

# **Detección del habla imaginada**

## **Reporte Final**

### **Alumnos**

Juan Daniel Rubio Camacho A01633924

Víctor Javier Aguayo Mendoza A01229260

### **Mentora**

Denise Magdalena Alonso Vázquez A00835272

## Introducción

El presente trabajo nació para cubrir la necesidad que es la incapacidad de comunicación de diversas personas con enfermedades o en un estado que les es físicamente imposible comunicarse mediante el habla, tal como lo es una de las motivaciones de este proyecto, la enfermedad de la esclerosis lateral amiotrófica (Figura 1), haciendo uso de la detección de ondas cerebrales en personas y la inteligencia artificial para poder tratar de traducir los pensamientos de una personas en palabras reales que podamos utilizar para apoyar a esas personas.



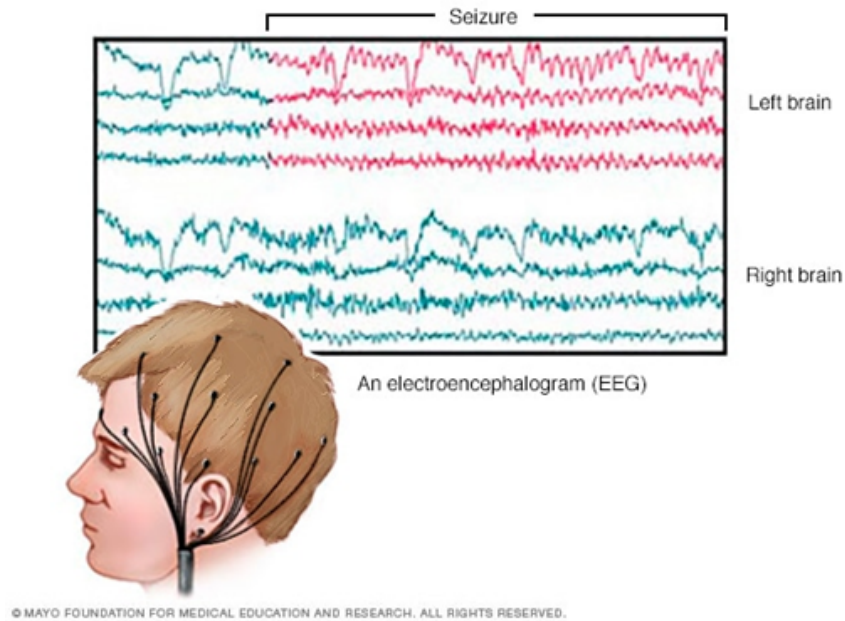
*Figura 1*

Las señales electroencefalográficas son un recurso utilizado para el análisis y estudio de diversas enfermedades, pero en años anteriores una nueva área se ha empezado a investigar y desarrollar en el ámbito enfocado a los pensamientos, para tratar de distinguir los procesos internos con los que contamos y previamente no se habían intentado predecir. Las nuevas tecnologías enfocadas en el ámbito de aprendizaje mediante computadoras han permitido un nuevo avance en esta área, al ya existir la capacidad de conseguir la información de estas ondas electroencefalográficas y traducirlas a datos manejables por computadora.

Por otro lado, la tecnología conocida como inteligencia artificial vio un crecimiento exponencial en los últimos años al desarrollar nuevos métodos para mejorar y al contar con un mayor nivel de procesamiento proveniente de las mejoras en procesadores y capacidad de almacenamiento. Es con esta tecnología que podemos entrenar a modelos de aprendizaje a reconocer patrones en diversas áreas utilizando información ya clasificada previamente, de manera que “enseñemos” a nuestro modelo a clasificar nuevas entradas de información con base a lo que ya le hemos enseñado, de manera que podamos predecir a qué clase pertenece esa nueva entrada. Idealmente esta información suele pasar por procesos previos que facilitan la lectura de estos datos y nos arroja mejores resultados en los porcentajes de predicción, sumado a esto existen parámetros que nos permiten modificar el comportamiento de nuestro modelo y que podrían mejorar o empeorar nuestro nivel de aciertos, de manera que también se debe realizar un estudio previo para comprobar qué opción se ajusta de mejor manera a la clase de información que estás utilizando.

Es de esta manera que planeamos utilizar estas dos herramientas para poder crear un sistema que nos permita recolectar la información proveniente de señales cerebrales de una persona para poder alimentar a nuestro modelo de clasificación y enseñarle la información requerida para poder detectar estos pensamientos imaginados, de manera que eventualmente pudiéramos desarrollar un mecanismo que lea las señales cerebrales y traduzca en tiempo real su significado.

