

# Reconocimiento de canto de aves basado en el análisis de componentes principales del espectrograma

## Bird chant's recognition based on principal components analysis of the spectrogram

Jessica González<sup>1</sup>, Jorge Padrón<sup>1</sup>, Iván Barbero<sup>1</sup>, Luis Custodio<sup>1</sup>, Fernando Merchán<sup>2\*</sup>

<sup>1</sup>Licenciatura en Ingeniería Electrónica y Telecomunicaciones, Facultad de Ingeniería Eléctrica, Universidad Tecnológica de Panamá

<sup>2</sup>Facultad de Ingeniería Eléctrica, Universidad Tecnológica de Panamá

**Resumen** En el presente artículo se propone un método de reconocimiento automático de canto de aves. Este método está basado en el análisis de componentes principales del espectrograma de los cantos, implementado mediante la técnica de eigenfaces. Para la preparación de la base de datos se implementó un método de detección de actividad acústica por medio de un filtro de media móvil. Se construyó una base de datos con tres cantos diferentes de tres aves: Formicivora Grisea, Harpia Harpyja y Reinita Protonotaria. Las pruebas realizadas muestran la efectividad del método propuesto, obteniéndose una tasa de reconocimiento de 100%.

**Palabras clave** Actividad acústica, análisis de componentes principales, canto de aves, *eigenfaces*, espectrograma.

**Abstract** This paper presents a method for automatic bird song recognition. This method is based on the principal components analysis (PCA) of bird chant's spectrograms, by using eigenfaces. For the database confection, we used a detection method for acoustic activity applying a moving average filter. We assembled a database of three chants of different birds: Formicivora Grisea, Harpia Harpyja and Reinita Protonotaria. The obtained results show the effectivity of the proposed method, obtaining a recognition rate of 100%.

**Keywords** acoustic activity, principal components analysis, bird chants, *eigenfaces*, spectrogram.

\* Corresponding author: fernando.merchan@utp.ac.pa

### 1. Introducción

Durante los últimos años el interés por preservar la fauna en una región ha aumentado y esto se debe al alto nivel de extinción que condena a las especies más importantes de una población. En Panamá se han desarrollado organizaciones en pro de la conservación de especies, como por ejemplo la Sociedad Audubón de Panamá, la cual promueve la conservación de aves y sus hábitats, por medio de educación, observación, investigación y participación comunitaria.

La definición oficial de una especie en peligro de extinción dada por la Ley N°24 del 7 de junio de 1995, es la siguiente: "Condición en la cual una población ha sido declarada como tal, por haber quedado reducida numéricamente a un nivel crítico, o cuyo hábitat ha experimentado una modificación considerable". Una especie es declarada en peligro de extinción cuando hay un decremento de seres por más de la mitad de su población en una década o por la restricción de su hábitat a menos de 5000 km<sup>2</sup>.

Las aves que se tomaron en consideración en este estudio fueron la *Harpia Harpyja*, la *Formicivora Grisea* y la *Reinita Protonotaria*.

El *Harpia Harpyja* es un ave que se encuentra en el nivel de peligro de extinción, categorizado por la UICN (Unión Internacional para la Conservación de la Naturaleza). El interés que existe por cuidar al Águila Arpía (*Harpia Harpyja*) radica en que es considerada el ave nacional de la República de Panamá, según la Ley 18 del 10 de abril de 2002.

En este trabajo se presenta un método de reconocimiento del canto de aves utilizando las componentes principales de su espectrograma. Este modo permite identificar o clasificar la especie avícola que se ingresa al sistema.

Las aves pueden reconocerse auditivamente mediante dos formas: cantos y llamados. Estos son sonidos no estacionarios. Los cantos de las aves son el medio que utilizan para relacionarse y esto tiende a ser complejo para identificar el ave que emitió el canto. Por otro lado, los llamados, son para comunicar información o señales de alarma entre miembros de la población. Normalmente los cantos están formados por "sílabas" o segmentos diferentes concatenados. Los llamados suelen estar compuestos por repeticiones de sonidos. Para el procesamiento de ambos tipos de formas se requiere detectar y segmentar los diferentes sonidos o sílabas, realizar

reconocimiento del mismo. Posteriormente se determina a que canto o llamado corresponde.

