

## Hacia un método para identificación del idioma a través de información acústica

Jesús A. Fortoul-Díaz<sup>1</sup>, Ana L. Reyes-Herrera<sup>1</sup>, Alejandro A. Torres-García<sup>2</sup>,  
Luis Villaseñor-Pineda<sup>2</sup>

<sup>1</sup> Tecnológico de Monterrey,  
Puebla, México

<sup>2</sup> Instituto Nacional de Astrofísica Óptica y Electrónica,  
México

{a01098295, alreyes}@itesm.mx, {alejandro.torres, villasen}@inaoep.mx

**Resumen.** En la actualidad existen diferentes métodos para la identificación automática del lenguaje hablado basados principalmente en dos enfoques: el primero utiliza información fonética para la tipificación del idioma, y el segundo enfoque utiliza la información suprasegmental extraída directamente de la señal acústica. A pesar de los buenos resultados de los métodos utilizados actualmente para cada uno de los enfoques, en el primero se depende de un estudio lingüístico previo para cada uno de los idiomas a identificar y en el segundo enfoque utiliza un considerable tiempo de cómputo para procesar y obtener la caracterización del ritmo de la señal acústica, haciendo el proceso de identificación del idioma difícil de obtener en tiempo real. En el presente trabajo se presenta un método de identificación del lenguaje hablado, el cual caracteriza la información suprasegmental usando el cálculo de la energía relativa wavelet de cada nivel de descomposición de la señal, así como el cálculo entre segmentos de la señal para capturar el cambio en el tiempo. Los resultados alcanzados son alentadores por lo que se espera continuar el trabajo para llevarlo a una configuración multiclase y con posibilidades de aplicar en tiempo real.

**Palabras clave:** Identificación automática del idioma, energía relativa wavelet, información suprasegmental.

### Toward a Method for Automatic Language Identification using Acoustic Information

**Abstract.** Nowadays there are different methods in the field of spoken language identification, mainly based in two approaches: the first uses phonetic information of languages and the second uses suprasegmental characteristic obtained from the acoustic signal. Despite the good results of these methods, the first one depends on a previous linguistic study for each language to identify, and the second one uses a considerable computation time to process and

characterize the rhythm of the acoustic signal, making language identification a little harder to obtain in real time. In this paper, we present a method that characterizes the suprasegmental information calculating the relative wavelet energy of each decomposition level of the acoustic signal, and the calculation between signal segments in order to capture the change in time. The results obtained are encouraging, and it is the reason to continue working with multiclass classification, and apply it in real time.

**Keywords:** Automatic language identification, relative wavelet energy, suprasegmental information.

## **1. Introducción**

Con el paso de los años, el hombre ha desarrollado técnicas que le permitan simplificar la vida diaria, uno de los puntos en los que se ha centrado principalmente en las últimas décadas es el procesamiento del lenguaje natural. Se ha logrado realizar la identificación de lenguajes, dialectos e idiomas haciendo uso de técnicas que requieren una transcripción fonética, esto quiere decir que la identificación se realiza por medio de similitudes entre sílabas, vocablos o palabras.

Pero qué pasaría si en vez de realizar todo el proceso, se pudiera realizar la identificación por medio de características que no dependan de realizar una transcripción, eso lograría agilizar el proceso, y serviría en casos, en los cuales no se tiene una transcripción fonética. Es por ello que varios investigadores se han dado la tarea de detectar nuevos métodos basados en características suprasegmentales, que se relacionen más con el ritmo, el acento, o la entonación del hablante.

En el presente trabajo se abordará la identificación de distintos idiomas a través de un método que permita su clasificación, tomando como base los proyectos de investigadores que trabajaron con identificación de características suprasegmentales. El método consiste en aplicar la transformada wavelet a las señales de audio, y posteriormente calcular la Energía Relativa Wavelet, con la cual se espera obtener una representación de la señal que proporcione información sobre los niveles en los que la energía es mayor, dependiendo del ritmo del idioma.

La ventaja que ofrece el método planteado en el trabajo es el hecho de requerir menor información para realizar la clasificación de idiomas, en comparación con otros trabajos –descritos en la siguiente sección– lo cual significa que los recursos computacionales a utilizar son menores.

