. Estas medidas incluyeron el Mini-Mental State Examination (MMSE), las pruebas Trail Making A y B, el Test de Símbolos y Dígitos, el Span de Dígitos (directo e inverso) y el Test del Dibujo del Reloj. Si bien en la práctica clínica estas pruebas se complementan con entrevistas cualitativas y reportes de informantes, que contribuyeron a la decisión diagnóstica original junto con los puntajes cuantitativos, aquí se utilizaron estrictamente las puntuaciones estandarizadas, lo que hace que este modelo represente un límite superior conservador de lo que la batería cuantificada puede lograr por sí sola. Se esperaba que este modelo alcanzara el mayor desempeño de clasificación, funcionando como techo contra el cual contrastar la contribución de los biomarcadores digitales.

Por último, entrenamos modelos utilizando las características digitales extraídas del cTMT mediante *NeuroTask*. Estos modelos constituyen el aporte central del trabajo y fueron comparados tanto con el límite inferior demográfico como con el límite superior neuropsicológico. Las variables extraídas incluyeron medidas clásicas del TMT (tiempo total, número de toques correctos e incorrectos, cruces sobre trazos previos), métricas de desviación espacial (distancia acumulada y área entre la trayectoria real e ideal), medidas cinemáticas (velocidad promedio, velocidad máxima e índices basados en la aceleración) y características derivadas de la segmentación por estados conductuales (búsqueda, desplazamiento y vacilación), identificados en cada ensayo (Fig. 14B). Para cada estado se calcularon medidas de distancia recorrida, tiempo transcurrido y velocidad media, permitiendo una caracterización detallada del comportamiento durante la tarea. La lista completa de características extraídas se presenta en la Tabla C3.S1.

Resultados

Análisis Exploratorio

Comenzamos dividiendo los datos en tres conjuntos de características: demográficas, no digitales (batería clínica) y digitales (~100 características extraídas del cTMT; Tabla C3.S1). Para las

características digitales, los valores de cada ensayo se promediaron entre los ensayos válidos del mismo tipo (Parte A o Parte B) para obtener un único valor por participante, tal como se describe en la sección de *Extracción de Características*. A su vez, como se reporta en la sección **Métodos**, los grupos DCL y CS no difirieron significativamente en edad, sexo o años de educación. En contraste, surgieron diferencias significativas en las pruebas no digitales realizadas por neuropsicólogos para el diagnóstico. Se observaron diferencias significativas en el rendimiento en varias variables, evaluadas mediante pruebas t de muestras independientes. Las puntuaciones del MMSE fueron más bajas en el grupo DCL ($t = 4.26$, $p < 0.001$), consistente con el deterioro cognitivo global. De manera similar, los participantes con DCL tuvieron tiempos de finalización de TMT-A ($t = -3.98$, $p < 0.001$) y TMT-B ($t = -5.37$, $p < 0.001$) significativamente más largos, reflejando una velocidad de procesamiento y funcionamiento ejecutivo reducidos. Se observaron diferencias adicionales en el Test de Símbolos y Dígitos ($t = 2.99$, $p = 0.004$), el Test de Span de Dígitos Inverso ($t = 2.61$, $p = 0.011$) y el Test del Dibujo del Reloj ($t = 2.50$, $p = 0.015$), indicando un peor rendimiento en el grupo DCL. Estos resultados confirman que los dos grupos difirieron consistentemente a través de la batería de referencia, proporcionando un punto de referencia válido para evaluar las medidas digitales en nuestro estudio.

Las características extraídas del cTMT también revelaron varias diferencias grupales significativas. En la Parte A, los participantes con DCL mostraron mayores desviaciones de la trayectoria ideal (*Area difference from ideal*, $t = -2.12$, $p = 0.038$), tiempos de vacilación más largos (*Hesitation time*, $t = -2.05$, $p = 0.045$) y tiempos intra-ítem más largos (*Intra-target time*, $t = -2.35$, $p = 0.022$). En la Parte B, los participantes con DCL exhibieron menos toques correctos completos (*Complete correct touches*, $t = 3.36$, $p = 0.001$), mayor tiempo completo en el ensayo (*Complete time in trial*, $t = -2.21$, $p = 0.03$) y menor ratio B/A de toques correctos completos (*Complete correct touches*, ratio B/A, $t = 2.23$, $p = 0.029$). Adicionalmente, el número total de ensayos correctos de tipo B fue significativamente mayor en el grupo CS ($t = 2.95$, $p = 0.004$). Estos resultados exploratorios sugieren que, más allá del tiempo de finalización bruto, los marcadores digitales de grano fino del cTMT pueden capturar diferencias entre grupos.

Clasificación del grupo diagnóstico

Entrenamos cuatro algoritmos de aprendizaje automático para predecir el grupo diagnóstico de los participantes: *Random Forest* (RF), *Support Vector Machine* (SVM), *Regresión Logística* (LR) y *XGBoost* (XGB). Cada algoritmo fue entrenado y evaluado utilizando cinco conjuntos de características: variables demográficas, características digitales derivadas del cTMT, medidas neuropsicológicas no digitales y dos conjuntos combinados (demográfico más digital y demográfico más no digital).

Para cada conjunto de características, seleccionamos el algoritmo con mayor AUC bajo LOOCV como mejor modelo, y graficamos las curvas ROC correspondientes (Fig. 16A). El AUC fue seleccionado como métrica principal, ya que es independiente del umbral de decisión, midiendo la capacidad global del modelo para ordenar los casos DCL por encima de los CS. Por esta razón los hiperparámetros fueron optimizados maximizando esta métrica en el bucle interno de validación cruzada. Esto es particularmente apropiado en un contexto de *screening*, donde el umbral óptimo depende del costo relativo de falsos negativos y falsos positivos y puede variar según el contexto clínico. Dado que solo se evaluó la significancia del mejor modelo por conjunto de características, se aplicó la corrección de Bonferroni entre los cinco conjuntos (umbral = 0.010). La significancia se

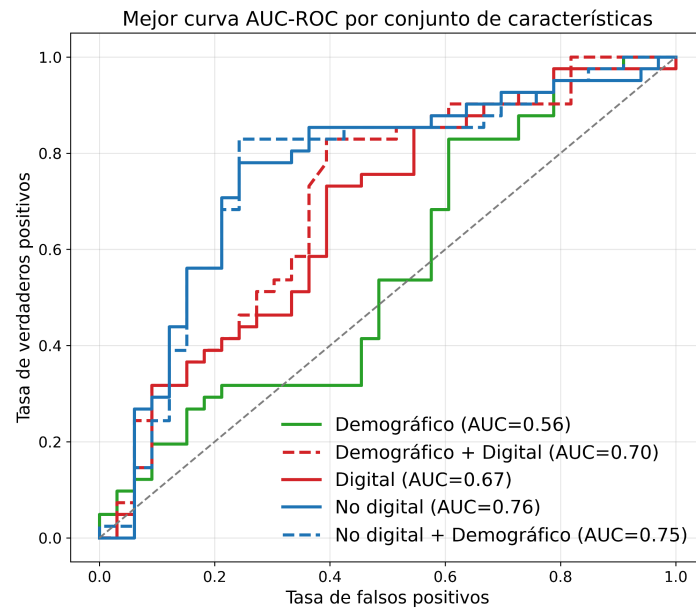
evaluó mediante pruebas de permutación no paramétricas con 1000 permutaciones de etiquetas; el p-valor empírico se definió como la proporción de permutaciones que arrojaron un AUC mayor o igual al observado. Los intervalos de confianza (IC) del 95% del AUC se estimaron mediante bootstrap con $N = 10000$ muestras, un número mayor que el de las permutaciones para asegurar estimaciones estables de los intervalos. La Tabla 2 reporta el AUC de los cuatro clasificadores para los cinco conjuntos de características. El *balanced accuracy* (media aritmética de sensibilidad y especificidad) se reporta junto al AUC para proporcionar una métrica extra teniendo en consideración el leve desbalance de clases de la muestra (41 DCL, 33 CS).

En línea con lo esperado, las variables demográficas por sí solas proporcionaron solo una discriminación modesta (RF, AUC = 0.56, IC 95% [0.42, 0.69], *balanced accuracy* = 0.60, $p = 0.189$), sin superar significativamente el nivel del azar. En contraste, las características digitales derivadas del cTMT mostraron una mejora clara (SVM, AUC = 0.67, IC 95% [0.54, 0.79], *balanced accuracy* = 0.66, $p = 0.007$). Combinar la demografía con características digitales aumentó aún más el desempeño (SVM, AUC = 0.70, IC 95% [0.57, 0.82], *balanced accuracy* = 0.69, $p = 0.002$), destacando el valor añadido del cTMT más allá de la información demográfica sola. Las medidas neuropsicológicas no digitales lograron el mayor desempeño discriminativo (RF, AUC = 0.76, IC 95% [0.64, 0.87], *balanced accuracy* = 0.77, $p = 0.001$), sin que la adición de datos demográficos aportara beneficio adicional (RF, AUC = 0.75, IC 95% [0.63, 0.87], *balanced accuracy* = 0.76, $p = 0.001$). Los cuatro conjuntos de características significativos superaron la corrección de Bonferroni.

En conjunto, estos resultados apoyan la interpretación de que las características digitales, cuando se integran con datos demográficos, mejoran el desempeño de clasificación más allá de la información demográfica, produciendo niveles de discriminación comparables a los obtenidos con evaluaciones neuropsicológicas tradicionales.

La importancia global de los predictores se expresó mediante valores SHAP absolutos medios a través de todos los *folds* (Fig. 16B). En resumen, estos valores se calcularon para cada sujeto excluido utilizando el modelo reajustado sobre el conjunto de entrenamiento de dicho fold; posteriormente, se agregaron promediando las contribuciones absolutas de cada variable en aquellas iteraciones donde fue retenida tras la selección de atributos. Estas métricas cuantifican el peso de cada elemento en las predicciones del modelo, donde valores absolutos más altos indican una mayor influencia en la probabilidad de clasificar a un sujeto como positivo (DCL). Cabe destacar que la edad, proveniente del set demográfico, se ubicó en la duodécima posición, mostrando que si bien la información demográfica es relevante, los marcadores digitales tienen un peso mayor en ese subconjunto. Los marcadores más influyentes, según sus altos valores SHAP, fueron: el número de ensayos con al menos ocho objetivos conectados correctamente (*Valid sum*) tanto para la parte A como para la B; el número total de toques correctos (*Complete correct touches*) y la proporción de las duraciones máximas de vacilación en ambas partes (*Hesitation max duration ratio*, B/A). Otro factor relevante fueron las transiciones de estado, que cuantifican la frecuencia con la que el cursor alterna entre vacilación, búsqueda y desplazamiento (*State transitions*).

A.



B.

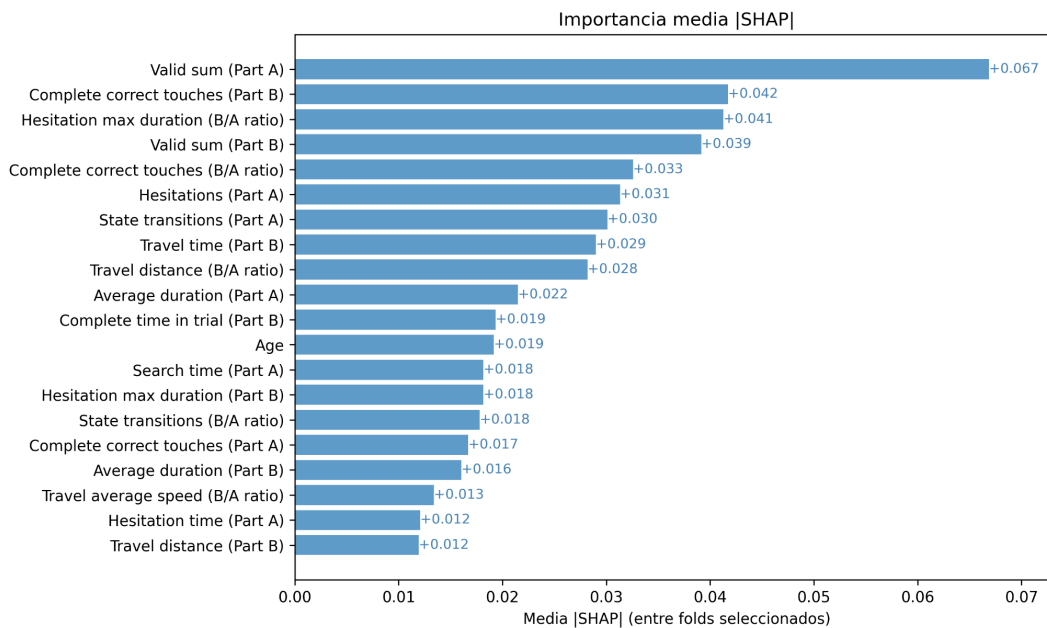


Figura 16. Desempeño de la clasificación del grupo de diagnóstico e importancia de las características. A. Curvas ROC en la muestra completa ($N = 74$, 41 MCI). Se indican los valores del área bajo la curva (AUC) para el modelo con mejor desempeño dentro de cada conjunto de características. Los datos demográficos por sí solos proporcionaron solo una discriminación modesta (RF, AUC = 0.56, $p = 0.189$). Las características digitales del cTMT mejoraron el desempeño (SVM, AUC = 0.67, $p = 0.007$), y la combinación de los datos demográficos con las características digitales aumentó aún más la precisión (SVM, AUC = 0.70, $p = 0.002$). Las medidas neuropsicológicas no digitales arrojaron el mayor valor discriminatorio (RF, AUC = 0.76, $p = 0.001$), sin que los datos demográficos aportaran ningún beneficio adicional (RF, AUC = 0.75, $p = 0.001$). **B.** Valores SHAP absolutos medios (media|SHAP|) para el modelo de clasificación. Las puntuaciones de importancia se promediaron entre *folds*, considerando únicamente aquellos en que cada característica fue seleccionada por *SelectKBest*.

Tabla 2. Desempeño de clasificación por conjunto de características y algoritmos. Para cada conjunto de características, el mejor modelo (según AUC más alto) se indica en negrita; los valores *p* provienen de pruebas de permutación (1.000 permutaciones de etiquetas) aplicadas únicamente al mejor modelo. Los IC 95% del AUC son bootstrap (N = 10000). La precisión balanceada (*Balanced accuracy*) se calculó utilizando un umbral para las probabilidades predichas en 0.5.

Conjunto de Características	Modelo	AUC	IC 95% (AUC)	p-valor (AUC)	Balanced accuracy
Demográfico	RF	0.56	[0.42, 0.69]	0.189	0.60
	SVC	0.31	[0.18, 0.44]	-	0.53
	LR	0.46	[0.32, 0.61]	-	0.54
	XGB	0.56	[0.42, 0.69]	-	0.48
Digital	RF	0.63	[0.50, 0.75]	-	0.61
	SVC	0.67	[0.54, 0.79]	0.007	0.66
	LR	0.62	[0.49, 0.75]	-	0.62
	XGB	0.57	[0.43, 0.70]	-	0.54
Demográfico + Digital	RF	0.64	[0.50, 0.77]	-	0.60
	SVC	0.70	[0.57, 0.82]	0.002	0.69
	LR	0.64	[0.51, 0.77]	-	0.62
	XGB	0.58	[0.45, 0.71]	-	0.56
No-digital	RF	0.76	[0.64, 0.87]	0.001	0.77
	SVC	0.75	[0.62, 0.86]	-	0.67
	LR	0.74	[0.61, 0.85]	-	0.67
	XGB	0.65	[0.51, 0.78]	-	0.67
No-digital + Demográfico	RF	0.75	[0.63, 0.87]	0.001	0.76
	SVC	0.75	[0.63, 0.86]	-	0.71
	LR	0.75	[0.63, 0.86]	-	0.69
	XGB	0.62	[0.49, 0.75]	-	0.60

Predicción de Puntuaciones de Pruebas Neuropsicológicas

De manera consistente con la sección de clasificación, entrenamos siete algoritmos de aprendizaje automático para predecir las puntuaciones de las tareas neuropsicológicas: *Regresión Lineal* (LR), *Ridge Regression* (RR), *Lasso*, *Elastic Net* (EN), *Support Vector Regression* (SVR), *Random Forest Regression* (RFR) y *XGBoost Regression* (XGBR). Cada algoritmo fue entrenado utilizando el conjunto de características combinado (demográfico más digital) para predecir una puntuación objetivo a la vez. La elección de este conjunto se debe a que fue el de mejor desempeño en clasificación entre los que incluyen características digitales. Se siguió el mismo esquema que en la clasificación, incluyendo validación cruzada anidada, optimización de hiperparámetros mediante grid search (Fig. 15) y pruebas de permutación para evaluar la significancia. Las grillas completas de hiperparámetros se proveen en la Tabla C3.S3 de los materiales suplementarios. Evaluamos la capacidad del modelo para predecir siete puntuaciones objetivo: MMSE, TMT-A, TMT-B, Test de Símbolos y Dígitos, Span de Dígitos (Directo e Inverso) y Test del Dibujo del Reloj.

Obtuvimos predicciones estadísticamente significativas para cuatro de las siete puntuaciones evaluadas tras la corrección de Bonferroni (7 comparaciones, umbral = 0.007). Para el MMSE, el mejor desempeño se logró con un regresor *Random Forest* (MAE = 1.45, $p = 0.004$). El Test de Símbolos y Dígitos fue predicho con menor error por un modelo Lasso (MAE = 9.47, $p = 0.001$). Asimismo, tanto el TMT-A como el TMT-B mostraron predicciones significativas, con valores de MAE de 11.95 (*Elastic Net*, $p = 0.001$) y 43.09 (*Support Vector Regressor*, $p = 0.001$), respectivamente. El Span de Dígitos Directo mostró un resultado nominalmente significativo que no alcanzó el umbral de corrección (*Elastic Net*, MAE = 0.85, $p = 0.027$). Por el contrario, los modelos para el Span de Dígitos Inverso y el Test del Dibujo del Reloj no alcanzaron significancia estadística, con valores p de 0.419 y 0.837, respectivamente.

Para poner en contexto estas predicciones, comparamos el Error Absoluto Medio (MAE) del modelo con la desviación estándar (DE) de la variable objetivo y su rango intercuartil (IQR) (Tabla 3). Como índice resumen de esta comparación, calculamos el error normalizado (MAE/DE): valores menores a 1 indican que el modelo predice con un error menor que la variabilidad natural de la medida. En todos los casos donde la predicción fue significativa, el MAE fue considerablemente menor que la DE de los valores objetivo (por ejemplo, para Símbolos y Dígitos: MAE = 9.47 vs DE = 13.7; para TMT-A: MAE = 11.95 vs DE = 17.2). Es necesario realizar una aclaración respecto al Test del Dibujo del Reloj, el cual presentó un comportamiento estadístico particular. Observamos que esta prueba sufre de una falta crítica de variabilidad entre el primer y el tercer cuartil, resultando en un IQR de cero. Esto se debe a una distribución de puntuaciones extremadamente sesgada hacia el rendimiento óptimo: el 76% de los participantes obtuvo la puntuación máxima de 3 (en una escala de 1 a 3). Esta homogeneidad (“efecto techo”) limita la capacidad de cualquier modelo de regresión para aprender patrones distintivos, ya que la varianza a explicar es mínima. En conjunto, estos resultados sugieren que nuestros modelos son capaces de capturar patrones informativos en la mayoría de las medidas neuropsicológicas, siendo particularmente precisos en aquellas pruebas que poseen suficiente variabilidad intrínseca para permitir la diferenciación entre sujetos.

Tabla 3. Predicción de pruebas individuales (objetivo) en el conjunto de características demográficas + digitales. RFR: regresor de random forest, SVR: regresor de vector de soporte, DE e IQR: desviación estándar y rango intercuartil de los valores objetivo, MAE: error absoluto medio de los valores predichos frente a los valores objetivo, IC 95%: intervalo de confianza del 95% del MAE (entre corchetes, estimado por bootstrap), Error normalizado: estimado como el *ratio* entre el MAE y la DE, valor p estimado a partir de la prueba de permutación (véase **Métodos**).

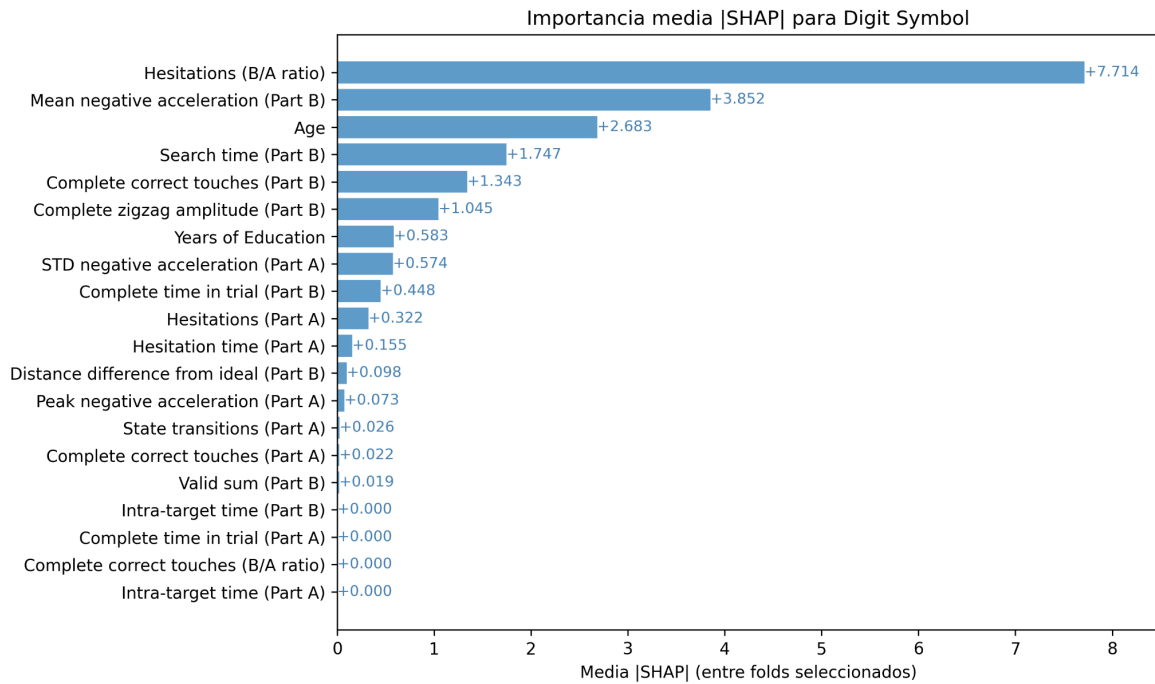
Variable objetivo	Mejor modelo	Valores de la variable objetivo		Métricas de predicción		
		DE	IQR	MAE [IC 95%]	Error normalizado (MAE/DE)	p-valor
MMSE [0, 30]	RFR	1.9	2.0	1.45 [1.20, 1.74]	0.76	0.004
Símbolos y Dígitos [0, 135]	Lasso	13.7	16.8	9.47 [7.78, 11.34]	0.69	0.001
TMT-A [0, 300] s	EN	17.2	20.1	11.95 [9.57, 14.64]	0.69	0.001
TMT-B [0, 300] s	SVR	64.0	69.8	43.09 [32.42, 55.40]	0.67	0.001
Span de Dígitos Directo [2, 9]	EN	1.0	1.0	0.85 [0.71, 1.00]	0.85	0.027
Span de Dígitos Inverso [2, 8]	SVR	0.8	1.8	0.62 [0.48,0.74]	0.78	0.419
Test del Reloj [1, 3]	SVR	0.5	0.0	0.29 [0.18,0.42]	0.58	0.837

De manera análoga al análisis de clasificación del grupo diagnóstico, estudiamos la importancia global de las características observando el valor absoluto medio de Shapley a través de los *folds* en los casos donde la predicción fue significativa (Fig. 17). Un hallazgo transversal fue la relevancia de la edad, que apareció dentro de las doce primeras posiciones en todos los casos, y se ubicó entre las tres primeras para el MMSE, Símbolos y Dígitos y TMT-A.

Para el Test de Símbolos y Dígitos, los predictores dominantes fueron el ratio B/A de vacilaciones y la aceleración negativa media (Fig. 17A). En el TMT-A, la edad fue el predictor más fuerte, seguida por el número de toques correctos en la parte A; las métricas de rendimiento y velocidad también contribuyeron, con características de vacilación mostrando influencia moderada (Fig. 17B). En el TMT-B, las características más relevantes provinieron de la propia parte B, con medidas de precisión y tiempo dominando y la distancia respecto a la trayectoria ideal entre las principales (Fig. 17C). Para el MMSE, la distribución de importancia fue más homogénea, con la edad y la proporción de transiciones de estado entre partes B y A mostrando los valores SHAP más altos, junto con medidas de velocidad, desempeño y tiempo de vacilación (Fig. 17D).

Cabe señalar que las predicciones del TMT-A y TMT-B están sujetas a una potencial circularidad, ya que estas variables objetivo comparten la misma tarea subyacente que las características digitales del cTMT. Sin embargo, esta preocupación no aplica a los tres objetivos restantes no relacionados con el TMT (MMSE, Símbolos y Dígitos y Span de Dígitos Directo), cuyas predicciones se basan en una relación independiente entre la dinámica del cursor y el rendimiento cognitivo. De estos, el MMSE y los Símbolos y Dígitos mostraron predicciones significativas tras la corrección de Bonferroni, mientras que el Span de Dígitos Directo obtuvo un resultado nominalmente significativo que no superó la corrección.

A.



B.

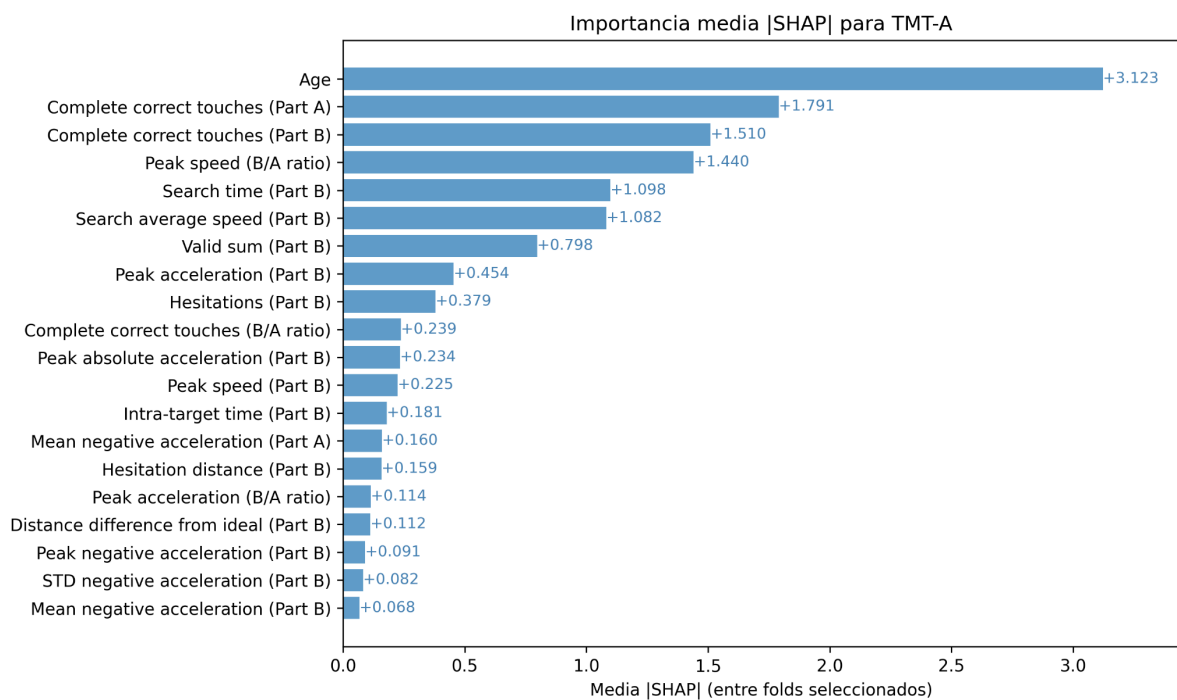
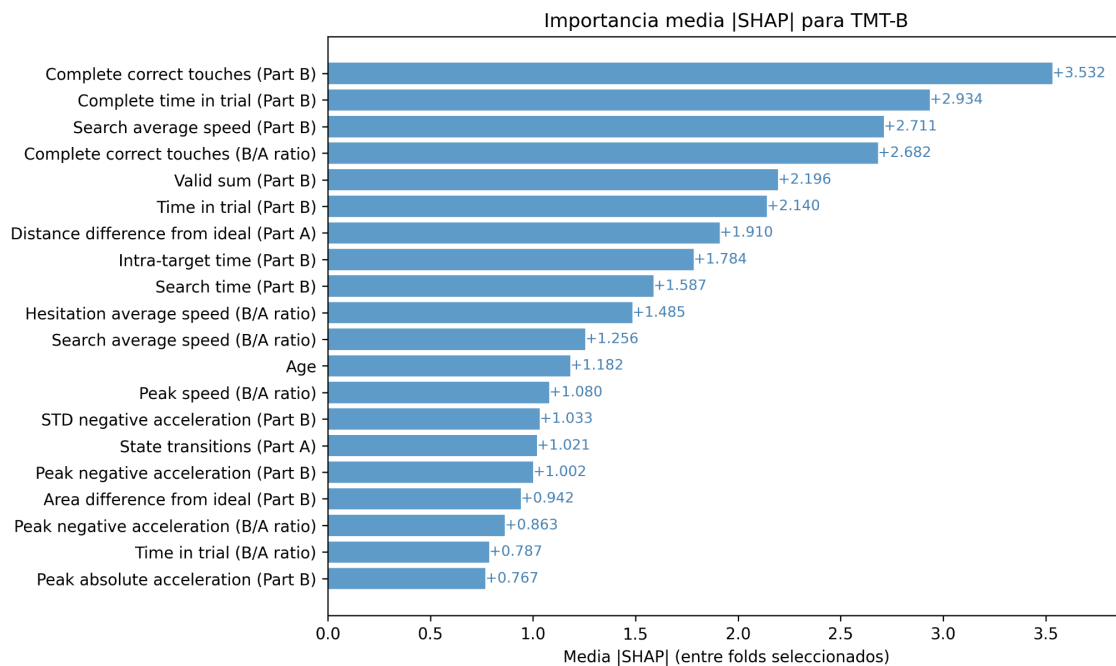


Figura 17. Valores SHAP absolutos medios (media |SHAP|) para los modelos de regresión significativos (ver Tabla 3). A. Símbolos y Dígitos. B. TMT-A. C. TMT-B. D. MMSE. Todos los modelos utilizaron el conjunto de características que combina características digitales y demográficas. Las puntuaciones de importancia se promediaron entre *folds*, considerando únicamente aquellos en que cada característica fue seleccionada por SelectKBest. (Continúa)

C.



D.

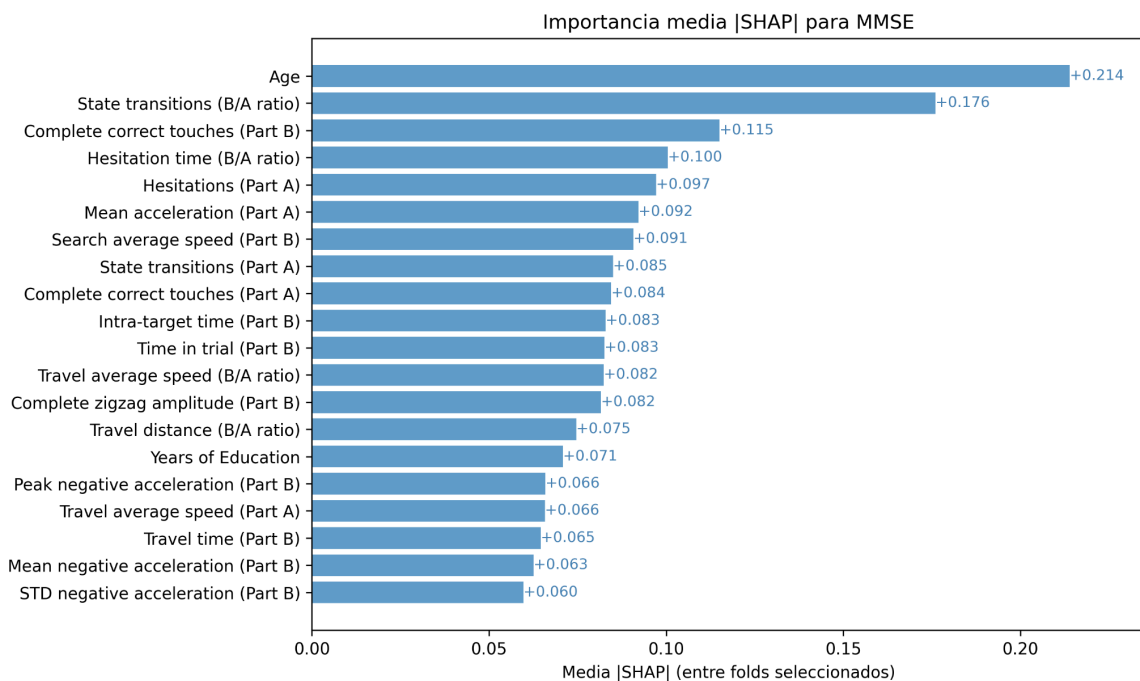


Figura 17 (Continuación). Valores SHAP absolutos medios (media |SHAP|) para los modelos de regresión significativos (ver Tabla 3). A. Símbolos y Dígitos. B. TMT-A. C. TMT-B. D. MMSE. Todos los modelos utilizaron el conjunto de características que combina características digitales y demográficas. Las puntuaciones de importancia se promediaron entre *folds*, considerando únicamente aquellos en que cada característica fue seleccionada por SelectKBest

Discusión

En este capítulo investigamos si las características detalladas del movimiento de la mano extraídas de un TMT computarizado (cTMT) podían discriminar entre personas con deterioro cognitivo leve (DCL) y controles cognitivamente sanos (CS). Para ello, incorporamos tres fuentes de información complementarias: (i) variables demográficas (edad, sexo, años de educación), (ii) evaluaciones neuropsicológicas estandarizadas utilizadas en el diagnóstico original de DCL, y (iii) características digitales de alta resolución derivadas de las trayectorias de la mano obtenidas a partir del *mouse*. Mediante la comparación sistemática de modelos que utilizaron estos tres conjuntos de características dentro de un proceso de aprendizaje automático de validación cruzada anidada, pudimos cuantificar el potencial valor diagnóstico añadido de los biomarcadores motores digitales en relación con cotas tanto mínimas (solo datos demográficos) como máximas (batería neuropsicológica completa).

Además, utilizamos el mismo esquema de aprendizaje automático para generar modelos de regresión de las puntuaciones de los tests neuropsicológicos estándar basándonos en las características extraídas del cTMT. Estos análisis revelaron que los patrones motores contenían información útil no solo para la clasificación, sino también para identificar asociaciones sistemáticas con la variabilidad individual en diferentes dominios cognitivos. Mediante el análisis SHAP, descubrimos que la edad se encontraba entre los tres principales predictores en los modelos MMSE, Digit Symbol y TMT-A, en consonancia con informes anteriores que relacionaban la edad con el rendimiento neuropsicológico general (Hamdan & Hamdan, 2009; Hoyer et al., 2004).

Los perfiles de importancia de características revelaron patrones específicos para cada prueba, interpretables en términos de sus demandas cognitivas. Para el Test de Símbolos y Dígitos, el tiempo de vacilación y la aceleración negativa media fueron los predictores dominantes, consistente con la dependencia de la tarea en la memoria de trabajo: las pausas prolongadas reflejan momentos en que los participantes escanean la pantalla por dificultad para recordar el siguiente ítem (Linari et al., 2022). En el TMT-A, la edad emergió como predictor principal, esperable dado que la Parte A mide predominantemente velocidad de procesamiento, que declina de forma robusta con la edad. Es interesante resaltar que la edad se ubicó en el duodécimo lugar para el TMT-B: esta disociación es consistente con los hallazgos de Salthouse y colaboradores (2011), quienes reportaron que la Parte B se relaciona más fuertemente con capacidad cognitiva fluida más allá de la velocidad, introduciendo una variabilidad que la edad sola no captura. Finalmente, el MMSE exhibió un perfil de importancia más homogéneo, con contribuciones de múltiples variables que incluyen transiciones de estado, medidas de velocidad y tiempo de vacilación, consistente con su carácter de batería de *screening* multidominio (Tombaugh & McIntyre, 1992). En conjunto, estos resultados respaldan la interpretación de que los biomarcadores digitales derivados de las trayectorias del cursor son indicadores sensibles del rendimiento cognitivo.