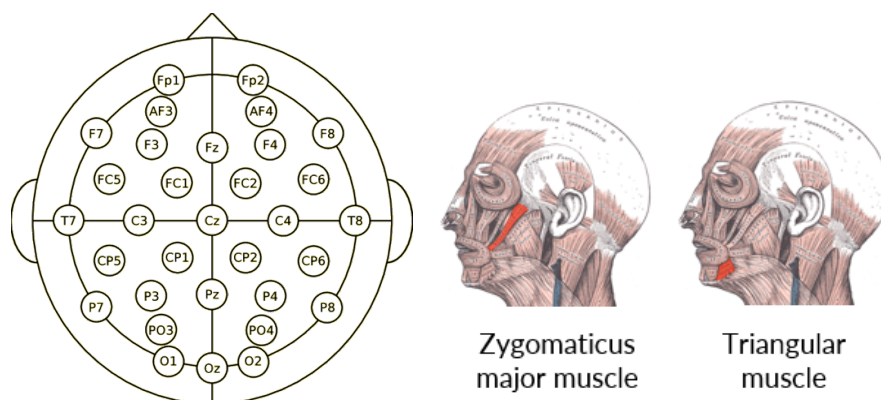
Con este proyecto apuntamos a poder interpretar las señales electroencefalográficas asociadas con acciones (Figura 2) y elementos propios de un vocabulario limitado, sin hacer uso del movimiento de articulaciones o músculo auditivos que provocarán algún sonido por parte de la personas (habla imaginada / interpretada). Las palabras elegidas para este experimento están dentro del vocabulario español y son las siguientes: “Sí”, “No”, “Agua”, “Comida”, “Dormir”. Para poder conseguir la información necesaria se realizaron múltiples experimentos con sujetos de prueba diferente de manera que con siguiéramos datos provenientes de más de una persona, con lo que nuestra información sería más amplia y diversa, junto a esto, es mencionable el hecho de que se siguió una serie de protocolos enfocados en separar posibles distracciones que pudieran surgir en el momento de la experimentación, ya que debido a la naturaleza de las ondas cerebrales, pequeños detalles tales como picazón en el sujeto de pruebas, un sonido fuera de lo común, o pensamientos sin relación al experimento pueden dañar la información que se está consiguiendo.



*Figura 2*

## Experimentos

Como se menciona más arriba, se realizaron experimentos en múltiples sujetos donde se realizarían tres procesos diferentes con el propósito de analizar y comparar cada uno de ellos, colocando electrodos en diferentes áreas de la cabeza para detectar las ondas necesarias con este dispositivo (Figura 3 y Figura 3.5), estos tres procesos son: habla imaginada, habla pronunciada y habla normal.



*Figura 3 y Figura 3.5*

El habla normal, como su nombre lo dice, se refiere simplemente a nuestros sujetos de prueba pronunciando de manera clara y en un tono de voz normal las palabras escogidas a la vez que se grababan las ondas cerebrales que se registraban en ese momento.

El habla gesticulada, por otro lado, es el proceso por el cual la persona trataría de decir la palabra sugerida por medio de la gesticulación, pero sin producir sonido alguno, con los movimientos de labios necesarios y cualquier otro músculo involucrado en este proceso.

El habla imaginada se refiere al proceso de pensar en una cierta palabra sin realizar ninguna pronunciación, movimiento o gesticulación que pudiera afectar a las ondas cerebrales en ese momento, haciendo uso igualmente de marcadores que permitieran a las personas descansos entre cada marca para tener la menor cantidad de ruido posible en nuestras tomas.

Dentro de estos tres experimentos se estableció un mismo patrón a seguir de manera que tuviéramos ciertas condiciones en las personas que nos ayudaron en estos experimentos, este patrón era el de dar un leve período de atención a la persona, de manera que estos pudieran calmar sus pensamientos y acciones antes del momento donde se realizar la pronunciación de la palabra, todo mediante una pantalla donde se podrían observar estas señales. Tras ese breve período de atención, aparecería en pantalla la palabra escogida en ese momento (cabe recordar que las palabras que se utilizaron para este experimento son “Sí”, “No”, “Agua”, “Comida” y “Dormir”), momento en el cual se haría la pronunciación de la palabra según el tipo de experimento llevado a cabo en ese momento, tras esto, habría un breve período de tiempo donde la persona podría descansar y hacer movimientos sin problema alguno.

Este ciclo se repetiría un cierto número de veces para almacenar todas las palabras con sus respectivas ondas registradas aproximadamente siete veces por palabra, para posteriormente almacenar toda la información registrada de las ondas y poder trabajar con ellas.