## **2. Trabajos relacionados**

En las últimas décadas se han desarrollado distintos métodos para la identificación de lenguajes, algunos de ellos utilizan la transcripción fonética, es decir se basan en la representación de la voz a través de un alfabeto escrito, realizando la identificación en base a sílabas, palabras o frases similares. Además de ese tipo de métodos existen otros que no requieren el uso de la transcripción y se basan principalmente en características suprasegmentales, es decir, en el ritmo del habla, la entonación, el acento, entre otros.

El trabajo más representativo usando características suprasegmentales es el realizado por Rouas et al. [1], [2]. En este trabajo la identificación de lenguajes se realiza a través del ritmo que posee cada idioma, obteniendo la relación entre los intervalos vocálicos y consonánticos de cada idioma y utilizando los modelos de mezclas Gaussianas (GMM) como técnica de clasificación. Y a través del análisis del ritmo es que logra obtener resultados sobre la identificación entre duplas de idiomas.

Otro trabajo relacionado es [3], el cual nos presentan un método basado en la identificación de idiomas usando características fonéticas y prosódicas, haciendo uso de modelos de mezclas gaussianas (GMM), clasificando los idiomas a través de máquinas de soporte vectorial (SVM) y utilizando un kernel de secuencia probabilística para obtener las características más destacadas de las señales de audio.

Un trabajo interesante es [4], principalmente porque en él se realiza la caracterización de la señal en base a los MFCC de la señal de audio; haciendo uso del algoritmo Fuzzy C-Means logra obtener nuevos vectores de datos, que permiten seleccionar de una mejor forma los atributos más significativos, para posteriormente realizar la identificación en base a modelos de Markov (HMM).

El trabajo presentado en [5] muestra una solución innovadora, basada en la transformada wavelet como método de caracterización; la forma en que realiza la identificación de idiomas es por medio del clasificador Naive Bayes, el cual proporciona muy buenos resultados. Se debe mencionar que en el presente trabajo se estará trabajando con un método similar, teniendo como ventaja una selección de atributos mucho más compacta. La siguiente sección presenta el método propuesto y describe la caracterización propuesta basada en la energía relativa wavelet.

### **3. Método propuesto**

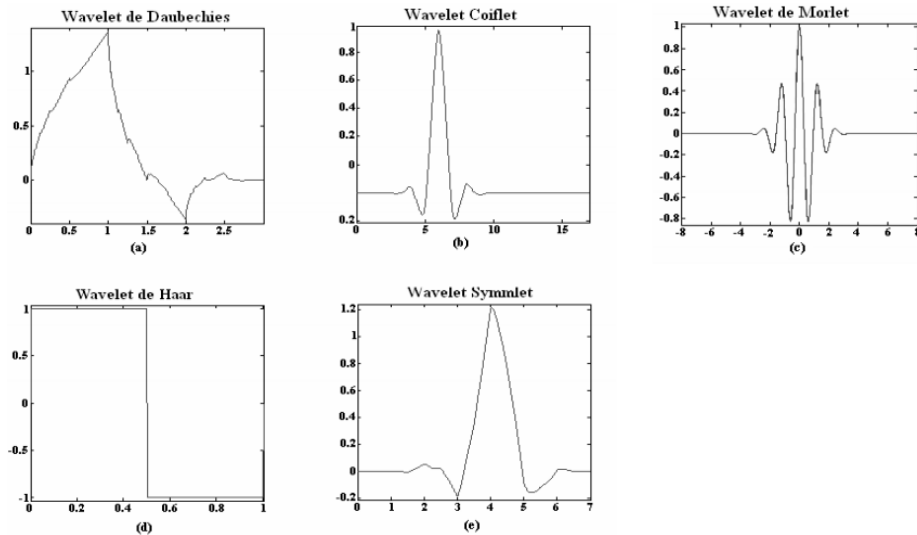
#### **3.1. Transformada wavelet**

Para el presente trabajo se hace uso de un método de extracción de características de la señal en el dominio de la frecuencia y al mismo tiempo conserve las características en el dominio del tiempo, este método es la Transformada Wavelet. En un inicio las wavelets eran utilizados en el campo de la geología, en la búsqueda de yacimientos de petróleo, y es ahí donde se observaron las ventajas que ofrecen sobre otros métodos, pues los componentes de la señal procesada siempre mantenían la misma forma, sin importar que la señal sufriera distorsiones como dilatación, compresión o desplazamiento en el tiempo; el único factor que llegaba a influenciar era la wavelet madre que se utilizaba para realizar el análisis [6].

