

# ESCUCHANDO A LA NATURALEZA: DEL RECONOCIMIENTO DE VOZ A LA BIOACÚSTICA

***Ángel David Pedroza Ramírez***

Universidad Autónoma de Zacatecas, Unidad Académica de Ingeniería Eléctrica  
P.A.D\_16@hotmail.com

***José Ismael de la Rosa Vargas***

Universidad Autónoma de Zacatecas, Unidad Académica de Ingeniería Eléctrica  
joseismaelrv@gmail.com, ismaelrv@ieee.org

***Rogelio Rosas Valdez***

Universidad Autónoma de Zacatecas, Unidad Académica de Ciencias Biológica  
rogrosas@gmail.com

## Resumen

La bioacústica es la rama de la ciencia que, mediante información acústica, se encarga del estudio de la forma de transmisión y recepción de información biológica con el fin de alcanzar desde la identificación de especies hasta la determinación de la salud de un ecosistema. Algunos desarrollos recientes se han enfocado en la adecuación y aplicación de técnicas de reconocimiento de voz tales como el uso de Modelos Ocultos de Markov, Redes Neuronales, entre otros; con el fin de lograr el reconocimiento automatizado de especies. En esta revisión se presentan algunos avances tecnológicos en el área así como una visión global sobre las herramientas matemáticas disponibles para, mediante ellas, lograr algunos de los objetivos que la bioacústica pretende alcanzar.

**Palabra(s) Clave(s):** Bioacústica, Biología, Identificación de especies, Reconocimiento Automático de Voz.

## 1. Introducción

La Bioacústica es la ciencia en donde confluyen la biología y la acústica para investigar de forma no invasiva la producción y recepción del sonido biológico, así

como los métodos para la transmisión de información por medio de la acústica y su respectiva propagación en el medio. En esta área de estudio se relacionan otras como la biología, taxonomía, sistemática, ecología del paisaje sonoro, computación y electrónica.

Esta ciencia provee de técnicas de bajo costo para el monitoreo de la biodiversidad de áreas remotas [1]. Algunos de los estudios más comunes se enfocan en la identificación de especies tales como murciélagos, ballenas, delfines, insectos, aves y anfibios [2]. Así como también en la determinación de la actividad biológica del fondo marino [3].

El continuo desarrollo, ha permitido crear estudios más avanzados sobre la población de aves (censos) y sobre la forma de prevenir la reducción de poblaciones del hábitat silvestre por actividad antropogénica (todo ello sin perturbar el ecosistema). El avance implica por tanto monitoreo de población, estimación de abundancia, riqueza y demografía así como la identificación de especies vulnerables o en riesgo de extinción, estudios taxonómicos, selección sexual y ecología del comportamiento [4].

Específicamente enfocado en el estudio de la conservación de las aves, México ocupa el décimo lugar a nivel mundial en riqueza de aves (según un estudio realizado en 2003) (figura 1). Esto coloca a México en el onceavo lugar entre los países mega diversos del mundo de acuerdo a su riqueza avifaunística (entre 1,123 y 1,150, y en el cuarto en cuanto a la proporción de especies endémicas, es decir entre 194 y 212 es el número de especies de aves que solo se encuentran en nuestro país [5]). En este sentido, las especies endémicas son las que poseen mayor importancia en la conservación; no obstante, entre 298 y 388 especies (26-33%) de la avifauna mexicana se encuentra en alguna categoría de amenaza de acuerdo a autoridades nacionales o internacionales [5,6]. Su extinción por tanto representa una pérdida global significativa.

La riqueza de especies endémicas en nuestro país se concentra en las zonas montañosas del eje Neo volcánico, la Sierra Madre Occidental y del Sur, y la Planicie Costera del Pacífico (figura 1).

En el año 1600 se extinguieron aproximadamente 103 especies de aves en el mundo (24 especies en los últimos 200 años en México). Una de cada 10 aves está en categoría de amenaza con lo cual se cree que para los próximos 100 años el 12% de la fauna mundial enfrentará riesgos (en México existen aproximadamente 40 especies amenazadas).