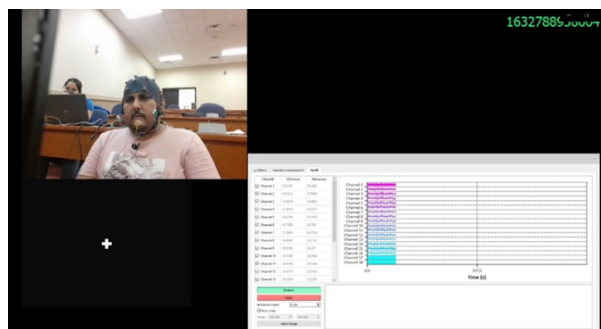
Estos experimentos se llevaron a cabo con cautela debido a factores como lo fueron el aprendizaje de este nuevo equipo como lo fueron los electrodos y la aplicación del gel necesario para detectar estas ondas, la distribución de tiempos generada para crear mejores resultados al momento de realizar cualquiera de los diferentes tres tipos de habla y usando frecuencias que nos permitieran aprovechar al máximo estos datos.

Las siguientes imágenes muestran el proceso de experimentación llevado a cabo en estos experimentos, realizados en un pequeño auditorio donde las interrupciones deberían ser mínimas y no existen distracciones que pudieran tomar la atención de la persona sobre la cual se está realizando el experimento.

De los primeros pasos se tenía que realizar la colocación de los electrodos en el sujeto de prueba, además de realizar la explicación del experimento, tras lo cual iniciaría la prueba donde se realizaría uno de los tres tipos de habla según se indicara, tras lo cual iniciaría el proceso (Figura 4 y Figura 5).

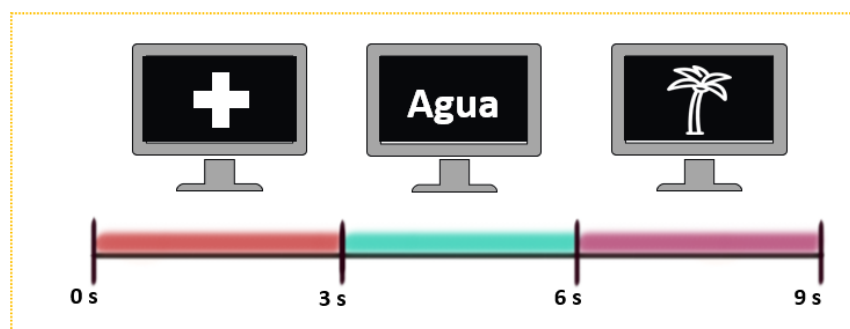


*Figura 4*



*Figura 5*

El procedimiento a llevarse a cabo durante el experimento consta de tres fases que se repiten 40 veces, 8 veces por cada palabra, donde se utilizan 3 segundos para llamar la atención del sujeto, 3 segundos para pronunciar la palabra según el experimento y 3 segundos de descanso, tras lo cual se repetiría el proceso con las demás palabras (Figura 6).

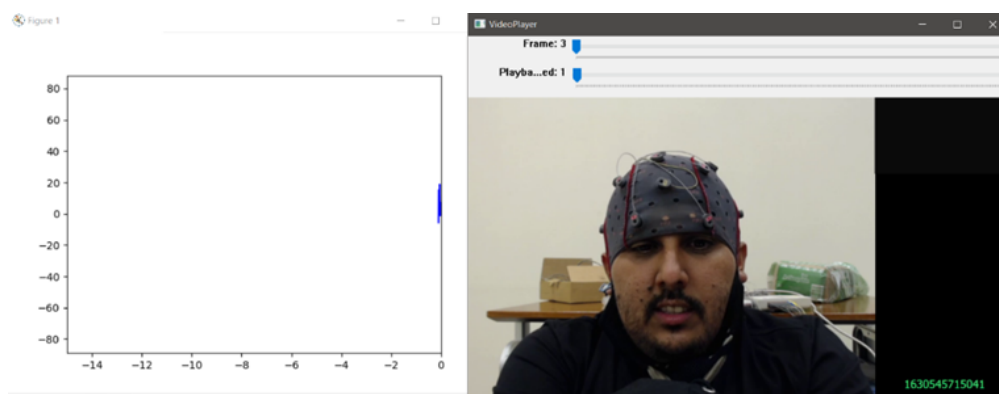


*Figura 6*

## Análisis

Una vez realizados los experimentos, se buscó analizar estas ondas primeramente en un simple ámbito como lo fueron gráficas para comparar los momentos en donde había actividad cerebral mayor al pronunciar la palabra deseada y en donde estábamos en los momentos de descanso y atención.