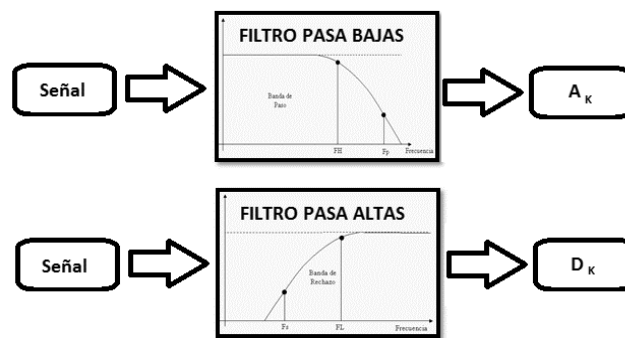
Existen diferentes familias de wavelets madre, entre ellas se encuentra la familia Daubechies, las cuales son wavelets ortonormales aplicadas en análisis de señales discretas; la familia Biortonormales con propiedades usadas en reconstrucción de imágenes. Algunas otras son: la wavelet Haar, la wavelet de Morlet, la familia de wavelet Coiflet, la familia wavelet Symmlet, entre otras. En la siguiente imagen, obtenida de [7], se muestra algunas de las wavelets madre más utilizadas en la actualidad.

El método implementado en el presente trabajo es la Transformada Discreta Wavelet, (DWT) la cual es utilizada principalmente en procesamiento de señales, con

la finalidad de lograr la codificación de la misma señal basándose principalmente en el uso de bancos de filtros para obtener distintos niveles de descomposición [5]. Utilizando un banco de filtros pasa altas se obtienen los coeficientes de detalle, mientras que con el filtro pasa bajas se puede obtener los coeficientes de aproximación.



**Fig. 1.** Muestra representativa de distintas familias de Wavelets: (a) Wavelet Daubechies 2, (b) Wavelet Coiflet 2, (c) Wavelet de Morlet, (d) Wavelet Haar, (e) Wavelet Symmlet 4.



**Fig. 2.** Proceso requerido para obtener los coeficientes de detalle ( $D_k$ ) y de aproximación ( $A_k$ ), a través de un banco de filtros.

La DWT usa un análisis de multi-resolución, en el cual se aplica el algoritmo denominado piramidal, y se puede expresar por medio de los filtros pasa altas y pasa bajas para obtener los coeficientes de detalle y aproximación respectivamente. Debido a que el análisis es del tipo multi-resolución, se obtienen coeficientes de detalle y aproximación en cada una de las diferentes bandas de frecuencia, lo cual significa que cada banda posee la mitad de las muestras de la banda superior [8] (véase la Fig. 2).

Un punto importante a tomar en cuenta durante el análisis de la señal es la forma en que los niveles de descomposición serán obtenidos. Tomando en cuenta el teorema de Nyquist, que expresa que la frecuencia de muestreo mínima necesaria para lograr una grabación de calidad siempre debe ser el doble de la frecuencia de la señal que se pretende grabar [9]:

$$F_s = 2F_0. \quad (1)$$

### 3.2. Energía relativa wavelet

Para obtener las características más destacadas relacionadas al ritmo del lenguaje se experimentó con los diferentes niveles de descomposición, y el de aproximación, al aplicar el método denominado, Energía Relativa Wavelet (EWR) [8]. La forma de calcular la EWR consiste en obtener la sumatoria del cuadrado de cada uno de los coeficientes para cada nivel “k”, ya sea de un nivel de detalle ( $D_k$ ) o del nivel de aproximación ( $A_n$ ).

$$E_j = \begin{cases} \sum_k |d_k|^2 & \text{para cada uno de los niveles de detalle} \\ \sum_k |a_n|^2 & \text{para el nivel de aproximacion.} \end{cases} \quad (2)$$

Debido a que este método asume que cada nivel (N) posee un porcentaje de la energía total, se procede a calcular la energía total de la señal, la cual consta de los “N” niveles, más la energía en el nivel de aproximación, esto quiere decir que la energía total se conforma de “N+1” componentes:

$$E_{Total} = \sum_{j=1}^{N+1} E_j. \quad (3)$$

Con el cálculo de  $E_{Total}$ , se procede a obtener la energía relativa de cada uno de los niveles “j” a través de la siguiente fórmula.

