



UNIVERSIDAD DE MURCIA

FACULTAD DE INFORMÁTICA

TRABAJO FIN DE MÁSTER

**Simulación de NeuroStrikes
en el Hipocampo**

Autor:

Juan Luis Serradilla Tormos

Tutores:

Sergio López Bernal

2 de julio de 2025

Índice

	Página
Declaración firmada sobre originalidad del trabajo	I
Resumen	II
Extended Abstract	III
1. Introducción	1
2. Background	4
2.1. El cerebro y el sistema nervioso	4
2.2. Hipocampo	6
2.2.1. Definición del hipocampo	6
2.2.2. Anatomía del hipocampo	7
2.2.3. Funciones del hipocampo	8
2.3. Memoria	9
2.3.1. Ritmos cerebrales en la consolidación de la memoria	10
2.4. BCIs	12
2.5. Guerra cognitiva	15
2.5.1. Síndrome de La Habana	15
3. Estado del Arte	17
3.1. Definición de los primeros ciberataques neuronales	17
3.2. Definición de nuevos ciberataques neuronales: ataques inhibitorios	18
3.3. Primeros resultados con una topología neuronal realista	19
3.4. guerra cognitiva y simulación de NeuroStrikes	19
4. Objetivos y Metodología	21
4.1. Objetivos	21
4.2. Metodología	21
5. Diseño de la Solución	22
5.1. Diseño	22
5.2. Implementación	25
5.2.1. Preparación del entorno de trabajo	25
5.2.2. Métodos de evaluación y comprobación de la estabilidad de la si- mulación	25
5.2.3. Archivos, funciones y estructura del proyecto	28

6. Análisis de Resultados	33
6.1. Análisis de la estabilidad de las simulaciones	33
6.2. Replicación y comparativa de los resultados de Töllke	35
6.2.1. Estado Saludable	35
6.2.2. Estados bajo ataques de modificación de parámetros	37
6.3. Análisis de los ataques con las nuevas métricas	43
6.4. Asociación de resultados con daños cognitivos	48
7. Conclusiones y Trabajos Futuros	52

Índice de figuras

1.	Esquema del sistema límbico. [8]	4
2.	Ilustración de una neurona cerebral. [31]	5
3.	Esquema de la sinapsis química. [31]	6
4.	Fotografía del hipocampo. [8]	7
5.	Esquema de las zonas del hipocampo [4].	8
6.	Ciclo de funcionamiento bidireccional de las BCIs que representa, en negro, las fases comunes para la adquisición de datos neuronales y la estimulación cerebral. (Lado izquierdo) Representación, en azul, de los procesos realizados y de los datos transferidos en cada fase del proceso de adquisición de datos neuronales. Este ciclo puede considerarse un proceso cíclico, pues comienza y termina en la misma fase. (Lado derecho) Representación, en rojo, de los procesos y transiciones de cada fase que conforman el proceso de estimulación. [17]	13
7.	Esquema del diseño de la solución en el que se observa cómo funciona el simulador y cómo se generan los diferentes ataques.	22
8.	Estructura de archivos del simulador.	29
9.	Resultados de 10 simulaciones individuales con los parámetros saludables. Se mide la duración de los SWRs y se representa con un diagrama de caja. El triángulo verde representa la media.	33
10.	Resultados de 10 simulaciones individuales con los parámetros saludables. Se mide la frecuencia pico de los SWRs y se representa con un diagrama de caja. El triángulo verde representa la media.	34
11.	Resultados de 10 simulaciones individuales con los parámetros saludables. Se mide la Potencia Espectral de las diferentes bandas de frecuencia y se representa con un diagrama de caja. El triángulo verde representa la media.	34
12.	Media de la frecuencia pico de los Sharp Wave Ripples en una simulación saludable. Resultados de Töllke [32]	35
13.	Potencia Espectral de las diferentes bandas en una simulación saludable. Resultados de Töllke [32]	35
14.	Media de la frecuencia pico de los Sharp Wave Ripples en cada experimento saludable, cada uno con cuatro simulaciones. Resultados propios.	36
15.	Potencia Espectral de las diferentes bandas en varios experimentos saludable, cada uno con cuatro simulaciones. Resultados propios.	36
16.	Frecuencia pico media para cada valor de N_{max} en simulaciones individuales. Resultados de Töllke [32].	37
17.	Duración media para cada valor de N_{max} en simulaciones individuales. Resultados de Töllke [32].	37

18.	Valores PS en cada banda de frecuencia para cada valor de N_{max} en simulaciones individuales. Resultados de Töllke [32].	37
19.	Frecuencia pico media para cada valor de g_{max_e} en simulaciones individuales. Resultados de Töllke [32].	37
20.	Duración media para cada valor de g_{max_e} en simulaciones individuales. Resultados de Töllke [32].	37
21.	PS en cada banda de frecuencia para cada valor de g_{max_e} en simulaciones individuales. Resultados de Töllke [32].	37
22.	Frecuencia pico media para cada valor de gCAN en simulaciones individuales. Resultados de Töllke [32].	38
23.	Duración media para cada valor de gCAN en simulaciones individuales. Resultados de Töllke [32].	38
24.	PS en cada banda de frecuencia para cada valor de gCAN en simulaciones individuales. Resultados de Töllke [32].	38
25.	Frecuencia pico media para cada valor de gACh en simulaciones individuales. Resultados de Töllke [32].	38
26.	Duración media para cada valor de gACh en simulaciones individuales. Resultados de Töllke [32].	38
27.	PS en cada banda de frecuencia para cada valor de gACh en simulaciones individuales. Resultados de Töllke [32].	38
28.	Frecuencia pico media para cada valor de N_{max} y g_{max_e} en simulaciones individuales. Resultados de Töllke [32].	39
29.	Duración media para cada valor de N_{max} y g_{max_e} en simulaciones individuales. Resultados de Töllke [32].	39
30.	PS en cada banda de frecuencia para cada valor de N_{max} y g_{max_e} en simulaciones individuales. Resultados de Töllke [32].	39
31.	Frecuencia pico media para cada valor de gCAN y gACh en simulaciones individuales. Resultados de Töllke [32].	39
32.	Duración media para cada valor de gCAN y gACh en simulaciones individuales. Resultados de Töllke [32].	39
33.	PS en cada banda de frecuencia para cada valor de gCAN y gACh en simulaciones individuales. Resultados de Töllke [32].	39
34.	Frecuencia pico media para cada valor de N_{max} , g_{max_e} , gCAN y gACh en simulaciones individuales. Resultados de Töllke [32].	40
35.	Duración media para cada valor de N_{max} , g_{max_e} , gCAN y gACh en simulaciones individuales. Resultados de Töllke [32].	40
36.	PS en cada banda de frecuencia para cada valor de N_{max} , g_{max_e} , gCAN y gACh en simulaciones individuales. Resultados de Töllke [32].	40

37. Resultados propios de los ataques con parámetros. En las etiquetas, cada vez que aparece el nombre de un parámetro se está indicando que ese parámetro se ha modificado: <code>N_max=6000, g_max_e=48, gCAN=25, gACh=3</code> . En rojo está el rango de oscilación de saludable. (a) muestra la frecuencia pico media, (b) la duración media.	40
38. PS de las diferentes bandas en varios experimentos de ataques de modificación de parámetros, cada uno con cuatro simulaciones. Resultados propios. En las etiquetas, cada vez que aparece el nombre de un parámetro se está indicando que ese parámetro se ha modificado. Los valores de los parámetros modificados son: <code>N_max=6000, g_max_e=48, gCAN=25, gACh=3</code> . En rojo se puede ver el rango en el cuál el estado saludable ha oscilado.	41
39. Resultados de los ataques basados en cambios de parámetros. En rojo se muestra el rango saludable.	43
40. Resultados de los ataques a archivos EEG, sin incluir Nonce. En rojo se muestra el rango saludable.	44
41. Resultados de los ataques Nonce. En rojo se muestra el rango saludable.	45
42. Resultados de los ataques Nonce Agresivo. En rojo se muestra el rango saludable.	46

Índice de tablas

1. Tabla comparativa entre la tesis de Töllke Töllke [32] y este trabajo. 20
2. Tabla comparativa con los resultados de las diferentes simulaciones de NeuroStrike. Cada fila representa un tipo de ataque, mientras que cada columna representa una característica y una banda de frecuencia. Se ha coloreado con rojo intenso los ataques que han provocado una excitación, en azul intenso los que han provocado una inhibición y en los mismos colores más claros los que han provocado los mismos efectos pero no de forma tan clara. 48

Declaración firmada sobre la originalidad del trabajo

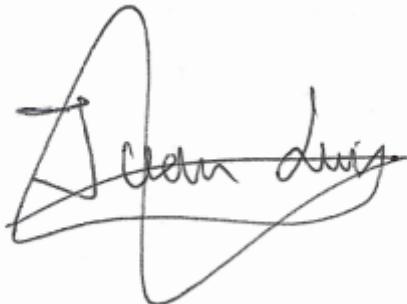
D./Dña. Juan Luis Serradilla Tormos, con DNI 58468843Z, estudiante de la titulación de Máster Universitario en Tecnologías de Análisis de Datos Masivos: Big Data de la Universidad de Murcia y autor del TFG titulado “Simulación de NeuroStrikes en el Hipocampo”.

De acuerdo con el Reglamento por el que se regulan los Trabajos Fin de Grado y de Fin de Máster en la Universidad de Murcia (aprobado C. de Gob. 30-04-2015, modificado 22-04-2016 y 28-09-2018), así como la normativa interna para la oferta, asignación, elaboración y defensa de los Trabajos Fin de Grado y Fin de Máster de las titulaciones impartidas en la Facultad de Informática de la Universidad de Murcia (aprobada en Junta de Facultad 27-11-2015)

DECLARO:

Que el Trabajo Fin de Grado presentado para su evaluación es original y de elaboración personal. Todas las fuentes utilizadas han sido debidamente citadas. Así mismo, declara que no incumple ningún contrato de confidencialidad, ni viola ningún derecho de propiedad intelectual e industrial

Murcia, a 28 de junio de 2025

A handwritten signature in black ink, appearing to read "Juan Luis Serradilla Tormos". The signature is fluid and cursive, with some loops and variations in line thickness.

Fdo.: Juan Luis Serradilla Tormos

Autor del TFG

Resumen

Este trabajo estudia la vulnerabilidad de los procesos neuronales, en particular los de dentro del hipocampo, frente a NeuroStrikes (ataques electromagnéticos de alta potencia). La guerra cognitiva y casos como el “Síndrome de La Habana” evidencian cómo los pulsos electromagnéticos pueden inducir alteraciones cognitivas significativas, lo que motiva una investigación pionera en simulaciones de estos ataques sobre modelos realistas.

Para abordar esta cuestión, primero se replican en Brian2 (un simulador de sistemas neuronales) los resultados de Töllke (2024). En su tesis, Töllke se centra en estudiar ataques de NeuroStrikes sobre el hipocampo, que es la zona encargada de gestionar la memoria en el cerebro. Para ello genera perturbaciones de parámetros neuronales y sinápticos (N_{max} , g_{max_e} , gCAN, gACh) del simulador. Tras replicar sus resultados, se visualiza una alta variabilidad en las simulaciones “saludables”, de forma que se establece el uso de promedios de cuatro ejecuciones por experimento para garantizar estabilidad estadística. A continuación, se propone un nuevo enfoque: en lugar de alterar los parámetros del simulador, se modifican directamente los archivos EEG de entrada que usa la simulación utilizada mediante modelos de ataque (Flooding, Jamming, Scanning, Selective Forwarding y Nonce, este último en dos variantes), inspirados en taxonomías de ciberataques neuronales, los cuales son capaces de afectar al comportamiento espontáneo de neuronas sanas.

La solución incluye una implementación de un módulo de ataque a archivos EEG en Brian2, capaz de escalar o inhibir segmentos de señal, crear pulsos discretos o introducir alteraciones aleatorias (Nonce); un rediseño de métricas de SWRs (Sharp Wave Ripples), midiendo cantidad, duración, frecuencia pico y densidad espectral de potencia (PSD) en bandas Theta, Gamma y Ripple, con superposición de rangos “saludables” para visualización comparativa y, finalmente, la definición de índices ad hoc (Índice de Ataque y Factor de Polarización) para cuantificar la eficacia y dirección de los ataques aleatorios Nonce.

Los resultados muestran que los ataques basados en archivos EEG (especialmente Scanning y Selective Forwarding) generan alteraciones más generalizadas en todo el espectro, con reducción de eventos Theta/Gamma y sobreestimulación de SWRs, lo que sugiere riesgos de deterioro en memoria de trabajo, espacial y potenciales crisis epilépticas. Los ataques a parámetros, en cambio, afectan principalmente a la potencia de los SWRs, reflejando vulnerabilidades en la consolidación de la memoria a largo plazo.

En suma, esta investigación avanza la literatura en el campo de la simulación de NeuroStrikes y la seguridad neuronal, aportando herramientas y métricas para evaluar y comparar diferentes tipos de ataque en modelos del hipocampo.

Extended Abstract

This work addresses the simulation of so-called NeuroStrikes (high-power, low-frequency electromagnetic radiation attacks) on realistic models of the human hippocampus, with the aim of evaluating their effects on memory consolidation processes and determining possible neuronal damage. Based on the growing concern surrounding Brain-Computer Interfaces (BCIs) and the use of high-powered electromagnetic pulses in the context of so-called cognitive warfare (which includes episodes such as the “Havana Syndrome”, affecting a large number of diplomats by causing anomalous health incidents), this research seeks to replicate and expand on the contributions of Lennart Töllke’s thesis (2024), providing new metrics, attack methods, and robust statistical analysis.

First, after a thorough review of the literature and fundamental works in neuroscience, BCIs, and neural cyberattacks, the Brian2 environment was selected as the simulation platform. This was done to take advantage of a detailed model of the hippocampus and entorhinal cortex developed and validated by Aussel et al. (2018, 2022). In addition, this model is the one used by Töllke in his work, so it was not necessary to implement a simulator from scratch, but rather Töllke’s code could be used to perform the experiments. This model, which incorporates more than thirty thousand neurons organized into CA3, CA1, and dentate gyrus regions, uses real EEG files recorded during sleep phases as input stimuli and allows for the synthetic generation of local field potentials (LFP) for analysis.

In the first phase of the work, Töllke’s experiments were reproduced. To do this, an exhaustive review of his thesis, the methods used and the metrics used, among others, was carried out. His repository was cloned and, after that, all his code was analyzed, so that it was possible to understand what functions he had implemented, why, and what they were used for, so that their operation could be expanded if necessary. These simulated attacks altered synaptic and cellular parameters, specifically `N_max` (maximum size of neural populations), `g_max_e` (maximum excitatory conductance), `gCAN` (acetylcholine-modulated calcium channel conductance), and `gACh` (cholinergic modulation factor), in order to emulate the biochemical and structural effects of electromagnetic pulses. Töllke’s idea was, therefore, to reproduce the neural conditions that would exist in the hippocampus after a high-powered electromagnetic attack (NeuroStrike), so that after performing a simulation with these parameters, we could visualize the repercussions on the brain. The replication of his results confirmed that the simulations have high intrinsic variability: measurements of peak frequency and duration of Sharp Wave Ripples (SWRs) varied significantly between independent runs. This made it difficult to compare results, since, having performed individual simulations, his results and those of this work could be clearly different due to a simple random factor. To mitigate this entropy, a protocol of four simulations per experiment was established, and formulas for combining means and standard errors were defined to ensure a reliable estimation of the metrics, allowing for a more reliable analysis without requiring prohibitive computational resources.

Once this first stage was completed, this work concluded that the results are similar to those in Töllke's thesis and, then, it was necessary to define new experiments to better understand the impact of Neurostrike attacks over the hippocampus. For that, this thesis provides new attacks and metrics in order to make a significant contribution to his work and to this emerging field of research. To this end, an innovative approach was introduced when simulating NeuroStrikes: instead of modifying the simulator parameters, the "healthy" EEG files were directly transformed to simulate signals "corrupted" by NeuroStrikes. The idea is that, instead of modifying the simulator parameters to try to recreate the neural conditions of the hippocampus after an electromagnetic attack, it might be possible to directly simulate the neural damage in the input EEG reading. In other words, the "healthy" input signal could be modified so that it was no longer healthy and became "corrupted," as if it had been attacked by a NeuroStrike, so that after performing a simulation with these "healthy" files and parameters in the simulator, we would obtain results of brain damage.

To develop these attacks, we drew inspiration from the taxonomy of neural cyberattacks proposed by López Bernal et al., implementing five signal alteration strategies: Flooding, Jamming, Scanning, Selective Forwarding, andNonce (the latter in two variants: standard and aggressive). The Flooding attack consisted of a complete scaling of the EEG signal during defined intervals (for example, scaling the signal value 10 times throughout its 60-second duration), while Jamming applied a reverse scaling (for example, decreasing 10 times, i.e., dividing the input signal power by 10 throughout its 60-second duration). Scanning and Selective Forwarding introduced discrete pulses of signal increase or decrease in minimal time windows, alternating attacks and rests to mimic sequential neuronal stimulation or inhibition (as in the previous attacks, multiplication and division by 10 was used to increase or decrease by an order of magnitude). Finally, Nonce caused random modifications at intervals, combining the probability of no effect with different degrees of excitation or inhibition, which generated a wide range of signal patterns. All these attacks have their own parameters, so they can be applied in very different ways. Although this work does not explore all these combinations due to lack of time and computing power, it is an interesting topic to see how different attacks behave with different combinations of parameters, even trying to find a transition from a discrete attack to a continuous attack in Scanning, Selective Forwarding, and Nonce attacks.

Once the attacks have been designed, the metrics and graphs that allow the impact of the NeuroStrikes simulations to be measured need to be created. In his study, Töllke focused primarily on the Sharp Wave Ripples band, as it is the most direct evidence of long-term memory consolidation within the hippocampus. However, other types of frequency bands (which Töllke only explored when calculating the Power Spectral Density or PSD), such as the Theta and Gamma bands, have also been considered important in detecting other types of cognitive damage (such as working memory or spatial memory). For all

these reasons, the extraction of metrics focused on the three key frequency bands for memory consolidation: Theta (4–8 Hz), Gamma (30–120 Hz), and SWRs (100–250 Hz). Rather than simply counting the occurrence of SWRs, the total number of events, average duration, average peak frequency, and spectral density power (PSD) were measured for each band. All these metrics were analyzed using bar charts and line graphs, which showed the mean values and standard errors, overlaid with the “healthy” range in which the values oscillated. This overlap of the healthy range was obtained from a set of ten reference experiments (each with four simulations), so that when combined in the different graphs with the results, the pathological deviations, either above or below the healthy range, induced by each type of attack could be identified visually and quantitatively.

Finally, after simulating all the attacks and generating the graphs, the results were analyzed, comparing the increases and decreases in the waves with the neurobiological literature on brain rhythms. These results reveal that parameter-based attacks mainly generate alterations in the power and presence of SWRs: they generally reduce their PSD and modify their frequency and duration in a variable manner depending on the parameter affected, evidencing vulnerabilities in long-term memory consolidation processes. This fits perfectly with Töllke’s results and explains why his metrics and graphs focused on the evaluation of Sharp Wave Ripples: the search for cognitive damage in long-term memory. In particular, cholinergic disabling (the gACh parameter) produced dramatic decreases in the number and duration of SWRs, in line with the fundamental role of these events in the “replay” of neural sequences and the transfer of memory traces to the cortex. Although this was not due to an actual decrease in their characteristics, i.e., it was not that the frequency or duration of Sharp Wave Ripples actually decreased, but rather that this type of attack completely eliminated Sharp Wave Ripples in the four simulations that were performed, which meant that no events were detected and, therefore, their characteristics were not measured. However, all these parameter attacks were less effective in disrupting the Theta and Gamma bands, whose temporal coupling is essential for the encoding of spatial and working information.

In contrast, attacks on EEG files showed a more dysfunctional and broad spectrum profile. Flooding and Jamming barely altered the metrics, as uniformly scaling the signal failed to disrupt the oscillatory patterns. In contrast, Scanning and Selective Forwarding caused drastic reductions in Theta and Gamma events (as in the case of gACh, this was due to the absence of events in these bands, which meant that none were detected), accompanied by overstimulation of SWRs. According to the literature, these disturbances would translate into deficits in spatial and working memory (due to theta-gamma decoupling), olfactory and learning impairment (linked to low gamma power), and risk of epileptic seizures due to pathological hypersynchronization of SWRs (“p-ripples”). Nonce attacks, despite their intrinsic randomness, exhibited consistent trends: increased gamma activity and modified theta stability, with slight changes in SWRs which, in their aggres-

sive version, favored episodes of low quantity and reduced PSD, suggesting an impact on synaptic plasticity and neuronal excitability. The fact that Flooding and Jamming attacks had no effect is probably due to the low variability of the modified signal compared to the original one. In other words, by scaling the signal proportionally over the 60 seconds of reading, the trends and behavior of the voltage did not change, so the brain was able to “adapt” to these new values without suffering any consequences. However, in the rest of the attacks, due to numerous changes in scale, the simulation was unable to adapt to these rapid changes, causing notable discrepancies in the results.

Nonce attacks, due to their randomness, posed a great challenge when measuring their results, as depending on how the input EEG reading was modified, they could give highly different results. To globally quantify the effectiveness and direction of these attacks, they defined two ad hoc indices: the Attack Index (I_A), which weights the proportion of metrics outside the healthy range by adjusting for the magnitude of the error, and the Polarization Factor (F_P), which distinguishes between excitatory or inhibitory effects normalized by the number of atypical events. These indicators allowed Nonce attacks to be compared with the rest. This showed that Scanning and Selective Forwarding are the most aggressive methods, followed by parametric attacks associated with conductances and neuronal population; Nonce attacks, on the other hand, also showed promising results with appreciable cognitive damage.

The projected cognitive impact of these findings is concerning. The decrease in theta waves points to a loss of temporal synchrony and association formation, with consequences analogous to those observed in dementias such as Alzheimer’s and mild cognitive impairment. The alteration of Gamma bands is related to deficiencies in associative learning and spatial memory, while the disregulation of SWRs compromises the nocturnal “replay” essential for fixing long-term memories and can trigger pathological epileptic activity. In short, the different NeuroStrikes modalities faithfully simulate possible Cognitive Warfare scenarios in which electromagnetic pulses or malicious BCI manipulations induce synaptic, structural, and oscillatory coordination damage.

The contributions of this work are manifold: first, the validation and improvement of the statistical protocol in Brian2 simulations to ensure robust results in the face of high intrinsic variability; second, the introduction of a novel attack method involving the modification of EEG files, which reveals risks of integral spectral disturbance that are not evident when changing simulation parameters; third, the expansion of evaluation metrics to include not only SWRs but also Theta and Gamma bands with adapted detection criteria; and fourth, the development of quantitative indices (I_A and F_P) that facilitate global comparison and characterization of excitatory or inhibitory effects.

