ARTÍCULO DE INVESTIGACIÓN
Vol. 36, No. 3, Sep-Dic 2015, pp. 233-247

REVISTA MEXICANA DE INGENIERÍA BIOMÉDICA

dx.doi.org/10.17488/RMIB.36.3.8

Aplicación de la Sonificación de Señales Cerebrales en Clasificación Automática

E.F. González Castañeda, A.A. Torres-García, C.A. Reyes-García, L. Villaseñor-Pineda
Instituto Nacional de Astrofísica Óptica y Electrónica (INAOE)

RESUMEN

En años recientes la sonificación de electroencefalogramas (EEG) ha sido utilizada como una alternativa para analizar señales cerebrales al convertir el EEG en audio. En este trabajo se aplica la sonificación a señales de EEG durante el habla imaginada o habla no pronunciada, con el objetivo de mejorar la clasificación automática de 5 palabras del idioma español. Para comprobarlo, se procesó la señal cerebral de 27 sujetos sanos. Estas señales fueron sonificadas para después extraer características con dos métodos diferentes: la transformada Wavelet discreta (DWT); y los coeficientes cepstrales en la escala de Mel (MFCC). Éste último comúnmente utilizado en tareas de reconocimiento de voz. Para clasificar las señales se aplicaron tres algoritmos distintos de clasificación Naive Bayes (NB), Máquina de vectores de soporte (SVM) y Random Forest (RF). Se obtuvieron resultados usando los 4 canales más cercanos a las áreas de lenguaje de Broca y Wernicke, así como también los 14 canales del dispositivo EEG utilizado. Los porcentajes de exactitud promedio para los 27 sujetos en los dos conjuntos de 4 y 14 canales usando sonificación de EEG fueron de 55.83% y 64.14% respectivamente, con lo que se logró mejorar ligeramente los porcentajes de clasificación de las palabras imaginadas respecto a no utilizar sonificación.

Palabras clave: **sonificación, electroencefalogramas (EEG), habla imaginada, palabras no pronunciadas, clasificación automática.**

Correspondencia:
Erick González
INAOE Luis Enrique Erro 1 Sta. María Tonantzintla, Puebla.
Correo electrónico: erick.gonzalezc@inaoep.mx

Fecha de recepción:
8 de mayo de 2015

Fecha de aceptación:
11 de septiembre de 2015

ABSTRACT

In recent years sonification of electroencephalograms (EEG) has been used as an alternative to analyze brain signals after converting EEG to audio. In this paper we applied the sonification to EEG signals during the imagined speech or unspoken speech, with the aim of improving the automatic classification of 5 words of Spanish. To check this, the brain signals of 27 healthy subjects were processed. These sonificated signals were processed to extract features with two different methods: discrete wavelet transform (DWT); and the Mel-frequencies cepstral coefficients (MFCC). The latter commonly used in speech recognition tasks. To classify the signals three different classification algorithms Naive Bayes (NB), Support Vector Machine (SVM) and Random Forest (RF) were applied. Results were obtained using the 4 channels closest to the language areas of Broca and Wernicke, as well as the 14 channels of the EEG device used. The percentages of average accuracy for the 27 subjects in the two sets of 4 and 14 channels using EEG sonification were 55.83% and 64.14% respectively, which are improvements in the classification rates of the imagined words compared with a scheme without sonification.

Keywords: **sonification, electroencephalograms (EEG), unspoken speech, unspoken words, classification.**

INTRODUCCIÓN

La sonificación de EEG es el uso de cualquier método de sonificación para transformar la lectura de ondas cerebrales a sonidos con el objetivo de facilitar su análisis. La sonificación de EEG permite, en general, distribuir la información en un amplio rango de frecuencias audibles. La sonificación de EEG ha sido utilizada para fines diversos: el estudio de la distribución espacial de la actividad cerebral [1], el análisis de la coherencia del sonido con la actividad cerebral de la banda alfa [2], la asistencia en rehabilitación [3] e incluso en la creación de composiciones musicales [4].

Diversos investigadores han desarrollado trabajos relacionados con las técnicas y aplicaciones de la sonificación de EEG. Thomas Hermann y sus colaboradores han demostrado las ventajas y desventajas de la sonificación de EEG [5]. Dichos investigadores han presentado métodos de sonificación de EEG para mostrar la correspondencia entre las actividades neurales y cognitivas, como por ejemplo con el método de sonificación por matriz de distancias entre filas de electrodos [6]. También han propuesto los métodos de sonificación basada en eventos [7], y el método de sonificación de EEG Vocal [8], los cuales han servido para analizar a pacientes con epilepsia.

Antecedentes

Los audios de un EEG sonificado han sido utilizados para ayudar en la toma de decisiones en los diagnósticos médicos, por ejemplo el diagnóstico de la enfermedad neurológica de Alzheimer.

En [9], los audios se generan a partir de electroencefalogramas (EEG) de sujetos sanos y sujetos con etapa temprana de la enfermedad de Alzheimer. En este trabajo se aplicó una técnica de sonificación de EEG para hacer clasificación entre las distintas condiciones de los pacientes. En una primera fase, la clasificación fue realizada manualmente por personas no especializadas, al percibir la diferencia en la retroalimentación audible del EEG de los pacientes con Alzheimer contra los audios de

pacientes saludables. En esta investigación se lograron reconocer las diferencias escuchando la sonificación de los sujetos sanos y los sujetos enfermos con un 95% de exactitud. En una segunda fase, se aplicó un método de clasificación automática mediante análisis discriminante lineal (LDA) para clasificar los mismos datos y compararlos con los obtenidos por los voluntarios, con lo que se obtuvo una exactitud promedio del 90%.

Dicho trabajo obtuvo resultados excelentes en la clasificación subjetiva y automática, lo que motivó nuestro interés en comprobar si al aplicar una técnica de sonificación de EEG se puede mejorar la clasificación automática en las señales cerebrales para la tarea de habla imaginada. Algunos antecedentes en la clasificación de señales EEG de palabras no pronunciadas son los trabajos reportados en [10, 11, 12]. Esta tarea de clasificación ha sido afrontada con distintos enfoques, donde por ejemplo en el trabajo [10] se buscó clasificar 5 palabras no pronunciadas del idioma inglés usando métodos de reconocimiento de voz pero sobre el EEG sin sonificar, aun cuando las señales de EEG poseen un espectro de frecuencias menor al de una señal de audio.