$$EWR_j = \frac{E_j}{E_{Total}}. \quad (4)$$

### 3.3. Delta de energía

El cálculo de la EWR se realiza a cada determinado tiempo sobre la señal, utilizando un intervalo de tiempo definido (por ejemplo cada segundo), la idea es obtener la energía de cada intervalo y donde el cambio de energía en los estos intervalos proporcione información sobre el ritmo del lenguaje. Para lograr esto, un primer paso es representar cada intervalo con la información más representativa para describir la información suprasegmental. Para ello se calculan los funcionales estadísticos de cada intervalo: la media aritmética ( $\mu$ ), la desviación estándar ( $\sigma$ ), el valor máximo, el valor

mínimo, todos ellos respecto a cada nivel de energía; esto quiere decir que la información en cuando a cada nivel se mantiene.

Para capturar la información de la EWR a través del tiempo se calcula el cambio en la energía (delta en la energía), en diferentes puntos del tiempo, la cual representa el cambio de energía de un nivel específico “j”, entre dos instantes de tiempo consecutivos:

$$\Delta E1_j = | EWR_j(t_1) - EWR_j(t_2) |. \quad (5)$$

De igual forma se puede calcular la delta de energía tomando instantes de tiempo que no sean estrictamente consecutivos, la cual representará el cambio de energía de un nivel específico “j”, entre dos instantes de tiempo distantes:

$$\Delta E2_j = | EWR_j(t_1) - EWR_j(t_3) |. \quad (6)$$

### 3.4. Clasificadores

Los atributos seleccionados para realizar la clasificación serán: los estadísticos de cada nivel y los estadísticos de los deltas de energía en el tiempo, con lo cual se espera tener una caracterización descriptiva representativa de cada idioma. En la actualidad existen varias herramientas que permiten llevar a cabo una buena clasificación de datos, para este caso se utilizará la herramienta Weka [10], [11], la cual permitirá realizar la minería de datos que permita encontrar patrones entre diferentes instancias y así determinar las relaciones que describan de una manera conjunta a los datos.

Existen diferentes algoritmos de clasificación en Weka, entre los más comunes se encuentran: J48, Naive Bayes, Random Forest, Máquinas de Soporte Vectorial. A su vez cada uno de ellos puede evaluarse utilizando diferentes opciones de prueba, como lo puede ser: usando el conjunto de entrenamiento, proporcionando un nuevo conjunto, o usar validación cruzada.

El resultado de una buena clasificación se puede observar a través del porcentaje de instancias clasificadas de manera correcta; otra forma en que se puede comprobar el resultado de la clasificación es observando los resultados obtenidos en la matriz de confusión, la cual provee de manera explícita la relación entre falsos positivos y falsos negativos en la clasificación de instancias.

### 3.5. Procedimiento

Para la realización del proyecto se decidió tomar de referencia los idiomas del corpus OGI que habían sido utilizados en diversas ocasiones dentro del marco referencial. Los idiomas a clasificar son: Alemán, Chino Mandarín, Español, Farsi, Inglés, Japonés, Coreano, Tamil, y Vietnamita.

Por cada uno de los idiomas se usaran 50 audios, con una duración de 10 segundos, en los cuales se escuchan a diferentes interlocutores realizando un relato de su interés. Dos de las características más relevantes de los audios son: la extensión de los archivos es “.wav”, y la frecuencia de muestreo es de 8kHz. Es por ello que se decidió trabajar el procesamiento de la señal con Matlab [12], [13], ya que cuenta con una función que permite leer archivos con extensión .wav, posee un toolbox para implementar la

transformada wavelet, y permite conectarse con el api de Weka, para realizar la clasificación necesaria.

Para implementar correctamente la DWT, se debe comenzar por definir los niveles de descomposición que tendrá la señal, es decir, que tan a detalle se analizará. Recordando que la frecuencia de muestreo es de 8kHz, se procede a aplicar el teorema de muestreo, el cual nos dice que la frecuencia de muestreo es el doble de la frecuencia fundamental (de la señal analógica que se digitalizó), por lo tanto al aplicar el teorema se obtendrá que la frecuencia fundamental de los audios es de 4kHz.