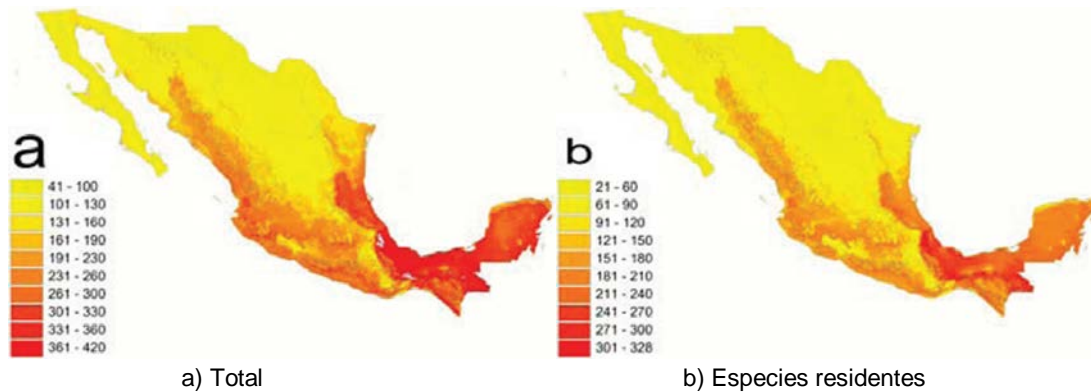


Figura 1 Distribución de riqueza potencial de aves en México [5].

El desarrollo continuo en el área de la bioacústica sigue siendo hoy promisorio “en teoría” [7]. Aunque existe más de una década de investigaciones en bioacústica, existen diversos enfoques para llevar a cabo la identificación debido a que la literatura concerniente se encuentra dispersa entre biología e ingeniería. A pesar de que existen avances tecnológicos para la aplicación en la conservación de especies, algunas de ellas resultan muy costosas tanto en espacio y tiempo (como el caso de la telemetría la cual es utilizada para el seguimiento de aves, seguimiento de rutas migratorias, resguardo de anidación, entre otros). Aunado a esto, aunque no existe un registro de la cantidad de investigadores dedicados al monitoreo y conservación de la biodiversidad, en México son escasos los estudios sobre sistemas que colecten e identifique cantos de manera automatizada creando un nicho de oportunidad para la creación de nuevas aplicaciones útiles para esta rama de la investigación [6].

El presente artículo tiene como objetivo principal el hacer una revisión sobre los avances tecnológicos así como la de brindar una visión global sobre las herramientas matemáticas disponibles para alcanzar algunos de los objetivos que

la bioacústica pretende alcanzar: lo cual es, desde el crear aplicaciones portables para teléfonos inteligentes (capaces de realizar la identificación de especies en tiempo real mediante la grabación, almacenamiento, identificación y geolocalización de vocalizaciones con el fin de ampliar las bases de información disponibles en los centros de investigación) hasta imágenes satelitales de monitoreo continuo para la recolección de información ambiental [8].

El artículo se organiza de la siguiente manera. En la sección 2 se presenta un resumen sobre los objetivos que la bioacústica pretende alcanzar. En la sección 3 se explica la estructura para el diseño de un reconocedor. En la sección 4 se presentan algunas de las herramientas matemáticas utilizadas en reconocimiento de voz. La sección 5 presenta algunos de los avances en materia de reconocimiento en bioacústica. Finalmente en la sección 6, se presentan algunas conclusiones.

## **2. Enfoques y aplicaciones en bioacústica**

En un inicio, las investigaciones enfocadas en la identificación de especies se realizaban por medio de inspección anatómica (observaciones en campo por un observador). Para realizar estas observaciones era necesario ver y capturar al espécimen para poder identificarlo por medio de una cámara (disponiendo siempre de una línea de vista, es decir sin obstrucción, entre el observador y el espécimen observado). Aunado a esto, el observador necesitaba contar con el camuflaje adecuado a una distancia considerable con el fin de no perturbar al entorno.

Como se puede notar para la identificación anterior, es necesario contar con la experiencia del investigador (para darle fiabilidad a la inspección). Algunos estudios se han centrado en probar la variabilidad que presentan los escuchas entrenados (bajo ciertas condiciones y para ciertas especies). En estos se pudo concluir que la mayoría de los escuchas entrenados solamente fallan en la identificación de aves poco comunes o raras [7]. Sin embargo, esto conlleva ciertas desventajas como la dispersión del conocimiento o el tiempo de adquisición para cubrir diferentes áreas de ecosistemas (con los cuál se hace necesario una mayor cantidad de observadores calificados) y por último, el posible sesgo

introducido por el observador al detectar, identificar y registrar las observaciones de un ecosistema.

