Reconocimiento de múltiples especies a partir del análisis automático del paisaje acústico

Maria José Guerrero Muriel

Tesis de maestría presentada para optar al título de Magíster en Ingeniería

Asesora
Claudia Victoria Isaza Narváez, PhD

Universidad de Antioquia
Facultad de Ingeniería
Maestría en Ingeniería
Medellín, Antioquia, Colombia
2022

Agradecimientos

Agradezco a mi asesora, la profesora Claudia Isaza por todo su acompañamiento, asesoría y paciencia durante el desarrollo de este trabajo, que fue un proceso lleno de aprendizajes tanto académicos, como personales. A los profesores José David López y Juan Manuel Daza por el acompañamiento, consejos y enseñanzas brindadas. Gracias a los estudiantes del semillero de inteligencia computacional del grupo SISTEMIC (Jonathan, Julián A, Julián S y Santiago) por su colaboración en las diferentes etapas de este trabajo de grado, y a Karen Jaramillo, por los aportes y análisis realizados desde el área de biología. Agradezco también al equipo de trabajo (estudiantes y profesores) del proyecto 852 por el apoyo y por hacer este proyecto ameno y disfrutable.

Gracias a mis papás, mi familia y mis amigos (Felipe, Laura, Sara y David) por el apoyo y la motivación brindada durante este proceso. Finalmente, gracias a la Universidad de Antioquia y al programa No. 111585269779 de Minciencias por la financiación proporcionada, con la cual fue posible realizar este trabajo de investigación.

Resumen

El monitoreo acústico pasivo se presenta como una alternativa para el análisis de las comunidades de vida silvestre y la evaluación de las condiciones del ecosistema. Los métodos automáticos de detección de especies apoyan el monitoreo y análisis de la biodiversidad al proporcionar información sobre la presencia-ausencia de especies, lo que permite entender la estructura del ecosistema. Por ello, se han propuesto diferentes alternativas para la identificación de especies. Sin embargo, los algoritmos que se encuentran en literatura están parametrizados para identificar especies específicas, no permite la identificación de múltiples especies de manera simultanea y se debe realizar un entrenamiento especifico para cada una de las especies de interés. El análisis de múltiples especies ayudaría a monitorear y cuantificar la biodiversidad, ya que incluye los diferentes grupos taxonómicos presentes en el paisaje sonoro. En este trabajo de investigación se presenta una metodología no supervisada para el reconocimiento de llamadas de múltiples especies de animales a partir del análisis automático de paisajes sonoros. La propuesta está basada en un algoritmo de clustering, específicamente el algoritmo LAMDA 3Π, que sugiere automáticamente el número de clústeres asociados a los sonotipos. Se hizo énfasis en el algoritmo de segmentación y en la selección de características del audio para analizar todo el paisaje sonoro sin parametrizar el algoritmo en función de cada grupo taxonómico. Para estimar el rendimiento de esta propuesta, se utilizaron cuatro conjuntos de datos correspondientes a diferentes lugares, años y habitats de Colombia. Estos conjuntos de datos contienen sonidos de los cuatro principales grupos taxonómicos que dominan los paisajes sonoros terrestres (aves, anfibios, mamíferos e insectos) en espectros audible y ultrasónico. La metodología presenta rendimientos entre el 75 % y el 96 % en el reconocimiento de la presencia-ausencia de las especies. Además, utilizando los clústeres propuestos por la metodología, se muestra que es posible cuantificar la biodiversidad presente en el paisaje sonoro. Para esto se comparó con la estimación de cuatro índices acústicos (ACI, NP, SO, BI). Este enfoque realiza evaluaciones de la biodiversidad de manera similar a la presentada en los índices acústicos, con la ventaja de proporcionar información específica de las especies sin necesidad de un conocimiento previo de las mismas en las grabaciones de audio.

Índice general

Índice de figuras

Índice de tablas

Capítulo 1

Introducción

Para la conservación de los ecosistemas es importante proponer planes de acción tendientes a cuidar la biodiversidad. El PAM (*passive acoustic monitoring*) se presenta como una alternativa costo eficiente que permite identificar los cambios en los ecosistemas recopilando la actividad presente en un paisaje sonoro [1]-[3], siendo captada por grabadoras de audio. El estudio de estos paisajes ha permitido obtener información sobre las poblaciones de un sitio, diferenciar especies que pueden ser morfológicamente similares, pero acústicamente diferentes; además de facilitar el monitoreo de especies crípticas (especies que son difíciles de distinguir morfológicamente).

Animales como las aves, anfibios, insectos, peces y algunos mamíferos usan señales acústicas como forma de comunicación y a través de estas pueden defender su territorio, reproducirse, localizar individuos de la misma especie y obtener recursos. El sonido registrado en una zona a través de grabaciones de audio es usado como herramienta para evaluar el comportamiento acústico de comunidades de animales. Dicha evaluación hace parte del PAM y permite detectar la presencia de las especies en un lugar y tiempo determinado. Además, este monitoreo brinda información sobre patrones de reproducción y permite identificar cambios en la riqueza y composición de las comunidades debido a impactos naturales y antrópicos [4]. Con el PAM es posible obtener datos de especies durante todo el día y noche, analizar largos periodos de tiempo y recopilar más información sobre el objeto de estudio que si se hiciera por observación directa (a través de salidas de campo) [3].

Existen herramientas que permiten procesar los audios para el análisis del paisaje acústico e identificación de llamados de especies como Raven (The Cornell Lab of Ornithology), Avisoft (Avisoft Bioacoustics), entre otras. Estas herramientas permiten analizar los audios a través de la visualización de los espectrogramas y partir de allí, realizar la anotación manual de la presencia de una especie especifica. Sin embargo, este tipo de detecciones (manuales) para grandes cantidades de datos, requieren tiempo y alta intervención por parte del experto por lo que es necesario contar con métodos de identificación automática que den soporte a los monitoreos a gran escala [5].