Aplicando el mismo teorema para obtener los posteriores niveles de descomposición, en conjunto con el algoritmo piramidal, se llega a obtener la descripción de las frecuencias que se analizan conforme se aumenta de nivel. Para el presente trabajo se decidió realizar una descomposición hasta el nivel 7, con lo cual se logró abarcar frecuencias bajas de hasta 31.25Hz; en la tabla 1 se da una descripción más detallada de los niveles obtenidos.

**Tabla 1.** Cada nivel de detalle posee límites de frecuencia basados en el teorema de muestreo; el nivel de aproximación abarca desde 0 Hz hasta el límite inferior del último nivel de detalle, en este caso D7.

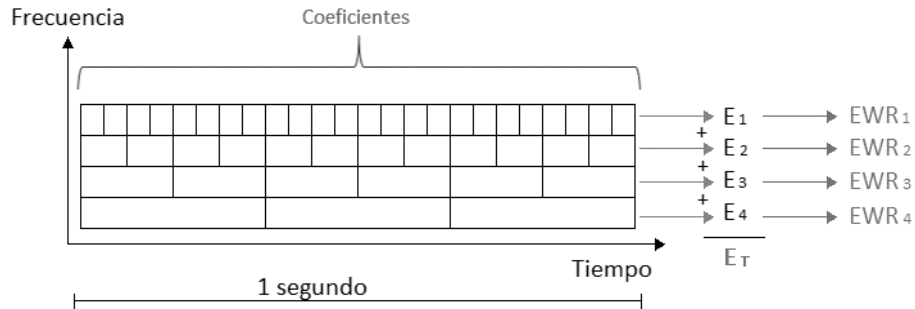
Nivel de descomposición	Rango de Frecuencias		
D1	2000 Hz	a	4000 Hz
D2	1000 Hz	a	2000 Hz
D3	500 Hz	a	1000 Hz
D4	250 Hz	a	500 Hz
D5	125 Hz	a	250 Hz
D6	62.5 Hz	a	125 Hz
D7	31.25 Hz	a	62.5 Hz
A7	0 Hz	a	31.25 Hz

Una vez que ya se tiene el número de niveles en los cuales se va a descomponer la señal, se procede a realizar la selección de la wavelet madre con la cual se va a realizar el procesamiento de la señal. Llevando a cabo varias pruebas con la familia Daubechies, y con la familia Biorthonormal, se llegó a la conclusión de que la mejor wavelet madre que se podría utilizar es la “Daubechies 8”, ya que proporciona mejores resultados a la hora de realizar la clasificación.

Automatizando el proceso en Matlab, se logró que el programa accediera a los archivos de audio, los cargara en el espacio de trabajo, y dividiera la señal en periodos de 1 segundo; una vez teniendo los segmentos de la señal se aplicó la transformada wavelet “Daubechies 8”, con 7 niveles de descomposición, y con ello se obtuvo un vector de coeficientes por cada nivel. Posteriormente se aplicó el método de la Energía Relativa Wavelet a los vectores, con la finalidad de obtener el porcentaje de energía total que posee cada uno de los niveles; la siguiente figura ilustra de una mejor manera este proceso.

Este proceso se repite para cada uno de los segmentos de la señal, obteniendo como resultado final una matriz en la cual los renglones representan los niveles de descomposición, mientras que las columnas representan cada periodo de la señal

(segundo a segundo). Una vez que ya se tiene la matriz se pueden obtener los estadísticos de los niveles; al mismo tiempo ya se pueden calcular las deltas de energía, ya sea la delta cercana  $\Delta E_{1j}$ , o la delta lejana  $\Delta E_{2j}$  para cada nivel.



**Fig. 3.** Representación del proceso que transforma los coeficientes de los diferentes niveles en valores de Energía Wavelet Relativa, tomando en cuenta que la señal analizada es de 1 segundo.