En cuanto a la comparación entre conjuntos de características, aunque los grupos estaban emparejados en edad y educación y no difirieron estadísticamente, esto no garantiza que las variaciones sutiles no afectaran de forma diferencial a las variables digitales y no digitales. En el modelo demográfico no observamos una clasificación significativa, lo cual es consistente con una visión general de las diferencias entre los grupos emparejados. Para el modelo no digital, en línea con lo anterior, la inclusión de variables demográficas no mejoró el desempeño (AUC: 0.76 a 0.75), lo cual también es consistente con que las puntuaciones neuropsicológicas ya incorporan información

demográfica en sus valores (incluimos tanto puntajes absolutos como z-scores corregidos por edad y educación). En cambio, las variables demográficas sí contribuyeron a mejorar el desempeño en el modelo digital (AUC: 0.67 a 0.70). Además, la edad apareció entre los predictores más importantes en los modelos de regresión, como característica que complementa la información de los atributos digitales. En conjunto, estos resultados sugieren que las medidas digitales capturan menos información relativa a la edad y los años de educación, o que lo hacen de forma menos explícita.

El modelo digital alcanzó un buen desempeño (AUC = 0.70) comparado con la batería neuropsicológica completa (AUC = 0.76), considerando su simplicidad y automaticidad. Vale destacar que concebimos el cTMT como una herramienta de *screening* o complemento, no como un reemplazo de la evaluación clínica. Dada la simplicidad del hardware necesario, este enfoque también fue pensado para poder implementarse de forma remota, lo que permitiría una mayor accesibilidad y escalabilidad, y que exploraremos en el **Capítulo 5**. Además, estos hallazgos complementan otros trabajos que exploran enfoques multimodales para las implementaciones de TMT digitales. Por ejemplo, se han integrado métodos de seguimiento ocular en marcos similares. En investigaciones anteriores, descubrimos que la combinación de métricas de seguimiento manual y ocular durante una versión computarizada del TMT se correlacionaba con componentes específicos de funciones ejecutivas (Linari et al., 2022). De manera similar, la herramienta ETMT (Chandrasekharan et al., 2023) emplea el seguimiento ocular y la inferencia neurodifusa adaptativa para producir una puntuación global de deterioro cognitivo, mientras que Jyotsna y colaboradores (2020) utilizaron también datos de seguimiento ocular para clasificar a los participantes en grupos de deterioro bajo, medio o alto. En conjunto, estos enfoques ilustran cómo el TMT puede adaptarse a contextos digitales y multimodales que capturan los procesos cognitivos con mayor detalle.

En comparación con estudios anteriores que utilizaron características digitales basadas en el movimiento TMT para la clasificación del DCL, nuestro trabajo difiere en su alcance, metodología y enfoque analítico. Por ejemplo, Hanczár et al. (2022) analizaron segmentos cortos de movimiento del *mouse* utilizando un software propietario desarrollado por la empresa a la que pertenecen los autores y un modelo de árbol de decisión entrenado con aumento de datos sintéticos y validado mediante divisiones aleatorias repetidas. Por el contrario, nuestro análisis se basa en trayectorias de ensayo completas y en la extracción de características estandarizadas a través de nuestra biblioteca de código abierto *NeuroTask*. También implementamos una evaluación más rigurosa utilizando validación cruzada anidada, pruebas de permutación y corrección de comparaciones múltiples con Bonferroni. Lara-Garduno et al. (2022) desarrollaron *SmartStrokes*, un sistema basado en tablet que incorpora el reconocimiento de trazos durante un TMT digital. Los autores también utilizaron modelos de regresión para predecir las puntuaciones de las pruebas neuropsicológicas, informando de un MAE medio de aproximadamente 2.4 (sd. 0.110) en MoCA, mientras que nuestras predicciones MMSE alcanzaron un valor inferior de 1.45 ($p = 0.004$). Si bien ambas pruebas utilizan una escala de 0 a 30, la comparación es aproximada dado que pueden presentar distribuciones de puntajes diferentes según la población: en muestras con DCL leve el MMSE tiende a mostrar menor variabilidad, lo que favorece MAE más bajos independientemente del modelo. Con esa salvedad, nuestro análisis demuestra una precisión comparable o mejorada. En conjunto, estos hallazgos sugieren que tanto los enfoques basados en el uso de lápiz óptico como en el *mouse* pueden captar una variabilidad cognitiva significativa, mientras que nuestro enfoque hace hincapié en la transparencia, la accesibilidad y el rigor metodológico.

Por otro lado, la rápida proliferación de modelos de aprendizaje automático para el diagnóstico y la predicción de la progresión del DCL a menudo ha superado al rigor metodológico.

Las revisiones han destacado que muchos estudios omiten detalles críticos sobre la selección de características, los procedimientos de validación o la reproducibilidad, lo que limita la traslación clínica. Al buscar adherirnos a las normas establecidas de entrenamiento y evaluación de modelos (Poldrack et al., 2020; Sasse et al., 2025), contribuimos a los esfuerzos en curso para mejorar la transparencia y la fiabilidad en la neuropsicología digital.

Es importante destacar que el uso de la validación cruzada anidada maximizó el uso de los datos al tiempo que mantuvo una estimación casi insesgada del desempeño de los modelos. Este enfoque es especialmente adecuado para conjuntos de datos pequeños, en los que mantener un conjunto de pruebas separado reduciría aún más el tamaño de la muestra y aumentaría la varianza de las estimaciones de desempeño (Varoquaux, 2018). Reconocemos que contar con más datos hubiera permitido reservar un conjunto de prueba independiente (*held-out set*), lo que habría sido beneficioso para obtener una estimación aún más conservadora del desempeño; ante el tamaño muestral disponible, optamos por el enfoque que la literatura recomienda para este escenario. Nuestro objetivo no era producir un modelo de diagnóstico implementable, sino comparar rigurosamente conjuntos de características en condiciones de validación coherentes.

Cabe considerar que el diagnóstico clínico del DCL incluyó el rendimiento en el TMT-A y TMT-B tradicionales, que comparten la misma tarea subyacente que las características digitales del cTMT, lo que plantea una posible preocupación de circularidad. Sin embargo, esta preocupación está sustancialmente limitada por los criterios diagnósticos aplicados. El criterio de Jak et al. (2009) requiere un rendimiento deteriorado en al menos dos pruebas dentro de un dominio cognitivo, combinado con una evaluación clínica cualitativa basada en una entrevista y el reporte de un informante. En nuestra muestra, solo 10 de los 41 participantes con DCL (24%) tuvieron al menos un puntaje TMT por debajo del umbral de 1.5 DE, lo que significa que para el 76% restante de los casos, el rendimiento en el TMT no fue un factor determinante en el diagnóstico.

Es importante remarcar algunas de las limitaciones que observamos en este capítulo de la presente tesis. En primer lugar, el rendimiento en las tareas digitales puede variar en función del dispositivo de entrada y su sensibilidad, lo que puede introducir una variabilidad no deseada. Aunque todos los participantes en este estudio utilizaron el mismo mouse y la misma pantalla, esto sigue siendo una consideración importante para futuras implementaciones o comparaciones entre estudios. En segundo lugar, el presente trabajo se centró en un enfoque unimodal; la combinación del movimiento de la mano con otras modalidades, como el seguimiento ocular, el habla o las medidas fisiológicas, podría proporcionar una visión más completa del funcionamiento cognitivo normal y patológico. En tercer lugar, aunque el tamaño de nuestra muestra es aceptable para estudios de este tipo, limita la potencia estadística y la generalización. Los intervalos de confianza (Bootstrap) reportados para AUC y MAE proveen una estimación de la incertidumbre, pero las estimaciones estables de rendimiento deberán confirmarse mediante replicación en cohortes más grandes y diversas. En cuarto lugar, el diseño transversal no permite evaluar si las características del cTMT pueden rastrear cambios cognitivos a lo largo del tiempo ni predecir la conversión de DCL a demencia, una pregunta clave que estudios longitudinales deberán abordar. En quinto lugar, si bien se excluyó a participantes con condiciones neurológicas o vasculares diagnosticadas, no se realizó un *screening* formal de dificultades motoras sutiles como temblor leve o artritis, ni se evaluó sistemáticamente la experiencia con la computadora. Estos factores podrían introducir variabilidad en las trayectorias del cursor no relacionada con la cognición. Estudios futuros se beneficiarían de incluir una tarea de control motor simple como covariable para desacoplar contribuciones motoras y cognitivas a la dinámica del cursor.

Una de las principales ventajas del enfoque propuesto es su bajo coste y su compatibilidad con el hardware estándar, lo que facilita su posible uso tanto en entornos clínicos como remotos. A diferencia de algunas pruebas estandarizadas (por ejemplo, el MMSE), el cTMT puede implementarse libremente sin restricciones de licencia. Además, la publicación en código abierto de *NeuroTask* promueve la transparencia metodológica y la reproducibilidad, lo que permite a otros grupos ampliarlo o adaptarlo.

Para concluir, nos interesa destacar que el deterioro cognitivo leve es una afección heterogénea que sigue siendo difícil de definir, identificar y supervisar en la práctica clínica (Sabbagh et al., 2020). La creciente adopción de herramientas neuropsicológicas digitales, incluidas las versiones informáticas del Trail Making Test, refleja un cambio más amplio hacia evaluaciones cognitivas accesibles y ricas en datos. Sin embargo, la utilidad clínica de estos enfoques depende del equilibrio entre la facilidad de uso, la interpretabilidad y la validación rigurosa.

Conclusiones

En este capítulo mostramos que las trayectorias de la mano registradas con un *mouse* común durante el cTMT contienen información cognitiva suficiente para discriminar entre personas con deterioro cognitivo leve y controles sanos, y para predecir el rendimiento en pruebas neuropsicológicas estándar. El modelo basado en la unión de características demográficas y digitales alcanzó un desempeño cercano al de la batería neuropsicológica completa, pese a apoyarse en hardware estándar y en un procedimiento automático y de bajo costo sin intervención de profesionales de la salud.

Concebimos este enfoque como una herramienta de *screening* o complemento, no como un reemplazo de la evaluación clínica. Su valor no reside solo en el desempeño predictivo, sino en el marco metodológico que lo sostiene: validación cruzada anidada, pruebas de permutación, interpretabilidad mediante SHAP y extracción de características con software de código abierto (*NeuroTask*). Su carácter transversal y unimodal, junto con el tamaño muestral, delimitan el alcance de estas conclusiones y señalan el camino hacia cohortes mayores, diseños longitudinales y abordajes multimodales.

Capítulo 4: DataPruebas: una plataforma para recolección de datos a escala

Este trabajo fue publicado en "ASAIID, Argentine Symposium on Artificial Intelligence and Big Data (2025):

Juantorena, G. E., Gauder, L., Laciana, P., Ferrer, L., & Kamienkowski, J. E. (2025). DataPruebas: una plataforma en línea para la recopilación de datos. JAIIO, Jornadas Argentinas de Informática, 11(1), 149-155.

Introducción

Recolectar datos de manera confiable es fundamental en la investigación en disciplinas como la ciencia cognitiva y la inteligencia artificial. A su vez, la posibilidad de hacerlo de manera remota abre un conjunto de oportunidades. Este potencial fue demostrado por el trabajo pionero de Luis von Ahn, creador de *reCAPTCHA* (von Ahn et al., 2008), *Duolingo* (von Ahn, 2013), y el concepto de *games with a purpose* (von Ahn, 2006). Su investigación mostró cómo la participación masiva en línea puede ser aprovechada para resolver problemas computacionales y recolectar datos a mayor escala.

El *crowdsourcing* ha surgido como un poderoso método para recolectar datos científicos a gran escala, permitiendo a los investigadores acceder a poblaciones diversas de participantes a través de internet. Para este fin, coexisten diversas plataformas, cada una con diferentes fortalezas y limitaciones. En el ámbito de los experimentos cognitivos: *Pavlovía* (Bridges et al., 2020) es compatible con estudios basados en *PsychoPy* pero cobra por participante y gestiona datos externamente. *Gorilla* (Anwyl-Irvine et al., 2020) ofrece un editor de experimentos intuitivo con alojamiento integrado, aunque requiere suscripción paga. *Cognition* (Vidal, 2020) es compatible con estudios basados en *jsPsych* con flexibilidad, aunque carece de gestión integrada de participantes. Opciones gratuitas como *JATOS* (K. Lange et al., 2015) e *IBEX* (Zehr & Schwarz, 2018) también existen: *JATOS* ofrece control completo pero requiere alojarlo en un servidor propio (*self-hosting*) y gestión de infraestructura propia, mientras que *IBEX* lo evita pero está limitado por un mini-lenguaje personalizado que restringe la flexibilidad y la integración moderna.

Por otro lado, plataformas como *Amazon Mechanical Turk*¹ y *Scale AI*² son ampliamente utilizadas para tareas de anotación de datos a gran escala, particularmente en el campo del aprendizaje automático. Son efectivas para distribuir microtareas, como etiquetado de imágenes, transcripción de audio o clasificación de texto, a una amplia cantidad de trabajadores. Sin embargo, estas plataformas presentan limitaciones importantes: ofrecen poco o ningún control experimental, carecen de seguimiento longitudinal de participantes, y proporcionan filtrado demográfico básico o limitado. *Prolific* (Palan & Schitter, 2018) se posiciona como una alternativa orientada específicamente a la investigación académica, ofreciendo pre-filtrado demográfico detallado y estándares de compensación justa que mejoran la calidad de los datos. No obstante, requiere pago por participante y su población está concentrada mayoritariamente en países angloparlantes.

Si bien estas herramientas han permitido un progreso significativo en la investigación basada en tareas remotas, muchas siguen siendo costosas o técnicamente exigentes y están diseñadas principalmente para

¹ <https://www.mturk.com/>

² <https://scale.com/>

participantes de habla inglesa, limitando aún más la accesibilidad y la inclusividad.

En respuesta a estas limitaciones, desarrollamos *DataPruebas* (ver **Apéndice A.1**; <https://datapuebas.org>), una plataforma de acceso abierto para optimizar el proceso de recolección de datos científicos. Diseñada específicamente para la comunidad hispanohablante y utilizando un conjunto de tecnologías web moderno (*Python/Django* para el *backend* y *JavaScript/React* para el *frontend*), la plataforma ofrece una interfaz accesible e intuitiva para que los participantes se registren, exploren estudios disponibles y contribuyan a investigaciones en curso.

Esta infraestructura tecnológica se concibió para superar las deficiencias de las herramientas existentes, estableciendo un entorno de recolección riguroso. La plataforma admite dos modos diferentes de participación: experimentos asíncronos y síncronos. En los primeros, los participantes completan tareas a su conveniencia, usando diferentes dispositivos con acceso a internet. Este formato es particularmente efectivo para obtener muestras grandes y diversas, ya que elimina la necesidad de coordinación en tiempo real. En contraste, los experimentos síncronos son sesiones programadas que implican interacción en tiempo real entre el participante y el investigador, a menudo usando herramientas de videoconferencia o funcionalidades de comunicación integradas. Este enfoque es más apropiado para tareas que requieren supervisión o retroalimentación inmediata. Al ofrecer ambos modos, la plataforma se adapta a una amplia variedad de diseños experimentales al tiempo que hace la participación más accesible.

Interfaz y Compatibilidad Experimental

Desde la perspectiva del participante, la plataforma ofrece un flujo de registro simple que recolecta variables demográficas clave (fecha de nacimiento, género, lugar de nacimiento y residencia, nivel educativo, entre otras) preservando el anonimato y permitiendo el emparejamiento con estudios relevantes (Figura 18). Los participantes pueden explorar los experimentos disponibles, filtrarlos por modalidad (asíncrona o síncrona) y participar a su conveniencia.

A

DATA PRUEBAS

¿Quiénes somos?

¡Ayúdanos a hacer ciencia!

Ayúdanos a investigar cómo funciona la mente a través de divertidos y sencillos experimentos online y presenciales.
Tus datos nos ayudarán a hacer avances científicos.

¡Hola de nuevo!
Ingresá para continuar.

Acceder con Google

Email

Password

¿Olvidaste tu contraseña?

Ingresar

Regístrate ahora y empezá a explorar

Regístrate

Cualquier duda podés comunicarte por mail a: somosdatapuebas@gmail.com

B

Completa tus Datos

Hola! Para encontrar los experimentos disponibles para usted, necesitamos que especifique algunos aspectos personales.

Seleccione su género

Argentina

Seleccione su nivel educativo

Ciudad Autónoma de Buenos Aires

Argentina

02/26/1937

Argentina

☒ Deseo recibir notificaciones por correo

Completar Registro

Figura 18. Interfaces de *DataPruebas*. **A.** Pantalla de inicio de sesión (*login*). **B.** Formulario demográfico que los participantes completan al registrarse por primera vez.

Desde la perspectiva del investigador, la plataforma es compatible con los *frameworks* de experimentación de código abierto más utilizados: *jsPsych* (de Leeuw, 2015), *lab.js* (Henninger et al., 2022) y *PsychoPy/PsychoJS* (J. Peirce et al., 2019). Los investigadores pueden alojar sus experimentos directamente en nuestros servidores o utilizar una URL externa, aprovechando en ambos casos la base de datos de participantes registrados y las funcionalidades integradas de la plataforma (filtros de reclutamiento, programación de sesiones, seguimiento de datos y exportación en formatos estándar). Todos los experimentos deben pasar por un proceso de moderación y revisión por parte de los administradores antes de ser publicados.

Para crear un experimento, los investigadores ingresan metadatos (título y descripción), suben materiales (por ejemplo, imágenes, audios, videos y/o formularios de consentimiento) y configuran parámetros como la duración estimada, la repetibilidad, los requisitos técnicos (por ejemplo, disponibilidad de cámara web) y los criterios de elegibilidad (edad, género, lugar de residencia). La actividad de los participantes puede monitorearse en tiempo real a través de un panel de ejecución. Los datos crudos de cada sesión (marcas de tiempo, respuestas, metadatos) pueden exportarse en formatos *JSON* o *ZIP* para su análisis y reproducibilidad.

Calidad de Datos Online

La transición de la experimentación de laboratorio a entornos en línea introduce desafíos metodológicos específicos que la plataforma busca abordar. Estos desafíos fueron particularmente relevantes para los estudios presentados en los **Capítulos 5 y 6** de esta tesis.

Un primer desafío es la variabilidad de hardware: los participantes acceden a los experimentos desde dispositivos heterogéneos con diferentes tamaños de pantalla, resoluciones, tasas de refresco y dispositivos de entrada (mouse, trackpad, webcam, micrófono, etc.). Para mitigarlo, la plataforma permite configurar requisitos técnicos mínimos por experimento, como el tipo de dispositivo, el tamaño mínimo de pantalla o la disponibilidad de cámara web. Adicionalmente, en los estudios remotos de esta tesis se implementó el procedimiento *Virtual Chinrest* (Li et al., 2020) para estandarizar las dimensiones de los estímulos en función del dispositivo de cada participante. El procedimiento consta de dos pasos. Primero, para estimar la densidad de píxeles de la pantalla (su escala física), se le pide al participante que ajuste el tamaño de un objeto en pantalla hasta hacerlo coincidir con el de una tarjeta física estándar (por ejemplo, una tarjeta de crédito) apoyada sobre el monitor, lo que permite convertir píxeles a centímetros. Segundo, para estimar la distancia de visualización, se emplea una tarea de punto ciego (Durgin et al., 1995): el participante se tapa un ojo, fija la mirada en un punto de referencia y un estímulo se desplaza hasta desaparecer en su punto ciego; como la posición angular del punto ciego respecto de la fovea es relativamente constante entre las personas, es posible inferir la distancia a la pantalla haciendo trigonometría. Combinando ambas estimaciones, los estímulos pueden expresarse en dimensiones físicas (centímetros o grados de ángulo visual) de forma consistente entre dispositivos heterogéneos. Cabe resaltar que la implementación oficial de este procedimiento para la librería *jsPsych* fue desarrollada como una contribución de código abierto en el marco de los desarrollos técnicos de este trabajo (ver **Apéndice A.2**).

La ausencia de un experimentador durante la sesión plantea además desafíos de control de calidad. *DataPruebas* contribuye a este objetivo al requerir el registro de cada participante con un identificador único y persistente, y al almacenar metadatos de sesión (dispositivo, navegador, duración) que facilitan la identificación de problemas técnicos. Los criterios de calidad específicos, como umbrales de rendimiento mínimo o detección de respuestas automatizadas, son responsabilidad del investigador responsable de cada experimento individual. En los **Capítulos 5 y 6** se detallan los controles aplicados en los estudios de esta tesis (plugin-virtual-chinrest; Juantorena & Kohler, 2021).

Captura de Trayectorias Motoras y Datos Multimodales

En función de los requerimientos de investigación de nuestro grupo y la comunidad científica en la que estamos inmersos, *DataPruebas* se diseñó específicamente para facilitar la recolección de datos multimodales, entre ellos: grabaciones de audio, seguimiento ocular mediante cámara web, y la adquisición precisa de trayectorias del cursor. Para simplificar la integración de experimentos externos, la plataforma ofrece un SDK en *JavaScript* que encapsula la comunicación con la API, gestiona la autenticación de forma

transparente y automatiza el registro de datos estructurados y el envío de archivos multimedia (audio y video).

En particular, la captura de trayectorias del cursor fue central para los estudios presentados en esta tesis. La misma se realiza mediante escuchas de eventos de *JavaScript* (*mousemove*), registrando las coordenadas (x, y) y la marca temporal de cada movimiento. La tasa de muestreo efectiva depende del navegador y el hardware del participante, variando típicamente entre 30 y 120 Hz. Si bien esta variabilidad es una limitación inherente al entorno *online*, los datos crudos pueden ser post-procesados mediante interpolación a una frecuencia uniforme (ver sección *preprocesamiento* en el **Capítulo 5**).

A partir de estas trayectorias, es posible extraer un amplio conjunto de características cinemáticas (velocidad, aceleración, distancia, etc.) que constituyen la base de los biomarcadores digitales explorados en los **Capítulos 3 y 5**. Además de las trayectorias motoras, *DataPruebas* fue utilizada para la realización de estudios de eye-tracking web (detallados en el **Capítulo 6**) y diversas investigaciones relacionadas con audio y video que, si bien no forman parte de los resultados de esta tesis, resaltan la capacidad multimodal de la plataforma.

Limitaciones Actuales y Trabajo Futuro

Si bien *DataPruebas* ofrece una infraestructura flexible para la realización de experimentos en línea y la recolección de datos, a continuación se presentan algunas de sus limitaciones actuales. Primero, la plataforma se encuentra alojada y es mantenida de forma centralizada por nuestro equipo, y aunque los investigadores tienen la posibilidad de almacenar datos fuera de nuestros servidores (por ejemplo, modificando el código del experimento para enviar datos a uno propio), este proceso aún no cuenta con soporte oficial a través de nuestra API, lo que limita la facilidad de integración con bases de datos externas o infraestructuras institucionales. Facilitar esta funcionalidad es un objetivo clave para el desarrollo futuro.

En segundo lugar, la creación de experimentos actualmente depende de conocimientos de programación o herramientas externas. *DataPruebas* no incluye una interfaz visual integrada para el diseño de experimentos, lo que puede presentar una barrera para investigadores sin formación técnica. Sin embargo, esta limitación puede ser abordada usando la interfaz gráfica de *Psychopy* (*PsychoPy Builder*) y exportando el código a *DataPruebas* como código en *JavaScript*. Nos encontramos facilitando y documentando este proceso para disminuir la fricción asociada al mínimo.

Mirando hacia el futuro, planeamos mejorar la experiencia del participante introduciendo elementos de *gamificación*, área donde ya iniciamos implementando un ranking basado en puntos obtenidos por la finalización de tareas. El objetivo es ofrecer a los participantes una sensación de progreso y reconocimiento por sus contribuciones a la ciencia.

Finalmente, aunque la plataforma ya está disponible de forma gratuita, planeamos publicarla como software de código abierto para permitir un uso más amplio y adaptado. Este paso favorecerá aumentar la transparencia, la colaboración y permitirá a las comunidades de investigación apropiarse de la infraestructura en sus contextos locales o institucionales.

Discusión

DataPruebas representa más que una solución técnica: es una comunidad creciente de investigadores y participantes comprometidos con el avance de la ciencia a través de la recolección de datos accesible, escalable y ética. La plataforma ya ha albergado más de 20 tareas experimentales diferentes, un número

superior a 950 ejecuciones de las mismas y casi 1000 usuarios registrados, contribuyendo a la creación de conjuntos de datos de alta calidad.

Creemos que el desarrollo local es crucial. Construir infraestructura de código abierto y sin fines de lucro en América Latina empodera a los investigadores en regiones subrepresentadas, reduce la dependencia de herramientas externas costosas, y ayuda a fomentar un ecosistema de investigación sostenible arraigado en las necesidades y realidades de nuestras comunidades. Esto también implica el desafío de mantener y mejorar continuamente estas herramientas.

La plataforma está diseñada con la flexibilidad como prioridad: es compatible con *frameworks* populares, permite tanto alojamiento interno como externo, y ofrece herramientas integradas para la gestión de participantes y exportación de datos. Este equilibrio entre bajas barreras técnicas y accesibilidad financiera la convierte en una opción ideal para investigadores que buscan control total sobre sus estudios sin sacrificar la usabilidad.

Conclusiones

DataPruebas es la infraestructura que hizo posible la recolección de datos multimodales a gran escala sobre la que se apoyan los estudios de esta tesis (**Capítulos 3, 5 y 6**). Más que una solución técnica, se concibió como una plataforma accesible y compatible con los *frameworks* de experimentación más usados, pensada desde y para la comunidad hispanohablante.

Quedan pasos pendientes que orientan el trabajo futuro: la publicación de la plataforma como software de código abierto, el soporte oficial para almacenamiento externo de datos y una interfaz visual de diseño de experimentos que reduzca la barrera técnica. Sostener y hacer crecer esta infraestructura, desarrollada de forma local y sin fines de lucro, es en sí mismo parte de la contribución de este capítulo.

Capítulo 5: Predicción de diferencias individuales en funciones ejecutivas a partir del TMT online

Este capítulo se basa en el siguiente trabajo que se encuentra en revisión:

Juantorena, G. E., Capelo, G., & Kamienkowski, J. E. (2026). Beyond completion time: predicting individual differences in executive functions using cursor trajectory features in an online Trail Making Test. En revisión.

Este capítulo también incluye trabajo de la siguiente tesis de grado en proceso, de la cual soy co-director:

Gianluca Capelo “Predicción de funciones ejecutivas con aprendizaje automático: Estudios de caso en cohortes online y contexto clínico presencial” (2026)

Introducción

En los capítulos anteriores demostramos que el cTMT permite extraer métricas a partir de los movimientos de la mano, y que estas métricas capturan información relevante sobre las funciones ejecutivas (FE), tanto en población neurotípica (**Capítulo 1**) como en la detección de deterioro cognitivo leve (**Capítulo 3**). Sin embargo, estos estudios fueron realizados en condiciones de laboratorio controladas, con muestras relativamente pequeñas y hardware estandarizado.

El presente capítulo extiende este enfoque a un contexto online y de gran escala, utilizando la plataforma *DataPruebas* (**Capítulo 4**), para investigar si las características del cursor pueden predecir diferencias individuales en funciones ejecutivas centrales (Diamond, 2013; Miyake et al., 2000) dentro de una población general.

La experimentación online ofrece ventajas complementarias al cTMT de laboratorio. Por un lado, permite recolectar muestras más grandes y diversas en ambientes naturales, mejorando la validez ecológica y la accesibilidad a poblaciones distintas a las WEIRD (*Western, Educated, Industrialized, Rich, Democratic*; Henrich et al., 2010). Por otro lado, incrementa el poder estadístico, lo cual es crítico para detectar asociaciones entre métricas motoras y capacidades cognitivas individuales. No obstante, introduce desafíos específicos como la variabilidad de hardware y la ausencia de supervisión directa (ver **Capítulo 4**).

Para evaluar rigurosamente el potencial predictivo de las características del cTMT, incluimos una batería de tareas estandarizadas como medidas de *ground truth* para dominios ejecutivos específicos: la tarea de *Change Detection Task* (CDT; Luck & Vogel, 1997) para la memoria visual de trabajo, la tarea *Go/No-Go* (Drewe, 1975; Luria, 1966) para la inhibición proactiva, y *Stop-Signal Task* (SST; Logan & Cowan, 1984) para la inhibición reactiva. Adicionalmente, exploramos la predicción de la edad, dada su estrecha relación con el rendimiento en el TMT (Tombaugh, 2004).

El presente estudio se enfoca en tres objetivos específicos: primero, extraer características de las trayectorias del cursor, como períodos de vacilación y perfiles de velocidad, que no son capturadas en evaluaciones tradicionales. Segundo, evaluar la capacidad de estas métricas para predecir la capacidad de memoria visual de trabajo, el control inhibitorio y la edad usando técnicas de aprendizaje automático. Finalmente, validar la viabilidad de la adquisición de datos en un entorno online y contribuir al desarrollo de biomarcadores digitales escalables para la detección temprana de disfunción ejecutiva.

Métodos

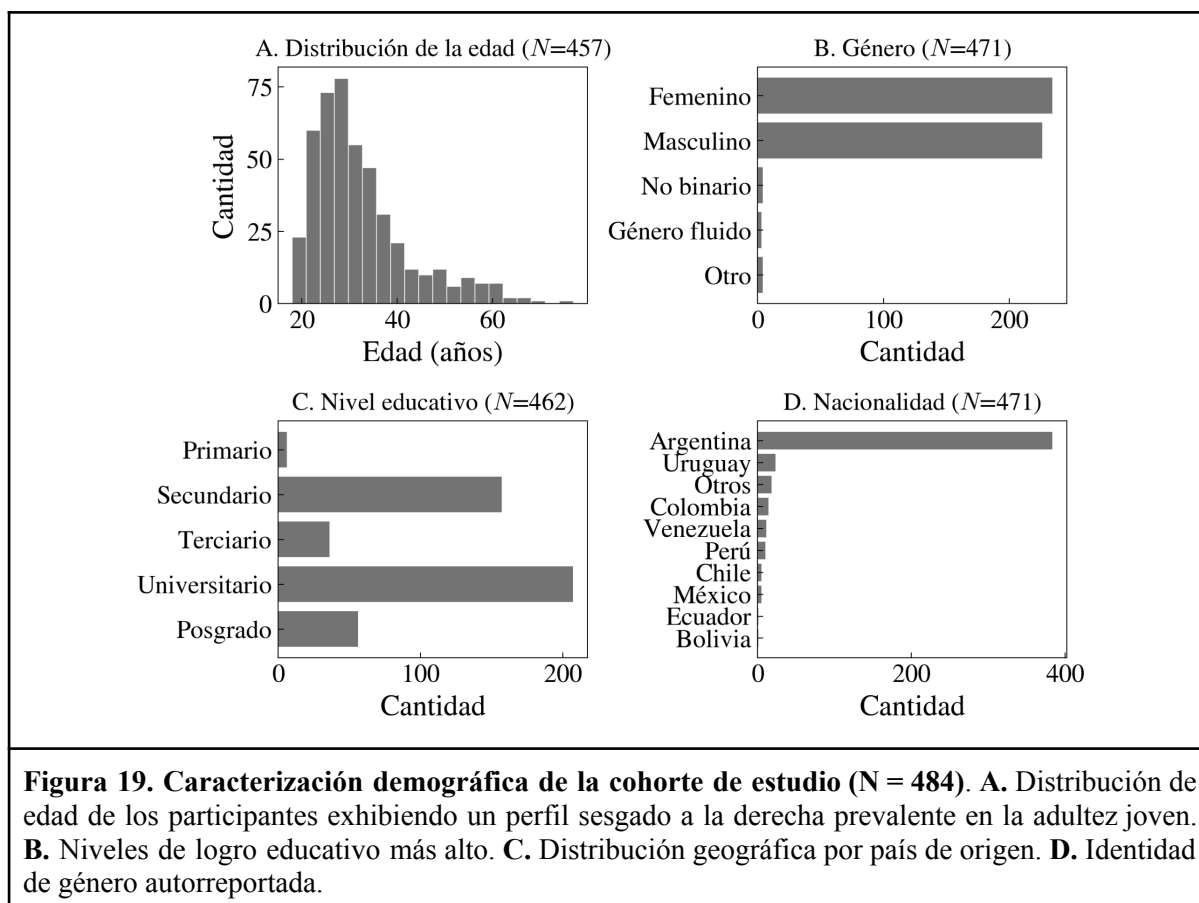
Participantes

Un total de 484 participantes únicos fueron identificados a través de las cuatro tareas cognitivas (cTMT, SST, CDT y Go/No-Go) después de aplicar controles de calidad específicos por tarea. No todos los participantes realizaron cada tarea: las muestras finales analizadas fueron de 368 para el cTMT, 246 para la SST, 248 para la CDT y 310 para la tarea Go/No-Go. Los participantes fueron reclutados en la universidad y a través de redes sociales, y los datos fueron recolectados usando *DataPruebas* (Juantorena et al., 2025; ver **Capítulo 4**). El *dataset* completo y anonimizado está disponible en el **Apéndice B.2**. Todos los participantes eran ingenuos respecto a los objetivos del experimento, tenían 18 años o más y otorgaron consentimiento informado electrónico de forma voluntaria. El estudio fue aprobado por el Comité de Ética correspondiente (Protocolo 295 del Instituto de Investigaciones Médicas Alfredo Lanari, Universidad de Buenos Aires).

De los 484 participantes que completaron al menos una tarea, se recolectaron datos demográficos autorreportados (Fig. 19). La disponibilidad de cada variable varió levemente: el género estuvo disponible para 471 participantes, la edad para 457, la nacionalidad para 471 y el nivel educativo para 462. Los porcentajes que se reportan a continuación se calcularon sobre el total de respuestas válidas de cada variable. La cohorte presentó una distribución de género equilibrada, con una leve prevalencia de participantes de género femenino (N=234, 49.7%) en comparación con el masculino (N=226, 48.0%); una minoría se identificó como género diverso (N=11, 2.3%). La edad de los participantes osciló entre 18 y 77 años [$M=32.1$, $DE=10.4$], con una distribución concentrada en la adultez joven. Respecto a la nacionalidad, la muestra fue predominantemente argentina (N=383, 81.3%), con representaciones menores de otros países latinoamericanos incluyendo Uruguay (N=23), Colombia (N=14), Venezuela (N=11) y Perú (N=10). En términos de educación, la muestra presentó un alto nivel educativo: el grupo más grande poseía un título universitario completo (N=207, 44.8%), seguido por aquellos con educación secundaria completa (N=157, 34.1%), de posgrado (N=56, 12.2%), formación terciaria o vocacional (N=36, 7.8%) y, en una minoría, solo educación primaria (N=6, 1.3%).

Criterios de inclusión

El estudio solo incluyó sujetos sin historial de trastornos psiquiátricos, epilepsia, convulsiones, abuso de drogas o alcohol, o uso de psicofármacos. Además, se les preguntó si tienen visión normal o corregida a normal. Dado que el enlace del experimento puede ser accedido fuera del Área Metropolitana de Buenos Aires (AMBA), se recolectó información sobre nacionalidad, país de residencia e idioma nativo. Solo se admitieron individuos de 18 años o más. Finalmente, el experimento no podía ser completado en teléfonos móviles, tablets o pantallas de dimensiones menores a 13 pulgadas.