Se han realizado muchas contribuciones en el campo de reconocimiento de sonidos de aves. Estos trabajos normalmente operan a nivel de una representación en frecuencial-temporal sobre la que obtienen un conjunto de parámetros y sobre los cuales operan un método de clasificación. Por ejemplo, en [1], se utilizan los MFCC (Mel-Frequency Cepstral Coefficients) del espectrograma de canto y se obtuvieron caracterizaciones efectivas de las señales con cambios abruptos y discontinuos. Por otro lado, se realizó un estudio para el reconocimiento de las diferentes especies de piquituertos en [2]. El método consistió en grabar el canto de los piquituertos, se generó el espectrograma y se separó el ruido de el canto identificándolo por su contenido de energía. Una vez se obtuvo una señal limpia se realizó el estudio de los componentes del primer y segundo armónico, colocando estas características en un vector y clasificando el canto según el método SOM (Self Organizing Map). Este estudio es poco eficiente debido a que solo contempló los dos armónicos principales de la señal.

En [3], se utilizan dos técnicas Hidden Markov Models (HMMS) y Dynamic Time Warping (DTW) para el reconocimiento del canto del azulejo índigo y el diamante cebra. Estos últimos dos métodos son eficientes, aunque presentan baja efectividad cuando se presentan cantos de baja duración.

Este trabajo se presenta como una alternativa para el reconocimiento de sonidos de aves. Al igual que el método presentado en [2], el mismo opera a nivel del espectrograma de la señal. Sin embargo, consideramos el espectrograma como un patrón en dos dimensiones que busca caracterizar y representar. Para ello, proponemos utilizar el método de *eigenfaces*, que ha sido utilizado previamente para el reconocimiento de rostros [4]. En el mismo se parte de una base de datos de espectrogramas de las diferentes clases bajo estudio, para el presente caso los diferentes sonidos, y se efectúa un análisis de componentes principales. Estos componentes principales vienen a ser una base ortogonal con la que podemos representar cada una de las clases de espectrograma. Facilitando implementar luego un esquema de clasificación basado en los coeficientes correspondientes de cada clase en esta base ortogonal. A diferencia del método presentado en [2], se aprovecha toda la información frecuencial-temporal presente en el espectrograma y no solo las dos primeras armónicas.

El resto del artículo está organizado de la siguiente manera. La sección 2 presenta la metodología implementada en cada una de las etapas de procesamiento de las señales, a saber, la detección de sonidos de aves, la supresión de componentes de ruido y el reconocimiento. La sección 3 presenta las pruebas realizadas y resultados obtenidos. La sección 4 concluye el artículo.

## 2. Metodología

Para la implementación de este método se utilizó la herramienta Matlab de la compañía Mathworks. Esta brinda un lenguaje de computación técnico y una interfaz interactiva para el desarrollo de algoritmos, cálculo numérico y visualización de datos.

A continuación, describimos las tres etapas de este método: detección, supresión de ruido y reconocimiento.

### 1.1 Detección de sonido de aves

Para realizar el proceso de identificación de sonidos es necesario extraer los mismos de las grabaciones obtenidas en [5], base de datos de cantos de aves de nivel internacional.

Un ejemplo de estas señales es la presentada en la figura 1.

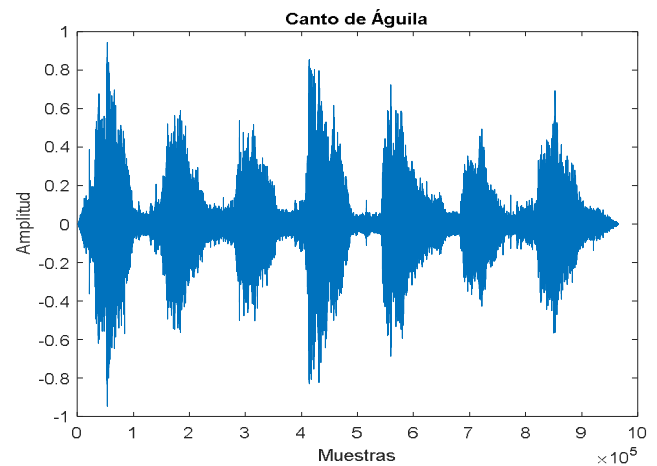


Figura 1. Grabación del canto de *Harpia Harpyja* en el dominio temporal.

