

# Análisis teóricos de señales EEG por medio de la STFT entre procesos cognitivos

Barriga C., Callata R., Casas A., Narro J.

**Abstract**—El presente trabajo consiste en analizar señales de un electroencefalograma (EEG) de 4 sujetos de prueba, donde se busca estudiar la diferencia que producen entre dos procesos cognitivos específicos. Para el análisis de los datos se emplea el Time/Frecuency Sub-regions of Interest (TFROI) y el Average Time-Frecuency energy within that TFROI (TFROI-AVG). Las tareas cognitivas ha analizar son *Baseline* y *Multiplication* El análisis de los datos se logró mediante el algoritmos desarrollado en Matlab,

**Index Terms**—Short-Time Fourier Transform (STFT), Time-Frecuency (TF), Time/Frecuency Sub-regions of Interest (TFROI), Average Time-Frecuency energy within that TFROI (TFROI-AVG), Ondas cerebrales, Procesos Cognitivos, Overlap.

## I. INTRODUCCIÓN

Se recopiló una data de un encefalograma (EEG), la cual es una prueba que detecta las actividades eléctricas del cerebro mediante electrodos colocados en el cuero cabelludo. El experimento de donde fue extraída la data consta de un grupo de personas que ejercen ciertas tareas, teniendo repeticiones de estas cada cierto periodo.

Para el siguiente análisis, se seleccionarán 4 sujetos de prueba y se hará una comparación entre dos tareas asignadas. Estas son: 'baseline' y 'multiplication'. Para cumplir con dicho estudio se empleará un lenguaje de alto nivel (MATLAB), el cual leerá la data, la segmentará, realizará los cálculos de la *Short-Time Fourier Transform* (STFT) mediante la función 'spectrogram' y se obtendrá la TFROI\_avg para cada segmento.

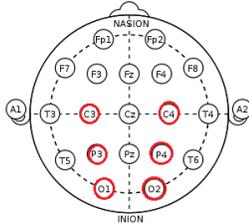


Fig. 1: Posicionamiento de los electrodos según el Sistema 10-20. [1].

## II. OBJETIVOS

### A. Generales

- Implementar un algoritmo en lenguaje de alto nivel (Matlab) para analizar la data a través de una metodología descrita.
- Comprender los conceptos de Short-Time Fourier Transform y espectrograma.

- Calcular la *Short-Time Fourier Transform* (STFT) y el TFROI\_avg de la data segmentada.

### B. Específicos

- Analizar las diferencias entre dos tareas realizadas por los sujetos de interés.
- Presentar los resultados experimentales mediante gráficas de barras de error.
- Explicar los resultados obtenidos y proponer opciones de mejora.

## III. FUNDAMENTOS TEÓRICOS

### A. Short-Time Fourier Transform (STFT)

Para analizar la data obtenida del electroencefalograma (EEG) utilizaremos la *STFT*. La *Short-Time Fourier Transform* es una secuencia de transformadas de Fourier. Además, este analiza de manera más exhaustiva las señales ya que proporciona información de frecuencia localizada en el tiempo para situaciones en las que los componentes de frecuencia de una señal varían en el tiempo [3]. Para el cálculo de la *STFT* de una señal 1 y su inversa 2, las ecuaciones podemos observarlas a continuación:

$$X_{STFT}[m, n] = \sum_{k=0}^{L-1} x[k]g[k-m]e^{-j2\pi nk/L} \quad (1)$$

$$x[k] = \sum_m \sum_n X_{STFT}[m, n]g[k-m]e^{-j2\pi nk/L} \quad (2)$$

### B. Procesamiento de señales basado en segmentación

Para el procesamiento de los datos se empleo la segmentación de ventanas, esta manera de procesar la data nos brinda algunas ventajas. Una de ellas es evitar la perdida de información cuando se segmenta la data al implementar un porcentaje de solapamiento, además de que permite una clasificación de los datos más confiable. La señal se divide en  $x(n)$  en  $L$  segmentos de longitud  $W$  y un overlap de  $O$ , el cual consiste en empezar tomando las últimas  $O$  muestras del segmento anterior. La segmentación esta expresada en (3).

$$x_l(n) = x(n + (l-1)(W - O + 1)) \quad (3)$$

$$n = 0, \dots, W - 1, l = 1, \dots, L$$

Para este trabajo se utilizó un porcentaje de overlap de 30% con respecto a la longitud del segmento ( $W$ ).