De acuerdo con lo anterior, en esta investigación se aplicó un método de aprendizaje automático para evaluar si la sonificación de la señal de EEG puede apoyar a mejorar los resultados de clasificación de palabras no pronunciadas. Este es el principal objetivo tanto de este trabajo como de nuestros trabajos previos (véase [13]). No obstante, el presente trabajo extiende la experimentación al evaluar la aplicación de técnicas de extracción de características comúnmente utilizadas para reconocimiento de voz; además de usar información de un mayor número de canales (14 canales).

METODOLOGÍA

En la presente investigación se siguieron las etapas mostradas en la figura 1. Se comenzó con una base de datos de palabras no pronunciadas, después en el montaje de referencia de la señal cerebral se aplicó la referencia promedio común. Luego las señales se sonificaron usando el método de sonificación de *EEG a tonos* [14]. Después se aplicaron los métodos para extraer las características de los audios resultantes de la sonificación, en este caso la transformada Wavelet discreta (DWT) y los coeficientes cepstrales en la escala de Mel (MFCC). Posteriormente se procedió a clasificar los vectores de características de las 5 palabras usando Naive Bayes, SVM y Random Forest.

Adquisición de la señal cerebral

La base de datos de señales cerebrales utilizada fue la misma que se adquirió en el trabajo [15], la cual cuenta con los registros de 27 sujetos durante la pronunciación imaginada de 5 palabras del idioma español (*arriba, abajo, derecha, izquierda, seleccionar*). Estas señales fueron registradas utilizando el dispositivo de EEG Epoc de la compañía Emotiv, este dispositivo cuenta con 14 electrodos de lectura los cuales poseen una frecuencia de muestreo de 128 Hz. En la figura 2 se muestra la distribución de los electrodos en el dispositivo de EEG.

En la sesión de registro de señales cada sujeto imaginó la pronunciación de cada una de las palabras 33 veces, delimitando la señal de interés mediante clics. Los registros fueron obtenidos en el Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), en un sala libre de distracciones y ruido externo audible, esto para evitar que la señal cerebral tuviera información no deseada que pudiera afectar el proceso de clasificación.

Montaje de referencia

Las señales de EEG obtenidas fueron procesadas con el método de referencia promedio común (CAR) (tal como se aplicó en otros trabajos sobre habla imaginada [16, 17, 12]). Este montaje mejora la relación señal a ruido de la señal de EEG.

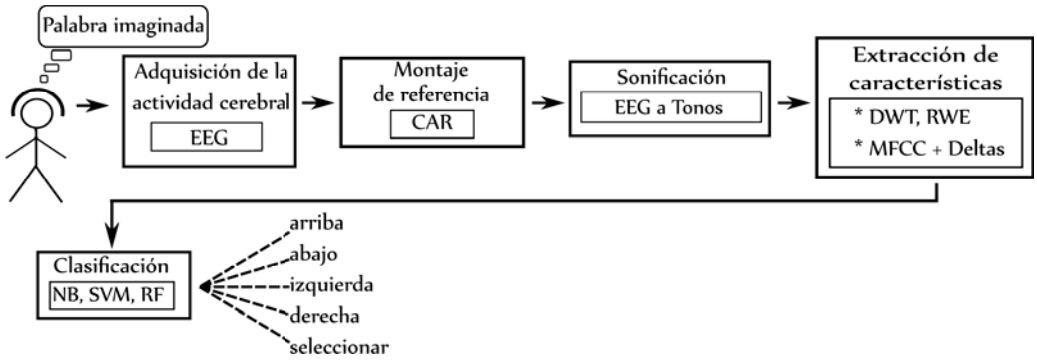


Figura 1. Metodología seguida para la clasificación de palabras no pronunciadas usando sonificación de EEG.

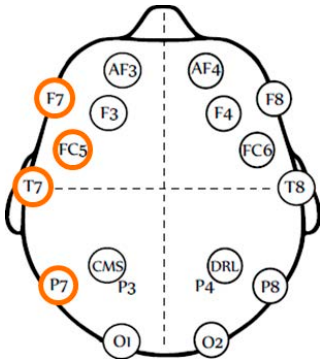


Figura 2. Localización de los electrodos en el dispositivo EPOC de Emotiv. Se resaltan en naranja los 4 electrodos seleccionados para representar las áreas que se activan durante el procesamiento de lenguaje según el modelo Geschwind-Wernicke.

La referencia promedio común busca quitar la información que es común en todas las lecturas simultáneas de los electrodos [18]. La CAR puede ser calculada mediante la resta del potencial entre cada electrodo y la referencia (el potencial promedio de todos los canales), se repite esto para cada instante de tiempo, tal como se muestra en 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 [12].

Sonificación de EEG

En el presente trabajo se utilizó la técnica de Sonificación *EEG to tones* (*EEG a tonos*) [14]. Este método genera audios basados en tonos (ondas sinusoidales con frecuencia audible) mediante la elección de amplitudes dominantes en las frecuencias de la señal cerebral. El pseudocódigo que explica el método de sonificación de EEG utilizado es mostrado en el algoritmo 1. A continuación se describe el proceso.

El algoritmo de sonificación *EEG a tonos*, toma de entrada una señal de EEG de un canal. Luego se crea un espectrograma de la señal de EEG, con el que se obtienen las amplitudes en cada frecuencia de la señal de EEG para cada segmento de tiempo.

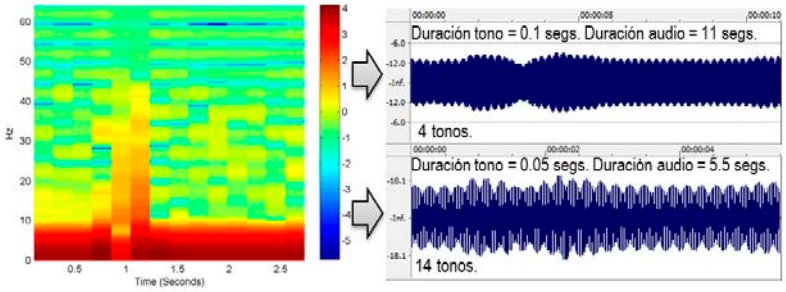


Figura 3. A la izquierda se muestra el espectrograma de una repetición de la palabra arriba. A la derecha el audio resultante de la sonificación de EEG con distinta duración de tono y distinto número de tonos.