Esta señal presenta espacios de tiempo en los cuales no tenemos información de interés, por lo que es necesario un método que nos permita automatizar el proceso de detección de sonido, se propone aprovechar los niveles de amplitud diferentes entre los momentos de canto y los momentos de silencio del ave. Se ha demostrado anteriormente el uso filtros promedio para el tratamiento de señales en dos dimensiones. El filtro promedio es utilizado para estimar una aproximación de la envolvente de la señal mediante la aplicación del filtrado de alto orden sobre el valor absoluto de la señal como podemos apreciar en la siguiente expresión:

$$y[n] = \sum_{i=0}^{10000} |x[n - i]| \quad (1)$$

Como se puede observar en la figura 2, se establece un umbral basado en la amplitud promedio de los periodos de silencio el cual nos sirve de referencia para poder diferenciar la amplitud en estos periodos de silencio, que corresponden al ruido de fondo, de la amplitud de los sonidos de aves, obtenido con el siguiente calculo:

$$umbral = \min(y[n]) \quad (2)$$

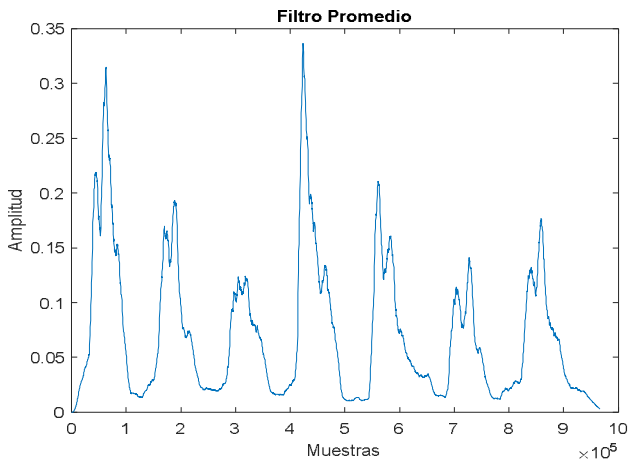


Figura 2. Señal resultante de filtro promedio.

Podemos establecer un comparador variable capaz de realizar una detección automática de estos espacios de tiempo que solo deje pasar los valores que se encuentren por encima del umbral obtenido.

En la figura 3 se puede observar una representación de los instantes de tiempo donde existen vocalizaciones de interés en la señal de audio del ave estudiada.



Figura 3. Reconocimiento de vocalización.

Con estas dos etapas, el cálculo de envolvente y la aplicación de umbral, podemos tener una buena representación del canto en el tiempo, pero no es suficiente para llegar a crear una clase de este canto.

### 1.2 Supresión de ruido

Una señal acústica contiene las componentes de interés en el dominio de la frecuencia por lo tanto se tiene que hacer un análisis de estos sonidos con respecto a su contenido en frecuencia a lo largo del tiempo. Para esto se generó un espectrograma de la vocalización que se muestra en la figura 5.

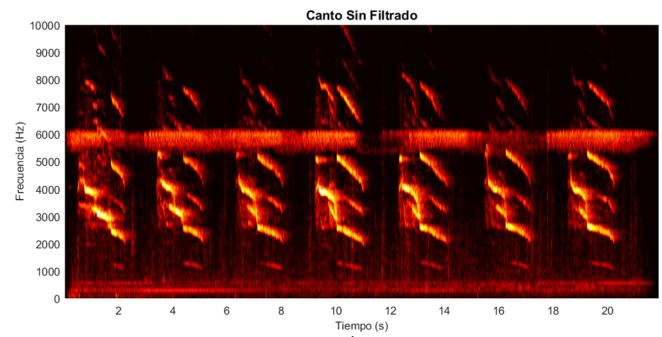


Figura 5. Espectrograma del canto del Águila Arpía.

En la figura 5 se puede observar que existe un componente de ruido de banda estrecha cerca de los 6000 Hz en esta señal. Este ruido afecta la clasificación adecuada del canto, por lo que se utilizó un filtro IIR rechaza banda con función de transferencia:

$$H(z) = \frac{(z - e^{j*2*\pi*fn})(z - e^{-j*2*\pi*fn})}{(z - 0.8e^{j*2*\pi*fn})(z - 0.8e^{-j*2*\pi*fn})} \quad (3)$$

$$\text{donde } fn = \frac{6000}{Fs}$$

Este filtro cuenta con una respuesta en frecuencia capaz de rechazar las frecuencias en la banda de 6000 Hz como se puede observar la figura 6.

