

Reconocimiento de habla interna usando señales EEG: Estado del arte y metodología propuesta

Luciana Espindola-Ulibarri, Olga Kolesnikova,
Rolando Menchaca-Méndez

Instituto Politécnico Nacional,
Centro de Investigación en Computación,
México

{lespindolau2024, kolesnikova, rmen}@cic.ipn.mx

Resumen. En este trabajo se abordarán las principales características del habla interna que mostrarán un panorama general para comprender los elementos que la conforman, aunado a la exposición de una técnica llamada electroencefalografía (EEG), apropiada para el desciframiento y reconocimiento del habla interna debido a que posee rasgos convenientes para esta tarea, como lo es su alta resolución temporal y su practicidad al ser un método no invasivo. Además, se propone una metodología para analizar, procesar y reconocer el habla interna a partir de señales EEG usando dos conjuntos de datos basados en un vocabulario reducido de palabras que representan comandos.

Palabras Clave: Habla interna, electroencefalografía (EEG) y alta resolución temporal.

Inner Speech Recognition Using EEG Signals: State of the Art and Proposed Methodology

Abstract. In this work we will address the main characteristics of internal speech that will show a general overview to understand the elements that make it up, together with the exposure of a technique called electroencephalography (EEG), appropriate for the decipherment and recognition of internal speech because it has convenient features for this task, such as its high temporal resolution and its practicality as it is a non-invasive method. In addition, a methodology is proposed to analyze, process and recognize internal speech from EEG signals using two data sets based on a reduced vocabulary of words that represent commands.

Keywords: Inner speech, electroencephalography (EEG), high temporal resolution.

1. Introducción

Actualmente se presenta el problema de llegar a descifrar el habla interna mediante técnicas que consisten en registrar la actividad neuronal, como lo es la electroencefalografía (EEG, por sus siglas en inglés) esto con el principal fin de crear mejores medios de comunicación para quienes padecen enfermedades como la esclerosis lateral amiotrófica (ELA), la embolia (ictus cerebral), las lesiones de médula espinal o cerebral, la parálisis cerebral, las distrofias musculares, la esclerosis múltiple, entre otros padecimientos. Por esta razón es de gran importancia conocer lo que caracteriza al habla interna, así como sus componentes y propiedades principales, para que partiendo de esta información sea posible tomar decisiones informadas acerca de los métodos y técnicas más pertinentes para su predicción. Destacando que el procesamiento, desciframiento y decodificación del habla interna puede ser la llave para introducir una nueva era de comunicación humana pasando de la comunicación común por voz o basada en texto a la comunicación basada en el cerebro [1].

Se ha demostrado que las señales EEG son apropiadas para descifrar el habla interna gracias a que tienen una alta resolución temporal, lo que les permite captar los cambios en las señales sobre el tiempo a nivel de milisegundos y producir una imagen de la actividad eléctrica en el cerebro, la cual es representada como ondas con variación de frecuencia, amplitud y forma basada en parámetros, además de ser un método no invasivo, que no daña el cerebro y que no requiere cirugía [2-5].

Para poder llegar a comprender los principales componentes y características conocidos hasta ahora del habla interna este trabajo está estructurado de la siguiente manera: La sección II introduce los antecedentes del habla interna, en la sección III se establecen un conjunto de definiciones del habla interna, la sección IV expone la relación entre el habla interna y el habla externa, en la sección V se abordan los principales componentes del habla interna, la sección VI presenta el habla interna en el cerebro y las principales zonas involucradas, en la sección VII se describen las ventajas y desventajas de la técnica de EEG para el problema de la predicción del habla interna, la sección VIII muestra el estado del arte con los resultados obtenidos de algunos trabajos basados en señales EEG para la predicción del habla interna, en la sección IX se describen dos conjuntos de datos para el reconocimiento del habla interna, en la sección X se presenta la metodología propuesta, la sección XI expone los resultados que se esperan obtener después de aplicada la metodología propuesta y finalmente en la sección XII se formulan las conclusiones.

2. Antecedentes

De acuerdo con el enfoque vygotskiano (Vygotsky 1896-1934), el pensamiento no se expresa simplemente en palabras, sino que existe a través de ellas, es decir, el lenguaje, el cual, si bien es adquirido en la interacción social, se interioriza en el curso del desarrollo, primero transformándose en un habla privada, que sigue siendo audible pero ya no dirigida a otros, y finalmente convirtiéndose en (inaudible) habla interna [6, 7].

3. Definiciones

Vygotsky (1896-1934) define el habla interna como la reutilización controlada e interna de una capacidad que se manifiesta primero en forma de discurso abierto. Oppenheim y Dell (2010) proponen que el habla interna “se caracteriza típicamente como la activación de representaciones lingüísticas abstractas o una simulación articulatoria detallada a la que sólo le falta la producción de sonido”. Alderson-Day y Fernyhough (2015) proponen que el habla interna puede definirse como “la experiencia subjetiva del lenguaje en ausencia de la articulación manifiesta y audible”. Grandchamp et al. (2019) también respaldan esta definición, ajustándola ligeramente a: “la experiencia subjetiva de la verbalización en ausencia de articulación o signo manifiesto”. Peter Langland-Hassan (2020) define al habla interna como un tipo de uso mental del lenguaje, o simulación del habla, que, con bastante frecuencia, ocurre conscientemente y en ausencia de cualquier articulación manifiesta [8]. Anna M. Borghi y Charles Fernyhough (2021) proponen que el habla interna es el resultado de la internalización de interacciones mediadas lingüísticamente que regulan la cognición y el comportamiento [7]. Nikola A. Kompa (2023) describe al habla interna como un episodio interno que involucra sustancialmente el sistema de producción del habla [9].