Recientemente se han desarrollado métodos para la identificación automática de especies a partir de sus vocalizaciones, incluyendo especies de aves [6]-[9], anuros [10]-[12], insectos [4] y algunas especies de mamíferos [13]-[15]. La mayoría de estos métodos de identificación automática están basados en modelos probabilísticos [16], técnicas de aprendizaje de máquina [17]-[19] y aprendizaje profundo [20]. En general, para la identificación de especies, se sigue un procedimiento común de cuatro pasos (i) preprocesamiento, (ii) segmentación, (iii) extracción de características y (iv) clasificación. En la mayoría de los modelos basados en aprendizaje profundo, varios de estos pasos se integran en un único flujo de trabajo, que incluye principalmente la extracción de características y la clasificación de vocalización de especies [20].

La mayoría de propuestas para identificación automática de especies usan técnicas de aprendizaje supervisado, requiriendo etiquetar datos para la etapa de entrenamiento de los modelos, lo cual puede ser una dificultad debido a que no en todos los ecosistemas se tiene información de las especies que están presentes [21]. Además, en estos trabajos se presentan modelos que realizan un análisis especie-especifico, es decir, limitan la búsqueda a bandas de frecuencia en las que una especie objetivo vocaliza. Si se busca monitorear la biodiversidad de un lugar y los patrones de comportamiento acústico de las especies que habitan allí, es preferible un método capaz de detectar varias especies de animales simultáneamente, además de tener la posibilidad de hacer detecciones de especies que no se esperaban previamente.

Por otro lado, para evaluar la biodiversidad desde la perspectiva acústica de una zona, se cuenta con los índices acústicos. Estos índices evitan la necesidad de entrenar modelos específicos para cada especie, generando una evaluación rápida de la biodiversidad acústica basada en la distribución de energía acústica de las especies [22], [23]. Sin embargo, estos índices brindan información acerca la actividad general del paisaje y no es posible conocer el comportamiento vocal de las especies que conforman ese paisaje.

En esta investigación, se propone una metodología no supervisada para la identificación de vocalizaciones de especies de animales y su aplicación para el análisis de la biodiversidad. Esta metodología es capaz de analizar múltiples bandas de frecuencia de una grabación, extraer los segmentos que por su intensidad en el espectrograma pueden ser vocalizaciones de animales, agrupar los segmentos de acuerdo a su similitud e identificar los sonotipos (patrones sonoros que podrían asociarse a vocalizaciones de especies de diferentes grupos taxonómicos) utilizando un enfoque no supervisado. Esta propuesta encuentra automáticamente los sonotipos que responden a la variabilidad acústica intraespecífica y permite la diferenciación entre grupos taxonómicos. Este método se constituye en una herramienta de apoyo para los expertos en bioacústica quienes asociaran los sonotipos con las llamadas de una especie y luego esta asociación servirá para identificar las vocalizaciones en nuevos registros acústicos.

El número de sonotipos propuestos por el algoritmo responde a la diversidad de especies acústicas del lugar analizado. Para ejemplificar la aplicación de este método en la estimación de la biodiversidad, se compararon los resultados obtenidos con cuatro índices acústicos comúnmente utilizados: índice de complejidad acústica (ACI), índice bioacústico (BI), número de picos (NP) y la ocupación espectral (SO). Luego, se evidencia que esta propuesta tiene la ventaja de identificar la estructura acústica de la comunidad como un indicador de biodiversidad. Además, se muestra que este método puede generalizarse a conjunto de datos independientes, espectros sónicos y ultrasónicos sin necesidad de ajustar parámetros del algoritmo ni hacer entrenamiento especifico por especie.

Esta metodología es la base de una herramienta computacional la cual permite hacer el análisis de los patrones de comportamiento acústico de especies que vocalicen entre 100 Hz y 70 KHz. Esta tesis aporta al monitoreo acústico pasivo a través de la construcción de herramientas que permitan obtener información sobre el cambio de los ecosistemas y el uso del espacio acústico para el entendimiento de las dinámicas de las comunidades y su futura conservación.

El contenido de este documento se presenta de la siguiente manera: el capítulo 2 describe las bases de datos usadas para el desarrollo de este trabajo, los casos de estudio analizados, la métrica usada para evaluar el desempeño y la definición de los cuatro índices acústicos usados para la comparación en la aplicación del método. Luego, el capítulo 3 presenta la metodología propuesta para la identificación de especies de animales. En el capítulo 4 se describen las dos primeras etapas de la metodología propuesta correspondiente a el pre-procesamiento y la segmentación de los audios, continuando con el capítulo 5 donde se presenta el análisis realizado para la etapa de extracción de características y el capítulo 6, donde se describe el algoritmo de agrupamiento propuesto. Luego, en el capítulo 7 se presentan los resultados obtenidos al evaluar la metodología de identificación de especies en tres casos de estudio. Finalmente, en el capítulo 8 se presentan las conclusiones derivadas de este trabajo de investigación.

1.1. Objetivos

Objetivo General

Proponer una metodología no supervisada para el reconocimiento automático de múltiples especies a partir de patrones de sonido encontrados en grabaciones de audio.

Objetivos Específicos

- Identificar algoritmos de reducción de ruido, segmentación y extracción de características que permitan detectar diferencias entre sonotipos y posteriormente, entre especies.

- Analizar técnicas de aprendizaje no supervisado para identificar cual es el modelo que más se ajusta a la clasificación de las especies.
- Diseñar una metodología para identificar patrones de actividad acústica de múltiples especies.
- Validar la metodología propuesta a través de pruebas con un número alto de grabaciones de bosque seco tropical y bosque húmedo colombiano.

1.2. Contribución del trabajo de investigación

- Método para identificar vocalizaciones de múltiples especies entre 100 Hz hasta 70 kHz de manera simultanea sin necesidad de parametrizar el algoritmo, el cual fue evaluado en zonas con alta biodiversidad donde no se conocen todas las especies presentes a priori.
- Este método da una medida de la riqueza de especies que vocalizan en un punto geográfico, de manera similar a lo que proponen los índices acústicos usados para análisis de biodiversidad. Al usar este método, se asocian sonotipos identificados automáticamente a especies presentes y, a través del sonido se puede descomponer el aporte de cada especie y conocer la estructura acústica del lugar, lo cual no es posible conocer con los índices acústicos ya que estos son una medida de complejidad acústica que puede estar relacionada con la riqueza de especies del sitio.