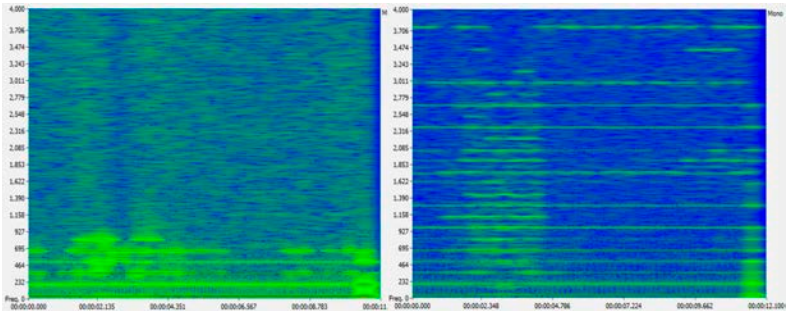


Figura 4. Espectrogramas del resultado de la sonificación de EEG correspondientes a utilizar 4 tonos (izquierda) y 14 tonos (derecha). En el audio con 14 tonos los picos de amplitud tienen mayor separación en el eje de frecuencias respecto al audio con 4 tonos, cuyas amplitudes se concentran en las bajas frecuencias.

Las amplitudes del espectrograma obtenido se re-escalan respecto a la máxima amplitud la cual pasa a representar el 100%. Las amplitudes escaladas del espectrograma se ordenan de manera descendente y se toman las de mayor valor. El número de amplitudes a tomar es el establecido por el parámetro de número de tonos a representar en el audio. Con lo anterior se crean vectores de frecuencia dominantes por cada segmento de tiempo.

A continuación se establece qué tono le corresponde a cada vector de frecuencias dominantes del espectrograma. Para ello se establecen las frecuencias mínima y máxima a representar en el archivo de audio. Luego a partir de ese rango se distribuyen proporcionalmente los tonos respecto al rango de frecuencias de la señal de EEG.

Por último, se forma el audio de salida. Para ello se recorre cada ventana del espectrograma y se verifica si existen frecuencias dominantes. De ser así, se asignan los tonos correspondientes con la duración previamente establecida en segundos. En otras palabras, se suman las señales sinusoidales de los tonos para cada segmento de tiempo.

En la figura 3 se muestra un ejemplo del resultado de la ejecución del algoritmo con dos configuraciones distintas, en el cual se muestra que al aplicar la sonificación de EEG podemos expandir o contraer la señal a partir de la modificación del parámetro de entrada de duración de tonos. Para escuchar ejemplos de la sonificacion de EEG visitar la página [1].

[1]Ejemplos de sonificación de EEG, goo.gl/NdNMft, Laboratorio de procesamiento de bioseñales y computación médica, INAOE, 2014

Algoritmo 1 Sonificación de EEG a tonos

Entradas: eeg //señal de EEG de un canal
fs //Frec. de muestreo EEG ntn //Número de tonos
fl //Frec. mínima en EEG fh //Frec. máxima en EEG
w //Longitud de ventana sf //Traslape de ventana
dur //Duración de los tonos fla //Frec. mínima del audio
fha //Frec. máxima del audio fsa //Frec. de muestreo del audio
Salida: $audio$ //Sonificación de la señal de EEG

```
//Obtener matriz de espectrograma del EEG
spec = espectrograma(eeg, w, sf, fl, fh, fs)
for c = 1 hasta columnas(spec) do
  //Escalar amplitudes respecto a la máxima amplitud
  for r = 1 hasta renglones(spec) do
    spec_{r,c} = spec_{r,c}/ amplitudMax(spec)
  end for
  //Ordenar amplitudes descendentemente
  ord_{:,c} = ordendesc(spec_{:,c})
  //Tomar las frec. de mayor amplitud por columna
  fda_{:,c} = ord_{1:ntn,c}
  //Escalar frecuencias únicas de fda a un tono audible
  for r = 1 hasta renglones(fda) do
    if unica(fda_{r,c}) then
      ftn = [(fda_{r,c} - fl)/(fh - fl)] * (fha - fla) + fla
      tn_{fda_{r,c}} = sin(2π/((fsa * ftn_i) * (fsa * dur)))
    end if
  end for
end for
//Crear audio con la suma de las señales en tonos
for c = 1 hasta columnas(fda) do
  for r = 1 hasta renglones(fda) do
    sumtn = sumtn + tn_{fda_{r,c}}
  end for
  audio = audio + sumtn
end for
return audio
```

La señal de audio generada ahora posee otras características en la distribución de las frecuencias como se puede apreciar en el espectrograma calculado a partir de las señales de audio. En los espectrogramas de la figura 4, se aprecia el impacto que tiene la cantidad de tonos en el audio de salida, pues para la sonificación con 4 tonos las frecuencias dominantes estarán con poca separación, mientras que con 14 tonos las frecuencias están mejor distribuidas.

Extracción de características

Después de aplicar el proceso de sonificación de EEG a las palabras no pronunciadas, la siguiente etapa es buscar la manera de extraer características de la señal de audio con el objetivo de que un clasificador pueda discriminar entre clases.

DWT y energía relativa

La transformada wavelet discreta (DWT, por su siglas en inglés), es una técnica que permite modelar las variaciones de las señales de EEG en el dominio tiempo-escala [19]

Una vez aplicada la DWT sobre la señal se obtienen coeficientes de aproximación y de detalle, desde los cuales es posible calcular la energía relativa wavelet. La energía relativa wavelet (RWE) representa la energía que algún nivel de descomposición aporta al total de la energía wavelet de la señal. La energía relativa provee información para caracterizar la distribución de energía de la señal en diferentes bandas de frecuencia, con lo que se obtiene independencia del tamaño de la ventana de señal de EEG o de audio, según sea el caso. En este trabajo se aplicó la RWE tal como se explica en [12].

Para N niveles de descomposición, la energía relativa wavelet para el nivel de descomposición j se define como:

$$RWE_j = \frac{E_j}{E_{total}}; \quad \text{Para } j = 1, \ldots, N+1; \qquad (2)$$

Donde E_j representa la energía en el j-ésimo nivel de descomposición, la cual se define como:
Para $i = 1, \ldots, N+1$.

$$E_j = \begin{cases} \sum_k |detalle_{j,k}|^2; & \text{Si } j <= N; \\ \sum_k |aprox_k|^2; & \text{en caso contrario.} \end{cases} \qquad (3)$$

Por su parte E_{total} representa la energía total de los coeficientes wavelet: $E_{total} \sum_{j=1}^{N+1} E_j$.

A partir de la descripción anterior, en las señales de EEG no sonificadas, se aplicó la transformada wavelet usando una wavelet de la familia *Daubechies*. Cada ventana de habla imaginada se representó mediante un conjunto de 5 valores de energía wavelet, 4 de los niveles de descomposición y uno de aproximación (D2-D5 y A5) con respecto a la energía wavelet total. Tal como se realizó en [12], el valor asociado con D1 es descartado.