However, this research opens up future avenues for improvement. It would be desirable to implement advanced algorithms for detecting Theta and Gamma events that go beyond

the V_{rms} -based threshold approach, as well as to explore functional connectivity analysis methods to capture more subtle coupling mismatches. Likewise, it would be valuable to study active countermeasures within the simulator that would mitigate the effects of NeuroStrikes. Finally, given the computational load, optimization and scaling to even more extensive and heterogeneous topologies will contribute to a more comprehensive understanding of Cognitive Warfare and its possible mitigation.

1. Introducción

La guerra cognitiva, como su nombre indica, consiste en la guerra de modificar la conciencia humana. Esto se puede hacer mediante herramientas políticas como la propaganda y la desinformación, modificando así el comportamiento de la población, o también se pueden utilizar tecnologías avanzadas que puedan perturbar el sistema nervioso humano. Este segundo tipo de guerra cognitiva puede parecer bastante lejano en el tiempo, pero China ya tiene en desarrollo proyectos de armas que puedan modificar la cognición humana; además, han habido ataques a lo largo de la historia que pueden achacarse a armas electromagnéticas de gran potencia. Un ejemplo de estos casos de ataque sería el Síndrome de La Habana, cuando varios ministros importantes de Estados Unidos experimentaron síntomas similares, los cuales no se podían achacar a ninguna enfermedad conocida. A este tipo de fenómenos de ataques electromagnéticos de alta potencia se los conoce como NeuroStrikes, y van a ser el objeto de estudio en este trabajo.

Si se revisa la literatura actual, se puede ver rápidamente que no hay investigaciones sobre las simulaciones de las consecuencias de los NeuroStrikes en el cerebro, simplemente se investiga cómo de factibles son y si se han usado. Lo más parecido que se puede encontrar son los artículos de López Bernal et al. [14, 15] y López Madejska et al. [26], que hablan de la seguridad neuronal de las interfaces cerebro-computadora (BCIs). En estas investigaciones analizan las repercusiones que pueden tener los ciberataques a las BCI con la capacidad de estimular o inhibir individual o grupalmente las neuronas de nuestro cerebro. A pesar de esto sí que encontramos un trabajo en la literatura sobre la simulación y repercusiones de los NeuroStrikes, la tesis de Töllke [32]. Lennart Töllke, en su trabajo, investiga la simulación de los NeuroStrikes en el cerebro usando el simulador Brian2 (un simulador de neuronas que funciona en Python) y una topología neuronal del hipocampo desarrollada por Aussel et al. [10, 11]. Lo que hizo Töllke fue modificar los parámetros del simulador para imitar los daños de un NeuroStrike en el hipocampo (disminuía el número de neuronas, disminuía la eficacia de la sinápsis...). Su objetivo era medir los daños en la memoria a largo plazo, por lo que para ello midió las características de los Sharp Wave Ripple (SWRs) del hipocampo, que son la clase de ondas del cerebro más relacionadas con la consolidación de la memoria a largo plazo. Por ello, midió solamente este tipo de ondas y concluyó que estas simulaciones de ataque “en base a parámetros” simulaba correctamente una disminución de los SWRs y, por tanto, un deterioro de la memoria.

Son, por tanto, las limitaciones de la tesis de Töllke lo que motiva el desarrollo de este trabajo. Lennart solo pudo medir las consecuencias en la memoria a largo plazo con los SWRs y realizando una sola simulación por experimento; sin embargo, en este trabajo se quiere ampliar su visión investigando más bandas de frecuencia (lo cuál nos permitirá ver más daños cognitivos en el cerebro), se realizarán más simulaciones para analizar el comportamiento del simulador, se idearán nuevas simulaciones de ataques y se desarrollarán

nuevas métricas y gráficas que nos permitan analizar estos nuevos daños.

Por lo tanto, para realizar este estudio, lo primero que se ha hecho es una revisión exhaustiva del estado del arte de la neurociencia, la guerra cognitiva y los ciberataques a los BCIs (ya que son las fuentes de información más cercanas a lo que se quiere hacer). También se ha investigado a fondo la tesis de Töllke, para tener bien claro hasta donde tenía que llegar este trabajo y qué más se podría aportar. Una vez se ha entendido la situación de la investigación, se empezaron a replicar los resultados de Töllke: clonar repositorios, instalar de herramientas, desglosar funcionamientos de scripts y ver los parámetros de la simulación. Tras una ardua revisión, se pudieron replicar los resultados de su investigación, aunque cabe destacar que durante esta inspección se detectó una alta variabilidad en la simulación de Brian2, de forma que resultados con los mismos parámetros tenían valores significativamente diferentes. A esto se le puso solución con una mayor cantidad de simulaciones por experimento, aunque un número limitado debido a la falta de potencia computacional, siendo a partir de aquí donde se empezó a ampliar el trabajo de Töllke.

Las contribuciones principales han sido las siguientes:

- Una mejora en la estabilidad de los experimentos, añadiendo mayor cantidad de simulaciones para mitigar la entropía del propio simulador.
- Un rediseño de las métricas de Lennart, además de nuevas gráficas que permiten visualizar nuevos los daños de los ataques simulados correctamente. Para ello, se han añadido los datos de otras bandas de frecuencia a parte de la Ripple (Theta y Gamma) a las métricas de cantidad de ondas, duración y frecuencia.
- Diseños de nuevos ataques simulados. La única forma que tenía Töllke de simular NeuroStrikes era modificando los parámetros del simulador; sin embargo, se ha propuesto que modificando los archivos EEG de entrada que usa el simulador se pueden replicar otros tipos de ataques. Es decir, el simulador usa lecturas EEG reales de entrada para funcionar, así que se ha propuesto que si se modifican estas entradas para imitar una lectura EEG tras un NeuroStrike y se introduce al simulador sin modificar los parámetros (parámetros saludables), la señal resultante debe presentar daños apreciables. Estos ataques se basan en el artículo de taxonomía de ciberataques neuronales de López Bernal et al. [16], siendo: Flooding, Jamming, Scanning, Selective Forwarding yNonce.
- Se han ideado índices ad hoc para poder analizar las repercusiones de los ataques Nonce. Dada a su naturaleza aleatoria, se han necesitado crear estos índices para medir cuánta repercusión tienen los ataques y de qué tipo.

Finalmente, se ha observado como se han podido replicar los resultados de Töllke, aunque con ciertas diferencias debido a las varaciones de la propia simulación. Los ataques en base a parámetros de Lennart, como dijo en su tesis, suelen causar una disminución de

SWRs, lo que crea un deterioro en la memoria a largo plazo. Sin embargo, los nuevos ataques diversifican más sus efectos, causando más daños en las bandas Theta y Gamma y, por lo general, ampliando los SWRs en vez de reduciéndolos. Esto se traduce en daños en otros tipos de memoria (como de trabajo y espacial) y en posibles pequeñas crisis epilépticas.

La estructura de este trabajo es la siguiente:

- En la Sección 2.5.1 revisaremos el conocimiento previo que se necesita saber para esta investigación, desde el funcionamiento de la memoria y el hipocampo hasta la definición de ciberataques neuronales en las BCIs o la guerra cognitiva.
- En la sección 3.4 se documenta el estado del arte de la ciberseguridad neuronal. Se analizarán los papers de ciberataques neuronales a las BCIs y la tesis de Töllke centrada en neurostrike sobre el hipocampo.
- En la Sección 4.2 se profundizará en los objetivos del trabajo, además de discutir la metodología aplicada.
- La Sección 5.2.3 explica el diseño de la solución y cómo se ha implementado para poder obtener los resultados.
- La Sección 6.4 muestra los resultados de los experimentos, además de analizarlos.
- Finalmente, en la Sección 7 se comentarán las conclusiones del trabajo. También se indicarán futuras vías de investigación dadas las limitaciones del estudio.

2. Background

2.1. El cerebro y el sistema nervioso

El cuerpo humano se compone de numerosos sistemas que regulan su funcionamiento. Para poder comunicarse con el entorno interviene el sistema nervioso, una red de señales eléctricas que comunica a todo el organismo. Dentro de esta gran red se encuentra el sistema límbico, donde a su vez se encuentra el hipocampo. En esta sección vamos a profundizar en cómo funciona el sistema nervioso.

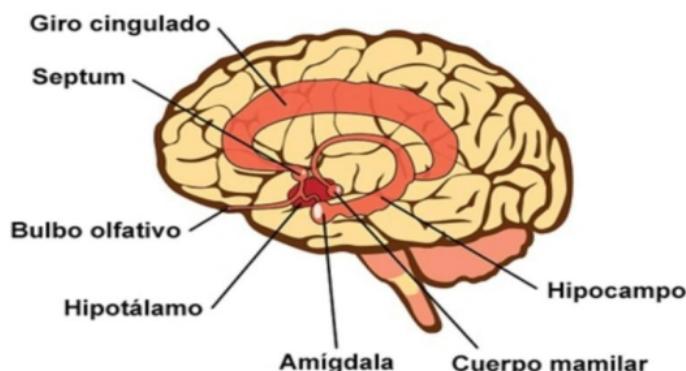


Figura 1: Esquema del sistema límbico. [8]

El sistema nervioso está formado por dos tipos de células [31]:

- **Neuronas:** Son el componente central del sistema nervioso en general y del cerebro en particular. Las neuronas son células eléctricamente excitables que procesan y transmiten información mediante señales electroquímicas, gracias a un proceso llamado sinapsis.
- **Neuroglías:** Estas células sirven para apoyar y proteger a las neuronas. Aunque solía pensarse que su función se limitaba al soporte físico, nutrición y reparación de las neuronas del sistema nervioso central, investigaciones más recientes sugieren que las glías, especialmente los astrocitos, desempeñan un papel mucho más activo en la comunicación cerebral y la neuroplasticidad.

Podemos ver las diferentes partes de la neurona en la Figura 2.

Las partes más importantes de la neurona son:

- **Soma (cuerpo celular):** Es la parte bulbosa de la célula que contiene el núcleo. Aquí es donde ocurre la mayor parte de la síntesis de proteínas. El soma es una de las principales ubicaciones que recibe señales.

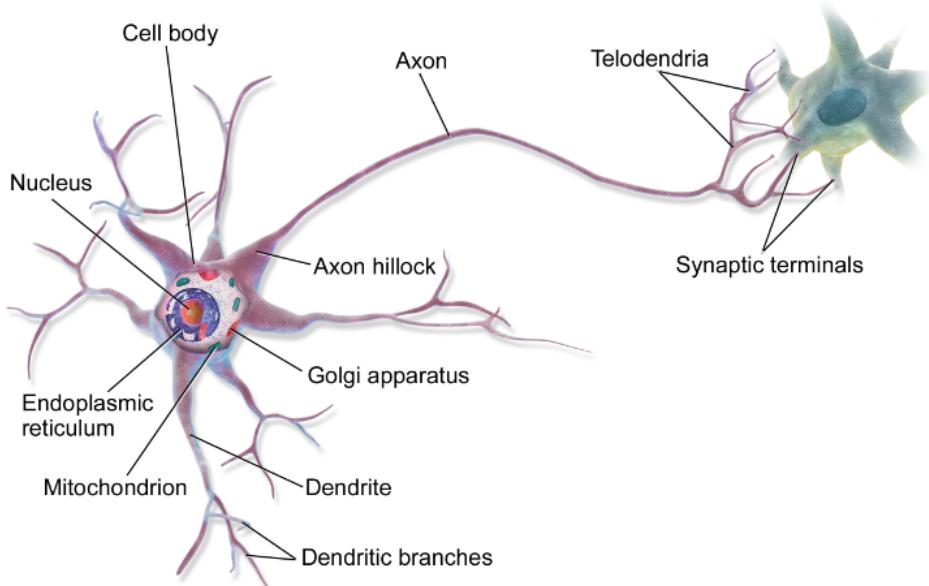


Figura 2: Ilustración de una neurona cerebral. [31]

- **Dendritas:** Son filamentos largos y ramificados que se extienden desde el cuerpo celular. Las dendritas son las estructuras por donde la neurona recibe la mayoría de las entradas o señales de otras neuronas, típicamente a través de las espinas dendríticas.
- **Axón:** Es una proyección fina, similar a un cable, que transporta las señales nerviosas fuera del soma y puede extenderse miles de veces el diámetro del soma en longitud.
- **Vainas de mielina:** Son sustancias ricas en lípidos que aislan los axones de las células nerviosas. Su función principal es aumentar la velocidad a la que la información viaja de una célula nerviosa a otra.

Para que las neuronas se puedan comunicar entre sí existe la sinapsis neuronal. Por lo tanto, se puede definir a la sinapsis como la estructura especializada en permitir la transmisión de información de una neurona a otra. Puede ser química o eléctrica:

- **Sinapsis eléctrica:** En una sinapsis eléctrica, las membranas de la célula presináptica y postsináptica están conectadas directamente por canales especiales llamados uniones gap o hendidura sináptica. Estos canales son capaces de pasar una corriente eléctrica, haciendo que los cambios de voltaje en la célula presináptica induzcan cambios de voltaje en la célula postsináptica. La principal ventaja de una sinapsis eléctrica es la rápida transferencia de señales de una célula a la siguiente.

- **Sinapsis química:** Es el tipo mayoritario de sinapsis. El proceso comienza cuando un impulso eléctrico (potencial de acción) llega a la terminal del axón de la neurona presináptica. Este impulso provoca la liberación de sustancias químicas llamadas neurotransmisores, que están almacenadas en vesículas sinápticas. Los neurotransmisores se difunden a través del espacio sináptico y se unen a receptores específicos en la membrana de la neurona postsináptica. Esta unión provoca cambios en la permeabilidad de la membrana de la célula receptora a iones específicos, lo que afecta su potencial de carga. Si las influencias excitatorias superan a las inhibitorias y el voltaje de la membrana postsináptica alcanza un potencial umbral (típicamente entre -40mV y -55mV), se genera un nuevo potencial de acción en esa neurona, propagando así la señal.

Podemos ver el proceso de la sinapsis química esquematizado en la Figura 3.

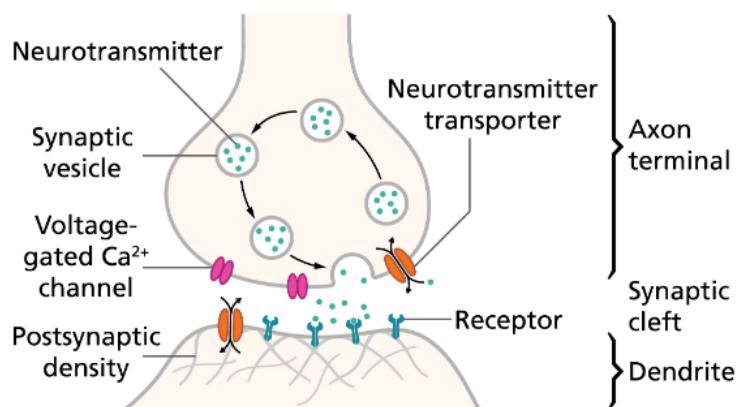


Figura 3: Esquema de la sinapsis química. [31]

2.2. Hipocampo

En esta sección se definirá el hipocampo, se describirá su anatomía y se explicarán sus funciones.

2.2.1. Definición del hipocampo

El hipocampo es una estructura pequeña, ubicada en el centro del cerebro, dentro del lóbulo temporal medial. Tiene una forma similar a un caballito de mar, la cual le da su nombre. Además, está relacionado con la formación de la memoria, la adquisición de nueva información y consolidación de relaciones espaciales, entre otras funciones, por lo que es imprescindible para el aprendizaje [8].



Figura 4: Fotografía del hipocampo. [8]

2.2.2. Anatomía del hipocampo

El hipocampo es una estructura subcortical que consta de tres regiones: los cuernos de Amón (CA1 y CA3), el hilus y el giro dentado (GD), conocido también como área dentada o fascia dentada. Podemos ver las diferentes zonas del hipocampo en la Figura 5.

En Bello-Medina et al. [13] se describe la anatomía de las diferentes regiones del hipocampo:

- **GD:** El GD está compuesto a su vez por tres regiones: la banda superpiramidal (limitando con el CA1), la banda infrapiramidal (limitando con el CA3) y el ginu (apéndice del GD), donde se unen las dos bandas anteriores. A su vez, las bandas del GD se pueden dividir en tres grandes capas: la capa molecular (contiene dendritas de neuronas glandulares), la capa de células granulares (conformada por los somas de las neuronas granulares) y la capa polimórfica (ubicada entre las dos bandas del GD). El GD es una de las dos únicas zonas del cerebro donde se generan nuevas neuronas (neurogénesis neuronal), que luego se integran con el resto de redes neuronales y estimulan el proceso de aprendizaje.
- **CA3:** Es la porción del cuerno de Amón más próxima al GD. A lo largo del eje anatómico del CA3 al longitud total de las dendritas y la organización de las células piramidales varía significativamente.
- **CA1:** Es la región del cuerno del Amón más distal al GD, siendo a su vez próxima al CA3. Las células piramidales del CA1 tienden a ser más pequeñas y homogéneas que las del CA3. Todas estas conexiones del CA1 con el CA3 sirven para relacionar diferentes informaciones sensoriales y para la codificación de nueva información en los procesos de aprendizaje y memoria.

También se describen los diferentes tipos de células que podemos encontrar:

- **Células piramidales:** Se encuentran en los cuernos de Amón (CA1 y CA3).

Hippocampus

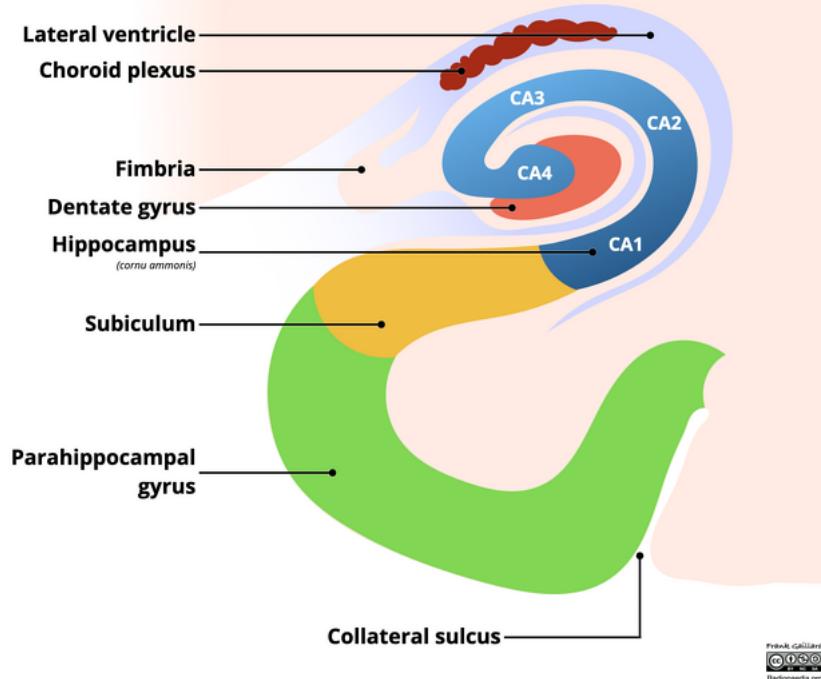


Figura 5: Esquema de las zonas del hipocampo [4].

- **Células granulares:** Se encuentran distribuidas homogéneamente en la banda superior e inferior del GD.
- **Células musgosas, intrínsecas y canasta:** Todas estas clases de células se encuentran distribuidas por el área del hilus.
- **Interneuronas:** Se encuentran ampliamente distribuidas por todas las regiones del hipocampo, representando entre un 10 y 15 % del total.

Bello-Medina et al. [13] también describen la anatomía de algunas de las diferentes regiones del hipocampo:

2.2.3. Funciones del hipocampo

El hipocampo posee numerosas funciones en los procesos cognitivos del cerebro, destacando sobre todo la memoria, el aprendizaje y la navegación. Vamos a realizar una descripción de las funciones cognitivas del hipocampo en base a los artículos de Olivares et al. [22],

Antepara et al. [8] y Amaral et al. [7].

Una de las funciones principales del hipocampo es la consolidación de memoria a corto plazo y de trabajo para convertirlas en memoria a largo plazo. Por lo tanto, las lesiones hippocampales pueden producir amnesia retrógrada de hechos y eventos. Sin embargo, la memoria procedimental (no declarativa), al no regularse por el hipocampo, permanece intacta, de forma que a pesar de no recordar hechos o eventos sí que se pueden aprender tareas (aunque no se recuerde que se han aprendido). Además, dentro del GD se generan nuevas neuronas, lo que contribuye a la plasticidad sináptica y el aprendizaje.

A parte de consolidar la memoria a largo plazo, el hipocampo tiene un papel importante en la navegación. Se han descubierto unos tipos de células especiales, células de lugar y de tiempo, localizadas en el CA1 y el CA3. La función de estas es crear un mapa cognitivo, estableciendo recuerdos antes y después en el tiempo o en localizaciones del espacio. Por ello, daños en el hipocampo pueden afectar a la navegación espacial y temporal.

2.3. Memoria

Según Carillo-Mora et al. [18], la memoria es el proceso cognitivo que permite codificar, almacenar y recuperar información, tanto de forma consciente (declarativa) como de forma inconsciente (no declarativa). Su función es fundamental para el aprendizaje, permitiéndonos experimentar el pasado y planificar el futuro.

La memoria a largo plazo es la memoria que nos permite visitar los recuerdos más lejanos. Esta puede dividirse en dos:

- **Memoria declarativa:** La memoria declarativa es el tipo de memoria que almacena hechos y eventos, permitiéndonos evocar recuerdos conscientes.
- **Memoria no declarativa:** La memoria no declarativa es la que engloba habilidades, hábitos y condicionamientos que se expresan de forma automática e inconsciente.

La memoria declarativa, a su vez, puede dividirse en memoria semántica y memoria episódica. La memoria semántica almacena hechos, conceptos y vocabulario, mientras que la memoria episódica recuerda experiencias personales con un contexto espacial y temporal.

Las estructuras cerebrales implicadas en la consolidación de la memoria declarativa, como se puede leer en Kenneth et al. [23], son el hipocampo y las áreas límbicas adyacentes (corteza entorrinal y parahipocampo). La función del hipocampo es codificar la información contextual y las relaciones espaciales, actuando así como un “punto de conmutación” para después transferir dichos recuerdos a la corteza cerebral para un almacenamiento a largo plazo. Debido a este papel del hipocampo en la memoria, estudios de lesiones hippocampales han demostrado que este tipo de daño produce amnesia anterógrada e incapacidad

para formar nuevos recuerdos declarativos [30].

