ARTÍCULO DE INVESTIGACIÓN
Vol. 34, No. 1, Abril 2013, pp. 23-39

Revista Mexicana de Ingeniería Biomédica
SOMIB

Análisis de Señales Electroencefalográficas para la Clasificación de Habla Imaginada

A.A. Torres-García
C.A. Reyes-García
L. Villaseñor-Pineda
J.M. Ramírez-Cortés

Instituto Nacional de Astrofísica Óptica y Electrónica (INAOE)

RESUMEN

El presente 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 imaginada 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. Posteriormente, se utiliza la transformada wavelet discreta (DWT) para extraer características de los segmentos que son usados para calcular la energía relativa wavelet (RWE) en cada una de los niveles en los que la señal es descompuesta, y se selecciona un subconjunto de valores RWE provenientes de los rangos de frecuencia menores a 32 Hz. Enseguida, éstas se concatenan en dos configuraciones distintas: 14 canales (completa) y 4 canales (los más cercanos a las áreas de Broca y Wernicke). Para ambas configuraciones se entrenan tres clasificadores: Naive Bayes (NB), Random Forest (RF) y Máquina de vectores de soporte (SVM). Los mejores porcentajes de exactitud se obtuvieron con RF cuyos promedios fueron 60.11 % y 47.93 % usando las configuraciones de 14 canales y 4 canales, respectivamente. A pesar de que los resultados aún son preliminares, éstos están arriba del 20 %, es decir, arriba del azar para cinco clases. Con lo que se puede conjeturar que las señales de EEG podrían contener información que hace posible la clasificación de las pronunciaciones imaginadas de las palabras del vocabulario reducido.

Palabras clave: electroencefalografía (EEG), interfaces cerebro-computadora(BCI), habla no pronunciada, habla imaginada, clasificación, transformada wavelet discreta (DWT), energía relativa wavelet, interfaces de habla silente (SSI).

Correspondencia:
Alejandro Torres-García
INAOE
Luis Enrique Erro # 1 Sta.
María Tonantzintla, Puebla.
Correo electrónico:
alejandro.torres@inaoep.mx

Fecha de recepción:
20 de Junio de 2012
Fecha de aceptación:
28 de Febrero de 2013

ABSTRACT

This work aims to interpret the EEG signals associated with actions to imagine the pronunciation of words that belong to a reduced vocabulary without moving the articulatory muscles and without uttering any audible sound (imagined or unspoken speech). Specifically, the vocabulary reflects movements to control the cursor on the computer, and consists of the Spanish language words: "arriba", "abajo", "izquierda", "derecha", and "seleccionar". To do this, we have recorded EEG signals from 27 subjects using a basic protocol to know a priori in what segments of the signal a subject imagines the pronunciation of the indicated word. Subsequently, discrete wavelet transform (DWT) is used to extract features from the segments. These are used to compute relative wavelet energy (RWE) in each of the levels in that EEG signal is decomposed and, it is selected a RWE values subset with the frequencies smaller than 32 Hz. Then, these are concatenated in two different configurations: 14 channels (full) and 4 channels (the channels nearest to the brain areas of Wernicke and Broca). The following three classifiers were trained using both configurations: Naive Bayes (NB), Random Forest (RF) and support vector machines (SVM). The best accuracies were obtained by RF whose averages were 60.11 % and 47.93 % using both configurations, respectively. Even though, the results are still preliminary, these are above 20 %, this means they are more accurate than chance for five classes. Based on them, we can conjecture that the EEG signals could contain information needed for the classification of the imagined pronunciations of the words belonging to a reduced vocabulary.

Keywords: brain-computer interface (BCI), electroencephalography (EEG), unspoken speech, discrete wavelet transform (DWT), imagined speech, relative wavelet energy, classification, silent speech interfaces (SSI).

INTRODUCCIÓN

En la República Mexicana, las discapacidades que se presentan con mayor número de ocurrencias son las de tipo motriz con un 58.3 %[1]. Dentro de este sector se encuentran discapacidades motrices severas 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. Estas discapacidades frecuentemente provocan que la persona no pueda controlar voluntariamente sus movimientos, incluyendo aquellos relacionados directa o indirectamente con la articulación del habla [1, 2]. En consecuencia, una persona en estas condiciones está prácticamente aislada de su entorno.

Con los avances tecnológicos, recientemente, se han propuesto diversas alternativas que permitan tener una mejor calidad de vida y reintegrar a estas personas a la sociedad. Las Interfaces Cerebro-Computadora (BCI, por sus siglas en inglés) son una de estas alternativas ya que intentan proveer a una persona de un

[1] XIII Censo General de Población y Vivienda 2010 realizado por el Instituto Nacional de Estadística, Geografía e Informática (INEGI)

nuevo canal, no muscular, de comunicación y control para transmitir mensajes y comandos al mundo exterior [1, 2]. Para adquirir datos acerca de la actividad cerebral, existen varias alternativas. Dentro de ellas se encuentra el electroencefalograma, el cual permite medir la actividad eléctrica proveniente del cerebro mediante un conjunto de electrodos colocados sobre el cuero cabelludo de las personas. Esta técnica es sensible al ruido provocado por otras actividades fisiológicas (la actividad cardiaca, el movimiento ocular, el movimiento de la lengua, la respiración, los potenciales de la piel, entre otras), e incluso por la corriente alterna y de los electrodos [3]. Sin embargo, a pesar de estas limitantes es la técnica más ampliamente utilizada en BCIs debido a que es no invasiva, requiere de un equipo relativamente sencillo y económico, y tiene buena resolución temporal.

PROBLEMÁTICA

En las BCIs basadas en EEG, a los mecanismos neurológicos o procesos empleados por el usuario para generar las señales de control, se les denomina fuentes electrofisiológicas. Las más utilizadas son: los potenciales corticales lentos (SCP, por sus siglas en inglés), los potenciales P300, las imágenes motoras (ritmos sensoriales motrices mu y beta) y los potenciales evocados visuales (VEP, por sus siglas en inglés) [4, 1, 2]. Con estas fuentes electrofisiológicas, la comunicación se realiza mediante dos paradigmas de control: discreto o continuo. En el paradigma discreto, el usuario puede elegir entre dos o más opciones discretas, por ejemplo elegir una tecla específica de un teclado virtual en el monitor de la computadora. En el paradigma continuo, un pequeño número de variables cinemáticas (por ejemplo, las coordenadas x y y de la posición del cursor en el monitor, o los valores de las primeras dos frecuencias formantes para una prótesis de habla) son controladas por el usuario [1].

Las BCIs descritas previamente, presentan los siguientes dos grandes problemas. El primero es el largo periodo de entrenamiento (algunas semanas hasta meses) requerido para que un usuario pueda utilizar una BCI. Lo anterior se debe a que, las fuentes descritas anteriormente (SCP, P300, imágenes motoras, y VEP) son generadas por el usuario de forma poco consciente [5]. Mientras que, el segundo son las bajas tasas de comunicación (una sola palabra procesada, o menos, por minuto) que resultan insuficientes para permitir una interacción natural. Este último problema se debe a que cada una de las fuentes electrofisiológicas usadas por las BCIs requieren un "mapeo" o traducción al dominio del habla [1].