### C. Time-Frequency sub-Regions of Interest(TFROI)

Para obtener los valores promedio de las regiones tiempo-frecuencia de interés para una banda gamma (30-100 Hz) se utilizó la expresión:

$$TFROI_{avg} = \frac{1}{(t_{max}-t_{min})(f_{max}-f_{min})} \sum_{t_{min}}^{t_{max}} \sum_{f_{min}}^{f_{max}} |STFT(t, f)|^2 \quad (4)$$

Donde STFT(t,f) representa la Short Time Fourier Transform y la frecuencia máxima y mínima están establecidas por la banda seleccionada.

### D. Banda gamma

La banda gamma ( $\gamma$ ) comprende las señales EEG en un rango de frecuencia de 30-100 Hz. La actividad de esta banda está ampliamente distribuida en todas las estructuras cerebrales, así como en la retina y el tracto olfatorio [4]. Esta banda participa en una variedad de funciones cerebrales, incluyendo la percepción, la memoria, la atención, la conciencia, la plasticidad sináptica, el procesamiento de información y el control motor [4]. Un desafío en la detección de señales EEG de bandas gamma es la eliminación de los artefactos musculares [5]. Esto debido a que ambas señales comparten el mismo rango de frecuencia, por lo que se podrían generar errores en las mediciones.

### E. The Wilcoxon Rank-Sum Test

El test de Wilcoxon Rank-Sum es una alternativa no paramétrica a la prueba de test de hipótesis entre dos muestras, basado solo en el orden en el que las observaciones de las dos muestras caen [6]. Este test se utiliza cuando se tienen dos variables nominales y una variable de medición. Una de las variables nominales presenta sólo dos valores, mientras que la otra variable nominal representa a un grupo de individuos [7]. Esta prueba debe usarse si existe una distribución anormal de las diferencias entre pares.

## IV. DESCRIPCIÓN DEL MÉTODO

### A. Extracción de información

Los datos utilizados para este análisis fueron extraídos de la base de datos 'eggdata.m' [x]. La información está almacenada en una matriz de celdas de una matriz de celdas, donde en cada celda se indica el sujeto, la tarea realizada, el número del intento y la señal obtenida por los siete electrodos. Esta última información está almacenada en una matriz de  $7 \times 2500$ , donde cada fila representa un electrodo y las columnas son las señales muestreadas a una frecuencia de 250 Hz durante 10 segundos. Para acceder a los datos de interés y realizar el análisis respectivo se utilizó el sistema de cómputo numérico Matlab. Para el análisis sólo se consideraron cuatro sujetos de interés (Sujeto 1, 3, 4 y 5) y las señales obtenidas por seis electrodos (c3, c4, p3, p4, o1, o2). Las tareas a comparar fueron 'Baseline' y 'Multiplication'. Todos los sujetos realizaron 10 repeticiones para cada tarea, a excepción del sujeto 5, quien realizó 15 repeticiones. Sin embargo, sólo se tendrán en cuenta las 10 primeras repeticiones para así homogeneizar el procedimiento.

### B. Segmentación de información

Una vez extraída la información de interés, el siguiente paso a realizar es la segmentación de las señales. La señal que obtuvo cada electrodo contiene 2500 muestras, por lo que al utilizar un porcentaje de segmentación de 10% y un porcentaje de overlap de 30%, se obtuvieron 14 segmentos que contenían 250 muestras cada uno. Sin embargo, el último segmento sólo contenía 225 muestras, por lo que se agregaron ceros para completar las 250 muestras restantes y así mantener el procedimiento homogéneo. Para cada canal, los segmentos fueron almacenados en una matriz de  $10 \times 14$ , donde las filas representan las 10 repeticiones y las columnas representan los 14 segmentos.

### C. Obtención de la Short-Time Fourier Transform (STFT)

Para calcular la STFT de cada segmento obtenido, se utilizó el comando spectrogram() de Matlab, con una frecuencia de muestreo igual a 250 Hz. Los valores de las amplitudes del espectrograma, de la frecuencia y del tiempo fueron almacenados en las variables  $s$ ,  $f$  y  $t$ . La Fig. 2 muestra como ejemplo el espectrograma para un segmento.