Plataforma web y aspectos de implementación

Los participantes accedieron al estudio a través de *DataPruebas* (ver **Capítulo 4**), una plataforma diseñada para la recolección de datos multimodales en entornos web (Juantorena et al., 2025). La plataforma permite a los investigadores configurar parámetros del experimento (como criterios de elegibilidad y requisitos técnicos del dispositivo) y monitorear la actividad de los participantes en tiempo real a través de un panel de ejecución. Al ingresar, los participantes aceptaron un consentimiento informado y completaron un cuestionario demográfico antes de realizar las cuatro tareas cognitivas, las cuales podían ser completadas en sesiones independientes. Las tareas fueron desarrolladas utilizando *jsPsych* (v6.3; de Leeuw, 2015), una biblioteca de *JavaScript* diseñada para la creación de experimentos conductuales en línea. El enlace al código de los experimentos se puede encontrar en el **Apéndice A.3**.

Para asegurar la consistencia de los estímulos visuales en distintos dispositivos, empleamos el procedimiento de *Virtual Chinrest* (Li et al., 2020), detallado en el **Capítulo 4**. Este método permite estimar la densidad de píxeles y la distancia de visualización, estandarizando la experiencia experimental entre participantes.

Tareas Experimentales

Trail-Making Test

Implementamos una versión online del TMT utilizando la biblioteca *jsPsych* y un *plugin* desarrollado a medida (ver código fuente en el **Apéndice A.3.**). Los participantes realizaron dos tipos de ensayos: cTMT-A, que requería conectar números en orden ascendente (1 a 20), y cTMT-B, que requería alternar entre números y letras (1-A-2-B...). Cada tipo de ensayo fue señalizado mediante una cruz de fijación de color distintivo (azul para cTMT-A, roja para cTMT-B). A diferencia de las versiones de laboratorio presentadas en los primeros capítulos, esta implementación fue diseñada para minimizar el abandono en entornos no supervisados. El protocolo incluyó una fase de entrenamiento (un ensayo por condición) donde los ítems se resaltaban visualmente al interactuar, familiarizando al usuario con la mecánica antes de la fase de test, compuesta por 10 ensayos por condición sin retroalimentación visual. Se instruyó a los participantes a completar los trazos con el mouse de la forma más rápida y precisa posible; ante un error, debían regresar al último ítem válido para retomar la secuencia. El sistema registró el tiempo total de completado, la precisión y las trayectorias del cursor de alta resolución (coordenadas x e y, y marcas temporales). Solo se analizaron los ensayos de la fase de testeo. El diseño básico de los arreglos se mantuvo consistente con lo descrito en el **Capítulo 1** (ver sección *Generación de estímulos*).

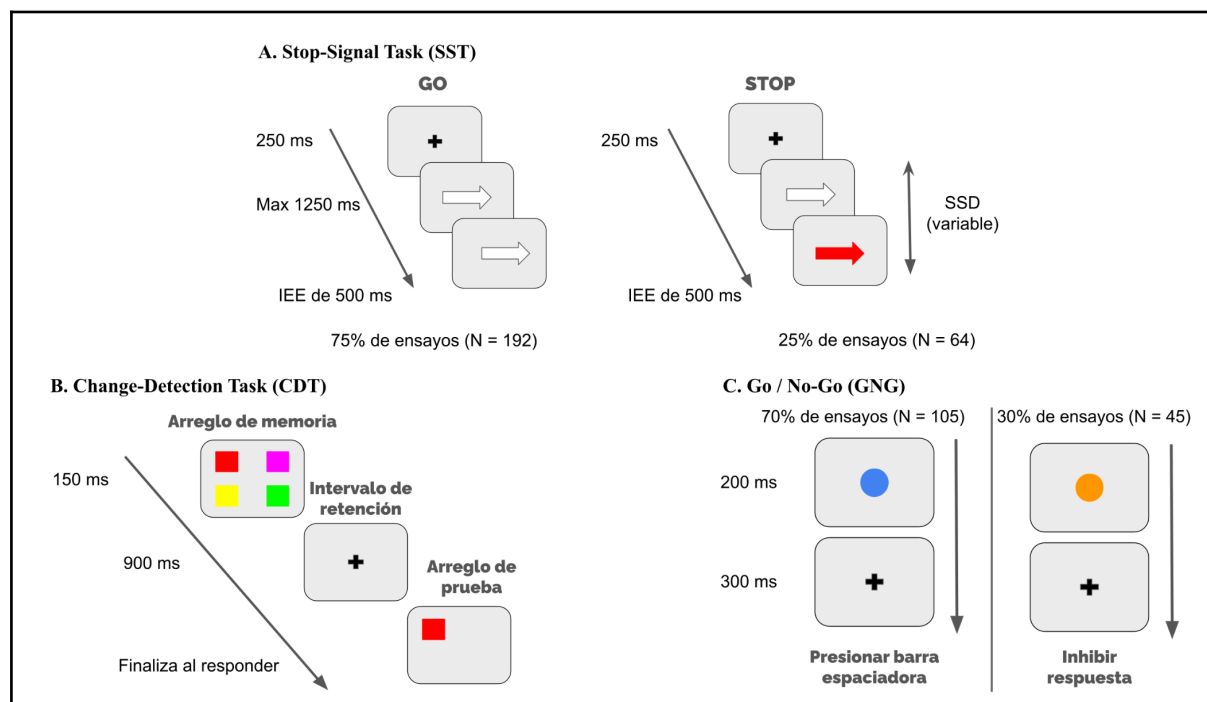


Figura 20. Representación gráfica de las tres tareas cognitivas implementadas en la plataforma junto al cTMT. A. Stop Signal Task (SST). B. Tarea de detección de cambios (CDT), que evalúa la memoria de trabajo visual. , que mide la inhibición reactiva. C. Tarea Go/No-Go, que evalúa la inhibición proactiva.

Stop-Signal Task

La tarea de *Stop-Signal* (SST) fue empleada para evaluar la inhibición reactiva, definida como la habilidad de cancelar una respuesta motora en curso ante la presentación de una señal de stop inesperada. Esta implementación se basó en el código de Verbruggen y colaboradores (STOP-IT; Verbruggen et al., 2008). La tarea está fundamentada en el *horse-race model* (Logan & Cowan, 1984), que conceptualiza el rendimiento como una carrera entre dos procesos competidores: un proceso "Go" disparado por el estímulo primario y un proceso "Stop" disparado por la señal de stop (Fig. 20A). Se instruyó a los participantes a responder lo más rápido y precisamente posible a un estímulo "Go" (una flecha blanca apuntando a izquierda o derecha) presionando la tecla de flecha correspondiente. En una minoría de ensayos (25%), el estímulo "Go" cambiaba de color a rojo (la señal de "Stop") tras un retardo variable, indicando que debían inhibir su respuesta. La tarea consistió en un bloque de práctica de 32 ensayos y cuatro bloques experimentales de 64 ensayos cada uno, resultando en un total de 256 ensayos experimentales. Cada ensayo comenzaba con una cruz de fijación (250 ms), seguida de la presentación del estímulo (duración máxima de 1250 ms) y un intervalo entre ensayos fijo de 500 ms. Se instruyó explícitamente a los participantes a no demorar sus respuestas en anticipación de la señal de stop, a fin de mantener distribuciones de tiempo de reacción válidas para el proceso "Go". Para asegurar una proporción aproximadamente equitativa de ensayos de inhibición exitosa y fallida (un prerequisite del *horse-race model*), el retardo de la señal de stop (*Stop Signal Delay*, SSD) se ajustó dinámicamente mediante un algoritmo de escalera adaptativa 1-up/1-down: el SSD inicial fue de 200 ms y se ajustó en ± 50 ms tras ensayos de inhibición exitosa y fallida, respectivamente. Durante las pausas entre bloques se proporcionó retroalimentación sobre el tiempo de reacción medio y la precisión.

Change-Detection Task

La capacidad de memoria visual de trabajo (VWM) fue evaluada usando un paradigma de detección de cambios (*Change detection Task*) implementado mediante un *plugin* de *jsPsych* desarrollado a medida (Fig. 20B). El diseño siguió el protocolo canónico establecido por Luck & Vogel (1997). Cada ensayo comenzó con una cruz de fijación central, seguida de un arreglo de 4 o 6 cuadrados coloreados presentados durante 150 ms, extraídos de un conjunto de ocho colores altamente discriminables (negro, azul, cian, verde, magenta, naranja, rojo y amarillo). Tras un intervalo de retención de 900 ms, aparecía un único cuadrado "target" en la ubicación de uno de los cuadrados previos, y los participantes debían juzgar si su color coincidía con el del cuadrado originalmente presentado en esa posición. Las respuestas se registraron por teclado ('L' para "igual" y 'D' para "diferente") dentro de una ventana máxima de 5000 ms. La sesión comenzó con una fase de entrenamiento con *feedback*, seguida de una fase de prueba de 120 ensayos divididos en tres bloques, balanceados entre tamaños de conjunto (60 ensayos con tamaño 4 y 60 con tamaño 6) y condiciones de cambio (50% de probabilidad de coincidencia). La medida de resultado primaria fue la capacidad de memoria (K de Cowan), junto con la precisión y el tiempo de respuesta.

Go/No-Go

Implementamos un paradigma Go/No-Go usando *jsPsych* para evaluar la inhibición proactiva (Jahfari et al., 2010), es decir, la habilidad de retener una respuesta antes de que sea iniciada (Fig.

20C) (Barbosa et al., 2020). Se les presentó a los participantes una serie de círculos coloreados en el centro de la pantalla y se les instruyó para presionar la barra espaciadora lo más rápido posible cuando apareciera un círculo azul (condición "Go") y retener su respuesta cuando apareciera un círculo naranja (condición "No-Go"). El experimento comenzó con un bloque de entrenamiento (6 ensayos) con un ritmo más lento y retroalimentación inmediata. La fase de prueba consistió en 150 ensayos divididos en 15 bloques. La probabilidad de señales "Go" fue establecida en 70% (105 ensayos) frente a 30% para señales "No-Go" (45 ensayos); esta alta frecuencia de ensayos Go fue diseñada para establecer una tendencia prepotente a responder, incrementando así la demanda cognitiva sobre los mecanismos de restricción inhibitoria. Durante la fase de prueba, el círculo se presentó durante 200 ms; inmediatamente después de su desaparición aparecía una cruz de fijación durante 300 ms, resultando en una ventana de respuesta total de 500 ms por ensayo. El ensayo concluía si el participante respondía dentro de esa ventana o si el tiempo se agotaba.

Procesamiento de datos y extracción de características

Controles de calidad

Para mitigar los riesgos asociados a la recolección de datos en línea no supervisada y asegurar la confiabilidad de las medidas conductuales, implementamos una serie de controles de calidad. En primer lugar, cada participante se registró bajo un identificador único y persistente, lo que evitó envíos duplicados y permitió vincular sus datos a través de las distintas tareas. En segundo lugar, el procedimiento de *Virtual Chinrest* nos permitió estimar la distancia de visualización y el tamaño de pantalla, posibilitando la exclusión de participantes con pantallas demasiado pequeñas para renderizar correctamente la disposición espacial del cTMT; además, la participación se restringió a usuarios de computadoras de escritorio o portátiles, excluyendo explícitamente dispositivos táctiles para garantizar que las trayectorias registradas reflejaran la dinámica precisa del mouse o el trackpad y no interacciones táctiles. En tercer lugar, aplicamos filtros de rendimiento para descartar respuestas de bajo esfuerzo o automatizadas. Para el cTMT, se consideraron válidos los ensayos en los que el participante conectó correctamente un mínimo de 10 ítems; los ensayos por debajo de ese umbral fueron excluidos.

De los 386 participantes que completaron el cTMT, 18 (4.7%) fueron excluidos de los análisis posteriores por datos incompletos, predominantemente por carecer de ensayos válidos en la Parte B. La muestra final quedó conformada por 368 participantes con datos completos del cTMT, es decir, con al menos un ensayo válido tanto en la Parte A como en la Parte B. Para las tareas sensibles al tiempo de reacción (SST, Go/No-Go y CDT), aplicamos un filtro de cota inferior fisiológica de 150 ms. Si bien Whelan (2008) señala que los tiempos de reacción genuinos tienen un mínimo fisiológico de al menos 100 ms, optamos por un umbral conservador de 150 ms dentro del rango de exclusión recomendado (100 a 200 ms) para remover rigurosamente las respuestas anticipatorias.

Preprocesamiento de las trayectorias

Las series temporales del cursor (coordenadas x e y , y marcas temporales) provenientes del cTMT se procesaron mediante *NeuroTask* (ver **Capítulo 3**), la librería de *Python* de código abierto que desarrollamos para realizar análisis conductual de alta resolución. Dado que los datos se recolectaron a través de navegadores web, la tasa de muestreo original era intrínsecamente irregular debido a variaciones en el disparo de eventos del navegador, la carga del sistema y el rendimiento del

hardware. Para estandarizar la resolución temporal entre todos los ensayos y participantes, aplicamos un procedimiento de interpolación lineal: primero se eliminaron las marcas temporales duplicadas para asegurar que las mismas fueran estrictamente crecientes y luego las coordenadas x e y se remuestrearon de forma independiente sobre una grilla temporal uniforme de 60 Hz (intervalos de aproximadamente 16.67 ms). Esta frecuencia se eligió para alinearse con la tasa de refresco estándar de la mayoría de las pantallas y para capturar el ancho de banda del movimiento motor intencional sin introducir ruido de alta frecuencia. Este procedimiento estableció una resolución temporal consistente preservando la integridad espacial y cinemática del movimiento original. Para los ensayos válidos pero incompletos (con al menos 10 ítems conectados), las características se calcularon a partir de la trayectoria hasta el último ítem correctamente conectado, garantizando la comparabilidad a lo largo de la muestra.

Extracción de características y segmentación

Utilizando *NeuroTask* extrajimos un conjunto de características de alta dimensionalidad organizado en tres dominios (ver Tabla C5.S1 en los materiales suplementarios): métricas de rendimiento y temporales (como el tiempo intra-ítem o dwell time y el tiempo inter-ítem), métricas de cinemática y dinámica (perfiles de velocidad y aceleración, velocidades media y máxima) y métricas de error espacial respecto a la trayectoria ideal (desviación euclidiana acumulada y área entre la trayectoria real y la óptima). Estas familias de características, comunes a las versiones de laboratorio, se describen en detalle en el **Capítulo 3**. Dado que el estudio presencial y este estudio en línea se realizaron en momentos diferentes, fueron procesados con versiones distintas de *NeuroTask*. En consecuencia, los conjuntos de características extraídos coinciden en su mayoría, aunque difieren en algunas métricas específicas (comparar la Tabla C3.S1, presencial, con la Tabla C5.S1, en línea).

Para capturar procesos cognitivos latentes durante la ejecución de la tarea, segmentamos el flujo continuo del cursor en tres estados mutuamente excluyentes (búsqueda, desplazamiento y vacilación) mediante el algoritmo descrito en detalle en el **Capítulo 3**, basado en un umbral de velocidad específico para cada participante (Lara-Garduno et al., 2022). El umbral de velocidad es específico para cada participante y se define como la mediana de las velocidades promedio registradas en el segmento del primer al segundo ítem en todos los ensayos válidos de la Parte A. Los tres estados son: *búsqueda* (períodos de baja velocidad, por debajo del umbral específico de cada participante, en los que el cursor permanece dentro de un ítem o en sus inmediaciones tras abandonarlo), *desplazamiento* (movimientos balísticos que superan el umbral de velocidad, dirigidos hacia el siguiente ítem) y *vacilación* (períodos en los que la velocidad cae por debajo del umbral tras un desplazamiento, o el cursor se mueve de forma errática fuera de las zonas de los ítems, compatibles con corrección de la trayectoria). A partir de estos estados segmentados extrajimos métricas específicas, como la fracción de tiempo en estado de búsqueda y la velocidad promedio durante el desplazamiento.

La totalidad de las características individuales ($N=26$; ver Tabla C5.S1) se extrajeron para el ensayo completo y hasta el ítem 10 (ver *Controles de calidad* en este capítulo), y de forma separada para la Parte A y la Parte B, completando un total de 104 características ($26 \times 2 \times 2$). Estas 104 características constituyeron la entrada del esquema de aprendizaje automático, que incluye un paso de selección de características (ver sección *Esquema de aprendizaje automático*) que hizo que a los modelos llegaran una cantidad mucho menor.

Variables objetivo

Para entrenar los modelos predictivos, redujimos la salida de alta dimensionalidad de las tareas de validación a métricas escalares individuales que representan funciones ejecutivas centrales. La memoria visual de trabajo se definió como la capacidad de memoria (K de Cowan) derivada del *Change Detection Task*, calculada por separado para los ensayos de 4 ($K4$) y 6 ($K6$) ítems, así como su promedio ($K_{promedio}$). La inhibición se cuantificó mediante el tiempo de reacción de la señal de stop (*Stop Signal Reaction Time*, SSRT) calculado a partir de la Stop-Signal Task, y mediante varias medidas derivadas de la tarea Go/No-Go, de las cuales reportamos el coeficiente c y la precisión (*accuracy*). Finalmente, exploramos la edad como variable objetivo, dado que es sabido que está estrechamente relacionada con el comportamiento en el TMT (Tombaugh, 2004). La Tabla 4 resume cada una de estas variables junto con su fórmula e interpretación cognitiva.

Tabla 4. Resumen de las medidas utilizadas como objetivos de predicción, derivadas de las tres tareas de validación y de los datos demográficos. Para cada variable objetivo se indican la fórmula y la interpretación cognitiva. HR: tasa de aciertos (*Hit Rate*); FA: tasa de falsas alarmas (*False Alarm rate*); $\overline{HR}$, $\overline{FA}$: medias muestrales; DE_HR, DE_FA: desviaciones estándar muestrales; S: tamaño del conjunto (*set size*); SSD: retardo de la señal de stop; p_r : probabilidad de responder en ensayos de stop.

Tarea	Variable objetivo	Fórmula	Interpretación
Change Detection Task	K de Cowan	$K = S \cdot (2 \cdot Accuracy - 1)$	Capacidad de memoria visual de trabajo: una K más alta posibilita una navegación más fluida y patrones de búsqueda más eficientes.
Stop-Signal Task	SSRT	$SSRT = F_{Go}^{-1}(p_r) - \overline{SSD}$	Control inhibitorio motor: tiempos más largos indican mayor dificultad para suprimir una respuesta prepotente.
Go/No-Go	Coeficiente c	$c = -0.5 \cdot \left(\frac{HR - \overline{HR}}{SD_{HR}} + \frac{FA - \overline{FA}}{SD_{FA}} \right)$	Sesgo de respuesta: vinculado a la eficiencia de rastreo de base y a la decisión en la planificación de la trayectoria.

Go/No-Go	<i>Accuracy</i>	$Accuracy = \frac{Correctos}{N_{trials}}$	Rendimiento global ante el desafío inhibitorio: balance entre aciertos en ensayos "Go" y falsas alarmas en "No-Go".
Demografía	Edad	<i>Edad</i> [18, 77] y.o.	Envejecimiento cerebral: caracterizado por mayor ruido motor, menor eficiencia espacial y períodos de planificación motora más prolongados.

Esquema de aprendizaje automático

Seguimos el mismo esquema de validación cruzada anidada empleado para el análisis de regresión del **Capítulo 3** (Fig. 15): un bucle externo Leave-One-Out (LOOCV) para estimar la generalización y un bucle interno de 10 *folds* para el ajuste de hiperparámetros mediante grid search, con todo el preprocesamiento ajustado exclusivamente sobre los datos de entrenamiento dentro de cada *fold* para prevenir el *data leakage* (Sasse et al., 2025). La única adición respecto al **Capítulo 3** fue un paso inicial de imputación de valores faltantes por la media del conjunto de entrenamiento (*SimpleImputer*), necesario dado que los datos provienen de sesiones en línea no supervisadas. Como en la regresión del **Capítulo 3**, la selección univariada de características (*SelectKBest*, $k \leq 20$; Pedregosa et al., 2018) usó el test F de regresión (*f_regression*), seguida de una estandarización a media cero y varianza unitaria (*StandardScaler*). Los hiperparámetros se optimizaron por el error absoluto medio negativo (Poldrack et al., 2020). Evaluamos siete modelos de regresión abarcando distintas familias: Regresión Lineal, Ridge, Lasso, Elastic Net, Support Vector Regression (SVR), Random Forest y XGBoost. Todos los experimentos usaron semillas aleatorias fijas para garantizar la reproducibilidad. Las predicciones de todos los *folds* de LOOCV se agregaron para calcular el MAE como métrica final. Como en el **Capítulo 3**, reportamos además el error normalizado (MAE/DE) como métrica de tamaño de efecto, que expresa el error de predicción relativo a la variabilidad natural de cada variable objetivo y permite comparar entre puntajes de distinta escala; asimismo, informamos el intervalo de confianza del 95% del MAE. Las grillas completas de hiperparámetros se proveen en la Tabla C3.S3 de los materiales suplementarios.

Análisis estadístico

El p-valor para cada modelo y variable objetivo (ver Tabla 5) se estimó mediante un test de permutación. El test compara el MAE observado con el MAE estimado en $N = 1000$ permutaciones de

la asociación entre los valores de la variable objetivo y el vector de características de cada individuo (Ecuación 2).

$$p = \frac{(\text{número de permutaciones con } MAE \leq MAE \text{ observado}) + 1}{N + 1} \quad (2)$$

Importancia de características: enfoque SHAP

Para la interpretación de los modelos aplicamos el mismo análisis SHAP descrito en el **Capítulo 3** (Lundberg & Lee, 2017): calculamos los valores de Shapley para cada sujeto reservado en cada *fold* de LOOCV, reajustando en cada uno el modelo seleccionado sobre su conjunto de entrenamiento, y los agregamos como el promedio de las contribuciones absolutas a través de los *folds* en que cada característica fue seleccionada.

Resultados

Datos conductuales

Siguiendo los procedimientos de control de calidad y extracción de características descritos anteriormente, las estadísticas descriptivas finales para cada tarea se presentan en la Figura 21. Para el cTMT, los datos de 368 participantes cumplieron los estrictos criterios de control de calidad (que requerían completar válidamente tanto la Parte A como la Parte B). El tiempo promedio para completar la Parte A fue de 21.93 s (DE = 2.10), mientras que la Parte B, cognitivamente más demandante, requirió un promedio de 23.82 s (DE = 1.18). Una prueba t pareada confirmó que el incremento en la duración de la Parte B fue estadísticamente significativo ($t(367) = -20.48$, $p < 0.001$, d de Cohen = 1.07), confirmando el costo ejecutivo esperado. Respecto al Porcentaje de Ensayos Completados (PC), es decir, aquellos en los que se pasó por al menos 10 ítems en el orden correcto, los participantes completaron exitosamente un promedio de 88.92% (DE = 15.36) de los ensayos de la Parte A y 73.55% (DE = 22.80) de los de la Parte B. Una prueba t pareada reveló que esta diferencia fue estadísticamente significativa ($t(367) = 16.10$, $p < 0.001$, d de Cohen = 0.84), indicando que las mayores demandas cognitivas de la Parte B se acompañaron de una menor tasa de ensayos completos.

Respecto a la Stop-Signal Task (SST), los datos de 246 participantes cumplieron los criterios de control de calidad. El Stop Signal Reaction Time (SSRT) promedio fue de 227.7 ms (DE = 49.3), con una probabilidad media de respuesta ante la señal de stop de 0.48. Este valor cercano a 0.5 indica un seguimiento exitoso del umbral de respuesta por parte del algoritmo de escalera adaptativa, asegurando la validez de la estimación del SSRT.

En la tarea Go/No-Go (GNG), un total de 310 participantes fueron incluidos en el análisis final, tras excluir a 18 sujetos por bajo *accuracy* (<60%). Las métricas mostraron una tasa de aciertos (*hit rate*) media de 0.82 (DE = 0.11) y una tasa de falsas alarmas de 0.27 (DE = 0.14), reflejando un nivel moderado de desafío inhibitorio. El tiempo de reacción medio en los ensayos Go fue de 333 ms (DE = 26), consistente con tareas estándar de control cognitivo.

Finalmente, los análisis para la Change Detection Task (CDT) incluyeron 248 participantes tras la aplicación de los procedimientos de control de calidad. Se excluyeron sujetos que exhibieron una tasa de omisión de ensayos superior al 20% o una precisión global inferior al 60%. El análisis

reveló una capacidad media de 2.83 (DE = 0.61) para la condición de baja carga (set size 4) y de 2.45 (DE = 1.00) para la de alta carga (set size 6). La precisión global en la muestra fue del 78%.

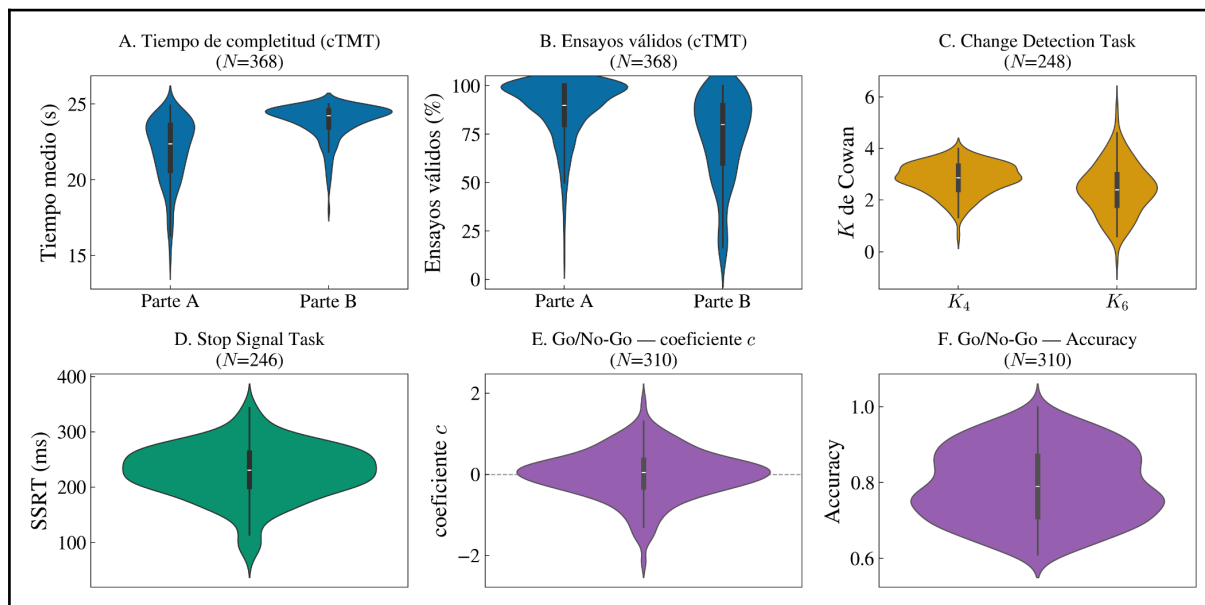


Figura 21. Distribución de las métricas de rendimiento a través de las tareas cognitivas. Los violín plots muestran la estimación de densidad kernel de cada métrica, con una caja intercuartil y un marcador de mediana. **A.** Tiempo medio de completado en el cTMT para ambas partes. **B.** Porcentaje de ensayos válidos (completados) en el cTMT. **C.** Capacidad de memoria visual de trabajo estimada por la K de Cowan para los tamaños de conjunto 4 y 6 en la tarea de *Change Detection Task* (CDT). **D.** Stop-Signal Reaction Time (SSRT), calculado mediante el método de integración, como medida de control inhibitorio. **E.** Coeficiente c de la tarea Go/No-Go. **F.** Accuracy en la tarea Go/No-Go, reflejando la proporción de respuestas correctas en ensayos Go y No-Go. Los colores denotan las tareas cognitivas: azul = cTMT (A, B), naranja = CDT (C), verde = SST (D), violeta = Go/No-Go (E-F). Los tamaños muestrales (N) varían entre paneles debido a los criterios de exclusión específicos de cada tarea.

Resultados de predicción

Se implementaron diferentes modelos predictivos en el mismo esquema de entrenamiento para distintas variables objetivo. Los mismos variaron en complejidad, desde la Regresión Lineal hasta Boosting (XGBRegressor), incluyendo regresión regularizada (Ridge, LASSO, Elastic Net), Support Vector Regression (SVR) y Random Forest Regression. La Tabla 5 resume los resultados significativos para las variables objetivo seleccionadas. La significancia se evaluó mediante un test de permutación, y el MAE constituye el tamaño del efecto. El MAE debe compararse con las estimaciones de dispersión de las variables objetivo, calculadas a partir de la muestra (DE e IQR).

Tabla 5. Resumen de resultados predictivos para cada variable objetivo. Cada modelo corresponde a un subconjunto de participantes que completaron ambas tareas (ver tamaños muestrales en **Métodos**). Las filas en negrita indican los modelos de mejor desempeño, los cuales fueron luego seleccionados para el análisis SHAP. "Subconjunto de sujetos" refiere a los participantes que completaron las tareas: 'cTMT $\cap$ CDT' indica los participantes que completaron tanto el cTMT como la CDT. El error normalizado se estimó como el ratio entre el MAE y la DE. El intervalo de confianza (IC 95%) del MAE se estimó mediante *bootstrap*. El desvío estándar (DE) y el rango inter-cuartil (IQR) se estimaron sobre los datos observados. El p-valor se estimó con un test de permutación (N=1000; ver **Métodos**).

Target	Subconjunto de sujetos	Modelo	MAE	IC 95% (MAE)	DE _{Target}	Error normalizado (MAE/DE)	IQR _{Target}	p-valor
edad	cTMT	SVR	6.78	[6.10, 7.53]	10.06	0.67	11.0	<0.001
edad	cTMT	Ridge	7.08	[6.46, 7.74]	10.06	0.7	11.0	<0.001
edad	cTMT	ElasticNet	7.08	[6.47, 7.74]	10.06	0.7	11.0	<0.001
edad	cTMT	LinearRegression	7.18	[6.55, 7.85]	10.06	0.71	11.0	<0.001
edad	cTMT	Lasso	7.19	[6.58, 7.84]	10.06	0.71	11.0	<0.001
edad	cTMT	RandomForestRegressor	7.33	[6.71, 7.98]	10.06	0.73	11.0	<0.001
edad	cTMT	XGBRegressor	7.61	[6.91, 8.34]	10.06	0.76	11.0	<0.001
K_{promedio}	cTMT $\cap$ CDT	RandomForestRegressor	0.57	[0.51, 0.64]	0.72	0.79	0.94	0.011
K _{promedio}	cTMT $\cap$ CDT	XGBRegressor	0.59	[0.53, 0.65]	0.72	0.82	0.94	0.003
K _{promedio}	cTMT $\cap$ CDT	Ridge	0.57	[0.51, 0.63]	0.72	0.79	0.94	0.024
accuracy	cTMT $\cap$ GNG	SVR	0.08	[0.07, 0.08]	0.09	0.89	0.16	0.003
accuracy	cTMT $\cap$ GNG	RandomForestRegressor	0.08	[0.07, 0.08]	0.09	0.89	0.16	0.004
accuracy	cTMT $\cap$ GNG	Ridge	0.08	[0.07, 0.08]	0.09	0.89	0.16	0.007
accuracy	cTMT $\cap$ GNG	ElasticNet	0.08	[0.07, 0.08]	0.09	0.89	0.16	0.012
accuracy	cTMT $\cap$ GNG	XGBRegressor	0.08	[0.07, 0.09]	0.09	0.89	0.16	0.015
accuracy	cTMT $\cap$ GNG	Lasso	0.08	[0.07, 0.08]	0.09	0.89	0.16	0.020
accuracy	cTMT $\cap$ GNG	LinearRegression	0.08	[0.07, 0.09]	0.09	0.89	0.16	0.053
coeficiente c	cTMT $\cap$ GNG	Ridge	0.43	[0.38, 0.48]	0.6	0.72	0.66	<0.001
coeficiente c	cTMT $\cap$ GNG	Lasso	0.43	[0.38, 0.48]	0.6	0.72	0.66	<0.001
coeficiente c	cTMT $\cap$ GNG	ElasticNet	0.44	[0.39, 0.49]	0.6	0.73	0.66	<0.001
coeficiente c	cTMT $\cap$ GNG	RandomForestRegressor	0.45	[0.40, 0.50]	0.6	0.75	0.66	<0.001
coeficiente c	cTMT $\cap$ GNG	LinearRegression	0.47	[0.41, 0.52]	0.6	0.78	0.66	0.004
coeficiente c	cTMT $\cap$ GNG	SVR	0.45	[0.40, 0.50]	0.6	0.75	0.66	0.006
coeficiente c	cTMT $\cap$ GNG	XGBRegressor	0.5	[0.45, 0.56]	0.6	0.83	0.66	0.040

El cTMT mostró un fuerte poder predictivo sobre la edad, ya que todos los modelos presentaron una predicción significativa con p-valores de $p < 0.001$ (N permutaciones = 1000) y MAEs un tercio menores que la dispersión de la variable objetivo. Respecto a la memoria visual de trabajo, las medidas del cTMT presentaron una predicción significativa sobre el K_{promedio}. Cabe destacar que esto involucra principalmente los modelos más complejos (XGBoost y Random Forest), aunque Ridge también alcanzó significancia marginal ($p=0.024$), lo que sugiere que la relación entre K y el comportamiento del cTMT es no lineal y probablemente depende de interacciones entre las variables. Por su parte, la tarea Go/No-Go fue caracterizada con varias variables objetivo y mostró efectos significativos para todas ellas. El *accuracy* y el coeficiente *c* mostraron los efectos más

fuertes: para el *accuracy*, SVR fue el modelo más significativo, mientras que para el coeficiente *c* los modelos lineales regularizados (Ridge y Lasso) lideraron, sugiriendo una dependencia más lineal en este último caso. Los tamaños del efecto fueron muy similares entre los distintos modelos. Para el caso de la tarea de Stop-Signal, caracterizada por el SSRT, no se observaron resultados significativos.

Análisis de importancia de características

La importancia de características se estimó mediante los valores SHAP absolutos medios, calculados para el modelo de mejor desempeño por variable objetivo, seleccionado según el menor MAE. Para la predicción de la edad (Fig. 22A), las características dominantes fueron métricas basadas en distancia, específicamente la distancia de búsqueda completa (*Complete search distance*) y la distancia total completa (*Complete total distance*), presentes tanto en la Parte A como en la Parte B, junto con métricas de velocidad y tiempo inter-ítem (*Inter-target time*). La relevancia de la longitud de la trayectoria es consistente con el concepto de *circuitousness* descrito en la literatura del TMT (Woods et al., 2015), lo que sugiere que la eficiencia espacial de las trayectorias del cursor porta información individual robusta sobre la edad. La presencia de características de la Parte A en lo más alto del ranking indica que la ejecución visomotora básica, independiente de las demandas de alternancia alfanumérica, es particularmente informativa, consistente con la conocida sensibilidad del comportamiento motor a los cambios a lo largo de la vida (Salthouse, 2011; Tombaugh, 2004).

Para la predicción de la capacidad de memoria visual de trabajo (K_{promedio}) (Fig. 22B), las transiciones de estado (*State transitions*) en la Parte A y las transiciones de estado completas (*Complete state transitions*) en la Parte B emergieron como los principales predictores, seguidas por características temporales y basadas en velocidad distribuidas en ambas partes, con predominio de la Parte B. La prominencia de las transiciones de estado, que cuantifican cuán frecuentemente el cursor alterna entre las fases de búsqueda, desplazamiento y vacilación, sugiere que este patrón de alternancia entre estados es particularmente informativo sobre la capacidad de memoria visual de trabajo. A diferencia de la predicción de la edad, donde tanto el predictor principal como el grueso del ranking correspondieron a la Parte A, en el caso del K_{promedio} el predictor individual más importante proviene de la Parte A (*State transitions*), mientras que la mayoría de los predictores subsiguientes derivan de la Parte B. Esta combinación sugiere que la capacidad de memoria de trabajo modula la dinámica del cursor a través de los componentes tanto visomotor como de alternancia de la tarea. Esto apunta a que el K_{promedio} no solo se relaciona con funciones como la alternancia de tareas (Parte B), sino que también define la eficiencia de base de la búsqueda, la navegación del cursor y la planificación de trayectorias a lo largo de toda la prueba.