Dentro del hipocampo hay varios procesos a través de los cuales se forma la memoria declarativa. En el artículo de Ortega et al. [25] podemos ver algunos de estos mecanismos:

- **Potenciación a largo plazo (LTP):** La LTP consiste en el fortalecimiento duradero de la sinapsis tras una estimulación repetitiva. Este proceso se regula a través de los receptores NMDA y AMPA (dos tipos principales de receptores de glutamato del sistema nervioso central).
- **Cambios estructurales:** El cambio y crecimiento de estructuras como las espinas dendríticas y la reorganización del citoesqueleto neuronal ayudan a la consolidación de la memoria.
- **Señalización intracelular:** Las señales de Ca^{2+} (a través de la proteína calmodulina) y las enzimas CAMKII, PKA y MAPK disparan la fabricación de las proteínas que hacen falta para que la consolidación tenga lugar.

Todos estos eventos moleculares tienen lugar primero en el hipocampo y, después, en la corteza parahipocampal y otras áreas corticales. De esta forma, se estabiliza el recuerdo en el tiempo.

2.3.1. Ritmos cerebrales en la consolidación de la memoria

En estos mecanismos de formación de la memoria, aparte de los eventos moleculares, tienen presencia los ritmos cerebrales. Estos consisten en diferentes tipos de ondas de potencial eléctrico, clasificadas según su frecuencia, que modifican la plasticidad de la sinapsis para consolidar la memoria. Dependiendo de su velocidad de oscilación las clasificamos, entre otras muchas, como ondas Theta, ondas Gamma y Sharp Wave Ripples (SWRs); cada una de ellas con un propósito diferente.

En la nota de prensa de Quijada [29] se puede ver que las ondas Theta se componen de frecuencias entre 4 y 8 Hz. Su función es sincronizar la actividad neuronal del hipocampo, actuando así como regulador temporal de ráfagas más rápidas, como las ondas gamma, que están en el rango de 30 a 120 Hz. Dentro de este ciclo theta, los paquetes gamma (eventos gamma breves) transportan fragmentos de información sensorial y espacial; esta división en “ventanas de codificación” permite que las neuronas de CA3 y CA1 refuerzen asociaciones sinápticas específicas, lo que es clave para la formación de memorias declarativas.

Por otro lado, los SWRs son fenómenos ondulatorios de alta frecuencia, portando frecuencias de 100 a 250 Hz. Se generan en la red CA3 y atraviesan CA1 como breves ráfagas ultrarrápidas [6]. Durante esos SWRs, las secuencias de espigas neuronales correspondientes a experiencias previas se reactivan, esto genera una “reproducción” que se sincroniza

con las oscilaciones corticales para transferir gradualmente los trazos de memoria desde el hipocampo a la corteza neocortical, afianzando así la memoria a largo plazo.

Conociendo todos los beneficios y funciones de las ondas Theta, Gamma y SWRs, cabe esperar que una disminución de estas ondas conllevará directamente a un déficit en la memoria. Por el lado contrario, un aumento de estos ritmos puede ser beneficioso para la memoria, mejorando su consolidación; aunque, como todo, un exceso puede dar lugar a síntomas patológicos.

- **Ondas Theta:** En Nuñez et al. [28], se discute que la disminución de la actividad theta podría perjudicar la capacidad del hipocampo para organizar y codificar secuencias de información y formar asociaciones. La reducción del acoplamiento theta-gamma, que es crucial para la transferencia de información espacial y las operaciones mnemónicas, se ha observado en modelos de la enfermedad de Alzheimer y en pacientes con deterioro cognitivo leve, lo que se relaciona con déficits en la memoria de trabajo y la navegación espacial. Por el lado contrario, se comenta que una actividad theta robusta en el hipocampo es fundamental para la formación de nuevas memorias, codificación de información espacial y no espacial y la recuperación de recuerdos. Por tanto, el aumento de la potencia theta puede sintonizar las propiedades espaciales y sincronizar las neuronas del hipocampo. Sin embargo, en Guan et al. [21], en el contexto de trastornos mentales y sentimientos negativos, se ha observado un aumento en la potencia theta y el acoplamiento theta-gamma (cuando las ondas theta modulan en fase o amplitud a las ondas gamma) en el hipocampo ventral de ratones, lo cual se correlaciona con ansiedad e hiperactividad. Aunque estas sean consecuencias conductuales, pueden llegar a afectar de forma indirecta a procesos superiores como la memoria.
- **Ondas Gamma:** En Guan et al. [21], pacientes con Alzheimer muestran disminuciones de la potencia gamma en el hipocampo, el giro dentado y el bulbo olfatorio, lo cuál se vincula a una alteración de la memoria espacial, un deterioro olfativo y trastornos generales de aprendizaje y memoria. Por el contrario, los acoplamientos de ritmos gamma a fases theta del hipocampo son cruciales para un rendimiento exitoso de la memoria, por lo que su aumento se vincula a mejor rendimiento en tareas de memoria mnemónicas y de aprendizaje asociativo. Sin embargo, un aumento aberrante de las oscilaciones gamma se ha relacionado también con defectos de aprendizaje y enfermedades del sistema nervioso central como el Alzheimer y el Parkinson.
- **SWRs:** Ego-Stengel et al. [19] destacan que interrumpir la actividad neuronal durante los eventos ripple afecta al aprendizaje espacial en ratas, además de conllevar una disminución del aprendizaje. En Liu et al. [24], por otro lado, se ha estudiado cómo un aumento en la duración de los SWRs del hipocampo mejora la memoria.

Sin embargo, debido a la naturaleza estrictamente síncrona de los SWRs, incluso pequeñas perturbaciones en los circuitos del hipocampo pueden transformarlos en oscilaciones de alta frecuencia con picos de población (picos de voltaje debidos a excitaciones simultáneas de poblaciones neuronales) más fuertemente sincronizados, conocidas como “ondas p-ripples” o rizos patológicos. Estas p-ripples pueden ocurrir de forma aislada o asociadas a descargas epilépticas interictales (IEDs), produciendo también deterioro en el aprendizaje espacial.

2.4. BCIs

Las Interfaces Cerebro-Computadora (BCI, por sus siglas en inglés) son sistemas que establecen un canal de comunicación bidireccional entre el cerebro y dispositivos externos. Tienen dos funciones principales: adquirir y procesar la actividad cerebral de los usuarios para realizar acciones específicas en máquinas o dispositivos externos y realizar estimulación neuronal para corregir problemas neuronales, replicar secuencias, etc.

En Bernal et al. [17] se engloban las diferentes definiciones de BCIs y se trata de estandarizar los pasos en el ciclo de grabación y estimulación de las BCIs.

Podemos observar cómo el ciclo BCI posee cinco fases:

1. **Generación de señales cerebrales:** Los procesos cerebrales del usuario producen actividad neural, que puede ser influenciada por estímulos externos.
2. **Adquisición y estimulación:** Las ondas cerebrales son capturadas por electrodos utilizando diversas tecnologías, como la Electroencefalografía (EEG) o la Resonancia Magnética Funcional (fMRI). Las señales analógicas brutas se transmiten para su procesamiento. También se puede producir estimulación neuronal, en la cuál las tecnologías provocarían pequeñas descargas eléctricas para estimular las células, ya sea individual o globalmente.
3. **Procesamiento y conversión de datos:** En esta fase, las señales analógicas se convierten a formato digital. El objetivo principal es maximizar la relación señal-ruido (SNR) para obtener la señal original con la mayor precisión posible. Para la estimulación, se transforman los patrones de disparo generales en parámetros específicos de la tecnología BCI. También se pueden convertir de digital a analógica, para así pasar de patrones de disparo digitales a estímulos analógicos que podrán ejecutarse en la siguiente fase (estimulación).
4. **Decodificación y codificación:** En esta fase se puede tanto decodificar las acciones a realizar que hay ocultas en los patrones de disparo de las neuronas, como codificar las acciones artificiales que se quieren imponer con el dispositivo BCI.
5. **Aplicaciones:** Las aplicaciones pueden ejecutar la acción prevista en el mundo

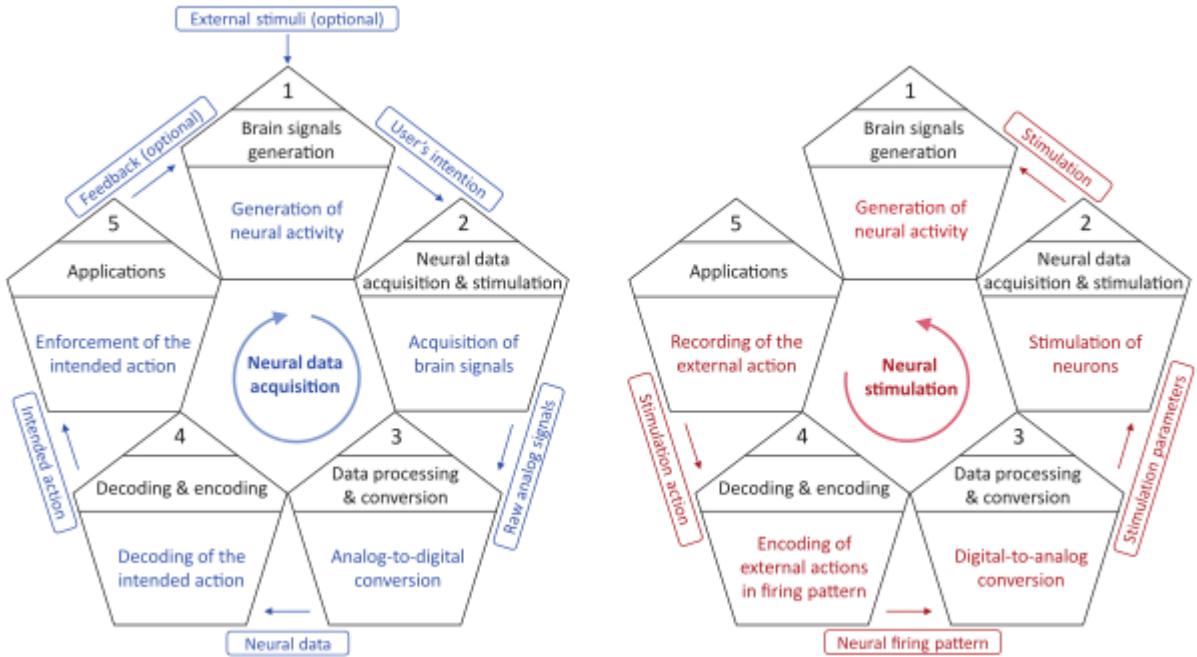


Figura 6: Ciclo de funcionamiento bidireccional de las BCIs que representa, en negro, las fases comunes para la adquisición de datos neuronales y la estimulación cerebral. (Lado izquierdo) Representación, en azul, de los procesos realizados y de los datos transferidos en cada fase del proceso de adquisición de datos neuronales. Este ciclo puede considerarse un proceso cíclico, pues comienza y termina en la misma fase. (Lado derecho) Representación, en rojo, de los procesos y transiciones de cada fase que conforman el proceso de estimulación. [17]

físico, por ejemplo, controlando un dispositivo externo. En el caso de la estimulación, las aplicaciones definen las acciones de estimulación deseadas que, posteriormente, se ejecutarán en el sistema nervioso.

Las fases son las mismas para los ciclos de adquisición y estimulación, siendo el sentido en el que se recorren estas fases lo que determina la naturaleza del ciclo BCI.

Dependiendo de cómo se incorpore la BCI al sistema nervioso para realizar los ciclos de adquisición y estimulación podemos clasificarlas en invasivas o no invasivas.

- **Invasivas:** Las BCIs invasivas requieren intervención quirúrgica para ser implantadas directamente en el cerebro. Son ampliamente utilizadas en terapia médica y han sido fundamentales para mejorar la calidad de vida de pacientes. Por ejemplo, se usan para el control de prótesis robóticas en personas con discapacidades o para la neuromodulación en el tratamiento de enfermedades neurodegenerativas como el

Parkinson. Un ejemplo de BCI invasiva es NeuraLink, el proyecto de Elon Musk para poder estimular individualmente las neuronas y tener un control muy amplio sobre el cerebro.

- **No invasivas:** Las BCIs no invasivas no requieren cirugía y capturan la actividad cerebral desde fuera del cuerpo, como la superficie del cuero cabelludo. Han ganado popularidad en los últimos años, expandiéndose desde escenarios médicos tradicionales a nuevos dominios como el entretenimiento y los videojuegos. Un ejemplo son las tecnologías de señales electroencefalográficas (EEG), las cuales usan electrodos colocados en la cabeza para grabar la actividad neuronal.

Las BCIs, al ser dispositivos electrónicos, pueden sufrir ataques malintencionados. Por lo tanto, en términos de seguridad las BCIs deben cumplir las siguientes dimensiones, ampliamente usadas en el ámbito de la ciberseguridad [17]:

- **Integridad:** Es la protección contra la modificación o destrucción no autorizadas de la información. En el contexto de las BCIs, esto significa asegurar que las señales cerebrales y los patrones de estimulación no sean alterados maliciosamente. Por ejemplo, los ciberataques basados en ruido pueden afectar la integridad de la señal EEG.
- **Confidencialidad:** Es la preservación de las restricciones autorizadas sobre el acceso y la divulgación, incluyendo los medios para proteger la privacidad personal y la información propietaria. En las BCIs, esto implica proteger información personal muy sensible, como pensamientos, emociones, orientación sexual o creencias religiosas, que podrían estar bajo amenaza si no se adoptan medidas de seguridad.
- **Disponibilidad:** Es la propiedad de que los datos o la información sean accesibles y utilizables a demanda por una persona autorizada. Un ataque a la disponibilidad podría impedir el funcionamiento correcto de un dispositivo BCI, por ejemplo, impidiendo que los electrodos capturen señales cerebrales o interrumpiendo un tratamiento médico vital.
- **Seguridad física:** Se refiere a todo daño causado sobre la salud o integridad personal de los usuarios de las BCIs. En concreto, en el ámbito de BCI se considera desde las perspectivas fisiológica, psiquiátrica y psicológica. Los ataques pueden tener un impacto significativo en la salud física y psicológica del usuario, desde la inducción de pensamientos y comportamientos no deseados hasta daños tisulares irreversibles o el agravamiento de enfermedades neurológicas.

2.5. Guerra cognitiva

La guerra cognitiva es una amenaza emergente que implica prácticas para interferir con la cognición de los objetivos humanos. Su objetivo es alterar la forma en que una población piensa y actúa, a menudo mediante el uso de herramientas tecnológicas [32].

Existen dos formas principales de guerra cognitiva:

- **Técnicas basadas en información:** Esta forma, más establecida, se centra en manipular las creencias de grandes poblaciones mediante el uso de propaganda y desinformación para influir en sus decisiones y comportamiento.
- **Ataques tecnológicos:** Esta forma se caracteriza por el uso de tecnologías avanzadas para influir de manera más fundamental en las funciones cognitivas y los estados emocionales. Los NeuroStrikes, en particular, emplean Radiación Electromagnética (REM), como las ondas de radio de baja intensidad, para alterar la actividad neurológica. Pueden causar un daño cognitivo significativo al interferir con la función cerebral del objetivo.

La guerra cognitiva está ganando cada vez más relevancia, siendo día a día una amenaza cada vez más real. Podemos analizar esta relevancia en tres aspectos diferentes: países que usan la guerra cognitiva, desarrollo de armas para la guerra cognitiva y mal uso en la investigación en neurociencia.

- **Países que usan la guerra cognitiva:** Como se ve en la investigación de Backes et al. [12], la guerra cognitiva ya forma parte del arsenal de muchos estados e instituciones. Se destaca cómo Rusia emplea campañas de desinformación sistemáticas y operaciones de influencia para moldear percepciones durante procesos electorales y conflictos armados.
- **Desarrollo de armas para la guerra cognitiva:** McCreight et al. [27] detalla proyectos destinados a diseñar NeuroStrikes basados en pulsos electromagnéticos y compuestos neuromoduladores. También recoge información que apunta a que China usa la investigación militar en estas tecnologías con fines de control cognitivo.
- **Mal uso de la investigación en neurociencia:** En el artículo “Neuroethics at 15: The Current and Future Environment for Neuroethics” [5], se advierte que avances legítimos en neurodispositivos pueden ser mal utilizados para fines militares dañinos. Se subraya la urgencia de marcos regulatorios que prevengan su aplicación como armas cognitivas.

2.5.1. Síndrome de La Habana

La “misteriosa enfermedad de los diplomáticos” (Golomb et al. [20]) comenzó en 2016, afectando a más de dos docenas de diplomáticos estadounidenses y sus familias en Cuba.

Posteriormente, algunos diplomáticos canadienses en Cuba y sus familias, y más tarde diplomáticos en China, también reportaron problemas similares. Muchos diplomáticos afectados informaron haber escuchado ruidos inusuales como chirridos o zumbidos durante los episodios que supuestamente desencadenaron los problemas de salud. Se reportó que estos ruidos eran localizados con precisión “similar a un láser” en algunas habitaciones o partes de ellas y que, en el área donde se percibía, el sonido parecía seguir a la persona. Golomb et al. [20] sugieren que estos sonidos percibidos se ajustan a las características del efecto Frey, el cual puede producir “sonidos” a través de la radiación de radiofrecuencia/microondas (RF/MW) pulsada. Este misterioso incidente pasó a llamarse como el Síndrome de La Habana, que a su vez se enmarca dentro de los incidentes conocidos en inglés Anomalous Healthy Incidents (incidentes anómalos de salud para los que no se ha encontrado una razón aparente).

Los síntomas de este incidente fueron variados. Podemos destacar los siguientes:

- **Síntomas auditivos distintivos:** La pérdida de audición, el tinnitus (zumbido en los oídos) y el dolor o la presión en los oídos son prominentes tanto en los diplomáticos afectados como en las personas que reportan síntomas por exposición a RF/MW. El tinnitus y la pérdida de audición de nueva aparición son particularmente distintivos entre los síntomas.
- **Otros síntomas:** Incluyen problemas de sueño, dolores de cabeza, problemas cognitivos, mareos, fatiga, problemas de equilibrio, y problemas de concentración y memoria. La náusea, problemas de visión y habla, y sangrado nasal (epistaxis) también se reportaron en algunos casos.
- **Sensaciones peculiares:** Los diplomáticos informaron sensaciones inusuales de presión y vibración.
- **Hallazgos cerebrales:** Se han reportado lesiones cerebrales, hinchazón cerebral y anomalías en la materia blanca en los diplomáticos. La lesión cerebral previa podría ser tanto un factor de predisposición genética como una consecuencia de la lesión por RF/MW.

Como se ha mencionado antes, Golomb et al. [20] sugieren que la coherencia con la exposición a RF/MW pulsada se establece a través del efecto Frey, también conocido como efecto auditivo de microondas o audición de RF. Este efecto explica cómo la RF/MW pulsada puede generar estos “sonidos” aparentes. La capacidad de oír estos “sonidos” inducidos por RF/MW depende de la audición de alta frecuencia del individuo y del bajo ruido ambiental, lo que concuerda con el hecho de que se escucharan principalmente por la noche y no por todas las personas expuestas. Además, la naturaleza del sonido percibido (clic, zumbido, silbido, golpe o chirrido) varía según las dimensiones de la cabeza y las características del pulso de la radiación, lo que explica las diferencias en los sonidos.

dos reportados por los diplomáticos. El hecho de que cubrirse los oídos no disminuyera el ruido también es consistente con los “sonidos” de RF/MW, ya que estos se generan internamente en la cabeza a través de la interacción de la radiación con el tejido cerebral, no por ondas de presión de aire que el oído externo captaría. Por ello, en caso de que el Síndrome de Habana sea consecuencia de RF/MW pulsada, estaríamos ante una de los pocos casos con consecuencias conocidas del efecto de NeuroStrikes en humanos.

3. Estado del Arte

En la literatura científica apenas hay investigaciones sobre los NeuroStrikes, al menos a la hora de intentar simular sus consecuencias con conjuntos neuronales. De hecho, solo se va a poder comparar este trabajo con la tesis de máster de Töllke (2024) [32]. Sin embargo, sí que hay trabajos que tratan temas parecidos, como la seguridad de las BCIs mediante la aplicación de ciberataques neuronales a grupos de neuronas simulados, como son López Bernal et al. 2020 [14], López Bernal et al. 2022 [15] y López Madejska et al. 2024 [26].

Vamos a analizar todos estos trabajos en orden cronológico, para ver cómo ha ido avanzando la investigación en este campo.

3.1. Definición de los primeros ciberataques neuronales

En este primer trabajo se introduce el concepto de “ciberataques neuronales” dirigidos a implantes cerebrales miniaturizados. Los autores identifican vulnerabilidades de seguridad y privacidad en las tecnologías BCI emergentes, como Neuralink y Neural Dust, que permiten la grabación, estimulación e inhibición de la actividad neural [14].

Además, se proponen dos ciberataques neuronales en base a ciberataques conocidos en la literatura de ciberseguridad. Estos ataques propuestos son el Neuronal Flooding y el Neuronal Scanning.

- **Neuronal Flooding (FLO):** Como su nombre indica, se trata de una “inundación” de estímulos neuronales, es decir, se estimulan de forma simultánea un gran número de neuronas a la vez.
- **Neuronal Scanning (SCA):** En lugar de estimular de forma simultánea un conjunto de neuronas, se estimulan discretamente neuronas individuales en instantes de tiempo diferentes. Es decir, en una selección de n instantes de tiempo seguidos, se estimula una neurona diferente en cada uno de estos.

Para poder realizar estos ataques y evaluar su impacto, en el estudio se usa el simulador Brian2 [1]. Dada la falta de topologías neuronales realista en ese momento, utilizan una topología basada en una red neuronal convolucional (CNN) entrenada para simular una

porción de la corteza visual de un ratón intentando escapar de un laberinto. En el mismo estudio se reconoce que no es una topología real, dando resultados dependientes de la misma, pero permite tener una primera aproximación al problema.

A la hora de medir el impacto de los ataques se definieron tres métricas: número de activaciones (spikes), porcentaje de desplazamientos (shifts) y dispersión de los spikes (tanto en el tiempo como en la cantidad de activaciones). Tras un análisis de los resultados en cada ataque, se llegó a la conclusión de que ambos ataques reducen el número de activaciones en comparación al comportamiento espontáneo, aumentando desplazamientos temporales y la dispersión. Además, se vio que FLO es el ataque más adecuado para generar daño inmediato y SCA para daño a largo plazo.

Al final de la investigación se señaló que, al tener resultados dependientes de una topología poco realista, se propone definir una taxonomía de ataques neuronales y ver cómo afectan a simulaciones más realistas.