Los problemas descritos anteriormente han motivado una serie de trabajos que tratan de utilizar los potenciales relacionados con la producción del habla, con diversos grados de éxito [1]. En estos trabajos, la fuente electrofisiológica es el habla imaginada, también referida como habla interna o habla no pronunciada (*unspoken speech*), donde el término habla imaginada se refiere a la pronunciación interna, o imaginada, de palabras pero sin emitir sonidos ni articular gestos para ello. Es importante mencionar que, Denby [6] incluye a estos trabajos dentro de un área de investigación denominada interfaces de habla silente (SSI, por *Silent Speech Interfaces*) cuya finalidad es desarrollar sistemas capaces de permitir la comunicación "hablada" que toman lugar cuando la emisión de una señal acústica entendible es imposible. Es importante remarcar que los trabajos que utilizan *habla imaginada* pueden dividirse, por la unidad de habla utilizada, en dos enfoques: palabras y sílabas. El primer enfoque es seguido en [7, 8, 9, 10, 11]. Mientras que, en [12, 13, 14] únicamente se tratan sílabas.

En el caso específico de los trabajos que exploran palabras -donde se ubica el presente trabajo-, se han identificado los siguientes problemas. En [9], se implementó un esquema de clasificación basada en prototipos que requiere de muchos ejemplos en el dominio del tiempo para generarlos, con lo que el método tiene un tasa de decisión lenta que resulta inadecuada para llevarse a procesamiento en linea. Mientras en [7, 8, 10, 11] se asume que las características extraídas pueden ser reconocidas con los modelos existentes para reconocimiento de habla común, no obstante las señales del habla y EEG presentan características muy diferentes. Por

ejemplo, las frecuencias del EEG llegan hasta 60 Hz, mientras que el habla humana está en el rango de 125 a 4000 Hz. Además, la señal del habla se capta en un sólo canal, mientras que la señal de EEG es registrada con un mayor número de canales (por ejemplo con 32, 64, 128, ó 256). Asimismo, en [8], el trabajo más reciente que utiliza el enfoque de palabras, se menciona la pertinencia de explorar un vocabulario distinto al usado en ese trabajo ("alpha", "'bravo", "charlie", "delta", "echo") que tenga mayor significado semántico con la finalidad de que dichas palabras puedan provocar una mayor actividad en el cerebro, y que se cuente con mayor número de repeticiones de cada palabra del vocabulario a usar (ellos utilizaron 20 repeticiones y registraron señales de EEG de 18 individuos). En consecuencia, se requiere de un modelo que permita tratar al reconocimiento de la pronunciación imaginada de palabras de manera adecuada. Además, es importante mencionar que este trabajo extiende los presentados por [15] y [16] con las siguientes diferencias. En la presente investigación se trabaja con las señales de EEG de un mayor número de individuos, y con un método de procesamiento y clasificación distinto a los descritos en [15, 16]. Asimismo, se trabaja con un mayor número de palabras que las usadas en [15].

La presente investigación tiene como objetivo poder interpretar las señales de EEG asociadas al habla imaginada. En específico, se orienta a interpretar las señales para reconocer la pronunciación imaginada de palabras de un vocabulario reducido compuesto por las siguientes cinco palabras en lenguaje español: "arriba", "abajo", "izquierda", "derecha" y "seleccionar". Estas palabras tienen mayor significado semántico que las utilizadas en [8], y cada una se repite 33 veces. Lo anterior retoma lo planteado por los trabajos previos con la idea en mente de que las palabras con mayor significado semántico puedan generar mayor actividad cerebral que permita su reconocimiento y correcta clasificación. Además, con ellas sería posible controlar la dirección del cursor de la computadora. El problema es tratado bajo un enfoque de clasificación, y se conoce a priori en qué parte de la señal de EEG la persona imagina la pronunciación de las palabras indicadas.

METODOLOGÍA

Las etapas de la metodología propuesta en este trabajo son: Adquisición de la actividad cerebral, Preprocesamiento, Extracción de características, Selección de características, y Clasificación. La metodología seguida en el trabajo se muestra de mejor manera en la Figura 1.

Los materiales y software utilizados en el trabajo se mencionan a continuación. 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 2). Además, se utiliza Matlab 2009b para la implementación de programas para las etapas de pre-procesamiento, extracción y selección de características. Mientras que, en la etapa de clasificación se utiliza Weka 3.6.8 [17]. A continuación se describe a detalle cada uno de los componentes de la metodología propuesta.

Adquisición de la actividad cerebral

En esta etapa se utiliza el EEG para adquirir las señales provenientes del cerebro. Como se menciona en la sección , se utiliza un kit de adquisición que consta de catorce canales. Sin embargo, de acuerdo al 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) [18].

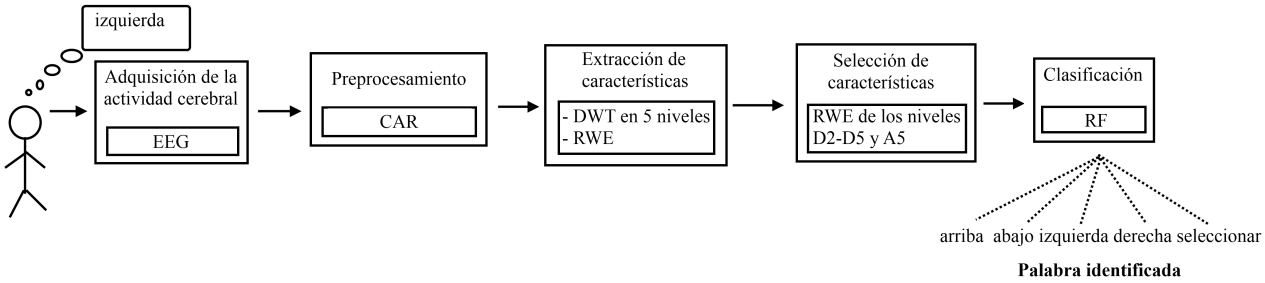

Figura 1: Diagrama de la metodología seguida en el trabajo. Los bloques internos detallan el método propuesto para la tarea de clasificación de habla imaginada registrada con electroencefalogramas. Como ejemplo, un individuo realiza la pronunciación imaginada de la palabra "izquierda" ante lo cual la tarea del método es inferir, con base a las señales de EEG registradas, cuál de las cinco palabras del vocabulario propuesto fue imaginada.

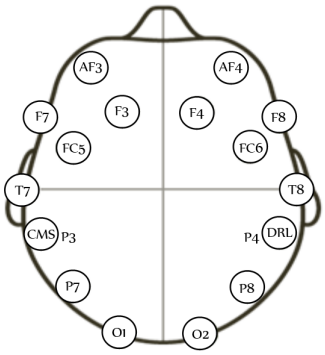

Figura 2: Localización de los electrodos en el kit Emotiv.

Particularmente, las áreas correspondientes a los canales F7, FC5, T7 y P7. Por lo tanto, en el presente trabajo, de manera similar que en [11], únicamente consideramos de interés las señales de EEG provenientes de estos canales que son los más cercanos a las regiones del modelo Geschwind-Wernicke.