4. Relación entre habla interna y habla externa

El habla interna está fonéticamente “empobrecida” en comparación con el habla externa, ocurriendo en un nivel más “abstracto” de procesamiento del lenguaje (es decir, antes de que se seleccionen las características fonémicas completas correspondientes al mensaje deseado). Puede ser que haya un componente articulatorio separado del habla interna: una representación o simulación [8]. Se ha demostrado que el habla interna involucra áreas similares del cerebro, al igual que la producción del habla manifiesta o externa [7]. Debido a las transformaciones que acompañan a la internalización, el discurso privado es funcional y formalmente distinto de su precursor, el habla social, lo que requerirá el refinamiento de los modelos de procesamiento predictivo del habla interna que la conciben como un discurso social atenuado. La evidencia de la abstracción proviene de los hallazgos de que el habla interna puede ocurrir en formas más condensadas y menos ricas en características. Un ejemplo reciente que apoya la abstracción mostró una disociación entre las velocidades en la lectura en voz alta y la lectura silenciosa en participantes adultos.

Mientras que la velocidad de lectura en voz alta se correlacionaba con la velocidad de articulación, la lectura silenciosa se correlacionaba más fuertemente con el vocabulario y el conocimiento conceptual [10]. Del mismo modo que el habla externa implica movimientos articulatorios complejos de la boca, los labios, la lengua y la caja vocal, el habla interna también puede implicar imágenes motoras de esos movimientos; o, si no se trata de imágenes, puede implicar la generación de un plan motor necesario para producir tal enunciado. Podemos llamar a esto el componente articulatorio del habla interna. La influyente teoría de Levelt (1989) sobre la producción del habla señala que el hablante comienza con un “mensaje” a transmitir, donde el mensaje se realiza en una estructura conceptual no lingüística de algún tipo, y luego lleva a cabo etapas

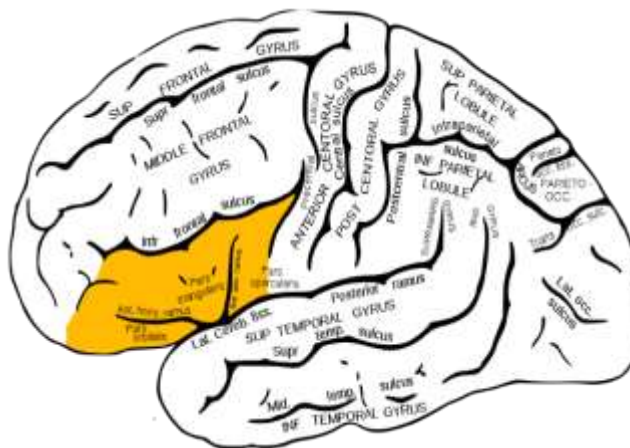


Fig. 1. Circunvolución frontal inferior izquierda (IIFGT) [13].

sucesivas de procesamiento lingüístico con el fin de generar palabras de lenguaje natural adecuadas para expresar el mensaje en voz alta [11].

5. Componentes del habla interna

Dentro de la categoría del habla interna hay otra distinción importante entre hablar (donde la experiencia subjetiva es la de producir un enunciado, análogo a hablar en una grabadora de audio) y la audición interna (donde la experiencia es la de percibir una expresión que ha sido hablada internamente, como al escuchar la propia voz en una grabación) [10]. Al menos a menudo, nuestros episodios internos del habla ("ISE", por sus siglas en inglés) parecen tener un componente auditivo y debido a esto podemos decir que los ISE tienen asociado un carácter sensorial auditivo-fonológico [11]. Scott presentó evidencia de que la generación del habla interna atenúa la percepción de los sonidos externos, lo que es consistente con la opinión de que una predicción sensorial del enunciado en cuestión interfiere con la percepción de un sonido externo [12].

6. Habla interna en el cerebro

Las áreas de producción del lenguaje, como la circunvolución frontal inferior izquierda (IIFG, Figura. 1), son las que se activan de manera más significativa durante el habla interna, aunque, en algunas condiciones, hay una activación significativa de la percepción del lenguaje (como la circunvolución temporal superior (STG, Figura 2) [8].

Es posible que el paso del habla interna condensada a la expandida implique la traducción de tales representaciones en algo más expresado, a través de la articulación en el giro frontal inferior izquierdo también llamado circunvolución frontal inferior izquierda y la representación fonológica en las estructuras del giro temporal superior o también conocido como circunvolución temporal superior [12].