3.2. Definición de nuevos ciberataques neuronales: ataques inhibitorios

En esta investigación, como continuación de la anterior, se diseña e implementa un nuevo ciberataque neuronal llamado Neuronal Jamming (JAM), que se enfoca en la inhibición de la actividad neuronal, impidiendo que las neuronas produzcan picos [15]. Esto contrasta con FLO y SCA, que se centran en la sobreestimulación de neuronas. Aún así, se continúa usando la topología neuronal generada a partir de una CNN. Otro cambio son las perspectivas desde las cuales se miden los impactos de los ataques, teniendo tanto un enfoque biológico (métricas neuronales como número de spikes, o dispersión temporal) como un enfoque comportamental (métricas de la capacidad del ratón como número de pasos o tasa de éxito).

Tras la adquisición de las métricas se ofrece una comparación de JAM y FLO. Se observa que JAM es más dañino en término de tasa de spikes neuronales y que FLO es más efectivo en las primeras posiciones del laberinto, mientras que JAM genera un mayor impacto cuando el número de posiciones consecutivas bajo ataque aumenta. Estos resultados, además, se acaban vinculando a enfermedades neurodegenerativas. Se discute la posibilidad de que ciberataques como JAM (inhibición) y FLO (hiperexcitabilidad) puedan recrear o exacerbar los efectos de enfermedades neurodegenerativas como el Alzheimer (desactivación del DMN) o la Esclerosis Lateral Amiotrófica (ELA) (hiperexcitabilidad cortical) si los implantes BCI tienen una cobertura cerebral suficiente.

Finalmente, se reitera la limitación de la topología neuronal no realista (derivada de CNN) y se propone explorar el uso de topologías realistas en el futuro.

3.3. Primeros resultados con una topología neuronal realista

Por primera vez este estudio supera la limitación de la falta de topologías neuronales realistas al evaluar el impacto de los ciberataques neuronales (FLO y JAM) en una representación neuronal realista de la corteza visual primaria de ratones. Utiliza una topología neuronal simplificada (450 neuronas) pero realista, basada en una reconstrucción de la capa 4 (L4) de la corteza visual primaria (V1) de ratones, desarrollada por Arkhipov et al. [9]. Este modelo incluye tanto neuronas excitatorias como inhibitorias y se realiza utilizando las herramientas NEST y BMTK.

Debido a tener una simulación topológicamente realista, hay diferencias clave con los estudios anteriores. En el ataque FLO se aumenta el número de spikes en vez de disminuir y en el ataque JAM se reduce el número de picos durante el ataque de inhibición seguido de un pico elevado al finalizar el ataque, debido a una resincronización de las neuronas. Otro hallazgo crucial es que, debido a la complejidad de la simulación, la actividad neuronal tiende a volver a su comportamiento espontáneo después de aproximadamente 500-600 ms después del ataque; lo cual sugiere que el cerebro posee mecanismos intrínsecos para estabilizar la actividad neuronal después de una anomalía, lo que no pasaba con modelos poco realistas.

Finalmente, aunque representa un avance significativo en realismo, la topología aún está simplificada. Se propone, por tanto, investigar el impacto cualitativo de estos ataques en funciones neuronales reales como la visión, como ceguera temporal, y escalar a topologías más grandes y complejas.

3.4. guerra cognitiva y simulación de NeuroStrikes

En último lugar tenemos la tesis de Töllke, que al igual que este trabajo, se centra en estudiar las consecuencias de los NeuroStrikes mediante simulaciones. Por lo tanto, en vez de centrarse en las BCIs y su seguridad, esta tesis ahonda en la guerra cognitiva como una amenaza emergente que busca interferir en la cognición del adversario, concretamente en los NeuroStrikes. El trabajo de Töllke explora los síntomas del Síndrome de La Habana (que usa como síntomas de los NeuroStrikes), centrándose en los daños relacionados con el proceso cognitivo de la memoria. Lo que propone Lennart, por tanto, es investigar y modelar los mecanismos subyacentes a los ataques electromagnéticos y su impacto en la memoria. Por ello, durante su trabajo busca identificar cómo estos ataques modifican la actividad cerebral, lo que le permite simularlos y analizar sus consecuencias.

A diferencia de las investigaciones anteriores, ya no se tiene una topología poco realista; por el contrario, Töllke selecciona un modelo de simulación del hipocampo (CA1, CA3 y GD) y la corteza entorrinal, desarrollado por Aussel et al. 2018 [10] y mejorado en 2022 [11]. Como se ha visto en la sección 2.5.1, el hipocampo es una estructura fundamental

para la consolidación de la memoria, por lo que la topología de la simulación es perfecta para analizar daños en esta función cognitiva tan esencial. Además, los datos cerebrales que se usan para la simulación proceden archivos EEG reales capturados durante la etapa de sueño en estudios clínicos, aportando un mayor valor a los resultados.

Para poder simular los ataques, en lugar de sobreestimular o inhibir neuronas, Töllke modificó los parámetros de la simulación para replicar alteraciones químicas y estructurales. Tras su investigación sobre los efectos de la radiación EM en seres vivos, concluyó que estos pueden manifestarse en la memoria como alteraciones de los neurotransmisores de acetilcolina (gCAN y gACh), la reducción de la conductancia sináptica (g_{max_e}) y la disminución en la cantidad total de neuronas (N_{max}).

La forma de evaluar el daño en la memoria fue medir las características de los SWRs, que como se ha visto en la sección 2.5.1, son imprescindible para la consolidación de la memoria. Se midió la ocurrencia, la frecuencia máxima, la duración media y el PSD; aunque esta última métrica media también las bandas theta y gamma para mayor información. Al analizar todas estas métricas, se veía, por lo general, una clara desviación del caso saludable, sugiriendo que ha habido un impacto negativo evidente en la memoria.

Finalmente, Töllke describe cuatro propuestas para trabajos futuros:

- Explorar diferentes ecuaciones y métodos de integración para asegurar un comportamiento más realista de la simulación.
- Mejorar la detección automática de SWRs en las grabaciones de campo local (LFP).
- Desarrollo de contramedidas específicas para ataques electromagnéticos.
- Investigación de otros aspectos cognitivos afectados.

Vamos a ver ahora una tabla comparativa entre la tesis de Töllke y este trabajo. En la Tabla 1 se observa cómo el simulador y la topología utilizados son los mismos en los dos trabajos. Sin embargo, los archivos de entrada usados no, ya que en este trabajo se amplían nuevos tipos de ataques. Además, las bandas estudiadas (señaladas con los puntos en la tabla) también son más, para así poder estudiar las repercusiones de los nuevos ataques.

Tabla 1: Tabla comparativa entre la tesis de Töllke [32] y este trabajo.

Trabajo	Simulador	Zona Simulada	Frecuencia (Hz)	Cantidad Neuronas	Input	Proceso Cognitivo	Cantidad			Duración			Frecuencia			PSD		
							T	G	R	T	G	R	T	G	R	T	G	R
Töllke [32]	Brian2 [1]	Hipocampo	1024	30000	EEG	Memoria		•			•			•	•	•	•	
Este Trabajo	Brian2 [1]	Hipocampo	1024	30000	EEG EEG Mod.	Memoria	•	•	•	•	•	•	•	•	•	•	•	

4. Objetivos y Metodología

4.1. Objetivos

Este trabajo tiene como objetivo la simulación de ataques de NeuroStrikes en el cerebro y el análisis de sus consecuencias.

Podemos definir los siguientes subobjetivos:

- Revisar el estado del arte de la guerra cognitiva, las BCIs, NeuroStrike, el hipocampo y la memoria para así tener una base sólida sobre la que trabajar.
- Estudiar el trabajo de Töllke y replicar sus resultados.
- Crear nuevas estrategias de ataque para simular NeuroStrikes a raíz de modificar las lecturas EEG de entrada del simulador.
- Obtener los resultados e interpretarlos, pudiendo explicar las consecuencias de los ataques (principalmente en la memoria).

4.2. Metodología

Para poder simular los NeuroStrikes se utiliza una topología del hipocampo, definida en el simulador Brian2, para simular ataques electromagnéticos de alta potencia. El trabajo partirá de la tesis de Töllke, intentando primero replicar sus resultados (con ataques simulados con los parámetros del simulador) y, después, se ampliará con ataques simulados a raíz de modificar los archivos EEG de entrada. La metodología utilizada para el trabajo ha sido la siguiente:

1. Lo primero que se hizo fue realizar una revisión exhaustiva del estado del arte en relación a los ciberataques neuronales y sus repercusiones en el cerebro. Tras ello, se profundizó en conocer cómo funcionan el hipocampo y la memoria, para así tener una visión de los mecanismos que permiten consolidar la memoria y poder extraer conclusiones de los ataques. Para acabar, se revisó en profundidad la tesis de Töllke y todo lo relacionado con la guerra cognitiva y los NeuroStrikes.
2. Una vez se obtuvo una visión general del problema, se clonó el repositorio de Töllke [3], se estudió su código y se replicaron sus resultados. Este proceso permitió obtener un mejor entendimiento sobre el problema y tener un punto de partida sólido sobre el que diseñar el presente trabajo.
3. Tras replicar los resultados se percibió una gran variación en las simulaciones del estado “saludable”. Por lo tanto, se lanzaron varias simulaciones para comprobar su variabilidad y, tras ver que era alta, se empezaron a usar promedios de varias simulaciones a la vez para reducir la entropía de los resultados.

4. Con los resultados de Töllke ya replicados, se discutieron nuevos métodos para simular NeuroStrikes. Se pensó en modificar los archivos “input” EEG del simulador para que aparentasen ser archivos EEG tras un ataque de NeuroStrike y lanzarlos con los parámetros “saludables”. Se idearon un total de cinco ataques: Jamming, Flooding, Scanning, Selective Forwarding y Nonce (este con dos variantes), inspirados en los ciberataques neuronales expuestos en [16].
5. Se rediseñaron las métricas que usaba Töllke y se crearon nuevas gráficas para poder visualizar los resultados.
6. Finalmente, con las gráficas ya generadas, se compararon los valores tras los ataques con el estado “saludable” de forma que, en base a la literatura, se relacionaron los resultados con daños cognitivos.

5. Diseño de la Solución

En esta sección vamos a analizar tanto el diseño de la solución, ofreciendo una visión a alto nivel del trabajo realizado, como la implementación llevada a cabo, donde se muestra un análisis a bajo nivel sobre los componentes y estructura de archivos.

5.1. Diseño

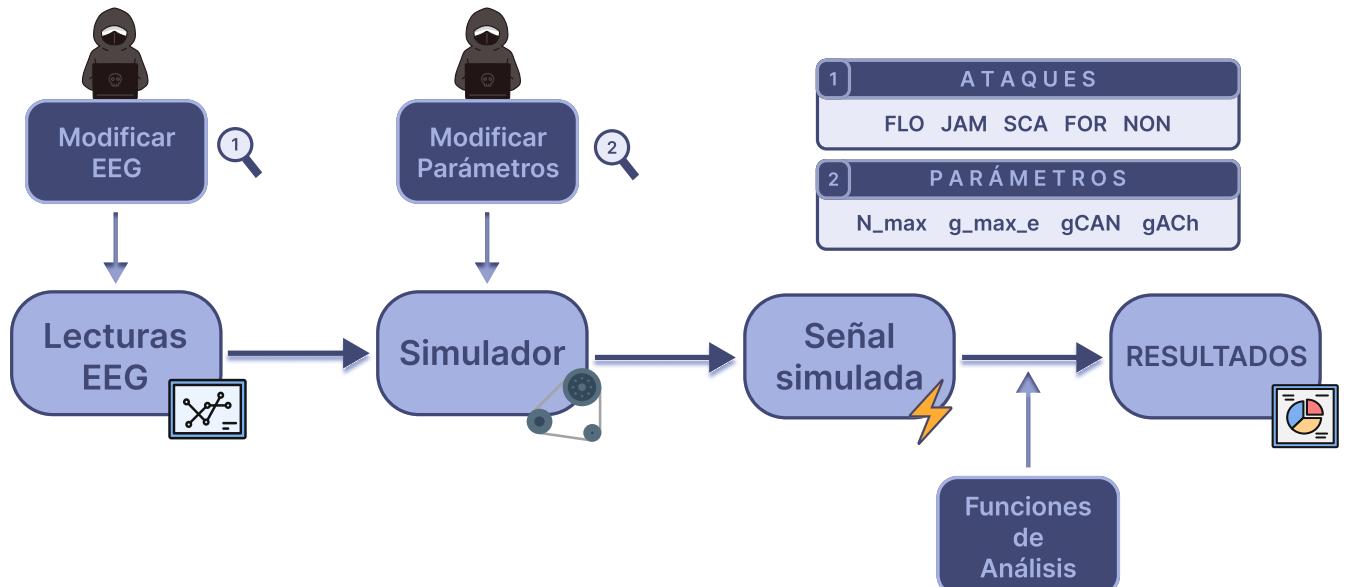


Figura 7: Esquema del diseño de la solución en el que se observa cómo funciona el simulador y cómo se generan los diferentes ataques.

En la Figura 7 se puede ver cómo funciona la solución implementada para simular NeuroStrikes. Vamos a explicar el flujo de trabajo para realizar las simulaciones (ya sean saludables o de ataques). Lo primero que se hace es introducir las lecturas EEG de entrada al simulador (pueden modificarse para simular ataques), para que tenga las señales iniciales; luego, se pone en marcha el simulador (pueden modificarse sus parámetros para simular ataques). Una vez realizadas las simulaciones, se generan archivos con valores de Potencial de Campo Local (LFP por sus siglas en inglés), medidos en voltios. Es a partir de estos archivos LFP de donde se sacan las métricas, es decir, extraemos las características de la señal resultante para poder determinar los daños. Las métricas que se miden son: la cantidad de ondas que hay en una simulación, la duración media de estas ondas, la frecuencia máxima media de las ondas y el PSD de la banda de frecuencias. Como no queremos medir todas las ondas, sino solamente las ondas Theta, Gamma y SWR, dentro de cada métrica habrá una división por espectro de frecuencia, de forma que en cada métrica tendremos un valor diferenciado para cada tipo de onda. Finalmente, cuando hemos extraído las características de la onda LFP y hemos calculado las métricas (usando las funciones de análisis que como se ve en la Figura 7) se realizan las gráficas, de forma que se obtienen los datos listos para analizar.

Se puede ver que, respecto a la generación de ataques, hay dos procedimientos alternativos:

- **Modificación de los parámetros de la simulación:** Esta es la forma en la que Töllke simula los NeuroStrikes. De lo que se trata es de modificar los parámetros `N_max`, `g_max_e`, `gCAN` y `gACh`. El parámetro `N_max` establece cuál es el tamaño máximo que pueden tener los grupos neuronales, de forma que los grupos más pequeños suponen un porcentaje de este tamaño; disminuir esta variable simula daños estructurales como muerte neuronal. El parámetro `g_max_e`, por su parte, establece la conductancia máxima de todas las sinapsis excitatorias de la red, por lo que disminuir su valor simula un daño estructural similar a daños sinápticos. Finalmente, los parámetros `gCAN` y `gACh` intentan simular daños químicos: `gCAN` representa la conductancia del canal iónico de calcio, que aumenta con la concentración de acetilcolina, por lo que una disminución representa una disminución de la acetilcolina. `gACh`, por su parte, es un factor que influye en una serie de conductancias sinápticas, lo cual escala también los niveles de acetilcolina. Esta disminución de acetilcolina supone una disminución de la plasticidad sináptica.
- **Modificación de archivos EEG:** Esta forma de simular NeuroStrikes es el nuevo método que se propone en este trabajo. El simulador necesita un input para empezar a funcionar, el cuál consiste en un conjunto de archivos EEG reales tomados de grabaciones durante un ciclo de sueño saludable. La idea sería modificar estos archivos “saludables” para que pasen a ser “corruptos”, de forma que se simule una actividad neuronal disfuncional debido a un gran ataque electromagnético. Los parámetros, por su parte, se mantendrán como los del estado “saludable”, de forma

que las repercusiones se deban exclusivamente a la modificación de los EEG. La forma de modificar los archivos EEG se ha inspirado en el la literatura de ciberataques neuronales, de forma que las taxonomías de los ciberataques dan nombre a las modificaciones [16]: Flooding (FLO), Jamming (JAM), Scanning (SCA), Selective Forwarding (FOR) y Nonce (NON). Vamos a explicar en qué consiste cada uno de ellos [16]:

- **Flooding (FLO):** El Flooding consiste en una estimulación completa de un grupo de neuronas durante un intervalo de tiempo especificado. En los archivos EEG, por tanto, el ataque FLO pasa a ser un aumento proporcional de toda la señal EEG durante un intervalo especificado (por ejemplo, aumentar $\times 10$ el valor de la señal durante 10 segundos).
- **Jamming (JAM):** Un ataque de Jamming, en contraposición al Flooding, consiste en la inhibición completa de un grupo de neuronas durante un periodo de tiempo. Por tanto, en nuestro caso consistirá en una disminución proporcional de toda la señal durante un tiempo establecido (como disminuir $\times 0,1$ la señal durante 10 s).
- **Scanning (SCA):** Este ataque, en la litratura de ciberataques neuronales, consiste en una estimulación individual de neuronas secuencialmente durante un periodo de tiempo, es decir, en cada instante de tiempo se estimula una neurona concreta. En los archivos EEG, como se captura la actividad de grandes cantidades de neuronas, lo que se hace es que se escala proporcionalmente la señal durante instantes muy cortos de tiempo, periódicamente.
- **Selective Forwarding (FOR):** La amenaza FOR, al contrario que el Scanning, consiste en una inhibición selectiva de neuronas en diferentes instantes de tiempo. Dadas todas las alternativas que puede haber a la hora de elegir cómo inhibir, la literatura propone un Scanning inhibitorio como la forma más básica de usar el FOR. Por tanto, en nuestro caso el ataque FOR se comportará como un Scanning inhibitorio, es decir, un Scanning que disminuya proporcionalmente la señal en vez de aumentarla.
- **Nonce (NON):** La estrategia Nonce se trata de un ataque aleatorio; es decir, en cada instante de tiempo se selecciona de forma aleatoria un grupo de neuronas que puede estimularse, inhibirse o no ser modificado. Para adaptarlo a los archivos EEG lo que se ha hecho es seleccionar un tamaño de intervalo de tiempo, dividir la duración del ataque en estos intervalos y, de forma aleatoria (con una probabilidad indicada), aumentar, disminuir o no hacer nada sobre la señal.

5.2. Implementación

En esta sección se va a comentar a bajo nivel todo lo que se ha hecho en el proyecto: qué programas se usan, cómo se prepara el entorno de simulación, qué funciones se han creado y los métodos que se usan para realizar los experimentos, entre otros.

5.2.1. Preparación del entorno de trabajo

Para ejecutar el simulador y el resto de código se ha partido del proyecto de Töllke en Github [3]. El proyecto está realizado en Python 3.6.9 y el entorno virtual está generado en Conda [2], con la lista de librerías en un archivo YAML llamado `env_ipp.yaml`.

Para poder instalar el entorno de Conda lo primero de todo hay que habilitar el canal `conda-forge` de Brian2 en Conda, como bien se indica en la guía de instalación dentro de su documentación [1].

```
conda install -c conda-forge brian2
conda config --add channels conda-forge # Para añadir permanentemente
```

Una vez tenemos el canal `conda-forge` listo, podemos proceder con la instalación del entorno de Conda. Para ello tenemos que estar en la misma carpeta que el archivo `env_ipp.yaml` y, una vez dentro, ejecutamos los siguientes comandos:

```
conda env create -f env_ipp.yaml      # Crea el entorno
conda hipp_sim activate                # Activa el entorno
```

Hecho todo esto, ya tenemos el entorno y ejecutable de Python listos para poder utilizar el simulador.

5.2.2. Métodos de evaluación y comprobación de la estabilidad de la simulación

Antes de analizar la estructura del proyecto y las funcionalidades de los archivos hay que comentar cómo se va a realizar la simulación de NeuroStrikes y cómo es la estabilidad de las propias simulaciones que se usan. Tras replicar los resultados de Töllke, se vio que simulaciones con los mismos parámetros daban resultados notablemente diferentes. Por ello, se creó un Jupyter Notebook en el cuál se evaluaron 10 simulaciones del estado saludable, de forma que se vieran las variaciones entre ellas. Tras ello, se confirmó que la variabilidad de la simulación era bastante grande, por lo que a partir de entonces se realizarían cuatro simulaciones por experimento para reducir la entropía de los resultados.

Por lo tanto, la idea en este trabajo es diferenciar entre “simulación” y “experimento”. Cuando se hable de “simulación” se referirá a solo una de las N tiradas que realiza el script `parallel_processing.py`, de forma que cada experimento tendrá cuatro simulaciones. Luego, la palabra “experimento” se asociará a un conjunto de características de ataque

(estas características serían las que definen el tipo de ataque y las variables dentro de este, es decir, mismas características indican mismo ataque con los mismos valores modificados) concreto, es decir, si ejecutamos un NeuroStrike de tipo “modificación de parámetros” con modificaciones en `N_max` y `g_max_e`, nos estaremos refiriendo a un solo experimento. Por otro lado, si nos referimos al conjunto de ataques de tipo “modificación de archivos EEG” estaremos hablando de un conjunto de seis experimentos (FLO, JAM, SCA, FOR y dos NON) con, a su vez, cuatro simulaciones cada uno.

Dentro de cada simulación todas las métricas se representarán con su media y desviación estándar (a excepción del número de eventos, que es solo una suma simple de eventos). Por lo tanto, si se quiere tener un valor representativo de cada conjunto de simulaciones, hay que calcular la media y el error estándar de estas métricas, lo cuál no es tan simple debido a que a su vez son valores medios con errores. Para ello, se aplicará las siguientes fórmulas:

$$\bar{\mu} = \frac{1}{N} \sum_{i=1}^N \mu_i \quad (1)$$

$$SE(\bar{\mu}) = \sqrt{Var(\bar{\mu})} = \frac{1}{N} \sqrt{\sum_{i=1}^N \frac{\sigma_i^2}{n_i}} \quad (2)$$

donde N es el número de simulaciones, $\bar{\mu}$ es la media final, μ_i es la media de la simulación i , $SE(\bar{\mu})$ el error estándar de la media, σ_i es la desviación estándar de la simulación i y n_i el número de elementos de la simulación i .