Por otra parte, se utiliza un protocolo básico para adquirir las señales de EEG de cada individuo mientras imagina la pronunciación de las palabras. 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. Con cada clic se envía un marcador al software de registro de las señales de EEG (ver Figura 3). Asimismo, al conjunto de muestras que se encuentran entre los marcadores de inicio y fin, se les denomina ventanas (épocas).

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 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. También, es importante mencionar que, el marcador de inicio podría agregar ruido a la señal de EEG; sin embargo, éste estará en todas las ventanas así que el sesgo en la etapa de clasificación será mínimo.

Con la finalidad de que el individuo no sepa cuantas veces se repetirá una palabra, en la sala de experimentos otra persona, alejada del campo visual y guardando el debido silencio, se encarga de realizar el conteo de repeticiones e indica cuando el individuo debe concluir. Esto con la finalidad de que el individuo no se distraiga contando el número de repeticiones ni se predisponga a saber que le falta poco o mucho para concluir el experimento.

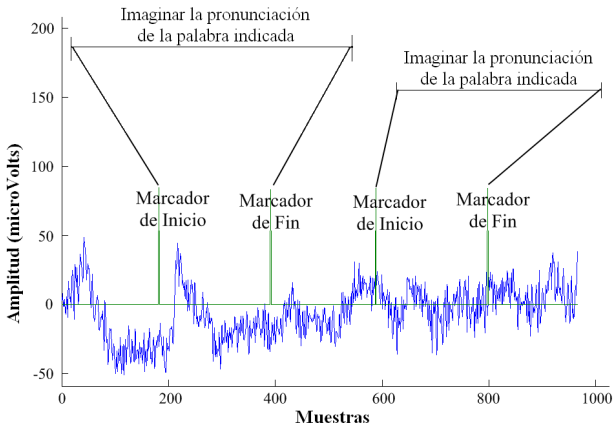

Figura 3: Señal de EEG del canal F7 del individuo S1 mientras imagina la dicción de la palabra "Abajo" siguiendo el protocolo de adquisición de datos

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.

La idea detrás del protocolo de adquisición es saber a priori en qué parte de la señal de EEG se debe buscar los patrones asociados con la imaginación de la pronunciación de la palabra indicada.

Preprocesamiento

Las señales de EEG obtenidas son preprocesadas con el método de referencia promedio común (CAR, por sus siglas en inglés). Este método tiene como fin mejorar la relación señal a ruido de la señal de EEG. Básicamente, se busca quitar todo aquello que es común en todas las lecturas simultáneas de los electrodos. La CAR puede ser calculada mediante la siguiente fórmula:

$$V_i^{CAR} = V_i^{ER} - \frac{1}{n}\sum_{j=1}^{n} V_j^{ER} \qquad (1)$$

donde V_i^{ER} es el potencial entre el i-ésimo electrodo y la referencia, y n es el número de electrodos en el montaje.

Extracción de características

Transformada Wavelet Discreta

En [19] se menciona que las características utilizadas en las BCI son no estacionarias ya que las señales de EEG pueden rápidamente variar con el tiempo. Además, estas características deben contener información del tiempo debido a que los patrones de actividad cerebral están generalmente relacionados a variaciones específicas del EEG en el tiempo. Lo anterior, hace necesaria una representación que considere eso.

Una técnica que permite modelar dichas variaciones, en el dominio tiempo-escala, es la transformada wavelet discreta (DWT, por su siglas en inglés). La DWT provee una representación wavelet altamente eficiente mediante la restricción de la variación en la traslación y la escala, usualmente a potencias de dos. En ese caso, la DWT es algunas veces llamada transformada wavelet diádica. La DWT se define mediante la siguiente ecuación [20]:

$$W(j,k) = \sum_j \sum_k f(x) 2^{-j/2} \psi(2^{-j}x - k). \qquad (2)$$

El conjunto de funciones $\psi_{j,k}(n)$ es referido como la familia de wavelets derivadas de $\psi(n)$, el cual es una función de tiempo con energía finita y rápido decaimiento llamada la wavelet madre. Las bases del espacio wavelet corresponden entonces, a las funciones ortonormales obtenidas de la wavelet madre después de las operaciones de escala y traslación. La definición indica la proyección de la señal de entrada en el espacio wavelet a través del producto interior, entonces, la función $f(x)$ puede ser representada en la forma:

$$f(x) = \sum_{j,k} d_j(k)\psi_{j,k} \qquad (3)$$

donde $d_j(k)$ son los coeficientes wavelet en el nivel j. Los coeficientes en diferentes niveles pueden ser vistos a través de la proyección de la señal en la familia de wavelets como:

$$\langle f, \psi_{j,k} \rangle = \sum_l d_l \langle f, \phi_{j,k+l} \rangle \qquad (4)$$

$$\langle f, \phi_{j,k} \rangle = \frac{1}{\sqrt{2}} \sum_l c_l \langle f, \phi_{j-1,2k+l} \rangle \quad (5)$$

El análisis DWT puede ser realizado usando un algoritmo piramidal rápido descrito en términos de bancos de filtros multi-tasa, es decir, aquellos donde se tiene más de una tasa de muestreo realizando conversiones mediante las operaciones de decimación e interpolación. La DWT puede ser vista como un banco de filtros con espacio de una octava entre ellos. Cada sub-banda contiene la mitad de las muestras de la frecuencia de la sub-banda vecina más alta. En el algoritmo piramidal la señal es analizada en diferentes bandas de frecuencias con diferentes resoluciones mediante la descomposición de la señal en una aproximación burda e información detallada. La aproximación burda es entonces adicionalmente descompuesta usando el mismo paso de descomposición wavelet. Esto se logra mediante un filtrado sucesivo de pasa-bajas y pasa-altas de la señal de tiempo, y un sub-muestreo por dos como se define en la siguientes formulas [21]:

$$a_j(k) = \sum_m \overline{h}(m-2k) a_{j+1}(m) \quad (6)$$

$$d_j(k) = \sum_m \overline{g}(m-2k) a_{j+1}(m) \quad (7)$$

Las señales $a_j(k)$ y $d_j(k)$ son conocidas como los coeficientes de aproximación y detalle, respectivamente. Este proceso puede ser ejecutado iterativamente formando un árbol de descomposición wavelet hasta algún nivel de resolución deseado.

En el presente trabajo se aplica la transformada wavelet discreta con 5 niveles de descomposición, utilizando como wavelet madre a una Daubechies de segundo orden (db2). El número de niveles de descomposición se selecciona a partir del número de máximo de niveles de descomposición que pueden ser obtenidos con la ventana de menor tamaño de todas las disponibles. Con lo anterior, se obtiene un vector de coeficientes wavelet para cada una de las ventanas en cada uno de los canales de interés. La Tabla 1 muestra los niveles de descomposición y los rangos de frecuencia en cada nivel. Esta transformación permite seguir teniendo hasta cierto grado una interpretación de las características en función de los rangos de frecuencia.

Tabla 1: Descripción de los niveles de descomposición de la DWT.

Nivel	Rango de frecuencias (Hz)
D1	32-64
D2	16-32
D3	8-16
D4	4-8
D5	2-4
A5	0-2

Como es evidente, el número de coeficientes wavelet en cada uno de los niveles variará dependiendo del tamaño de la señal de EEG delimitada entre los marcadores. Esto debido a que, de manera similar al habla convencional, la duración de las ventanas de pronunciación imaginada de una palabra es variable tanto en ventanas de un sólo individuo como en ventanas de individuos distintos. Para tratar con este problema, los coeficientes wavelets son normalizados mediante la energía relativa wavelet que se describe a continuación.