Sin embargo, tras el nacimiento de la bioacústica como método de identificación de aves, algunas de las investigaciones, han optado por el utilizar las vocalizaciones de las aves como característica para lograr dicha tarea. Del avance en estos estudios, dentro de la bioacústica, se pueden identificar grandes áreas de aplicación, entre las cuales (algunas de ellas aún sin resolver de forma eficiente) se pueden identificar las siguientes [1]:

- **Identificación de especies en campo.** En la mayoría de las aves se ha podido identificar que son vocalmente más activas durante los periodos de reproducción. Esto hace posible la detección de aves aún a grandes distancias, con lo cual además de su identificación, se hace posible llegar a determinar el tamaño de población de especies (lo cual representa un método económico para estimar biodiversidad).

Ejemplo de la eficiencia de este método, es el mencionado sobre la identificación de 85% de las especies de aves mediante un registro de apenas 5 días realizado en el amazonas boliviano (el método de captura con redes requirió 54 días de campaña llevada a cabo por 7 ornitólogos).

- **Identificación de individuos, sexo y mapeo de territorios.** Para el caso de la identificación individual es necesario, dentro de la generalidad uniforme de sonidos, poder diferenciar matices de canto y por tanto, con base en esas diferencias, lograr una identificación individual (lo cual posibilitaría el conteo de animales de una especie por zonas).

Para poder diferenciar entre sexos en una población de una misma especie es necesario conocer diferencias específicas en los parámetros de canto. Los sistemas de identificación de sexos han de poder por tanto tomar en cuenta diferencias en características, las cuales van desde la frecuencia fundamental (relacionado con el tamaño de la siringe, órgano que permite la variación del canto en las aves), presencia de armónicos, entre otros.

En cuanto al mapeo de territorios, una vez que se llevó a cabo la identificación individual, se puede realizar mediante “perchas” (registros de

identificación) o por cambios de respuesta (parecidos a los de identificación de especies en campo) en diferentes posiciones geográficas.

- **Estimación de densidad absoluta y relativa.** En algunas condiciones será necesario el conocer la densidad absoluta de la población, sin embargo, el costo para llevar a cabo esta tarea resulta elevado (aunado a la variabilidad reducida presente en algunas ocasiones). La estimación de densidad absoluta se calcula mediante la detección de cada uno de los sujetos que están próximos al observador así como la distancia a la cual se encuentra cada uno de ellos respecto de esta referencia. Por otro lado, para llevar a cabo el índice relativo de la densidad, solo es necesaria que una misma proporción de individuos de la población sea registrada.

Aunque la segunda tarea puede parecer más sencilla que la primera, la detectabilidad de los sujetos (en este caso aves) no es siempre la misma y varía según sea el sexo, época del año, condición climática, hora del día, tipo de ambiente, entre otros. En este punto es posible estandarizar esta tarea tomando en cuenta solo los cantos presentes a cierta distancia del observador (por ejemplo los cantos detectados a menos de 50 metros) y evitando condiciones climatológicas extremas (viento y lluvia).

- **Estimación de la tasa de mortalidad.** La existencia de los machos a través del tiempo en un área, es un índice que permite estimar la tasa de mortalidad (para llevar a cabo esta identificación de machos se utilizaban tradicionalmente medios artificiales).

Esto solo es posible para aquellas especies cuya marca sea invariante a través del tiempo, es decir, siempre la misma a lo largo de toda la vida del sujeto estudiado.

- **Atracción y ahuyentamiento por playback .** La atracción por playback se puede aplicar en la marcación de individuos y para una captura selectiva (dado que la grabación es especie-específica). De esta forma las aves podrían ser atraídas hacia sitios seguros y por ende ayudar en la conservación de especies las cuales son atraídas por los parlantes.

Por su parte, en cuanto al ahuyentamiento, este método es utilizado cuando se requiere evitar en cierta área la presencia de aves. Este método es el sustituto de técnicas que implicaban barreras o métodos pirotécnicos. En su lugar, se producen sonidos naturales o de gran intensidad capaces de espantar a las aves y evitar así pérdidas económicas o posibles accidentes en áreas donde su presencia sería perjudicial. Para este último caso, es necesario tomar en cuenta la posible extinción de respuesta asociada al estímulo; por lo que será necesario cambiar de playback con frecuencia, aunque se recomienda la combinación de varios métodos para asegurar la efectividad del espantamiento.

### **3. Visión global y específica**

Con el fin de determinar el estado de un ecosistema se ha buscado diagnosticar a las especies en el hábitat, así como sus relaciones macro sistemáticas y densidades de población relativa, lo cual conlleva al desarrollo de herramientas tecnológicas que permitan lograr la identificación de las especies con el fin de crear sistemas más complejos lo más automatizados posibles [9, 10, 11].