Luego, tendremos que tener alguna forma de poder comparar los estados alterados con el estado saludable. Para ello, se ha implementado una característica dentro de las funciones de gráficas, la cuál, si se pide al programa, devuelve el rango del conjunto de experimentos que va desde el valor mínimo, restando el error estándar, hasta el valor máximo, sumando el error estándar. Si devolvemos este rango después de introducir un conjunto de experimentos en estado saludable obtenemos el rango en el cuál los valores saludables pueden variar (se usarán 10 experimentos). A parte de esto, se han implementado rectángulos horizontales rojos dentro de las gráficas que muestran los ataques no saludables para representar este límite, de forma que se pueda ver en un solo vistazo qué valores quedan claramente fuera del estado saludable.

Las gráficas que se crearán serán: gráficas de un conjunto de 10 experimentos en el estado saludable, gráficas del conjunto de experimentos con parámetros, gráficas del conjunto de experimentos de modificación de archivos EEG (a excepción de NON), gráficas de un conjunto de 10 experimentos Nonce Estándar y gráficas de un conjunto de 10 experimentos Nonce Agresivo. La razón por la que se han realizado las gráficas de Nonce separadamente a los otros ataques es debido a su naturaleza aleatoria. Es decir, cada vez que se realiza un experimento Nonce se usa un archivo EEG modificado con el ataque NON; sin embargo,

como la naturaleza de este ataque es aleatoria (a diferencia del resto de ataques) el archivo resultantes es completamente diferente cada vez, por lo que el resultado de cada ataque va a variar. Por esto, en los resultados, las métricas de NON pueden estar por encima, por debajo o dentro del rango saludable, característica que no puede representarse haciendo simplemente la media de múltiples experimentos. Es por todo esto por lo que se ha decidido evaluar los ataques NON en gráficas separadas del resto, donde un conjunto de 10 experimentos nos dará una muestra suficientemente grande como para describir su comportamiento.

Como se ha estado diciendo, la forma final de comparar los resultados es mediante gráficas. Sin embargo, después de analizar las gráficas se realizará una tabla comparativa entre todos los ataques, de forma que se pueda ver de la manera más sencilla cuáles son las repercusiones de los ataques sobre las características de los eventos. Esto supone un problema, ya que los ataques Nonce son 10, por lo que hay que idear una métrica que evalúe si el conjunto de estos ataques tiene repercusiones y cuáles son. Es decir, queremos evaluar si suele haber repercusiones con los ataques Nonce y de qué tipo son, pero sin hacer una media simple ya que sustrae gran parte del comportamiento de los ataques. Por esto, se han ideado dos índices basados en cálculos y observaciones simples de las gráficas: índice de ataque y factor de polarización. A continuación se describen los pasos seguidos para obtener cada uno de ellos:

- **Índice de ataque (I_A):** Este índice mide si un conjunto de ataques ha tenido repercusiones significativas o no.
 1. Primero se elige la característica dentro de la cuál se quiere hacer el recuento. Por ejemplo, decidimos que vamos a analizar si ha habido repercusiones en la frecuencia de las ondas.
 2. Se cuentan todos los resultados que estén por encima o por debajo del rango saludable.
 3. Si su error no está dentro de ese rango se suma 1, si no, se suma 0.5. Se guarda el resultado en una variable (por ejemplo A).
 4. Se cuenta todos los resultados que estén dentro del rango saludable.
 5. Si su error está completamente dentro del rango saludable se suma 1, si no, se suma 0.5. Se guarda el resultado en una variable (por ejemplo B).
 6. Se divide A entre la suma de A y B : $I_A = \frac{A}{A+B}$.
- **Factor de polarización (F_P):** Este índice mide qué tipo de repercusiones han tenido los ataques, si mayormente inhibitorias o excitatorias.
 1. Primero se elige la característica dentro de la cuál se quiere hacer el recuento

(igual que en el índice anterior).

2. Se cuentan todos los resultados que estén por encima o por debajo del rango saludable.
3. Si el resultado está por encima del rango saludable se suma 1.
4. Si el resultado está por debajo del rango saludable se resta 1.
5. Se divide el resultado entre el número de experimentos por encima o debajo del rango saludable.

Estos índices I_A y F_P oscilan entre (0, 1) y (-1, 1) respectivamente. Si el I_A tiene un valor menor a 0.5, entonces indica que no ha habido daños, mientras que si es superior a este umbral entonces sí que el ataque ha tenido repercusiones notables. Sin embargo, no es un espectro discreto de sí o no, sino un indicador de ayuda, por lo que los valores cercanos a 0.5 están en un rango de duda, donde los resultados no son seguros. El F_P , por su parte, indica el tipo de ataque según su signo: si es positivo es excitatorio y si es negativo es inhibitorio. Sin embargo, si su valor es cercano a 0 (ya sea positivo o negativo), quiere decir que el ataque tiene las dos características, es tantas veces excitatorio como es inhibitorio.

5.2.3. Archivos, funciones y estructura del proyecto

Una vez está claro cómo se va a proceder con las simulaciones, vamos a ver la estructura de archivos que se ha creado. Esta estructura la podemos ver en la Figura 8.

Podemos distinguir dos carpetas principales dentro del proyecto: la carpeta `results` y la carpeta `simulation-eeg`. La primera es la carpeta en la cuál se van generando los resultados, mientras que la segunda es la carpeta que tiene los archivos correspondientes para poder realizar las simulaciones y gráficas pertinentes.

Vamos a empezar desglosando la carpeta `simulation-eeg`. Dentro de esta carpeta podemos ver una gran cantidad de archivos: fichero YAML de configuración del entorno, scripts gráficos y scripts de detección, entre otros. Veamos qué hace cada uno de ellos:

- **`parallel_processing.py`:** Este script se encuentra dentro de los archivos de Töllke. Es el script de la simulación, el cuál contiene todos los parámetros para simular el hipocampo. Cada uno de los parámetros de la simulación es una lista de Python, de forma que puede contener varios valores; esto se hace para que el script coja todos los valores de cada parámetro y haga una simulación con cada combinación de valores. Por ejemplo, si tenemos 1 valor para cada variable y una con 4 valores, habrá 4 simulaciones; si hay 1 variable con 2 valores, otra con 3 y otra con 5, se harán $2 \times 3 \times 5 = 30$ simulaciones. Todos estos archivos se guardarán dentro de una misma carpeta, dentro de la carpeta `results` (se explicará más adelante esta estructura).

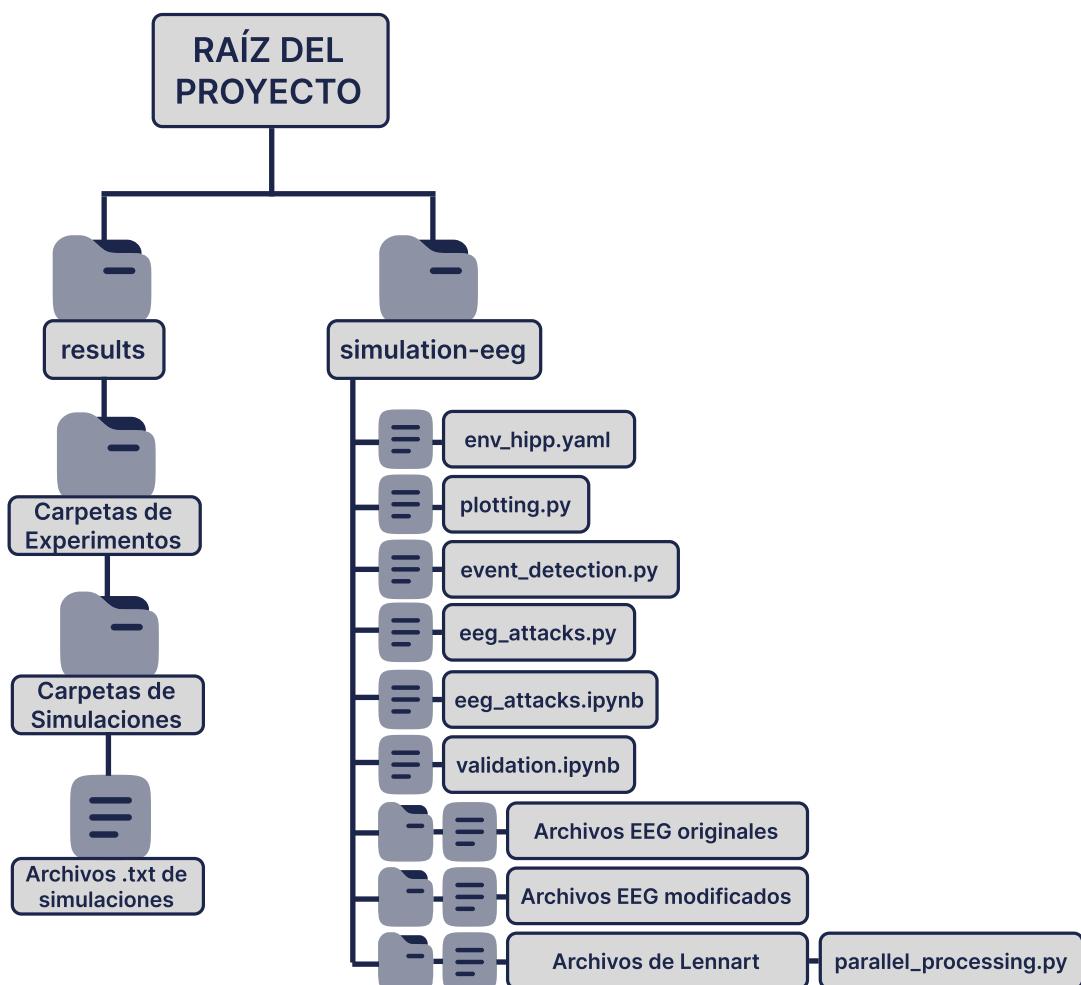


Figura 8: Estructura de archivos del simulador.

Para reducir la entropía de las simulaciones, como se ha dicho anteriormente, se han realizado cuatro ejecuciones por experimento.

- **event_detection.py:** Este es el script que contiene las funciones de detección de ondas. Está basado en el script de detección de Töllke, pero se ha modificado para poder detectar también ondas Theta y Gamma. Tiene 3 funciones: la función **frequency_band_analysis**, cuyo propósito es realizar un análisis y filtrado de una banda de frecuencia; la función **sharp_wave_detection**, cuya función es detectar ondas que se elevan por encima del V_{rms} y la función **event_detection**, cuyo objetivo es utilizar las funciones anteriores para detectar todos los eventos Theta, Gamma y Ripple.

- **eeg_attacks.py:** Este script contiene las funciones para modificar los archivos EEG, es decir, los ataques a las lecturas de entrada. Cada función tiene sus parámetros y funciona de forma diferente:

- **flooding_jamming_attack:** Esta función tiene como objetivo realizar el ataque FLO y JAM (dependiendo de los valores de las variables de entrada). Los parámetros que usa son: **signal**, la señal a modificar; **t**, vector de tiempo de la señal en segundos; **factor**, factor de escala de la señal, si es mayor que uno es un FLO y si es menor a 1 es un JAM y **t_attack**, tiempo durante el cuál se realiza el ataque. Lo que hace la función es localizar la señal dentro del tiempo de ataque especificado y multiplicar los valores de potencial por el factor de escala. Los valores que se han usado para el factor de escala son 10 y 0,1, ya que así asegurábamos aumentar un orden de magnitud la señal, creando un ataque con un gran impacto. Por ello, estos factores de escala se usarán en el resto de ataques también.
- **scanning_forwarding_attack:** El objetivo de la función es realizar un ataque SCA o FOR dependiendo de sus parámetros. Estos son: **signal**, la señal a modificar; **fs**, frecuencia de muestreo de la señal; **t**, vector de tiempo de la señal en segundos; **factor**, factor de escala de la señal, si es mayor que uno es un SCA y si es menor a 1 es un FOR; **t_attack**, tiempo durante el cuál se realiza el ataque; **dt**, tamaño de las ventanas individuales de ataque; **gap**, cada cuantas ventanas individuales se realiza un ataque: 1 = ataque FLO o JAM, 2 = 1 de cada 2, 3 = 1 de cada 3...).

Lo que se hace es dividir el tiempo de ataque en ventanas individuales de tamaño **dt** y multiplicar por el factor de escala una ventana cada número **gap** de estas. Determinar los valores de **dt** y **gap** es complicado, ya que cuanto menor sea **dt** y mayor sea **gap** más cercano será el ataque a un ataque a instantes de tiempos discretos (y por tanto más realista), pero si la diferencia es muy grande al final apenas vamos a tener pulsos y, por tanto, repercusiones. Al final se ha usado un **dt** = $1/fs * 100$ y **gap** = 4, de forma que se produzca un ataque de 100 unidades mínimas de tiempo cada 400 de estas mismas (una relación 1 a 5).

- **nonce_attack:** Esta función trata de realizar un ataque Nonce a la señal. Sus variables son: **signal**, la señal a modificar; **fs**, frecuencia de muestreo de la señal; **t**, vector de tiempo de la señal en segundos; **t_attack**, tiempo durante el cuál se realiza el ataque; **dt**, tamaño de las ventanas individuales de ataque; **prob**, tupla de probabilidades de realizar cada ataque, por ejemplo, (0.1, 0.5, 0.2)); **power**, tupla con los factores de escalado asociados a cada probabilidad, por ejemplo, (0.1, 1, 10)). Para ser consistentes con el resto de ataques se

ha escogido un $dt = 100/\text{fs}$.

Además, se han hecho dos tipos de ataques Nonce jugando con las probabilidades y los factores de escala: un ataque estándar y un ataque agresivo. El ataque estándar es el ataque que se muestra en el paper de la taxonomía de ciberataques neuronales [16], de forma que hay un 50 % de probabilidad de que no se cause ningún daño, un 50 % de que sí se cause daño y, dentro de esta última, un 50 % de que sea estimulante o inhibidor ($\text{prob} = (0,25, 0,5, 0,25)$, $\text{power} = (10, 1, 0,1)$). Por otro lado, el ataque agresivo tiene muchas más probabilidades de causar daño, siendo un paso intermedio entre Nonce estándar y SCA/FOR ($\text{prob} = (0,4, 0,2, 0,4)$, $\text{power} = (10, 1, 0,1)$).

- **eeg_attacks.ipynb:** Este es el Jupyter Notebook mediante el cual se realizan los diferentes ataques. En este Notebook se explica cómo funciona cada uno de los ataques y se usan las funciones del script `eeg_attacks.py` para crear los archivos con las lecturas EEG corruptas.
- **plotting.py:** Este es el script que contiene las funciones que permiten realizar los gráficos a partir de las métricas. Estas son:
 - **data_to_dict:** Esta función tiene como objetivo convertir los archivos de datos (`LFP.txt`) en un diccionario con las características de las ondas. Como entrada no tendrá la dirección del archivo, sino la carpeta que contiene, a su vez, otras carpetas (diferentes simulaciones) con los archivos de datos, quedando al final la siguiente estructura de diccionario:

```
{  
    caracteristica1: [simulacion1, simulacion2, ..., simulacionK],  
    caracteristica2: [simulacion1, simulacion2, ..., simulacionK],  
    ...  
    caracteristicaN: [simulacion1, simulacion2, ..., simulacionK]  
}
```

- **n_detects_plot:** Es la función que grafica la cantidad de eventos de cada banda de frecuencias que ha habido en la simulación. Se le introduce a la función una lista con las direcciones de los experimentos (donde cada uno de estos tiene a su vez cuatro simulaciones cada uno, como se ha mencionado anteriormente), de forma que se genere una gráfica de barras con todos los experimentos proporcionados, para poder compararlos entre sí. Para poder realizar esto la función, mediante un bucle, desglosa cada carpeta de experimento en un diccionario, que añade a una lista con todos los diccionarios (uno por experimento). Estos diccionarios tienen el atributo `x_list` (donde `x` es el nombre de la banda

de frecuencias) de forma que, contando el número de elementos de la lista, se tiene la cantidad de eventos. Finalmente, se hace una media de los recuentos ($\bar{n} = \frac{1}{4} \sum_{i=1}^4 n_i$) y se calcula el error estándar ($SE(\bar{n}) = \frac{\sigma}{\sqrt{N}}$). Estos valores y errores son los que se representarán, junto con el área que indica la variación del estado saludable.

- **durations_plot:** Es la función que grafica la duración media de los eventos de cada banda de frecuencias. Se le deben introducir las direcciones de los experimentos en forma de lista, de forma que la propia función desglosa la información en un diccionario y calcula la media global junto a su error estándar (para lo cuál usa las ecuaciones 1 y 2). Al igual que la anterior genera un gráfico de barras, en el cuál se pueden comparar de un solo vistazo todos los valores de los experimentos proporcionados.
- **peak_freqs_plot:** Es la función que grafica la frecuencia pico media de los eventos de cada banda de frecuencias. Se le deben introducir las direcciones de los experimentos en forma de lista, de forma que la propia función desglosa la información en un diccionario y calcula la media global junto a su error estándar (para lo cuál usa las ecuaciones 1 y 2). A diferencia de los casos anteriores no es un gráfico de barras, sino un gráfico de líneas. Debido a que las diferentes bandas tienen frecuencias pico muy alejadas unas de otras, no se representa todo en un mismo gráfico, sino que se hace una representación para cada banda. Al final tenemos tres gráficos con n etiquetas en el eje X, siendo este n el número de experimentos.
- **psd_plot:** Es la función que grafica la Densidad de Potencia Espectral (PSD) de cada banda de frecuencias. Al igual que las fórmulas anteriores se le proporciona las direcciones de los experimentos, de forma que se pasa a un diccionario y de ahí se calculan las medias y los errores con 1 y 2. Es un gráfico de barras, por lo que las bandas se representan todas juntas. Tiene un parámetro para alternar entre dos modos de Densidad Espectral. `scale` puede ser "density" (en cuyo caso estaremos representando el PSD) o "spectrogram" (en cuyo caso representaremos solo PS, es decir, la potencia espectral). Esto se ha hecho ya que Töllke usaba PS y no PSD (ya que las medidas estaban en V^2 y no en V^2/Hz), por lo que a la hora de comparar los resultados con los suyos es necesario estar en PS.
- **validation.ipynb:** Este Jupyter Notebook tiene dos funciones: comprobar la estabilidad de la simulación en el estado saludable y realizar las gráficas de los diferentes ataques (usando las funciones de `plotting.py`).

6. Análisis de Resultados

Vamos a empezar con el análisis de resultados. Podemos dividir esta sección en 5 grandes subsecciones: análisis de la estabilidad de la simulación, la comparación con los resultados de Töllke, la evaluación de los ataques y, finalmente, la explicación de las secuelas cognitivas provocadas por los ataques.

6.1. Análisis de la estabilidad de las simulaciones

En esta sección vamos a analizar la variabilidad de las diferentes simulaciones del estado healthy. Si este estado varía, el resto (que dependen de este) también lo harán. Vamos a realizar un análisis de las métricas que proporcionaba Töllke en su script de detección, ya que para el momento en el que se hizo este análisis aún no se habían ampliado las funciones. Por ello, se verá la variación de la duración y frecuencia pico de los SWRs, y la potencia espectral de todas las bandas. Esto se ve en las Figuras 9, 10 y 11.

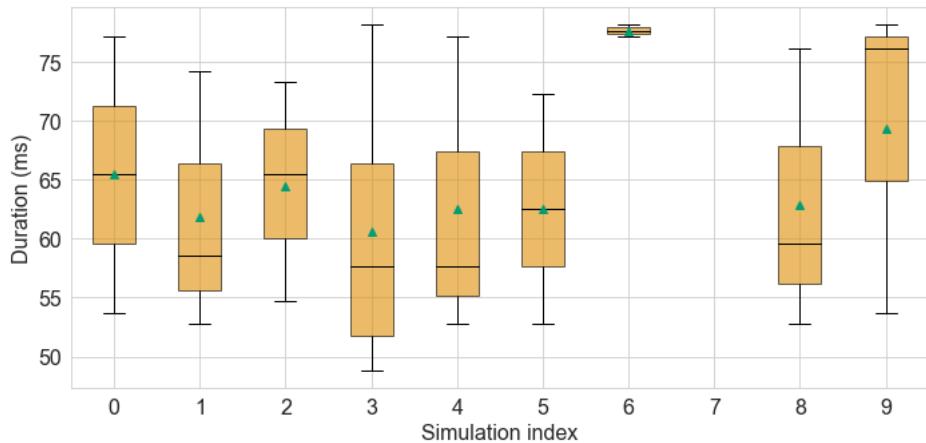


Figura 9: Resultados de 10 simulaciones individuales con los parámetros saludables. Se mide la duración de los SWRs y se representa con un diagrama de caja. El triángulo verde representa la media.

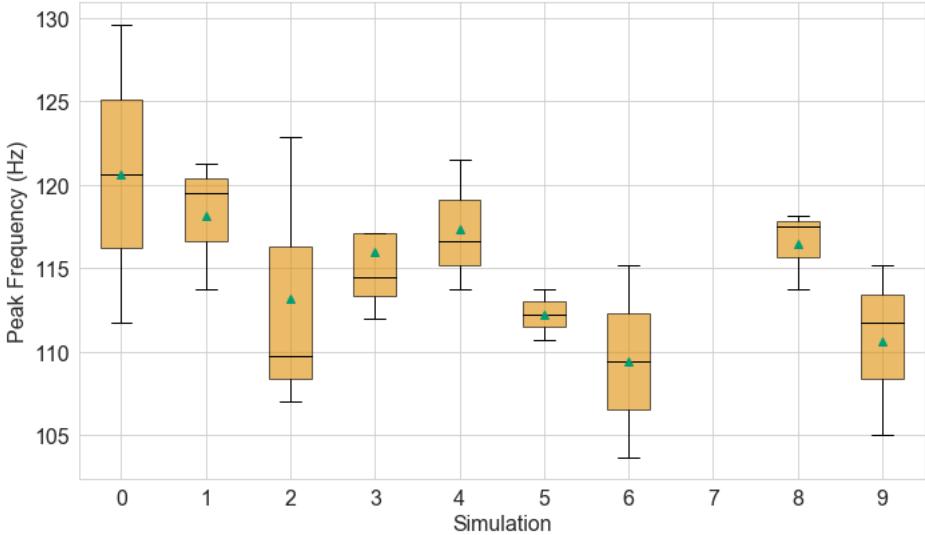


Figura 10: Resultados de 10 simulaciones individuales con los parámetros saludables. Se mide la frecuencia pico de los SWRs y se representa con un diagrama de caja. El triángulo verde representa la media.