Nuestro primer acercamiento a este análisis fue mediante la graficación de un canal de EEG disponible a la vez que se reproducía el video (Figura 7), de manera que pudiéramos comparar en tiempo real los nuevos picos detectados al tiempo que vemos cómo el sujeto realizar el experimento, con el objetivo de comprobar que sí existen picos de actividad cerebral al pronunciar la palabra.



*Figura 7*

En esta imagen podemos comprobar como sí existen momentos en donde parecen existir ondas cerebrales mayores (Figura 8), de manera que sí podríamos trabajar con esta información para encontrar los momentos precisos en donde pasan estas palabras y cómo se observa la información obtenida en ese momento, de manera que pudiéramos alimentar esa información a nuestro clasificador,

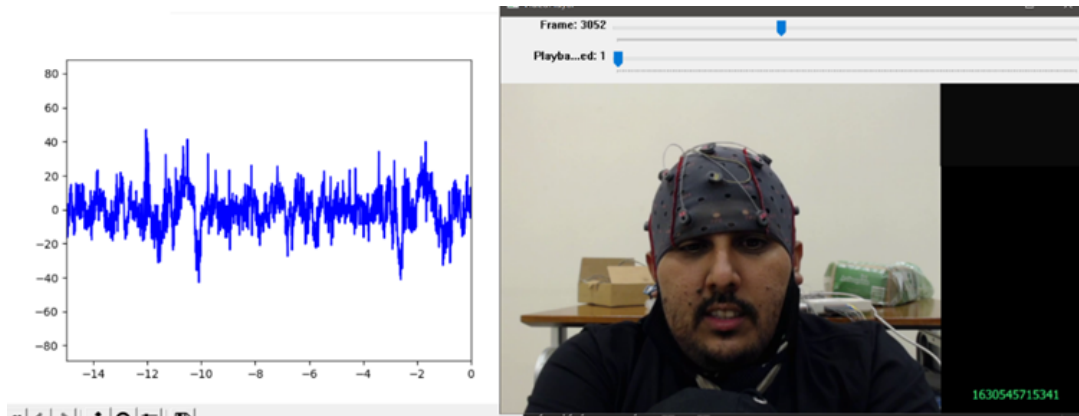
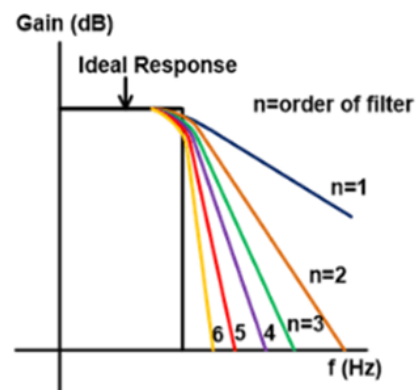
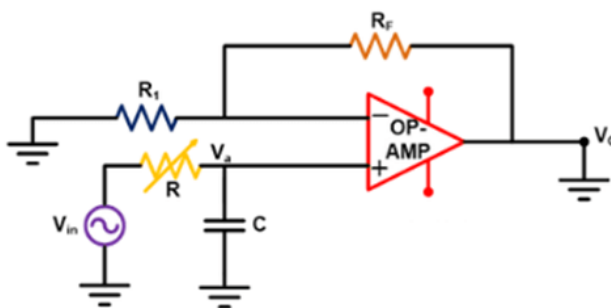


Figura 8

Una vez comprobada esta información, el siguiente paso sería manipular nuestro archivo que contiene estos datos pasándolo por un filtro (Figura 9) especializado en aprovechar estos picos de actividad para hacerlos aún más claros.

## What is a Butterworth Filter?



**Electrical 4 U**

Figura 9

El filtro elegido fue uno de Butterworth, el cuál nos ayudará a disminuir el ruido que tenemos en las grabaciones de manera que podamos eliminar parte de los datos que no son relevantes para nuestro experimento, esto fue realizado con Python y la librería de signal (Figura 10), la cual nos permitió aplicar este filtro a la información con escogiendo la frecuencia en Hz que debería tener nuestros datos ya filtrados, la frecuencia de muestreo, el pase de banda, la frecuencia actual de los datos y finalmente el tipo de filtro que deseamos aplicar, como se muestra en esta imagen:



```

column_names = ["Tiempo", "Canal 1", "Canal 2", "Canal 3", "Canal 4", "Canal 5", "Canal 6",
                "Canal 7", "Canal 8", "Canal 9", "Canal 10", "Canal 11", "Canal 12",
                "Canal 13", "Canal 14", "Canal 15", "Canal 16", "EMG 1",
                "EMG 2", "Timestamp", "Marca"]

csv_data = pandas.read_csv('./9no Semestre/Estancias/Programas/files/Sujeto_1/13-09-21/S1_habla_imaginada_commas.csv', names=column_names)

#filter for EMG 4-200 Hz EMG
b, a = signal.iirfilter(4, [4, 200], fs=1200, rs=60, btype="bandpass", ftype='butter')

#filter for EEG 1-40 Hz Canales
d, c = signal.iirfilter(4, [1, 40], fs=1200, rs=60, btype="bandpass", ftype='butter')