La mayoría de las metodologías que se utilizan en el reconocimiento del habla son la base para el monitoreo bioacústico [12]. Sin embargo, algunas de las técnicas utilizadas en sistemas de reconocimiento automático (cuyo enfoque es la conversión de voz a texto) fallan en la capacidad de reconocimiento en ambientes naturales por lo que es necesario hacer uso de modificaciones a las técnicas avanzadas de reconocimiento para adaptarlas al reconocimiento en bioacústica (como el utilizado en [13]).

Los sistemas de reconocimiento de voz utilizados en la actualidad pueden variar en estructuras tan diversas como el diseñador lo desee [14]. Sin embargo, de entre la variabilidad existente, algunos son los procesos fundamentales que deben de poseer para poder ser considerados como sistemas de reconocimiento. Al respecto, se pueden considerar los siguientes procesos:

#### **Adquisición**

Para el caso del reconocimiento en bioacústica este proceso consiste entre otros, en la aplicación de filtros capaces de suprimir el ruido externo en las

grabaciones, así como del conocimiento especie-específica para realizar una segmentación y ventaneo apropiado.

### **Extracción de características**

Una vez que la señal de voz ha sido adecuada en el pre-procesamiento, la etapa siguiente corresponde a la extracción de la información contenida en la señal de voz. Esto se realiza con el fin de obtener una secuencia de vectores o parámetros que describan la información de interés contenida en la señal. Es necesario aclarar que entre más parámetros posean estos vectores, más difícil se vuelve su tratamiento e implementación.

A este proceso se le conoce como parametrización o caracterización y consiste en obtener las características propias o distintivas de la señal de voz en análisis con el fin de poder realizar de manera más sencilla las comparaciones necesarias para llevar a cabo el reconocimiento. Sin embargo, el método utilizado para llevar a cabo este proceso depende de la aplicación en que se busca aplicar el reconocimiento.

En el caso de la bioacústica, este parámetro es determinado según el tipo de reconocimiento que se lleva a cabo (llamados, cantos, etc.), lo cual está directamente relacionado con la especie-específica a identificar y qué parámetros considerar como característicos

### **Clasificación**

En esta etapa se lleva a cabo propiamente la identificación de la palabra (o sonido). Este proceso requiere de dos fases: entrenamiento y prueba (en este punto es posible utilizar los mismos métodos de reconocimiento del habla en bioacústica pero, como se mencionó anteriormente, con sus respectivas modificaciones).

## **4. Herramientas disponibles (visión general)**

De entre las principales técnicas aplicadas en reconocimiento de voz se tienen:

- **Cuantificación Vectorial (VQ).** El resultado de utilizar ya sea bancos de filtros, análisis LPC, etc., son vectores característicos que representan las



características espectrales variantes en el tiempo de acuerdo a la señal de voz. Denotaremos a los vectores espectrales como  $X_i$ , en donde típicamente cada vector es de una dimensión  $p$  (o al menos  $p$ ), ecuaciones 1.

$$\begin{aligned} X_1 &= A_0^1, A_1^1, A_2^1, \dots, A_p^1 \\ X_2 &= A_0^2, A_1^2, A_2^2, \dots, A_p^2 \\ &\vdots \\ X_L &= A_0^L, A_1^L, A_2^L, \dots, A_p^L \end{aligned} \quad (1)$$

La cuantificación vectorial es la generalización de la cuantificación escalar y en ella la cantidad total de información es tratada y almacenada en vectores los cuales representarán al total de información [15]. Entre las ventajas de la representación por Cuantificación Vectorial se encuentran el que se reduce el almacenamiento para el análisis de la información, el tiempo de cálculo para determinar los vectores de análisis espectral y que los cálculos de similitud son reducidos a solo una tabla de estimación. Por otro lado, entre las desventajas se encuentra el que se crea una distorsión espectral inherente en la representación del análisis del cada vector.

Se habla de un cuantificador óptimo cuando la distorsión promedio es mínima al cuantificar las secuencias de vectores. Para poder saber la eficiencia de un cuantificador es necesario conocer la distorsión promedio producida por dicho cuantificador, la cual se expresa como:

$$D = \frac{1}{L} \sum_{l=1}^L d(X_l, q(x-l)), \quad (2)$$

En donde  $D$  es la distorsión promedio y  $d( )$  es una medida de distorsión o distancia espectral.