Productos derivados de este trabajo de investigación

- Guerrero, M. J., Bedoya, C. L., López, J. D., Daza, J. M., Isaza, C. (2023). Acoustic animal identification using unsupervised learning. Methods in Ecology and Evolution, 00, 1–15. `https://doi.org/10.1111/2041210X.14103`
- Guerrero, M. J., Restrepo, J., Nieto-Mora, D.A., Daza, J.M., Isaza, C. (2022). Insights from Deep Learning in Feature Extraction for Non-supervised Multispecies Identification in Soundscapes. In: Advances in Artificial Intelligence – IBERAMIA 2022. Lecture Notes in Computer Science, vol 13788. Springer, Cham. `https://doi.org/10.1007/978-3-031-22419-5_19`

Capítulo 2

Bases de datos y casos de estudio

2.1. Bases de datos

Cuatro bases de datos correspondientes a regiones tropicales proporcionadas por el Grupo Herpetológico de Antioquia, fueron usadas para este trabajo con el fin de evaluar el rendimiento del algoritmo en diferentes temporadas, años y tipos de hábitat.

Las bases de datos A y B fueron adquiridas en la zona protegida de la central hidroeléctrica de Jaguas (06°26′ N, 075°05′ W; 06°21′ N, 074°59′ W) ubicada en la vertiente oriental de la Cordillera Central del norte de Antioquia, Colombia (ver Figura 2.1). El área protegida abarca 50 Km^2, incluyendo el embalse de San Lorenzo con 10.2 Km^2 y cubre un gradiente de elevación de 850 a 1.300 m.s.n.m. La cobertura vegetal dentro del área protegida está dominada por diferentes estados de sucesión de bosque secundario (70 %), seguida por mosaico de tierras de cultivo/vegetación natural (23 %), superficies no naturales o degradadas (5 %) y pastizales (2 %). El área protegida de Jaguas mantiene comunidades de vertebrados terrestres, incluyendo especies amenazadas y endémicas, se considera primordial para la conservación de la biodiversidad a escala regional [24], [25].

El conjunto de datos A consta de 50 grabaciones de audio tomadas con un dispositivo Song Meter SM2 (Wildlife Acoustics, Inc.) durante los meses de noviembre de 2012 y febrero de 2013. Estos datos fueron usados para evaluar la propuesta de reconocimiento de especies de diferentes grupos taxonómicos en un sitio de gran biodiversidad biológica. Cada grabación tenía una duración de un minuto, adquirida cada diez minutos, con una tasa de muestreo de 44,1 kHz utilizando un canal con una resolución de 16 bits. Las especies presentes en el paisaje sonoro fueron etiquetadas manualmente por tres expertos en bioacústica, escuchando los audios y revisando los espectrogramas. El conjunto de datos contiene llamadas de 41 especies diferentes entre ellas aves, anuros e insectos.

El conjunto de datos B se utilizó para comparar el rendimiento de esta propuesta con otras librerías y software disponibles como Autodetec del paquete WarbleR [26],

MonitoR [27] y Kaleidoscope Pro [28]. El conjunto de datos consta de 1400 grabaciones de un minuto de duración obtenidas cada quince minutos, durante los meses de mayo a junio de 2017, mediante un dispositivo Song Meter SM4 (Wildlife Acoustics, Inc.). Las señales se adquirieron con una tasa de muestreo de 24 kHz a una resolución de 16 bits. En esta base de datos se cuenta con la información de cuatro especies de anuros y una especie de ave que fueron etiquetados manualmente por tres expertos en la fauna de la zona utilizando Raven (The Cornell Lab of Ornithology) y Sonic Visualiser (Queen Mary University of London).