En el caso de la predicción del *accuracy* de Go/No-Go, una predominancia de métricas derivadas del cTMT-B apareció en las posiciones superiores del ranking de importancia (Fig. 22C). Las variables más influyentes correspondieron a esa parte del test: velocidad media de búsqueda completa (*Complete search average speed*), la distancia de búsqueda completa (*Complete search distance*) y la velocidad media completa (*Complete mean speed*), sugiriendo que la eficiencia de ejecución durante la tarea de alternancia del cTMT-B constituye el dominio funcional más informativo para predecir el rendimiento inhibitorio. Las métricas de la Parte A se ubicaron predominantemente en posiciones medias o bajas, indicando una menor contribución de los componentes de secuenciación simple. El rango general de valores SHAP fue estrecho y de baja magnitud, reflejando un patrón de predicción distribuido a través de múltiples características sin una dependencia marcada de un único predictor. Un patrón similar se observa si los valores SHAP se estiman a partir del modelo Ridge, que

también es significativo y provee una comparación más directa con los resultados del coeficiente c (ver Fig. C5.S1).

Respecto a la predicción del coeficiente c de Go/No-Go, los valores SHAP mostraron una predominancia casi absoluta de métricas derivadas del cTMT-A, que representaron 14 de las 15 variables del ranking de importancia (Fig. 22D). Las variables más influyentes fueron el tiempo de búsqueda completo (*Complete search time*), la velocidad media de búsqueda completa (*Complete search average speed*) y las transiciones de estado completas (*Complete state transitions*). Todas para la parte A del test. Esto indicó que el rendimiento durante la tarea de secuenciación simple del cTMT-A constituye el dominio funcional más informativo para predecir el sesgo de respuesta.

Considerando conjuntamente los resultados de *accuracy* y del coeficiente c , donde las métricas del cTMT-B y del cTMT-A dominaron respectivamente el ranking de importancia, los resultados sugieren un patrón diferencial en el que los procesos perceptivo-motores capturados por la Parte A parecen estar más estrechamente vinculados al sesgo de respuesta, mientras que los procesos ejecutivos evaluados por la Parte B están más fuertemente asociados con el rendimiento inhibitorio general. Esta disociación se alinea con la evidencia de que las dos partes del cTMT capturan procesos cognitivos diferenciados (Sánchez-Cubillo et al., 2009).

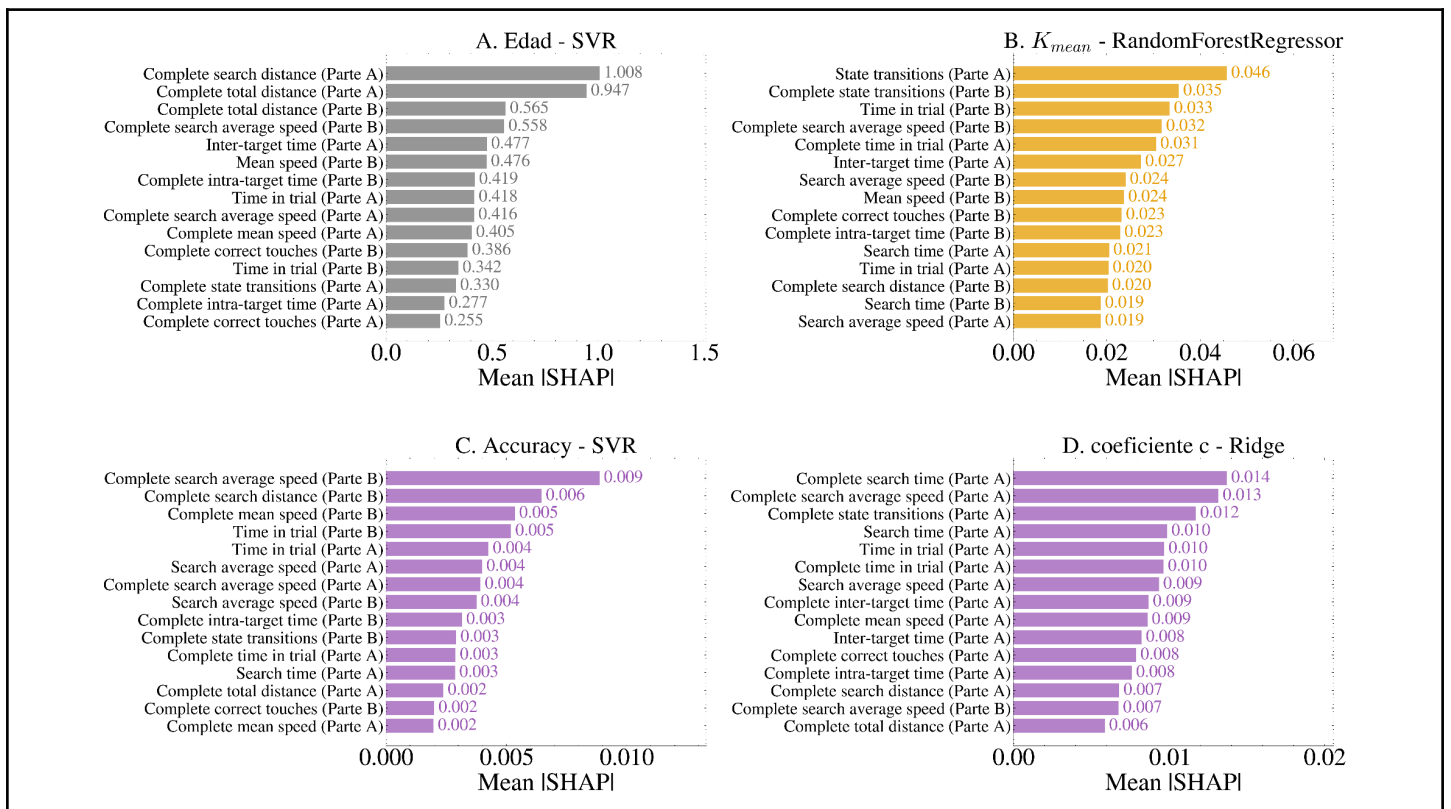


Figura 22. Valores medios |SHAP| a través de los *folds* LOOCV para las 15 características principales por modelo. **A.** Modelo SVR, predicción sobre la Edad. **B.** Modelo RandomForestRegressor, predicción sobre $K_{promedio}$. **C.** Modelo SVR, predicción sobre el *accuracy*. **D.** Modelo Ridge, predicción sobre el coeficiente c . Los colores de las barras corresponden a la tarea sobre la cual se calculó la métrica objetivo (Edad: gris; CDT: ámbar; Go/No-Go: violeta). Todos los modelos usaron el conjunto de características del cTMT. Las puntuaciones de importancia se promediaron entre *folds*, considerando los distintos subconjuntos de características debido a la selección de características (ver Métodos).

Discusión

En este capítulo exploramos el potencial del Trail Making Test computarizado (cTMT) para identificar diferencias individuales en funciones ejecutivas centrales mediante la extracción de características de alta resolución de las trayectorias del cursor, pero a diferencia de lo expuesto en el **Capítulo 3**, se trató de una tarea remota donde las personas realizaron el experimento desde sus casas, sin supervisión de ningún investigador y usando su propio *hardware*. Nuestros resultados muestran que la dinámica del cursor, particularmente aquella relacionada con la fluidez motora y las transiciones de estado, proveyó una señal conductual rica para predecir estas capacidades cognitivas, extendiendo la información capturada por las métricas tradicionales de tiempo de completado global. En resumen, los hallazgos clave incluyeron: (1) una fuerte relación entre la edad y las características del cTMT, especialmente aquellas que reflejan la sinuosidad (*circuitousness*); (2) la predicción exitosa de la capacidad de memoria visual de trabajo (K de Cowan) usando modelos no lineales, indicando que la VWM actuó como una restricción generalizada sobre la eficiencia de búsqueda, incluso bajo condiciones de baja carga ejecutiva (Parte A); y (3) un patrón de predicción diferencial para los resultados de Go/No-Go, donde las métricas de la Parte B predijeron mejor el *accuracy* mientras que las métricas de la Parte A predijeron mejor el sesgo de respuesta (coeficiente c).

Implicaciones teóricas: el cTMT como herramienta de evaluación cognitiva continua

Tradicionalmente, el TMT es visto como una medida de flexibilidad cognitiva y alternancia de tareas, con la Parte B indexando principalmente el costo de cambiar entre conjuntos alfanuméricos. Sin embargo, nuestro análisis de importancia de características reveló que el componente "motor" (Parte A) está más cargado cognitivamente de lo que se asumía previamente. Notablemente, las transiciones de estado en la Parte A emergieron como el predictor más fuerte de la capacidad de memoria de trabajo visual (K_{promedio}), superando incluso a su contraparte de la Parte B, mientras que características adicionales de la Parte A, como el tiempo inter-ítem (*Inter-target time*) y la velocidad promedio de búsqueda (*search average speed*), contribuyeron de forma moderada. Esto respalda la idea de que el TMT es esencialmente una tarea de búsqueda visual y planificación de trayectorias donde la demanda de monitoreo dirigido a metas es constante (Sánchez-Cubillo et al., 2009).

Observamos que las predicciones más robustas para el K_{promedio} se lograron usando modelos no lineales, como XGBoostRegressor o Random Forest. Esto sugiere que la capacidad de VWM influye en el rendimiento del cTMT a través de interacciones no lineales entre la planificación motora y la búsqueda visual, en lugar de una relación lineal directa. La prominencia de las transiciones de estado indica que el patrón de alternancia entre estados es altamente sensible a la capacidad de memoria. Este hallazgo se alinea con la "teoría de la velocidad de procesamiento" (Sommers & Dessenberger, 2022), pero le agrega una dimensión espacial.

Respecto al control inhibitorio, el patrón predictivo diferencial observado entre el *accuracy* global y el sesgo de respuesta (coeficiente c) puso de manifiesto la multidimensionalidad tanto de la tarea Go/No-Go como del cTMT. Mientras que el *accuracy* fue mejor predicho por métricas de la Parte B (como la velocidad y la distancia de búsqueda durante la tarea de alternancia), el sesgo de respuesta se asoció casi exclusivamente con métricas de la Parte A (por ejemplo, el tiempo de búsqueda y las transiciones de estado). Esta disociación sugiere que la secuenciación visomotora básica y la eficiencia de búsqueda capturadas por la Parte A están estrechamente vinculadas a la tendencia basal de un individuo a responder (sesgo de respuesta). Por el contrario, la capacidad de

inhibir exitosamente una respuesta prepotente bajo exigencia cognitiva (*accuracy*) depende en gran medida de la flexibilidad cognitiva y los recursos de alternancia adicionales indexados por la Parte B. En consecuencia, las características precisas de la trayectoria permitieron desenredar estos procesos inhibitorios superpuestos pero distintos.

En contraste, nuestros modelos predictivos no arrojaron resultados significativos para el Stop-Signal Reaction Time (SSRT). Este hallazgo nulo puede sugerir que la inhibición reactiva (Logan & Cowan, 1984) no comparte fuertemente las mismas firmas cinemáticas que la búsqueda visual continua y la planificación motora proactiva requeridas en el TMT. Si bien estudios previos han demostrado que las trayectorias motoras capturan ajustes espaciales y temporales sutiles durante paradigmas de control proactivo como la tarea Go/No-Go (Benedetti et al., 2020; Bravi et al., 2022), los ajustes dinámicos ensayo a ensayo inherentes al algoritmo de escalera adaptativa de la SST podrían introducir una variabilidad más difícil de mapear sobre las características de trayectoria globales y basadas en estados que extrajimos. Este resultado se alinea con la literatura que enfatiza que la inhibición reactiva y proactiva, aunque relacionadas, dependen de mecanismos temporales y neurales distintos (H. C. Meyer & Bucci, 2016), sugiriendo la necesidad de características de trayectoria diferentes para capturar el control reactivo.

Finalmente, los resultados de predicción del comportamiento del cTMT con la edad no solo obtuvieron los p-valores más bajos, sino que además todos los modelos lograron resultados significativos. Esto está en línea con resultados previos en el TMT de lápiz y papel (Tombaugh, 2004) y, junto con la fuerte asociación del rendimiento del TMT con varias funciones cognitivas (Woods et al., 2015), sugiere una relación de los atributos extraídos con el envejecimiento cerebral. Esto establece a las características de la mano del cTMT y al pipeline propuesto aquí como un potencial biomarcador del envejecimiento y enfermedades neurodegenerativas relacionadas. Los biomarcadores del envejecimiento derivados de técnicas como fMRI o EEG alcanzan niveles de desempeño comparables (Al Zoubi et al., 2018; Liem et al., 2017; Yu et al., 2024), pero requieren evaluaciones presenciales, dispositivos más costosos y una logística más compleja. Las revisiones sistemáticas enfatizan que la precisión de la predicción depende fuertemente de la modalidad, el tamaño muestral, el rango etario y la elección del modelo (More et al., 2023).

Entre los modelos de mejor rendimiento, el éxito de las regresiones lineales sugiere que el declive asociado a la edad se manifiesta de forma progresiva y proporcional en la dinámica del cursor. Como mostró nuestro análisis SHAP, las características de mayor impacto estuvieron relacionadas con métricas basadas en distancia (por ejemplo, la distancia de búsqueda completa y la distancia total) en ambas partes, A y B. El rol prominente de la longitud de la trayectoria respalda el fenómeno de sinuosidad (*circuitousness*; (Woods et al., 2015)) donde una mayor edad se caracteriza por una menor eficiencia espacial, movimientos exploratorios más largos y mayor ruido motor durante la navegación dirigida a metas, independientemente de la carga de alternancia cognitiva.

Limitaciones y direcciones futuras

El enfoque en línea nos permitió acceder a muestras más grandes y diversas que en el laboratorio, habilitando un primer paso en la evaluación de factibilidad de este tipo de sistemas en entornos remotos. Además, la implementación de un esquema de aprendizaje automático diseñado cuidadosamente para evitar problemas como por ejemplo *data leakage*, nos permite un primer acercamiento a la cuantificación de nuestras características derivadas del movimiento del mouse para

predecir diferentes funciones ejecutivas. No obstante, existen limitaciones que deben considerarse. En primer lugar, la variabilidad en el hardware y el rendimiento del navegador introdujo tasas de muestreo no uniformes de las trayectorias del mouse. En entornos en línea, los navegadores registran las posiciones del mouse a tasas más bajas que el sistema operativo, y la latencia de la red agrava aún más esta discontinuidad (T. Meyer et al., 2024). La interpolación lineal mitigó este problema siguiendo la práctica estándar (Spivey et al., 2005), aunque el trabajo futuro se beneficiaría de procedimientos de muestreo más estandarizados, ya que las tasas de muestreo siguen siendo muy variables y poco reportadas entre estudios (Schoemann et al., 2021). En segundo lugar, nuestra muestra consistió mayoritariamente en adultos jóvenes y de mediana edad con alto nivel educativo, lo que limita la generalización a poblaciones clínicas o cohortes de mayor edad con alto riesgo de neurodegeneración. Cabe señalar que esta distribución también impidió analizar la influencia de la reserva cognitiva asociada a la educación, dada la sobrerrepresentación de niveles educativos altos en la muestra.

Estudios futuros deberían expandir la variabilidad de la muestra, validar los modelos sobre un conjunto de prueba independiente (*held-out set*), explorar la confiabilidad test-retest de estas características y evaluar su sensibilidad en diseños longitudinales. Adicionalmente, integrar estas métricas con otros flujos de datos, como dinámicas de pulsaciones de teclas o evaluaciones ecológicas multimodales (incorporando movimientos oculares o habla), podría mejorar la precisión predictiva de los modelos cognitivos digitales. En particular, los movimientos oculares podrían proporcionar una medida más directa de las fases identificadas en la segmentación de trayectorias (exploración, monitoreo, planificación) y también de comportamientos directamente relacionados con la inhibición o la memoria de trabajo como se mostró en el **Capítulo 1** y en otros trabajos de la literatura (Linari et al., 2022; Wölwer et al., 2003, 2012). Desarrollos recientes han habilitado el registro de movimientos oculares en línea, aunque se necesitan mejoras adicionales en la precisión para mediciones confiables en el TMT (M. X. Huang et al., 2016; Juantorena et al., 2023; Papoutsaki et al., 2016; Xu et al., 2015), tema que abordamos en el **Capítulo 6**.

Conclusiones

Al ir más allá del tiempo de completado, mostramos que el cTMT en línea puede servir como una herramienta poderosa de administración remota para evaluar los matices del control ejecutivo, y también puede ser útil para rastrear el declive cognitivo asociado a la edad. Nuestra capacidad de predecir capacidades específicas, como la memoria de trabajo, procesos inhibitorios distintos y la edad, basándonos en trayectorias del cursor de alta resolución, destaca la riqueza de la señal presente en el comportamiento motor continuo. En última instancia, aprovechar esta complejidad en entornos no supervisados o de administración remota, tiene un potencial significativo para el futuro de la psiquiatría computacional, las evaluaciones remotas escalables y la neuropsicología personalizada.

Capítulo 6: Seguimiento ocular basado en web

Este capítulo se basa en el siguiente trabajo:

Juantorena, G. E., Figari, F., Petroni, A., & Kamienkowski, J. E. (2023). Web-based eye-tracking for remote cognitive assessments: The anti-saccade task as a case study. bioRxiv, 2023-07.

Este capítulo también incluye trabajo de las siguientes tesis de grado, de las cuales fui co-director:

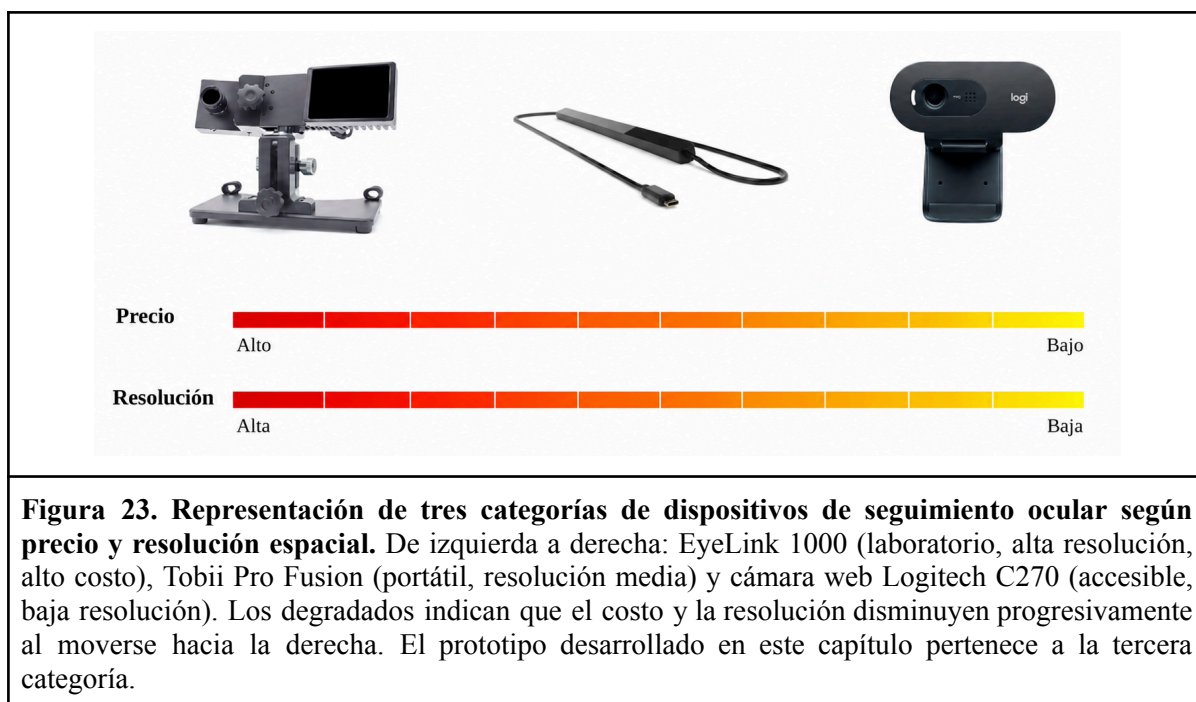
Francisco Figari, “Aplicabilidad de eye tracking web remoto al diagnóstico de condiciones neuropsicológicas”, Tesis de Licenciatura en Ciencias de la Computación (DC, FCEyN, UBA, 2022).

Agustín Penas, “Eye-tracking remoto: detección automática de eventos y mejoras en la performance”, Tesis de Licenciatura en Ciencias de la Computación (DC, FCEyN, UBA, 2026).

Introducción y motivación

En los **Capítulos 1 y 2** demostramos que el registro de movimientos oculares durante el cTMT aporta información única sobre los procesos ejecutivos involucrados en la tarea: el número de fijaciones y los patrones de exploración visual discriminaron entre grupos donde las métricas clásicas no lo lograron. Sin embargo, la adquisición de datos de *eye tracking* en poblaciones clínicas presentó desafíos importantes de escalabilidad, tanto en términos del costo del hardware como de la necesidad de presencia física en el laboratorio. En el **Capítulo 5**, dimos un primer paso hacia la evaluación remota extendiendo el cTMT a un entorno online a través del registro de trayectorias del cursor, haciendo uso de la infraestructura de toma de datos online que desarrollamos en el **Capítulo 4** (*DataPruebas*). El presente capítulo completa ese recorrido: construimos sobre software existente y validamos un sistema de *eye tracking* con cámara web que permitió registrar movimientos oculares de forma remota y sin hardware especializado, habilitando la visión multimodal con la que abrimos esta tesis. Durante este capítulo se definirán las mejoras implementadas, los experimentos para su evaluación, un caso mínimo viable para la evaluación neuropsicológica remota haciendo uso de la tarea de antisacadas (Hallett, 1978), y las limitaciones que encontramos en el desarrollo de este enfoque.

Una posible taxonomía de los dispositivos de *eye tracking* (ET) puede hacerse teniendo en cuenta su resolución espacial y temporal (Fig. 23): equipos de laboratorio de alta resolución (como EyeLink-1000, *SR Research*), equipos portátiles de resolución media (como Tobii Pro Fusion, *Tobii AB*), y sistemas basados en cámara web estándar. Esta tercera categoría, que incluye el prototipo presentado en este capítulo, elimina la necesidad de hardware especializado y permite que cualquier computadora con cámara web realice seguimiento ocular de forma asíncrona y sin supervisión presencial (M. X. Huang et al., 2016; Hutt et al., 2023; Papoutsaki et al., 2016; Semmelmann & Weigelt, 2018; Slim & Hartsuiker, 2022; vanWell & Tanaka, 2022; Xu et al., 2015; Yang & Krajchich, 2020). Aunque presenta limitaciones inherentes de resolución espacial y frecuencia de muestreo, podría ser suficiente para ciertas tareas cognitivas.



Con esta tecnología, no es necesario contar con un laboratorio ni con la supervisión de un investigador (presencial o en línea), permitiendo, además obtener datos de manera simultánea de un gran número de participantes. Estos avances implican que cualquier computadora estándar (portátil o de escritorio) con cámara web puede realizar ET, habilitando a cualquier potencial participante acceso a esta tecnología de seguimiento ocular y abriendo las puertas a posibles aplicaciones web, como por ejemplo su uso en telemedicina. Por supuesto, este tipo de sistema de ET tiene importantes limitaciones, como una menor tasa de fotogramas (*frames per second* o *FPS*), una menor resolución espacial y una menor precisión. No obstante, para algunas tareas cognitivas sencillas de ET basta con contar con una resolución moderada, por lo que son buenas candidatas para este tipo de configuraciones.

Los objetivos de este capítulo son: evaluar el estado del arte de las implementaciones de ET web, desarrollar y mejorar un prototipo funcional, y validarlo con una tarea neuropsicológica clásica (antisacadas; Munoz & Everling, 2004) buscando replicar efectos bien documentados en la literatura.

Entre las opciones de software existentes para utilizar como base, elegimos *WebGazer.js* (Papoutsaki et al., 2016) una biblioteca de código abierto, escrita totalmente en el lenguaje *JavaScript* y diseñada para operar directamente en navegadores web. *WebGazer* aplica técnicas de regresión y aprendizaje automático para mapear puntos faciales (*facial landmarks*) del usuario y estimar la posición de la mirada en tiempo real. Esta tecnología ha demostrado ser particularmente útil en escenarios educativos, como la evaluación de la atención en estudiantes, y en estudios de interacción humano-computadora (Madsen et al., 2021; Papoutsaki et al., 2016). Además de las razones antes expuestas, tiene una gran adhesión por la comunidad (más de 600 citas y aproximadamente 3900 estrellas en Github³) y ha sido incluida en diversas aplicaciones como también puede observarse en base a la actividad de su repositorio.

³ <https://github.com/brownhci/WebGazer>

A su vez, esta herramienta no está exenta de aspectos negativos, por un lado *WebGazer* está influenciado por varios factores externos, como la calidad de la cámara, las condiciones de iluminación y el movimiento del usuario. Estos factores pueden afectar significativamente la precisión del seguimiento ocular y limitar su aplicabilidad en entornos reales. En cuanto a la precisión y robustez, *WebGazer* es inferior a los dispositivos tradicionales debido principalmente a la dependencia de cámaras de baja resolución, la falta de iluminación específica (falta de emisores infrarrojos) y el uso de algoritmos menos sofisticados. Esto resulta en un rendimiento inconsistente, especialmente en tareas muy demandantes en tiempo real.

Otro aspecto crítico es la robustez frente a los movimientos del usuario. Los cambios en la postura y los desplazamientos de la cabeza tienden a desviar la calibración del sistema, reduciendo la fiabilidad de las predicciones en sesiones prolongadas. Si bien algunos sistemas comerciales incorporan algoritmos de compensación para mantener la precisión ante estos movimientos, *WebGazer* carece de mecanismos equivalentes, lo que limita la calidad de la señal adquirida. Asimismo, *WebGazer* no implementa de manera nativa la detección de parpadeos (*blink detection*) ni la estimación de la distancia a la pantalla, dos características comunes en los dispositivos estándar. La ausencia de estas métricas limita tanto la evaluación de la calidad de los datos recolectados como su aplicabilidad en estudios de atención, fatiga cognitiva o ergonomía visual (Biondi et al., 2023; Maffei & Angrilli, 2018).

En términos de frecuencia de muestreo, este sistema también se encuentra en desventaja. Mientras que los dispositivos de gama media alcanzan entre 60-200 Hz y los de alta alrededor de 1000 Hz, al evaluar esta biblioteca estaban en el entorno de los 13-19 predicciones por segundo en pruebas de laboratorio. Esta limitación compromete su fluidez y restringe su uso en aplicaciones que requieren respuestas en tiempo real o alta interactividad como en experimentos de control motor. Siendo un límite superior realista los 30 Hz de frecuencia de muestreo al que llegan la mayoría de las cámaras web.

En síntesis, aunque *WebGazer* constituye un aporte significativo hacia la democratización del ET, presenta limitaciones estructurales en precisión, robustez y funcionalidad. Estos desafíos han motivado múltiples propuestas de mejora orientadas a ampliar su aplicabilidad en entornos reales sin sacrificar la accesibilidad que caracteriza a las tecnologías web, y abren la puerta a aplicaciones menos exploradas en telemedicina, diagnóstico remoto y recolección masiva de datos.

Para que el *eye tracking* web trascienda las pruebas de concepto, sin embargo, no basta con mejorar el software: es necesario demostrar que puede replicar efectos neurocognitivos bien documentados en condiciones de baja resolución y ruido ambiental.

La tarea de antisacadas (Hutton & Ettinger, 2006; Munoz & Everling, 2004) ofrece un escenario de prueba ideal para ello. Este paradigma contrasta dos condiciones: la prosacada (mirar de forma refleja hacia un estímulo) y la antisacada (inhibir ese reflejo y dirigir la mirada voluntariamente hacia el lado opuesto). La diferencia entre ambas condiciones se conoce como costo de antisacada, y se manifiesta en mayor latencia y mayor tasa de errores en la condición inhibitoria.

La elección de esta tarea no es arbitraria. A nivel cerebral, el costo de antisacada refleja la competencia entre el sistema de orientación reflejo, mediado por el colículo superior, y el sistema de control voluntario arriba hacia abajo (*top-down*, en idioma inglés), dependiente de la corteza prefrontal dorsolateral (Munoz & Everling, 2004). Debido a esta dependencia de redes frontales, la

tarea es altamente sensible a disfunciones ejecutivas y se utiliza frecuentemente como biomarcador en patologías como la esquizofrenia, enfermedades neurodegenerativas y trastornos de atención.

Por lo tanto, la tarea de antisacadas funciona como una prueba de estrés ideal para esta tecnología. El sistema debe ser lo suficientemente sensible para discriminar diferencias sutiles en los tiempos de reacción, generalmente en el rango de las decenas de milisegundos ($\approx 40\text{-}80$ ms), y lo suficientemente preciso espacialmente para distinguir una sacada correcta de un error de dirección, todo ello a pesar de la baja resolución temporal y el ruido inherente a una cámara web estándar. Si el prototipo propuesto logra recuperar el patrón de costos (latencias mayores y más errores en antisacadas) con significancia estadística, constituye una evidencia de que esta clase de herramientas puede ser viable para evaluaciones neuropsicológicas remotas.

Desarrollo e implementación del prototipo

Arquitectura base y selección de herramientas

El desarrollo del prototipo (ver **Apéndice A.6**) partió de la biblioteca de código abierto *WebGazer.js* (Papoutsaki et al., 2016). Esta elección se fundamentó en su capacidad para resolver los pasos necesarios íntegramente en el navegador (*client-side*), eliminando la latencia de transmisión de video a un servidor y garantizando la privacidad de los datos biométricos del usuario, sin que los mismos salgan nunca de su computadora. Siendo los únicos datos almacenados las estimaciones de la posición de la mirada en la pantalla y su correspondientes tiempos asociados.

El funcionamiento del prototipo sigue un *pipeline* de cinco etapas (Fig. 24). En primer lugar, se captura el video desde la cámara del usuario a través del navegador web. Luego, para cada cuadro de video, se detecta la región facial utilizando un modelo de malla facial (*Face Mesh*, en inglés) implementado en *TensorFlow.js* (Smilkov et al., 2019), que identifica 468 puntos de referencia distribuidos sobre el rostro. A continuación, estos puntos faciales se emplean para localizar y recortar las regiones correspondientes a ambos ojos, las cuales son posteriormente normalizadas para reducir la variabilidad asociada a la posición y escala del rostro.

En la cuarta etapa, las imágenes oculares son preprocesadas mediante conversión a escala de grises, ecualización de histograma para atenuar diferencias de iluminación y redimensionamiento a una resolución de 6×10 píxeles. Finalmente, las matrices correspondientes a ambos ojos se aplanan y concatenan para formar un vector de 120 características, que se utiliza como entrada de un modelo de regresión *Ridge* encargado de estimar las coordenadas horizontales y verticales de la mirada (x, y) sobre la pantalla.

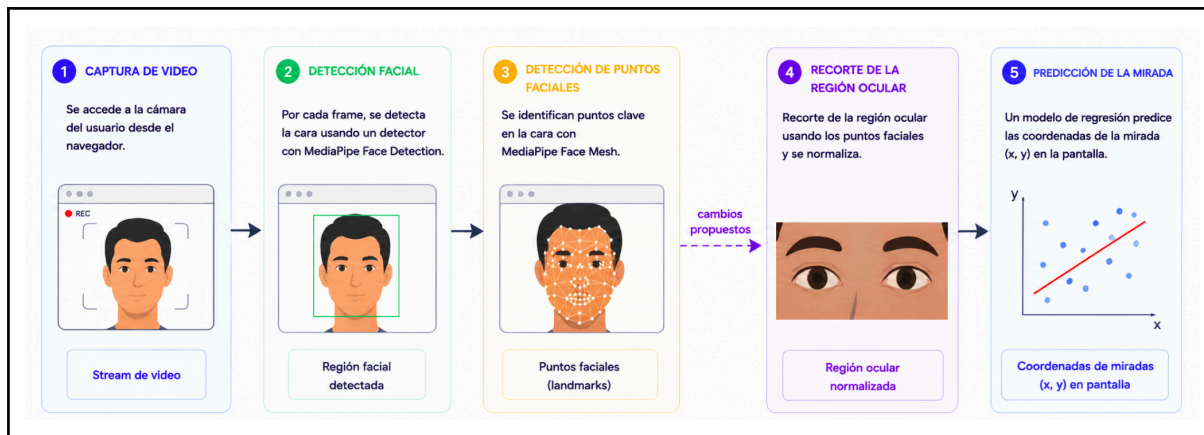


Figura 24. Pipeline de estimación de la mirada a partir de video facial. El proceso comienza con la captura de video desde la cámara del usuario (1), seguida de la detección de la región facial en cada cuadro mediante MediaPipe Face Detection (2). A continuación, se identifican puntos faciales clave utilizando MediaPipe Face Mesh (3), que se emplean para recortar y normalizar la región ocular (4). Finalmente, un modelo de regresión estima las coordenadas de la mirada en la pantalla a partir de las imágenes oculares procesadas (5).

Optimización y adaptación al contexto experimental

Para transformar la biblioteca base en una herramienta apta para la investigación clínica, fue necesario realizar cuatro modificaciones estructurales críticas.

En primer lugar, se abordó la estabilidad temporal. La implementación original dependía de un modelo de malla facial deprecado que sufría fallos de memoria (*crashes*) en sesiones superiores a los 10 minutos. Dado que para el buen desarrollo del protocolo de antisacadas se recomienda una duración mayor a ese límite (Antoniades et al., 2013), se migró el núcleo de detección a una versión actualizada y más estable de la biblioteca que modela los puntos de la cara⁴. Esto no solo eliminó los fallos silenciosos, sino que requirió la reescritura de la lógica de extracción de coordenadas oculares, una mejora que fue posteriormente incorporada al repositorio oficial de *WebGazer*⁵.

En segundo lugar, se reemplazó el protocolo de calibración. El método original de calibración continua basado en la interacción con el mouse resultaba inadecuado para tareas cognitivas pasivas donde no existe tal interacción (Fig. 25.A). Se implementó una fase de calibración explícita de 9 puntos, similar a la utilizada en equipos comerciales (e.g., EyeLink, Tobii), donde el sujeto fija la mirada y confirma mediante una tecla (Fig. 25.B). A nivel computacional, esto permitió optimizar el rendimiento: en lugar de recalcular los coeficientes de regresión en cada cuadro (*frame*), estos se computan una única vez al finalizar la fase de calibración.

⁴ Específicamente, se reemplazaron los paquetes deprecados originales por `@tensorflow-models/face-landmarks-detection v0.0.3` y `@tensorflow/tfjs v3.13.0`, que soluciona los problemas de gestión de memoria observados en versiones anteriores.