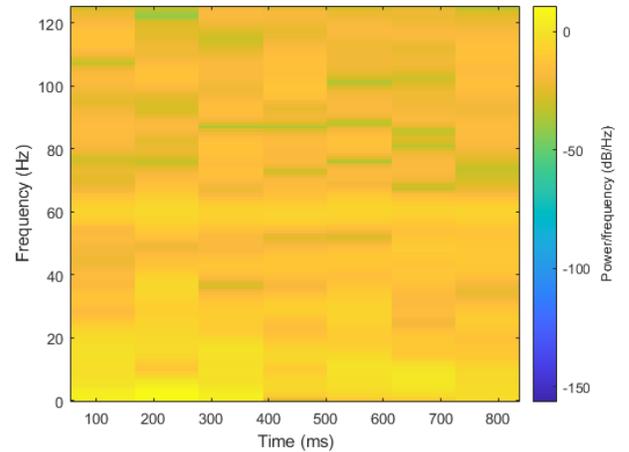


Fig. 2: Espectrograma del primer segmento (sujeto 1, canal 1, Tarea B).

### D. Obtención de Time-Frequency sub-Regions of Interest (TFROI)

Para el cálculo del TFROI\_avg de cada segmento, solo se consideraron los datos presentes en la banda  $\gamma$  (30-100 Hz) de su respectivo espectrograma. Una vez que se identificaron estos elementos, el TFROI\_avg es calculado con la expresión (4). Los valores de  $t_{min}$  y de  $t_{max}$  fueron extraídos de la variable  $t$ , mientras que los valores de  $f_{min}$  y de  $f_{max}$  fueron extraídos de la variable  $f$  una vez que se aplicó el filtro por la banda  $\gamma$ .

### E. Aplicación de The Wilcoxon Rank-Sum Test

Una vez obtenidos y almacenados los TFROI\_avg, se realizó el test de hipótesis two-side Wilcoxon Signed Rank test

utilizando el comando `signrank()` de Matlab. Para ello, se agruparon los TFROI\_avg de uno de los segmentos de cada repetición de un canal de la tarea 'Baseline' en la población A. Igualmente, los TFROI\_avg de la tarea 'Multiplicación' fueron agrupados en la población B. Para este test estadístico, se consideraron dos hipótesis:

- H0: La diferencia media es cero
- H1: La diferencia media es positiva

Los valores de p-value y de H se almacenaron en las variables  $p$  y  $h$  respectivamente. A un nivel de significancia del 5%, si  $h=1$ , significa que existe una diferencia significativa entre ambas poblaciones y se rechaza la hipótesis nula; si  $h=0$ , significa que no existe una diferencia significativa entre las poblaciones y no se rechaza la hipótesis nula. Adicionalmente, para ambas poblaciones se calculó la media aritmética y la desviación estándar utilizando los comandos `mean()` y `std()` respectivamente. Por último, se realizó un gráfico de barras de error para ambas poblaciones y se identificaron con un asterisco a aquellos segmentos que presentaron diferencias significativas ( $h=1$ ). Este mismo proceso se realizó para los 6 canales de los 4 sujetos de interés, obteniendo así 24 gráficas de barras de error.

## V. EVALUACIÓN EXPERIMENTAL

### A. Implementación en Matlab

Para implementar esta operación en Matlab, se hicieron uso de funciones creadas para los sujetos. Primero, se escogía la data a usar y se segmentaba, obteniendo como resultado 14 segmentaciones para cada intento. Donde el último segmento solo contenía 225 muestras. Entonces, para corregir dicho error, se le aumentó una cantidad de ceros para que conserve una dimensión homogénea con las demás. Dicha segmentación se encuentra almacenada en las 10 repeticiones (matrices) y estas a su vez están almacenadas en los 6 canales (celdas).

Luego, para hallar la STFT en cada segmento se usó el siguiente comando:

$$[s, f, t] = \text{spectrogram}(x, [], [], [], Fs, 'yaxis')$$

Donde:

- $x$ : segmento.
- $F_s$ : frecuencia de muestreo.
- 'yaxis': plasma la frecuencia en el eje y y el tiempo en el eje x.

Después, calculamos la TFROI-AVG para cada STFT de los segmentos. Para así, agruparlos en poblaciones, las cuales almacenan los segmentos(n) de las 10 repeticiones obteniendo 14 matrices de 14x1 para cada canal. Calculamos la media aritmética y la desviación estándar de cada matriz y las almacenamos en dos vectores para luego mostrarlos en gráficas.

Finalmente, se hizo uso de *The Wilcoxon Rank Sum-Test* [6] para realizar un test de hipótesis y comparar dos diferentes actividades para cada sujeto de interés. También se hallaron los valores de la media aritmética y de la desviación estándar para ambas poblaciones. Con estos datos se obtuvieron las gráficas de barras de error.