```

Figura 10

Como también podemos notar, tenemos dos filtros diferentes que aplicar para nuestros diferentes canales de EEG y EMG, con la principal variación que es la frecuencia a la cual queremos convertir nuestros datos (4-200 para los canales de EMG y 1-40 para los canales de EEG).

Después de aplicar este filtro a la información que tenemos, nos dedicamos a graficar la información para conocer si realmente había funcionado nuestro filtro, y podemos notar cómo existen picos en las muestras los cuáles corresponden a los momentos en los que se pronunció o pensó una palabra en nuestro experimento (Figura 11).

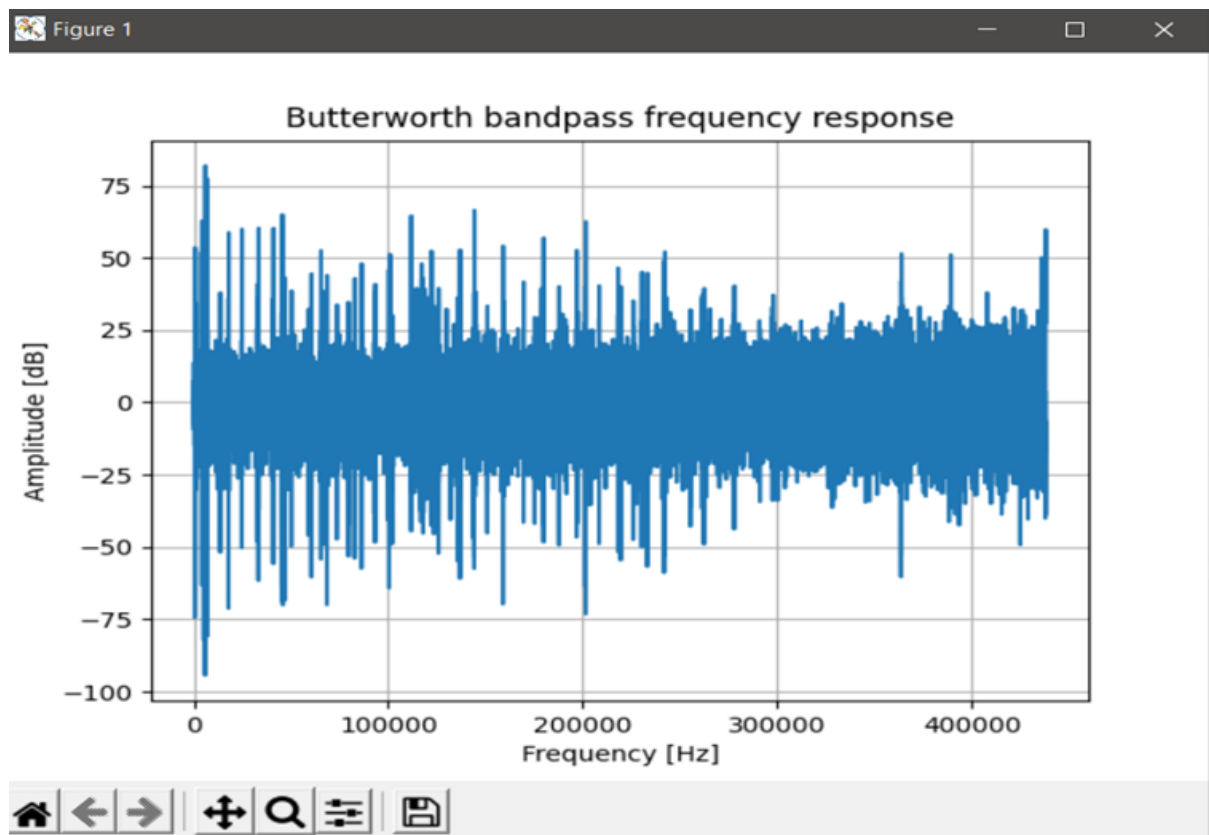


Figura 11

Una vez obtenidos estos datos, decidimos almacenarlos en un archivo formateado por comas para poder realizar la parte del entrenamiento de un modelo para tratar de predecir las palabras correspondientes a los períodos de tiempo donde se realizó la pronunciación o imaginación de una palabra en el experimento.