⁵ <https://github.com/brownhci/WebGazer/issues/171#issuecomment-1078676817>

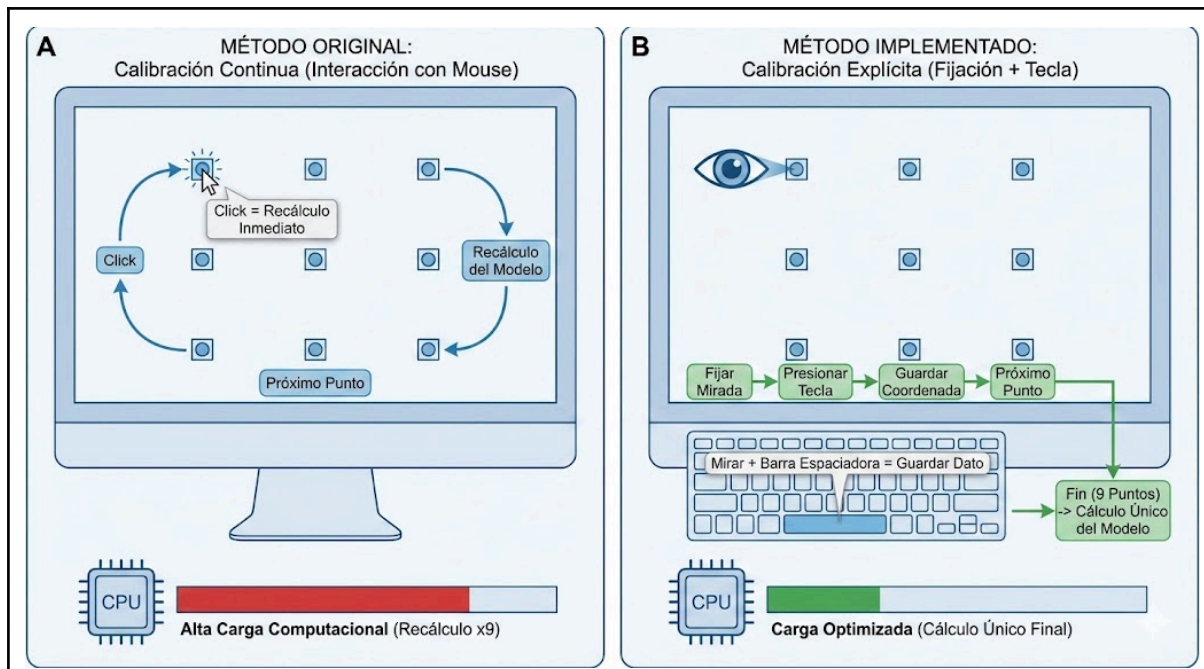


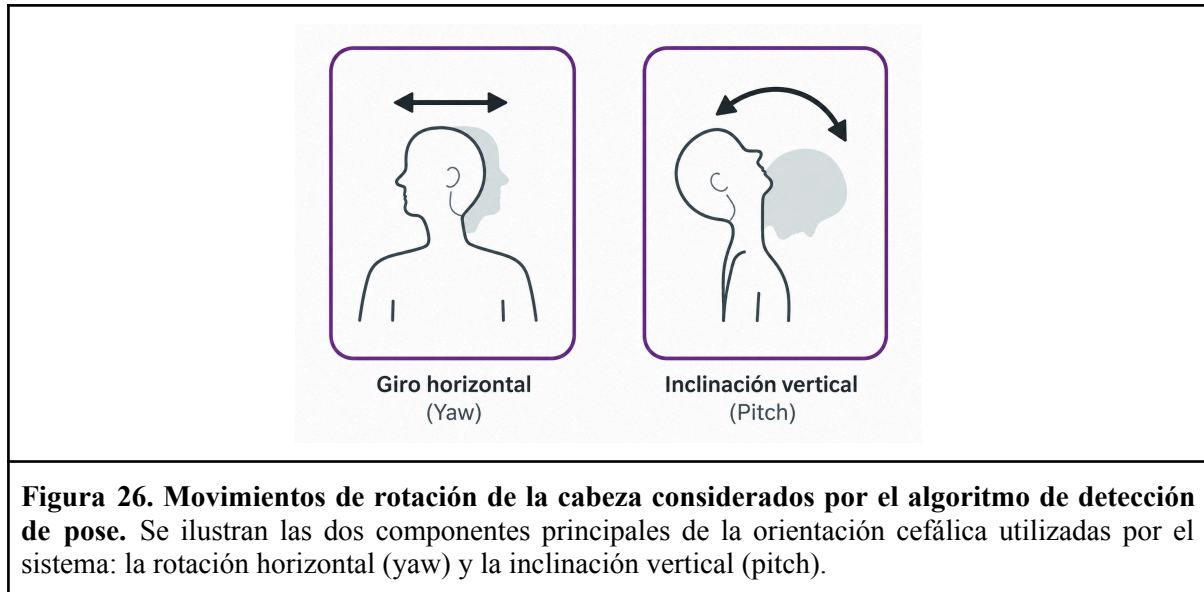
Figura 25. Comparación entre el método de calibración original de *WebGazer* y el implementado en el prototipo. A. La calibración continua recalcula el modelo de regresión en cada clic del usuario, generando alta carga computacional. **B.** La calibración explícita de 9 puntos requiere que el participante fije la mirada y presione una tecla; el modelo se computa una única vez al finalizar, reduciendo la carga de procesamiento y eliminando la dependencia de la interacción con el mouse.

En tercer lugar, se optimizó el rendimiento del sistema para sostener una tasa de muestreo estable en tiempo real. Un análisis de los cuellos de botella del ciclo de procesamiento identificó al cálculo de la malla facial como la operación más costosa. Dado que la cámara entrega alrededor de 30 cuadros por segundo, ejecutar ese cálculo a mayor frecuencia implicaba procesar repetidamente la misma imagen sin ganancia alguna. Por ello se desacopló el cómputo del bucle principal y se lo sincronizó a intervalos fijos de 34 ms (equivalentes a unas 30 veces por segundo), alineándose con la frecuencia de actualización de la cámara y eliminando llamadas redundantes.

En cuarto lugar, se desarrolló un módulo de monitoreo del comportamiento del usuario para la validación continua de la calidad del dato. Mientras que las primeras iteraciones dependían de heurísticas posicionales que alertaban sobre cambios en la posición absoluta de los ojos, la versión final incorporó tres submódulos de detección basados en geometría facial relativa: rotación de cabeza, parpadeos y distancia a pantalla.

Detección de movimiento de cabeza

Para la estimación de la rotación de cabeza, se reemplazó el criterio de desplazamiento por un algoritmo geométrico basado en la simetría de puntos de referencia faciales. Al monitorear la relación entre distancias clave, como la asimetría entre la nariz y ambas orejas, el sistema logró inferir movimientos de rotación (*yaw*) e inclinación (*pitch*) sin incurrir en el costo computacional de un modelado 3D completo (Fig. 26).



Detección de parpadeos

Se implementó un análisis comparativo de cuatro estrategias algorítmicas. La distancia euclidiana mide la separación vertical entre los párpados superior e inferior cuadro a cuadro; la correlación de cuadros compara la similitud entre fotogramas consecutivos de la región ocular, requiriendo una ventana de fotogramas almacenados; la intensidad de color opera de forma similar pero reemplaza la correlación por la suma de valores de píxel en esa misma ventana; y la relación de aspecto ocular (EAR, por sus siglas en inglés) cuantifica la apertura del ojo mediante la geometría de sus puntos de referencia faciales (Soukupova & Čech, 2016).

El EAR se calcula a partir de las distancias euclidianas entre puntos faciales clave que delimitan el ojo como se puede observar en la ecuación 3:

$$EAR = \frac{||P_2 - P_6|| + ||P_3 - P_5||}{2 \cdot ||P_1 - P_4||} \quad (3)$$

donde P_1 y P_4 son los extremos horizontales del ojo, y P_2 , P_3 , P_5 , P_6 , son puntos del párpado superior e inferior. Al ser una relación de distancias, el EAR permanece relativamente estable cuando el ojo está abierto y cae abruptamente hacia cero al cerrarse, produciendo un pico detectable en la señal temporal.

La Tabla 6 resume las principales ventajas y desventajas de los cuatro métodos implementados:

Tabla 6. Comparación cualitativa de los métodos de detección de parpadeo.

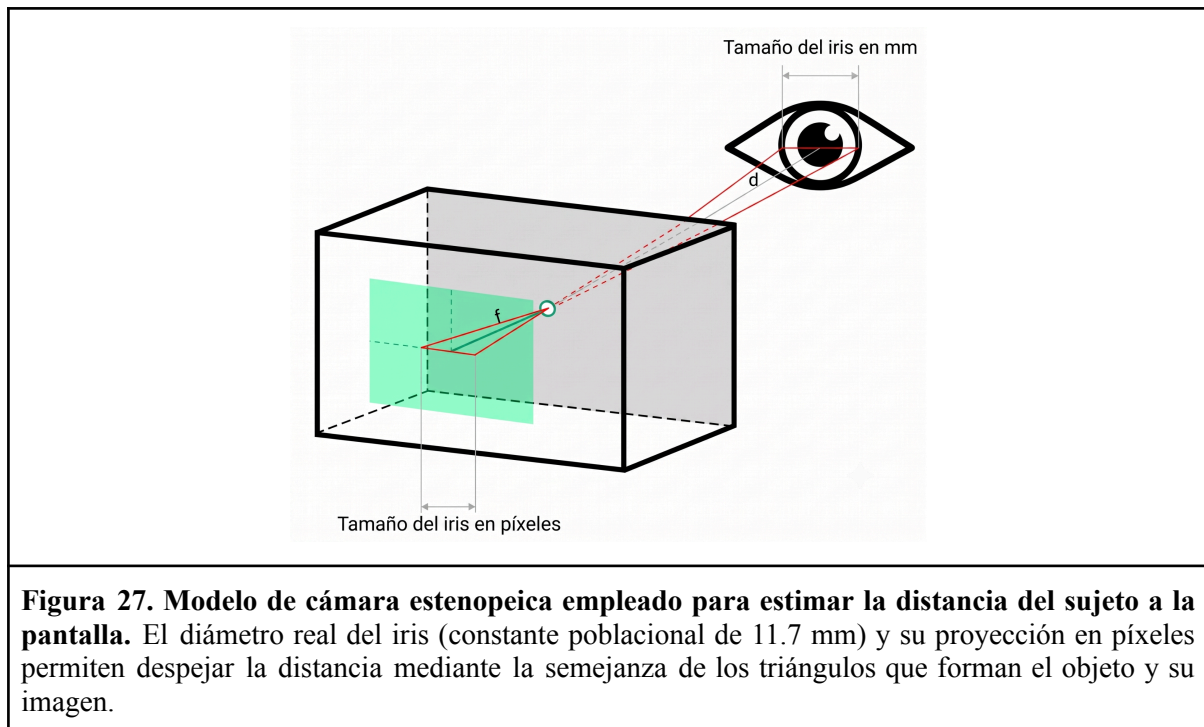
Método	Ventajas	Desventajas
EAR	Muy utilizado y robusto	Dependencia de landmarks precisos
Distancia euclidiana	Simple de implementar	Sensibilidad a rotaciones de cabeza
Correlación	Independiente de puntos de referencia (<i>landmarks</i>)	No detecta cierres oculares prolongados
Color	Computacionalmente simple	Menor robustez ante variaciones de iluminación

Estimación de distancia a pantalla

Se implementó un módulo de estimación continua de la distancia a la pantalla. El método se basa en el modelo de cámara estenopeica (Fig. 27), que relaciona el tamaño real de un objeto con su proyección en píxeles mediante la ecuación 4:

$$H = \frac{D}{D'} \cdot H' \quad (4)$$

donde H es la distancia del sujeto a la cámara (incógnita), D es el diámetro real del iris (constante poblacional de 11.7 ± 0.5 mm), D' es el diámetro del iris medido en píxeles en la imagen, y H' es la distancia focal de la cámara expresada en píxeles. La única variable desconocida es H' , ya que los navegadores web no exponen los parámetros ópticos del hardware de captura. Como solución práctica, se utilizó un valor aproximado derivado empíricamente a partir de la calibración experimental de una cámara de referencia. Si bien esta estimación no reemplaza a un sistema calibrado con parámetros ópticos exactos, provee una medida relativa útil: al monitorear cambios en H a lo largo del tiempo, el módulo detecta desplazamientos significativos del sujeto respecto a la pantalla que podrían comprometer la calibración, permitiendo descartarlos o corregirlos sin requerir intervención del usuario.



Finalmente, se empleó el *plugin-virtual-chinrest* de *jsPsych* (Juantorena & Kohler, 2021), basado en el procedimiento de Li et al. (2020), para estandarizar las dimensiones de los estímulos y convertir las coordenadas de la mirada a grados de ángulo visual, asegurando la consistencia metodológica con los capítulos previos.

Validación técnica de los módulos de monitoreo

Con el objetivo de ajustar los algoritmos de detección de parpadeos, movimientos de cabeza y estimación de distancia, se diseñaron una serie de experimentos controlados. En cada sesión, los participantes realizaron tareas específicas frente a su cámara web. Es importante destacar que no se almacenaron las grabaciones de video, sino únicamente las coordenadas de los puntos faciales generados en tiempo real.

Estos datos numéricos fueron procesados posteriormente en un entorno *offline* mediante scripts de *Python* que replicaban la lógica de los algoritmos de *WebGazer*. Esta metodología permitió "reproducir" los experimentos múltiples veces modificando sistemáticamente los umbrales de decisión (thresholds). De esta forma, fue posible evaluar el desempeño de los algoritmos bajo distintas configuraciones y encontrar los parámetros que ofrecían el mejor balance entre precisión y estabilidad, sin necesidad de volver a recolectar datos.

Todos los participantes tenían visión normal o corregida. En lugar de buscar una muestra demográficamente representativa, se priorizó, cualitativamente, la diversidad: diferentes computadoras, calidades de cámara y condiciones de iluminación. El propósito fue poner a prueba la robustez de los algoritmos frente a la variabilidad real que enfrentaría la herramienta en una implementación web masiva.

La validación combinó una fase presencial de calibración y una fase online de despliegue. La fase presencial, realizada en condiciones controladas y sobre un mismo equipo, permitió comparar de manera homogénea las distintas implementaciones algorítmicas y seleccionar las más robustas. La fase online, en cambio, expuso a los algoritmos a la variabilidad real de hardware, iluminación y postura que enfrentaría la herramienta en un despliegue masivo. A continuación se describen los procedimientos específicos de cada módulo. El número de participantes difirió entre módulos y entre modalidades; en particular, en la validación online de la detección de parpadeos se excluyeron algunos casos con detección casi nula, según se detalla en los resultados correspondientes.

Conviene aclarar que estas validaciones tuvieron un carácter exploratorio, a modo de prueba piloto de las modificaciones introducidas a *WebGazer*. El objetivo no fue contrastar hipótesis sino caracterizar el comportamiento de cada módulo y seleccionar sus parámetros, por lo que los análisis presentados a continuación son de naturaleza descriptiva y no incluyen pruebas de estadística inferencial.

Experimento de detección de parpadeos

Con el objetivo de calibrar y comparar la robustez de distintos algoritmos de detección de parpadeo, se diseñó un protocolo de tareas controladas adaptado de Witte y Lins (2023). El protocolo varía de manera sistemática la dinámica del parpadeo y las condiciones de movimiento: incluye parpadeos voluntarios rápidos y relajados, cierres oculares prolongados de 2 segundos, y secuencias realizadas tanto en condiciones estáticas como dinámicas, esto es, acompañadas de movimientos de la mirada y rotaciones de cabeza (Tabla 7). La cantidad de parpadeos a ejecutar en cada prueba era conocida de antemano y se utilizó como valor de referencia (*ground truth*) contra el cual contrastar las detecciones de cada algoritmo.

Tabla 7. Protocolo de pruebas para la validación de la detección de parpadeo (adaptado de (Witte & Lins, 2023).

Prueba	Instrucción y objetivo
1	Parpadear 5 veces rápido, esperar 2 s y repetir (control de velocidad).
2	Parpadear 5 veces de forma relajada, esperar 2 s y repetir (control de naturalidad).
3	Cerrar los ojos 2 s, esperar 4 s y volver a cerrarlos 2 s (control de cierre prolongado).
4	Fijar la mirada y parpadear 10 veces (sin movimiento de cabeza).
5	Mover la mirada sin mover la cabeza y parpadear 10 veces (movimiento ocular).
6	Mover la cabeza y la mirada mientras se parpadea 10 veces (movimiento combinado).

El experimento se ejecutó en las dos modalidades mencionadas. En la presencial participaron 14 voluntarios (7 mujeres y 7 varones, de 25 a 65 años) sobre la misma computadora, lo que permitió comparar de forma controlada las cuatro implementaciones: EAR, Distancia Euclidiana, Correlación e Intensidad de color. En la versión online, 24 participantes completaron la tarea desde sus propios dispositivos, en condiciones heterogéneas de hardware e iluminación. En esta segunda instancia se procesaron únicamente los métodos EAR y Distancia Euclidiana, dado que las implementaciones basadas en Correlación e Intensidad de color requieren almacenar los píxeles de la región ocular, lo que generaría archivos de gran tamaño y comprometería el resguardo de datos sensibles.

El desempeño de cada algoritmo se cuantificó mediante dos métricas de error ampliamente utilizadas: el Error Absoluto Medio (EAM, equivalente al MAE empleado en los **Capítulos 3 y 5**), que promedia en valor absoluto la diferencia entre la cantidad de parpadeos detectada y la real, y el Error Cuadrático Medio (ECM), que penaliza con mayor peso los errores grandes. El procesamiento offline de las marcas faciales permitió reevaluar cada algoritmo bajo múltiples configuraciones de umbrales, determinando el valor que minimizaba el error de detección (EAM y ECM) para cada método.

Experimento de estimación de distancia a la pantalla

Para validar el módulo de estimación de distancia se contrastó el método de estimación continua propuesto, basado en el modelo de cámara estenopeica descrito previamente, contra el procedimiento de referencia del *Virtual Chinrest* (Li et al., 2020), que infiere la distancia a partir de la localización del punto ciego (Durgin et al., 1995). Mientras el participante completaba tres ejecuciones consecutivas del *Virtual Chinrest*, el sistema estimaba en paralelo la distancia a la pantalla cuadro a cuadro. Este diseño permitió comparar ambos enfoques bajo las mismas condiciones y, sobre todo, evaluar la estabilidad temporal de la estimación a lo largo de la sesión.

Experimento de detección de rotaciones de cabeza

El módulo de detección de rotaciones de cabeza se validó mediante un protocolo de tareas guiadas por video. En cada prueba se presentaba al participante un video instructivo que indicaba una rotación específica; el participante presionaba la barra espaciadora, ejecutaba el movimiento solicitado y volvía a presionar la barra al finalizar, garantizando la sincronización entre la ejecución del movimiento y el registro de los datos. Se evaluaron ocho direcciones: hacia cada hombro y hacia puntos intermedios entre el frente y cada hombro (rotación horizontal o *yaw*), y hacia arriba y hacia abajo en dos magnitudes, mirando por encima o por debajo del monitor y, de forma más pronunciada, hacia el techo o hacia el ombligo (inclinación vertical o *pitch*). Las rotaciones hasta puntos intermedios se consideraron parciales y el resto, totales, lo que permitió evaluar no solo la detección del movimiento sino también la capacidad del sistema para discriminar su magnitud.

Resultados

Desempeño del sistema y de los módulos de monitoreo

Antes de evaluar el desempeño neurocognitivo, se analizaron los resultados de la optimización del sistema y la validación de los algoritmos de monitoreo en entornos controlados y remotos.

Optimización del rendimiento (*frame rate*)

Para identificar los cuellos de botella, se utilizó la herramienta *Performance Insights* de Google Chrome, que permite analizar el tiempo de cómputo de cada componente durante la ejecución. Este análisis señaló al método *TFFaceMesh.getEyePatches* como el de mayor costo computacional por iteración. El diagnóstico reveló que ejecutarlo en cada ciclo del bucle principal era redundante: dado que la cámara provee aproximadamente 30 imágenes por segundo, correrlo con mayor frecuencia implicaría procesar repetidamente la misma imagen. La solución fue desacoplar su invocación del bucle continuo y reemplazarla por una ejecución periódica mediante *setInterval*, función nativa de *JavaScript* que ejecuta un bloque de código a intervalos temporales fijos, configurada a 34 ms (equivalente a ~30 Hz), en sincronía con el flujo de video.

El análisis de rendimiento confirmó la efectividad de las estrategias de optimización implementadas sobre el ciclo de procesamiento de *WebGazer* (Fig. 28). En las versiones iniciales (sin

optimización), la tasa de predicción era irregular, ubicándose en un promedio de 16 a 18 predicciones por segundo. Tras desacoplar el cálculo de la malla facial (*facemesh*) del ciclo principal y sincronizarlo a intervalos de 34 ms, la frecuencia de muestreo se estabilizó notablemente, alcanzando de forma consistente entre 23 y 27 predicciones por segundo. Esta mejora validó la capacidad del sistema para operar en tiempo real con la fluidez necesaria para tareas cognitivas.

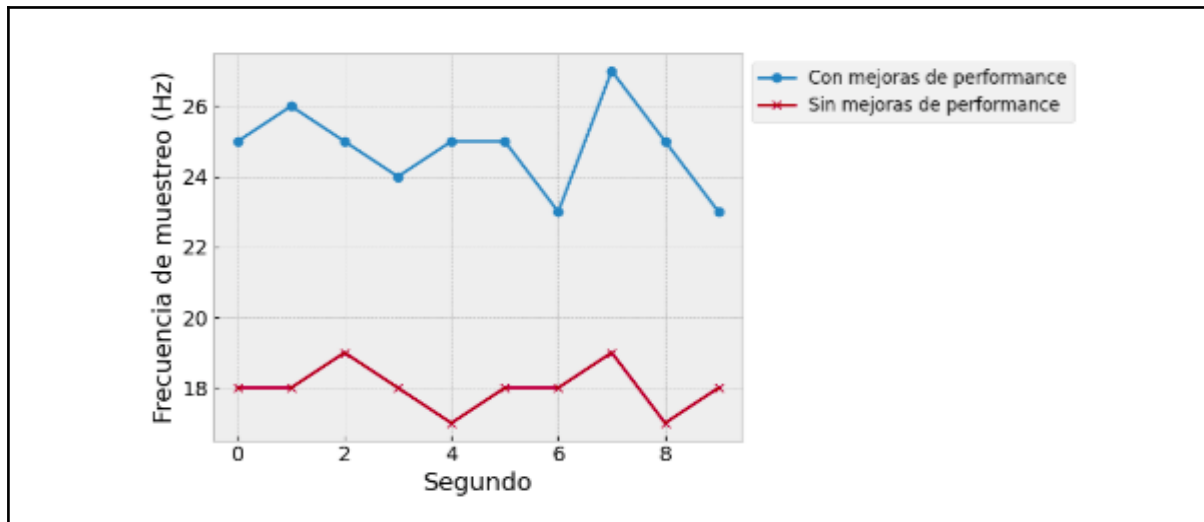
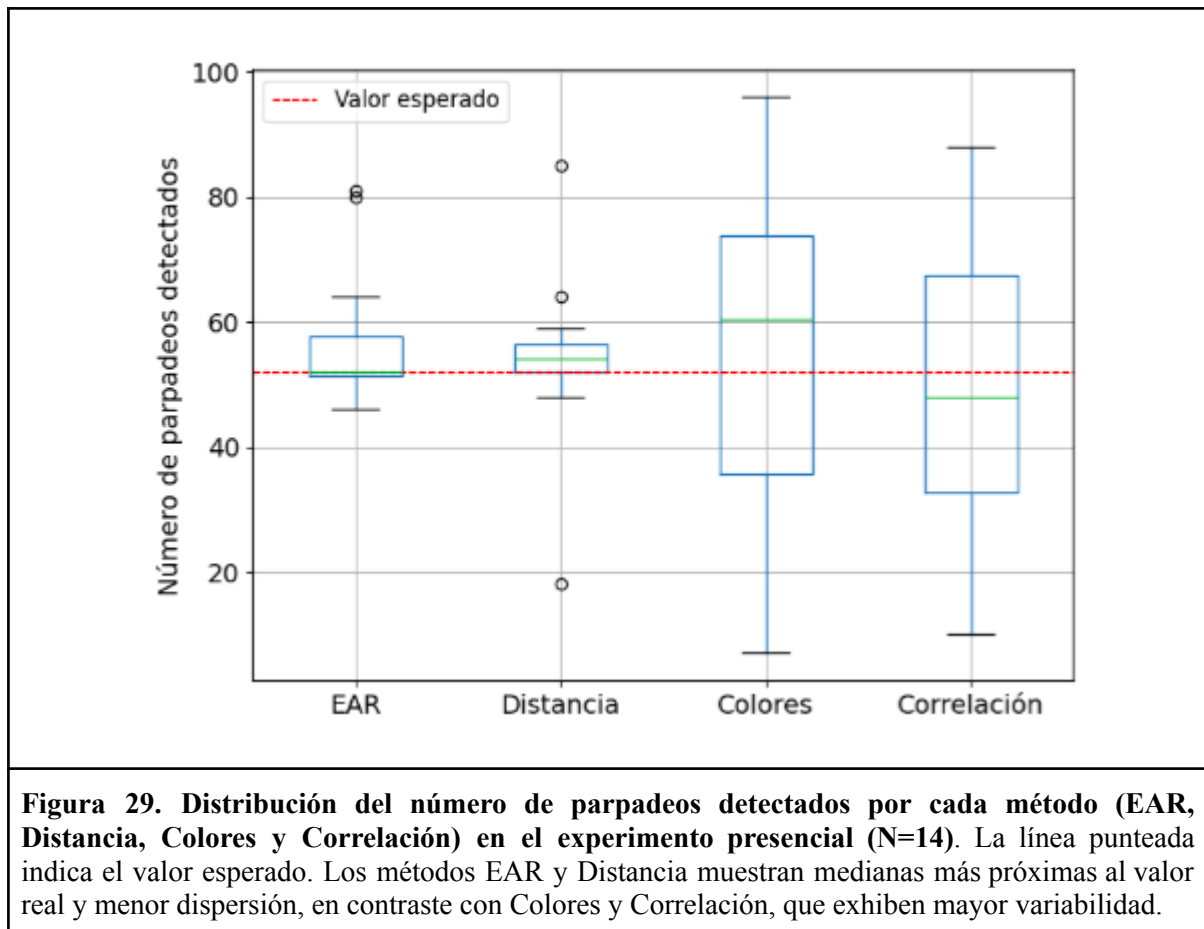


Figura 28. Frecuencia de muestreo (Hz) con y sin las optimizaciones de rendimiento, durante los primeros diez segundos de un mismo experimento ejecutado en la misma computadora. Tras desacoplar el cálculo del *facemesh* del bucle principal y sincronizarlo a la frecuencia de la cámara, la tasa de predicción aumentó de unas 16-18 Hz a unas 23-27 Hz de forma estable.

Detección de Parpadeo: Comparación de Algoritmos

Se evaluaron cuatro métodos de detección (EAR, Distancia, Colores y Correlación) para determinar cuál de ellos ofrecía mayor robustez. Los resultados experimentales indicaron una superioridad clara del método basado en la Relación de Aspecto Ocular (EAR).

En el experimento presencial, el método EAR presentó el menor error absoluto medio (7.57) y la menor dispersión (10.36) en comparación con los métodos basados en Colores y Correlación, cuyos errores promedio ascendieron a 22.57 y 19.00 respectivamente (Fig. 29).



En la validación online (N=24), en condiciones heterogéneas de hardware, el EAR mantuvo su estabilidad, registrando un error promedio de 2.04 parpadeos (DE = 2.96). El método de Distancia Euclidiana se posicionó como una alternativa válida aunque ligeramente inferior (Error promedio = 2.64; DE = 3.49; Fig. 30). Estos valores excluyen a 3 participantes que registraron una cantidad de parpadeos nula o casi nula de manera consistente en ambos métodos, atribuible a fallas de captura propias de un entorno remoto no supervisado (iluminación insuficiente, oclusión de la cámara o falta de adherencia a la consigna).

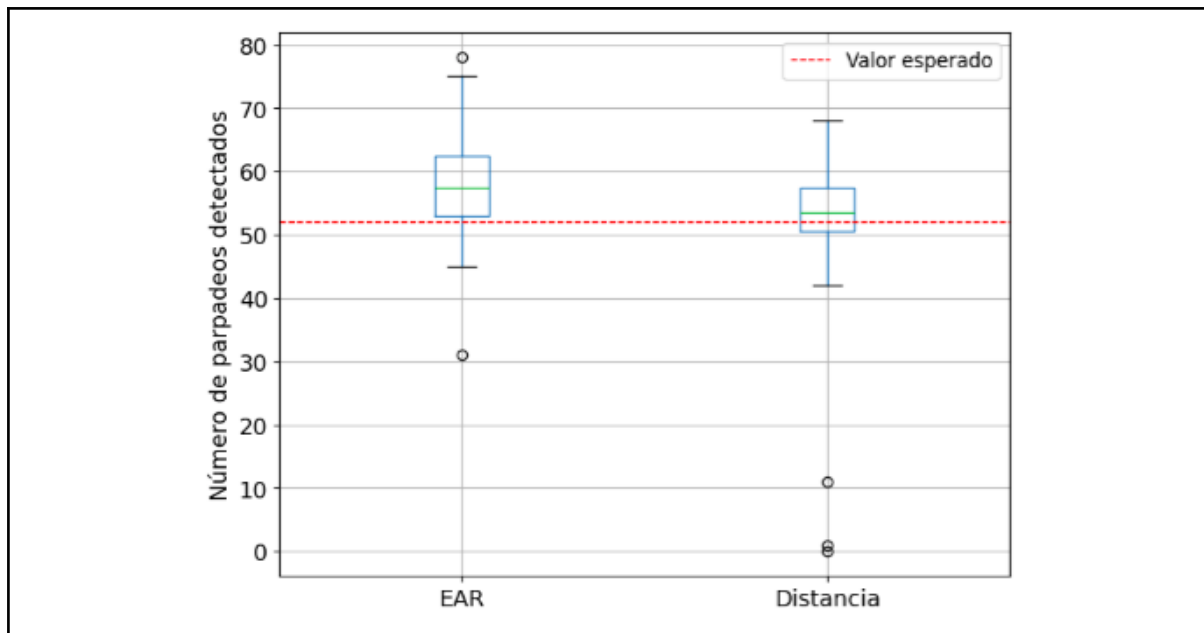
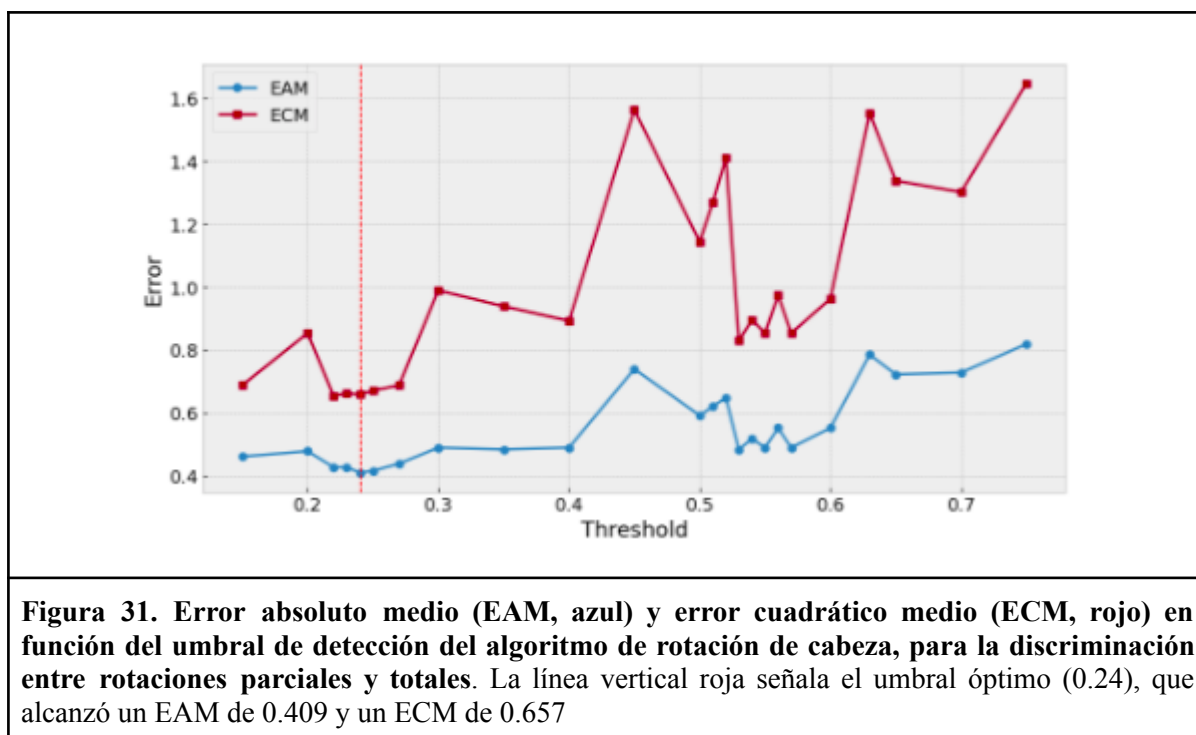


Figura 30. Distribución del número de parpadeos detectados por los métodos EAR y Distancia en el experimento online (N=24, excluyendo 3 participantes con detección nula). El EAR presentó menor error promedio (2.04; DE 2.96) y mayor cercanía al valor esperado que el método de Distancia Euclidiana (2.64; DE 3.49), confirmando su superioridad en condiciones heterogéneas de hardware

Estos hallazgos permitieron seleccionar al EAR como el estándar para el filtrado de datos en el prototipo final. Esta métrica es la que mejor logra diferenciar entre una pérdida de señal de seguimiento (tracking loss) y el cierre fisiológico de los ojos, mitigando la posibilidad de confundir el parpadeo natural con errores de captura.

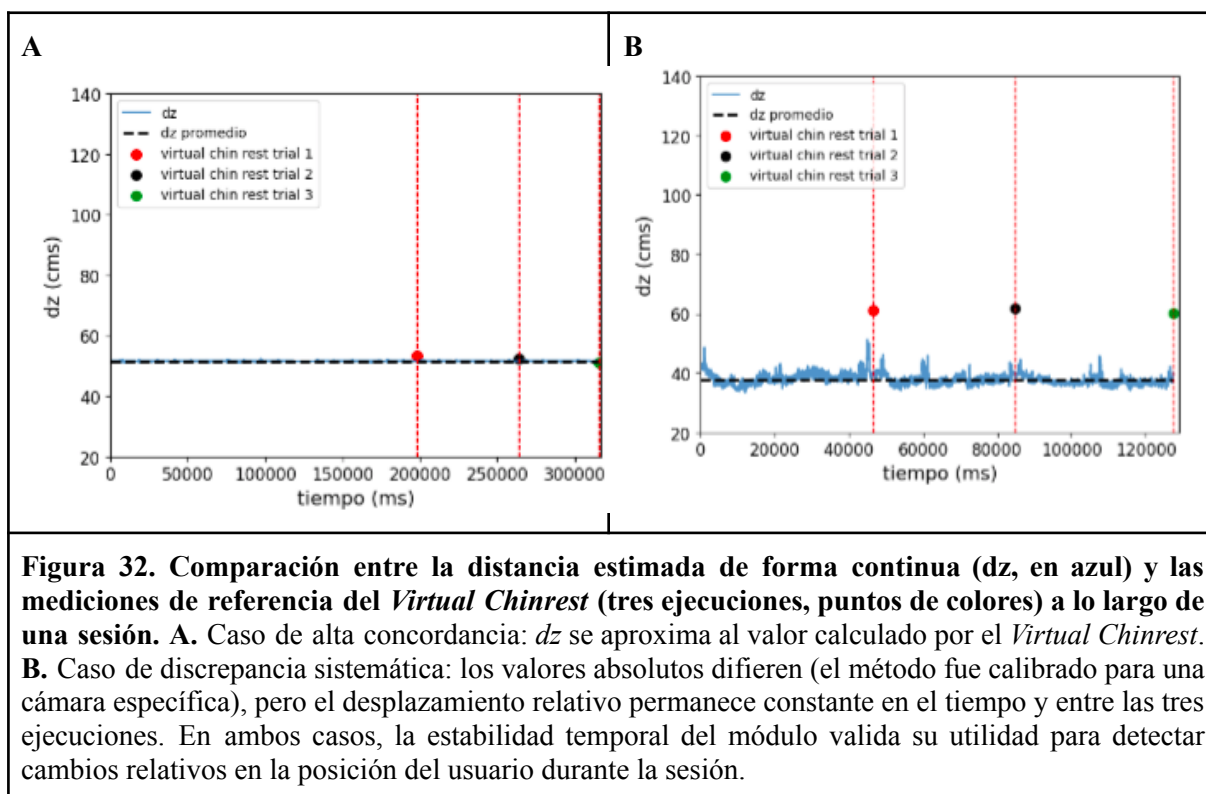
Detección de Rotación de Cabeza

La validación del algoritmo geométrico de rotación demostró su capacidad para distinguir entre movimientos totales y parciales en distintas direcciones (Fig. 31). Tras calibrar los umbrales de decisión, se determinó que un valor de 0.24 ofrecía el mejor balance de desempeño. Bajo esta configuración, el sistema alcanzó un Error Absoluto Medio de 0.409 y un Error Cuadrático Medio de 0.657. Esto confirma que el monitoreo de las distancias relativas entre puntos faciales es una estrategia efectiva para identificar cambios posturales que podrían comprometer la calibración.