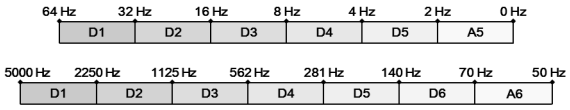


Figura 5. Descomposición en niveles de la transformada Wavelet de la señal de EEG (arriba) y la señal de EEG sonificado (abajo).

Por otra parte, en las señales EEG sonificadas se usó la misma familia de wavelets y se determinó usar 7 valores que representan la energía wavelet en todos los niveles de descomposición y el último de aproximación (D1-D6 y A6).

En la figura 5 se muestran los niveles de descomposición y detalle obtenidos de la transformada wavelet, donde cada nivel representa un rango de frecuencias en la señal de EEG o de audio. En la sonificación de EEG se aplicó la DWT usando una wavelet de diferente orden y con diferentes niveles de descomposición para intentar adaptarlos al nuevo dominio y al diferente rango de frecuencias. Se aplicó de dicha forma debido a que para la señal de EEG se calcularon los niveles a partir del rango de frecuencias de 0 a 60 Hz, mientras que en la señal de EEG sonificado los niveles van de 50 a 5000 Hz.

MFCC y coeficientes Delta

Los coeficientes cepstrales en la escala de Mel (Mel-frequency Cepstral Coefficients) son características espectrales que se calculan desde un análisis espectral en tiempo corto derivado de la transformada rápida de Fourier, comunmente se utilizan en señales de audio. Estos coeficientes aproximan el comportamiento de la señal al uso de la escala de frecuencia lineal.

En esta técnica se crea un espectrograma basado en los parámetros de tamaño de ventana y traslape entre ventanas. Luego a partir del espectrograma se establecen filtros en la escala de Mel por medio de funciones triangulares. A partir de las frecuencias filtradas el algoritmo obtiene coeficientes cepstrales que indican la energía en las distintas bandas del espectro. Con lo anterior se obtiene una matriz de coeficientes cepstrales por cada segmento de tiempo.

Los coeficientes cepstrales totales obtenidos son de gran dimension por lo que es conveniente hacer reducciones o representar los valores con sus atributos estadísticos. En nuestro caso se optó por representar los coeficientes con los siguientes atributos estadísticos: Valor máximo, valor mínimo, media y desviación estándar.

A partir de los coeficientes cepstrales, se implementó un esquema para la extracción de los cambios diferenciales entre segmentos de tiempo, para obtener los valores conocidos como valores *delta*. Además se extrajeron los cambios diferenciales entre segmentos de valores delta, para obtener los valores conocidos como valores *doble delta*. Para obtener los valores delta se tiene que establecer una ventana que indicará cuántas muestras vecinas se tomarán para calcular los cambios en el tiempo [20]. La ventana es representada por coeficientes obtenidos de un tiempo anterior y de un tiempo posterior de la muestra central. Estos coeficientes se definen como: $C = \{k, k-1, k-2, \cdots, -k\}$. Donde k es el tamaño de la ventana. Se repiten las muestra inicial y final k veces como ajuste. Después para calcular la derivada d_i en el tiempo de cada muestra, se recorre el vector aplicando una operación filtro usando los coeficientes calculados:

$$\Delta_i = C_1 X_n + C_2 X_{n-1} + \ldots + C_{|C|} X_{n-|C|} \quad (4)$$

Donde X_n es la muestra en la posición final de la ventana. Una vez calculadas las derivadas para cada muestra del vector, se divide cada valor entre la suma de los cuadrados de los coeficientes C.

$$\Delta F_i = \frac{\Delta_i}{\sum_{n=1}^{|C|} {C_n}^2} \quad (5)$$

Al final se quitan las muestras de ajuste agregadas. Los valores delta representan a un segmento de MFCC por lo que tienen

que ser representados igualmente en forma de atributos estadísticos. Por otra parte, los valores doble delta se calculan con el mismo proceso que los coeficientes delta, pero con los coeficientes delta como vector de entrada y también son representados con sus respectivos atributos estadísticos.

Clasificación

En la etapa de clasificación de este trabajo se emplearon dos configuraciones de experimentos con diferente número de canales. Un enfoque clasifica solo 4 canales correspondientes al modelo Geschwind-Wernicke correspondientes al hemisferio izquierdo en las áreas que involucran el procesamiento del lenguaje. Los electrodos seleccionados se muestran en la figura 2. El otro enfoque buscará clasificar usando todos los 14 canales disponibles, esto con la finalidad de evaluar si un conjunto más amplio de canales permite obtener información adicional que ayude a mejorar los porcentajes de exaactitud en la clasificación. Cabe señalar que las características fueron obtenidas para cada canal del dispositivo de EEG, por lo que en la etapa de clasificación las características (de los 4 ó 14 canales) fueron concatenadas en un solo archivo.

En este trabajo se evaluó el desempeño de los clasificadores: Naive Bayes (NB), Support Vector Machine (SVM) y Random Forest (RF). Se eligieron estos clasificadores para comparar los resultados con el trabajo previo [12]. Cada clasificador se implementó bajo un enfoque de validación cruzada [21] el cual hace particiones sobre el conjunto de datos de entrada. En nuestro caso para mantener las comparaciones de manera justa, se ejecutó la validación cruzada con 10 particiones bajo la misma semilla para el generador aleatorio de particiones.

Naive Bayes

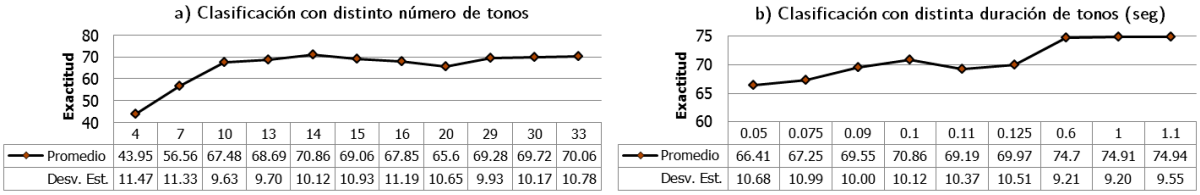
Naive Bayes es un clasificador probabilista, el cual es muy utilizado por ser sencillo y eficiente en muchos casos. Este clasificador funciona bajo la suposición de que las características en un conjunto de datos son independientes entre sí. Para clasificar una instancia, se determinan las probabilidades de cada valor de los atributos de esa instancia para cada clase disponible. Después se aplica el teorema de Bayes para conocer la probabilidad que tiene la instancia de pertenecer a una clase y se elige la clase con el valor de probabilidad más alto. Naive Bayes se describe con más detalle en [22]. En el presente trabajo se utilizó el clasificador Naive Bayes en su version tradicional.