Para poder realizar nuestro modelo predictor, necesitamos de una primera etapa conocida como entrenamiento, en la cual proveeremos de información a nuestro modelo de aprendizaje de manera que tenga conocimiento sobre cuáles son las clases que tiene que identificar y qué valores son usuales en ella.

Para esto, nos dedicamos a realizar “ventanas” (Figura 12) de nuestra información en las cuales existían los datos registrados en las ondas sobre los momentos de la experimentación, es decir, se podría decir que tomamos los momentos de la grabación donde el experimentado hacía pronunciación de la palabra, de manera que podamos etiquetar ese período de tiempo y esos datos con un identificador que pertenezca a la palabra que se dijo, ya que al ser estos datos conocidos podemos usar toda esta información para entrenar a nuestro modelo, al terminar con un formato de: INFORMACIÓN --> (pertenece) a 4 (4 es la palabra Agua). Es así que nuestro clasificador termina con una base sobre la cual puede determinar cómo se vería un set de información que podría o no ser la palabra “Agua”.

## Sliding window vs. forward chaining cross validation

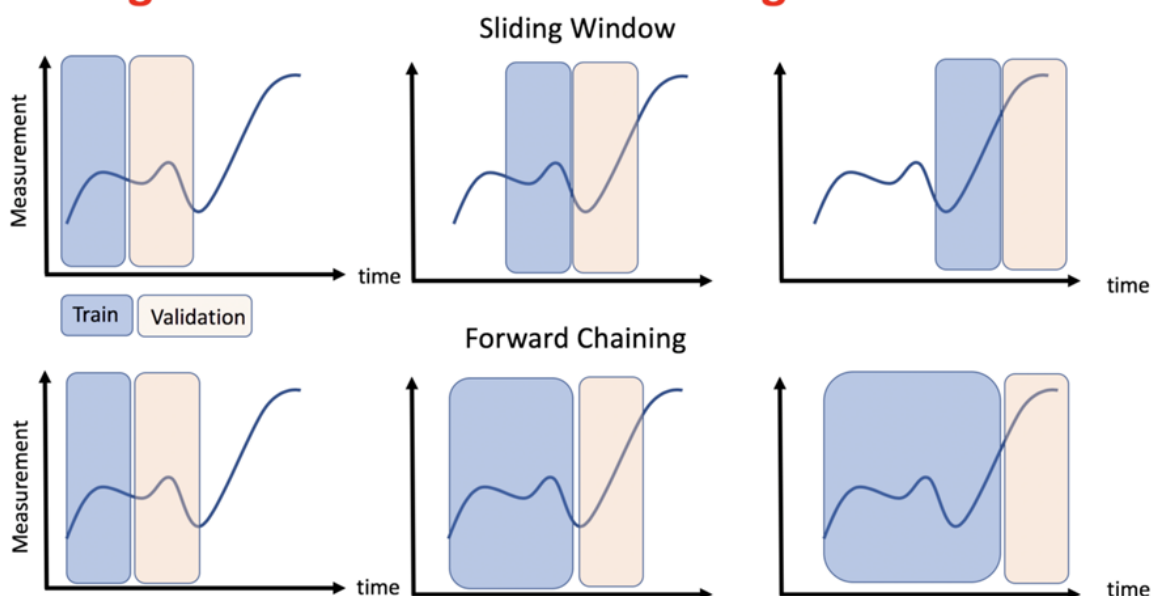


Figura 12

Este ventaneo se realizó para cada uno de los archivos que generamos, sobre el habla pronunciada, habla normal y habla imaginada, de manera que el procesamiento duró algo de tiempo, ya que cada experimento para cada palabra se realizaba la pronunciación la cantidad de cinco veces, de manera que terminamos con más de 500 ventanas en total las cuáles se pasan como entrenamiento al modelo para predecir nuevos datos, la manera en que esto funciona es tomando una parte de esos datos como información de validación y la otra parte como información de entrenamiento, de manera que comparamos una parte de los datos contra la otra conociendo qué es cada parte y así podemos determinar un porcentaje de precisión para nuestro clasificador.