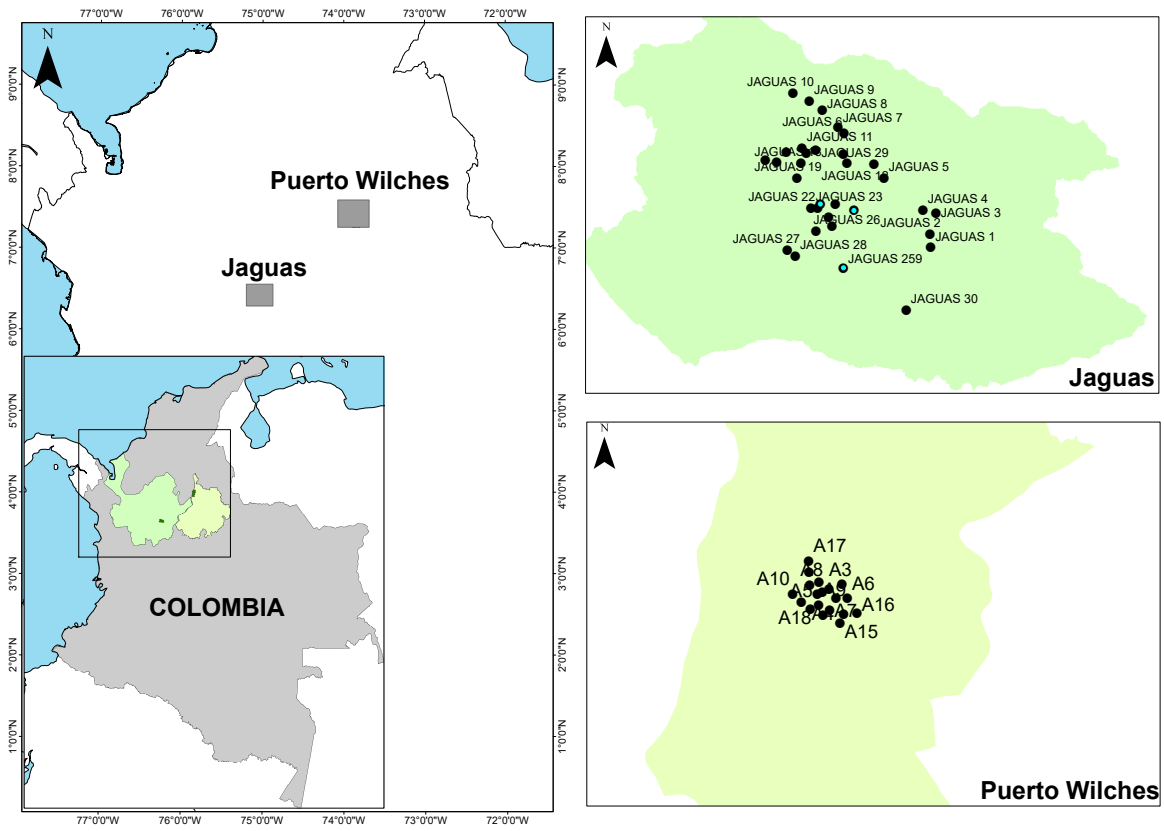

Figura 2.1: Mapa con ubicaciones geográficas de los set de datos A,B,C,D. El mapa de **Jaguas** corresponde a las ubicaciones (cada punto) de las bases de datos A (resaltado en azul) y B. El mapa de **Puerto Wilches** corresponde a las ubicacciones (cada punto) de las bases de datos C y D.

Las bases de datos C y D fueron adquiridas en una zona rural del municipio de Puerto Wilches, Santander, Colombia (7°21′52,5”N, 73°51′33,0”W) (ver Figura 2.1). La zona caracterizada corresponde a un área circular delimitada por un radio de 1500 m. En el sitio dominan los cultivos de palma de aceite de diferentes edades (75 %), vegetación secundaria (7,6 %), bosque (6,13 %), pastizales (5,5 %) y vegetación acuática (3,2 %). En la zona hay algunas edificaciones y una red de caminos secundarios vinculados a la

producción de palma y ganado.

El conjunto de datos C se recolectó usando un dispositivo Song Meter Mini (Wildlife Acoustics, Inc.) programado para grabar durante un minuto cada diez minutos con una tasa de muestreo de 48 kHz. Este conjunto consta de 2517 grabaciones de audio obtenidas entre marzo y junio de 2021. Este conjunto de datos se usó para evaluar la identificación de la presencia de especies utilizando la metodología no supervisada en diferentes condiciones y comparar nuestro enfoque con los índices acústicos asociados al análisis de biodiversidad. En este conjunto de datos se realizó un trabajo estricto de etiquetado por expertos en bioacústica donde 11 especies de interés, incluyendo aves, anuros y un primate fueron seleccionados y etiquetados.

El conjunto de datos D se utilizó para probar la capacidad del método propuesto para detectar especies en el espectro ultrasónico (20 kHz - 70 kHz). Este set de datos consta de 197 grabaciones de audio obtenidas entre marzo y junio de 2021 y fueron adquiridos utilizando un dispositivo Song Meter Mini bat (Wildlife Acoustics, Inc) programado para grabar quince segundos cada quince minutos con una frecuencia de muestreo de 384 kHz. En este conjunto de datos se encontraron 13 especies de murciélagos y 6 ortópteros.

2.2. Casos de estudio

Usando las bases de datos descritas anteriormente, se evaluó el método en cuatro casos de estudio diferentes con el fin de demostrar la capacidad de identificación de especies de animales y analizar los patrones acústicos que representen cada lugar.

- Caso 1 (conjunto de datos A): este caso de estudio prueba la calidad del enfoque de agrupación para el reconocimiento de llamadas de múltiples especies en un sitio altamente biodiverso (41 especies con vocalizaciones/estridulaciones entre 900 Hz y 15 kHz asociadas a especies de aves, anuros e insectos).

- Caso 2 (conjunto de datos B): compara el rendimiento de la metodología propuesta con otras tres metodologías disponibles para la identificación de especies. Dos de ellas son librerías de R que no realizan agrupamiento pero son ampliamente utilizadas por expertos en bioacústica.

- Caso 3 (conjunto de datos C): en este caso de estudio, probamos las capacidades de la propuesta multi-especie para el reconocimiento de llamadas de animales en diferentes condiciones (diferentes especies, sitios, años y grabadoras). Esto se conoce como *method-based validation* [29]. Este tipo de validación externa pone a prueba la estabilidad de la partición del clúster y se centra en las similitudes estructurales de los resultados de los clústeres generados por el método propuesto [29] por ende, permite verificar que los resultados obtenidos no son un efecto del

conjunto de datos inicial. Además, como este conjunto de datos proporciona información sobre las especies y su ubicación especifica, se analizaron las similitudes con los índices acústicos para generar una medición de la biodiversidad y ver el aporte de las especies a la riqueza de un sitio.

- Caso 4 (conjunto de datos 4): se evaluó las capacidades de la metodología para detectar especies en el espectro ultrasónico como murciélagos y ortópteros.

Las bases de datos usadas en los cuatro casos de estudio contienen especies vocalizando en diferentes frecuencias, localizadas en diferentes tipos de hábitat y fueron grabadas usando diferentes grabadoras.

2.3. Métricas de desempeño

El desempeño de cada caso de estudio fue evaluado de acuerdo a la detección presencia de cada especie en una grabación, usando como métricas la sensibilidad (Ec.(2.1)) y la especificidad (Ec.(2.2)):

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \tag{2.1}$$

$$\text{Especificidad} = \frac{VN}{VN + FP} \tag{2.2}$$

Para el caso de estudio tres, se usaron cuatro índices acústicos relacionados con la estimación de la riqueza de especies en paisajes sonoros para comparar con la aplicación de este método para estimar la biodiversidad acústica de un lugar: índice de Complejidad Acústica (ACI) [30], índice Bioacústico (BI) [31], Número de Picos (NP) [32], y Ocupación Espectral (SO) [33], [34]. Se eligieron estos cuatro índices porque miden las contribuciones de los elementos bióticos al espectro acústico. Se comparó la tendencia de estos índices con los resultados generados por la metodología propuesta, sugiriendo automáticamente unos sonotipos cuya cantidad representa la riqueza de un sitio.