Energía wavelet relativa

Con la DWT aplicada sobre una señal usando una wavelet madre ψ y un número de niveles de descomposición N se obtiene un conjunto de coeficientes de detalle ($d_{j,k}$; para $j = 1, \ldots, N$) y un conjunto de coeficientes de aproximación $a_{N,k}$ a la que se denotará como los coeficientes a_k en el nivel $N+1$ con el objetivo de simplificar la notación. A partir de estos coeficientes es posible calcular la energía relativa wavelet (RWE, por sus siglas en inglés) en cada uno de los niveles de descomposición. La energía relativa wavelet para j-ésimo nivel de descomposición se define como:

$$RWE_j = \frac{E_j}{E_{total}}; \quad \text{Para } j = 1, \ldots, N+1; \quad (8)$$

donde E_j representa la energía en el j-ésimo nivel de descomposición y E_{total} representa la energía total de los coeficientes wavelet de una

señal dada. La energía en el j-ésimo nivel de descomposición E_j se define como:

$$E_j = \begin{cases} \sum_k |d_{j,k}|^2; & \text{Si } j <= N; \\ \sum_k |a_k|^2; & \text{en caso contrario.} \end{cases} \quad (9)$$

$$\text{Para } j = 1, \ldots, N+1.$$

Mientras que, la energía total E_{total} se calcula como sigue:

$$E_{total} = \sum_{j=1}^{N+1} E_j. \quad (10)$$

Claramente, $\sum_j RWE_j = 1$ y la distribución RWE_j puede ser considerada como una densidad tiempo-escala. Esto provee información para caracterizar la distribución de energía de la señal en diferentes bandas de frecuencia.

A partir de la descripción anterior, cada ventana de habla imaginada de las señales de EEG se representa mediante un conjunto de 6 valores que representan las energías wavelet de cada uno de los niveles de descomposición(D1-D5 y A5) con respecto a la energía wavelet total. Con lo anterior, se beneficia una independencia del tamaño de la ventana de señal de EEG.

Selección de características

El problema de la selección de características implica seleccionar un mínimo subconjunto, con M características, $S = (S_1, ..., S_M)$ del conjunto de características original $F = (F_1, \cdots, F_N)$, donde $M \leq N$ y $S \subseteq F$, de manera que el espacio de características sea óptimamente reducido y el desempeño de la clasificación sea mejorada o no se degrade significativamente [22].

En esta etapa, el subconjunto mínimo de características se selecciona con base en el trabajo descrito en [12], donde se menciona que las frecuencias de la señal de EEG que son mayores a 25 Hz. están más relacionadas a actividad electromiográfica (EMG). De aquí y de acuerdo a la Tabla 1 se puede observar que los rangos de frecuencias de los coeficientes en el nivel de detalle (D1) exceden dicha frecuencia, por lo cual los valores de energía relativa wavelet obtenidos a partir de dichos coeficientes se descartan. Por lo tanto, el subconjunto de características seleccionado se compone de los valores de energía relativa wavelet obtenidos a partir de los coeficientes de detalle (D2-D5) y el de aproximación (A5) con lo que se reduce la dimensión de los vectores de características y se busca reducir el impacto del problema de la dimensionalidad (*curse of dimensionality*) en la etapa de clasificación. Con lo anterior, cada una de las ventanas de cada uno de los canales está representada con 5 valores de energía relativa wavelet.

Clasificación

De acuerdo con [23], la clasificación cubre cualquier contexto en el que alguna decisión o pronóstico es hecho sobre la base de información histórica disponible. Esta base de información disponible D se define de la siguiente forma:

$$D = \{\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle, \cdots, \langle x_m, y_m \rangle\} = \langle X, Y \rangle \quad (11)$$

donde los valores $x_i \in X$ son típicamente vectores multi-dimensionales de la forma: $x_i = \{z_1, z_2, \cdots, z_n\}$ cuyos elementos pueden tomar valores reales o discretos. Estos componentes se denominan atributos (o características). El objetivo es inferir una función (o relación) f.

$$f : X \rightarrow Y. \quad (12)$$

donde los valores de Y están contenidos en un conjunto finito de clases $C = \{C_1, ..., C_k\}$ que caracterizan los datos dados. Los modelos aprendidos de los datos de entrenamiento son, entonces, evaluados con un conjunto de prueba distinto para determinar si los modelos pueden ser generalizados a nuevos casos [24].

En el presente trabajo se entrenan y prueban los siguientes tres clasificadores de naturaleza diversa: Naive Bayes (NB), Random Forest (RF) y Máquina de vectores de soporte (SVM, por sus siglas en inglés). En las siguiente secciones se describe brevemente sus principales características.

Algoritmo 1 Generación del Random Forest

Requerir: IDT (un árbol de decisión), T (el número de iteraciones), S (el conjunto de entrenamiento), μ (el tamaño de la submuestra), N (número de atributos usados en cada nodo)
Asegurar: M_t; $t = 1, \ldots, T$
for $t \leftarrow 1$ hasta T **do**
  $S_t \leftarrow$ muestra con μ instancias de S con remplazo
  Construir el clasificador M_t usando $IDT(N)$ en S_t.
end for

Algoritmo 2 Clasificación del RF

Requerir: $x \in X$ (una instancia x del conjunto de instancias de prueba X), $|dom(y)|$ (el número de clases del problema en cuestión), M_i (el i-ésimo árbol de decisión generado con el algoritmo 1), y T (el número de árboles miembros de RF).
Asegurar: C (predicción de la clase)
$Contador_1, \ldots, Contador_{|dom(y)|} \leftarrow 0$ {inicializar los contadores de votos para las clases}
for $i \leftarrow 1$ hasta T **do**
  $voto_i \leftarrow M_i(x)$ {obtener la clase predicha por el clasificador M_i para la instancia x}
  $Contador_{voto_i} \leftarrow Contador_{voto_i} + 1$ { incrementar uno al contador de la clase correspondiente}
end for
$C \leftarrow$ la clase con el mayor número de votos.
Retornar C.

Naive Bayes

Aunque simples, frecuentemente exhiben alta exactitud en la clasificación, comparable en rendimiento con los mejores árboles de decisión y redes neuronales. Está basado en las probabilidades determinadas de los datos, los nuevos objetos puedes ser determinados para pertenecer a las clases con diversos grados de probabilidad [24].

Naive Bayes asume que los atributos son independientes, es decir, el efecto de un atributo sobre una clase específica es independiente de los valores de los otros atributos (variables o características). Además, Naive Bayes está basado en el teorema de Bayes. En este trabajo se utiliza la implementación de Naive Bayes desarrollada en Weka versión 3.6.8.

Máquinas de vectores de soporte

Una SVM utiliza un hiperplano discriminante para identificar las clases. Sin embargo, en el caso de SVM, el hiperplano seleccionado es el que maximiza los márgenes, es decir, la distancia entre los puntos de entrenamiento más cercanos. Maximizar los márgenes se sabe que aumenta la capacidades de generalización [25, 26]. Además, SVM utiliza un parámetro de regularización C que es capaz de acomodar valores atípicos (*outliers*) y permite errores en el conjunto de entrenamiento.