- **Modelos Ocultos de Markov (HMM).** La aproximación de patrones no modela estadísticamente a las señales de voz por lo que posee ciertas restricciones. Por otro lado las técnicas estadísticas han sido aplicadas en el agrupamiento para crear patrones de referencia. Mediante ello se mejora

la clasificación de patrones y se realiza una simplificación dado que se utilizan múltiples señales de referencia y se caracterizan mejor las variaciones de diferentes pronunciaciones.

Una de las herramientas que permiten caracterizar estadísticamente las propiedades de un patrón o una señal de voz son los Modelos Ocultos de Markov (MOM ó HMM) [16,17]. Los HMM son considerados una extensión de las Cadenas Ocultas de Markov caso continuo o discreto de las Cadenas de Markov).

Para entender el proceso de las cadenas de evidentes de Markov, se considera como tal a una sucesión de pruebas a intervalos regulares de tiempo. Este proceso genera una sucesión de estados  $A_1, A_2, \dots, A_n$  llamado espacio de estados. En cada prueba ocurre un cambio de estado de acuerdo a un conjunto de probabilidad que se asocia a ese estado. Estos cambios ocurren en tiempos discretos  $t = 1, 2, 3, \dots, n$  y su estado asociado es denominado  $X_t$

Para aplicar un HMM, como se mencionó anteriormente, es necesario tomar en cuenta dos etapas: entrenamiento y prueba. Por otro lado, un modelo oculto de Markov está caracterizado por [18]:

- ✓ *El número de estados en un modelo (N):* En este sentido, a los estados individuales se les describe con la letra  $q$  en un determinado tiempo  $t$ .
- ✓ *El número de símbolos distintos de observación en cada estado (M):* Al respecto, al conjunto de observaciones se le denota como  $O = o_1, o_2, \dots, o_M$ , mientras que a los símbolos individuales se les denota como  $V = v_1, v_2, \dots, v_M$ .
- ✓ *La matriz de probabilidad de transición de estados ( $A = [a_{ij}]$ ):* Esta matriz expresa la probabilidad de pasar de un estado a otro. Es decir:

$$a_{ij} = P \left[ q_{t+1} = \frac{j}{q_t} = i \right] \quad (3)$$
$$1 \leq i \leq N$$

$$1 \leq j \leq N$$

- ✓ La distribución de probabilidad de los símbolos en el estado

( $B = [b_j(k)]$ ): Es decir:

$$b_j(k) = P \left[ O_t = \frac{v_k}{q_t} = j \right] \quad (4)$$

$$1 \leq k \leq M$$

- ✓ La distribución de probabilidad inicial de estados ( $\pi = [\pi_i]$ ): Este parámetro indica la probabilidad de iniciar en un estado cualquiera:

$$\pi_i = P[q_i = i] \quad (5)$$

$$1 \leq i \leq N$$

Con estos valores el HMM puede generar una secuencia de observaciones. De manera compacta se puede expresar como:

$$\lambda = (A, B, \pi) \quad (6)$$

- **Modelos de Mezcla de Gaussianas (GMM).** Para el caso de HMM continuos, las observaciones se modelan con PDFs que son una mezcla o combinación de funciones de densidad de probabilidad gaussianas [19]. El utilizar funciones de este tipo mejora el modelado estadístico del conjunto de vectores de observación que se pueden producir desde un estado cuando la dispersión de esos vectores es grande.

En el caso específico de los HMM continuos se calcula un vector de medias y de covarianza para cada una de las gaussianas de la mixtura de cada estado. Esto se logra conociendo el vector de observaciones correspondiente a la gaussiana de un estado específico. Experimentalmente esta técnica es adecuada para el reconocimiento del habla pero es necesario una gran cantidad de datos de entrenamiento (dada la cantidad de parámetros a entrenar) y utilizar un tiempo de ejecución significativamente mayor que con otros métodos.

- **Redes Neuronales.** Las redes neuronales (Neural Networks - NN) consisten en un gran número de unidades de cálculo simple, las cuales se encuentran altamente interconectados entre sí [20]. Ellas pretenden

interconectar un conjunto de unidades de proceso (o neuronas) en paralelo de forma similar a como lo hace el ser humano (obteniendo también prestaciones similares en reconocimiento tanto en tiempo de respuesta como en tasa de error). Este método es útil cuando se desean evaluar varias hipótesis en paralelo y como herramienta de clasificación de vectores característicos así como filtrado de señales de voz.