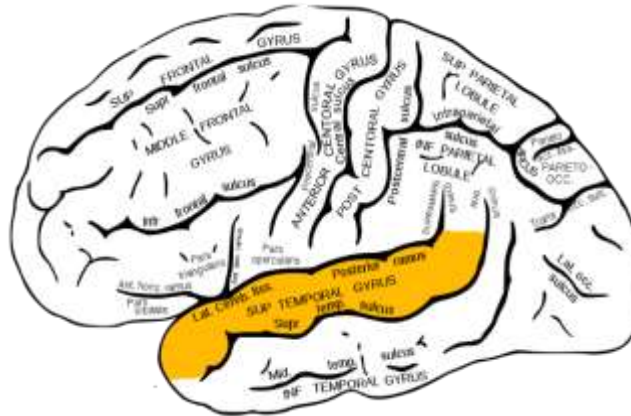


Fig. 2. Circunvolución temporal superior (STG) [13].

7. Electroencefalografía

La electroencefalografía es una técnica de imagen médica que lee la actividad eléctrica del cuero cabelludo generada por las estructuras cerebrales. El electroencefalograma (EEG) se define como la actividad eléctrica de tipo alterno registrada en la superficie del cuero cabelludo después de ser captada por electrodos metálicos y medios conductores [2]. El EEG medido directamente desde la superficie cortical se denomina electrocortigrama, mientras que cuando se utilizan sondas de profundidad se denomina electrograma. Por lo tanto, la lectura electroencefalográfica es un procedimiento completamente no invasivo que se puede aplicar repetidamente a pacientes, adultos normales y niños, prácticamente sin riesgo ni limitación. Los patrones cerebrales forman formas de onda que suelen ser sinusoidales. Por lo general, se miden de pico a pico y normalmente oscilan entre 0,5 y 100 μV de amplitud. Por medio de la transformada de Fourier se deriva el espectro de potencia de la señal de EEG en bruto.

En el espectro de potencia, la contribución de las ondas sinusoidales con diferentes frecuencias es visible. Aunque el espectro es continuo, desde 0 Hz hasta la mitad de la frecuencia de muestreo, el estado cerebral del individuo puede hacer que ciertas frecuencias sean más dominantes. Las ondas cerebrales se han clasificado en cuatro grupos básicos [14]:

- Ondas Gamma (>35 Hz): Los cambios en la alta frecuencia gamma (70-150 Hz) se asocian con el habla abierta y encubierta. Durante el habla abierta el lóbulo temporal, el área de Broca, el área de Wernicke, la corteza premotora y la corteza motora primaria presentan cambios gamma elevados. Por otro lado, este estudio también presenta evidencia de altos cambios gamma durante el habla encubierta en la circunvolución supramarginal y el lóbulo temporal superior.

Tabla 1. Estado del arte.

Autor, Año	Características y métodos	Resultados
Moctezuma et al. (2017) [15]	Clasificación binaria entre actividad e inactividad lingüística imaginada utilizando dos conjuntos de datos, el primero que contiene las palabras en español: arriba, abajo, izquierda, derecha y seleccionar. Mientras que el segundo contiene las palabras en español: arriba, abajo, izquierda y derecha. Métodos: Bosques aleatorios	-92% de exactitud
Balaji et al. (2017) [16]	Clasificación multiclase de las palabras imaginadas “yes”, “no”, “haan” y “na” en el idioma hindi e inglés como primer y segundo lenguaje respectivamente. Métodos: Máquinas de soporte vectorial Árboles aleatorios AdaBoost Redes neuronales artificiales	-92.18% de exactitud
Sereshkeh et al. (2017) [17]	Clasificación binaria de las palabras imaginadas “yes” y “no”. Métodos: Máquina de soporte vectorial	-69.3% de exactitud
Alsaleh et al. (2018) [18]	Clasificación multiclase de las palabras imaginadas en español: arriba, abajo, izquierda, derecha y seleccionar. Métodos: Máquina de soporte vectorial Naive Bayes Bosques aleatorios Análisis discriminante lineal	-87.4% de exactitud
D. Vorontsova et al. (2021) [4]	Clasificación multiclase de 8 palabras imaginadas en el idioma ruso que expresan los comandos: forward, backward, up, down, help, take, stop y reléase. Métodos: Redes neuronales convolucionales Redes neuronales recurrentes	-84.5% de exactitud
Varshney and Khan (2022) [19]	Clasificación multiclase de seis palabras imaginadas: could, yard, give, him, there y toe. Métodos: Bosques aleatorios Máquinas de soporte vectorial	-28.61% de exactitud
Agarwal and Kumar (2022) [20]	Clasificación multiclase de cuatro palabras y una frase imaginadas, las cuales son, “sos”, “stop”, “medicine”, “washroom” y “come-here” respectivamente. Métodos: Redes neurales recurrentes de tipo memoria a corto y largo plazo	-73.56% de exactitud

	Redes neuronales convolucionales	
	Unidades recurrentes cerradas	
<i>Einizade et al. (2022) [21]</i>	Clasificación multiclase de tres palabras imaginadas las cuales son “piedra”, “papel” o “tijeras” y el estado de reposo. Métodos: Clasificadores basados en grafos	-50.10% de exactitud
<i>Liwicki et al. (2022) [22]</i>	Clasificación multiclase de dos grupos de palabras imaginadas en español, el primer grupo se compone de las vocales y el segundo grupo de seis palabras que expresan comandos: arriba, abajo, izquierda, derecha, atrás y adelante. Acceso abierto. Métodos: Redes neurales recurrentes de tipo memoria a corto y largo plazo Redes neuronales convolucionales Unidades recurrentes cerradas	-35.20% de exactitud para las vocales -29.21% de exactitud para las palabras que expresan comandos