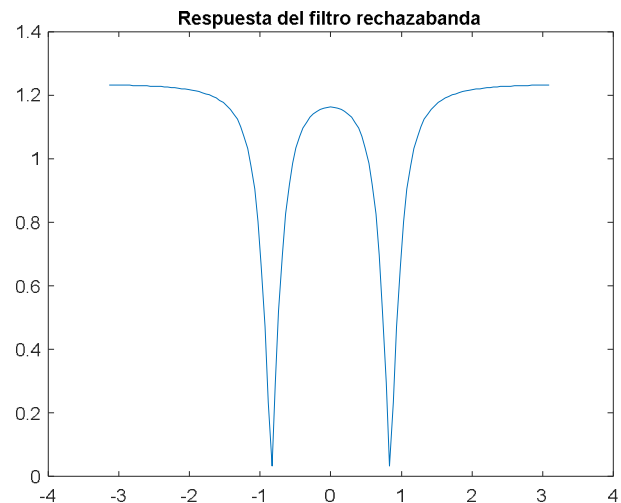


Figura 6. Respuesta en frecuencia del filtro rechaza banda.

Con estas etapas de detección de sonidos de aves y supresión de ruido se preparó una base de datos de espectrograma de tres tipos de sonidos de aves. Los sonidos son los siguientes: llamado de pedido del *Harpia Harpyja*, una sílaba del canto de la *Formicivora Grisea* y una sílaba de canto de la *Reinita Protonotaria*. Podemos ver ejemplos de estos sonidos en la figura 7, 8 y 9.

### 1.3 Método de reconocimiento de sonido

Para el reconocimiento de los distintos sonidos de aves se implementa un algoritmo basado en el método de *eigenfaces*, el cual está basado en el Análisis de Componentes Principales (PCA, por sus siglas en inglés). Este método de *eigenfaces* es ampliamente utilizado en el reconocimiento de rostros debido a su simplicidad y los buenos resultados obtenidos bajo condiciones controladas [4].

En nuestro análisis se utilizaron los espectrogramas característicos de cada caso, figura 7, 8 y 9, de donde se obtuvo los componentes principales, o *eigenvectors* de la matriz covarianza que mejor los representa.

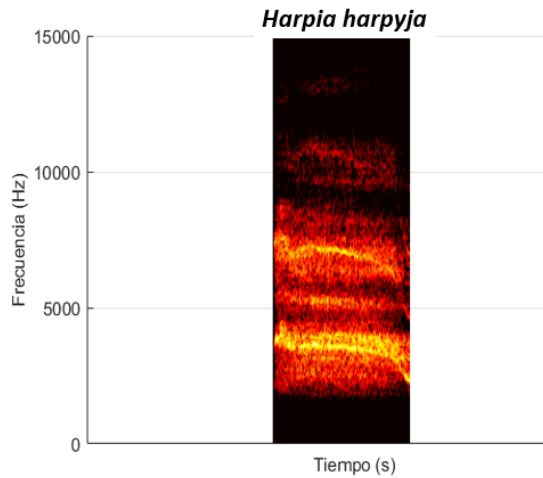


Figura 7. Tono del *Harpia Harpyja*.

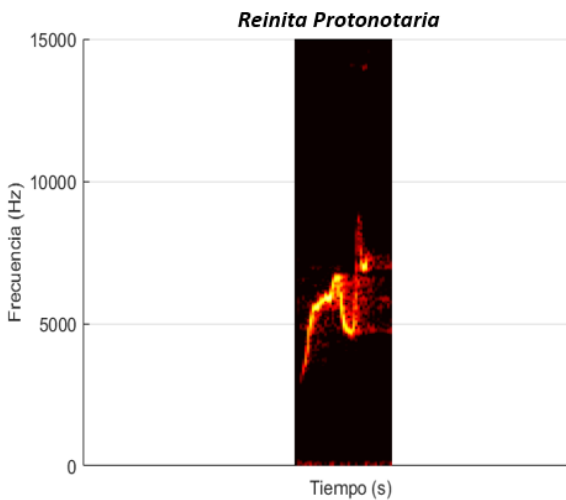


Figura 8. Sonido de la *Reinita Protonotaria*.