Estimación de Distancia a la Pantalla

Finalmente, se comparó el método de estimación continua propuesto contra el estándar de *Virtual Chinrest* (Li et al., 2020; ver **Capítulo 4**). Si bien los valores absolutos de distancia mostraron discrepancias debido a la falta de acceso a la distancia focal física de las cámaras web de los usuarios, la estimación se mantuvo constante a lo largo del tiempo y entre distintas ejecuciones (Fig. 32). Esto validó la utilidad del módulo para medir desplazamientos relativos (Δz) del sujeto durante la sesión, permitiendo corregir o descartar ensayos donde el participante modifica significativamente su posición frente al monitor.



Estudio de caso: validación de la tarea de antisacadas

Dada su sensibilidad a diferencias sutiles en los tiempos de reacción y su relevancia como biomarcador clínico, se seleccionó la tarea de antisacadas para validar el prototipo. A continuación se detallan los aspectos metodológicos de su implementación.

Participantes

Un total de 26 participantes completaron una única sesión experimental (31.8 ± 6.1 años de edad, 7 mujeres). Todos los sujetos contaban con visión normal o corregida a normal. El estudio fue revisado y aprobado por el Comité de Ética del Instituto de Investigaciones Médicas Alfredo Lanari (IDIM), y todos los participantes proporcionaron su consentimiento informado por escrito de acuerdo con la Declaración de Helsinki. Los datos anonimizados se encuentran disponibles en el **Apéndice B.3**.

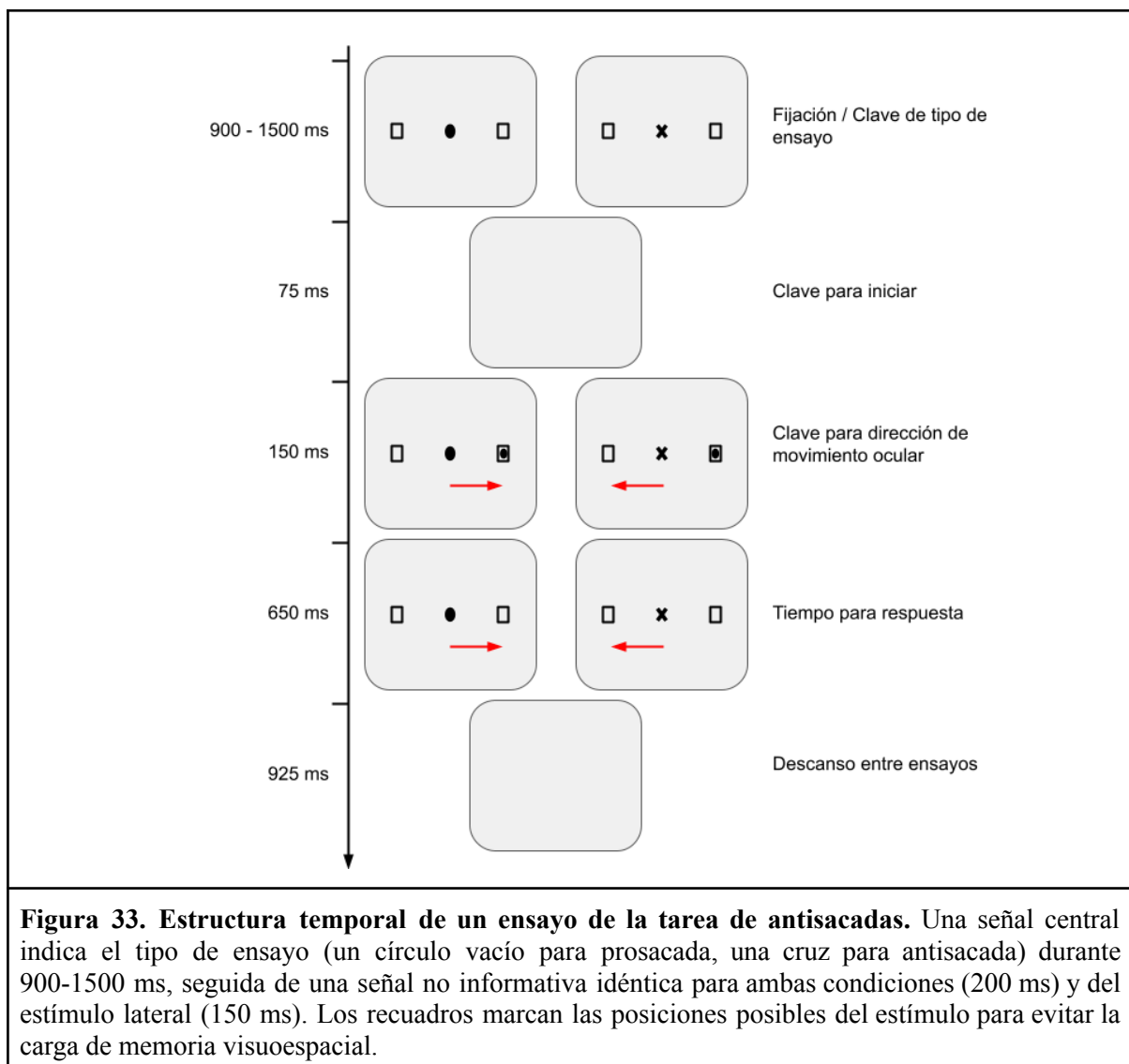
Materiales

Para asegurar la consistencia técnica, se utilizó una configuración de computadora estándar (Ubuntu 20.04.5 LTS / 16 GB RAM / Intel Core i5-4460 CPU @ 3.20GHz x 4) y un monitor LG (22MP48HQ-P, 1920X1080, 60 Hz). Los experimentos se realizaron utilizando una cámara web Logitech C270 (30 FPS, 720p).

Protocolo Experimental

Al inicio del protocolo, cada participante habilitó el acceso a su cámara web y completó las dos tareas del *Virtual Chinrest* (calibración con tarjeta de crédito y punto ciego) descritas previamente. Posteriormente, se les mostró la señal de video capturada por su cámara web y se les solicitó centrar su cabeza dentro de un recuadro en la pantalla, asegurando así que el modelo de malla facial (*facemesh*) pudiera capturar todos los puntos durante el experimento.

Una vez completados estos pasos, se presentaron las instrucciones para ambas condiciones (prosacadas y antisacadas). Cada ensayo (Fig. 33) comenzaba con una señal visual central que indicaba el tipo de tarea: un círculo vacío para ensayos de prosacada o una cruz para ensayos de antisacada, la cual permanecía en pantalla por un tiempo aleatorio entre 900 y 1500 ms. Esta señal era reemplazada por una señal no informativa de 200 ms (un cuadrado vacío en el centro, idéntico para ambas condiciones). Inmediatamente después, aparecía un estímulo lateral (un círculo lleno) durante 150 ms. Para evitar priorizar la memoria visuoespacial, se dibujaron recuadros marcando las ubicaciones donde aparecerían los estímulos laterales. El tiempo permitido para la respuesta del participante fue de 650 ms, seguido de un intervalo entre ensayos (ITI) de 925 ms.



La estructura de la sesión consistió en una serie de bloques conteniendo siempre ensayos del mismo tipo (pro o antisacada), alternando entre los dos tipos de tareas. Los participantes recalibraron el prototipo entre cada bloque, momento en el cual se verificaba también la posición de la cabeza. Una ejecución completa del experimento incluyó un bloque de tutorial al inicio con 10 ensayos de cada tipo (los cuales fueron descartados para los análisis), seguido de 160 antisacadas y 160 prosacadas, distribuidas en bloques de 20 repeticiones cada uno.

Preprocesamiento de datos

Para garantizar la calidad de los datos obtenidos en el entorno web, se aplicaron los pasos de procesamiento descriptos a continuación. El procesamiento inicial de los datos crudos se realizó con *Pyxations* (v0.3.0; ver **Apéndice A.5**), una biblioteca de código abierto para el manejo de datos de seguimiento ocular en formato BIDS (Gorgolewski et al., 2016). Esta herramienta consolidó distintos desarrollos de procesamiento de movimientos oculares que fueron apareciendo a lo largo de esta tesis (en particular en los capítulos dedicados al cTMT) y que terminaron convergiendo en una única biblioteca, construida en colaboración con otros integrantes del laboratorio. En este capítulo, *Pyxations* se encargó de parsear los archivos CSV generados por *WebGazer* en un DataFrame por muestra con metadatos a nivel de ensayo, convertir los datos crudos a la estructura de directorios BIDS-ET, y propagar los metadatos conductuales (tipo de sacada, posición de la señal, coordenadas de los puntos de validación) hacia los archivos de muestras de mirada. Sobre esta salida estandarizada se aplicaron los pasos específicos de la tarea descriptos a continuación.

Primero, se abordó la variabilidad de la frecuencia de muestreo inherente a los navegadores web: las estimaciones de cada ensayo fueron remuestreadas a 30 Hz mediante interpolación lineal. A continuación se realizó una sustracción de la línea de base utilizando el promedio de la coordenada *x* estimada durante la fase de fijación (definida como la mediana entre -200 y 100 ms) para centrarla en 0. Posteriormente, la señal se normalizó estableciendo la mediana entre 500 y 700 ms como valor 1. Se excluyeron del análisis los últimos 100 ms del ensayo para evitar artefactos anticipatorios del siguiente intento.

Se descartaron los ensayos con valores de posición fuera del rango [-1.5, 1.5] (correspondientes a una pérdida excesiva de precisión). Esto resultó en la exclusión del $6.9\% \pm 4.7\%$ de las prosacadas y el $8.4\% \pm 6.8\%$ de las antisacadas, sin diferencias significativas entre condiciones (prueba de suma de rangos de Wilcoxon, $p=0.797$). Adicionalmente, se aplicó un filtro basado en la tasa de error por bloque: si un participante cometía más del 50% de errores en un bloque, este se descartaba por completo, asumiendo una falla en la comprensión de la consigna. Tras aplicar este filtro, solo se descartaron 3 bloques en toda la muestra.

Para la visualización y análisis de las curvas, la mitad de los ensayos fueron "espejados" (multiplicando por -1 las estimaciones de ensayos izquierdos), de modo que todas las respuestas correctas de antisacada mostraran un salto hacia valores negativos, y las prosacadas correctas hacia valores positivos. Dado que la frecuencia de muestreo de las cámaras web estándar (~30 Hz) y el nivel de ruido de la señal limitan la eficacia de los algoritmos clásicos basados en velocidad (como el I-VT; Salvucci & Goldberg, 2000), se evaluó también el uso de *RemoDNaV* (Dar et al., 2021), un detector adaptativo más robusto; sin embargo, la calidad de la señal no permitió identificar sacadas de forma fiable. Por ello, para cuantificar los movimientos oculares se implementó un criterio de detección basado en umbrales espaciales. Se definieron límites normalizados en el intervalo [-0.5, 0.5] respecto

al centro de la pantalla: cuando la señal de la mirada superaba estos límites (a partir de los 100 ms del estímulo) y se estabilizaba en la región del objetivo, se consideraba que el sujeto había mirado hacia uno de los objetivos laterales. Esto permitió clasificar automáticamente cada ensayo en cuatro categorías (prosacadas correctas/incorrectas y antisacadas correctas/incorrectas). El tiempo de reacción (RT) se definió operativamente como el intervalo temporal entre la aparición del estímulo y el primer instante en que la señal superaba dicho umbral espacial.

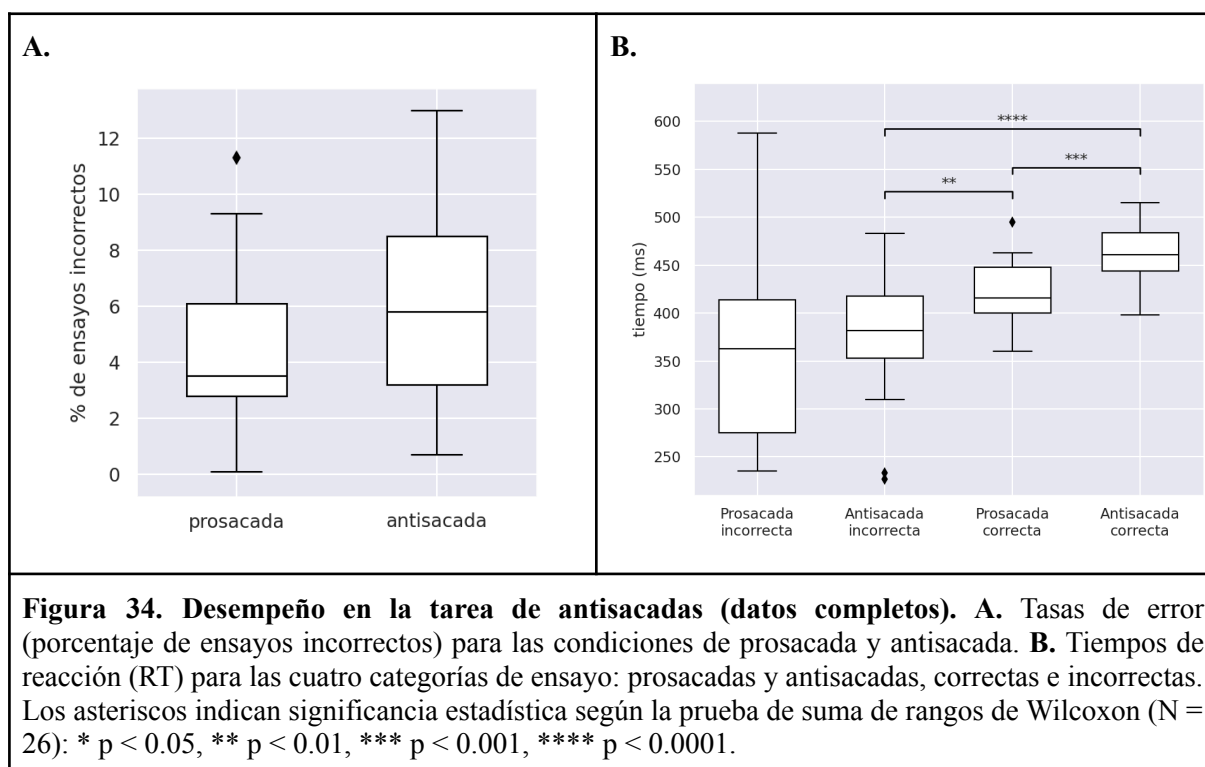
Análisis estadístico

El objetivo principal fue comparar el desempeño en prosacadas y antisacadas. Para ello, se realizaron comparaciones entre ambas condiciones utilizando el test no paramétrico de Wilcoxon (rank-sum test): el porcentaje de respuestas incorrectas y los tiempos de reacción (RT) de las respuestas correctas. En un segundo nivel de análisis, se exploró el efecto de anticipación en la condición de antisacada comparando los RT de los ensayos correctos frente a los incorrectos. En todos los casos se reportan la mediana y el rango intercuartil entre participantes, junto con el p-valor proporcionado por el test de Wilcoxon.

Resultados de la tarea de antisacadas

Cuantificamos la diferencia en las tasas de error (Fig. 34A) y los tiempos de reacción (Fig. 34B) para todas las pruebas de prosacadas y antisacadas. En cuanto a las tasas de error, el porcentaje de ensayos incorrectos fue mayor para los ensayos de antisacadas que para los de prosacadas, tal y como se ha descrito anteriormente en la bibliografía (Hallett, 1978; Hutton & Ettinger, 2006; Munoz & Everling, 2004). Esta diferencia no alcanzó significancia estadística al nivel de 0.05, pero mostró una clara tendencia (mediana [rango intercuartil]; prosacadas: 3.5 [2.8, 6.1] %; antisacadas: 5.8 [3.2, 8.5] %; prueba de suma de rangos de Wilcoxon (N = 26): $p = 0.087$). En la misma línea, encontramos que en las pruebas de antisacadas correctas los tiempos de reacción eran más largos en comparación con las pruebas de prosacadas (mediana [rango intercuartil]; prosacadas correctas: 416 [400, 448] ms; antisacadas correctas: 461 [444, 484] ms; prueba de suma de rangos de Wilcoxon (N = 26): $p = 0.00019$), tal y como se esperaba según la bibliografía antes mencionada. Esta diferencia fue significativa y refleja el coste cognitivo adicional que implica la inhibición necesaria para realizar pruebas de antisacadas.

Notablemente, observamos tiempos de reacción más bajos para las antisacadas en las pruebas incorrectas en comparación con las correctas (mediana [rango intercuartil]; tiempo de reacción de las antisacadas incorrectas 382 [353, 418] ms; tiempo de reacción de las antisacadas correctas 461 [444, 484] ms; prueba de suma de rangos de Wilcoxon (N = 26): $p = 2.0e-06$), y también con las pruebas de prosacadas correctas (mediana [rango intercuartil]; RT de antisacadas incorrectas 382 [353, 418] ms; RT de prosacadas correctas 416 [400, 448] ms; prueba de suma de rangos de Wilcoxon (N = 26): $p = 0.0030$). Este no fue el caso en las pruebas de prosacadas incorrectas (mediana [rango intercuartil]; RT antisacadas incorrectas 382 [353, 418] ms; RT prosacadas incorrectas 363 [275, 414] ms; prueba de suma de rangos de Wilcoxon (N = 26): $p = 0.24$). Este hallazgo es consistente con la presencia de errores rápidos (*fast errors*) característicos de esta tarea: la menor latencia de las antisacadas incorrectas sugiere que reflejan sacadas reflejas hacia el estímulo que escaparon al control inhibitorio, en línea con la idea de que los errores se deben a un fallo en la inhibición de la respuesta automática (Munoz & Everling, 2004).



En general, los datos obtenidos replican los patrones clásicos de la literatura: se observó una clara tendencia hacia una mayor tasa de errores en la condición de antisacadas, acompañada de una diferencia significativa en los tiempos de reacción. Específicamente, las antisacadas correctas mostraron latencias mayores que las prosacadas, reflejando el costo cognitivo de la inhibición, mientras que los errores en antisacadas fueron notablemente más rápidos, consistentes con la falla en suprimir una respuesta automática.

Discusión y Conclusiones

El presente capítulo tuvo como principal objetivo transformar una librería de seguimiento ocular de código abierto (*WebGazer.js*) en una herramienta para la evaluación neurocognitiva remota. Los resultados obtenidos confirman que luego de la optimización del *framerate*, la integración de módulos de registro de la distancia a la pantalla, parpadeos y diversas rotaciones de la cabeza, y la posibilidad de analizar la distancia a la pantalla de manera continua, fue posible replicar efectos sutiles, como el costo de antisacada, utilizando únicamente la cámara web de computadoras estándar.

Uno de los mayores desafíos del *eye tracking* online es la variabilidad del entorno del usuario (Bánki et al., 2021; Zhang & Yin, 2025). La validación técnica demostró que las heurísticas simples de las versiones originales de la librería eran insuficientes para garantizar la calidad del dato en sesiones prolongadas.

La implementación del algoritmo EAR (*Eye Aspect Ratio*) marcó una diferencia crítica en el desempeño del sistema. A diferencia de los métodos basados en intensidad de color o correlación, que

arrojaron tasas de error promedio más elevadas, el EAR mostró una consistencia superior en los resultados generales. Esta mejora se atribuye a la naturaleza geométrica del algoritmo: al calcular una relación de distancias entre puntos faciales en lugar de depender de los valores absolutos de los píxeles, el método resulta intrínsecamente más robusto frente a la variabilidad ambiental. Esto permitió distinguir con mayor precisión entre pérdidas de señal (*tracking loss*) y parpadeos fisiológicos, reduciendo drásticamente la tasa de falsos positivos que, de otro modo, habrían contaminado el análisis.

Del mismo modo, el estimador geométrico de postura (*head pose*) basado en distancias relativas permitió controlar la segunda fuente mayor de ruido: el movimiento del sujeto. Al detectar en tiempo real los desplazamientos y rotaciones excesivas de la cabeza, el sistema permite identificar y descartar los tramos en los que la calibración deja de ser válida, compensando así la ausencia de los elementos de sujeción física (como las mentoneras) propios de un laboratorio tradicional.

Por otro lado, la replicación del efecto de antisacada en una muestra remota constituye el hallazgo empírico más significativo de este capítulo. La diferencia observada de aproximadamente 45 ms entre las latencias de antisacadas y prosacadas correctas es consistente con los rangos reportados en la literatura de laboratorio (Munoz & Everling, 2004). Esta consistencia sugiere que, pese a las limitaciones del hardware, el prototipo conserva la sensibilidad temporal suficiente para detectar el costo del procesamiento inhibitorio *top-down*.

Asimismo, el análisis de los errores aporta evidencia sobre la capacidad del sistema para capturar dinámicas rápidas. La observación de que los errores en antisacadas presentan latencias significativamente menores (382 ms) que las respuestas correctas es compatible con la teoría del "reflejo de agarre visual" (visual grasp reflex; Coe & Munoz, 2017; Guitton et al., 1985). Estos datos respaldan la hipótesis de que la herramienta puede registrar fallos de inhibición automáticos impulsados por la captura atencional exógena, validando su utilidad potencial no solo para métricas de precisión espacial, sino también para la caracterización de perfiles de respuesta temporal.

Es importante destacar que la validez de los resultados no depende exclusivamente del software, sino de su interacción con el protocolo. Si bien se respetaron los lineamientos del estándar internacional (Antoniades et al., 2013), se introdujeron adaptaciones para el contexto online. El incremento del número de ensayos a 320 fue pensado como una medida compensatoria para contrarrestar la menor relación señal-ruido del equipo. Del mismo modo, la decisión de fragmentar la sesión en bloques breves con recalibración frecuente parece haber sido una estrategia efectiva para gestionar dos problemas concurrentes: la fatiga atencional del participante, que tiende a deteriorar el control inhibitorio en protocolos extensos, y la deriva instrumental (*drift*) de los algoritmos de calibración, cuya precisión decae temporalmente en ausencia de reajustes (Sammelmann & Weigelt, 2018).

A pesar de las mejoras implementadas en la robustez del sistema, este estudio presenta limitaciones inherentes a la naturaleza de la tecnología web y la variabilidad del hardware de los participantes.

Adicionalmente, las validaciones de los módulos de monitoreo (parpadeo, rotación de cabeza y distancia a pantalla) tuvieron carácter exploratorio y descriptivo, sin incluir estadística inferencial; estudios futuros con muestras más amplias deberán confirmar la robustez de los parámetros seleccionados.

Las principales limitaciones detectadas están asociadas a la dependencia de las características ópticas de las cámaras web, tales como la resolución, la distancia focal y la calidad del sensor. Específicamente, la baja resolución espacial sigue siendo un obstáculo significativo para trasladar este tipo de tecnologías del laboratorio a la web. En el caso del módulo de estimación de distancia, la falta de acceso a la distancia focal física de la cámara dificulta la obtención de valores absolutos precisos sin una calibración previa estricta. Asimismo, aunque el impacto se mitigó mediante el algoritmo EAR, los métodos de monitoreo basados en color y correlación mostraron una alta sensibilidad a las variaciones de iluminación, lo que restringe su utilidad fuera de condiciones controladas. Es importante destacar, sin embargo, que el uso de teléfonos inteligentes (*smartphones*) podría subsanar este problema, siendo un ejemplo el kit de desarrollo de software (SDK) del sistema operativo *Android*⁶, el cual permite el acceso a las especificaciones ópticas del sensor. Si se migrara este tipo de tareas a dichos dispositivos, sería posible utilizar los parámetros intrínsecos reales en lugar de depender de estimaciones.

Por otro lado, para asegurar un acceso masivo a este tipo de tareas, es necesario abordar la compatibilidad entre plataformas. Dado que Google Chrome es el navegador más utilizado globalmente, nuestros experimentos se centraron en su motor (Chromium). Sin embargo, la generalización de la herramienta requiere una validación exhaustiva en otros navegadores populares como Mozilla Firefox y Safari, los cuales pueden presentar comportamientos distintos en la gestión de recursos de video y ejecución de *JavaScript*.

Consideramos que varias de estas limitaciones, especialmente las referidas a la resolución espacial y el rendimiento, podrán ser mitigadas progresivamente gracias a la disponibilidad de mejor hardware (cámaras web más rápidas o sincronización con *smartphones*) y al avance de las capacidades de los navegadores, como la implementación de la API *WebGPU*, la cual permite exponer las capacidades de aceleración gráfica del hardware directamente al entorno web (Sengupta et al., 2025; Wang et al., 2025).

En cuanto a la potencial aplicabilidad clínica, la validación de este enfoque modular, capaz de arrojar datos robustos sin requerir equipamiento especializado, es crucial para su traslación clínica.

En los últimos años, múltiples estudios han destacado el potencial diagnóstico de la tarea de antisacadas, en particular su capacidad para discriminar entre la enfermedad de Alzheimer (EA), el deterioro cognitivo leve (DCL) y las poblaciones sanas (Eraslan Boz et al., 2023; Opwonya et al., 2022, 2023). Además, cabe mencionar que, aunque nos hemos centrado en la tarea antisacádica, existen otras tareas relevantes para la práctica clínica que no requieren un seguimiento ocular de alta resolución espacial (Eizenman et al., 2003), y que también pueden utilizarse para mejorar los resultados de las intervenciones dependientes de la mirada en personas con diversos trastornos cerebrales (Carelli et al., 2022). Basándonos en este conjunto de investigaciones, creemos que una metodología de seguimiento ocular basada en la web y totalmente funcional podría ser extremadamente útil para monitorizar el grado de progresión de este tipo de patologías, brindando una oportunidad a las poblaciones normalmente excluidas a la atención de salud mental. Además, podría servir como un sistema de alerta temprana que permita informar al profesional de la salud sobre el

6

https://developer.android.com/reference/android/hardware/camera2/CameraCharacteristics#LENS_INFO_AVAILABLE_FOCAL_LENGTHS

estado del paciente, y que este solo deba acudir a la consulta médica en caso de ser estrictamente necesario, ayudando a disminuir las complicaciones asociadas a la falta de profesionales y, especialmente, optimizando los recursos de las personas que se encuentran lejos de centros de salud de alta complejidad.

En conclusión, este capítulo presenta un prototipo mejorado de seguimiento ocular web que resulta prometedor en términos de confiabilidad y usabilidad, quedando todo el código disponible de manera *open source* (Ver **Apéndice A.6.**). Los resultados se vieron favorecidos por la incorporación de módulos de monitoreo que complementan la librería original: la implementación del algoritmo EAR ofreció una mayor consistencia en la detección de parpadeos frente a variaciones, mientras que el estimador de postura colaboró en el control de la posición de la cabeza, un factor crítico en ausencia de mentonera.

Asimismo, la tarea de antisacadas utilizada como caso de estudio evidenció mayores tasas de error y tiempos de reacción más prolongados en la condición de antisacada respecto a la de prosacada, un patrón consistente con la literatura existente. Esto sugiere que la combinación de la herramienta de software con un diseño experimental compensatorio (caracterizado por bloques breves y un mayor número de ensayos) es una estrategia viable para mitigar el ruido inherente a los registros domésticos.

En suma, el seguimiento ocular basado en web, respaldado por estas estrategias de control de calidad, mantiene su potencial para expandir la accesibilidad de la investigación, permitiendo alcanzar poblaciones diversas y eventualmente asistir en el monitoreo del estado cognitivo de pacientes a gran escala de manera remota. Se requieren aún mejoras e investigaciones adicionales para incrementar las capacidades y la precisión de esta tecnología, pero los presentes hallazgos constituyen un paso metodológico relevante hacia su utilización rigurosa fuera del laboratorio.

Conclusiones generales

Esta tesis se propuso transformar un instrumento neuropsicológico clásico, el Trail Making Test, junto con la medición de trayectorias visuales y manuales, en un conjunto de herramientas digitales capaces de evaluar funciones ejecutivas con mayor precisión, rigor y alcance. El recorrido de los seis capítulos expone un camino que va desde el registro de alta resolución en el laboratorio hasta la evaluación cognitiva remota, accesible y escalable.

El punto de partida fue la medición y la descomposición de la tarea. En el **Capítulo 1** desarrollamos y validamos el cTMT, una versión computarizada que controla las configuraciones espaciales de los ítems y registra de manera simultánea y con alta resolución los movimientos de la mano y de los ojos, lo que permitió descomponer una tarea aparentemente unitaria en fases de monitoreo, planificación y exploración, derivar marcadores internos de memoria de trabajo y, de forma preliminar, de control inhibitorio, y replicar los efectos clásicos del test. En el **Capítulo 2** llevamos esa herramienta a pacientes con deterioro cognitivo leve (DCL) y encontramos que las métricas oculares, en particular el número de fijaciones, discriminaban entre grupos allí donde las medidas conductuales clásicas no lo lograban. Ese mismo estudio puso de manifiesto una dificultad práctica: aunque el registro ocular es viable y aporta información valiosa, adquirir datos de calidad en población clínica resulta más costoso y demandante, lo que plantea un desafío de escalabilidad. Esa experiencia motivó un giro estratégico hacia los movimientos de la mano, una señal con menores requerimientos de hardware y, por lo tanto, más fácil de desplegar a gran escala.

Ese cambio de rumbo organiza la segunda mitad del recorrido, orientada a la predicción, la infraestructura y la posibilidad de evaluar de manera remota. En el **Capítulo 3** mostramos que un modelo de aprendizaje automático basado en los movimientos de la mano alcanzaba un valor diagnóstico cercano al de una batería neuropsicológica completa (AUC de 0.67 con las características digitales solas, que ascendió a 0.70 al sumar variables demográficas básicas, frente a 0.76 de la batería completa), con la ventaja de no requerir hardware especializado y de apoyarse en una metodología rigurosa y reproducible. El **Capítulo 4** presentó *DataPruebas*, una de las contribuciones a las que más esfuerzo de desarrollo dedicamos: una plataforma web de acceso abierto y gratuita, diseñada específicamente para la comunidad hispanohablante, que buscó superar las barreras de costo e idioma de las herramientas de recolección de datos existentes, en su mayoría pagas y orientadas a participantes angloparlantes. Más que una herramienta puntual, funcionó como la infraestructura que sostuvo los estudios remotos de esta tesis. Sobre esa base, en el **Capítulo 5** extendimos el cTMT a un entorno online con 484 participantes y demostramos que las trayectorias del cursor predicen diferencias individuales en memoria de trabajo, inhibición de la respuesta y edad. Finalmente, el **Capítulo 6** cerró el círculo abierto en los primeros capítulos: recuperamos el registro ocular, pero esta vez de forma remota y sin hardware especializado, mejorando la biblioteca *WebGazer.js* y validando el sistema con la tarea de antisacadas.

Más allá de los resultados particulares de cada capítulo, la tesis se sostiene sobre cuatro pilares. El primero es la medición multimodal de alta resolución temporal, que combina movimientos oculares y manuales para acceder a procesos cognitivos que una medida global como el tiempo de respuesta no alcanza a capturar. El segundo es el rigor metodológico y el compromiso con la ciencia abierta, expresado tanto en los procedimientos de validación (validación cruzada anidada, pruebas de permutación, interpretabilidad mediante SHAP) como en el ecosistema de herramientas y conjuntos de datos liberados bajo licencia abierta que se recopilan en los **Apéndices A y B**. El tercero es el foco explícito en la accesibilidad, el bajo costo y la escalabilidad, con especial atención a América Latina y a los entornos de recursos limitados, donde el hardware asequible y las metodologías abiertas pueden marcar la diferencia en el acceso a la evaluación cognitiva. El cuarto es la contribución a los campos emergentes de la neuropsicología digital y la psiquiatría computacional, mediante el desarrollo de biomarcadores digitales objetivos, cuantificables y ecológicamente válidos.

En particular, los conjuntos de datos recolectados a lo largo de esta tesis constituyen un aporte en sí mismos. Más allá de los análisis aquí presentados, quedan disponibles de forma abierta para su reutilización, replicación y reanálisis por parte de la comunidad. Se trata de un recurso especialmente valioso para el estudio de poblaciones de habla hispana, históricamente subrepresentadas en los grandes conjuntos de datos conductuales, y un insumo para líneas de investigación futuras.

Estas contribuciones deben leerse a la luz de las limitaciones del trabajo. Los estudios clínicos se apoyaron en diseños transversales y en tamaños muestrales acotados, lo que restringe la potencia estadística y no permite evaluar la progresión del deterioro en el tiempo. Las implementaciones remotas, por su parte, dependen del hardware y las condiciones del entorno de cada participante, y varios de los módulos desarrollados en el último capítulo fueron validados de manera exploratoria. Estas limitaciones no invalidan los hallazgos, pero delimitan su alcance y señalan las direcciones que el trabajo futuro deberá recorrer.

La continuación natural de esta línea de investigación pasa por los estudios longitudinales, que permitirían evaluar la confiabilidad *test-retest* de los marcadores propuestos y, sobre todo, su capacidad para rastrear la conversión de DCL a demencia. Un segundo eje es la integración

multimodal genuina, que combine en un mismo protocolo los movimientos de la mano y los ojos con otras señales como el habla o las medidas fisiológicas, con el fin de construir perfiles cognitivos más completos. A esto se suma la necesidad de validar las herramientas en cohortes más grandes y diversas, incluyendo su despliegue remoto en poblaciones clínicas, y de seguir mejorando la precisión del seguimiento ocular web mediante avances técnicos como la API *WebGPU*, la sincronización con smartphones y los sistemas de calibración adaptativa.

En conjunto, esta tesis avanza en dos direcciones complementarias. Por un lado, profundiza la medición de las funciones ejecutivas, diseccionando con alta resolución procesos complejos que los instrumentos clásicos resumen en una única medida. Por el otro, demuestra que es factible trasladar esa evaluación fuera del ámbito tradicional del laboratorio, preservando el rigor metodológico y asumiendo, de manera controlada, los márgenes de error inherentes a la medición en entornos remotos. Al combinar mediciones ricas, modelos transparentes e infraestructura abierta, abre un camino hacia evaluaciones cognitivas válidas, accesibles y escalables, con el potencial de alcanzar a poblaciones habitualmente excluidas del sistema de salud y de contribuir, desde la psiquiatría computacional, a una distribución más equitativa de las herramientas de diagnóstico y monitoreo cognitivo.