N \ t	Coeficientes										
	1s	2s	3s	4s	5s	6s	7s	8s	9s	10s	
Nivel 1	EWR <sub>1</sub>	EWR <sub>1</sub>	EWR <sub>1</sub>	EWR <sub>1</sub>	EWR <sub>1</sub>	EWR <sub>1</sub>	EWR <sub>1</sub>	EWR <sub>1</sub>	EWR <sub>1</sub>	EWR <sub>1</sub>	→ $\mu_1, \sigma_1, \max_1, \min_1$
Nivel 2	EWR <sub>2</sub>	EWR <sub>2</sub>	EWR <sub>2</sub>	EWR <sub>2</sub>	EWR <sub>2</sub>	EWR <sub>2</sub>	EWR <sub>2</sub>	EWR <sub>2</sub>	EWR <sub>2</sub>	EWR <sub>2</sub>	→ $\mu_2, \sigma_2, \max_2, \min_2$
Nivel 3	EWR <sub>3</sub>	EWR <sub>3</sub>	EWR <sub>3</sub>	EWR <sub>3</sub>	EWR <sub>3</sub>	EWR <sub>3</sub>	EWR <sub>3</sub>	EWR <sub>3</sub>	EWR <sub>3</sub>	EWR <sub>3</sub>	→ $\mu_3, \sigma_3, \max_3, \min_3$
Nivel 4	EWR <sub>4</sub>	EWR <sub>4</sub>	EWR <sub>4</sub>	EWR <sub>4</sub>	EWR <sub>4</sub>	EWR <sub>4</sub>	EWR <sub>4</sub>	EWR <sub>4</sub>	EWR <sub>4</sub>	EWR <sub>4</sub>	→ $\mu_4, \sigma_4, \max_4, \min_4$

**Fig. 4.** Representación de una señal de audio de 10 segundos ya procesada; la cual permite obtener datos de la señal como: la media, desviación estándar, máximo, mínimo, y deltas de energía en el tiempo.

Cada una de las deltas de energía generará más información sobre como es el cambio de EWR a lo largo del tiempo; de ellas podemos extraer los mismos estadísticos como se hizo anteriormente en cada nivel. Al igual que en la etapa de lectura de la señal, se programó un algoritmo que permitiera desarrollar un archivo “.arff” de forma automática, usando como atributos los estadísticos de nivel, y los estadísticos de las deltas de energía, y como clase los dos idiomas que se deseaba comparar. Es por ello que cada señal de audio debe contar con el siguiente conjunto de atributos para ser correctamente clasificada.

	Estadísticos de Nivel					Estadísticos de $\Delta E_1$					Estadísticos de $\Delta E_2$			
	$\mu$	$\sigma$	max	min		$\mu$	$\sigma$	max	min		$\mu$	$\sigma$	max	min
A7					A7					A7				
D7					D7					D7				
D6					D6					D6				
D5					D5					D5				
D4					D4					D4				
D3					D3					D3				
D2					D2					D2				
D1					D1					D1				

**Fig. 5.** Se muestran los 96 atributos distintivos de una señal de audio, los cuales se utilizarán para realizar la clasificación correspondiente.



Una vez que se tiene construido el archivo, se carga al entorno de Weka, donde se realizaron diferentes pruebas con los clasificadores: J48, Random Forest, Naive Bayes, y Maquinas de Soporte Vectorial (SMO), cada uno fue evaluado utilizando validación cruzada con 5 pliegues; con ello se descubrió que los mejores resultados los daba el clasificador SMO, ya que el porcentaje de instancias correctamente clasificada era un poco más alto en comparación a los otros clasificadores.

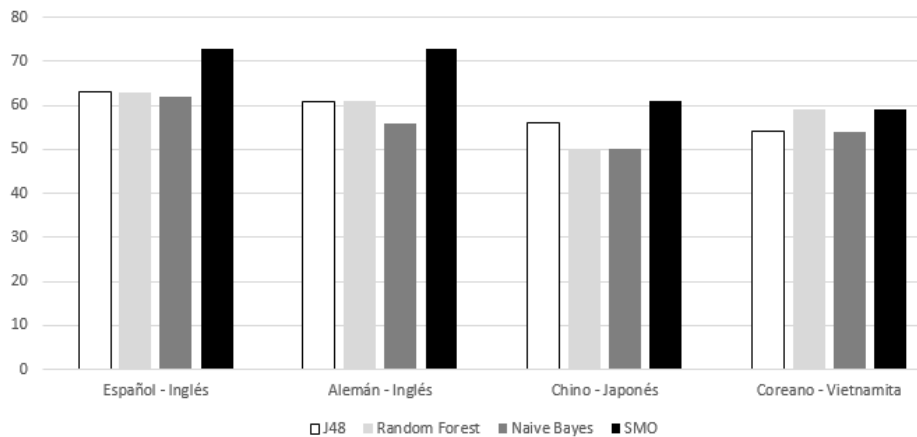


Fig. 6. Resultados obtenidos con diferentes clasificadores, para 4 duplas de idiomas; se observa como en promedio los mejores resultados son obtenidos con el clasificador SMO.