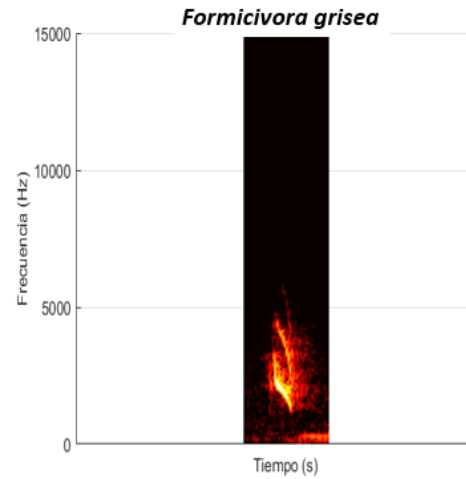


Figura 9. Sonido de la *Formicivora Grisea*.

En términos matemáticos se parte de un conjunto de imágenes de entrenamiento que corresponden a cada uno de los sonidos en la base de datos. Estas imágenes representadas por matrices cuadradas  $N \times N$  son reorganizadas como vectores  $N^2$  y definidas por la notación  $\Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_{M-1}, \Gamma_M$ . En este conjunto de vectores puede haber más de una imagen por tipo de canto.

El promedio  $\Psi$  de los vectores es calculado como:

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i \quad (4)$$

Al conjunto de vectores  $\Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_{M-1}, \Gamma_M$  se le resta la media como se muestra:

$$\Phi_i = \Gamma_i - \Psi \quad (5)$$

Estos vectores se agrupan en una matriz  $A$  de dimensiones  $N^2 \times M$  como se muestra:

$$A = [\Phi_1, \Phi_2, \Phi_3, \dots, \Phi_{M-1}, \Phi_M] \quad (6)$$

De la matriz  $A$  se calcula la covarianza, definida de la siguiente manera:

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = AA^T \quad (7)$$

La matriz covarianza  $C$  tiene un tamaño  $N^2 \times N^2$ , lo cual la hace muy grande como para ser procesada computacionalmente de una manera eficiente. Por ello se crea una nueva matriz  $L$  de dimensiones  $M \times M$  definida de la siguiente manera:

$$L = A^T A \quad (8)$$

De la ecuación (8) se obtienen los vectores propios  $v_k$  ( $k = 1, 2, 3, \dots, M$ ) de la matriz  $L$ . A partir de la combinación de estos vectores propios y los vectores  $\Phi_i$  se obtiene los vectores propios  $u_k$  ( $k = 1, 2, 3, \dots, M$ ) de la matriz covarianza  $C$ , como se muestra:

$$U = [u_1, u_2, \dots, u_M] = [\Phi_1, \Phi_2, \dots, \Phi_M][v_1, v_2, \dots, v_M] \quad (9)$$

De la matriz  $U$  solo se toman en consideración los vectores que presentan mayor energía, lo que permite reducir el tamaño de los vectores que representan cada tipo de canto en el subespacio PCA. Para el reconocimiento de cada uno de los casos se comparan los valores de entrenamiento versus los valores de prueba con la siguiente ecuación:

$$\omega_k = u_k^T [\Gamma - \Psi]; \quad k = 1, 2, 3, \dots, P \quad (10)$$

De la ecuación (11) se obtiene un nuevo vector  $\Omega^T = [\omega_1, \omega_2, \dots, \omega_P]$ , el cual describe la contribución de cada componente principal de los espectrogramas de entrada. Este vector es utilizado en nuestro algoritmo para reconocer los patrones característicos de cada tipo de canto.

El método seguido en este caso, por su simplicidad, es la búsqueda del vector de canto que minimice la distancia euclidiana definida como:

$$E_k = \|\Omega - \Omega_k\| \quad (11)$$

donde el vector  $\Omega_k$  describe la una de las clases de canto.

### 3. Pruebas y resultados

Se obtuvieron resultados positivos al utilizar el método de detección explicado en la sección 2.1. Este método se utilizó para las tres clases de aves dando como resultado la, señales como la mostrada en la figura 10.