El índice ACI cuantifica las variaciones espectrales en el espectro acústico mediante la penalización de valores de energía similares en los intervalos de frecuencia adyacentes. Cuanto mas heterogéneo sea el paisaje sonoro, mayor será el valor del ACI. Se utilizó el índice ACI_{ft}, el cual es el índice de complejidad acústica calculado a lo largo de las frecuencias [30]. BI cuantifica la energía acústica entre 2-8 kHz, la cual suelen ser las bandas de frecuencias con mayor contenido de señales biofónicas [31]. NP mide el número de elementos que contribuyen al espectro acústico, es decir, el número de picos en la densidad espectral de potencia [32]. De manera similar, SO mide el porcentaje del espectro acústico que se está utilizando. Se hace sumando los anchos de banda de las bandas de frecuencias ocupadas y luego dividiendo entre el espectro acústico total disponible. [33], [34].

Capítulo 3

Metodología para identificación de vocalizaciones de múltiples especies

El método propuesto (figura 3.1) agrupa los segmentos identificados como posibles vocalizaciones en función de sus similitudes acústicas y propone sonotipos que pueden asociarse a los llamados de especies de animales presentes en un paisaje sonoro.

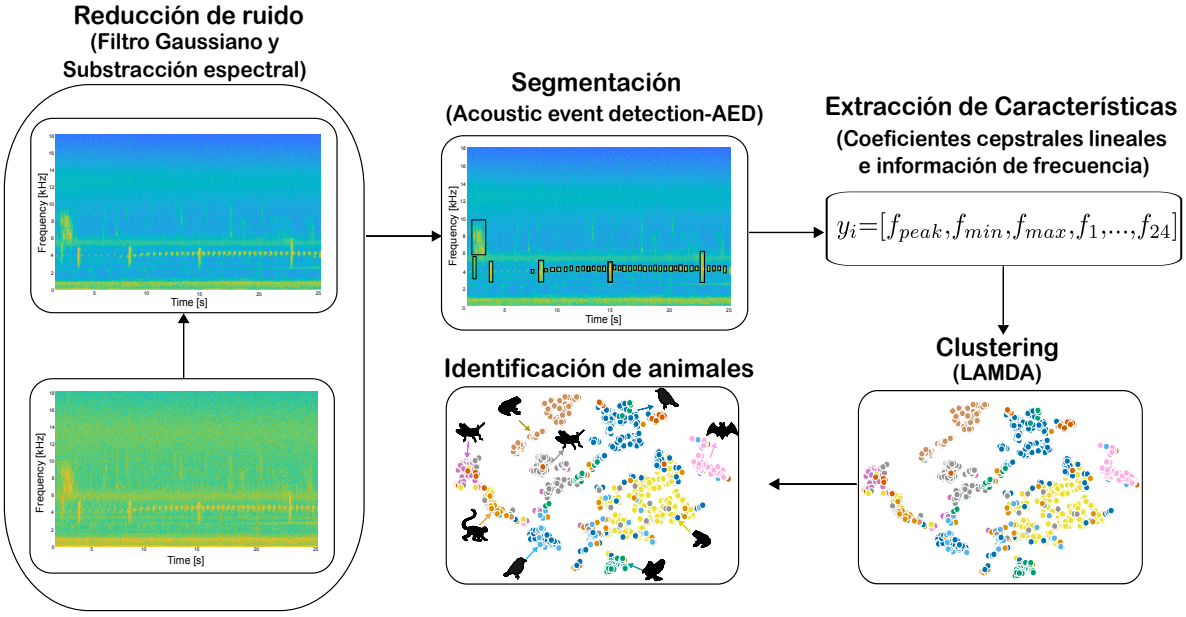

Figura 3.1: Metodología propuesta para la agrupación de sonidos de animales en paisajes sonoros. Cada grabación pasa por una etapa de reducción de ruido y segmentación. Luego, se extrae la información de frecuencia y los coeficientes cepstrales escalados linealmente y se agrupan usando el algoritmo de clustering LAMDA 3Π. Los segmentos analizados se representan como un punto para el clustering y cada color corresponde a un clúster. Cada clúster tiene un patrón (sonotipo) que se asociará con las llamadas de las diferentes especies por los expertos.

La primera etapa (pre-procesamiento) reduce el ruido de fondo y destaca la actividad acústica asociada a biofonía. Esto facilita la segunda etapa (segmentación), en la cual se extraen los segmentos usando métodos de procesamiento digital de imágenes como la umbralización de Otsu y operaciones morfológicas. La tercera etapa (extracción de características) estima las frecuencias pico, mínima, máxima, así como los coeficientes cepstrales de escala lineal, que se utilizan como entrada para la ultima etapa (agrupación). En esta etapa, un algoritmo de agrupación no iterativo analiza las características extraídas anteriormente y agrupa los sonidos del paisaje en función de sus similitudes acústicas. Por último, los grupos, de acuerdo al patrón de llamado, son asignados a especies de animales.

Cada etapa se presenta como un capítulo, esto debido a que se analizó el estado del arte respectivo y se realizaron pruebas especificas que permitieron tomar las decisiones para cada algoritmo. Los capítulos 4 y 5 presentan los resultados obtenidos que llevaron a concluir sobre los algoritmos necesarios en ambas etapas. En el capítulo 6 se presenta el algoritmo de agrupamiento usado y finalmente en el capítulo 7 se presentan los resultados de la implementación de la metodología para identificación de vocalizaciones de múltiples especies completa.