- Ondas Beta (12–35 Hz): Estas ondas a menudo están relacionadas con el movimiento muscular y la retroalimentación. Por lo tanto, se puede considerar que están involucrados durante las tareas auditivas y la producción del habla.
- Ondas Alfa (8–12 Hz): Durante el procesamiento del lenguaje, estas ondas están involucradas en la retroalimentación auditiva y la percepción del habla. Además, la frecuencia alfa durante el habla encubierta se ha identificado como débil en comparación con su comportamiento durante el discurso abierto.
- Ondas Theta (4–8 Hz): Estas ondas se activan durante la restauración fonémica y el procesamiento de las señales de coarticulación para componer palabras. Además, otro estudio identificó que las ondas theta pueden ayudar a identificar las consonantes en las sílabas.
- Ondas Delta (0.5–4 Hz): Se ha encontrado que la entonación y el ritmo durante la percepción del habla caen en rangos de frecuencia que pertenecen a la banda de oscilación delta inferior. Asimismo, diversos estudios han encontrado otros procesos del habla en los que intervienen las ondas delta, como el fraseo prosódico, la estructura de sílabas, las sílabas largas, entre otros [2].

8. Estado del arte

En la Tabla 1 se muestra el estado del arte, el cual está organizado por año y autor, características de los conjuntos de datos y los modelos utilizados para cada caso, y finalmente los resultados por porcentaje de exactitud.

Tabla 2. Características del primer conjunto de datos.

Características	
Palabras	Arriba, abajo, izquierda, derecha y seleccionar
Sujetos	27
Número de ensayos por palabra	33
Técnica	EEG
Canales	14
Frecuencia de muestreo	128 Hz
Número de registros	1 569 207

9. Conjuntos de datos para el reconocimiento del habla interna

Existen una serie de conjuntos de datos para el reconocimiento del habla interna, sin embargo, actualmente se cuenta con escasos conjuntos de acceso abierto que puedan ser utilizados para el análisis de dicho problema y el mejoramiento de los modelos. Entre estos, se encuentran los siguientes dos conjuntos.

Conjunto 1: Este trabajo tiene como objetivo interpretar las señales de EEG registradas durante la pronunciación imaginada de palabras de un vocabulario reducido, sin emitir sonidos ni articular movimientos (habla interna o no pronunciada) con la intención de controlar un dispositivo. Específicamente, el vocabulario permitiría controlar el cursor de la computadora, y consta de las palabras del lenguaje español: “arriba”, “abajo”, “izquierda”, “derecha”, y “seleccionar”. Para ello, se registraron las señales de EEG de 27 individuos utilizando un protocolo básico para saber a priori en qué segmentos de la señal la persona imagina la pronunciación de la palabra indicada.

Para la adquisición de la actividad cerebral se registraron las señales EEG utilizando un kit EMOTIV. Este kit es inalámbrico y consta de catorce electrodos (canales) de alta resolución (más las referencias CMS/DRL en las posiciones P3/P4 respectivamente) cuya frecuencia de muestreo es de 128 Hz. Los nombres de los canales, de acuerdo con el sistema internacional 10-20, son: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 (ver Figura 3).

El protocolo consiste en colocar a la persona cómodamente sentada con los ojos abiertos cerca de un escritorio, y con la mano derecha sobre el mouse de una computadora. Con un clic al mouse, el usuario delimita tanto el inicio como el fin de la pronunciación imaginada de alguna de las cinco palabras del vocabulario. La pronunciación imaginada de cada una de las cinco palabras fue repetida 33 veces consecutivas durante el registro del EEG, es decir, cinco bloques de 33 repeticiones por palabra. Antes de cada bloque se le indicó al individuo cuál es la palabra que debía pronunciar internamente. Asimismo, todas las épocas de las cinco palabras

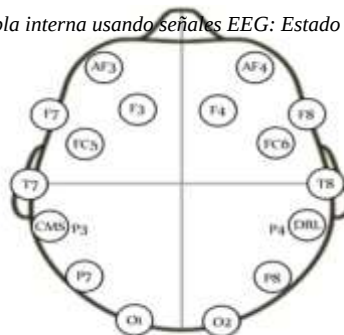


Fig. 3. Localización de los electrodos en el kit Emotiv [23].

pertenecientes a un individuo fueron registradas en una sola sesión. Además, al inicio del registro de las señales de EEG, se le indicó a la persona que evitara parpadear o realizar movimientos corporales mientras imaginaba la pronunciación de la palabra indicada, ya que después de cada marcador de fin podía tomarse un tiempo de descanso para dichos movimientos.