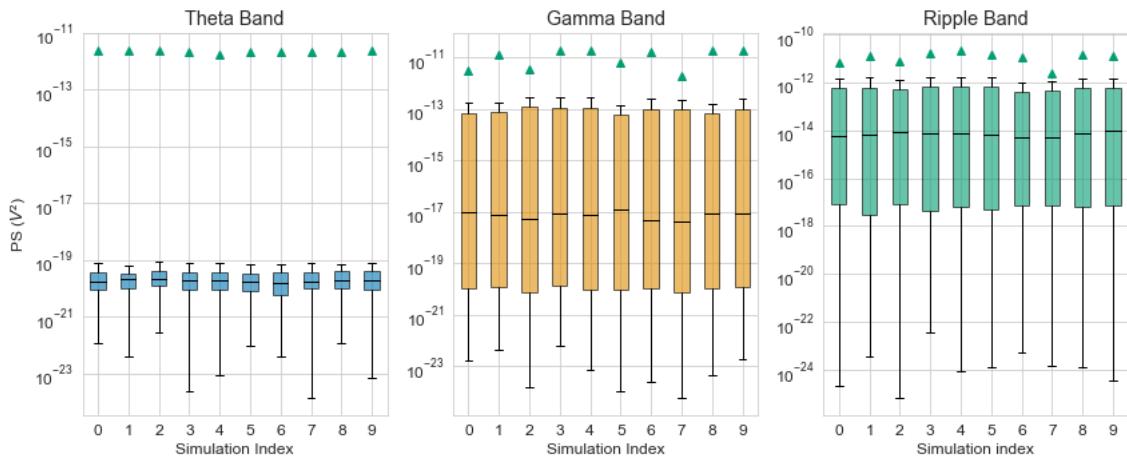


Figura 11: Resultados de 10 simulaciones individuales con los parámetros saludables. Se mide la Potencia Espectral de las diferentes bandas de frecuencia y se representa con un diagrama de caja. El triángulo verde representa la media.

En la Figura 10 se puede ver la variación de la frecuencia pico de los SWRs. En los diagramas de caja se ve que los resultados de los primeros labels del eje x parecen relativamente estables, pero conforme avanzamos por este eje las últimas simulaciones demuestran que podemos tener resultados algo erráticos (como las simulaciones 6 y 7), de forma que lan-

zando una sola simulación no podemos tener la certeza de que el resultado vaya a ser típico. Lo mismo pasa en la Figura 9, donde los valores de la duración de los SWRs varían bastante. En contraposición, la Figura 11 muestra cómo la Potencia Espectral no tiene tanta variación, sobre todo en la mediana. Aún así, como en el resto de métricas sí que hay una variación notable (incluso con la mediana), se ha decidido realizar un promedio de 4 simulaciones para cada experimento. No se ha escogido un número mayor debido al coste computacional.

6.2. Replicación y comparativa de los resultados de Töllke

En esta sección vamos a comprar los resultados de Töllke con los propios. Vamos a realizar diferentes secciones en función del tipo de ataque: saludable, N_{max} , g_{max_e} , $gCAN$, $gACh$, combinación de daño a neurotransmisores, combinación de daño estructural y ataque completo.

6.2.1. Estado Saludable

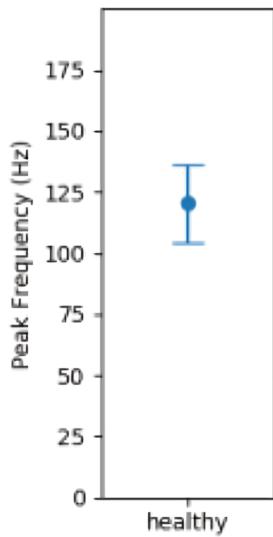


Figura 12: Media de la frecuencia pico de los Sharp Wave Ripples en una simulación saludable. Resultados de Töllke [32]

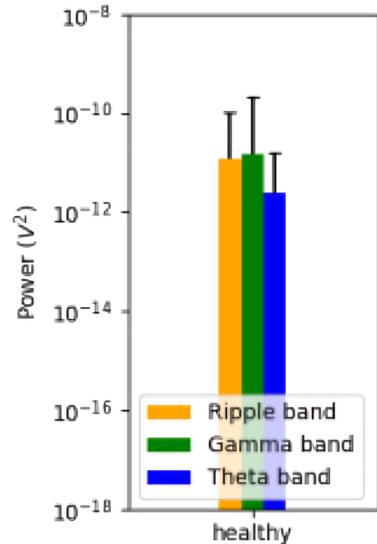


Figura 13: Potencia Espectral de las diferentes bandas en una simulación saludable. Resultados de Töllke [32]

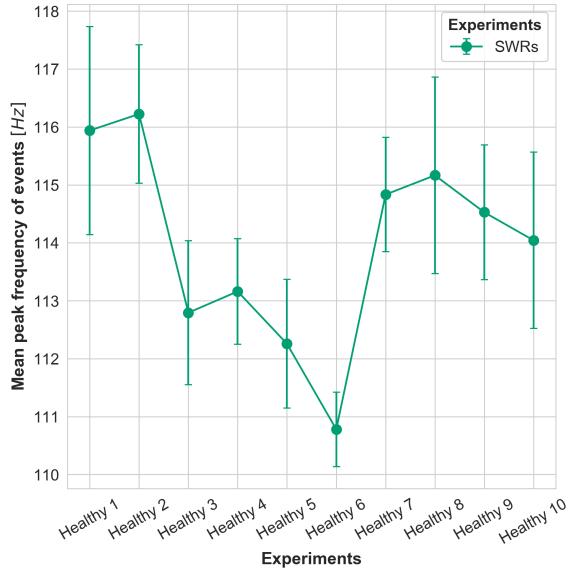


Figura 14: Media de la frecuencia pico de los Sharp Wave Ripples en cada experimento saludable, cada uno con cuatro simulaciones. Resultados propios.

Si nos fijamos en la frecuencia pico que calcula Töllke (Figura 12), esta se encuentra ligeramente por debajo de 125 Hz, mientras que en nuestro caso (Figura 14) se encuentra, de forma general, entre 110 y poco más de 115 Hz. Es perfectamente posible que, debido a la variabilidad de los resultados, el resultado de Töllke esté por encima de la media típica. Esto se puede ver en la Figura 10, donde la media de la simulación 0 se encuentra un poco por encima de 120, mientras que el resto se sitúan, por lo general, bastante por debajo de esta. Por otro lado, en la Figura 13, se puede apreciar cómo el valor PS de la banda Theta se encuentra entre 10^{-12} y 10^{-11} V^2 y los de las bandas Gamma y Ripple ligeramente por encima de 10^{-11} V^2 . Además, vemos que la banda Gamma es ligeramente superior a la banda Ripple. En nuestro caso (Figura 15), los valores de las bandas coinciden con los de Töllke, aunque se aprecia cómo las bandas Theta y Ripple están bastante igualadas; esto hace que, debido a la alta entropía de la simulación, a veces la banda Gamma esté por encima de la Ripple y viceversa.

Podemos afirmar, por tanto, que nuestros resultados del estado saludable coinciden con los de Töllke.

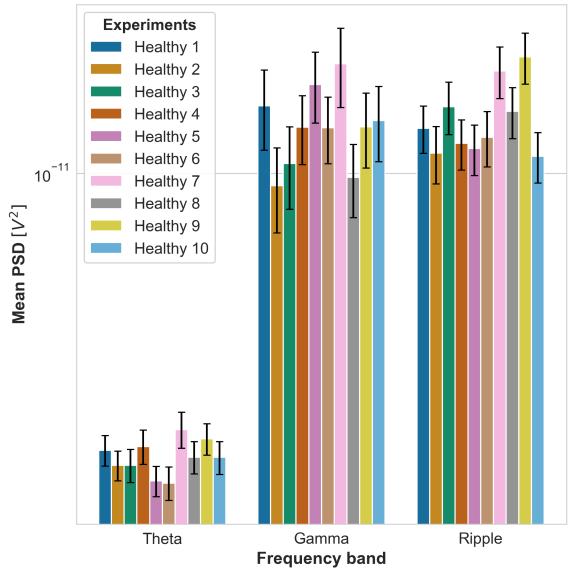


Figura 15: Potencia Espectral de las diferentes bandas en varios experimentos saludable, cada uno con cuatro simulaciones. Resultados propios.

6.2.2. Estados bajo ataques de modificación de parámetros

Ahora vamos a comparar los resultados de Töllke de los diferentes ataques con los propios de este trabajo. A diferencia del estado saludable, también se usará como métrica comparativa la duración de los SWRs.

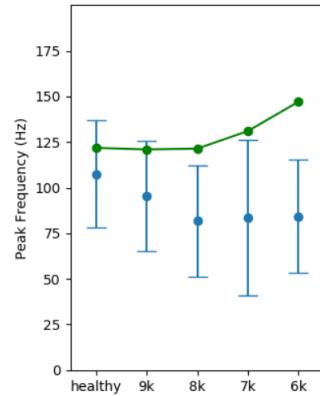


Figura 16: Frecuencia pico media para cada valor de N_{max} en simulaciones individuales. Resultados de Töllke [32].

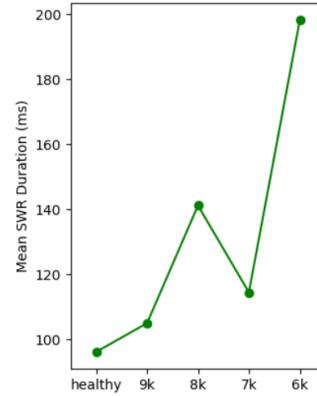


Figura 17: Duración media para cada valor de N_{max} en simulaciones individuales. Resultados de Töllke [32].

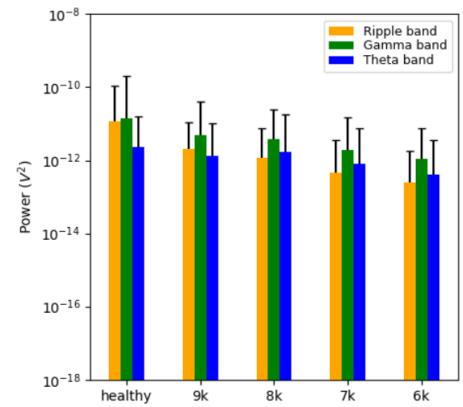


Figura 18: Valores PS en cada banda de frecuencia para cada valor de N_{max} en simulaciones individuales. Resultados de Töllke [32].

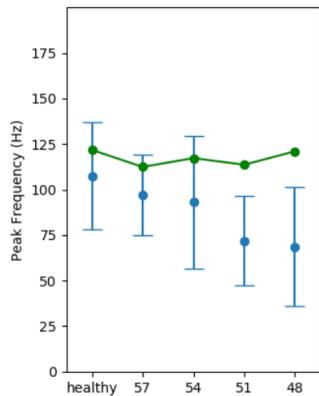


Figura 19: Frecuencia pico media para cada valor de g_{max_e} en simulaciones individuales. Resultados de Töllke [32].

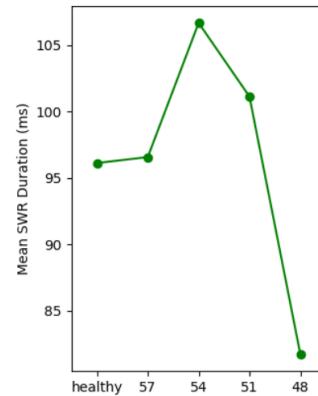


Figura 20: duración media para cada valor de g_{max_e} en simulaciones individuales. Resultados de Töllke [32].

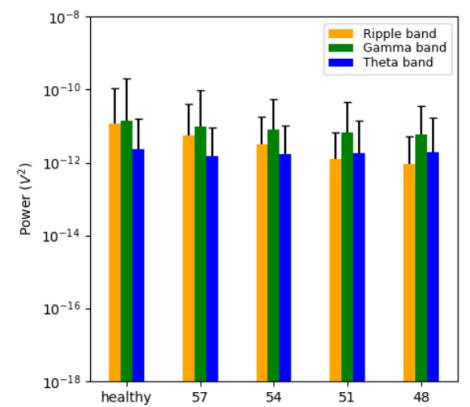


Figura 21: PS en cada banda de frecuencia para cada valor de g_{max_e} en simulaciones individuales. Resultados de Töllke [32].

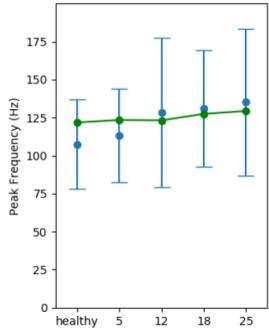


Figura 22: Frecuencia pico media para cada valor de gCAN en simulaciones individuales. Resultados de Töllke [32].

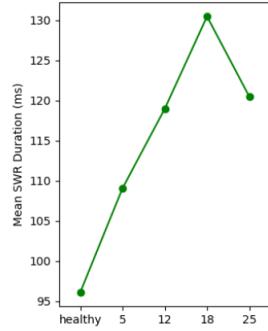


Figura 23: Duración media para cada valor de gCAN en simulaciones individuales. Resultados de Töllke [32].

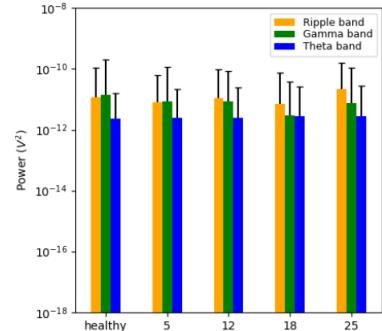


Figura 24: PS en cada banda de frecuencia para cada valor de gCAN en simulaciones individuales. Resultados de Töllke [32].

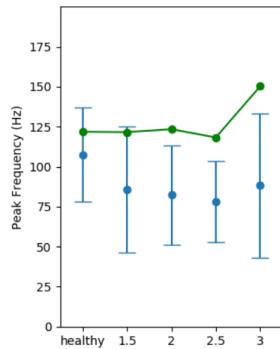


Figura 25: Frecuencia pico media para cada valor de gACh en simulaciones individuales. Resultados de Töllke [32].

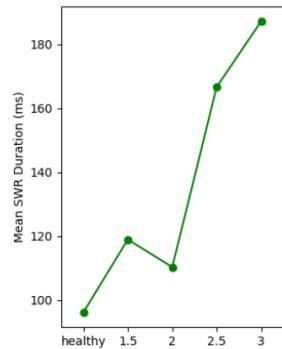


Figura 26: Duración media para cada valor de gACh en simulaciones individuales. Resultados de Töllke [32].

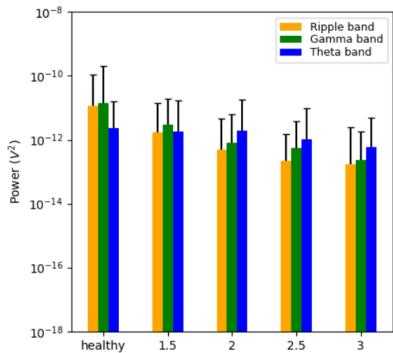


Figura 27: PS en cada banda de frecuencia para cada valor de gACh en simulaciones individuales. Resultados de Töllke [32].

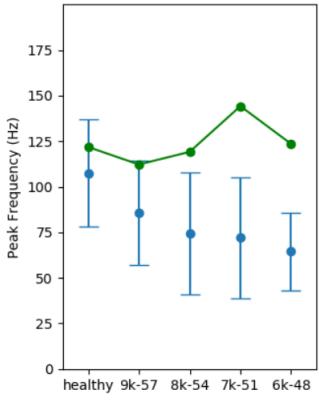


Figura 28: Frecuencia pico media para cada valor de N_{max} y g_{max_e} en simulaciones individuales. Resultados de Töllke [32].

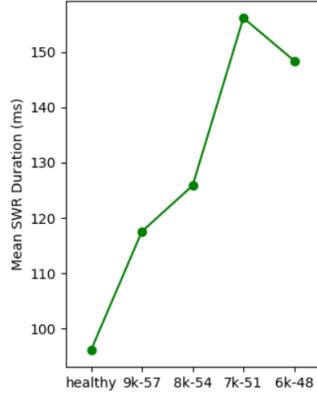


Figura 29: Duración media para cada valor de N_{max} y g_{max_e} en simulaciones individuales. Resultados de Töllke [32].

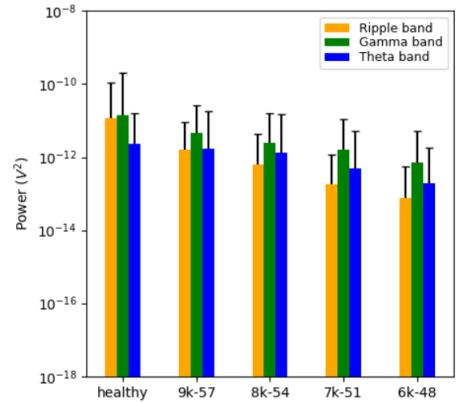


Figura 30: PS en cada banda de frecuencia para cada valor de N_{max} y g_{max_e} en simulaciones individuales. Resultados de Töllke [32].

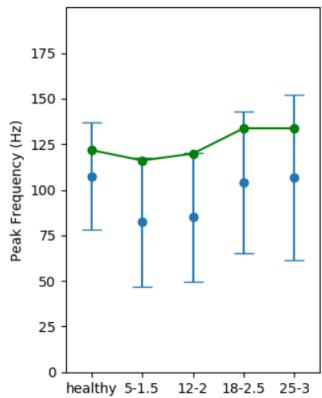


Figura 31: Frecuencia pico media para cada valor de gCAN y gACh en simulaciones individuales. Resultados de Töllke [32].

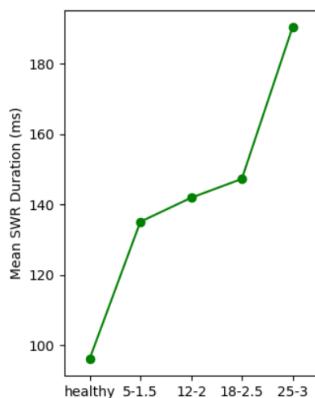


Figura 32: Duración media para cada valor de gCAN y gACh en simulaciones individuales. Resultados de Töllke [32].

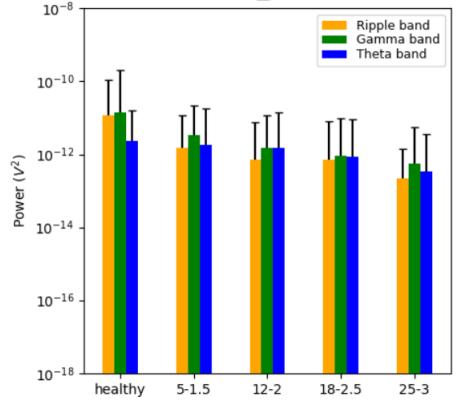


Figura 33: PS en cada banda de frecuencia para cada valor de gCAN y gACh en simulaciones individuales. Resultados de Töllke [32].

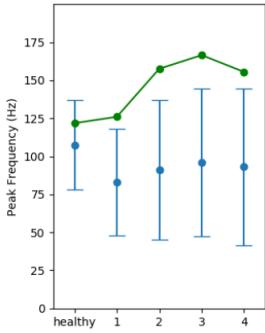


Figura 34: Frecuencia pico media para cada valor de N_{\max} , g_{\max_e} , gCAN y gACh en simulaciones individuales. Resultados de Töllke [32].

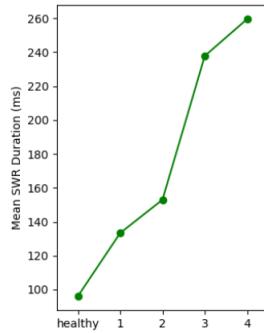


Figura 35: Duración media para cada valor de N_{\max} , g_{\max_e} , gCAN y gACh en simulaciones individuales. Resultados de Töllke [32].

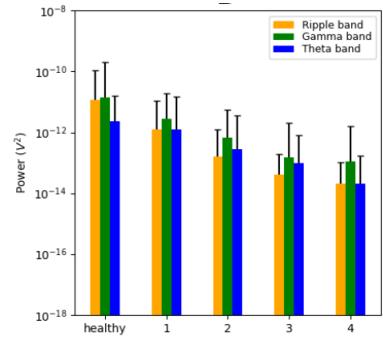
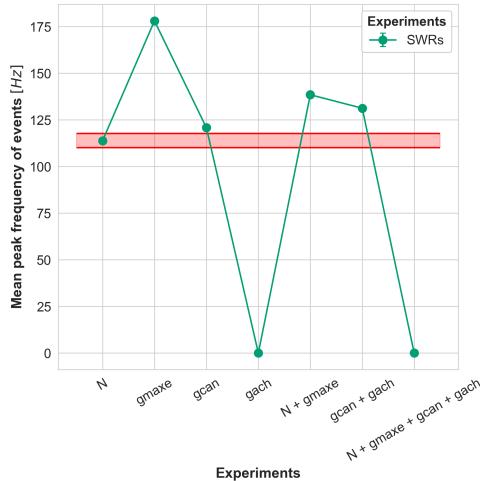
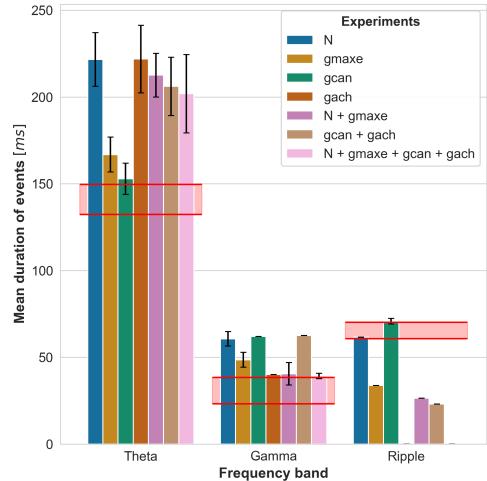


Figura 36: PS en cada banda de frecuencia para cada valor de N_{\max} , g_{\max_e} , gCAN y gACh en simulaciones individuales. Resultados de Töllke [32].



(a) Frecuencia pico



(b) Duración

Figura 37: Resultados propios de los ataques con parámetros. En las etiquetas, cada vez que aparece el nombre de un parámetro se está indicando que ese parámetro se ha modificado: $N_{\max}=6000$, $g_{\max_e}=48$, gCAN=25, gACh=3. En rojo está el rango de oscilación de saludable. (a) muestra la frecuencia pico media, (b) la duración media.