Capítulo 4

Pre-procesamiento y segmentación de los audios

4.1. Pre-procesamiento

En esta etapa se realiza la estimación de una representación de tiempo-frecuencia (como el espectrograma) y la eliminación del ruido de la señal [20] buscando resaltar los llamados de las especies de interés. El ruido puede considerarse un sonido de interés dependiendo del objetivo de la investigación, como la lluvia, por ejemplo [35], sonidos asociados a antropofonía o incluso biofonía asociada a especies que no son de interés para el problema. En aplicaciones de bioacústica, los métodos de reducción de ruido propuestos son dependientes de la especie animal que se desee analizar. Xie et al. [36] presenta diferentes alternativas como la implementación de filtros FIR para vocalizaciones presentes en frecuencias altas (filtros pasa-altas) y en frecuencias bajas (pasa-bajas). Para casos más complejos como el análisis de múltiples especies, donde podría existir superposición de las señales, propone el uso de métodos de eliminación de ruido como la sustracción espectral, reducción de ruido basado en procesamiento de imágenes y reducción de ruido basado en técnicas de aprendizaje profundo o una combinación de estas si el problema lo requiere. Sin embargo no se encuentran propuestas con la implementación para múltiples especies.

Para el enfoque multi-especies que se propone en este trabajo, primero se genera la representación tiempo-frecuencia de la señal del audio aplicando la transformada de Fourier de tiempo corto (STFT). El espectrograma $\mathbf{S} \in \mathbb{R}^{N_\mathrm{s} \times N_\mathrm{t}}$ es calculado usando una ventana tipo Hamming de tamaño $N_\mathrm{ws} \in \mathbb{R}$ y con un solapamiento de 0.5. Donde $N_\mathrm{s} = N_\mathrm{ws}/2 + 1$ y $N_\mathrm{t} = 2N_\mathrm{x}/N_\mathrm{ws} - 1$ representan los dominios de frecuencia y tiempo respectivamente y $N_\mathrm{x} \in \mathbb{R}$ representa el número de muestras de la señal de audio. A continuación, se crea el espectrograma $\mathbf{S}' \in \mathbb{R}^{N_\mathrm{s} \times N_\mathrm{t}}$ cuyo ruido de fondo ha sido reducido aplicando el método de reducción de ruido propuesto por Xie et al. en [37] en el que la imagen de los espectrogramas se convolucionan usando un kernel gaussiano para eliminar la granulidad de la señal. Para este caso, el tamaño del kernel gaussiano se fijó

en 3×3. Adicional a esto, se implementó la técnica de sustracción espectral propuesta también por Xie et al. [37] con la finalidad de eliminar el ruido que no fue posible reducir con el filtro gaussiano. Para esta propuesta, solo se consideró como ruido de fondo los sonidos generados por geofonías como el viento y la lluvia. Este tipo de sonido tiene componentes espectrales en todas las bandas de frecuencia [35].

4.2. Segmentación

Esta etapa consiste en la detección y aislamiento de las posibles vocalizaciones presentes en el audio. Las vocalizaciones corresponden a segmentos importantes de la señal que pueden representar a las diferentes especies de la comunidad acústica a monitorear [38]. Se han encontrado propuestas donde realizan la segmentación seleccionando manualmente en el espectrograma el rango de frecuencias y la ubicación temporal en la que un individuo emite el sonido [15], [39]; otra alternativa consiste en crear plantillas con ejemplos de las vocalizaciones de las especies de interés [26], [27], [40]. Sin embargo, se considera importante automatizar este proceso debido al tiempo que puede tomar analizar grandes cantidades de datos, además, de la precisión que debe tener el investigador al momento de realizar la selección manual de la vocalización o la plantilla.

Dentro de las propuestas de segmentadores automáticos se encuentran segmentadores basados en el análisis de la energía de la señal como se presenta en [5], [19], [41] y propuestas utilizando métodos de segmentación automática basados en técnicas de análisis de imágenes como las presentadas por [18], [42]. El desarrollo de esta etapa, al igual que en la etapa de pre-procesamiento, depende de la especie de interés, no se ha estudiado ninguna técnica especifica para la detección simultanea de especies pertenecientes a diferentes grupos taxonómicos. Una segmentación automática adecuada potencializa las etapas posteriores (extracción de características y *clustering*) y permite hacer la asociación de las diferentes vocalizaciones a las especies.

4.2.1. Segmentador implementado

Para abordar esta tarea, dado que no hay una metodología que segmente de manera automática y simultanea diferentes especies, se propone una modificación del algoritmo de detección de eventos acústicos (AED, por sus siglas en inglés) propuesto por Xie et al. [37], de manera que pueda trabajar en todas las bandas de frecuencia.

El enfoque de segmentación propuesto utiliza el umbral de Otsu [43] y operaciones morfológicas como *opening* y *closing* para aislar los posibles eventos acústicos. El *opening* (*erosion/dilation*) se realizó utilizando un kernel rectangular de 9×7 para eliminar los objetos pequeños en el espectrograma $\mathbf{S}'$, seguido del *closing* (*dilation/erosion*) con un kernel cuadrado de 6×6 para rellenar los huecos pequeños. A continuación, se aplica una función que permite medir las propiedades de las regiones resaltadas (*regionprops*)

y a partir de esto, poder determinar los segmentos usando el área de la región resaltada, la extensión y la excentricidad. Estos segmentos son representados con recuadros delimitadores (*bounding boxes*). Finalmente, la información proveniente de los recuadros delimitadores es almacenada en una matriz $\mathbf{G} \in \mathbb{R}^{N_\mathrm{b} \times 4}$ donde N_b es el número de segmentos y se utilizan como delimitadores espectrales y temporales en el espectrograma $\mathbf{S}'$ durante el proceso de extracción de características.

El método original propuesto por [37] fue desarrollado para la identificación automática de especies de ranas cuyas vocalizaciones tienen una frecuencia máxima de 6 kHz. Por ende, todos los segmentos que se encuentren en frecuencias superiores de la frecuencia máxima de interés, son considerados como ruido y se eliminan. Para este trabajo, no se descartan, ya que es de interés el análisis de múltiples grupos taxonómicos y para ello es necesario conservar los segmentos en todas las bandas de frecuencia.