Un SVM que permite la clasificación utilizando fronteras de decisión lineal es conocido como SVM lineal. En este trabajo se utiliza la versión de SVM lineal desarrollada en Weka versión 3.6.8 etiquetada como SMO. Sin embargo, es posible crear límites no lineales de decisión, con un aumento de la complejidad del clasificador, utilizando el "truco del *kernel*". Éste consiste en la transformación de los datos a otro espacio, generalmente de dimensión mayor, usando una función núcleo o *kernel*.

SVM ha sido aplicado a problemas multiclase utilizando la estrategia conocida como OVR (*One Versus the Rest*), que consiste en separar cada clase de las otras [27]. También, es importante mencionar que, SVM tiene muchas ventajas como: buenas propiedades de generalización, es insensible al sobre-entrenamiento y al problema de la dimensionalidad. Además, necesita que pocos hiper-parámetros sean definidos manualmente [28, 25, 26].

Random Forest

Random Forest (RF) es una combinación de árboles predictores tal que cada uno de los árboles depende de los valores de un vector aleatorio muestreado independientemente y con

la misma distribución para todo los árboles en el bosque. Cada árbol arroja un único voto para la clase más popular para una entrada x dada, y al final la salida de RF se realiza usando voto mayoritario. Los árboles individuales son construidos usando el Algoritmo 1. Mientras que, el error de generalización por bosques converge casi seguramente a un límite cuando el número de árboles en el bosque llega a ser grande [29].

De acuerdo con [30], en el algoritmo 1, IDT representa un árbol de decisión con las siguientes modificaciones: el árbol de decisión no se poda, y en cada nodo, en vez de seleccionar la mejor división entre todos los atributos, el inductor de manera aleatoria muestrea N atributos y selecciona entre ellos la mejor división.

Después de que RF ha sido generado, éste puede ser usado para clasificar una nueva instancia x perteneciente a un conjunto de instancias de prueba X. Para ello, cada árbol miembro de RF retorna la predicción de la clase para la instancia desconocida x. Al final de este proceso, RF devuelve la clase con el mayor número de predicciones. Esto es conocido como voto mayoritario. Lo anterior se puede resumir en el Algoritmo 2.

Algunas de las características de Random Forest son: su rapidez, y su capacidad para manejar, fácilmente, un gran número de atributos de entrada. En el presente trabajo se utilizaron los siguientes hiper-parámetros para la implementación del clasificador en Weka 3.6.8: el número de árboles $T = 50$, el número de atributos considerados en cada nodo es $N = log_2(numeroCaracteristicas) + 1$, el árbol IDT base es un random tree, y el tamaño de μ es igual al tamaño del conjunto de entrenamiento S. Es importante mencionar que se realizaron experimentos con 10 [2], 50, 100, 500, 1000 y 5000 árboles en el bosque. Sin embargo, con 50 árboles se obtuvo el mejor balance entre el número de árboles y la exactitud obtenida por lo que con este hiper-parámetro fueron calculados los porcentajes de exactitud de la Tabla 3.

EXPERIMENTACIÓN Y RESULTADOS

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. A cada uno se le registraron 33 épocas de cómo imaginan la pronunciación de cada una de las cinco palabras ("arriba", "abajo", "izquierda", "derecha","seleccionar").

En el primer experimento se busca evaluar si el habla imaginada puede ser identificada, con tasas de exactitud promedio arriba del azar, independientemente del clasificador elegido. Para ello, las épocas registradas pasan por las siguientes etapas de la metodología: pre-procesamiento, extracción de características, y selección de características. Posteriormente, se concatenan los coeficientes de los cuatro canales de interés en el orden F7-FC5-T7-P7 que están en el mismo intervalo de tiempo, es decir la misma época. Con lo anterior, cada época está descrita por un vector que consta de 20 características que representa los valores de energía relativa wavelet provenientes de los canales de interés más su etiqueta de clase. Por último, se utilizan los datos de cada uno de los individuos de manera separada para entrenar y probar a tres clasificadores de distinta naturaleza (RF, SVM, y NB).

La medida para evaluar el desempeño de los clasificadores es el porcentaje de exactitud que se define como el número de épocas correctamente clasificadas entre el número de épocas presentadas al clasificador. Los porcentajes de exactitud se obtienen mediante validación cruzada de 10 particiones. Ésta se realiza dividiendo el corpus de datos disponible en diez particiones. Posteriormente, se utilizan nueve particiones para entrenar y una para probar al clasificador. Lo anterior se repite diez veces, de tal manera que cada una de las diez particiones se utilice únicamente una vez para probar. Los diez resultados obtenidos son promediados para dar una estimación de la exactitud. En la Tabla 2 se pueden observar dichos porcentajes.

[2] El número de árboles por defecto es 10. Este número fue utilizado para obtener los porcentajes de exactitud de RF en la Tabla 2 para una comparación justa.

Tabla 2: Porcentajes de exactitud (± desviación estándar) obtenidos por los clasificadores RF, NB y SVM después de aplicar validación cruzada de diez particiones usando únicamente los 4 canales de interés

Individuo	RF		SVM		NB	
S1	**60.76**	±10.11	42.66	±5.19	48.4	±12.91
S2	**32.26**	±13.44	22.56	±6.86	28.83	±8.62
S3	**36.52**	±8.69	26.92	±8.07	32.41	±9.53
S4	**42.18**	±11.65	28.78	±9.09	22.69	±8.55
S5	56.44	±12.75	**59.2**	±13.46	47.08	±8.33
S6	32.26	±9.42	**33.72**	±11.24	33.42	±9.26
S7	43.09	±9.74	**44.6**	±14.16	35.96	±11.24
S8	**61.28**	±9.09	43.85	±6.87	49.59	±7.77
S9	**54.85**	±14.99	40.54	±11.03	39.54	±11.02
S10	32.55	±11.46	**34.97**	±8	29.72	±9.02
S11	**69.49**	±6.56	45.56	±7.65	50.51	±12.37
S12	**44.66**	±11.35	43.79	±12.97	38.61	±8.31
S13	**49.36**	±10.49	44.13	±9.05	40.1	±10.91
S14	30	±8.55	**32.74**	±9.9	29.23	±7.72
S15	54.89	±11.11	**56.66**	±11.36	46.51	±7.29
S16	**40**	±17.3	39.83	±11.68	28.44	±10.38
S17	45.4	±12.9	42.49	±15.85	**46.08**	±14.93
S18	**49.94**	±12.41	21.84	±8.63	33.63	±9.29
S19	28.88	±9.99	**29.43**	±10.84	27	±9.84
S20	**70.95**	±8.18	64.35	±11.53	58.06	±9.03
S21	27.1	±11.86	**29.94**	±11.5	25.95	±9.15
S22	**55.14**	±12.54	52.59	±9.29	51.56	±6.57
S23	**40.98**	±11.81	33.4	±9.28	30.12	±8.43
S24	33.68	±9.27	33.97	±10.76	**34**	±11.58
S25	26.88	±13.58	29.74	±11.59	**30.11**	±9.25
S26	**36.22**	±12.06	29.95	±6.95	31.12	±8.62
S27	**43.85**	±11.9	42.42	±7.9	40.11	±6.53
Promedio	**44.43**	±11.23	38.91	±10.03	37.36	±9.5