## 4. Resultados

### 4.1. Corpus OGI

Para realizar las pruebas, se hizo uso del “OGI Multilanguage Corpus” [14], el cual consta de diversos audios grabados a través de conversaciones realizadas por medio de teléfonos comerciales. Algunos de los idiomas que incluye el corpus son: Inglés, Español, Francés, Alemán, Japonés, Chino Mandarín, Tamil, Coreano, Vietnamita, entre otros; incluyendo un total de 175 llamadas por idioma. Se debe mencionar que la frecuencia de muestreo a la que se grabaron los audios fue a 8KHz, y la grabación fue realizada por medio de una contestadora automática.

### 4.2. Prueba experimental

Haciendo uso de la wavelet Daubechies 8, y aplicando el método de la EWR, se obtienen los estadísticos necesarios para que por medio del clasificador SMO, se evalúen las clases utilizando 5 pliegues. Con ello se procedió a aplicar el método a todas las duplas de idiomas para obtener el porcentaje de clasificación.

Tomando como referencia la tabla comparativa utilizada en los trabajos descritos en el marco de referencia, [1], [5], se logró determinar que en algunos casos los resultados obtenidos superaban el porcentaje obtenido en otros trabajos.

	Inglés	Alemán	Español	Chino	Vietnamita	Japonés	Coreano	Tamil	Farsi
Inglés	X	73	73	57	53	74	58	78	58
Alemán		X	58	59	53	48	56	62	60
Español			X	61	69	61	48	62	69
Chino				X	49	61	61	71	57
Vietnamita					X	53	59	59	53
Japonés						X	54	65	57
Coreano							X	64	54
Tamil								X	67
Farsi									X

Fig. 7. Resultados obtenidos en la aplicación del método descrito, a las duplas de idiomas.

	Inglés	Alemán	Español	Chino	Vietnamita	Japonés	Coreano	Tamil	Farsi
Inglés	X	73 (59.5)	73 (67.7)	57 (75)	53 (67.7)	74 (67.6)	58 (79.4)	78 (77.4)	58 (76.3)
Alemán		X	58 (59.4)	59 (62.2)	53 (65.7)	48 (65.8)	56 (71.4)	62 (69.7)	60 (71.8)
Español			X	61 (80.6)	69 (62.1)	61 (62.5)	48 (75.9)	62 (65.4)	69 (66.7)
Chino				X	49 (50)	61 (50)	61 (73.5)	71 (74.2)	57 (76.3)
Vietnamita					X	53 (68.6)	59 (56.2)	59 (71.4)	53 (66.7)
Japonés						X	54 (65.7)	65 (59.4)	57 (66.7)
Coreano							X	64 (62.1)	54 (75)
Tamil								X	67 (69.7)
Farsi									X

Fig. 8. Imagen comparativa entre los resultados obtenidos utilizando el método propuesto, y los resultados obtenidos por Rouas et al, que se encuentran ubicados entre paréntesis. Si el resultado obtenido supera al estado del arte, se sombrea en gris.

## 5. Conclusiones

A lo largo de este trabajo se presentó una técnica innovadora con la cual se pretende realizar la identificación de idiomas, basada principalmente en la obtención de la transformada wavelet de la señal. Complementando a la transformada wavelet, se realizó el cálculo de la energía wavelet relativa, con la que se puede determinar el grado de influencia de cada uno de los niveles de descomposición a la hora de realizar la clasificación.

Observando los resultados obtenidos durante la prueba, se puede comprobar que el método implementado tiene resultados alentadores, ya que en algunos casos los resultados obtenidos en este trabajo lograban superar a los obtenidos por el trabajo de referencia. La ventaja del método propuesto en el presente trabajo respecto a los que ya existen, es que para obtener los resultados de clasificación de instancias correctas solo fueron necesarios pocos datos por idioma (solamente los estadísticos de cada nivel, y los estadísticos de las deltas de tiempo), lo cual abre una puerta para continuar trabajando en él ya que esto permitirá que el procesamiento computacional sea menor, pudiendo dar paso a la realización de un sistema que trabaje en tiempo real.