Con la finalidad de que el individuo no supiera cuantas veces se repetiría una palabra, en la sala de experimentos otra persona, alejada del campo visual y guardando el debido silencio, se encargó de realizar el conteo de repeticiones e indicó al individuo cuándo debía concluir. Esto con la finalidad de que el individuo no se distrajera contando el número de repeticiones ni se predispusiera a saber que le faltaba poco o mucho para concluir el experimento. Las sesiones se registraron en un laboratorio alejado de ruido audible externo, ruido visual (como la transición del día y la noche, y distracciones), entre otras. Para formar un corpus de datos para los experimentos, se registraron las señales de EEG de 27 individuos sanos (S1-S27), de los cuales 2 de ellos son zurdos y el resto son diestros. Sin embargo, de acuerdo con el modelo Geschwind-Wernicke, las señales de EEG relacionadas con la producción del habla afectan diferentes áreas en la parte izquierda del cerebro (a excepción de algunas personas zurdas) [23]. Las características dichas anteriormente, así como el número de registros que contiene el conjunto se pueden apreciar más claramente en la Tabla 2.

Conjunto 2: El objetivo principal de este trabajo es proporcionar a la comunidad científica una base de datos de electroencefalografía multiclase de acceso abierto de comandos de voz interna que podría utilizarse para una mejor comprensión de los mecanismos cerebrales relacionados. Cada participante realizó entre 475 y 570 ensayos en un solo día de registro, obteniendo un conjunto de datos con más de 9 horas de registro continuo de datos de EEG, que resultaron siendo más de 5600 ensayos en total.

El protocolo experimental fue aprobado por el Comité Asesor de Ética y Seguridad en el Trabajo Experimental (CEySTE, CCT-CONICET, Santa Fe, Argentina). Diez participantes diestros sanos, cuatro mujeres y seis hombres con edad media de 34 años, sin pérdida de audición, sin pérdida del habla y sin trastornos neurológicos, del movimiento o psiquiátricos, se unieron al experimento y dieron su consentimiento informado por escrito. Todos los participantes eran hablantes nativos de español.

Ninguno de los individuos tenía experiencia previa en BCI y participaron en aproximadamente dos horas de grabación. El estudio se llevó a cabo en una habitación

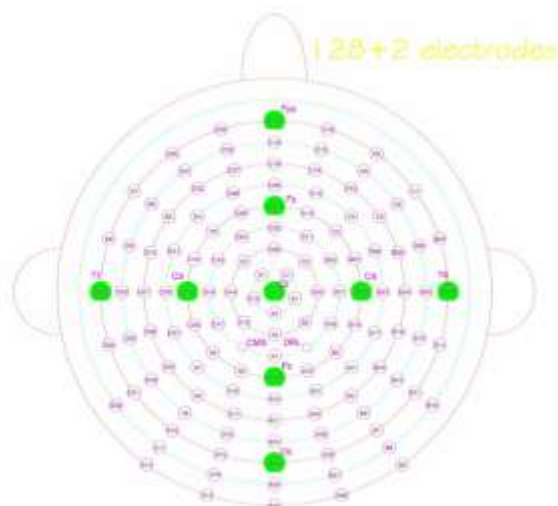


Fig. 4. Distribución de los electrodos en el casco BioSemi cap_128_layout_medium.jpg (970×900) (biosemi.com)[24].

con blindaje eléctrico. Los participantes se sentaron en una silla cómoda frente a una pantalla de computadora donde se presentaron las señales visuales. Con el fin de familiarizar al participante con el procedimiento experimental y el ambiente de la sala, se explicaron todos los pasos del experimento, mientras se colocaba el casco del EEG y los electrodos externos. El proceso de configuración duró aproximadamente 45 minutos. En la Figura 4 se muestra la configuración principal del experimento.

El fondo de la pantalla era de color gris claro para evitar el deslumbramiento y la fatiga ocular. Cada individuo participó en un solo día de grabación que comprendió tres sesiones consecutivas. Se dio un período de descanso autoseleccionado entre sesiones, para evitar el aburrimiento y la fatiga. Al comienzo de cada sesión, se registró una línea de base de segundos en la que se indicaba al participante que se relajara y permaneciera lo más quieto posible. Dentro de cada sesión, se presentaron cinco corridas de estimulación. Las ejecuciones corresponden a las diferentes condiciones propuestas: habla pronunciada, habla interna y condición visualizada. Al comienzo de cada corrida, la condición se anunciaba en la pantalla de la computadora durante un período de 3 segundos. En todos los casos, el orden de las corridas fue: un discurso pronunciado, dos discursos internos y dos condiciones visualizadas. Se dio un descanso de un minuto entre corridas. Las clases se seleccionaron específicamente considerando una aplicación de control BCI natural con las palabras en español: "arriba", "abajo", "derecha", "izquierda".

La clase (palabra) del ensayo se presentó al azar. Cada participante tuvo 200 ensayos tanto en la primera como en la segunda sesión. Sin embargo, dependiendo de la voluntad y el cansancio, no todos los participantes realizaron el mismo número de ensayos en la tercera sesión. Para evaluar la atención de cada participante, se agregó aleatoriamente un control de concentración al habla interna y se ejecutó la condición visualizada. La tarea de control consistió en preguntar al participante, después de algunos ensayos seleccionados al azar, cuál era la dirección de la última clase mostrada.

El participante tenía que seleccionar la dirección usando las flechas del teclado. No se dio un límite de tiempo para responder a estas preguntas y el protocolo continuó después de que el participante presionara cualquiera de las cuatro teclas de flecha. Se proporcionó información visual que indicaba si la pregunta se había respondido correcta o incorrectamente.