La Tabla 2 muestra que, generalmente, los porcentajes de exactitud obtenidos por los tres clasificadores se encuentran por encima del azar para cinco clases, el cual es del 20 %. Este porcentaje de exactitud se toma como cota inferior debido a que, de acuerdo con [31], un buen clasificador es aquel que obtiene tasas de error menores que el azar en la etapa de generalización (prueba). Asimismo, los porcentajes de exactitud promedio de la Tabla 2 permiten conjeturar que, a pesar de la complejidad inherente a la tarea, las señales de EEG registradas durante el habla imaginada contienen información que permiten su identificación inclusive independientemente del clasificador utilizado. También, en la Tabla 2 se puede observar que para la mayoría de los individuos, los mejores porcentajes de exactitud se obtuvieron para RF. Esto se vio reflejado en el desempeño promedio de RF, entre los diferentes individuos, ya que fue el mayor con un 44.43 %. Es por esta razón que RF es elegido como clasificador base para el siguiente experimento.

El segundo experimento busca evaluar si los 10 canales restantes pueden aportar información adicional para el proceso de clasificación de habla imaginada. Para evaluar lo anterior, estos diez canales fueron conjuntados con los cuatro canales más cercanos a las áreas de Wernicke y Broca. A esta unión se le denomina configuración de 14 canales. En resumen, se compara la configuración de 14 canales con respecto a la de 4.

Tabla 3: Porcentajes de exactitud (± desviación estándar) y tasa de contribución (± desviación estándar) obtenidos por el clasificador RF, con 50 árboles, después de aplicar validación cruzada de 10 particiones usando las configuraciones de 14 canales y 4 canales

Individuo	14 canales		4 canales		tasa de contribución	
S1	**79.96**	±9.12	68.49	±12.48	0.86	±0.18
S2	**49.08**	±14.34	41.07	±17.1	0.85	±0.33
S3	**63.09**	±12.16	36.95	±9.62	0.61	±0.23
S4	**57.87**	±12.42	46.25	±11.47	0.85	±0.3
S5	**67.68**	±10.44	62.35	±8.26	0.93	±0.12
S6	**42.98**	±5.43	34.93	±10.94	0.82	±0.26
S7	**63.86**	±11.17	46.25	±7.59	0.74	±0.14
S8	**86.07**	±7.5	66.1	±5.45	0.77	±0.05
S9	**62.5**	±9.4	56.95	±12.05	0.92	±0.19
S10	**60.96**	±11.94	36.43	±9.95	0.62	±0.22
S11	**83.6**	±9.09	71.43	±6.14	0.86	±0.12
S12	**60.59**	±13.73	51.73	±11.02	0.89	±0.27
S13	**65.62**	±12.83	46.76	±9.4	0.73	±0.16
S14	**43.2**	±11.86	32.94	±13.9	0.8	±0.35
S15	60.55	±9.36	**61.8**	±12.25	1.04	±0.27
S16	**50.99**	±12.85	40.15	±15.76	0.78	±0.2
S17	**67.32**	±8.07	48.53	±12.28	0.72	±0.15
S18	**71.87**	±14.67	55.7	±9.15	0.79	±0.14
S19	**51.51**	±11.14	30.18	±8.42	0.59	±0.15
S20	**76.32**	±6.23	72.87	±12.22	0.96	±0.16
S21	**36.8**	±12.86	27.79	±6.6	0.81	±0.21
S22	**65.33**	±9.81	52.06	±9.66	0.81	±0.13
S23	**54.08**	±16.93	49.85	±11.57	0.98	±0.32
S24	**45.55**	±12.24	38.86	±8.84	0.93	±0.39
S25	**43.64**	±9.6	27.43	±13.83	0.63	±0.3
S26	**52.68**	±11.2	39.52	±15.54	0.75	±0.27
S27	**59.15**	±8.66	50.77	±9.85	0.87	±0.16
Promedio	**60.11**	±10.93	47.93	±10.79	0.79	±0.21

Para tal fin, además de presentar los porcentajes de exactitud de RF (con 50 árboles) para ambas configuraciones; se introduce la tasa de contribución que es un estimado de cuánto aportan los cuatro canales en la exactitud obtenida usando los 14 canales. Esta medida puede ser mayor a uno, igual a uno o menor a uno, según la exactitud obtenida con cuatro canales sea mayor, igual o menor que usando catorce, respectivamente. La tasa de contribución se calcula como sigue:

$$tasa_{contribucion} = \frac{exactitud_{4canales}}{exactitud_{14canales}} \qquad (13)$$

Para obtener los vectores de características para la configuración de 14 canales, los datos del corpus pasan por las mismas etapas que en el primer experimento. Sin embargo, la diferencia radica en que, en este experimento se usan los valores de energía relativa wavelet obtenidos a partir de señales de EEG en un mismo intervalo de tiempo provenientes de los 14 canales. Estos valores de energía son concatenados de la siguiente manera: AF3-F7-F3-FC5-T7-P7-O1-O2-P8-T8-FC6-F4-F8-AF4. Con lo que resulta un vector con 70 valores de energía relativa wavelet más su respectiva etiqueta de clase.

La Tabla 3 presenta la tasa de contribución y los porcentajes de exactitud del clasificador RF cuando se utilizan las configuraciones de 14 y 4 canales, después de aplicar validación cruzada de 10 particiones. En ella se aprecia que, el

clasificador RF tiene mejor desempeño usando la configuración de 14 canales salvo el caso del individuo S20 (aunque la diferencia es mínima), y para ambas configuraciones los porcentajes de exactitud se mantienen por encima del azar para cinco clases. Sin embargo, es importante recalcar que la tasa de contribución para todos los individuos está por arriba de 0.5 lo que quiere decir que la información proveniente de los cuatro canales de interés representa cuando menos el 50 % de la clasificación usando todos los catorce canales.

Por último, es importante mencionar que el presente trabajo obtiene resultados comparables a trabajos previos. En el caso de RF usando la configuración de todos los canales, se obtuvo una exactitud promedio de 60.11 % que supera a lo descrito en [11] y en [8] donde se reportan exactitudes de 47.24 % y el 45.50 %, respectivamente. Inclusive con RF usando únicamente cuatro canales, se logra en promedio una exactitud de 47.93 % que supera a lo reportando en ambos trabajos para cinco palabras. La comparación es bajo reserva puesto que no se tuvo acceso a las señales de EEG usadas ni por [11], ni por [8]. Esto hace que existan diferencias en los datos procesados, tales como: el protocolo de adquisición del habla imaginada registrada con EEG, el vocabulario reducido, el número de individuos, el número de repeticiones por palabra, la relación entre el idioma del vocabulario y el de los individuos, y el número de canales procesados.

Sin embargo, la principal diferencia con respecto a ambos trabajos radica en que ellos procesan a las señales de EEG con modelos de reconocimiento de habla convencional, asumiendo que hallarán los mismos patrones de actividad en ambas señales. Esto a pesar de que las señales de habla y EEG son de diferente naturaleza.