La idea básica detrás de una red neuronal es: dados una serie de parámetros, combinarlos con la finalidad de predecir un cierto resultado. El modelo de la neurona queda descrito mediante:

$$a = f(pw + b) \quad (7)$$

En (7),  $f\{\}$  representa la función de transferencia o de activación la cual puede ser lineal o no lineal,  $w$  representa el peso (o sinapsis de entrada),  $p$  es la señal de entrada,  $b$  es la señal umbral y  $a$  es la salida o axón. En esta, las funciones de transferencia pueden ser lineal, simétricas, log-sigmoidal, positivo lineal, entre otros.

El tipo de red se define según la forma en que se conectan las neuronas, el tipo de neurona que lo conforma y la forma de entrenamiento de la red. Para el caso de reconocimiento de patrones, se tienen tres tipos de redes principales: Red Perceptrón (feed forward), Red Hamming y Red de memoria asociativa (Hopfield).

Las unidades de proceso son de varios tipos pero, la más utilizada, dispone de varias entradas y la salida (figura 2) es el resultado de una transformación no lineal a la combinación lineal de las entradas. La ecuación correspondiente son dos ecuaciones 8 y 9.

$$n = w_{1,1}P_1 + w_{1,2}P_2 + \dots + w_{1,R}P_R \quad (8)$$

$$a = f(n + b) \quad (9)$$

Sin embargo, estructuras simples, no será suficiente cuando se tenga un mayor número de entradas por lo que se requerirá una estructura más compleja. A este tipo de estructura se le llama capa (la cual opera en paralelo). Si se aplica esta estructura en varias neuronas se puede llegar a estructuras más complejas (figura 3).

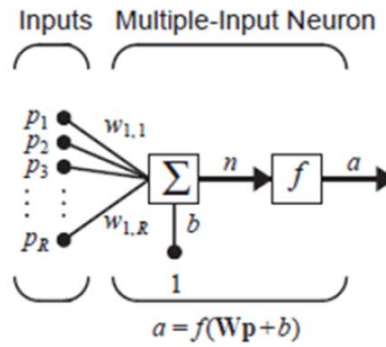


Figura 2 Neurona de entrada múltiple [21].

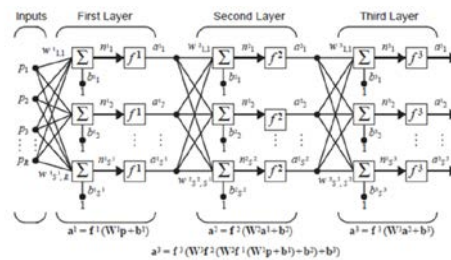


Figura 3 Red de múltiples capas [21].

De esta, el tipo de red que mejor resultado ha obtenido en reconocimiento automático del habla es el “perceptrón multicapa” (MLP) [22]. En ella las neuronas poseen capas: una capa de entrada (operando con los vectores de observación directamente), una capa de salida (apuntando a la palabra reconocida), y las respectivas capas intermedias. El enlace entre capas se da en base a ciertos pesos específicos entre neuronas. En el caso en que se requiere de un número de capas intermedias con varias neuronas cada una se clasifica como “red neuronal profunda” [23]. En ellas la idea básica es que con más capas y neuronas cada una produzca una mejor predicción del conjunto de datos.

## 5. Avances tecnológicos de reconocimiento en bioacústica

Al igual que en reconocimiento de voz, y como se mencionó anteriormente, existe una gran variabilidad de estructuras para llevar a cabo el reconocimiento en bioacústica las cuales van desde los métodos de adquisición hasta los esquemas completos de reconocimiento. Algunas investigaciones, por ejemplo, se han enfocado en utilizar información frecuencial para clasificar las señales bioacústica.

En una investigación, se se utilizó como método de codificación de datos una vectorización por espectrogramas [24]. Esto consiste básicamente en transformar la información del dominio temporal al dominio frecuencial, lo cual es denominado el análisis de espectro (mediante la transformada de Fourier), para después codificar dicha información en vectores. Este tipo de representación permite el observar elementos (frecuencias) que componen a un canto de ave y determinar mediante ellos, diferencias individuales, de poblaciones, entre otros. Este análisis ya ha sido estudiando con anterioridad sin embargo, mediante la vectorización propuesta, se logra mejorar el proceso de extracción de información.

Para dicha investigación, se utilizaron redes neuronales como método de clasificación siendo, la que mejor resultado brindó, la red perceptrón multicapa (MLP) con un 100% de éxito y 91% para la generalización (datos diferentes a aquellos mediante los cuales se entrena la red). Se entrenó con el algoritmo de retro propagación y gradiente descendiente con momentum.