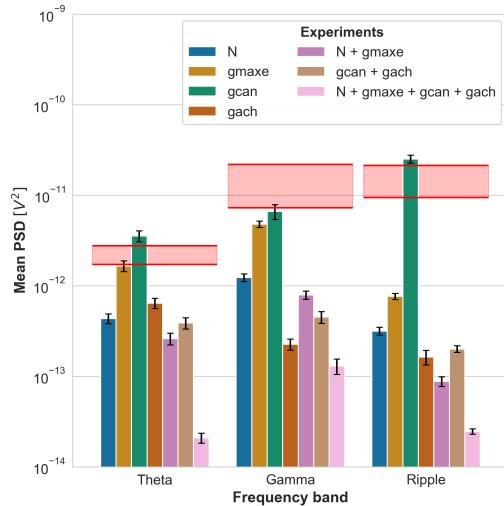


Figura 38: PS de las diferentes bandas en varios experimentos de ataques de modificación de parámetros, cada uno con cuatro simulaciones. Resultados propios. En las etiquetas, cada vez que aparece el nombre de un parámetro se está indicando que ese parámetro se ha modificado. Los valores de los parámetros modificados son: $N_{max}=6000$, $g_{max_e}=48$, $gCAN=25$, $gACh=3$. En rojo se puede ver el rango en el cuál el estado saludable ha oscilado.

Vamos a analizar uno a uno los resultados de los diferentes ataques.

- **Ataque con parámetro N_{max} :** En las gráficas de Töllke (Figuras 16, 17 y 18) se puede ver cómo la frecuencia Ripple asciende a 150 Hz, la duración a 200 ms y la PS desciende a 10^{-12} V^2 . En las Figuras 37 y 38, podemos ver cómo la frecuencia pico se mantiene a unos 115 Hz, la duración se mantiene al borde del rango saludable a unos 60 ms y la PS entre 10^{-12} y 10^{-13} V^2 . Se ve, por tanto, cómo los resultados propios son bastante más conservadores, ya que lo único que se ve perjudicado es la PS.
- **Ataque con parámetro g_{max_e} :** En las gráficas de Töllke (Figuras 19, 20 y 21) se observa cómo la frecuencia Ripple se mantiene cerca de 125 Hz, la duración cae a menos de 85 ms y la PS se queda entorno a 10^{-12} V^2 . Por otro lado, en las Figuras 37 y 38, podemos ver cómo la frecuencia pico asciende a 175 Hz, la duración desciende a unos 25 ms y la PS desciende a poco menos de 10^{-12} V^2 . Por lo tanto, el ataque se comporta de forma similar a los resultados de Töllke, a excepción de la frecuencia, que asciende en vez de mantenerse.
- **Ataque con parámetro $gCAN$:** En las gráficas de Töllke (Figuras 22, 23 y 24) se puede ver cómo la frecuencia Ripple se mantiene entorno a 125 Hz, la duración aumenta a 120 ms y la PS se mantiene igual con un ligero aumento. Por otro lado, en las Figuras 37 y 38, se aprecia cómo la frecuencia Ripple es ligeramente superior

al rango saludable (casi 125 Hz), la duración se mantiene en el borde superior del estado saludable (unos 70 Hz) y la PS asciende ligeramente por encima de los valores saludables. Por lo tanto, a excepción de la duración, los ataques coinciden.

- **Ataque con parámetro gACh:** En las gráficas de Töllke (Figuras 25, 26 y 27) se puede ver cómo la frecuencia Ripple asciende a 150 Hz, la duración asciende a 180 ms, y la PS Ripple desciende a menos de 10^{-12} V². Por otro lado, en las Figuras 37 y 38, se puede observar cómo la frecuencia y la duración caen en picado a 0 Hz, y la PS desciende a poco más de 10^{-13} V². Por lo tanto, el ataque propio ha sido mucho más destructivo que el de Töllke.
- **Ataque con parámetros N_max y g_max_e:** En las gráficas de Töllke (Figuras 28, 29 y 30) se puede ver cómo la frecuencia se mantiene en 125 Hz, la duración asciende a 150 ms y la PS desciende a unos 10^{-13} V². Por otro lado, en las Figuras 37 y 38, la frecuencia pico asciende entre los 130 y 140 Hz, la duración baja a unos 25 ms y la PS desciende por debajo de 10^{-13} . Los resultados de frecuencia y duración difieren, pero los de PS coinciden.
- **Ataque con parámetros gCAN y gACh:** En las gráficas de Töllke (figuras 31, 32 y 33) puede apreciar cómo la frecuencia Ripple aumenta ligeramente, la duración tiene una gran subida por encima de los 180 ms y la PS desciende a 10^{-12} V². Por otro lado, en las Figuras 37 y 38, la frecuencia pico asciende a unos 130 Hz, la duración desciende a menos de 25 ms y la PS se mantiene entre 10^{-13} y 10^{-12} V². Se ve que los valores de frecuencia y PS difieren pero mantienen las mismas tendencias, mientras que la duración es completamente diferente.
- **Ataque completo:** En las gráficas de Töllke (figuras 34, 35 y 36) se aprecia cómo la frecuencia Ripple aumenta a 150 Hz, la duración aumenta a 260 ms y la PS desciende por debajo de 10^{-13} V². Por otro lado, en las Figuras 37 y 38, la frecuencia y la duración descienden a 0 Hz y 0 ms, mientras que la PS desciende entre 10^{-14} y 10^{-13} V². Los valores de PS coinciden, pero la frecuencia y la duración, debido a que no se han llegado a registrar eventos Ripple, dan resultados completamente diferentes.

Hay que destacar que nos hemos centrado en analizar solo los resultados de los Ripples ya que el trabajo de Töllke [32] se centra concretamente en las repercusiones de los SWRs. Por eso, aunque en algunas gráficas como las de PS tengan las bandas de Theta y Gamma, no las hemos comparado. Después de analizar todos los resultados, se puede ver como aunque algunos coincidan otros dan resultados diferentes o incluso llegan a mostrar tendencias contrarias. Esto, sin embargo, seguramente se deba a la aleatoriedad de las simulaciones. Al tener solo una simulación de referencia en los resultados de Töllke, es posible que los casos en los que las tendencias no coincidan se deban a valores atípicos en sus resultados.

6.3. Análisis de los ataques con las nuevas métricas



Figura 39: Resultados de los ataques basados en cambios de parámetros. En rojo se muestra el rango saludable.

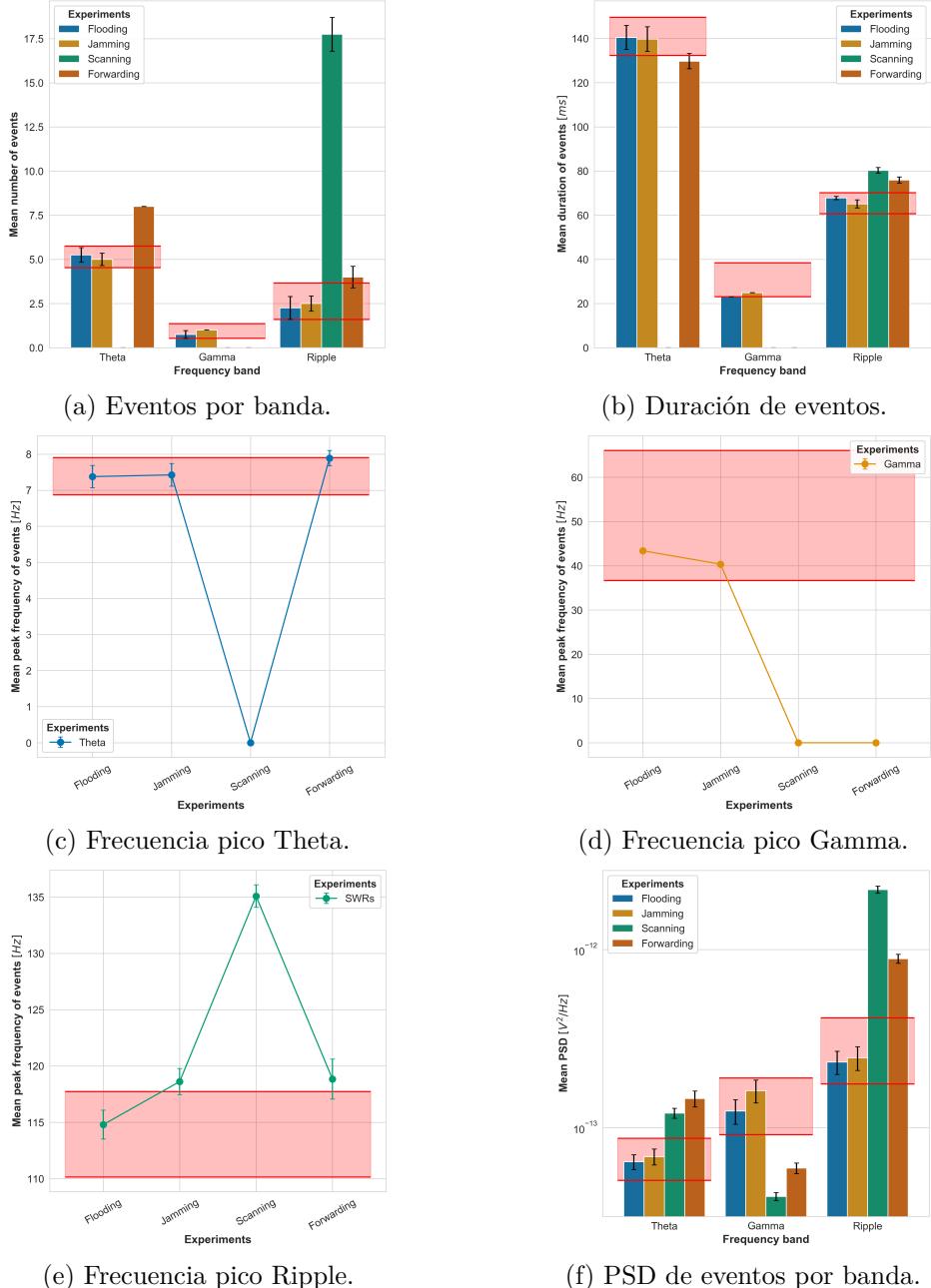


Figura 40: Resultados de los ataques a archivos EEG, sin incluir Nonce. En rojo se muestra el rango saludable.

Ahora vamos a explicar los resultados de los ataques. Para ello, primero explicaremos de forma general cómo se ha comportado cada grupo de ataques (parámetros, EEG,

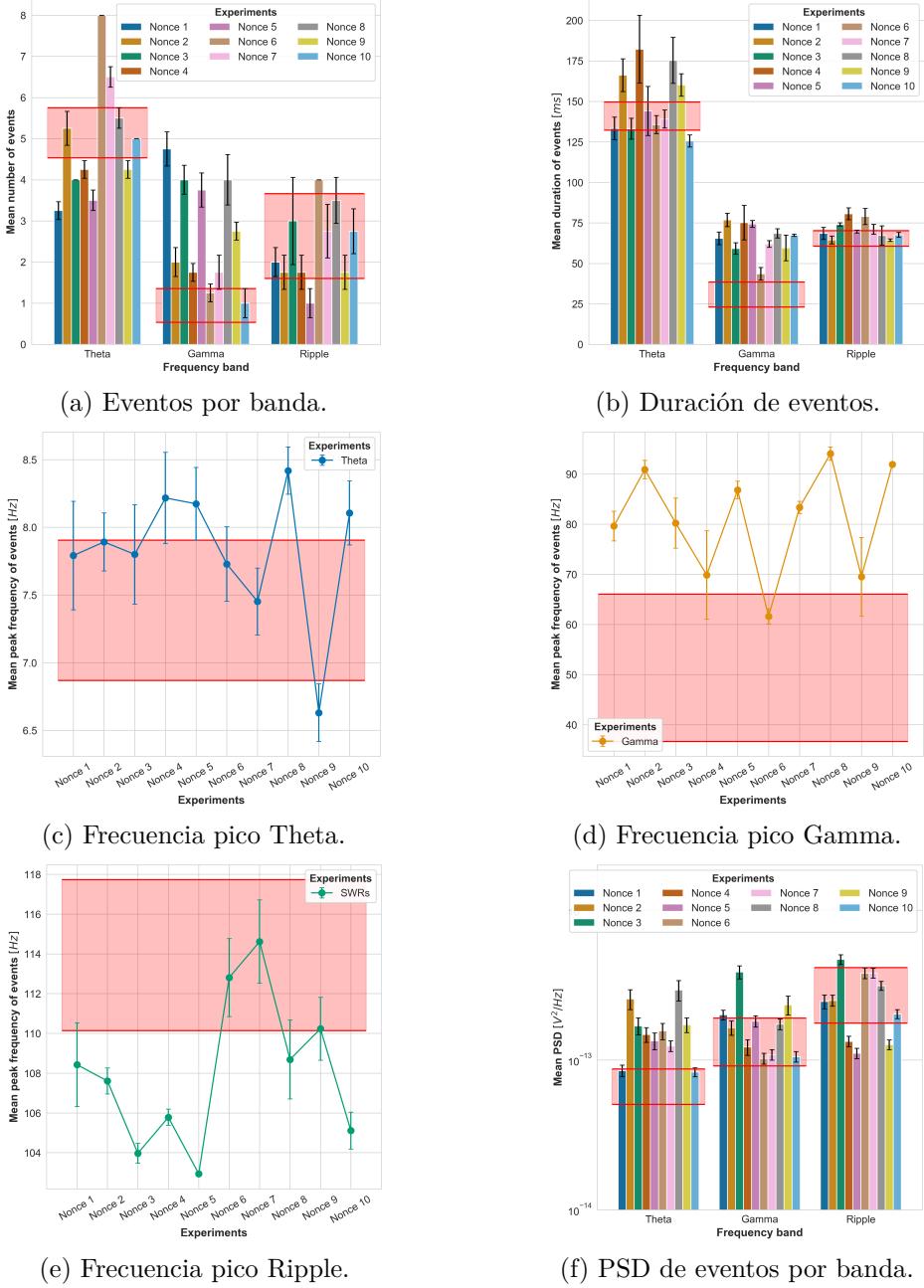


Figura 41: Resultados de los ataques Nonce. En rojo se muestra el rango saludable.

NON y NON agresivo). Luego, se mostrará una tabla donde se compararán los ataques individuales, cuánto daño hacen y de qué tipo (excitatorio o inhibitorio).

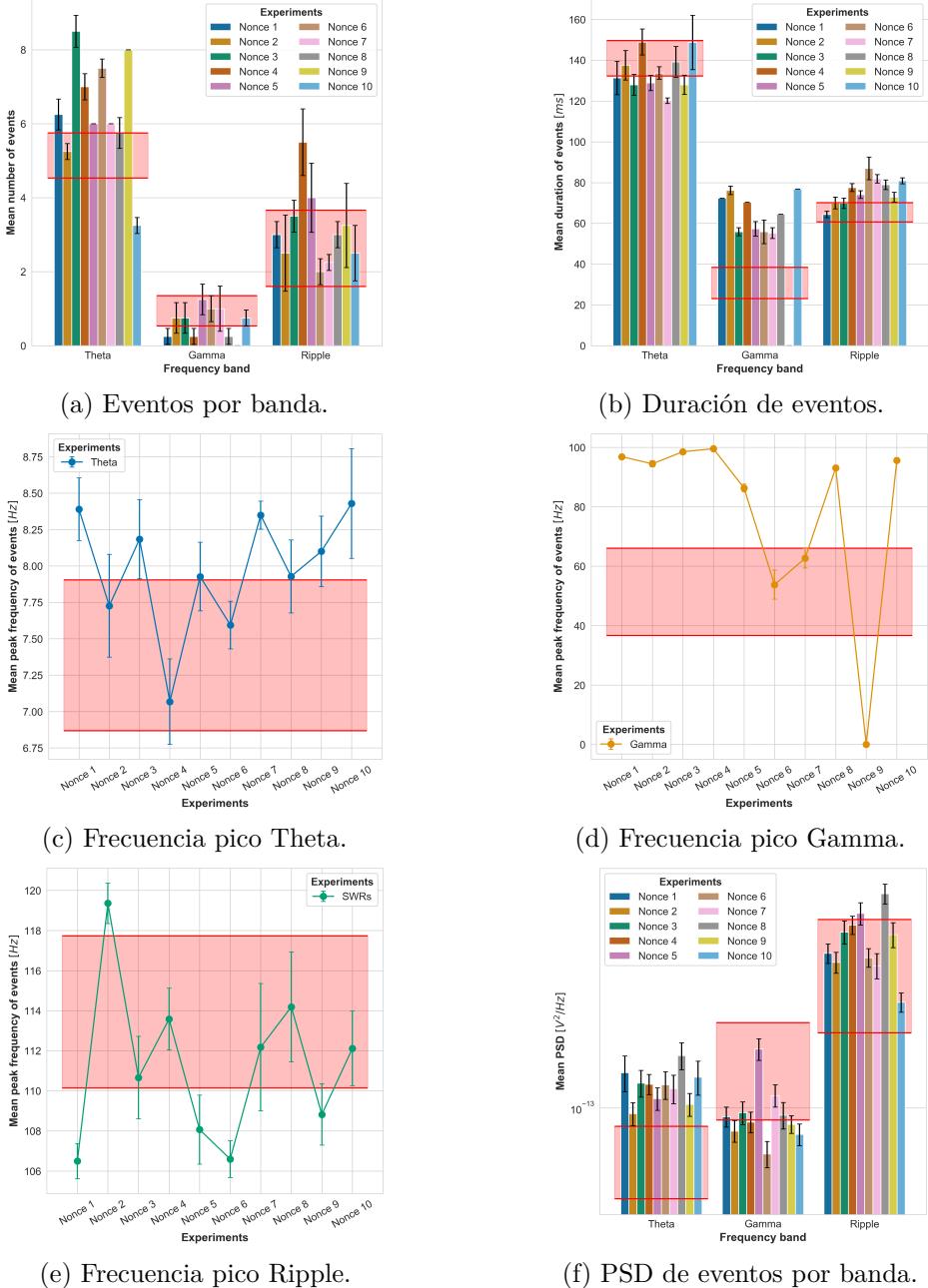


Figura 42: Resultados de los ataques Nonce Agresivo. En rojo se muestra el rango salu-dable.

- **Ataques de parámetros:** Podemos ver los ataques de parámetros en la Figura 39. Estos ataques, como se ha mencionado anteriormente, tratan de causar daños neu-

rológicos modificando las características de la simulación. Si hacemos zoom a las diferentes subfiguras, vemos que en (a), (b), (e) y (f), la mayoría de los ataques modifican las características medidas. Sin embargo, tienden a no cambiar (o hacerlo muy poco) las frecuencias pico de las bandas Theta y Gamma ((c), (d)). Como norma general, los ataques tienden a disminuir la cantidad, duración y PSD de los SWRs (a excepción del ataque de gACh). También tienden a disminuir el PSD de Gamma y Theta, y la cantidad de eventos Theta. Por otro lado, aumenta la cantidad de ondas Gamma y la duración de las ondas Theta y Gamma.

- **Ataques a archivos EEG:** Podemos ver los ataques a archivos EEG en la Figura 40 (a excepción de los ataques Nonce). Estos ataques, como se menciona en apartados anteriores, tratan de modificar los archivos EEG de entrada para simular daños neuronales. Se puede observar cómo los resultados del Flooding y el Jamming tienden a ser conservadores, manteniéndose siempre dentro del rango saludable. Por otro lado, los ataques Scanning y Selective Forwarding causan daños más notables, alterando enormemente el resto de características. El ataque de Scanning elimina completamente los eventos Theta y Gamma, aunque luego aumenta el PSD de Theta. El ataque FOR tiene le mismo impacto en la banda Gamma, mientras que en Theta da resultados más conservadores (aunque con modificaciones igualmente notables en ciertas características). Por otro lado, los dos ataques, en mayor o menor medida, tienden a aumentar las características de los Ripples.
- **Ataque Nonce:** Podemos ver los diferentes ataques Nonce en la Figura 41. Debido a su naturaleza aleatoria, se han tenido que realizar varios ataques Nonce con las mismas características pero con diferentes semillas aleatorias. En los resultados se puede observar esa aleatoriedad, pero dentro de esta aleatoriedad hay un cierto orden. Las características de los Ripples (a excepción de la frecuencia) no suelen cambiar, al igual que duración Theta, la frecuencia Theta y el PSD Gamma. Por otro lado, hay características con tendencias muy claras a aumentar o disminuir, como la cantidad, duración y frecuencia de los eventos Gamma o el PSD de las ondas Theta, que tienden a aumentar.
- **Ataque Nonce:** Podemos ver los diferentes ataques Nonce en la Figura 42. Este, al ser un tipo de ataque Nonce pero con características más agresivas, tiene los mismos problemas de aleatoriedad. Al igual que el ataque anterior, no tiene mucha repercusión sobre la banda Ripple (ni siquiera en la frecuencia, a diferencia del Nonce estándar). Sin embargo, la duración de Gamma, la frecuencia de Gamma y el PSD de Theta tienden a aumentar de forma sólida. Aunque parece que este ataque tiene una ligera tendencia a aumentar la frecuencia pico de Theta.

Finalmente, a modo comparativo, se muestra una tabla con los índices y factores descritos en la Sección 5.2.3. En el caso de Nonce, en cada casilla habrá 2 valores: I_A y F_P (el primero

encima del segundo). Para leer los resultados habrá primero que fijarse en el I_A , si este es mayor que 0.5 entonces se puede decir que el ataque ha funcionado (aunque si está muy cerca de 0.5 la evidencia no es tan clara); luego, el signo del F_P indicará si el ataque ha sido excitatorio o inhibitorio. Para el resto de ataques simplemente se usará un valor (-1, 0 o 1), el cuál indica si el ataque ha tenido efecto o no (± 1 o 0) y la naturaleza de este (excitatoria o inhibitoria).

Los resultados de la Tabla 2 se analizarán en la siguiente sección.

Tabla 2: Tabla comparativa con los resultados de las diferentes simulaciones de NeuroStrike. Cada fila representa un tipo de ataque, mientras que cada columna representa una característica y una banda de frecuencia. Se ha coloreado con rojo intenso los ataques que han provocado una excitación, en azul intenso los que han provocado una inhibición y en los mismos colores más claros los que han provocado los mismos efectos pero no de forma tan clara.

Ataques		Cantidad			Duración			Frecuencia			PSD		
		T	G	R	T	G	R	T	G	R	T	G	R
Parámetros	N	-1	1	-1	1	1	0	0	1	0	-1	-1	-1
	g_max_e	0	1	-1	1	1	-1	0	0	1	0	-1	-1
	gCAN	-1	-1	1	1	1	0	0	1	1	1	0	1
	gACh	-1	-1	-1	1	0	-1	1	1	-1	-1	-1	-1
	Estructural	0	1	-1	1	1	-1	0	0	1	-1	-1	-1
	Químico	-1	-1	-1	1	1	-1	0	0	1	-1	-1	-1
	Completo	-1	1	-1	1	0	-1	0	0	-1	-1	-1	-1
EEG	Flooding	0	0	0	0	0	0	0	0	0	0	0	0
	Jamming	0	0	0	0	0	0	0	1	0	0	0	0
	Scanning	-1	-1	1	-1	-1	1	-1	-1	1	1	-1	1
	Forwarding	1	-1	1	-1	-1	1	0	-1	1	1	-1	1
	Nonce	0,7 -0,43	0,83 1	0,27 0	0,57 0,6	1 1	0,38 1	0,57 0,6	0,89 1	0,71 -1	0,89 1	0,28 1	0,4 -0,5
	Nonce	0,84	0,43	0,19	0,58	1	0,78	0,73	0,8	0,53	1	0,53	0,17
	Agresivo	0,75	-1	1	-1	0,8	1	1	0,75	-0,6	1	-1	1

6.4. Asociación de resultados con daños cognitivos

Para tener claros los posibles efectos de los ataques, vamos a realizar un pequeño esquema de qué ocurre al aumentar o disminuir las características de las diferentes ondas según la bibliografía discutida en la Sección 2.5.1.