Los datos de electroencefalografía (EEG), electrooculografía (EOG) y electromiografía (EMG) se adquirieron utilizando un sistema de medición de alta resolución BioSemi ActiveTwo. Para la adquisición de datos, se utilizaron 128 canales activos de EEG y 8 canales externos activos de EOG/EMG con una resolución de 24 bits y una frecuencia de muestreo de 1024 Hz. BioSemi también proporciona gorros de cabeza de EEG estándar de diferentes tamaños con posiciones de electrodos pre-fijados. Se eligió un gorro de tamaño adecuado para cada participante midiendo la circunferencia de la cabeza con una cinta métrica.

Con el objetivo de localizar las principales fuentes de activación y analizar sus conexiones, pedimos a los participantes que realizaran el experimento bajo tres condiciones diferentes: habla interna, habla pronunciada y condición visualizada:

- *Habla interna.* El habla interna es la condición principal en el conjunto de datos y tiene como objetivo detectar la actividad eléctrica del cerebro relacionada con el pensamiento de un participante sobre una palabra en particular. En las ejecuciones de discurso interno, se indicó a cada participante que imaginara su propia voz como si estuviera dando una orden directa a la computadora, repitiendo la palabra correspondiente hasta que el círculo blanco se volviera azul. A cada participante se le pidió explícitamente que no se centrara en los gestos de articulación. Además, a cada participante se le indicó que permaneciera lo más quieto posible y que no moviera la boca ni la lengua. En aras de la imaginación natural, no se proporcionó ninguna señal rítmica.

- *Habla pronunciada.* La condición pronunciada del habla se propuso con el propósito de encontrar regiones motoras involucradas en la pronunciación que coincidan con las activadas durante la condición interna del habla. En las ejecuciones de discurso pronunciado, se indicó a cada participante que pronunciara repetidamente en voz alta la palabra correspondiente a cada señal visual, como si estuviera dando una orden directa a la computadora. Al igual que en las ejecuciones de voz internas, no se proporcionó ninguna señal rítmica.

- *Condición visualizada.* Dado que las palabras seleccionadas tienen un alto componente visual y espacial, y con el objetivo de encontrar cualquier actividad relacionada que se produce durante el habla interior, se propuso la condición visualizada. Es oportuno mencionar que los principales centros neuronales relacionados con el pensamiento espacial se localizan en las regiones occipital y parietal. En las corridas de condiciones visualizadas, se indicó a los participantes que se concentraran en mover mentalmente el círculo que aparecía en el centro de la pantalla en la dirección indicada por la señal visual [24].

Las características dichas anteriormente, así como el número de registros que contiene el conjunto se pueden apreciar más claramente en la Tabla 3.

Tabla 3. Características del segundo conjunto de datos.

Características	
Palabras	Arriba, abajo, izquierda y derecha
Sujetos	10
Número de ensayos por palabra	Habla externa: 25-30
	Habla interna: 50-60
Técnica	EEG
Canales	128
Frecuencia de muestreo	1024 Hz
Número de registros	47 775 570

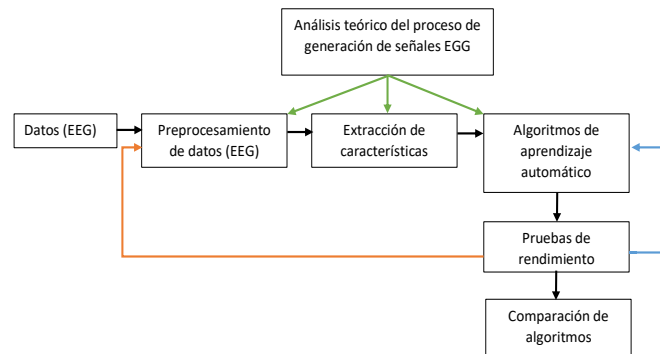


Fig. 5. Diagrama de la metodología propuesta.

10. Metodología propuesta

La metodología que se propone en este trabajo y con base en la información y conjuntos de datos descritos en las secciones anteriores, es la siguiente:

Como se aprecia en el diagrama de la Figura 5, la propuesta comienza con la adquisición de los datos, posteriormente se realiza una investigación teórica sobre el proceso de generación de las señales EEG, y para este caso particular en qué consiste el habla interna, así como las zonas del cerebro involucradas y las características y peculiaridades que la distinguen. Lo anterior, con el objetivo de conocer sus principales propiedades y tomar decisiones informadas para la selección del modelo de clasificación (flechas verdes →). El paso siguiente es la realización del preprocesamiento de los datos recaudados, así como la extracción de características, que bien puede ser mediante medidas estadísticas [15] aplicando además un método que por medio de raíces aplicadas a éstas medidas se saca una relación entre todas ellas a partir de la sustitución consecutiva del índice y el radicando por cada uno de los

valores de las medidas estadísticas [25] u otros métodos afines para luego aplicar clasificadores basados en algoritmos de aprendizaje automático como redes neuronales recurrentes de tipo memoria a corto y largo plazo (LSTM, por sus siglas en inglés) o técnicas de aprendizaje con grafos. Posteriormente se propone realizar pruebas de rendimiento para comparar los resultados con trabajos anteriores, de forma que, si el porcentaje no supera a los anteriores o no se obtiene un resultado relevante en ningún aspecto se vuelva a la etapa de algoritmos de aprendizaje automático (representado por la flecha azul →) o en su peor defecto al preprocesamiento de los datos (representado por la flecha naranja →).