## Resultados

Para la clasificación de datos se utilizaron tres modelos diferentes, KNN Neighbours, lineal y RB, entrenando cada modelo para los tres tipos de habla que tenemos, de manera que terminamos con tres modelos para habla imaginada, habla gesticulada y habla pronunciada, lo cual corresponde a los doce archivos recibidos de los experimentos realizados.

Para cada uno realizamos una validación cruzada de K Folds utilizando 5 splits para determinar la precisión del modelo, además de utilizar una función implementada de sklearn para obtener nuestra matriz de confusión (Figura 13).

25	0	0	0	0
0	21	0	0	0
3	0	20	0	0
0	0	0	15	0
0	3	0	0	19

*Figura 13*

Además de obtener esos datos, también decidimos que para un mejor análisis de nuestros resultados obtendremos los datos individuales de recall para cada clase, el cual se usa como una medida de para cuantificar , además de la puntuación F1 para una mejora.

Las siguientes tablas muestran detalladamente el accuracy que obtuvimos para cada tipo de habla en nuestra primer ronda de resultados (Figura 14 y Figura 15), se obtuvieron los mejores resultados para ambos tipos de señales EMG y EEG en el clasificador KNN Neighbours, mientras que de manera individual podemos observar un claro porcentaje de acierto muy grande para el clasificador lineal SOLO en el caso de las ondas EEG.

EMG	Linear	RBF	K-NN	EEG	Linear	RBF	K-NN
Habla pronunciada	64.36%	16.5%	97.03%	Habla pronunciada	99.09%	19.84%	91.61%
Habla gesticulada	64.58%	15.90%	95.56%	Habla gesticulada	99.09%	19.09%	91.08%
Habla imaginada	64.74%	14.81%	94.69%	Habla imaginada	99.39%	18.53%	91.61%

*Figura 14 y 15*

El siguiente paso que tomamos fue el de realizar una selección de características de manera que pudiéramos observar si existía alguna mejora clara, lo cual sí fue el caso, como podemos observar en las siguientes tablas (Figura 16 y 17) que nos vuelven a mostrar los porcentajes de precisión de nuestros modelos, solo que esta vez aplicando esa selección.

EMG	Linear	RBF	K-NN	EEG	Linear	RBF	K-NN
Habla pronunciada	98.23%	20.25%	95.14%	Habla pronunciada	99.69%	19.09%	93.34%
Habla gesticulada	98.05%	21.04%	94.43%	Habla gesticulada	99.70%	20.04%	91.54%
Habla imaginada	98.11%	20.03%	95.26%	Habla imaginada	99.84%	19.28%	93.07%

*Figura 16 y 17*

Los siguientes datos que quisimos obtener para tener un mejor análisis del comportamiento y qué tan precisos son los modelos que realizamos, fueron como ya mencionamos arriba la exhaustividad correspondiente a cada clase/palabra (Figura 18 y 19) además de obtener la precisión por clase (Figura 20 y 21) y la puntuación F1 (Figura 22 y 23).

El color azul en las tablas indica la superioridad en resultados sobre el otro clasificador, de esta manera podemos ver de manera visual los resultados y determinar el mejor clasificador

para cada una de nuestras palabras. El color dorado indica la perfección en el valor f1, lo cual nos indica que el valor de precisión y exhaustividad fueron también perfectas.

EMG	Linear					K-NN				
Palabras	Si	Dormir	No	Agua	Comida	Si	Dormir	No	Agua	Comida
Habla pronunciada	100%	96.24%	96.52%	95.55%	96.52%	96.84%	97.17%	100%	93.93%	94.37%
Habla gesticulada	100%	96.27%	100%	95.80%	100%	97.77%	95.12%	97%	97.69%	100%
Habla imaginada	100%	96.41%	100%	96.06%	95.83%	89.22%	86.38%	94.61%	92.34%	94.78%

*Figura 18*

EEG	Linear					K-NN				
Palabras	Si	Dormir	No	Agua	Comida	Si	Dormir	No	Agua	Comida
Habla pronunciada	100%	100%	100%	97.57%	100%	97.39%	92.23%	97.93%	90.60%	92.97%
Habla gesticulada	100%	96.66%	100%	96.92%	100%	96.66%	88.86%	97.77%	83.77%	94.69%
Habla imaginada	100%	100%	100%	100%	100%	94.54%	95.27%	96.84%	82.73%	95.38%

*Figura 19*

EMG	Linear					K-NN				
Palabras	Si	Dormir	No	Agua	Comida	Si	Dormir	No	Agua	Comida
Habla pronunciada	96.36%	94.02%	100%	96.02%	100%	91.60%	100%	100%	94.41%	97.27%
Habla gesticulada	100%	96.01%	100%	96.54%	100%	97.27%	100%	94%	96.84%	97.85%
Habla imaginada	98.94%	96.34%	100%	93.26%	100%	91.53%	89.49%	94.44%	91.97%	91.37%