### B. Gráficas obtenidas en Matlab

Se obtuvieron 24 gráficas de la media aritmética y la desviación estándar de los 4 sujetos de estudio. A continuación, se mostrarán las 5 gráficas en donde se puede apreciar mejor la diferencia entre las tareas designadas.

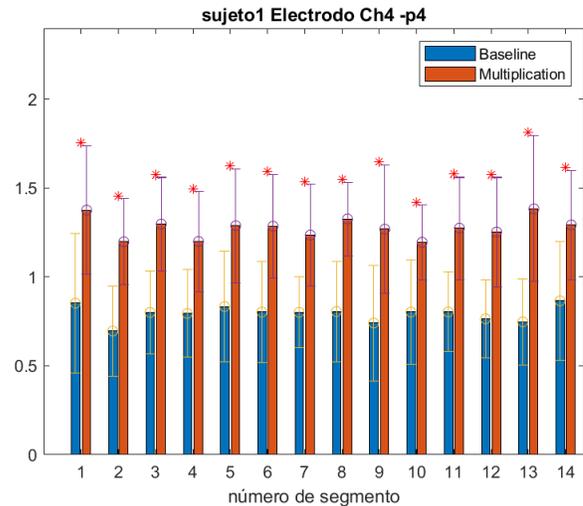


Fig. 3: Barras de error. Sujeto 1, electrodo p4.

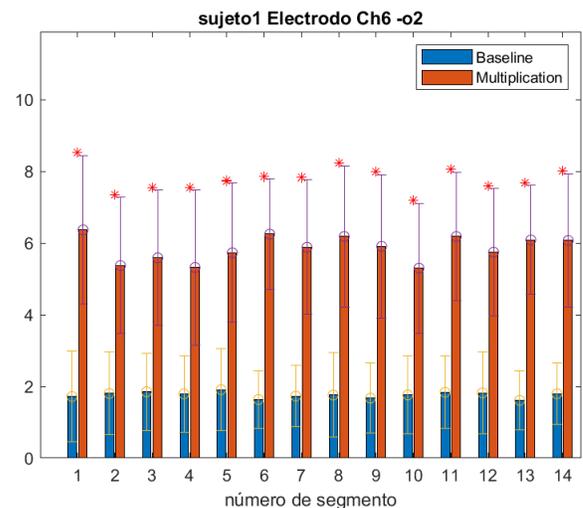


Fig. 4: Barras de error. Sujeto 1, electrodo o2.

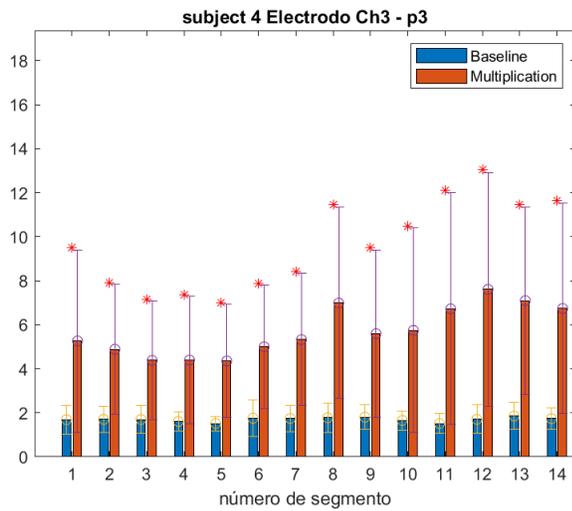


Fig. 5: Barras de error. Sujeto 4, electrodo p3.

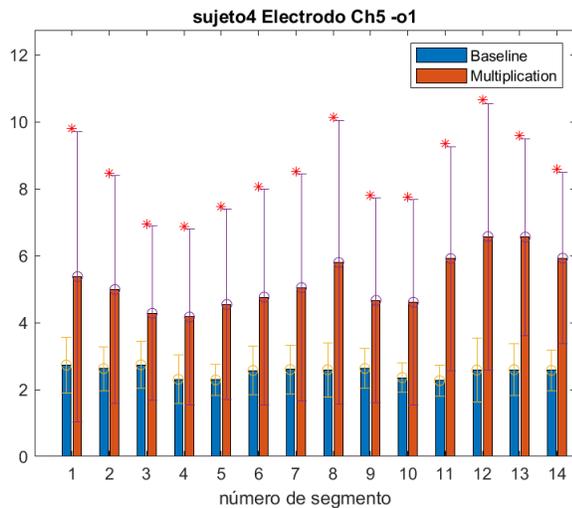


Fig. 6: Barras de error. Sujeto 4, electrodo o1.