Como trabajo a futuro se espera implementar diferentes técnicas para lograr una mejor clasificación entre las duplas de los idiomas, con lo cual se espera aumentar el porcentaje de instancias correctamente clasificadas; la primera será realizar un cambio

en como representar la energía de la señal de audio, ya que si se describe mejor se podrá mejorar la exactitud con que se clasifican los datos. Además es necesario incluir un paso de pre procesamiento de la señal para eliminar silencios, ya que eso afecta considerablemente el cálculo de cambio de la energía.

Otro experimento a realizar es en cuanto al cálculo de las deltas de tiempo, las cuales pueden calcularse por medio de una delta más generalizada, pues con ella se englobará más información de la señal, y no solamente representará el cambio entre dos puntos del tiempo, consecutivos o distantes, sino el cambio de energía entre más puntos de tiempo.

Por último se pueden aplicar técnicas de selección de atributos sobre la caracterización actual de la señal de audio, es decir se debe realizar un análisis de los atributos, tal como ganancia de información, mejorar la clasificación, ya que como se sabe, en algunos casos, unos atributos pueden reducir la precisión con la que un modelo puede llegar a clasificar las instancias de forma correcta.

## Referencias

1. Rouas, J.-L., Farinas, J., Pellegrino, F., André-Obrecht, R.: Modeling Prosody for Language Identification on Reas and Spontaneous Speech. IEEE ICASSP, Francia, pp. 40–43 (2003)
2. Rouas, J.-L., Farinas, J., Pellegrino, F., André-Obrecht, R. Rhythmic unit extraction and modeling for automatic language identification. Journal Speech Communication, Elsevier, pp. 436–456 (2005)
3. Hosseini Amereei, S.A., Homayounpour, M. M.: Using probabilistic characteristic vector based on both phonetic and prosodic features for language identification. Telecommunications (IST), 5th International Symposium on Tehran: IEEE, pp. 750–754 (2010)
4. Sadanandam, M., Prasad, V.K., Ramana, N., Rao, E.J.: New features using fuzzy c-means algorithm for automatic language recognition. Computational Intelligence and Computing Research (ICIC) Coimbatore: IEEE, pp. 1–5 (2014)
5. Reyes-Herrera, A.L.: Un Método para la Identificación Automática del Lenguaje Hablado Basado en Características Suprasegmentales (Tesis doctoral). Tonanzintla Puebla: INAOE (2007)
6. National Academy of Science: Wavelets: Ver el Bosque y los Árboles. Recuperado, Beyond Discovery: [http://www7.nationalacademies.org/spanishbeyonddiscovery/mat\\_008276-03.html](http://www7.nationalacademies.org/spanishbeyonddiscovery/mat_008276-03.html) (2003)
7. Cortazar-Martinez, O.: Procesamiento Digital de Imágenes Usando Wavelets (Tesis de Ingeniería en Electrónica y Telecomunicaciones). Pachuca de Soto, Hidalgo: Universidad Autónoma del Estado de Hidalgo (2006)
8. Torres, A.: Análisis de Señales Electroencefalográficas para la Clasificación de Habla Imaginada. Revista mexicana de Ingeniería Biomédica, Vol. 34, No. 1, (2013)
9. Morales-Mendoza, L.: Teorema de Muestreo. Procesamiento Digital de Señales, pp. 22–30, DICIS-UG (2011)
10. The University of Waikato: Use Weka in your Java code. Obtenido de <https://weka.wikispaces.com/Use+WEKA+in+your+Java+code#Classification-Building+a+Classifier> (2009)
11. Abernethy, M. (s.f.): Data mining with WEKA, Part 2: Classification and clustering. <http://www.ibm.com/developerworks/library/os-weka2/> (2015)
12. MathWorks: Multilevel 1-D wavelet decomposition. Obtenido de <http://www.mathworks.com/help/wavelet/ref/wavedec.html> (2006)

*Jesús A. Fortoul-Díaz, Ana L. Reyes-Herrera, Alejandro A. Torres-García, Luis Villaseñor-Pineda*

13. Jang, R.: Audio Signal Processing and Recognition. Obtenido de Reading Wave Files <http://mirilab.org/jang/books/audiosignalprocessing/matlab4waveRead.asp?title=4-2%20Reading%20Wave%20Files> (2005)
14. Cole, R., Muthusamy, Y.: OGI Multilanguage Corpus. Recuperado el 26 de Octubre de 2015, de University of Pennsylvania, Linguistic Data Consortium: <https://catalog ldc.upenn.edu/LDC94S17> (1994)