En la figura 4.1 se muestra el proceso de segmentación del audio descrito anteriormente partiendo del espectrograma con reducción de ruido $\mathbf{S}'$

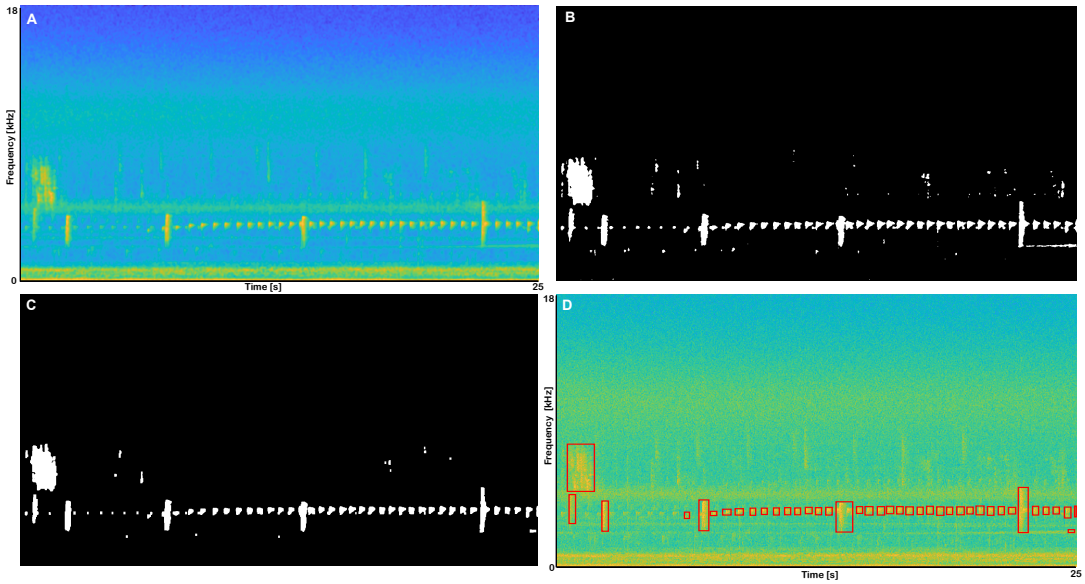

Figura 4.1: Propuesta para la detección de eventos acústicos (AED modificado) partiendo del espectrograma $\mathbf{S}'$ (**A**). **B** presenta la binarización del espectrograma usando el umbral de Otsu, seguido por las operaciones morfológicas (*opening-closing*) en **C**. Finalmente en **D** se presenta el espectrograma con los segmentos obtenidos que son representados con recuadros delimitadores de color rojo.

4.2.2. Selección del segmentador y pruebas

Se analizaron diferentes metodologías propuestas en trabajos anteriores para abordar el problema de segmentación de vocalizaciones en múltiples bandas de frecuencia de

manera simultanea. Se analizó la propuesta de [5] debido a que su enfoque trabaja con la detección de regiones interés realizando el análisis de todas las bandas de frecuencia y la propuesta de [37] cuya metodología se basa en la detección de eventos acústicos usando técnicas de análisis de imagen, técnica que según [36] podría funcionar para un enfoque multi-especie. Estas propuestas fueron implementadas y comparadas con el segmentador propuesto por [41] cuya metodología tiene una estructura similar a la de este trabajo y usa una técnica de segmentación basada en la energía de la señal de audio. Se evaluó el desempeño en la segmentación de las vocalizaciones de las especies del conjunto de datos A (los datos se describen en la sección 2) con cada una de estas propuestas y los resultados presentados en la tabla 4.1. Como se mencionó anteriormente, la propuesta de Xie et al. [37] fue modificada con el propósito de obtener los segmentos de todas las bandas de frecuencia y por ello se presenta como AED modificado.

Desempeño segmentación de vocalizaciones			
	Bedoya et. al(2014)	Ulloa et. al(2018)	AED Modificado
Mediana	0.50	0.63	1.00
Promedio	0.65	0.63	0.88
Desviación estándar	0.29	0.21	0.19

Tabla 4.1: Resultados de desempeño de las diferentes metodologías aplicadas para la etapa de segmentación.

El enfoque de segmentación propuesto (AED modificado) superó a las demás propuestas tanto en precisión como en la calidad de la segmentación permitiendo encontrar especies de manera simultanea en todas las bandas de frecuencia. La figura 4.2 muestra la comparación en la etapa de segmentación del algoritmo basado en energía de la señal propuesto por [41] y el algoritmo propuesto en este trabajo basado en análisis de imagen. Las figuras 4.2A y 4.2B muestran los resultados en segmentación de los dos métodos en el espectro audible (100 Hz - 20 kHz) mientras que las figuras 4.2C y 4.2D muestran el desempeño de los segmentadores para el espectro ultrasónico (20 kHz - 80 kHz). Los recuadros azules corresponden a los segmentos esperados y los rojos a aquellos que no están asociados a posibles vocalizaciones de especies de animales, estos segmentos se espera que en la etapa de clustering se agrupen en clusters no asociados a especies. En esta comparación es posible observar que con el método AED modificado se pueden obtener mayor número de segmentos acústicos de manera simultanea en todo el paisaje acústico, además de permitir la detección de especies en el espectro ultrasónico.

4.2.3. Conclusiones

En este capítulo se presentaron los métodos propuestos para la etapa de preprocesamiento y segmentación de la señal de audio. La identificación de múltiples especies es una tarea compleja que depende del rendimiento adecuado de todas las etapas. Las