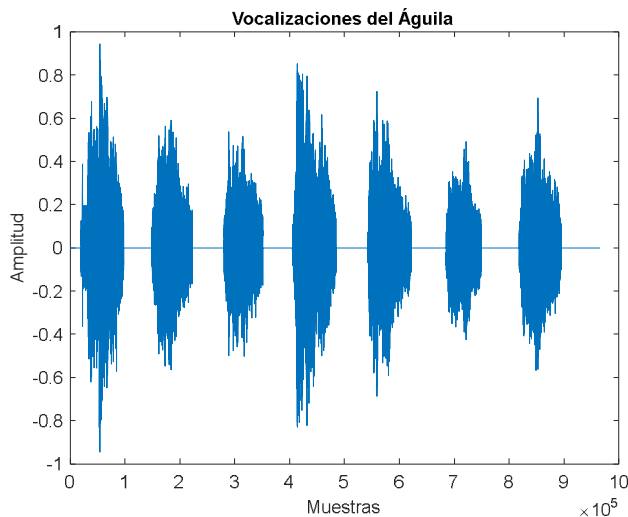


Figura 10. Vocalización del Águila Harpia Harpyja procesada.

La implementación de filtros rechaza banda fue de suma importancia para eliminar componentes no deseados que pueden llevar a una mala representación y clasificación del canto. Se puede observar el resultado de la aplicación del filtro sobre el canto del Águila *Harpia harpyja* en la figura 11. Este procedimiento se aplicó en aquellas señales que así lo requerían.

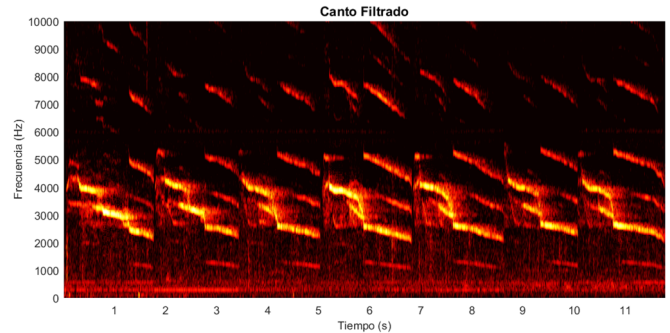


Figura 11. Canto del Águila *Harpia harpyja* filtrado.

En el experimento se utilizaron  $M = 12$  muestras de espectrogramas de sonidos como base de datos de entrenamiento. Estos son: cuatro muestras de la *Formicivora Grisea*, 4 de la *Harpia Harpyja* y 4 de la *Reinita Protonotaria*. Se utilizaron  $P = 12$  coeficientes asociados a los componentes principales de mayor propio. En base a esta información se generó una base de datos de pruebas con 2 sonidos de prueba para cada tipo de ave con un total de 6 señales de prueba. Cabe destacar que esta base de datos se manejó en el dominio de la frecuencia por medio del espectrograma de cada una de las señales de prueba. Además de utilizar señales de entrada con la misma cantidad de muestras con el fin de asegurar que los espectrogramas tuviesen las mismas dimensiones. Para esto se agregaron ceros antes y después de los vectores de cada señal según fuese requerido, asegurando que el sonido de ave se encontrase centrado. Esto es necesario ya que el método de *eigenfaces* es sensible a la traslación de la imagen.

Con la utilización de la ecuación (10) se pudo obtener los vectores característicos de cada ave como se muestra en la figura 12, 13 y 14.

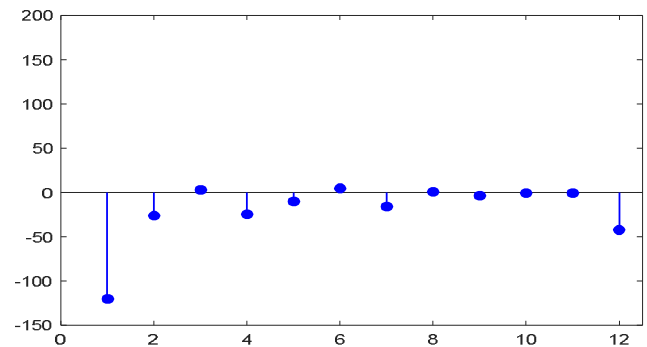


Figura 12. Vector característico de la *Harpia Harpyja*.