SVM

SVM es un algoritmo de clasificación que busca separar lo mejor posible las clases de un conjunto de datos. Además la SVM tiene buenas propiedades de generalización, es insensible al sobre-entrenamiento y puede tratar conjuntos de datos de gran dimensionalidad. La SVM utiliza un hiperplano para maximizar los margenes que forman las fronteras de cada clase y un kernel el cual puede modificar la forma en que el hiperplano realiza la separación de clases. (en [23] se describe SVM a detalle). En esta investigación se aplicó un clasificador SVM que usa optimización mínima secuencial, bajo los siguientes parámetros: kernel tipo lineal y un error de $1.0E^{-12}$.

Random Forest

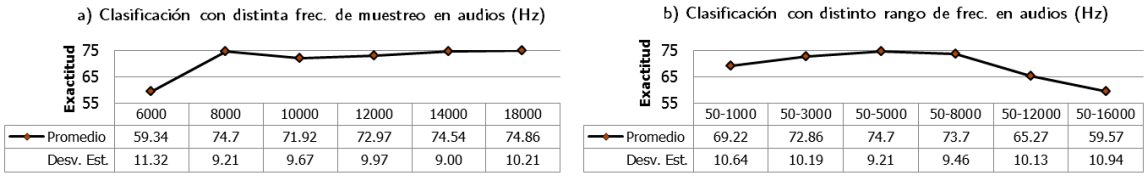
Random Forest es un algoritmo de aprendizaje basado en arboles de decisión que puede clasificar en corto tiempo bases de datos que pueden tener ruido y una gran cantidad de atributos. RF es una combinación de árboles predictores tal que cada uno de los árboles depende de los valores de un vector aleatorio 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



	4	7	10	13	14	15	16	20	29	30	33
Promedio	43.95	56.56	67.48	68.69	70.86	69.06	67.85	65.6	69.28	69.72	70.06
Desv. Est.	11.47	11.33	9.63	9.70	10.12	10.93	11.19	10.65	9.93	10.17	10.78

	0.05	0.075	0.09	0.1	0.11	0.125	0.6	1	1.1
Promedio	66.41	67.25	69.55	70.86	69.19	69.97	74.7	74.91	74.94
Desv. Est.	10.68	10.99	10.00	10.12	10.37	10.51	9.21	9.20	9.55

Figura 6. Promedios de exactitud en la clasificación con distintas configuraciones de número de tonos (a) y de duración de los tonos (b).



	6000	8000	10000	12000	14000	18000
Promedio	59.34	74.7	71.92	72.97	74.54	74.86
Desv. Est.	11.32	9.21	9.67	9.97	9.00	10.21

	50-1000	50-3000	50-5000	50-8000	50-12000	50-16000
Promedio	69.22	72.86	74.7	73.7	65.27	59.57
Desv. Est.	10.64	10.19	9.21	9.46	10.13	10.94

Figura 7. Promedios de exactitud en la clasificación con distintas configuraciones de distinta frec. de muestreo en audios (a) y de rango de frecuencias en audios (b).

dada, y al final la elección de la clase en RF se realiza usando voto mayoritario. En [24] se describe en detalle el algoritmo de clasificación de RF incluyendo el proceso para construir los arboles individuales. En la implementación de este clasificador se establecieron los siguientes parámetros: el número de árboles es 50 y el número de atributos considerados en cada nodo fue establecido a $log_2(numeroCaracteristicas)+1$.

RESULTADOS Y DISCUSIÓN

Configuración de parámetros del método de Sonificación de EEG

Los procesos de sonificación de EEG, de extracción de características usando la DWT y de extracción de características usando MFCC, requieren diversos parámetros de los cuales se desconocían sus efectos en la clasificación de audios provenientes de la sonificación de EEG. Para conocer dichos efectos, se ejecutaron diversas simulaciones que permitieran obtener una tendencia en la variación de los porcentajes de exactitud promedio al clasificar las palabras. En nuestro caso se tomó un subconjunto de 3 sujetos para poder efectuar las simulaciones en corto tiempo, sin buscar optimizar los parámetros para todos los sujetos, sino que sólo se buscó obtener valores cuya exactitud en la clasificación fuera aproximada a la mejor. La clasificación fue realizada con características extraídas aplicando la DWT y usando el clasificador RF, bajo un enfoque de validación cruzada con 10 particiones. Por ejemplo, el algoritmo de sonificación de *EEG a tonos* posee parámetros que afectan las características del audio que se genera, por lo que fue conveniente evaluar su comportamiento para seleccionar los parámetros.

El número de tonos en el proceso de sonificación indica el número de frecuencias dominantes del espectrograma de EEG que estarán representadas en la señal de audio EEG (ver las figuras 3 y 4). El número de tonos afecta también a cómo se percibe por el oído. Por ejemplo, un audio con pocos tonos tendrá pocas variaciones en tipos de sonido al escucharla y será más difícil encontrar detalles en una frecuencia en específico.

En la figura 6a se muestran los resultados de clasificación de los sujetos usando distinta cantidad de tonos en la señal de audio. Se

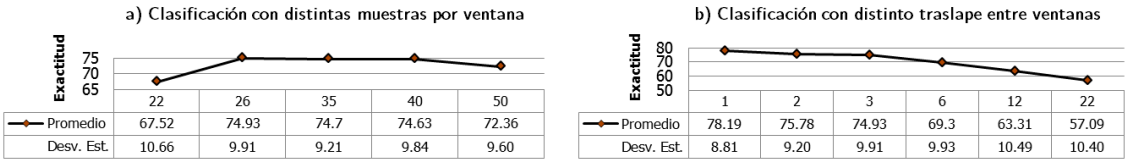


Figura 8. Promedios de exactitud en la clasificación con distintas muestras por ventana (a) y distintas muestras de traslape entre ventanas (b).