DISCUSIÓN Y CONCLUSIONES

En la búsqueda de una fuente electrofisiológica alternativa para controlar BCIs, se exploró la posibilidad de usar para tal fin al habla imaginada. Ésta tiene como ventajas el ser más intuitiva y multi-clase (más de dos opciones), permitiría la comunicación sin necesidad de una traducción al dominio del lenguaje [1], podría ser útil cuando la corteza motora tenga lesiones [32], y no requiere de un estímulo externo. Específicamente, en la presente investigación se desarrolló un método de procesamiento y clasificación de las señales de EEG asociadas al habla imaginada. Éste no utiliza técnicas de reconocimiento de habla convencional para tratar a la señal de EEG debido a que las características del EEG y del habla son distintas. Además, se exploró un vocabulario reducido de mayor carga semántica que los reportados en el estado del arte compuesto de las palabras del idioma español: “arriba”, “abajo”, “izquierda”, “derecha” y “seleccionar”. Esto se debe tanto al significado claro que cada una de las palabras tienen para el usuario como al hecho de que éstas reflejan órdenes con las que se podría controlar el desplazamiento y el clic del mouse. Asimismo, en búsqueda de disminuir el impacto del problema de la dimensionalidad en la etapa de clasificación, se calcularon y se seleccionaron los valores de energía relativa wavelet de los coeficientes wavelet en los niveles de detalle (D2, D3, D4, D5 y D6), y en nivel de aproximación (A6); con lo que se obtiene una representación compacta de cada una de las épocas registradas de los individuos durante la pronunciación imaginada de una palabra.

Actualmente, las principales fuentes electrofisiológicas aprovechadas para la realización de sistemas BCI (P300, VEP, imágenes motoras y SCP) pueden proporcionar porcentajes de reconocimiento elevados [33, 34, 35]; específicamente, cuando se trata de decodificar dos intenciones -por ejemplo, los movimientos imaginados de la mano derecha e izquierda-. Esos porcentajes son mayores a los porcentajes obtenidos en el presente estudio utilizando habla imaginada, los cuales aún distan de la situación ideal para llevar el trabajo a un nivel de aplicación práctico. Sin embargo, es importante recalcar que los porcentajes obtenidos por los clasificadores (RF, SVM, NB) utilizando habla imaginada registrada únicamente de los canales F7-FC5-T7-P7, en todos los individuos, son superiores al 20 %, es

decir, están arriba del azar para cinco clases (ver Tabla 2). Esto es especialmente interesante si se considera el hecho de haber utilizado palabras semánticamente similares. Con base en todo lo anterior, se puede conjeturar que a pesar de la complejidad inherente a la tarea, las señales de EEG registradas durante el habla imaginada podrían contener información para ser utilizada en tareas de clasificación de palabras de un vocabulario reducido. Actualmente se está realizando un análisis de las características utilizadas por el método con la intención de ampliar nuestro conocimiento sobre cuáles de ellas son las más apropiadas para la correcta clasificación del habla imaginada.

Por otra parte, en la Tabla 3 se puede observar que los canales más relacionados con el habla imaginada (F7-FC5-T7-P7) aportan siempre más del 50 % de la clasificación obtenida usando la configuración de 14 canales. Sin embargo, también se puede observar que para prácticamente todos los individuos es mejor utilizar las señales registradas de todos los canales. De aquí es importante remarcar que el porcentaje de exactitud promedio utilizando la configuración de 14 canales es de 60.11 %, es decir, tres veces mayor al azar para cinco clases. Esta mejoría en las exactitudes entre la configuración de 14 canales con respecto a la de 4 canales permite conjeturar que los demás canales contribuyen al proceso de reconocimiento del habla imaginada. Lo anterior está en sintonía con modelos cerebrales más recientes que discuten la contribución de regiones del hemisferio izquierdo fuera de las áreas de Wernicke y Broca a la generación del lenguaje [36]. Asimismo, en [37] se sugiere que el habla imaginada podría generarse en la corteza auditiva o en el área de Wernicke, si se toma como cierto el hecho de que se produce en la corteza auditiva, en [38] se ha mostrado que existe actividad en está región asociada a la percepción de lenguaje en ambos hemisferios con lo que se justificaría el uso de algunos canales del hemisferio derecho del cerebro.

No obstante, el usar todos los canales no necesariamente implica mayor información para obtener el mejor desempeño posible; por lo que, queda por explorar cuál es la mejor combinación de canales, dentro del espacio de $2^{14} - 1$ posibles combinaciones, que permita obtener el mejor porcentaje de exactitud utilizando el menor número de canales posible. Esto con el objetivo de mitigar el costo de procesar la información proveniente de todos los canales. El hallar la solución que minimice el error (diferencia entre el total y la exactitud) y el número de canales es considerado por los investigadores de computación como un problema NP-difícil, es decir, el tipo de problemas que no son de decisión para los cuales no se conoce un algoritmo "eficiente" que resuelva el problema en un tiempo polinomial. Para tratar de solventar estas limitaciones existen algunas heurísticas bio-inspiradas como los algoritmos genéticos que permiten explorar el espacio de búsqueda y obtener buenas soluciones, es decir, aquellas donde el conjunto de canales seleccionados sea el menor posible, y la exactitud mejora o se degrada lo menos posible. Asimismo, la calidad de las soluciones puede verificarse en función de qué tan cercanos están los canales seleccionados a las áreas del lenguaje.

Por último, en todas las etapas existe la posibilidad de trabajo futuro. En la etapa de preprocesamiento se podría aplicar Análisis de Componentes Independientes (ICA, por sus siglas en inglés) y evaluar cada componente independiente usando el coeficiente Hurst. Con lo anterior, de acuerdo a la evidencia experimental, es posible eliminar artefactos como latidos del corazón y parpadeos. En esta misma etapa, también se podría aplicar Análisis de Componentes Principales (PCA, por sus siglas en inglés) o selección de canales por métodos de envoltura basados en algoritmos evolutivos. En lo que a extracción de características se refiere, seleccionar a otra familia wavelet (Symlet, Mexican Hat, Morlet entre otras) como madre así como variar el orden de la wavelet también podrían ayudar. Otra acción en la misma etapa podría ser, usar otras representaciones de la señal de EEG - por ejemplo, los coeficientes auto-regresivos, coeficientes cepstrales en las frecuencias de Mel (MFCC), o codificadores de predicción lineal (LPC) - y combinarlos con los coeficientes DWT. En la clasificación, seleccionar de forma automática los hiper-parámetros óptimos de los

clasificadores utilizados es un tema importante a tratar. Asimismo, la aplicación de métodos híbridos del área de inteligencia computacional, como sistemas neuro-difusos o fuzzy-genéticos, etc. puede también ayudar a mejorar la exactitud final.

Agradecimientos

Los autores agradecen al pueblo de México que, mediante el Consejo de Ciencia y Tecnología (CONACyT), apoyó la investigación con la beca # 329011. Asimismo, al INAOE por el apoyo brindado para realizar este trabajo.

REFERENCIAS

1. Brumberg J. S., Nieto-Castanon A., Kennedy P. R. and Guenther F. H., "Brain–computer interfaces for speech communication," *Speech Communication*, no. 52, p. 367–379, 2010.