■ Ondas Theta:

- **Disminución:** Perjudica la capacidad del hipocampo para organizar y codificar secuencias de información y formar asociaciones, reduciendo el acoplamiento theta-gamma; esto se observa en modelos de enfermedad de Alzheimer y en

pacientes con deterioro cognitivo leve, asociándose con déficits en memoria de trabajo y navegación espacial [28].

- **Aumento patológico:** Un incremento patológico de la potencia theta y su acoplamiento con gamma en el hipocampo ventral de ratones se asocia con ansiedad e hiperactividad, pudiendo repercutir indirectamente en la memoria [21].

■ Ondas Gamma:

- **Disminución:** Se vincula a alteraciones de la memoria espacial, deterioro olfativo y trastornos de aprendizaje y memoria en Alzheimer, debido a la pérdida de potencia en el giro dentado y el bulbo olfatorio [21].
- **Aumento patológico:** Oscilaciones gamma demasiado fuertes se han asociado también a defectos de aprendizaje y a patologías como Alzheimer, reflejando un desajuste entre excitación e inhibición neuronal [21].

■ Ondas Ripple:

- **Disminución:** Bloquear los SWRs durante el reposo impide el “replay” de secuencias neuronales y deteriora el aprendizaje espacial en ratas [19].
- **Aumento patológico:** Pequeñas perturbaciones pueden convertir SWRs en “p-ripples” (oscilaciones patológicas supersincrónicas), aisladas o asociadas a descargas epilépticas interictales, que también dañan el aprendizaje espacial [24] [19].

Ahora que tenemos un esquema como referencia, vamos a ver las consecuencias que ha tenido cada ataque. Cabe destacar que dadas las características de la implementación hecha para poder evaluar los ataques, cuando tenemos resultados por encima del rango saludable, no podemos determinar si se está en una zona patológica o saludable. Como se ha visto en la Sección 2.5.1, tener resultados superiores a la normalidad no indica patología, sino que muchas veces indica incluso un mejor desempeño del cerebro. Sin embargo, al no poder saber si el rango es saludable o patológico, se va a suponer que cualquier aumento es perjudicial, ya que es la situación más interesante para este estudio.

■ Ataques de parámetros:

- **Ataque al número máximo de neuronas:** Este ataque disminuye la cantidad de ondas Theta, reduciendo el acoplamiento con Gamma y provocando deterioro cognitivo leve y déficits en la memoria de trabajo y la navegación espacial. Su alta excitación Gamma empeoraría el acople con Theta y, la inhibición Ripple, aumentaría el deterioro de la memoria.

- **Ataque a la sinápsis neuronal:** Este ataque tiene casi las mismas repercusiones que el anterior, de forma que los daños cognitivos serían los mismos que al reducir la cantidad máxima de neuronas.
- **Ataque con gCAN:** Este ataque disminuye la cantidad de eventos Theta y Gamma a cambio de aumentar su duración. Esto no tiene que provocar ningún desacople de las ondas, ya que hay la misma cantidad y con una duración extendida proporcionalmente. Sin embargo, provoca un aumento de las características de los Ripples, lo que podría provocar daños epilépticos.
- **Ataque con gACh:** Este ataque es parecido al anterior en las bandas Theta y Gamma, salvo que disminuye el PSD y aumenta la frecuencia. A diferencia del ataque anterior, este aumento de frecuencia puede tener que ver con la disminución de ondas Theta, de forma que una cantidad de estas se acercan al rango de las Gamma. Todo esto puede desembocar en un desajuste de excitación e inhibición, lo que provoca deterioro cognitivo como en enfermedades como el Alzheimer. A todo esto se le suma la disminución de los SWRs, aumentando el daño a la memoria.
- **Ataque con daño estructural:** Las repercusiones son prácticamente las mismas que los ataques N_max y g_max_e. Hay algunas excitaciones que ocurren en uno y en otro no, pero no tiene gran repercusión.
- **Ataque con daño químico:** Las repercusiones son muy parecidas a modificar solo gACh, pero sin aumentar la frecuencia de Theta y Gamma. El daño, por tanto, se debe solo a la disminución de Ripple, provocando daño en la memoria y el aprendizaje.
- **Ataque completo:** Este ataque disminuye la potencia de todas las bandas, pero sus efectos claros se ven solo en los Ripples, que disminuye todo respecto a estos. También hay un desbalanceo entre la cantidad de ondas Theta y Gamma, que sumado a los Ripples puede provocar daños en la memoria como en enfermedades como Alzheimer.

▪ Ataques a archivos EEG:

- **Flooding:** No tiene ninguna repercusión. Esto puede deberse a que el resultado de hacer esta modificación es un escalado simple de toda la señal. Esto hace que el comportamiento sea el mismo pero con diferentes valores.
- **Jamming:** Al igual que el FLO no tiene repercusiones. La explicación sería la misma, que es un escalado simple de toda la señal.
- **Scanning:** Este ataque hace lo que no ha hecho ningún ataque anterior, modificar todas las características de las ondas. Esto ya indica que seguro que

habrá algún tipo de daño. Tiende a disminuir las características de los eventos Theta y Gamma y aumentar los Ripples. Esto perjudica la memoria espacial, desajusta el acoplamiento theta-gamma y perturba los Ripples; lo que se traduce en daño en la memoria espacial y de trabajo, en el deterioro olfativo, en el aprendizaje y puede provocar daños epilépticos.

- **Selective Forwarding:** Es parecido al Scanning pero con repercusiones un poco menores, ya que aumenta la cantidad de Theta en vez de disminuirla, además de no cambiar la frecuencia de Theta. Aún así, las repercusiones serían muy similares al SCA.
- **Nonce:** El Nonce claramente aumenta las características de Gamma y Theta (a excepción de la cantidad de ondas Theta). Esto se puede traducir en hiperactividad y ansiedad, y quizás en un desajuste theta-gamma, dañando también la memoria.
- **Nonce Agresivo:** Este Nonce más agresivo tiene resultados algo diferentes al Nonce normal. Disminuye la cantidad de Theta a cambio de aumentar su duración, frecuencia y potencia, mientras que aumenta la cantidad, duración y frecuencia de Gamma sin cambiar su potencia. Los Ripples, por su parte, son más largos y de menor frecuencia. Esto difícilmente provocaría crisis epilépticas, pero dañaría la memoria debido al desajuste theta-gamma.

En conclusión, podemos ver que los ataques de parámetros tienden a ser muy dañinos con los Ripple, siendo más efectivos a la hora de dañar la memoria de forma más directa, ya que los SWRs son las ondas más directamente correlacionadas con la formación de la memoria a largo plazo. Sin embargo, ataques como el Scanning y el Selective Forwarding provocan un daño más general a todo el espectro, que no solo provoca daños en memorias más a corto plazo como la espacial o la de trabajo, sino que podría causar daños epilépticos debido a una hiper-sincronización de los SWRs. Por otro lado, los ataques Nonce, a pesar de su naturaleza aleatoria tienen consecuencias claras, además de resultados diferentes a todo lo demás, pero que igualmente se pueden traducir en daños en memoria debido a un desequilibrio theta-gamma.

Si analizamos el por qué los ataques SCA, FOR y NON han funcionado y FLO y JAM no, seguramente se deba a los numerosos cambios de voltaje. Es decir, como se ha discutido antes, los ataques FLO y JAM escalan la señal entera por un factor común, lo que hace que las tendencias y cambios de la señal sean exactamente los mismos a la original. Por otro lado, los ataques SCA, FOR y NON van creando regiones con factores de escala diferentes. Esto puede hacer que al cerebro no le de tiempo a adaptarse, ya que cuando está volviendo al estado saludable se realiza otro cambio de potencial abrupto. Esta tendencia de adaptarse a nuevos estados alterados y recuperarse se ha visto en el artículo de López Madejska et al. [26], cuando en una topología realista tras un ataque el cerebro tenía a

volver al estado espontáneo.

7. Conclusiones y Trabajos Futuros

Como se ha visto a lo largo del trabajo, la guerra cognitiva no es un problema del futuro, sino un dilema actual con el cuál hay que tratar. Los NeuroStrikes, además, son ataques posibles dada la tecnología actual, incluso en vías de investigación en algunos países. Es por tanto que la investigación de sus repercusiones es algo importante.

En este trabajo se ha visto un acercamiento a la problemática de los NeuroStrikes desde el punto de vista de las simulaciones cerebrales, intentando replicar los daños cognitivos que tendrían estos en nuestro cerebro. Se empezó replicando los resultados de la tesis de Töllke y, una vez se consiguió, se propusieron nuevos tipos de simulaciones de NeuroStrikes.

Las aportaciones de este trabajo al estudio de los NeuroStrikes en el ser humano son varias. Se empezó analizando la estabilidad de las simulaciones dada topología usada por Töllke, viendo que tenían una alta entropía y causando así resultados altamente diferentes dependiendo de la suerte que se tuviera al realizar las simulaciones. Es por ello que se minimizó el error aumentando la cantidad de simulaciones por experimento, con un total de cuatro, lo que calmó las tendencias erráticas de los resultados. Tras un análisis de los resultados de los ataques de Töllke, se empezó a pensar una nueva forma de simular ataques, proponiendo así la idea de modificar los archivos de entrada al simulador para aparentar daños en la actividad cerebral inicial. Con estos nuevos ataques en mente, se idearon nuevas métricas y gráficas, las cuales pudieran captar los daños de los nuevos ataques y características que no se hubieran percibido de los antiguos, rediseñando de paso las gráficas para mostrar los resultados. Finalmente, debido a la naturaleza aleatoria de un tipo de ataque, se tuvo que idear una forma de medir las repercusiones. Todos estos pasos a lo largo del trabajo han sido aportaciones a la simulación de los NeuroStrikes y sus daños.

Sin embargo, no está todo acabado. Hay características mejorables en esta investigación, las cuales son necesarias destacar para la realización de trabajos futuros. Lo primero a destacar es la forma de medir las ondas Theta y Gamma. Debido a la naturaleza continua de estas ondas, no es posible discretizarlas de forma sencilla para medir cantidad, duración o frecuencia. Es por ello que, debido a que se partía de las funciones de detección de Töllke, se ha utilizado la función de detección de los SWRs para detectar estas ondas (pero adaptada a las bandas de frecuencias Theta y Gamma). Esto quiere decir que para poder detectar estos eventos el potencial medio de la onda a detectar tiene que ser mayor que el V_{rms} de la señal general y, una vez se encuentra esta señal que está por encima del nivel promedio, se mide la frecuencia, catalogando el evento en Theta, Gamma o Ripple. Aunque esto pueda medir los eventos más importantes y significativos no detecta todas

las ondas en su totalidad, por lo que encontrar una forma mejor de medir estas ondas sería un gran avance para el análisis de daños. Además, este estudio ha supuesto que cuando en la evaluación de los ataques se supera el límite máximo del estado saludable, este ha sido un resultado patológico. Sin embargo, esto no tiene por qué ser así, sino que puede ser un resultado saludable, por lo que sería necesario realizar otro tipo de gráficas (como diagramas de caja) o usar otro tipo de métricas que permitan discernir si se está en una situación patológica o no. Otro problema en este trabajo ha sido el alto coste computacional, lo que ha limitado el número de pruebas que se han podido realizar. No se ha podido jugar con las combinaciones de características de los ataques EEG, lo cuál habría dado resultados muy interesantes debido a su alta variabilidad. Además, se han hecho ataques a la totalidad de la señal, cuando hubiera estado bien poder probar qué pasaría al atacar a una fracción de esta.

Por todo esto, quedan pendientes trabajos futuros: implementar mejores funciones de detección de eventos, mejorar e implementar nuevas métricas de daños, probar combinaciones de los nuevos ataques e implementar ataques diferentes aún no vistos, diseñar posibles contramedidas a los ataques y mejorar la topología utilizada.

Referencias

- [1] Brian 2 documentation — brian 2 0.0.post128 documentation. URL: <https://brian2.readthedocs.io/en/stable/>.
- [2] Conda documentation — conda-docs documentation. URL: <https://docs.conda.io/en/latest/>.
- [3] Github - lexutros/ba_bcis-to-mitigate-cognitive-warfare. URL: https://github.com/LexuTros/BA_BCIs-to-Mitigate-Cognitive-Warfare.
- [4] Hippocampus (illustration) | radiology case | radiopaedia.org. URL: <https://radiopaedia.org/cases/hippocampus-illustration>.
- [5] A job neuroscience neuroethics at 15: The current and future environment for neuroethics neuroethics at 15: The current and future environment for neuroethics emerging issues task force, international neuroethics society Å. 2019. URL: <https://www.tandfonline.com/action/journalInformation?journalCode=uabn20>, doi: 10.1080/21507740.2019.1632958.
- [6] Un estudio descubre la estructura de las oscilaciones cerebrales vinculadas con la consolidación de recuerdos. 11 2023. doi:10.1038/s41593-023-01471-9.
- [7] D. G. Amaral and M. P. Witter. The three-dimensional organization of the hippocampal formation: a review of anatomical data. 31:571–591, 1989.
- [8] Alvarado Antepara, Barberán Jiménez, Cassanello Junco, Fabricio Andrés Alvarado Antepara, Fiorella Camila Barberán Jiménez, and Nina Sofia Cassanello Junco. Funciones cognitivas y el papel del hipocampo en la memoria. doi:10.53734/mj.vol5.id273.
- [9] Anton Arkhipov, Nathan W. Gouwens, Yazan N. Billeh, Sergey Gratiy, Ramakrishnan Iyer, Ziqiang Wei, Zihao Xu, Reza Abbasi-Asl, Jim Berg, Michael Buice, Nicholas Cain, Nuno da Costa, Saskia de Vries, Daniel Denman, Severine Durand, David Feng, Tim Jarsky, Jérôme Lecoq, Brian Lee, Lu Li, Stefan Mihalas, Gabriel K. Ocker, Shawn R. Olsen, R. Clay Reid, Gilberto Soler-Llavina, Staci A. Sorensen, Quanxin Wang, Jack Waters, Massimo Scanziani, and Christof Koch. Visual physiology of the layer 4 cortical circuit in silico. *PLOS Computational Biology*, 14:e1006535, 11 2018. URL: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006535>, doi:10.1371/JOURNAL.PCBI.1006535.
- [10] Amélie Aussel, Laure Buhry, Louise Tyvaert, and Radu Ranta. A detailed anatomical and mathematical model of the hippocampal formation for the generation of sharp-wave ripples and theta-nested gamma oscillations. *Journal of Computational Neuroscience*, 45:207–221, 12 2018. doi:10.1007/s10827-018-0704-x.

- [11] Amélie Aussel, Radu Ranta, Olivier Aron, Sophie Colnat-Coulbois, Louise Maillard, and Laure Buhry. Cell to network computational model of the epileptic human hippocampus suggests specific roles of network and channel dysfunctions in the ictal and interictal oscillations. *Journal of Computational Neuroscience*, 50:519–535, 11 2022. URL: <https://link.springer.com/article/10.1007/s10827-022-00829-5>. doi:10.1007/S10827-022-00829-5/METRICS.
- [12] Oliver Backes and Andrew Swab. Cognitive warfare the russian threat to election integrity in the baltic states. 2019. URL: www.belfercenter.org.
- [13] Paola Cristina Bello-Medina, Diego Alexander González-Franco, and Cristina Medina Andrea. El hipocampo: historia, estructura y función. *TEPEXI Boletín Científico de la Escuela Superior Tepeji del Río*, 5(10), jul 2018. URL: <https://repository.uaeh.edu.mx/revistas/index.php/tepxi/article/view/3303>, doi:10.29057/estr.v5i10.3303.
- [14] Sergio Lopez Bernal, Alberto Huertas Celtran, Lorenzo Fernandez Maimo, Michael Taynnan Barros, Sasitharan Balasubramaniam, and Gregorio Martinez Perez. Cyberattacks on miniature brain implants to disrupt spontaneous neural signaling. *IEEE Access*, 8:152204–152222, 2020. doi:10.1109/ACCESS.2020.3017394.
- [15] Sergio López Bernal, Alberto Huertas Celdrán, and Gregorio Martínez Pérez. Neuronal jamming cyberattack over invasive bcis affecting the resolution of tasks requiring visual capabilities. *Computers & Security*, 112:102534, 1 2022. doi:10.1016/J.COSE.2021.102534.
- [16] Sergio López Bernal, Alberto Huertas Celdrán, and Gregorio Martínez Pérez. Eight reasons to prioritize brain-computer interface cybersecurity. *Communications of the ACM*, 66:68–78, 3 2023. doi:10.1145/3535509.
- [17] Sergio López Bernal, Alberto Huertas Celdrán, Gregorio Martínez Pérez, Michael Taynnan Barros, and Sasitharan Balasubramaniam. Security in brain-computer interfaces: State-of-the-art, opportunities, and future challenges. *ACM Computing Surveys*, 54, 1 2021. doi:10.1145/3427376.
- [18] Paul Carrillo-Mora and Manuel Velasco Suárez. Sistemas de memoria: reseña histórica, clasificación y conceptos actuales. primera parte: Historia, taxonomía de la memoria, sistemas de memoria de largo plazo: la memoria semántica. *Salud mental*, 33:85–93, 2010. URL: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-33252010000100010&lng=es&nrm=iso&tlang=es.

- [19] Valérie Ego-Stengel and Matthew A. Wilson. Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat. *Hippocampus*, 20:1, 1 2010. URL: <https://PMC.ncbi.nlm.nih.gov/articles/PMC2801761/>, doi:10.1002/HIPO.20707.
- [20] Beatrice Alexandra Golomb. Diplomats' mystery illness and pulsed radiofrequency/microwave radiation. *Neural Computation*, 30:2882–2985, 11 2018. URL: https://dx.doi.org/10.1162/neco_a_01133, doi:10.1162/NECO_A_01133.
- [21] Ao Guan, Shaoshuang Wang, Ailing Huang, Chenyue Qiu, Yansong Li, Xuying Li, Jinfei Wang, Qiang Wang, and Bin Deng. The role of gamma oscillations in central nervous system diseases: Mechanism and treatment. *Frontiers in Cellular Neuroscience*, 16:962957, 7 2022. doi:10.3389/FNCEL.2022.962957/XML/NLM.
- [22] Juan David Olivares Hernández, Enrique Juárez Aguilar, and Fabio García García. El hipocampo: neurogénesis y aprendizaje hippocampus: neurogenesis and learning. Technical report. URL: www.uv.mx/rm.
- [23] Kenneth A. Koenigshofer. 10.4: Mecanismos sinápticos de aprendizaje y memoria - libretexts español. URL: https://espanol.libretexts.org/Ciencias_Sociales/Psicologia/Biopsicolog%C3%ADa_%280ERI%29_-_PROYECTO_DE_REVISI%C3%93N/10%3A_Aprendizaje_y_memoria/10.04%3A_Mecanismos_sin%C3%A1pticos_de_aprendizaje_y_memoria?utm_source=chatgpt.com.
- [24] Anli A. Liu, Simon Henin, Saman Abbaspoor, Anatol Bragin, Elizabeth A. Buffalo, Jordan S. Farrell, David J. Foster, Loren M. Frank, Tamara Gedankien, Jean Gotman, Jennifer A. Guidera, Kari L. Hoffman, Joshua Jacobs, Michael J. Kahana, Lin Li, Zhenrui Liao, Jack J. Lin, Attila Losonczy, Rafael Malach, Matthijs A. van der Meer, Kathryn McClain, Bruce L. McNaughton, Yitzhak Norman, Andrea Navas-Olive, Liset M. de la Prida, Jon W. Rueckemann, John J. Sakon, Ivan Skelin, Ivan Soltesz, Bernhard P. Staresina, Shennan A. Weiss, Matthew A. Wilson, Kareem A. Zaghloul, Michaël Zugaro, and György Buzsáki. A consensus statement on detection of hippocampal sharp wave ripples and differentiation from other fast oscillations. *Nature Communications* 2022 13:1, 13:1–14, 10 2022. URL: <https://www.nature.com/articles/s41467-022-33536-x>, doi:10.1038/s41467-022-33536-x.
- [25] Christian Ortega Loubon and Julio César Franco. Neurofisiología del aprendizaje y la memoria. plasticidad neuronal. 6, 2010. doi:10.3823/048.
- [26] Victoria Magdalena López Madejska, Sergio López Bernal, Gregorio Martínez Pérez, and Alberto Huertas Celrá. Impact of neural cyberattacks on a realistic neuronal topology from the primary visual cortex of mice. *Wireless Networks*,

30:7391–7405, 12 2024. URL: <https://link.springer.com/article/10.1007/s11276-023-03649-2>, doi:10.1007/S11276-023-03649-2/TABLES/5.

- [27] Robert McCreight, Academic Editor, Gabriel Gutie, and Valerio Napolioni. The war inside your mind: unprotected brain battlefields and neuro-vulnerability. *ACADEMIA BIOLOGY*, 2024, 2024. doi:10.20935/AcadBiol6156.
- [28] Angel Nuñez and Washington Buño. The theta rhythm of the hippocampus: From neuronal and circuit mechanisms to behavior. *Frontiers in Cellular Neuroscience*, 15:649262, 3 2021. URL: www.frontiersin.org, doi:10.3389/FNCEL.2021.649262/XML/NLM.
- [29] Pilar Quijada. El cerebro se sirve de distintos ritmos de ondas lentas y rápidas para adaptarse a las demandas cognitivas. 8 2020. doi:10.7554/eLife.57313.
- [30] Hugo Solís and Estela López-Hernández. Neuroanatomía funcional de la memoria. *Arch Neurocienc (Mex)*, 14:176–187, 2009.
- [31] Jiawei Zhang. Basic neural units of the brain: Neurons, synapses and action potential.
- [32] Lennart Töllke Zurich, Alberto Huertas Celdrán, Chao Feng, Sergio López Bernal, Victoria López Madejska, and Burkhard Stiller. A solution based on brain-computer interfaces to mitigate cognitive warfare. Technical report, 2024.