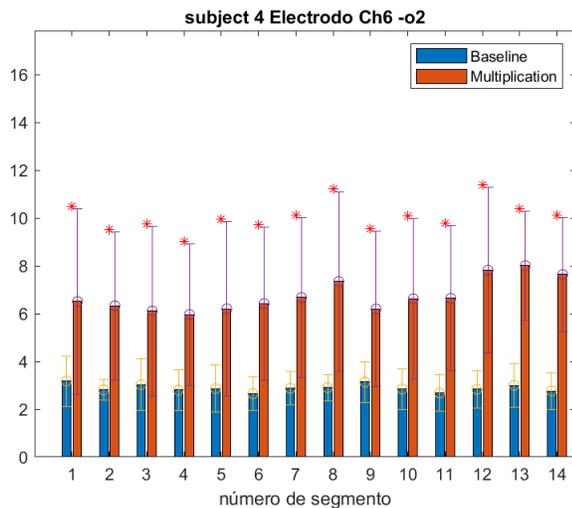


Fig. 7: Barras de error. Sujeto 4, electrodo o2.

## VI. DISCUSIÓN Y CONCLUSIONES

De las 5 figuras mostradas anteriormente, se puede observar que todos los electrodos registran una mayor actividad en la tarea 'Multiplication'. Además, en la mayoría de las 24 gráficas obtenidas se observó que la actividad cerebral de la tarea 'Multiplication' excede a la de 'Baseline'. Sin embargo, el caso del sujeto 3 resulta bastante particular, ya que hay ocasiones en las que la actividad cerebral durante la tarea 'Baseline' es mayor a la de 'Multiplication'. Debido a que esto no sucede con los demás sujetos, es muy probable que el sujeto 3 no haya colaborado completamente con el experimento y no haya realizado correctamente la tarea de 'Multiplication'. Otro caso particular es el del sujeto 5, ya que en gran parte de los casos no se logra observar una diferencia significativa entre la realización de ambas tareas ( $h=0$  en la mayoría de los casos). Por lo tanto, es probable que el sujeto no haya realizado adecuadamente la tarea 'Baseline' o durante el desarrollo de ambas tareas el sujeto realizó la misma actividad. Otro factor a tener en cuenta es la banda gamma (30-100 Hz). Como se mencionó anteriormente, esta banda está relacionada a procesos de memoria y aprendizaje. Por el contrario, la banda alfa (8-13 Hz) está relacionada a estados de reposo del cerebro. Debido a esto, es posible que el uso de la banda gamma no sea el más adecuado para la medición de la tarea de 'Baseline', ya que el sujeto se encuentra en un estado de relajación. Por este motivo, es posible que exista alguna pérdida significativa de información al calcular las regiones de interés (TFROI). Por lo tanto, un opción de mejora sería realizar el cálculo de los TFROI\_avg con la banda correspondiente a la tarea a realizar, para que de esta forma se eviten la pérdida de información. Sin embargo, esto podría ser un método que consuma más recursos computacionales, por lo que se tendría que evaluar si es que la información que se pierde es relevante o insignificante.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] Revista de estudiantes de Terapia Ocupacional. (2016). *Posicionamiento de Electrodo 10-20*. Extraído de: <http://revistacellphone.blogspot.com/2016/11/como-funciona-el-auricular.html>
- [2] Base de datos de señales EEG. Disponible en: [https://www.cs.colostate.edu/eeg/main/data/1989\\_Keirn\\_and\\_Aunon](https://www.cs.colostate.edu/eeg/main/data/1989_Keirn_and_Aunon) [Accedido: nov-2020]
- [3] N. Kehtarnavaz, *Digital Signal Processing System Level Design Using LabVIEW*, 2008.
- [4] C. Amo et al. *Induced gamma band activity from EEG as a possible index of training-related brain plasticity in motor tasks*, April 2017. Recuperado de: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5469342/>.
- [5] G. Shou et al. *Resting-state Gamma-band EEG Abnormalities in Autism*, October 2018. Recuperado de: <https://ieeexplore.ieee.org/document/8512718>
- [6] C. Wild, *The Wilcoxon Rank Sum-Test*.
- [7] J. McDonald, *Wilcoxon Signed-Rank Test*. Recuperado de: <http://www.biostathandbook.com/wilcoxonsignedrank.html>.
- [8] K. Jones et al. *S-transform time-frequency analysis of P300 reveals deficits in individuals diagnosed with alcoholism*, August 2006