2. Wolpaw J., Birbaumer N., McFarland D., Pfurtscheller G. and Vaughan T., "Brain-computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

3. Sánchez de la Rosa J. L., *Métodos para el procesamiento y análisis estadístico multivariante de señales multicanal: aplicación al estudio del EEG.* PhD thesis, Universidad de La Laguna, España, 1993.

4. Bashashati A., Fatourechi M., Ward R. and Birch G., "A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals," *Journal of Neural engineering*, vol. 4, pp. R32–R57, 2007.

5. Pfurtscheller G., "Brain-computer interfaces: State of the art and future prospects," in *Proceedings of the 12th European Signal Processing Conference: EUROSIPCO 04*, pp. 509–510, 2004.

6. Denby B., Schultz T., Honda K., Hueber T., Gilbert J. and Brumberg J., "Silent speech interfaces," *Speech Communication*, vol. 52, no. 4, pp. 270–287, 2010.

7. Calliess J., "Further investigations on unspoken speech," Master's thesis, Institut für Theoretische Informatik Universität Karlsruhe (TH), Karlsruhe, Germany, 2006.

8. Porbadnigk A. and Schultz T., "EEG-based Speech Recognition: Impact of Experimental Design on Performance," Master's thesis, Institut für Theoretische Informatik Universität Karlsruhe (TH), Karlsruhe, Germany, 2008.

9. Suppes P., Lu Z. and Han B., "Brain wave recognition of words," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 26, p. 14965, 1997.

10. Wand M., "Wavelet-based Preprocessing of Electroencephalographic and Electromyographic Signals for Speech Recognition," *Studienarbeit Lehrstuhl Prof. Waibel Interactive Systems Laboratories Carnegie Mellon University, Pittsburgh, PA, USA and Institut für Theoretische Informatik Universität Karlsruhe (TH), Karlsruhe, Germany*, 2007.

11. Wester M. and Schultz T., "Unspoken Speech - Speech Recognition Based On Electroencephalography" Master's thesis, Institut für Theoretische Informatik Universität Karlsruhe (TH), Karlsruhe, Germany, 2006.

12. Brigham K. and Kumar B., "Imagined Speech Classification with EEG Signals for Silent Communication: A Preliminary Investigation into Synthetic Telepathy" in *Bioinformatics and Biomedical Engineering (iCBBE), 2010 4th International Conference on*, pp. 1–4, IEEE, 2010.

13. DaSalla C. S., Kambara H., Koike Y. and Sato M., "Spatial filtering and single-trial classification of EEG during vowel speech imagery" in *i-CREATe '09: Proceedings of the 3rd International Convention on Rehabilitation Engineering & Assistive*

Technology, (New York, NY, USA), pp. 1–4, ACM, 2009.

14. D'Zmura M., Deng S., Lappas T., Thorpe S. and Srinivasan R., "Toward EEG sensing of imagined speech" *Human-Computer Interaction. New Trends*, pp. 40–48, 2009.

15. Torres-García A. A., Reyes-García C. A. and Villaseñor-Pineda L., "Hacia la clasificación de habla no pronunciada mediante electroencefalogramas (EEG)" in *XXXIV Congreso Nacional de Ingeniería Biomédica*, pp. 9–12, 2011.

16. Torres-García A. A., Reyes-García C. A. and Villaseñor-Pineda L., "Toward a silent speech interface based on unspoken speech" in *BIOSTEC - BIOSIGNALS* (S. V. Huffel, C. M. B. A. Correia, A. L. N. Fred, and H. Gamboa, eds.), pp. 370–373, SciTePress, 2012.

17. Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., and Witten I., "The weka data mining software: an update" *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

18. Geschwind N., "Language and the brain.," *Scientific American*, 1972.

19. Lotte F., Congedo M., Lécuyer A., Lamarche F. and Arnald B., "A review of classication algorithms for EEG-based brain-computer interfaces" *Journal of Neural Engineering*, vol. 4, pp. r1–r13, 2007.

20. Priestley M., "Wavelets and time-dependent spectral analysis," *Journal of Time Series Analysis*, vol. 17, no. 1, pp. 85–103, 2008.

21. Pinsky M., *Introduction to Fourier analysis and wavelets*, vol. 102. Amer Mathematical Society, 2002.

22. Zhu Z., Jia S. and Ji Z., "Towards a Memetic Feature Selection Paradigm [Application Notes]," *Computational Intelligence Magazine, IEEE*, vol. 5, no. 2, pp. 41–53, 2010.

23. Michie D., Spiegelhalter D. and Taylor C., *Machine Learning, Neural and Statistical Classification.* 1994.

24. Jensen R. and Shen Q., *Computational intelligence and feature selection: rough and fuzzy approaches.* IEEE Press Series On Computational Intelligence, 2008.

25. Burges C., "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.

26. Bennett K. and Campbell C., "Support vector machines: hype or hallelujah?," *ACM SIGKDD Explorations Newsletter*, vol. 2, no. 2, pp. 1–13, 2000.

27. Schlögl A., Lee F., Bischof H. and Pfurtscheller G., "Characterization of four-class motor imagery EEG data for the BCI-competition 2005," *Journal of Neural Engineering*, vol. 2, p. L14, 2005.

28. Jain A., Duin R. and Mao J., "Statistical pattern recognition: A review," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 1, pp. 4–37, 2000.

29. Breiman L., "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

30. Rokach L., *Pattern Classification Using Ensemble Methods.* World Scientific, 2009.

31. Dietterich T., "Ensemble methods in machine learning," *Multiple classifier systems*, pp. 1–15, 2000.

32. Lal T. N., Schroder M., Hinterberger T., Weston J., Bogdan M., Birbaumer N., and Scholkopf B., "Support vector channel selection in BCI," *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 6, pp. 1003–1010, 2004.

33. Guger C., Edlinger G., Harkam W., Niedermayer I., and Pfurtscheller G., "How many people are able to operate an EEG-based brain-computer interface (BCI)?," *Neural Systems and Rehabilitation*

Engineering, IEEE Transactions on, vol. 11, no. 2, pp. 145–147, 2003.

34. Allison B., Luth T., Valbuena D., Teymourian A., Volosyak I., and Graser A., "BCI Demographics: How many (and what kinds of) people can use an SSVEP BCI?," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 18, no. 2, pp. 107–116, 2010.

35. Volosyak I., Valbuena D., üth T. L, Malechka T., and Gräser A., "BCI Demographics II: How many (and what kinds of) people can use an SSVEP BCI," *IEEE Trans. Neural Syst. Rehabil. Eng*, 2011.

36. Binder J., Frost J., Hammeke T., Cox R., Rao S., and Prieto T., "Human brain language areas identified by functional magnetic resonance imaging," *The Journal of Neuroscience*, vol. 17, no. 1, pp. 353–362, 1997.

37. Hesslow G., "Conscious thought as simulation of behaviour and perception," *Trends in cognitive sciences*, vol. 6, no. 6, pp. 242–247, 2002.

38. Hickok G. and Poeppel D., "The cortical organization of speech processing," *Nature Reviews Neuroscience*, vol. 8, no. 5, pp. 393–402, 2007.