Por otro lado, en otra investigación diferente, se utiliza el análisis puramente temporal mediante la TDSC (time domain signal coding [25]). Esto consiste en crear un alfabeto único mediante la segmentación por cruce en cero, duración de muestras y formas (todo esto en el dominio temporal). Entre las ventajas principales de esta técnica se tienen que se pueden generar características invariantes en el tiempo aunado a que la información frecuencial puede ser extraída sin sufrir limitantes de tiempo-banda de los métodos espectrales.

Esta investigación se enfocó en la identificación de 13 especies de grillos diferentes con un 100 por ciento de éxito y cero erróneas bajo condiciones de bajo nivel de ruido. Esta misma estructura se planteó para 25 especies más y 10 especies de aves japonesas (lo cual indica que una misma red puede ser utilizada para la identificación de diferentes especies). Para este caso se utilizaron una red neuronal con 784 neuronas en la capa de entrada, 20 en la capa oculta y 13 a 25 en la capa de salida. Para aves se clasificó por encima del 90% mientras que, para el caso de los grillos (bajo condiciones de poco ruido) la identificación es muy alta, entre 99 y 100% (para una identificación con un SNR variable ver [8]).

Algunos otros han optado por utilizar representaciones totalmente distintas a las frecuenciales o temporales. En donde, por medio de LPC, MFCC, etc., se identificaron 8 perros, 11 ranas y 13 aves (obteniendo 94% de eficiencia para perros, 90% para ranas y 71% en aves mediante el uso de redes neuronales) [13].

## **6. Conclusiones**

Con base en la investigación anterior se puede concluir que las técnicas de reconocimiento de voz pueden ser usadas como herramientas básicas para sistemas de reconocimiento sofisticados enfocados en bioacústica. Sin embargo, es necesario realizar algunas modificaciones debido a las condiciones de ruido en las que se llevará a cabo la identificación acústica.

Los métodos utilizados en reconocimiento enfocado en bioacústica poseen fortalezas y debilidades para cada caso según sea el enfoque de reconocimiento deseado. Sin embargo, algunos de los problemas encontrados en la mayoría de las investigaciones de la literatura consultada radican en la imposibilidad para lograr una identificación en condiciones reales. Esto conlleva a que algunas de las investigaciones posean buenos desempeños en pruebas de laboratorio (y simulaciones) pero bajo desempeño en pruebas de tiempo real debido al uso de esquemas ineficientes para extraer el ruido de la señal de interés.

Al igual que en reconocimiento de voz, las estructuras de reconocimiento en bioacústica han de ser diseñadas dependiendo el reconocimiento que se pretende alcanzar lo cual a su vez determina las técnicas de reconocimiento a utilizar. Es por ello que, el reconocimiento enfocado en bioacústica sigue siendo hoy en día un área de investigación con gran potencial.

## **7. Bibliografía y Referencias**

- [1] P. Tubaro, "Bioacústica aplicada a la sistemática, conservación y manejo de poblaciones naturales de aves". *Etología*. Vol. 7. 1999. Pp.19-32.
- [2] M. Aide,"Real-time bioacoustics monitoring and automated species identification". *PeerJ*. 2013. Pp. 1-19.

- [3] R. Carbó, “Bioacustica de fondos marinos. Acústica aplicada a la pesca”. *Revista de Acústica*. Vol. 31. No. 3 y 4. Pp. 1-4.
- [4] L. Baptista, J. Martínez, “La investigación bioacústica de las aves del Archipiélago de Revillagigedo: un reporte de avance”. *Revista Mexicana de Ornitología*. Vol. 3. No. 2. 2002. Pp. 33-41.
- [5] A. Navarro et al., “Biodiversidad de Aves en México”. *Revista Mexicana de biodiversidad*. 2014. Pp. 476-495.
- [6] E. Villarreal, Identificación del canto de *Turdus migratorius* (Aves) utilizando un modelo acústico estadístico. Tesis licenciatura. Universidad Nacional Autónoma de México. Facultad de estudios superiores Zaragoza. 2014.
- [7] R. Rempel, “Bioacoustic monitoring of forest songbirds: interpreter variability and effects of configuration and digital processing methods in the laboratory”. *Journal of Field Ornithology*. Vol. 76. No. 1. 2005. Pp. 1-11.
- [8] D. Chesmore, “Automated bioacoustic identification of species”. *Annals of the Brazilian Academy of Sciences*. Vol. 76. No. 2. 2004. Pp. 435-440.
- [9] G. Moreno, “El comportamiento de las aves como herramienta para su identificación”. *Acta Granatense*. 2006. Pp. 85-93.
- [10] C. Duncan et al., “The quest for a mechanistic understanding of biodiversity–ecosystem services relationships”. *The royal society publishing*. 2015. Pp.1-10.
- [11] N. Toro, S. Giraldo, T. Salazar, “Reconocimiento de especies de anuros por sus cantos, en archivos de audio, mediante técnicas de procesamiento digital de señales”. *Scientia et Technica*. No. 32. 2006. Pp. 1-6.
- [12] P. Catalina, J. Ruiz, M. Orozco, “Reconocimiento automatizado de señales bioacústicas: Una revisión de métodos y aplicaciones”. *Ingeniería y Ciencia*. Vol. 9. No. 18. 2013. Pp. 171-195.
- [13] J. Pabico, “Automatic Identification of Animal Breeds and Species Using Bioacustics and Artificial Neural Networks”. *CoRR*. 2015.
- [14] A. Pedroza, J. de la Rosa, “El invisible y asombroso proceso de la comunicación oral: bases sobre reconocimiento de voz”. *Pistas Educativas*. No. 112. Noviembre 2015. Pp. 1310-1330.