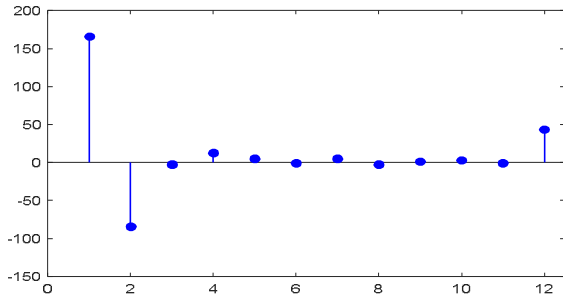


Figura 13. Vector característico del *Reinita Protonotaria*.

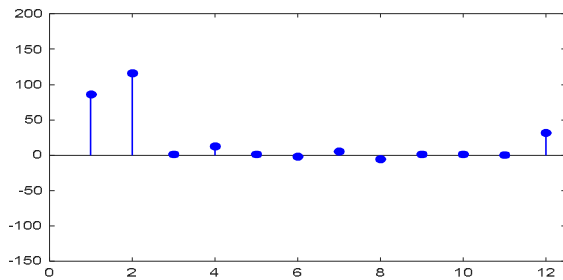


Figura 14. Vector característico de la *Formicivora Grisea*.

Para el análisis de estas señales se realizó la comparación de los vectores característicos con las señales de prueba y se usó la ecuación (11) para poder obtener la distancia euclidiana mínima, siendo esta última la base de la clasificación y determinación del tipo de canto que se registró en el sistema.

Al aplicar el esquema de clasificación propuesto, basado en la selección de la clase de sonido según distancia euclidiana mínima, las 6 señales de prueba fueron clasificadas correctamente sin obtener ningún falso positivo como puede observarse en la Tabla 1.

Tabla 1. Clasificación de 6 señales de prueba

		Señales de prueba					
		Harpia Harpyja		Reinita Protonotaria		Formicivora Grisea	
		P1	P2	P1	P2	P1	P2
Señales de entrenamiento	Harpia Harpyja	280	268	251	274	247	232
	Reinita Protonotaria	562	549	235	254	329	309
	Formicivora Grisea	525	512	299	320	233	225

#### 4. Conclusiones

En este trabajo se ha implementado y probado el método de *eigenfaces* para el reconocimiento de espectrogramas de sonidos de aves. Los resultados obtenidos demuestran la

efectividad del método en la base de pruebas utilizadas. Quedan como perspectivas de este trabajo probar el método propuesto una base de datos de sonidos de datos más amplia y comparar el rendimiento del mismo con otros métodos.

Este método se puede mejorar complementándolo con un sistema de GPS que ayude a filtrar las probabilidades de las especies de ave según el lugar geográfico donde se detecte el canto. Esto puede ser implementado como aplicación móvil para ser utilizado en teléfonos inteligentes y otros dispositivos. El detector de cantos de ave además de tener un objetivo científico para el estudio de la fauna existe la posibilidad de un uso recreativo para los fanáticos de avistamiento de aves y de la ornitología.

#### AGRADECIMIENTO

Agradecemos a los docentes de la Facultad de Ingeniería Eléctrica por su asistencia y colaboración en el desarrollo de este trabajo de investigación.

#### REFERENCIAS

- [1] S. S. Chen and Y. Li, "Automatic Recognition of Bird Songs Using Time-Frequency Texture" 2013 5th International Conference and Computational Intelligence and Communication Networks, 2013, pp. 262-266.
- [2] Juha T. Tantu, Jari Turunen, Mikko Ojanen. "Automatic Classification of Flight Calls of Crossbill Species (*Loxia spp.*)", In Proceedings of the 1st International Conference on Acoustic Communication by Animals, 2003.
- [3] Joseph A. Kogan and Daniel Margoliash, "Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study", Journal of Acoustic Society of America, 103(4), 2185-96, 1998.
- [4] M. Turk, A. Pentland, "Eigenfaces for Recognition" Journal of Cognitive Neuroscience, vol. 3, number 1, pp.71-86, 1991.
- [5] Xeno-Canto- página web oficial. [Online]. Disponible en: <https://www.xeno-canto.org/>