Existe también una segunda alternativa de comparación que consiste en utilizar Transformers [26][27] específicamente un transformador generativo pre-entrenado para series de tiempo, mejor conocido como TimeGPT por sus siglas en inglés, que es el primer modelo base para series de tiempo, capaz de generar predicciones precisas para diversos conjuntos de datos sin necesidad de entrenamiento adicional. Lo anterior, gracias a que TimeGPT se entrenó con la mayor colección de series de tiempo disponibles públicamente (alrededor de 100 billones de datos) y puede pronosticar series de tiempo no vistas sin volver a entrenar sus parámetros. El entrenamiento de TimeGPT, con este rico conjunto de datos lo equipa para manejar una amplia gama de escenarios, lo que mejora su solidez y capacidades de generalización. Además de las características temporales, el conjunto de datos muestra variabilidad en los niveles de ruido y la presencia de valores atípicos, lo que proporciona una condición de entrenamiento sólida.

Esto es importante porque algunos conjuntos de datos exhiben patrones ordenados y predecibles, mientras que otros exhiben picos de ruido significativos o eventos inesperados, lo que proporciona una amplia gama de escenarios para la asimilación de modelos lo que permite a TimeGPT pronosticar con precisión series temporales no vistas, al tiempo que elimina la necesidad de entrenamiento y optimización de modelos individuales. TimeGPT se beneficia del entrenamiento previo en conjuntos de datos de series temporales masivos y diversos, incluso en escenarios con datos escasos, puede aprovechar este conocimiento previo para tener un buen rendimiento. En contraste, los modelos tradicionales de aprendizaje automático tienen dificultades para capturar patrones complejos debido a la insuficiencia de datos de entrenamiento, lo que resulta en una menor precisión de pronóstico. Con este nuevo modelo se hace innecesaria la implementación del preprocesamiento a los datos, la optimización de hiper parámetros y la selección de modelos [26, 27].

11. Resultados esperados

Los resultados que se esperan obtener dada la metodología propuesta, son, en primera instancia, una métrica de exactitud que supere la obtenida en trabajos anteriores relacionados o al menos los que se sirven de los mismos conjuntos de datos citados en la sección IX, así como también sustraer conocimientos significativos acerca de las características más primordiales y fundamentales de las señales del habla interna para el caso particular de un vocabulario reducido de palabras que expresan comandos, sean estas características propiedades de las señales como frecuencia, amplitud, formas

basadas en parámetros, caracterización de patrones, etc.. De igual manera llegar a conclusiones importantes sobre los modelos aplicables a problemas de series de tiempo basados en señales electroencefalográficas para la clasificación del habla interna o incluso para construir una hipótesis bien fundamentada para el reconocimiento del habla interna fluida o el discurso interno.

12. Conclusiones

Las señales electroencefalográficas poseen características útiles para el análisis del habla interna que pueden aportar información relevante a cerca de las propiedades que distinguen al habla interna del habla externa, debido a esto los modelos de procesamiento predictivo del habla interna que la conciben como un discurso social atenuado requerirán su refinamiento. Por lo que se espera que la presente metodología propuesta aporte información relevante a tal fin, con la que sea posible la clasificación más precisa de un vocabulario reducido de palabras que expresan comandos, así como sentar algunas bases hipotéticas para el reconocimiento del habla interna fluida a partir de señales electroencefalográficas.

Agradecimientos. El trabajo se realizó con el apoyo de la Secretaría de Investigación y Posgrado del Instituto Politécnico Nacional mediante los proyectos SIP 20240951 y 20240950. Además, los autores agradecen al CONAHCYT por su apoyo.

Referencias

1. Lee, Y.E., Lee, S. H., Kim, S. H.: Towards Voice Reconstruction from EEG During Imagined Speech. In: Proc. 37th AAAI Conf. Artif. Intell., 37, pp. 6030–6038 (2023) doi: 10.1609/aaai.v37i5.25745.
2. Lopez-Bernal, D., Balderas, D., Ponce, P.: A State-of-the-Art Review of EEG-Based Imagined Speech Decoding. *Front. Hum. Neurosci.*, 16, pp. 1–14 (2022) doi: 10.3389/fnhum.2022.867281.
3. Gu X., *et al.*: EEG-Based Brain-Computer Interfaces (BCIs): A Survey of Recent Studies on Signal Sensing Technologies and Computational Intelligence Approaches and Their Applications. In: *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, 18(5), pp. 1645–1666 (2021) doi: 10.1109/TCBB.2021.3052811.
4. Vorontsova, D., *et al.*: Silent Eeg-Speech Recognition Using Convolutional And Recurrent Neural Network With 85% Accuracy Of 9 Words Classification. *Sensors*, 21(20), pp. 1–19 (2021) doi: 10.3390/s21206744.
5. Nurhadi, J., Rahma, R., Rahayu, W.: Mechanism of Inner Speech in Silent Reading Activity. In: *Proc. Fifth Int. Conf. Lang. Lit. Cult. Educ. (ICOLLITE)*, 595, pp. 379–385 (2022) doi: 10.2991/assehr.k.211119.059.
6. L. S. Vigotsky, *Pensamiento y lenguaje*. México: Ediciones Quinto Sol (1988)
7. Kompa, N. A.: Inner Speech and ‘Pure’ Thought – Do we Think in Language?. *Rev. Philos. Psychol* (2023) doi: 10.1007/s13164-023-00678-w.
8. Langland-Hassan, P.: Inner speech. *Wiley Interdiscip. Rev. Cogn. Sci.*, 12(2), pp. 1–18 (2021) doi: 10.1002/wcs.1544.
9. Borghi A. M., Fernyhough, C.: Concepts, abstractness and inner speech. *Philos. Trans. R. Soc. B Biol. Sci.*, 378(1870) (2023) doi: 10.1098/rstb.2021.0371.