Bibliografía

- Al Zoubi, O., Ki Wong, C., Kuplicki, R. T., Yeh, H., Mayeli, A., Refai, H., Paulus, M., & Bodurka, J. (2018). Predicting Age From Brain EEG Signals—A Machine Learning Approach. *Frontiers in Aging Neuroscience*, 10. <https://doi.org/10.3389/fnagi.2018.00184>
- Antoniades, C., Ettinger, U., Gaymard, B., Gilchrist, I., Kristjánsson, A., Kennard, C., John Leigh, R., Noorani, I., Pouget, P., Smyrnis, N., Tarnowski, A., Zee, D. S., & Carpenter, R. H. S. (2013). An internationally standardised antisaccade protocol. *Vision Research*, 84, 1–5. <https://doi.org/10.1016/j.visres.2013.02.007>
- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1), 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Aranda, M. P., Kremer, I. N., Hinton, L., Zissimopoulos, J., Whitmer, R. A., Hummel, C. H., Trejo, L., & Fabius, C. (2021). Impact of dementia: Health disparities, population trends, care interventions, and economic costs. *Journal of the American Geriatrics Society*, 69(7), 1774–1783. <https://doi.org/10.1111/jgs.17345>
- Arbuthnott, K., & Frank, J. (2000). Trail Making Test, Part B as a Measure of Executive Control: Validation Using a Set-Switching Paradigm. *Journal of Clinical and Experimental Neuropsychology*, 22(4), 518–528. [https://doi.org/10.1076/1380-3395\(200008\)22:4;1-0;FT518](https://doi.org/10.1076/1380-3395(200008)22:4;1-0;FT518)

- Badrulhisham, F., Pogatzki-Zahn, E., Segelcke, D., Spisak, T., & Vollert, J. (2024). Machine learning and artificial intelligence in neuroscience: A primer for researchers. *Brain, Behavior, and Immunity*, 115, 470–479. <https://doi.org/10.1016/j.bbi.2023.11.005>
- Baez, S., Alladi, S., & Ibanez, A. (2023). Global South research is critical for understanding brain health, ageing and dementia. *Clinical and Translational Medicine*, 13(11), e1486. <https://doi.org/10.1002/ctm2.1486>
- Bánki, A., de Eccher, M., Falschlehner, L., Hoehl, S., & Markova, G. (2021). Comparing Online Webcam- and Laboratory-Based Eye-Tracking for the Assessment of Infants' Audio-Visual Synchrony Perception. *Frontiers in Psychology*, 12, 733933. <https://doi.org/10.3389/fpsyg.2021.733933>
- Barbosa, A., Kamienkowski, J. E., & Ison, M. J. (2020, October 7). *Investigating the effects of contextual information in hybrid visual search (poster presentation, XXXV Annual Meeting of the Argentinian Society for Neuroscience Research)*. XXXV Annual Meeting of the Argentinian Society for Neuroscience Research. <https://san2020.saneurociencias.org.ar/>
- Benedetti, V., Gavazzi, G., Giovannelli, F., Bravi, R., Giganti, F., Minciacchi, D., Mascalchi, M., Cincotta, M., & Viggiano, M. P. (2020). Mouse Tracking to Explore Motor Inhibition Processes in Go/No-Go and Stop Signal Tasks. *Brain Sciences*, 10(7), 464. <https://doi.org/10.3390/brainsci10070464>
- Biondi, F. N., Saberi, B., Graf, F., Cort, J., Pillai, P., & Balasingam, B. (2023). Distracted worker: Using pupil size and blink rate to detect cognitive load during manufacturing tasks. *Applied Ergonomics*, 106, 103867. <https://doi.org/10.1016/j.apergo.2022.103867>
- Bowie, C. R., & Harvey, P. D. (2006). Administration and interpretation of the Trail Making Test. *Nature Protocols*, 1(5), 2277–2281. <https://doi.org/10.1038/nprot.2006.390>
- Bravi, R., Gavazzi, G., Benedetti, V., Giovannelli, F., Grasso, S., Panconi, G., Viggiano, M. P., & Minciacchi, D. (2022). Effect of different sport environments on proactive and reactive motor inhibition: A study on open- and closed-skilled athletes via mouse-tracking procedure. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.1042705>
- Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, 8, e9414. <https://doi.org/10.7717/peerj.9414>
- Brown, E. C., Casey, A., Fisch, R. I., & Neuringer, C. (1958). Trail Making Test as a screening device for the detection of brain damage. *Journal of Consulting Psychology*, 22(6), 469–474. <https://doi.org/10.1037/h0039980>

- Bulloch, M. C., Prime, S. L., & Marotta, J. J. (2015). Anticipatory gaze strategies when grasping moving objects. *Experimental Brain Research*, 233(12), 3413–3423. <https://doi.org/10.1007/s00221-015-4413-7>
- Carelli, L., Solca, F., Tagini, S., Torre, S., Verde, F., Ticozzi, N., Ferrucci, R., Pravettoni, G., Aiello, E. N., Silani, V., & Poletti, B. (2022). Gaze-Contingent Eye-Tracking Training in Brain Disorders: A Systematic Review. *Brain Sciences*, 12(7), 931. <https://doi.org/10.3390/brainsci12070931>
- Cha, G.-E., & Min, B.-C. (2022). Correlation between Unconscious Mouse Actions and Human Cognitive Workload. *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems, CHI EA '22*, 1–7. <https://doi.org/10.1145/3491101.3519658>
- Chandrasekharan, J., Joseph, A., Ram, A., & Nollo, G. (2023). ETMT: A Tool for Eye-Tracking-Based Trail-Making Test to Detect Cognitive Impairment. *Sensors*, 23(15), Article 15. <https://doi.org/10.3390/s23156848>
- Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 21, 217–227. <https://doi.org/10.1016/j.acn.2005.12.002>
- Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R., Dwyer, D., & Choi, K. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*, 20(2), 154–170. <https://doi.org/10.1002/wps.20882>
- Chudzik, A., Nowacki, J. P., & Przybyszewski, A. W. (2024). Recognizing Patterns of Parkinson's Disease Using Online Trail Making Test and Response Dynamics – Preliminary Study. *Pattern Recognition: 27th International Conference, ICPR 2024, Kolkata, India, December 1–5, 2024, Proceedings, Part II*, 193–207. https://doi.org/10.1007/978-3-031-78166-7_13
- Coe, B. C., & Munoz, D. P. (2017). Mechanisms of saccade suppression revealed in the anti-saccade task. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1718), 20160192. <https://doi.org/10.1098/rstb.2016.0192>
- Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, 34(4), 613–617. <https://doi.org/10.3758/BF03195489>

- Custodio, N., Wheelock, A., Thumala, D., & Slachevsky, A. (2017). Dementia in Latin America: Epidemiological Evidence and Implications for Public Policy. *Frontiers in Aging Neuroscience*, 9. <https://doi.org/10.3389/fnagi.2017.00221>
- Dahmen, J., Cook, D., Fellows, R., & Schmitter-Edgecombe, M. (2017). An Analysis of a Digital Variant of the Trail Making Test Using Machine Learning Techniques. *Technology and Health Care : Official Journal of the European Society for Engineering and Medicine*, 25(2), 251–264. <https://doi.org/10.3233/THC-161274>
- Dar, A. H., Wagner, A. S., & Hanke, M. (2021). REMoDNaV: Robust eye-movement classification for dynamic stimulation. *Behavior Research Methods*, 53(1), 399–414. <https://doi.org/10.3758/s13428-020-01428-x>
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12. <https://doi.org/10.3758/s13428-014-0458-y>
- Del Gatto, C., Indraccolo, A., Delogu, F., May, M., Pedale, T., & Brunetti, R. (2025). Investigating visual search mechanisms and enhancing the diagnostic potential of the trail making test using eTMT. *Scientific Reports*, 15(1), 33445. <https://doi.org/10.1038/s41598-025-19125-0>
- Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, 64(1), 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Dowd, C. (2020). *A New ECDF Two-Sample Test Statistic* (arXiv:2007.01360). arXiv. <https://doi.org/10.48550/arXiv.2007.01360>
- Drewe, E. A. (1975). Go—No go learning after frontal lobe lesions in humans. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 11(1), 8–16. [https://doi.org/10.1016/s0010-9452\(75\)80015-3](https://doi.org/10.1016/s0010-9452(75)80015-3)
- Du, M., Andersen, S. L., Cosentino, S., Boudreau, R. M., Perls, T. T., & Sebastiani, P. (2022). Digitally generated Trail Making Test data: Analysis using hidden Markov modeling. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 14(1), e12292. <https://doi.org/10.1002/dad2.12292>
- Duchowski, A. T. (2017). *Eye Tracking Methodology*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-57883-5>
- Durgin, F. H., Tripathy, S. P., & Levi, D. M. (1995). On the filling in of the visual blind spot: Some rules of thumb. *Perception*, 24(7), 827–840. <https://doi.org/10.1068/p240827>
- Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive*

- Neuroscience, Sensitive Periods across Development*, 25, 69–91.
<https://doi.org/10.1016/j.dcn.2016.11.001>
- Eizenman, M., Yu, L. H., Grupp, L., Eizenman, E., Ellenbogen, M., Gemar, M., & Levitan, R. D. (2003). A naturalistic visual scanning approach to assess selective attention in major depressive disorder. *Psychiatry Research*, 118(2), 117–128. [https://doi.org/10.1016/S0165-1781\(03\)00068-4](https://doi.org/10.1016/S0165-1781(03)00068-4)
- Eraslan Boz, H., Koçoğlu, K., Akkoyun, M., Tüfekci, I. Y., Ekin, M., & Akdal, G. (2023). Eye movement patterns during viewing face images with neutral expressions in patients with early-stage Alzheimer's disease and amnesic mild cognitive impairment. *Brain and Behavior*, 13(11), e3232. <https://doi.org/10.1002/brb3.3232>
- Fellows, R. P., Dahmen, J., Cook, D., & Schmitter-Edgecombe, M. (2017). Multicomponent analysis of a digital Trail Making Test. *The Clinical Neuropsychologist*, 31(1), 154–167. <https://doi.org/10.1080/13854046.2016.1238510>
- Fisk, J. D., & Goodale, M. A. (1985). The organization of eye and limb movements during unrestricted reaching to targets in contralateral and ipsilateral visual space. *Experimental Brain Research*, 60(1), 159–178. <https://doi.org/10.1007/BF00237028>
- Fossum, B., Holmberg, H., & Reinvang, I. (1992). *Spatial and symbolic factors in performance on the Trail Making Test*. <https://doi.org/10.1037/0894-4105.6.1.71>
- Gaudino, E. A., Geisler, M. W., & Squires, N. K. (1995). Construct validity in the trail making test: What makes part B harder? *Journal of Clinical and Experimental Neuropsychology*, 17(4), 529–535. <https://doi.org/10.1080/01688639508405143>
- Germine, L., Reinecke, K., & Chaytor, N. S. (2019). Digital neuropsychology: Challenges and opportunities at the intersection of science and software. *The Clinical Neuropsychologist*, 33(2), 271–286. <https://doi.org/10.1080/13854046.2018.1535662>
- Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., ... Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3(1), 160044. <https://doi.org/10.1038/sdata.2016.44>

- Goset, J., Vinuela-Navarro, F., Mestre, C., Aldaba, M., & Vilaseca, M. (2024). Eye movements during the performance of the neuropsychological Trail Making Test. *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications, ETRA '24*, 1–2. <https://doi.org/10.1145/3649902.3655653>
- Gouveia, P. A. R., Brucki, S. M. D., Malheiros, S. M. F., & Bueno, O. F. A. (2007). Disorders in planning and strategy application in frontal lobe lesion patients. *Brain and Cognition*, 63(3), 240–246. <https://doi.org/10.1016/j.bandc.2006.09.001>
- Guillon, D., Buchtel, H. A., & Douglas, R. M. (1985). Frontal lobe lesions in man cause difficulties in suppressing reflexive glances and in generating goal-directed saccades. *Experimental Brain Research*, 58(3), 455–472. <https://doi.org/10.1007/BF00235863>
- Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, 18(10), 1279–1296. [https://doi.org/10.1016/0042-6989\(78\)90218-3](https://doi.org/10.1016/0042-6989(78)90218-3)
- Hamdan, A. C., & Hamdan, E. M. L. R. (2009). Effects of age and education level on the Trail Making Test in a healthy Brazilian sample. *Psychology & Neuroscience*, 2(2), 199–203. <https://doi.org/10.3922/j.psns.2009.2.012>
- Hanczár, G., Griechisch, E., Ovád, N., Törteli, O. M., Tóth, G., Hanák, D., Vértes, B., Horváth, A., & Kamondi, A. (2022). Detection of mild cognitive impairment based on mouse movement data of trail making test. *Informatics in Medicine Unlocked*, 35, 101120. <https://doi.org/10.1016/j.imu.2022.101120>
- Hartman, D. E. (2009). Wechsler Adult Intelligence Scale IV (WAIS IV): Return of the Gold Standard. *Applied Neuropsychology*, 16(1), 85–87. <https://doi.org/10.1080/09084280802644466>
- Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). lab.js: A free, open, online study builder. *Behavior Research Methods*, 54(2), 556–573. <https://doi.org/10.3758/s13428-019-01283-5>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, 33(2–3), 61–83; discussion 83–135. <https://doi.org/10.1017/S0140525X0999152X>
- Hicks, S. L., Sharma, R., Khan, A. N., Berna, C. M., Waldecker, A., Talbot, K., Kennard, C., & Turner, M. R. (2013). An Eye-Tracking Version of the Trail-Making Test. *PLOS ONE*, 8(12), e84061. <https://doi.org/10.1371/journal.pone.0084061>
- Hoyer, W. J., Stawski, R. S., Wasylshyn, C., & Verhaeghen, P. (2004). Adult Age and Digit Symbol Substitution Performance: A Meta-Analysis. *Psychology and Aging*, 19(1), 211–214. <https://doi.org/10.1037/0882-7974.19.1.211>

- Huang, C., Vilotijević, A., Theeuwes, J., & Donk, M. (2021). Proactive distractor suppression elicited by statistical regularities in visual search. *Psychonomic Bulletin & Review*, 28(3), 918–927. <https://doi.org/10.3758/s13423-021-01891-3>
- Huang, M. X., Kwok, T. C. K., Ngai, G., Chan, S. C. F., & Leong, H. V. (2016). Building a Personalized, Auto-Calibrating Eye Tracker from User Interactions. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*, 5169–5179. <https://doi.org/10.1145/2858036.2858404>
- Hutt, S., Wong, A., Papoutsaki, A., Baker, R. S., Gold, J. I., & Mills, C. (2023). Webcam-based eye tracking to detect mind wandering and comprehension errors. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-022-02040-x>
- Hutton, S. B., & Ettinger, U. (2006). The antisaccade task as a research tool in psychopathology: A critical review. *Psychophysiology*, 43(3), 302–313. <https://doi.org/10.1111/j.1469-8986.2006.00403.x>
- Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, Eye Tracking as a Tool to Study and Enhance Multimedia Learning*, 20(2), 172–176. <https://doi.org/10.1016/j.learninstruc.2009.02.013>
- Insel, T. R. (2017). Digital Phenotyping: Technology for a New Science of Behavior. *JAMA*, 318(13), 1215–1216. <https://doi.org/10.1001/jama.2017.11295>
- Insel, T. R. (2018). Digital phenotyping: A global tool for psychiatry. *World Psychiatry*, 17(3), 276–277. <https://doi.org/10.1002/wps.20550>
- Jack Jr., C. R., Andrews, J. S., Beach, T. G., Buracchio, T., Dunn, B., Graf, A., Hansson, O., Ho, C., Jagust, W., McDade, E., Molinuevo, J. L., Okonkwo, O. C., Pani, L., Rafii, M. S., Scheltens, P., Siemers, E., Snyder, H. M., Sperling, R., Teunissen, C. E., & Carrillo, M. C. (2024). Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimer's & Dementia*, 20(8), 5143–5169. <https://doi.org/10.1002/alz.13859>
- Jahfari, S., Stinear, C. M., Claffey, M., Verbruggen, F., & Aron, A. R. (2010). Responding with Restraint: What Are the Neurocognitive Mechanisms? *Journal of Cognitive Neuroscience*, 22(7), 1479–1492. <https://doi.org/10.1162/jocn.2009.21307>
- Jak, A. J., Bondi, M. W., Delano-Wood, L., Wierenga, C., Corey-Bloom, J., Salmon, D. P., & Delis, D. C. (2009). Quantification of Five Neuropsychological Approaches to Defining Mild Cognitive Impairment. *The American Journal of Geriatric Psychiatry*, 17(5), 368–375. <https://doi.org/10.1097/JGP.0b013e31819431d5>

- Juantorena, G. E., Capelo, G., Leon Vallejos, B. D., Ibáñez, A., Petroni, A., Berrios, W., Fernández, M. C., & Kamienkowski, J. E. (2026). Machine learning based digital assessment of mild cognitive impairment using mouse trajectories during the trail making test. *Scientific Reports*. <https://doi.org/10.1038/s41598-026-62955-9>
- Juantorena, G. E., Figari, F., Petroni, A., & Kamienkowski, J. E. (2023). *Web-based eye-tracking for remote cognitive assessments: The anti-saccade task as a case study* (p. 2023.07.11.548447). bioRxiv. <https://doi.org/10.1101/2023.07.11.548447>
- Juantorena, G. E., Gauder, L., Laciana, P., Ferrer, L., & Kamienkowski, J. E. (2025). DataPruebas: Una plataforma en línea para la recopilación de datos. *JAIIO, Jornadas Argentinas de Informática*, 11(1), 149–155. <https://revistas.unlp.edu.ar/JAIIO/article/view/19778>
- Juantorena, G. E., & Kohler, P. J. (2021). *jsPsych virtual-chinrest plugin* [Computer software]. <https://www.jspsych.org/v7/plugins/virtual-chinrest/>
- Jyotsna, C., Amudha, J., Rao, R., & Nayar, R. (2020). Intelligent gaze tracking approach for trail making test. *Journal of Intelligent & Fuzzy Systems*, 38(5), 6299–6310. <https://doi.org/10.3233/JIFS-179711>
- Kalaria, R. N., Maestre, G. E., Arizaga, R., Friedland, R. P., Galasko, D., Hall, K., Luchsinger, J. A., Ogunniyi, A., Perry, E. K., Potocnik, F., Prince, M., Stewart, R., Wimo, A., Zhang, Z.-X., & Antuono, P. (2008). Alzheimer's disease and vascular dementia in developing countries: Prevalence, management, and risk factors. *The Lancet Neurology*, 7(9), 812–826. [https://doi.org/10.1016/S1474-4422\(08\)70169-8](https://doi.org/10.1016/S1474-4422(08)70169-8)
- Kim, S. Y., Park, J., Choi, H., Loeser, M., Ryu, H., & Seo, K. (2023). Digital Marker for Early Screening of Mild Cognitive Impairment Through Hand and Eye Movement Analysis in Virtual Reality Using Machine Learning: First Validation Study. *Journal of Medical Internet Research*, 25(1), e48093. <https://doi.org/10.2196/48093>
- Kobayashi, M., Yamada, Y., Shinkawa, K., Nemoto, M., Nemoto, K., & Arai, T. (2022). Automated Early Detection of Alzheimer's Disease by Capturing Impairments in Multiple Cognitive Domains with Multiple Drawing Tasks. *Journal of Alzheimer's Disease*, 88(3), 1075–1089. <https://doi.org/10.3233/JAD-215714>
- Kortte, K. B., Horner, M. D., & Windham, W. K. (2002). The Trail Making Test, Part B: Cognitive Flexibility or Ability to Maintain Set? *Applied Neuropsychology*, 9(2), 106–109. https://doi.org/10.1207/S15324826AN0902_5

- Kourtis, L. C., Regele, O. B., Wright, J. M., & Jones, G. B. (2019). Digital biomarkers for Alzheimer's disease: The mobile/wearable devices opportunity. *Npj Digital Medicine*, 2(1), 9. <https://doi.org/10.1038/s41746-019-0084-2>
- Kuiper, N. H. (1960). Tests concerning random points on a circle. *Indagationes Mathematicae (Proceedings)*, 63, 38–47. [https://doi.org/10.1016/S1385-7258\(60\)50006-0](https://doi.org/10.1016/S1385-7258(60)50006-0)
- Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online Studies. *PLOS ONE*, 10(6), e0130834. <https://doi.org/10.1371/journal.pone.0130834>
- Lange, R. T., Iverson, G. L., Zakrzewski, M. J., Ethel-King, P. E., & Franzen, M. D. (2005). Interpreting the Trail Making Test Following Traumatic Brain Injury: Comparison of Traditional Time Scores and Derived Indices. *Journal of Clinical and Experimental Neuropsychology*, 27(7), 897–906. <https://doi.org/10.1080/1380339049091290>
- Langenecker, S. A., Zubieta, J.-K., Young, E. A., Akil, H., & Nielson, K. A. (2007). A task to manipulate attentional load, set-shifting, and inhibitory control: Convergent validity and test-retest reliability of the Parametric Go/No-Go Test. *Journal of Clinical and Experimental Neuropsychology*, 29(8), 842–853. <https://doi.org/10.1080/13803390601147611>
- Lara-Garduno, R. A., Jia, Y., Deutz, N. E., Engelen, M., Leslie, N., & Hammond, T. (2022). Detecting mild cognitive impairment through digitized trail-making test interface. *Graphics Interface 2022*.
- Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for Participants' Viewing Distance in Large-Scale, Psychophysical Online Experiments Using a Virtual Chinrest. *Scientific Reports*, 10(1), Article 1. <https://doi.org/10.1038/s41598-019-57204-1>
- Liem, F., Varoquaux, G., Kynast, J., Beyer, F., Kharabian Masouleh, S., Huntenburg, J. M., Lampe, L., Rahim, M., Abraham, A., Craddock, R. C., Riedel-Heller, S., Luck, T., Loeffler, M., Schroeter, M. L., Witte, A. V., Villringer, A., & Margulies, D. S. (2017). Predicting brain-age from multimodal imaging data captures cognitive impairment. *NeuroImage*, 148, 179–188. <https://doi.org/10.1016/j.neuroimage.2016.11.005>
- Linari, I., Juantorena, G. E., Ibáñez, A., Petroni, A., & Kamienkowski, J. E. (2022). Unveiling Trail Making Test: Visual and manual trajectories indexing multiple executive processes. *Scientific Reports*, 12(1), 14265. <https://doi.org/10.1038/s41598-022-16431-9>

- Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, 91(3), 295–327. <https://doi.org/10.1037/0033-295X.91.3.295>
- Luck, S. J., & Vogel, E. K. (1997a). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), Article 6657. <https://doi.org/10.1038/36846>
- Luck, S. J., & Vogel, E. K. (1997b). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), Article 6657. <https://doi.org/10.1038/36846>
- Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, 17(8), 391–400. <https://doi.org/10.1016/j.tics.2013.06.006>
- Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
- Luria, A. R. (1966). *Higher Cortical Functions in Man*. Springer Science & Business Media.
- Madsen, J., Júlio, S. U., Gucik, P. J., Steinberg, R., & Parra, L. C. (2021). Synchronized eye movements predict test scores in online video education. *Proceedings of the National Academy of Sciences*, 118(5), e2016980118. <https://doi.org/10.1073/pnas.2016980118>
- Maffei, A., & Angrilli, A. (2018). Spontaneous eye blink rate: An index of dopaminergic component of sustained attention and fatigue. *International Journal of Psychophysiology*, 123, 58–63. <https://doi.org/10.1016/j.ijpsycho.2017.11.009>
- Margulis, L. E., Louhau, M. R. S., & Ferreres, A. R. (2018). Baremo del Trail Making Test para Capital Federal y Gran Buenos Aires. *Revista Argentina de Ciencias del Comportamiento*, 10(3), 54–63. <https://doi.org/10.32348/1852.4206.v10.n3.19741>
- McKeown, K., Richards, E., Richardson, J., & Tales, A. (2021). The Trails Making Test. Does a Single Trial Reflect Performance Capability? *OBM Neurobiology*, 5(2), 1–29. <https://doi.org/10.21926/obm.neurobiol.2102100>
- Meyer, H. C., & Bucci, D. J. (2016). Neural and behavioral mechanisms of proactive and reactive inhibition. *Learning & Memory*, 23(10), 504–514. <https://doi.org/10.1101/lm.040501.115>

- Meyer, T., Kim, A. D., Spivey, M., & Yoshimi, J. (2024). Mouse tracking performance: A new approach to analyzing continuous mouse tracking data. *Behavior Research Methods*, 56(5), 4682–4694. <https://doi.org/10.3758/s13428-023-02210-5>
- Miner, T., & Ferraro, F. R. (1998). The Role of Speed of Processing, Inhibitory Mechanisms, and Presentation Order in Trail-Making Test Performance. *Brain and Cognition*, 38(2), 246–253. <https://doi.org/10.1006/brcg.1998.1034>
- Miyake, A., Emerson, M. J., & Friedman, N. P. (2000). ASSESSMENT OF EXECUTIVE FUNCTIONS IN CLINICAL SETTINGS: PROBLEMS AND RECOMMENDATIONS. *Seminars in Speech and Language*, 21(2), 169–183. <https://doi.org/10.1055/s-2000-7563>
- Moher, J., & Egeth, H. E. (2012). The ignoring paradox: Cueing distractor features leads first to selection, then to inhibition of to-be-ignored items. *Attention, Perception, & Psychophysics*, 74(8), 1590–1605. <https://doi.org/10.3758/s13414-012-0358-0>
- Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences*, 16(1), 72–80. <https://doi.org/10.1016/j.tics.2011.11.018>
- More, S., Antonopoulos, G., Hoffstaedter, F., Caspers, J., Eickhoff, S. B., & Patil, K. R. (2023). Brain-age prediction: A systematic comparison of machine learning workflows. *NeuroImage*, 270, 119947. <https://doi.org/10.1016/j.neuroimage.2023.119947>
- Mukadam, N., Wolters, F. J., Walsh, S., Wallace, L., Brayne, C., Matthews, F. E., Sacuiu, S., Skoog, I., Seshadri, S., Beiser, A., Ghosh, S., & Livingston, G. (2024). Changes in prevalence and incidence of dementia and risk factors for dementia: An analysis from cohort studies. *The Lancet. Public Health*, 9(7), e443–e460. [https://doi.org/10.1016/S2468-2667\(24\)00120-8](https://doi.org/10.1016/S2468-2667(24)00120-8)
- Munoz, D. P., & Everling, S. (2004). Look away: The anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, 5(3), Article 3. <https://doi.org/10.1038/nrn1345>
- Nichols, E., Steinmetz, J. D., Vollset, S. E., Fukutaki, K., Chalek, J., Abd-Allah, F., Abdoli, A., Abualhasan, A., Abu-Gharbieh, E., Akram, T. T., Al Hamad, H., Alahdab, F., Alanezi, F. M., Alipour, V., Almustanyir, S., Amu, H., Ansari, I., Arabloo, J., Ashraf, T., ... Vos, T. (2022). Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health*, 7(2), e105–e125. [https://doi.org/10.1016/S2468-2667\(21\)00249-8](https://doi.org/10.1016/S2468-2667(21)00249-8)

- Niehörster, D. C., Cornelissen, T. H. W., Holmqvist, K., Hooge, I. T. C., & Hessels, R. S. (2018). What to expect from your remote eye-tracker when participants are unrestrained. *Behavior Research Methods*, 50(1), 213–227. <https://doi.org/10.3758/s13428-017-0863-0>
- Opwonya, J., Ku, B., Lee, K. H., Kim, J. I., & Kim, J. U. (2023). Eye movement changes as an indicator of mild cognitive impairment. *Frontiers in Neuroscience*, 17. <https://doi.org/10.3389/fnins.2023.1171417>
- Opwonya, J., Wang, C., Jang, K.-M., Lee, K., Kim, J. I., & Kim, J. U. (2022). Inhibitory Control of Saccadic Eye Movements and Cognitive Impairment in Mild Cognitive Impairment. *Frontiers in Aging Neuroscience*, 14. <https://doi.org/10.3389/fnagi.2022.871432>
- Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27. <https://doi.org/10.1016/j.jbef.2017.12.004>
- Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable Webcam Eye Tracking Using User Interactions. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence - IJCAI 2016*. <https://par.nsf.gov/biblio/10024076-webgazer-scalable-webcam-eye-tracking-using-user-interactions>
- Park, S.-Y., & Schott, N. (2022). The trail-making-test: Comparison between paper-and-pencil and computerized versions in young and healthy older adults. *Applied Neuropsychology: Adult*, 29(5), 1208–1220. <https://doi.org/10.1080/23279095.2020.1864374>
- Partington, J. E., & Leiter, R. G. (1949). Partington's Pathways Test. *Psychological Service Center Journal*, 1, 11–20.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2018). *Scikit-learn: Machine Learning in Python* (arXiv:1201.0490). arXiv. <https://doi.org/10.48550/arXiv.1201.0490>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Piau, A., Wild, K., Mattek, N., & Kaye, J. (2019). Current State of Digital Biomarker Technologies for Real-Life, Home-Based Monitoring of Cognitive Function for Mild Cognitive Impairment to Mild

- Alzheimer Disease and Implications for Clinical Care: Systematic Review. *Journal of Medical Internet Research*, 21(8), e12785. <https://doi.org/10.2196/12785>
- Poldrack, R. A., Huckins, G., & Varoquaux, G. (2020). Establishment of Best Practices for Evidence for Prediction: A Review. *JAMA Psychiatry*, 77(5), 534–540. <https://doi.org/10.1001/jamapsychiatry.2019.3671>
- Poreh, A., Miller, A., Dines, P., & Levin, J. (2012). Decomposition of the Trail Making Test—Reliability and Validity of a Computer Assisted Method for Data Collection. *Archives of Assessment Psychology*, 2, 57–72.
- Raschka, S. (2020). *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning* (arXiv:1811.12808). arXiv. <https://doi.org/10.48550/arXiv.1811.12808>
- Recker, L., & Poth, C. H. (2023). Test–retest reliability of eye tracking measures in a computerized Trail Making Test. *Journal of Vision*, 23(8), 15. <https://doi.org/10.1167/jov.23.8.15>
- Reitan, R. M. (1958). Validity of the Trail Making Test as an Indicator of Organic Brain Damage. *Perceptual and Motor Skills*. (Sage CA: Los Angeles, CA). <https://doi.org/10.2466/pms.1958.8.3.271>
- Remington, R. W., Wu, S.-C., & Pashler, H. (2011). What determines saccade timing in sequences of coordinated eye and hand movements? *Psychonomic Bulletin & Review*, 18(3), 538–543. <https://doi.org/10.3758/s13423-011-0066-0>
- Revankar, G. S., Salian, A. C., Ozono, T., Suzuki, M., Kanemoto, H., Shigenobu, K., Nishio, Y., Yoshiyama, K., Yamamoto, Y., Revankar, S. S., Watari, A., Furuya, K., Ogasawara, I., Yoshida, N., Saeki, C., Kajiyama, Y., Nakatani, D., Hashimoto, M., Nagai, Y., ... Mochizuki, H. (2025). *Smartphone-based behavioral profiling for distinguishing Dementia with Lewy bodies from Alzheimer's Disease* (p. 2025.03.31.25324871). medRxiv. <https://doi.org/10.1101/2025.03.31.25324871>
- Rostagno, A. A. (2023). Pathogenesis of Alzheimer's Disease. *International Journal of Molecular Sciences*, 24(1), 107. <https://doi.org/10.3390/ijms24010107>
- Sabbagh, M. N., Boada, M., Borson, S., Chilukuri, M., Doraiswamy, P. M., Dubois, B., Ingram, J., Iwata, A., Porsteinsson, A. P., Possin, K. L., Rabinovici, G. D., Vellas, B., Chao, S., Vergallo, A., & Hampel, H. (2020). Rationale for Early Diagnosis of Mild Cognitive Impairment (MCI) supported by Emerging Digital Technologies. *The Journal of Prevention of Alzheimer's Disease*, 7(3), 158–164. <https://doi.org/10.14283/jpad.2020.19>

- Salemme, S., Lombardo, F. L., Lacorte, E., Sciancalepore, F., Remoli, G., Bacigalupo, I., Piscopo, P., Zamboni, G., Rossini, P. M., Cappa, S. F., Perani, D., Spadin, P., Tagliavini, F., Vanacore, N., & Ancidoni, A. (2025). The prognosis of mild cognitive impairment: A systematic review and meta-analysis. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 17(1), e70074. <https://doi.org/10.1002/dad2.70074>
- Salthouse, T. A. (2011). What cognitive abilities are involved in trail-making performance? *Intelligence*, 39(4), 222–232. <https://doi.org/10.1016/j.intell.2011.03.001>
- Salthouse, T. A., & Fristoe, N. M. (1995). Process analysis of adult age effects on a computer-administered Trail Making Test. *Neuropsychology*, 9(4), 518–528. <https://doi.org/10.1037/0894-4105.9.4.518>
- Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications, ETRA '00*, 71–78. <https://doi.org/10.1145/355017.355028>
- Sánchez-Cubillo, I., Periañez, J. A., Adrover-Roig, D., Rodríguez-Sánchez, J. M., Rios-Lago, M., Tirapu, J., & Barceló, F. (2009). Construct validity of the Trail Making Test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society*, 15(3), 438–450. <https://doi.org/10.1017/S1355617709090626>
- Sasse, L., Nicolaisen-Sobesky, E., Dukart, J., Eickhoff, S. B., Götz, M., Hamdan, S., Komeyer, V., Kulkarni, A., Lahnakoski, J. M., Love, B. C., Raimondo, F., & Patil, K. R. (2025). Overview of leakage scenarios in supervised machine learning. *Journal of Big Data*, 12(1), 135. <https://doi.org/10.1186/s40537-025-01193-8>
- Schoemann, M., O'Hora, D., Dale, R., & Scherbaum, S. (2021). Using mouse cursor tracking to investigate online cognition: Preserving methodological ingenuity while moving toward reproducible science. *Psychonomic Bulletin & Review*, 28(3), 766–787. <https://doi.org/10.3758/s13423-020-01851-3>
- Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, 50(2), 451–465. <https://doi.org/10.3758/s13428-017-0913-7>
- Sengupta, S., Wu, N., Varvello, M., Jana, K., Chen, S., & Han, B. (2025). From WebGL to WebGPU: A Reality Check of Browser-Based GPU Acceleration. *Proceedings of the 2025 ACM Internet Measurement Conference, IMC '25*, 1018–1024. <https://doi.org/10.1145/3730567.3764504>
- Shalom, D. E., & Sigman, M. (2015). Freedom and rules in human sequential performance: A refractory period in eye-hand coordination. *Journal of Vision*, 13(3), 4. <https://doi.org/10.1167/13.3.4>

- Shulman, K. I. (2000). Clock-drawing: Is it the ideal cognitive screening test? *International Journal of Geriatric Psychiatry*, 15(6), 548–561.
[https://doi.org/10.1002/1099-1166\(200006\)15:6<548::AID-GPS242>3.0.CO;2-U](https://doi.org/10.1002/1099-1166(200006)15:6<548::AID-GPS242>3.0.CO;2-U)
- Slim, M. S., & Hartsuiker, R. J. (2022). Moving visual world experiments online? A web-based replication of Dijkgraaf, Hartsuiker, and Duyck (2017) using PCIBex and WebGazer.js. *Behavior Research Methods*.
<https://doi.org/10.3758/s13428-022-01989-z>
- Smilkov, D., Thorat, N., Assogba, Y., Nicholson, C., Kreeger, N., Yu, P., Cai, S., Nielsen, E., Soegel, D., Bileschi, S., Terry, M., Yuan, A., Zhang, K., Gupta, S., Sirajuddin, S., Sculley, D., Monga, R., Corrado, G., Viegas, F., & Wattenberg, M. M. (2019). TensorFlow.js: Machine Learning For The Web and Beyond. *Proceedings of Machine Learning and Systems*, 1, 309–321.
https://proceedings.mlsys.org/paper_files/paper/2019/hash/acd593d2db87a799a8d3da5a860c028e-Abst
[ract.html](https://proceedings.mlsys.org/paper_files/paper/2019/hash/acd593d2db87a799a8d3da5a860c028e-Abst)
- Smith, K. A., Blease, C., Faurholt-Jepsen, M., Firth, J., Daele, T. V., Moreno, C., Carlbring, P., Ebner-Priemer, U. W., Koutsouleris, N., Riper, H., Mouchabac, S., Torous, J., & Cipriani, A. (2023). Digital mental health: Challenges and next steps. *BMJ Mental Health*, 26(1).
<https://doi.org/10.1136/bmjment-2023-300670>
- Sommers, M. S., & Dessenberger, S. (2022). The Effects of Aging on Processing Speed, Inhibitory Control, Executive Function, and Working Memory: An Integrated Account of Age-Cognition Relations Across the Adult Lifespan. In *Comprehensive Clinical Psychology, Second Edition* (pp. 40–51). Elsevier.
<https://doi.org/10.1016/B978-0-12-818697-8.00090-X>
- Song, J., Cho, E., Lee, H., Lee, S., Kim, S., & Kim, J. (2025). Development of Neurodegenerative Disease Diagnosis and Monitoring from Traditional to Digital Biomarkers. *Biosensors*, 15(2), 102.
<https://doi.org/10.3390/bios15020102>
- Soukupova, T., & Čech, J. (2016). Real-Time Eye Blink Detection using Facial Landmarks. *21 St Computer Vision Winter Workshop*.
- Sperl, L., Ambrus, G. G., Kaufmann, J. M., Schweinberger, S. R., & Cañal-Bruland, R. (2021). Electrophysiological correlates underlying interference control in motor tasks. *Biological Psychology*, 163, 108138. <https://doi.org/10.1016/j.biopsycho.2021.108138>

- Spikman, J. M., Kiers, H. A., Deelman, B. G., & van Zomeren, A. H. (2001). Construct validity of concepts of attention in healthy controls and patients with CHI. *Brain and Cognition*, 47(3), 446–460. <https://doi.org/10.1006/brcg.2001.1320>
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, 102(29), 10393–10398. <https://doi.org/10.1073/pnas.0503903102>
- Theeuwes, J., Belopolsky, A., & Olivers, C. N. L. (2009). Interactions between working memory, attention and eye movements. *Acta Psychologica, Spatial Working Memory and Imagery: From Eye Movements to Grounded Cognition*, 132(2), 106–114. <https://doi.org/10.1016/j.actpsy.2009.01.005>
- Thornton, I. M., & Horowitz, T. S. (2020). Searching Through Alternating Sequences: Working Memory and Inhibitory Tagging Mechanisms Revealed Using the MILO Task. *I-Perception*, 11(5), 2041669520958018. <https://doi.org/10.1177/2041669520958018>
- Thorpe, A., Friedman, J., Evans, S., Nesbitt, K., & Eidels, A. (2022). Mouse Movement Trajectories as an Indicator of Cognitive Workload. *International Journal of Human–Computer Interaction*, 38(15), 1464–1479. <https://doi.org/10.1080/10447318.2021.2002054>
- Tombaugh, T. N. (2004). Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, 19(2), 203–214. [https://doi.org/10.1016/S0887-6177\(03\)00039-8](https://doi.org/10.1016/S0887-6177(03)00039-8)
- Tombaugh, T. N., & McIntyre, N. J. (1992). The Mini-Mental State Examination: A Comprehensive Review. *Journal of the American Geriatrics Society*, 40(9), 922–935. <https://doi.org/10.1111/j.1532-5415.1992.tb01992.x>
- Torralva, T., Roca, M., Gleichgerricht, E., López, P., & Manes, F. (2009). INECO Frontal Screening (IFS): A brief, sensitive, and specific tool to assess executive functions in dementia—CORRECTED VERSION. *Journal of the International Neuropsychological Society*, 15(5), 777–786. <https://doi.org/10.1017/S1355617709990415>
- vanWell, A., & Tanaka, J. (2022). Gazer: A New Tool for Web-based Eye Tracking. *PsyArXiv Preprints*. <https://doi.org/10.31234/osf.io/2pckv>
- Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7(1), 91. <https://doi.org/10.1186/1471-2105-7-91>

- Varoquaux, G. (2018). Cross-validation failure: Small sample sizes lead to large error bars. *NeuroImage, New Advances in Encoding and Decoding of Brain Signals*, 180, 68–77. <https://doi.org/10.1016/j.neuroimage.2017.06.061>
- Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Pedregosa, F., & Mueller, A. (2015). Scikit-learn: Machine Learning Without Learning the Machinery. *GetMobile: Mobile Comp. and Comm.*, 19(1), 29–33. <https://doi.org/10.1145/2786984.2786995>
- Verbruggen, F., Logan, G. D., & Stevens, M. A. (2008). STOP-IT: Windows executable software for the stop-signal paradigm. *Behavior Research Methods*, 40(2), 479–483. <https://doi.org/10.3758/BRM.40.2.479>
- Verma, R., & Dhyani, I. (2025). Development and Validation of an Android-based Trail Making Test Application (TMT App) for Cognitive Evaluation in Healthy and Depressed Individuals. *Indian Journal of Psychological Medicine*, 47(3), 268–274. <https://doi.org/10.1177/02537176241247789>
- Verón, G. L., Juantorena, G. E., Keller, G., Crivelli, L., & Kamienkowski, J. E. (2025). Eye tracking as a diagnostic tool in Alzheimer's disease, mild cognitive impairment, and related dementias: A systematic review. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 18(1), e70238. <https://doi.org/10.1002/dad2.70238>
- Vidal, J. (2020). *Cognition.run* [Computer software]. <https://www.cognition.run>
- von Ahn, L. (2006). Games with a purpose. *Computer*, 39(6), 92–94. <https://doi.org/10.1109/MC.2006.196>
- von Ahn, L. (2013). Duolingo: Learn a language for free while helping to translate the web. *Proceedings of the 2013 International Conference on Intelligent User Interfaces, IUI '13*, 1–2. <https://doi.org/10.1145/2449396.2449398>
- von Ahn, L., Maurer, B., McMillen, C., Abraham, D., & Blum, M. (2008). reCAPTCHA: Human-Based Character Recognition via Web Security Measures. *Science*, 321(5895), 1465–1468. <https://doi.org/10.1126/science.1160379>
- Wang, Q., Jiang, S., Chen, Z., Cao, X., Li, Y., Li, A., Ma, Y., Cao, T., & Liu, X. (2025). Anatomizing Deep Learning Inference in Web Browsers. *ACM Transactions on Software Engineering and Methodology*, 34(2), 47:1-47:43. <https://doi.org/10.1145/3688843>
- Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, 58(3), 475–482.