Tabla 1. Exactitud promedio en la clasificación con y sin sonificación de EEG en 4 canales

Clasificador/ Enfoque	**NB**	**SVM**	**RF**	*Atributos*
DWT	37.84 ± 21.32	38.69 ± 14.85	48.10 ± 12.77	*20*
Sonif. DWT	45.50 ± 18.89	48.17 ± 13.69	**55.83 ± 11.45**	*28*
Sonif. MFCC	44.96 ± 21.60	52.08 ± 13.33	52.37 ± 12.21	*368*

Tabla 2. Exactitud promedio en la clasificación con y sin sonificación de EEG en 14 canales

Clasificador/ Enfoque	**NB**	**SVM**	**RF**	*Atributos*
DWT	40.36 ± 22.65	52.30 ± 13.41	58.41 ± 11.45	*70*
Sonif. DWT	47.95 ± 20.0	59.67 ± 12.67	63.82 ± 10.51	*98*
Sonif. MFCC	49.90 ± 19.94	**64.14 ± 12.23**	60.66 ± 11.42	*1288*

observa que cuando hay menos de 10 tonos aproximadamente en la señal de audio los resultados de clasificación son bajos respecto a cuando se eligen 10 o más. Además al usar 14 tonos se obtuvo el mejor promedio de clasificación, esto nos indica que no necesariamente una sonificación de EEG con gran cantidad de tonos obtiene los mejores resultados de clasificación.

En la figura 6b se muestra que una duración de tonos baja no ayuda al proceso de clasificación, mientras que conforme se va aumentando la duración, la exactitud también aumenta.

En la figura 7a se muestra que para frecuencias de 8 KHz en adelante la exactitud de los promedios de clasificación aumenta en pequeñas cantidades, aunque cabe señalar que tanto la duración de los tonos, como la frecuencia de muestreo en los audios afecta directamente al tamaño del archivo de audio almacenado, por lo que su incremento tiene un alto costo.

En la figura 7b se muestran los resultados de clasificación variando el tamaño del rango de frecuencias, estableciendo un mínimo de 50 Hz y máximos de 1 KHz hasta 16 KHz, donde los rangos entre 3 KHz y 8 KHz obtienen los mejores resultados.

El tamaño de las ventanas para crear el espectrograma influye en el detalle de los cambios capturados en el tiempo de la señal. En la figura 8a se muestra que con una ventana formada por 26 muestras se obtiene el mejor porcentaje de clasificación de las simulaciones realizadas. El traslape entre ventanas influye en la longitud y detalle de los cambios capturados por el espectrograma. En la figura 8b se muestran los resultados de variar el traslape entre ventanas manteniendo el tamaño de ventana en 26. Se muestra que cuando existe el mínimo traslape entre ventanas se obtiene el mejor resultado de clasificación.

Resumen de configuraciones utilizadas

Después de efectuar simulaciones como las mencionadas anteriormente se obtuvieron las configuraciones con las cuales se realizaron los experimentos finales. Las configuraciones de parámetros utilizadas son descritas a continuación:

- Sonificación de EEG a tonos: Número de tonos = 14, frecuencia mínima de la señal de EEG = 1 Hz, frecuencia máxima de la señal de EEG = 60 Hz, tamaño de la ventana del espectrograma = 26 muestras, traslape entre ventanas del espectrograma = 1, duración de los tonos = 0.6 seg., frecuencia mínima del audio = 50 Hz, frecuencia máxima del audio = 5000 Hz, frecuencia de muestreo del audio = 8000 Hz.
- Extracción de características usando DWT en EEG: Wavelet madre daubechies orden 2, niveles de descomposición = 5.
- Extracción de características usando DWT en EEG sonificado: Wavelet madre daubechies orden 20, niveles de descomposición = 6.
- Extracción de características usando MFCC en EEG sonificado: Duración de la ventana = 20 ms, duración del traslape entre ventanas = 10 ms, frecuencia inferior = 50 Hz, frecuencia superior = 5000 Hz, coeficientes cepstrales = 23. Valores delta con ventana = 2 y valores doble delta con ventana = 1. Atributos estadísticos: Máximo, mínimo, media, desv. estándar. Al final solo toman los 23 coeficientes de los valores doble delta.

Experimento de clasificación para los 27 sujetos

Después de extraer las caracterísitcas de los audios se procedió a clasificar las palabras imaginadas de los 27 sujetos. Luego de obtener los resultados de la clasificación, se procedió a realizar una comparación entre los resultados de aplicar la sonificación de EEG y no aplicarla. En las tablas 1 y 2 se muestran los resultados de clasificación promedio obtenidos para los 27 sujetos, así como los resultados reportados al no utilizar sonificación de EEG.

Analizando las tablas se observa que el enfoque de extracción de características de los audios con DWT obtuvo el mejor desempeño al usar 4 canales clasificando con RF. Mientras que para 14 canales el enfoque de extracción de características con MFCC logra los mejores porcentajes de clasificación clasificando con SVM. Los enfoques propuestos incluyendo sonificación mejoran ligeramente los porcentajes de exactitud respecto a los resportados en [12], obteniendo una diferencia de 7.73% y 5.73% para 4 y 14 canales respectivamente.

En la última columna de cada una de las tablas 1 y 2 se muestra el número de características o atributos obtenidos para cada enfoque. La cantidad de atributos obtenidos puede afectar la clasificación si estos son muy pocos o demasiados, por lo que seleccionar el valor adecuado de atributos es un problema de optimización (fuera del alcance de este trabajo). En nuestro caso, más allá de la cantidad, toma relevancia lo que representan los atributos obtenidos. Donde por ejemplo al clasificar con 14 canales, en el enfoque de sonificación de EEG con MFCC se obtuvo una cantidad alta de atributos (1288), y éstos fueron clasificados con buenos resultados usando SVM sin caer en el problema de la alta dimensionalidad.

En las figuras 9 y 10 se comparan los promedios de clasificación de los 27 individuos usando 4 y 14 canales, respectivamente.

En la figura 9 se comparan los resultados obtenidos por el enfoque de EEG sin sonificación y extracción de características con DWT contra el enfoque de sonificación de EEG con DWT, ambos clasificando con RF. Mientras que en la figura 10 se comparan

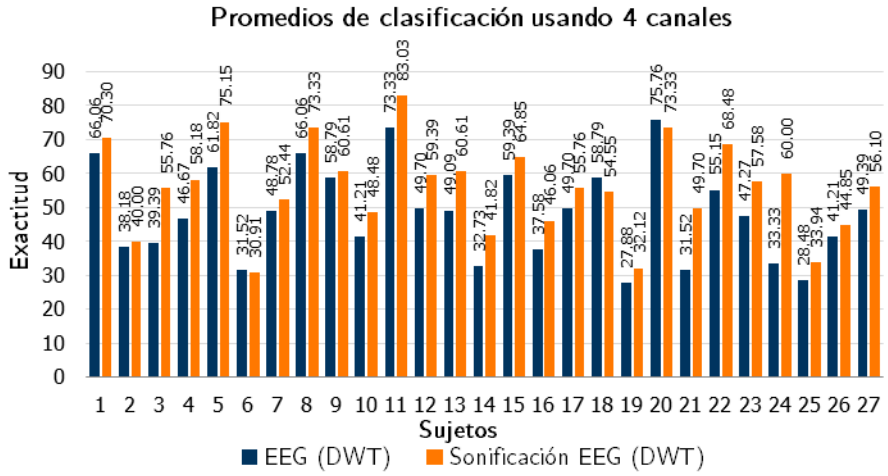


Figura 9. Exactitud de la clasificación para los 27 sujetos usando 4 canales.