10. Fernyhough, C., Borghi, A. M.: Inner speech as language process and cognitive tool. *Trends Cogn. Sci.*, 27 (12), pp. 1180–1193 (2023) doi: 10.1016/j.tics.2023.08.014.
11. Gregory, D.: Inner Speech: New Voices. *Anal. (United Kingdom)*, 80(1), pp. 164–173 (2020) doi: 10.1093/ANALYS/ANZ096.
12. Fernyhough C., Alderson-Day, B.: Inner Speech: Development, Cognitive Functions, Phenomenology, and Neurobiology. *Psychol. Bull.*, 141(5), pp. 931–965 (2015)
13. Wikimedia: Wikimedia Commons (2020) url: <https://commons.wikimedia.org/wiki/MainPage>.
14. Teplan, M.: xxFundamentals of EEG Measurement. *Meas. Sci. Rev.*, 2, pp. 1–11 (2002)
15. Alfredo-Moctezuma, L., Carrillo, M., Villaseñor Pineda, L.: Hacia la clasificación de actividad e inactividad lingüística a partir de señales de electroencefalogramas (EEG). *Res. Comput. Sci.*, 140 (1), pp. 135–149 (2017) doi: 10.13053/rcs-140-1-11.
16. Balaji A., et al.: EEG-based Classification of Bilingual Unspoken Speech Using ANN. In: *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 1022–1025 (2017) doi: 10.1109/EMBC.2017.8037000.
17. Sereshkeh, A.R., Trott, R., Bricout, A.: Online EEG Classification of Covert Speech for Brain-Computer Interfacing. *Int. J. Neural Syst.*, 27(8), pp. 1–16 (2017) doi: 10.1142/S0129065717500332.
18. Alsaleh, M., Moore, R., Christensen, H.: Examining Temporal Variations in Recognizing Unspoken Words Using EEG Signals. In: *Proc. IEEE Int. Conf. Syst. Man, Cybern. SMC 2018*, pp. 976–981 (2018) doi: 10.1109/SMC.2018.00173.
19. Varshney Y.V., Khan, A.: Imagined Speech Classification Using Six Phonetically Distributed Words. *Front. Signal Process.*, 2, pp. 1–11 (2022) doi: 10.3389/frsip.2022.760643.
20. Agarwal P., Kumar, S.: Electroencephalography-Based Imagined Speech Recognition Using Deep Long Short-Term Memory Network. *ETRI J.* 44(4), pp. 672–685 (2022) doi: 10.4218/etrij.2021-0118.
21. Einizade, A., Mozafari, M., Jalilpour, S.: Neural Decoding of Imagined Speech From EEG Signals Using the Fusion of Graph Signal Processing and Graph Learning Techniques. *Neurosci. Informatics*, 2(3), pp. 100091 (2022) doi: 10.1016/j.neuri.2022.100091.
22. Simistira-Liwicki, F., Gupta, V., Saini, R.: Rethinking the Methods and Algorithms for Inner Speech Decoding and Making Them Reproducible. *NeuroSci*, 3(2), pp. 226–244 (2022) doi: 10.3390/neurosci3020017.
23. Torres, A., Reyes, C.A., Villaseñor, L.: Análisis de Señales Electroencefalográficas para la Clasificación de Habla Imaginada. *Rev. Mex. Ing. Biomédica*, 34(1), pp. 23–39 (2013)
24. Nieto, N., Peterson, V., Rufiner, H. L.: Thinking Out Loud, an Open-Access EEG-Based BCI Dataset for Inner Speech Recognition. *Sci. Data*, 9(1) (2022) doi: 10.1038/s41597-022-01147-2.
25. Cerón Figueroa, S.: Clasificador asociativo basado en conceptos lineales y no lineales. Instituto Politécnico Nacional, CIC (2018) <https://tesis.ipn.mx/jspui/bitstream/123456789/26705/1/T1977.pdf>.
26. Garza, A., Mergenthaler-Canseco, M.: TimeGPT-1. pp. 1–12, (2023) <http://arxiv.org/abs/2310.03589>.
27. Liao, W., et al.: TimeGPT in Load Forecasting: A Large Time Series Model Perspective. (2024) <http://arxiv.org/abs/2404.04885>.