- Witte, A., & Lins, C. (2023). A Real-Time Eye Tracking System for the Detection of Eye Blinks. In R. Röhrig, N. Grabe, M. Haag, U. Hübner, U. Sax, C. Oliver Schmidt, M. Sedlmayr, & A. Zapf (Eds.), *Studies in Health Technology and Informatics*. IOS Press. <https://doi.org/10.3233/SHTI230719>
- Wolpert, D. H. (1996). The Lack of A Priori Distinctions Between Learning Algorithms. *Neural Computation*, 8(7), 1341–1390. <https://doi.org/10.1162/neco.1996.8.7.1341>
- Wölwer, W., Falkai, P., Streit, M., & Gaebel, W. (2003). Trait Characteristic of Impaired Visuomotor Integration during Trail-Making Test B Performance in Schizophrenia. *Neuropsychobiology*, 48(2), 59–67. <https://doi.org/10.1159/000072878>
- Wölwer, W., & Gaebel, W. (2002). Impaired Trail-Making Test-B performance in patients with acute schizophrenia is related to inefficient sequencing of planning and acting. *Journal of Psychiatric Research*, 36(6), 407–416. [https://doi.org/10.1016/S0022-3956\(02\)00050-X](https://doi.org/10.1016/S0022-3956(02)00050-X)
- Wölwer, W., Stroth, S., Brinkmeyer, J., & Gaebel, W. (2012). Electrophysiological correlates of planning and monitoring in first episode schizophrenia. *Psychiatry Research: Neuroimaging*, 203(1), 83–88. <https://doi.org/10.1016/j.psychresns.2011.11.009>
- Woodman, G. F., & Luck, S. J. (2007). Do the contents of visual working memory automatically influence attentional selection during visual search? *Journal of Experimental Psychology: Human Perception and Performance*, 33(2), 363–377. <https://doi.org/10.1037/0096-1523.33.2.363>
- Woods, D. L., Wyma, J. M., Herron, T. J., & Yund, E. W. (2015). The Effects of Aging, Malingering, and Traumatic Brain Injury on Computerized Trail-Making Test Performance. *PLOS ONE*, 10(6), e0124345. <https://doi.org/10.1371/journal.pone.0124345>
- Wulff, D. U., Kieslich, P. J., Henninger, F., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2026). Publisher Correction: Movement tracking of psychological processes: A tutorial using mousetrap. *Behavior Research Methods*, 58(2), 53. <https://doi.org/10.3758/s13428-025-02894-x>
- Xu, P., Ehinger, K. A., Zhang, Y., Finkelstein, A., Kulkarni, S. R., & Xiao, J. (2015, April 25). *TurkerGaze: Crowdsourcing Saliency with Webcam based Eye Tracking*. arXiv.Org. <https://doi.org/10.48550/arXiv.1504.06755>
- Yang, X., & Krajbich, I. (2020). *Webcam-based online eye-tracking for behavioral research*. <https://doi.org/10.31234/osf.io/qhme6>
- Yarbus, A. L. (1967). *Eye Movements and Vision*. Springer US. <https://doi.org/10.1007/978-1-4899-5379-7>

- Yu, S., Srimaharaj, W., & Chaisricharoen, R. (2024). EEG Age Prediction via Elastic Net Linear Regression. *2024 8th International Conference on Information Technology (InCIT)*, 266–271.
<https://doi.org/10.1109/InCIT63192.2024.10810600>
- Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)*.
<https://osf.io/md832/overview>
- Zhang, H., & Yin, L. (2025). Robust Camera-Based Eye-Tracking Method Allowing Head Movements and Its Application in User Experience Research. *Journal of Eye Movement Research*, 18(6), 71.
<https://doi.org/10.3390/jemr18060071>

Apéndice A: Repositorios de Código y Recursos Digitales

En este apéndice se reúnen las herramientas de software desarrolladas o mejoradas durante esta tesis. Cada entrada enlaza al recurso correspondiente: código fuente, plataforma o documentación técnica. Todos los recursos están reunidos en una página de acceso centralizado: <https://gej1.github.io/tesis-apendice/>



Escanear para acceder

A.1. Plataforma DataPruebas (Caps. 4, 5 y 6)

Infraestructura para la recolección de datos multimodales en entornos web. Incluye el sistema de gestión de participantes, la integración de tareas cognitivas y el almacenamiento persistente.

Plataforma: datapuebas.org

A.2. Implementación del Virtual Chinrest (Li et al., 2020) (Caps. 5 y 6)

Permite estimar la distancia entre el participante y la pantalla, y ajustar el contenido de la página de *jsPsych* a una dimensión física conocida (por ejemplo, garantizando que un estímulo de 200 píxeles de ancho mida 2,2 cm en el monitor del participante).

Plugin oficial: <https://www.jspsych.org/v8/plugins/virtual-chinrest/>

A.3. Repositorio de tareas cognitivas en jsPsych (Cap. 5)

Go/No-Go, Change Detection Task y Trail Making Test, este último con su propio *plugin* de *jsPsych* que permite armar distintas versiones de la tarea. La cuarta tarea utilizada en el Capítulo 5 (Stop Signal Task) fue integrada a partir de la implementación de código abierto STOP-IT (github.com/fredvbrug/STOP-IT).

Repositorio: https://github.com/NeuroLIAA/Neuro_psych_tasks

A.4. Biblioteca NeuroTask (Caps. 3 y 5)

Librería en *Python* para el preprocesamiento de series temporales conductuales y la extracción de características cinemáticas de alta resolución. Incluye interpolación de señales, segmentación de estados basada en velocidad (Búsqueda, Desplazamiento, Vacilación).

Repositorio: github.com/NeuroLIAA/neurotask

A.5. Biblioteca Pyxations (Cap. 6)

Biblioteca en *Python* para el análisis de datos de seguimiento ocular, tanto en datos crudos como en conjuntos de datos procesados. Facilita la extracción de información a partir de movimientos oculares complejos dentro del ecosistema de *Python*. Disponible en *PyPI* (*pip install pyxations*).

Repositorio: github.com/NeuroLIIA/pyxations

Documentación: <https://neuroliaa.github.io/pyxations/>

A.6. Web Eye-Tracker: seguimiento ocular por cámara web (Cap. 6)

Prototipo de eye-tracking remoto por cámara web, construido en torno a la tarea de antisacadas e integrado a *jsPsych*. Incluye las modificaciones técnicas sobre *WebGazer*: la migración a *TensorFlow.js* (*Face mesh*), la detección de parpadeos (EAR) y la calibración explícita de 9 puntos.

Prototipo (sistema de experimentos): github.com/agustin-penas/rastreador-ocular

Núcleo WebGazer: github.com/NeuroLIIA/WebGazer

Apéndice B: Repositorios de datos

B.1. Datos del cTMT en pacientes con DCL (Cap. 3)

Trayectorias del cursor y evaluaciones neuropsicológicas de la cohorte italiana de pacientes con deterioro cognitivo leve y controles.

Repositorio (OSF): <https://osf.io/nm9xy/>

B.2. Datos del cTMT online (Cap. 5)

Trayectorias del cursor y desempeño en las tareas cognitivas recolectadas en el estudio online a gran escala.

Repositorio (OSF): <https://osf.io/9rnp4/>

B.3. Datos de la tarea de antisacadas (Cap. 6)

Datos anonimizados del experimento de seguimiento ocular web (prosacadas y antisacadas).

Repositorio (GitHub):
https://github.com/GEJ1/et_webcam_experiments/tree/main/antisaccade_experiment/raw_data

Material suplementario

Figura C1.S1: Medidas de rendimiento para dos umbrales de objetivos diferentes

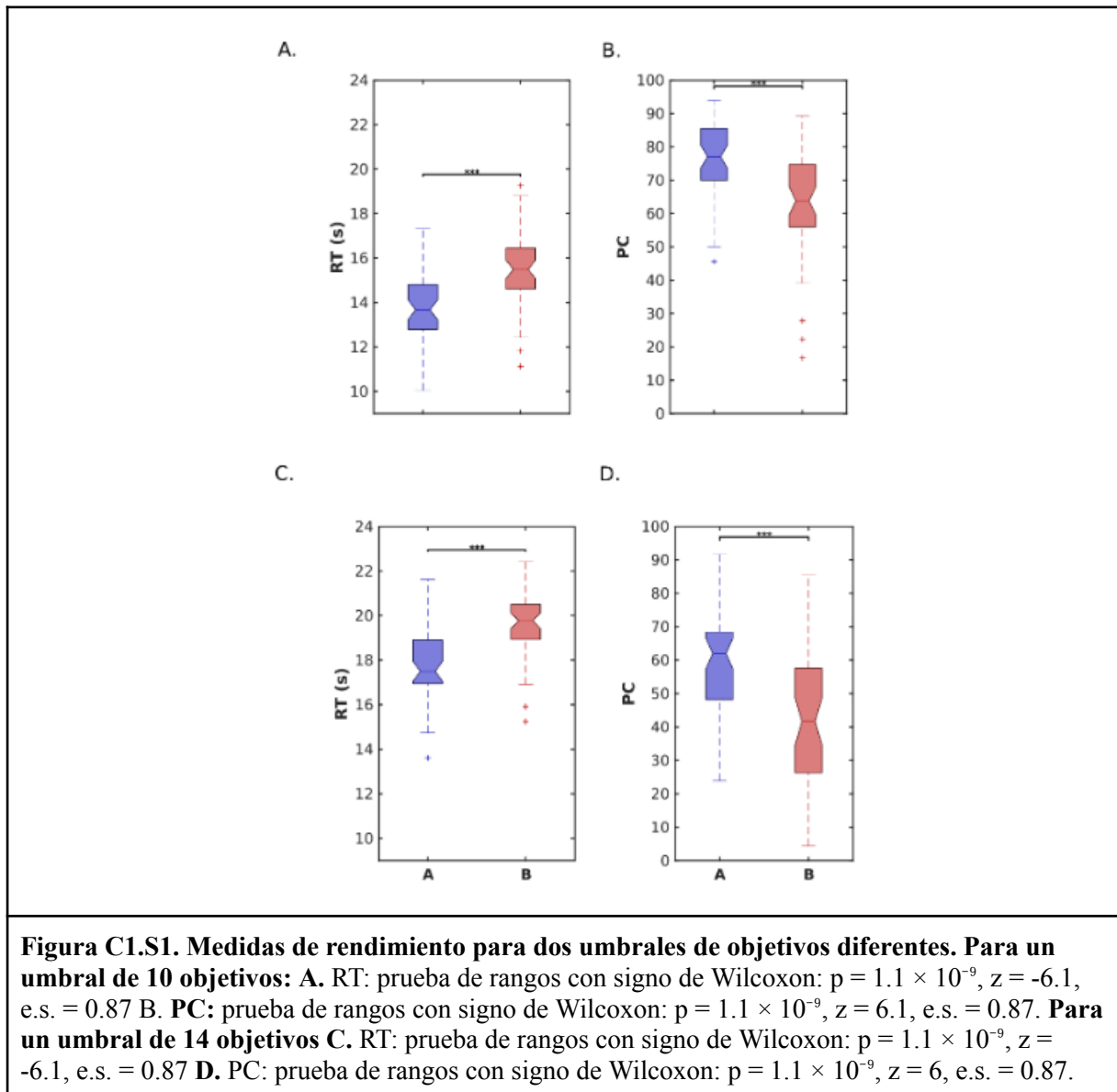


Figura C1.S2: Medidas de rendimiento para el primer tercio de los ensayos

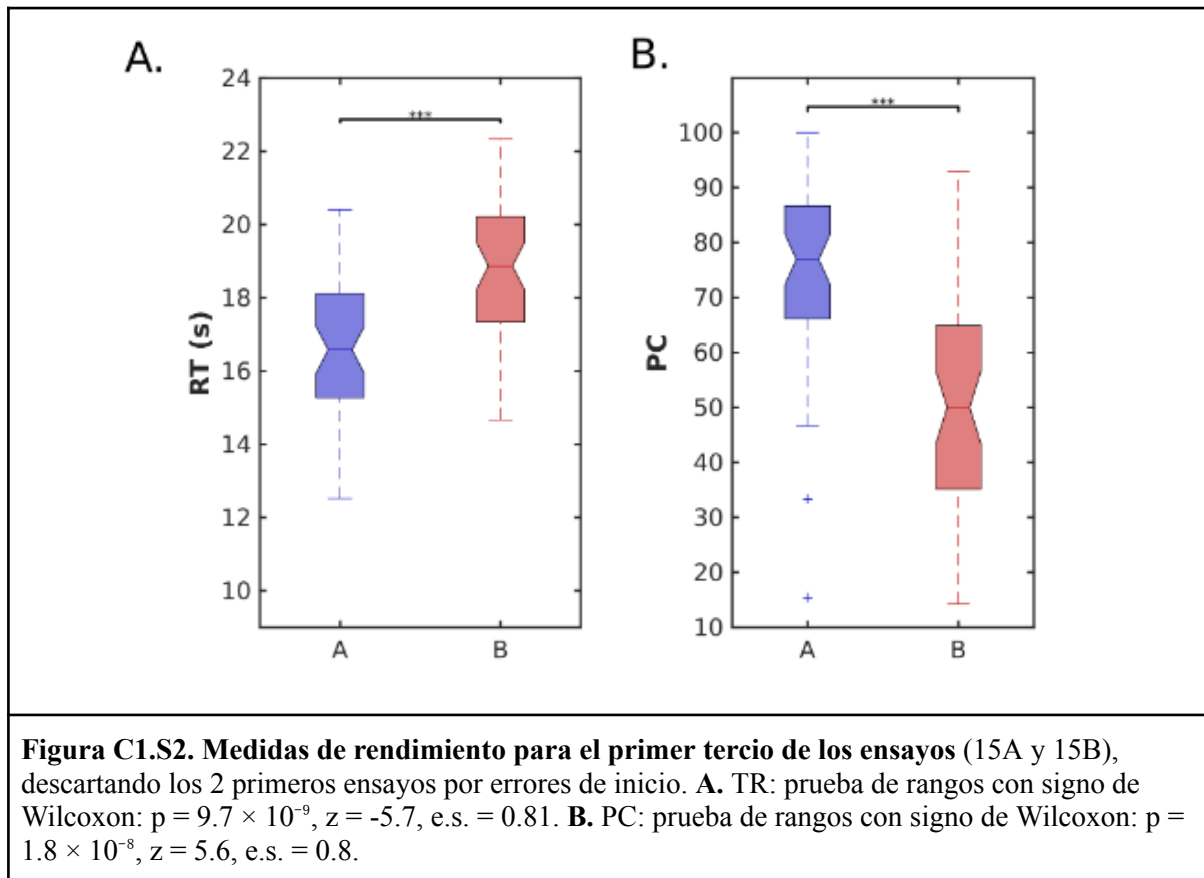


Figura C1.S3: Distribución espacial de las trayectorias exploradas por la mano al fijar la vista en un nuevo ítem para un participante

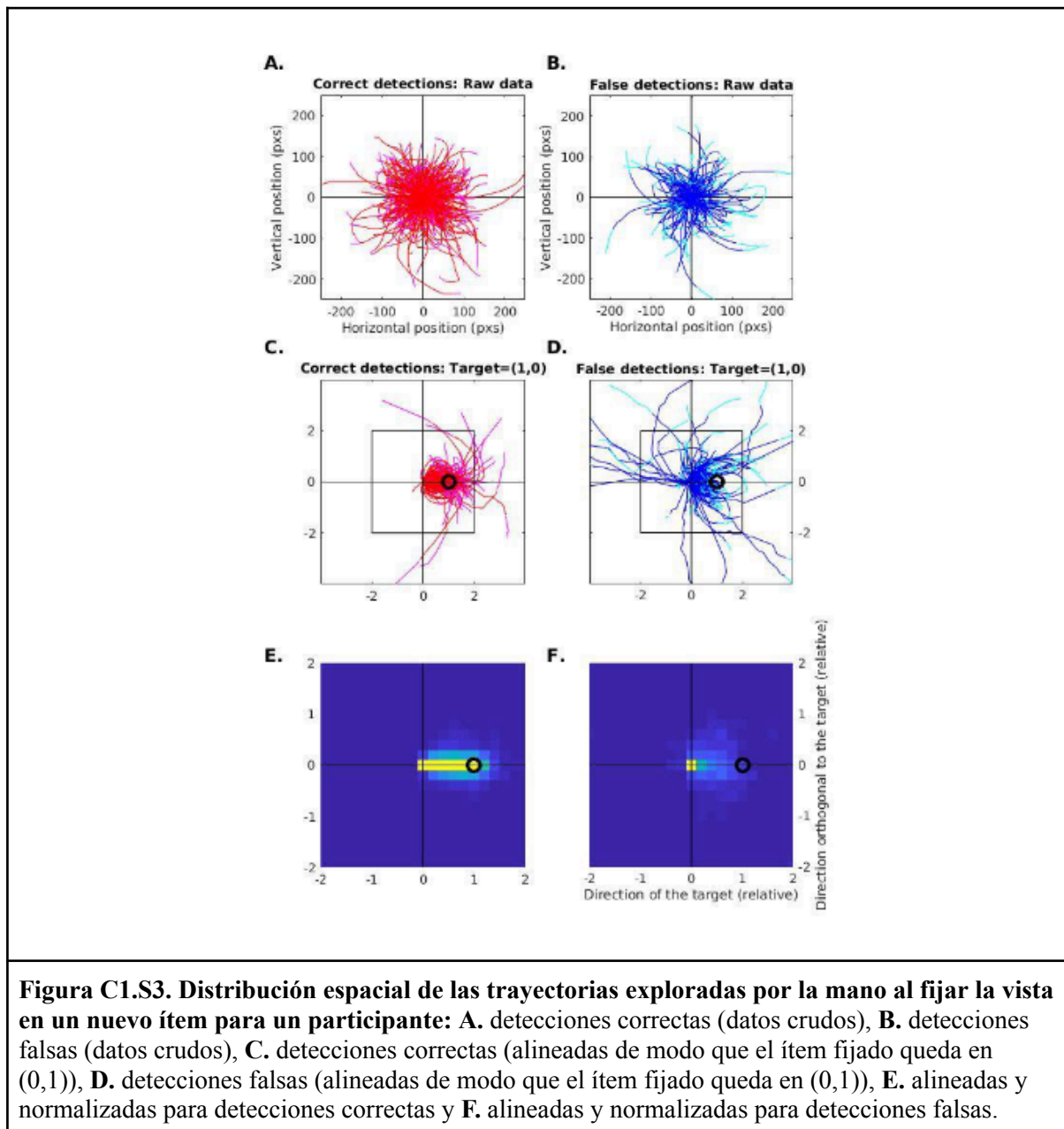


Figura C5.S1: Valores medios |SHAP| - Ridge - Accuracy

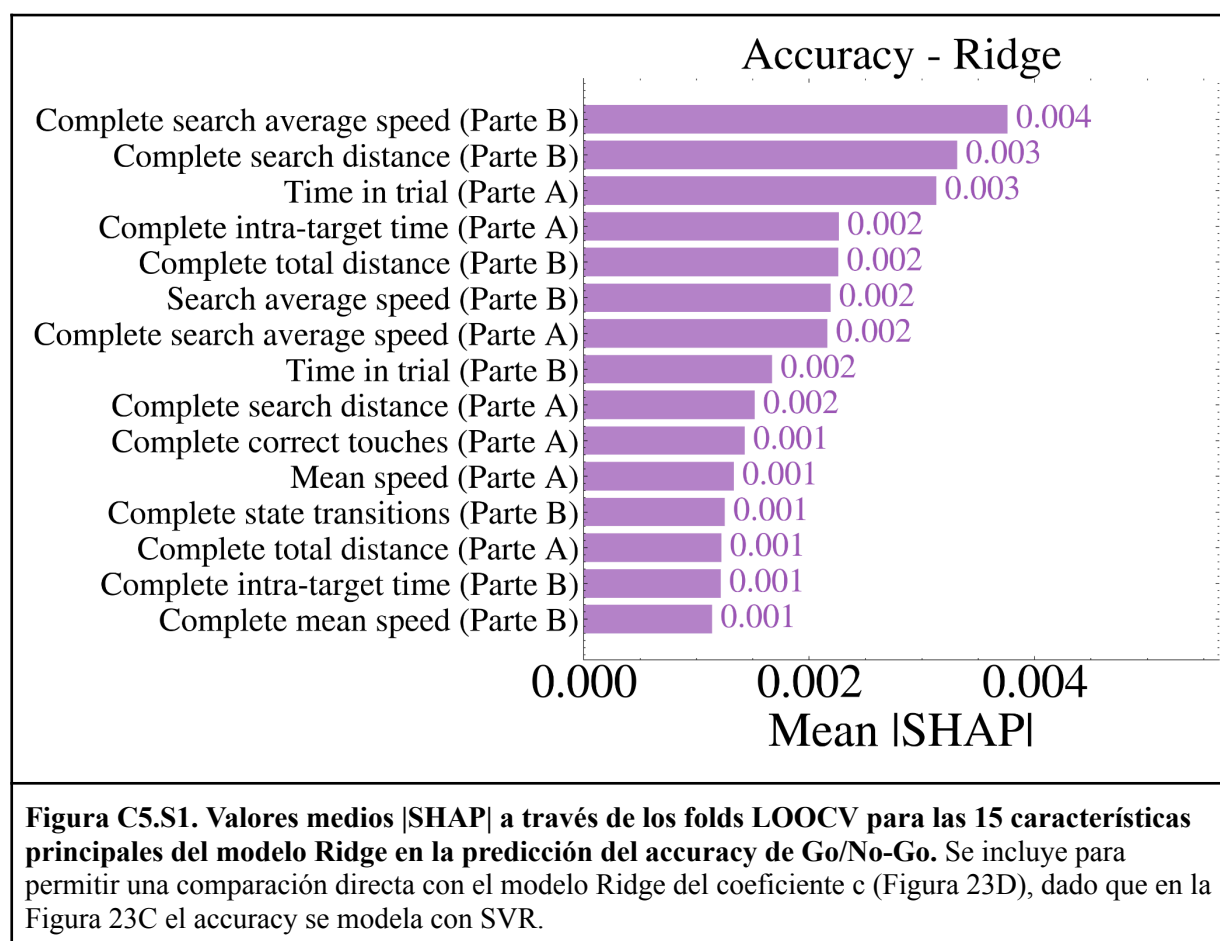


Tabla C3.S1: Atributos extraídos del cursor (estudio presencial)

Tabla C3.S1. Atributos extraídos de la serie temporal del movimiento del cursor (estudio presencial). Las mismas variables se calcularon por separado para la Parte A, la Parte B y el cociente B/A, lo que da lugar a tres variantes de cada atributo. El prefijo <i>Complete</i> indica la variante sobre el ensayo completo (véase Métodos).	
Atributos	Descripción
Valid sum	Indica la cantidad de ensayos válidos (es decir, aquellos con al menos ocho objetivos conectados correctamente).
Total distance	Distancia total recorrida por el cursor.
Time in trial	Duración total del ensayo.
Correct touches	Cantidad de objetivos conectados correctamente.
Mean speed	Velocidad media del cursor.
STD speed	Desvío estándar de la velocidad del cursor.
Peak speed	Velocidad máxima del cursor observada.
Mean acceleration	Aceleración media del cursor.
STD acceleration	Desvío estándar de la aceleración del cursor.
Peak acceleration	Aceleración máxima observada.
Mean absolute acceleration	Media de los valores absolutos de aceleración.
STD absolute acceleration	Desvío estándar de los valores absolutos de aceleración.
Peak absolute acceleration	Aceleración absoluta máxima observada.
Mean negative acceleration	Media de los valores de aceleración negativa (desaceleración).
STD negative acceleration	Desvío estándar de los valores de aceleración negativa.
Peak negative acceleration	Valor de aceleración negativa más extremo observado.
Hesitations	Cantidad de períodos de <i>hesitation</i> distintos detectados durante el ensayo.
Hesitation average speed	Velocidad media del cursor durante los segmentos de <i>hesitation</i> .
Hesitation distance	Distancia total recorrida durante los segmentos de <i>hesitation</i> .
Hesitation time	Suma del tiempo transcurrido en los segmentos de <i>hesitation</i> .
Hesitation max duration	Duración del período de <i>hesitation</i> más largo durante el ensayo.
Hesitation average duration	Duración media de los períodos de <i>hesitation</i> individuales detectados durante el ensayo. En las figuras del estudio presencial aparece rotulada como “Average duration”; es la misma métrica.
Travel distance	Distancia total recorrida durante los segmentos de <i>travel</i> .
Travel time	Suma del tiempo transcurrido en los segmentos de <i>travel</i> .
Travel average speed	Velocidad media del cursor durante los segmentos de <i>travel</i> .
Search distance	Distancia total recorrida durante los segmentos de <i>search</i> .
Search time	Suma del tiempo transcurrido en los segmentos de <i>search</i> .
Search average speed	Velocidad media del cursor durante los segmentos de <i>search</i> .
Inter-target time	Suma del tiempo transcurrido fuera del área de los objetivos.
Intra-target time	Suma del tiempo permanecido dentro del área de un objetivo antes de abandonarlo.
Area difference from ideal	Diferencia entre el área de la trayectoria real y la de la trayectoria ideal.

Distance difference from ideal	Diferencia entre la longitud de la trayectoria recorrida y la longitud de la recta ideal que conecta con el siguiente objetivo.
State transitions	Cantidad total de transiciones de un estado (<i>hesitation</i> , <i>search</i> o <i>travel</i>) a otro.
Zigzag amplitude	Diferencia temporal media entre transiciones consecutivas de número y letra en la Parte B.

Tabla C3.S2: Grilla de hiperparámetros de clasificación

Tabla C3.S2. Modelos de clasificación del estudio presencial.		
Nombre	Hiperparámetro	Valores
Random Forest Classifier	n_estimators	100; 200; 500
	max_depth	None; 8; 16
	min_samples_leaf	1; 2; 5; 10
	max_features	sqrt; log2
Support Vector Machine (SVC)	kernel	linear; rbf
	C	0,1; 1; 10
	gamma (para kernel rbf)	scale; auto
Logistic Regression	penalty	l1; l2; elasticnet
	C	0,1; 1; 10
	solver	lbfgs; liblinear; saga
	l1_ratio (para penalización elasticnet)	0,5
XGBoost Classifier	n_estimators	100; 200
	max_depth	3; 6
	learning_rate	0,1; 0,3
	subsample	0,8; 1,0
	colsample_bytree	0,8; 1,0

Tabla C3.S3: Grilla de hiperparámetros de regresión (capítulos 3 y 5)

Tabla C3.S3. Modelos de regresión (estudio presencial, Capítulo 3, y estudio online, Capítulo 5). Las grillas son idénticas en ambos estudios, salvo el XGBoost Regressor: en el presencial solo se varió <i>n_estimators</i> (100; 200); en el online se ajustaron además <i>max_depth</i> , <i>learning_rate</i> , <i>subsample</i> y <i>colsample_bytree</i> .		
Nombre	Hiperparámetro	Valores
Random Forest Regressor	n_estimators	100; 200; 500
	max_depth	None; 8; 16
	min_samples_leaf	2; 5; 10
	max_features	sqrt; log2
Support Vector Machine (SVR)	C	0,01; 0,1; 1; 10; 100; 1000
	epsilon	0,01; 0,1; 0,5; 1,0
	gamma (para kernel rbf)	scale; auto
Linear Regression	Sin hiperparámetros	-
Ridge Regression	alpha	0,0001; 0,001; 0,01; 0,1; 1,0; 5; 10; 100; 1000; 10000
Lasso Regression	alpha	0,0001; 0,001; 0,01; 0,1; 1,0; 5; 10,0
XGBoost Regressor	n_estimators	100; 200
	max_depth (solo online)	3; 6
	learning_rate (solo online)	0,1; 0,3
	subsample (solo online)	0,8; 1,0
	colsample_bytree (solo online)	0,8; 1,0
Elastic Net	alpha	0,0001; 0,001; 0,01; 0,1; 1; 10; 100
	l1_ratio	0,05; 0,2; 0,5; 0,8; 0,95

Tabla C5.S1: Atributos extraídos del cursor (estudio online)

Tabla C5.S1. Atributos extraídos de la serie temporal del movimiento del cursor (estudio online). Las mismas variables se calcularon por separado para la Parte A y la Parte B, lo que da lugar a dos variantes de cada atributo. El prefijo <i>Complete</i> indica la variante sobre el ensayo completo (véase Métodos).	
Atributos	Descripción
Time in trial	Duración total del ensayo.
Correct touches	Cantidad de objetivos conectados correctamente.
Wrong target touches	Cantidad de objetivos incorrectos alcanzados, es decir, distintos del siguiente esperado en la secuencia.
Area difference from ideal	Diferencia entre el área de la trayectoria real y la de la trayectoria ideal.
Distance difference from ideal	Diferencia entre la longitud de la trayectoria recorrida y la longitud de la recta ideal que conecta con el siguiente objetivo.
Inter-target time	Suma del tiempo transcurrido fuera del área de los objetivos.
Intra-target time	Suma del tiempo permanecido dentro del área de un objetivo antes de abandonarlo.
Mean speed	Velocidad media del cursor.
SD speed	Desvío estándar de la velocidad del cursor.
Peak speed	Velocidad máxima del cursor observada.
Number of crosses	Cantidad de veces que la trayectoria del cursor se cruza consigo misma, excluyendo las intersecciones entre segmentos temporalmente adyacentes.
State transitions	Cantidad total de transiciones de un estado (<i>hesitation</i> , <i>search</i> o <i>travel</i>) a otro.
Total distance	Distancia total recorrida por el cursor.
Hesitations	Cantidad de períodos de <i>hesitation</i> distintos detectados durante el ensayo.
Hesitation max duration	Duración del período de <i>hesitation</i> más largo durante el ensayo.
Hesitation average duration	Duración media de los períodos de <i>hesitation</i> individuales detectados durante el ensayo.
Hesitation ratio	Proporción del tiempo en <i>hesitation</i> respecto del tiempo total (<i>travel</i> + <i>hesitation</i>): $\text{hesitation_time} / (\text{travel_time} + \text{hesitation_time})$.
Hesitation time	Suma del tiempo transcurrido en los segmentos de <i>hesitation</i> .
Hesitation average speed	Velocidad media del cursor durante los segmentos de <i>hesitation</i> .
Hesitation distance	Distancia total recorrida durante los segmentos de <i>hesitation</i> .
Search average speed	Velocidad media del cursor durante los segmentos de <i>search</i> .
Search distance	Distancia total recorrida durante los segmentos de <i>search</i> .
Search time	Suma del tiempo transcurrido en los segmentos de <i>search</i> .
Travel average speed	Velocidad media del cursor durante los segmentos de <i>travel</i> .
Travel distance	Distancia total recorrida durante los segmentos de <i>travel</i> .
Travel time	Suma del tiempo transcurrido en los segmentos de <i>travel</i> .