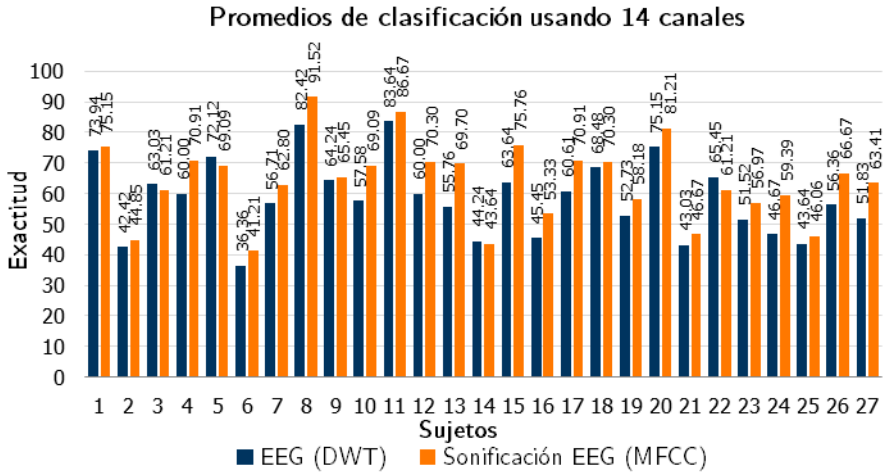


Figura 10. Exactitud de la clasificación para los 27 sujetos usando 14 canales.

los resultados obtenidos del enfoque de EEG sin sonificación con DWT clasificando con RF contra el enfoque de sonificación de EEG con MFCC clasificando con SVM. En dichas gráficas se observa que existen sujetos que obtuvieron una amplia diferencia en la exactitud de clasificación al ser procesados mediante los enfoques de sonificación de EEG, pero también existen sujetos que no mejoraron sus promedios de clasificación al aplicar la sonificación de EEG. En la tabla 3 se detallan los resultados de las gráficas y 10 incluyendo la exactitud y desviación estándar obtenida en la clasificación por cada sujeto. En esta tabla se muestra que el mejor resultado de clasificación que obtuvo un sujeto fue de 83.03% y 91.52% para 4 y 14 canales, respectivamente. Esto sugiere que existe información suficiente en la señal cerebral para que por medio de técnicas como las aquí propuestas se puedan tener buenos porcentajes de clasificación. Se muestra que el método tiene variaciones amplias en la exactitud entre sujetos. Se aplicó una prueba T con un valor de significancia de 0.05, para comparar los promedios de la tabla 3, para los pares de 4 canales y 14 canales respectivamente. Con esta prueba se obtuvo

Tabla 3. Mejores exactitudes promedio en la clasificación con y sin sonificación de EEG en 4 y 14 canales

	EEG Sin Sonif. 4 canales	**EEG Sonif. 4 Canales**	**EEG Sin Sonif. 14 canales**	**EEG Sonif. 14 Canales**
S1	66.06 ± 9.54	70.30 ± 8.78	73.94 ± 8.28	75.15 ± 11.35
S2	38.18 ± 14.44	40.00 ± 14.73	42.42 ± 13.98	44.85 ± 14.06
S3	39.39 ± 14.44	55.76 ± 11.05	63.03 ± 10.50	61.21 ± 12.67
S4	46.67 ± 13.28	58.18 ± 10.74	60.00 ± 11.24	70.91 ± 11.63
S5	61.82 ± 9.95	75.15 ± 8.22	72.12 ± 9.80	69.09 ± 11.67
S6	31.52 ± 15.24	30.91 ± 15.57	36.36 ± 14.54	41.21 ± 14.49
S7	48.78 ± 13.15	52.44 ± 12.15	56.71 ± 11.41	62.80 ± 12.44
S8	66.06 ± 9.46	73.33 ± 8.01	82.42 ± 7.74	91.52 ± 10.37
S9	58.79 ± 11.35	60.61 ± 10.90	64.24 ± 11.05	65.45 ± 12.24
S10	41.21 ± 14.36	48.48 ± 13.28	57.58 ± 11.89	69.09 ± 11.68
S11	73.33 ± 7.79	83.03 ± 5.37	83.64 ± 5.82	86.67 ± 10.56
S12	49.70 ± 12.89	59.39 ± 10.61	60.00 ± 11.57	70.30 ± 11.71
S13	49.09 ± 11.93	60.61 ± 10.30	55.76 ± 11.40	69.70 ± 11.73
S14	32.73 ± 15.4	41.82 ± 14.30	44.24 ± 14.26	43.64 ± 13.92
S15	59.39 ± 10.99	64.85 ± 9.97	63.64 ± 10.41	75.76 ± 11.39
S16	37.58 ± 14.29	46.06 ± 13.18	45.45 ± 12.93	53.33 ± 12.96
S17	49.70 ± 12.45	55.76 ± 11.48	60.61 ± 10.61	70.91 ± 11.48
S18	58.79 ± 11.60	54.55 ± 11.53	68.48 ± 10.22	70.30 ± 11.99
S19	27.88 ± 15.99	32.12 ± 15.36	52.73 ± 13.11	58.18 ± 13.08
S20	75.76 ± 7.40	73.33 ± 7.83	75.15 ± 8.04	81.21 ± 10.87
S21	31.52 ± 15.67	49.70 ± 12.60	43.03 ± 14.08	46.67 ± 13.80
S22	55.15 ± 10.4	68.48 ± 8.97	65.45 ± 9.33	61.21 ± 12.14
S23	47.27 ± 13.73	57.58 ± 11.71	51.52 ± 13.24	56.97 ± 12.54
S24	33.33 ± 14.7	60.00 ± 10.63	46.67 ± 13.36	59.39 ± 12.91
S25	28.48 ± 17.32	33.94 ± 16.65	43.64 ± 16.06	46.06 ± 12.43
S26	41.21 ± 14.23	44.85 ± 13.39	56.36 ± 12.41	66.67 ± 11.82
S27	49.39 ± 12.71	56.10 ± 11.90	51.83 ± 11.80	63.41 ± 12.29
Avg.	**48.10 ± 12.77**	**55.83 ± 11.45**	**58.41 ± 11.45**	**64.14 ± 12.23**

que existe diferencia significativa entre los promedios de exactitud obtenidos por los métodos que si aplican sonificación respecto a los que no aplican sonificación, esto para ambos pares de 4 y 14 canales.