- [15] A. Buzó, "Speech Coding Based Upon Vector Quantization". IEEE Transactions on Acoustics, Speech, and Signal Processing. Vol. assp-28. No. 5. 1980. Pp. 562-574.
- [16] G. Bendezú, Filtro adaptivo LMS y su aplicación en el reconocimiento de palabra aisladas para el control de equipo de sonido por medio de la voz. Tesis de licenciatura. Pontificia Universidad Católica del Perú: Facultad de Ciencias e Ingeniería. 2004.
- [17] M. Saerens, H. Boullard, "Linear and nonlinear prediction for speech recognition with Hidden Markov Models". En: 3rd European Conference on speech communication and technology. Berlin, Alemania. 21-26 de Septiembre de 1993. Eurospeech. 1993.
- [18] L. Rabiner, "A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition". Proceedings of the IEEE. Vol. 77. No. 2. Febrero de 1989. Pp. 257-286.
- [19] J. Pech, Desarrollo de un sistema de reconocimiento de voz para el control de dispositivos utilizando mixturas gaussianas. Tesis de Maestría. Instituto Politécnico Nacional: Centro de Investigación en Computación. 2006.
- [20] G. Hinton, "Deep Neural Networks for Acoustic Modeling in Speech Recognition". IEEE Signal Processing Magazine. 2012. Pp. 82-97.
- [21] M. Hagan, Neural Network Design. 2da. Edición, editorial Martin T. Hagan and Howard B. Demuth. 1012 páginas.
- [22] Las redes neuronales: qué son y por qué están volviendo. <http://goo.gl/GNFV>. Junio de 2014.
- [23] J. Schmidhuber, "Deep learning in neural networks: An overview". Neural Networks. Vol. 61. 2015. Pp. 85-117.
- [24] A. Castro, Construcción de una red neuronal artificial para clasificar cantos de aves: una aplicación de la inteligencia artificial a la biología. Tesis Maestría. Universidad de Costa Rica. 2006.
- [25] D. Chesmore, "Application of time domain signal coding and artificial neural networks to passive acoustical identification of animals". Applied acoustics. Vol. 62. 2001. Pp. 1359-1374.

## **8. Autores**

M.C. Ángel David Pedroza Ramírez obtuvo el grado de Maestro en Ciencias de la Ingeniería por la Universidad Autónoma de Zacatecas en 2015. Actualmente se encuentra cursando el segundo semestre del Doctorado en Ciencias de la Ingeniería con especialidad en Procesamiento de Señales y Mecatrónica en la misma institución. Su línea de investigación actual es en reconocimiento de voz así como identificación automática enfocada en bioacústica.

Ph.D. José Ismael de la Rosa Vargas obtuvo el grado de Doctor en Ciencias con especialidad en Procesamiento de Señales y Control (noviembre de 2002), por parte de la Universidad Paris Sud (XI) y de la Escuela Superior de Electricidad (SUPELEC) al sur de Paris (Gif-sur-Yvette), Francia. Trabaja actualmente en procesamiento de imágenes y voz, métodos estocásticos en problemas inversos e instrumentación.

D.C. Rogelio Rosas Valdez obtuvo el grado de doctor en Ciencias Biológicas en la Universidad Nacional Autónoma de México (Agosto 2008). Su línea de investigación actual es la Sistemática y Biogeografía de vertebrados de vida silvestre del estado de Zacatecas.