*Figura 20*

EEG	Linear					K-NN				
Palabras	Si	Dormir	No	Agua	Comida	Si	Dormir	No	Agua	Comida
Habla pronunciada	100%	100%	96.52%	100%	100%	97.39%	84.60%	97.93%	94.35%	93.79%
Habla gesticulada	100%	100%	93.04%	100%	100%	82.27%	97.27%	97.6%	93.64%	94.81%
Habla imaginada	100%	100%	100%	100%	100%	91.99%	86.56%	91.68%	96.25%	97.39%

*Figura 21*

EMG	Linear					K-NN				
Palabras	Si	Dormir	No	Agua	Comida	Si	Dormir	No	Agua	Comida
Habla pronunciada	98%	94.98%	98.09%	95.60%	98.09%	94.01%	98.46%	100%	93.71%	95.75%
Habla gesticulada	100%	96.09%	100%	96.12%	100%	97.36%	97.40%	94.84%	97.06%	98.86%
Habla imaginada	99.45%	96.35%	100%	94.59%	97.67%	90.07%	87.03%	94.11%	91.17%	92.32%

*Figura 22*

EEG	Linear					K-NN				
Palabras	Si	Dormir	No	Agua	Comida	Si	Dormir	No	Agua	Comida
Habla pronunciada	100%	100%	98.09%	98.70%	100%	97.39%	92.23%	97.93%	90.60%	92.97%
Habla gesticulada	100%	98.18%	95.78%	98.33%	100%	88.87%	92.73%	97.54%	87.87%	94.17%
Habla imaginada	100%	100%	100%	100%	100%	92.36%	90.25%	93.64%	88.45%	96.32%

*Figura 23*

Todos estos parámetros de medida y evaluación del modelo tomarán más sentido e importancia conforme se amplíe el alcance y las palabras prueba del proyecto, pero nos pareció muy interesante mostrar los resultados para dimensionar la capacidad de ajuste que aun podemos tener dentro del clasificador.

## Conclusiones

Con base a los resultados obtenidos en nuestro modelo de clasificación, nos podemos dar cuenta que la precisión de la mayoría de modelos fue bastante alta, en especial al momento de aplicar la selección de características para aumentar los porcentajes en cada uno de los modelos.

En cuanto a resultados específicos en nuestros modelos, nos dimos cuenta que sin aplicar la selección de características, el KNN Neighbors resultó ser el mejor modelo (o al menos, el que arrojó una mejor precisión) de entre los tres que probamos, superando al lineal y al rb. Pero en el momento que aplicamos esta última parte del proyecto para ver cómo cambiaban nuestros porcentajes, nos dimos cuenta que filtrar nuestros datos utilizando una misma función de sklearn podíamos mejorar todos los resultados, con un gran énfasis en los modelos lineales que utilizaban canales de EMG.

También podemos afirmar que las señales que parecen más aptas para este proyecto son las EEG, ya que mostraron un porcentaje de exactitud prácticamente perfecto.

Después de realizar todo este proceso, al equipo le pareció muy buena metodología para probar nuestros clasificadores, y creemos que antes de seleccionar uno deberíamos analizar más palabras, algunas más complejas para comparar los resultados y poder hacer una decisión precisa en cuanto a la elección del clasificador se refiere.

Es así que podemos concluir que la detección del habla imaginada es posible si realizamos experimentos con condiciones claras, parámetros establecidos previamente y mantenemos un orden y estabilidad en la duración del experimento, ya que de esta manera los datos obtenidos pueden ser fácilmente manipulados para luego ser alimentados a un modelo de clasificación y predecir nuevas entradas, como se probó con los datos recolectados a lo largo del semestre.

## Referencias

A.A. Torres-García, C.A. Reyes-García, L. Villaseñor-Pineda & J.M. Ramírez-Cortés. (Abril 2013). Análisis de Señales Electroencefalográficas para la Clasificación de Habla Imaginada. Revista Mexicana de Ingeniería Biomédica, 34, 23-39. Noviembre 2, 2021, De Scielo Base de datos.

Mashael M. AISaleh\* Mahnaz Arvaneh\* Heidi Christensen\* and Roger K. Moore\*. (2016). Brain-Computer Interface Technology for Speech Recognition: A Review. University of Sheffield, 1, 5. Noviembre 5, 2021, De IEEEExplore Base de datos.