CONCLUSIONES

Este trabajo explora una representación distinta de las señales de EEG en el dominio acústico mostrando su impacto en la clasificación automática de estas señales.

Se muestra que la clasificación de la señal de EEG aplicando la sonificación, mediante la elección de las frecuencias dominantes del espectrograma de la señal de EEG y el mapeo de las frecuencias de EEG a frecuencias del audio, logró resaltar patrones de la señal que ayudaron a tener mejores resultados respecto a trabajos previos en la correcta clasificación de los datos de 27 sujetos. Por otra parte se observó que los métodos propuestos varían su exactitud dependiendo del sujeto y no hay un método dominante, esto como se observó en la sección de resultados, da pie a probar una posible unión de enfoques o esquema de selección de métodos. Otra idea a explorar es personalizar la configuración del método, usando un algoritmo de selección de modelo para optimizar los parámetros de la sonificación, extracción de características y clasificación, los cuales otorguen la mejor exactitud de clasificación para cada sujeto. También se muestra que cuando se extraen características usando MFCC los porcentajes de clasificación son relativamente mejores,

posiblemente esto se debe a que bajo esta representación se incluyó información temporal, representada con los valores doble delta. Es por ello que en trabajos futuros se espera experimentar con otros métodos que caractericen de mejor forma la secuencia de eventos en la señal.

REFERENCIAS

[1] T. M. Rutkowski, A. Cichocki, and D. Mandic, "Information fusion for perceptual feedback: A brain activity sonification approach," *Signal Processing Techniques for Knowledge Extraction and Information Fusion*, pp. 261–273, 2008.

[2] G. Marco-Zaccaria, "Sonification of EEG signals: A study on alpha band instantaneous coherence," Master thesis, 2011.

[3] E. Miranda, A. Brouse, B. Boskamp, and H. Mullaney, "Plymouth brain-computer music interface project: Intelligent assistive technology for music-making," *International Computer Music Conference*, 2005.

[4] J. Eaton and E. Miranda, "Real-time notation using brainwave control," *Sound and Music Computing Conference*, 2013.

[5] G. Baier and T. Hermann, "The sonification of rhythms in human electroencephalogram." *International Conference on Auditory Display (ICAD)*, 2004.

[6] T. Hermann, P. Meinicke *et al.*, "Sonification for EEG data analysis," *Proceedings of the 2002 International Conference on Auditory Display*, 2002.

[7] G. Baier, T. Hermann, S. Sahle, and U. Stephani, "Event based sonification of EEG rhythms in real time," *Clinical Neurophysiology*, vol. 118, no. 6, pp. 1377–1386, 2007.

[8] T. Hermann, G. Baier, U. Stephani, and H. Ritter, "Kernel regression mapping for vocal EEG sonification," *Proceedings of the International Conference on Auditory Display*, 2008.

[9] M. Elgendi, J. Dauwels *et al.*, "From auditory and visual to immersive neurofeedback: Application to diagnosis of alzheimerś disease," *Neural Computation, Neural Devices, and Neural Prosthesis*, pp. 63–97, 2014.

[10] M. Wester and T. Schultz, "Unspoken Speech - Speech Recognition Based On Electroencephalography," Master's thesis, Institut fur Theoretische Informatik Universitat Karlsruhe (TH), Karlsruhe, Germany, 2006.

[11] A. A. Torres-García, C. A. Reyes-García, and L. Villaseñor Pineda, "Toward a silent speech interface based on unspoken speech," *BIOSTEC - BIOSIGNALS*, pp. 370–373, 2012.

[12] A. A. Torres-García, C. A. Reyes-García, and L. Villaseñor-Pineda, "Análisis de Señales Electroencefalográficas para la Clasificación de Habla Imaginada," *Revista Mexicana de Ingeniería Biomédica*, vol. 34, no. 1, pp. 23–39, 2013.

[13] E. F. González-Castañeda, A. A. Torres-García, C. A. Reyes-García, and L. Villaseñor-Pineda, "Sonificación de EEG para la clasificación de palabras no pronunciadas," *Research in Computing Science*, vol. 74, pp. 61–72, 2014.

[14] C. Anderson, "Sonification - Brain Computer Interfaces Laboratory," *Department of Commputer Science, Colorado State University, www.cs.colostate.edu/eeg/main/projects/sonification*, 2005.

[15] A. A. Torres-Garcia, "Clasificación de palabras no pronunciadas presentes en Electroencefalogramas (EEG)," Tesis de Maestría, 2011.

[16] Xuemin, Chi and Hagedorn, John and others, "EEG-based discrimination of imagined speech Phonemes," *International Journal of Bioelectromagnetism*, vol. 13, no. 4, pp. 201–206, 2011.

[17] X. Pei, D. L. Barbour, E. C. Leuthardt, and G. Schalk, "Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans," *Journal of neural engineering*, vol. 8, no. 4, p. 046028, 2011.

[18] J. Mourino, J. del R Millan *et al.*, "Spatial filtering in the training process of a brain computer interface," *Engineering in Medicine and Biology Society. Proceedings of the 23rd Annual International Conference of the IEEE*, vol. 1, pp. 639–642, 2001.

[19] M. A. Pinsky, *Introduction to Fourier analysis and wavelets*. Amer Mathematical Society, 2002, vol. 102.

[20] X. Hu, H. Zhang *et al.*, "Isolated word speech recognition system based on FPGA," *Journal of Computers*, vol. 8, no. 12, pp. 3216–3222, 2013.

[21] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *IJCAI*, vol. 14, no. 2, pp. 1137–1145, 1995.

[22] G. H. John and P. Langley, "Estimating continuous distributions in bayesian classifiers," in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1995, pp. 338–345.

[23] J. C. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in kernel methods*. MIT press, 1999, pp. 185–208.

[24] L. Rokach, *Pattern Classification Using Ensemble Methods*. World Scientific, 2009.