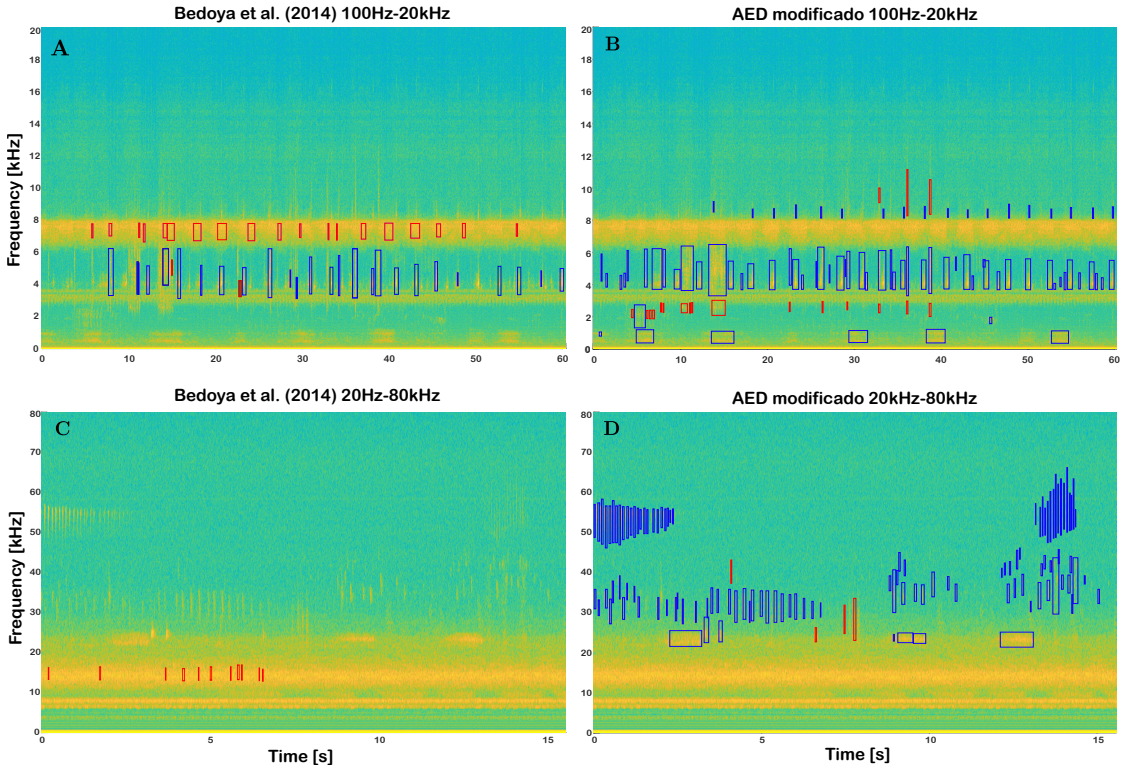

Figura 4.2: Representación en tiempo-frecuencia (espectrograma) de un audio. Cada rectángulo azul representa una región de interés, mientras que los rectángulos rojos representan segmentos asociados a ruido de fondo. **A - C** corresponde a la segmentación del espectrograma utilizando umbrales de energía en las bandas de frecuencia 100Hz-20kHz (A) y 20kHz-70kHz (C). **B - D** corresponde a la segmentación del espectrograma utilizando el enfoque propuesto en este trabajo, que a su vez, permite la detección de especies en altas frecuencias, incluidos murciélagos e insectos.

etapas de pre-procesamiento y segmentación de la señal son fundamentales cuando se van a analizar múltiples bandas de frecuencia de manera simultanea. En esta etapa se realizó un análisis de tres diferentes métodos propuestos en trabajos anteriores encontrando que la modificación realizada al método propuesto por [37], fue la indicada para la segmentación de los eventos acústicos permitiendo la detección en múltiples bandas de frecuencia, incluyendo el espectro ultrasónico, de manera simultanea. Esta etapa se considera fundamental pues la correcta segmentación de la señal de audio implicará una correcta identificación del patrón de vocalización de las especies. Los métodos mostrados en esta sección y elegidos con mejor desempeño fueron implementado en la metodología propuesta, mostrada en el capítulo 3

Con este capítulo se aporta en el cumplimiento del objetivo específico 1: "Identificar algoritmos de reducción de ruido, segmentación y extracción de características que permitan detectar diferencias entre sonotipos y posteriormente, especies"

Capítulo 5

Extraccion de Características

La extracción de características es una etapa donde se capturan las propiedades más informativas y significativas de los segmentos para un mejor entendimiento de la señal. Entre las características mas usadas para la identificación de especies se encuentran los Coeficientes Cepstrales de Mel (MFCC) [41], [44], Coeficientes Cepstrales escalados Linealmente (LFCC) [45], *Gammatone Cepstral Coefficients* (GTCC) [46], características basadas en la energía de la señal, cruce por cero y características espectrales y temporales de la señal como la frecuencia máxima, frecuencia mínima, frecuencia pico y duración promedio del canto [47]. Recientemente, se han utilizado métodos de aprendizaje profundo como Redes Neuronales Convolucionales (CNN) y Autoencodificadores Variacionales (VAE) para extraer características discriminativas en diferentes aplicaciones de audio. Estas características son la entrada de un algoritmo de agrupación que ha permitido diferenciar entre especies de aves [48], [49], especies de ranas [50] y entre individuos de una misma especie de ave [6]. Sin embargo, la aplicación de estas técnicas en bioacústica, concretamente en la detección de múltiples especies mediante aprendizaje no supervisado, está poco estudiada.

En este trabajo se realizó el análisis de diferentes métodos de extracción de características para la identificación automática de múltiples especies que implica el análisis de todas las bandas de frecuencia presentes en el paisaje sonoro (100 Hz - 70 kHz). Se analizaron las características que se extraen habitualmente en bioacústica: coeficientes cepstrales y características de frecuencia como la frecuencia de pico, frecuencia mínima y frecuencia máxima. Desde el área de aprendizaje profundo, se trabajó con Autoencoders Variacionales (VAE), que es un método no supervisado y recientemente usado en bioacústica [48], además de explorar con un enfoque supervisado utilizando una arquitectura de red neuronal convolucional (CNN) pre-entrenada VGG19 [6] como extractor de características. Estas características son la entrada de un algoritmo de clustering que agrupa las especies según sus similitudes.

5.1. Coeficientes Cepstrales escalados linealmente

Los Coeficientes Cepstrales de Frecuencia Mel se utilizan ampliamente como características discriminatorias en el reconocimiento del habla humana. Redistribuyen la frecuencia a lo largo del espectro de forma logarítmica, en la escala Mel, con el fin de beneficiar bandas de frecuencia específicas en las que trabaja el aparato vocal humano [37]. En algunos casos, el rango espectral de las especies estudiadas se encuentra en el espectro del habla humana (30Hz-3kHz) y el reconocimiento de las especies puede realizarse con éxito; sin embargo, generar una escala tipo Mel para identificar las vocalizaciones de diferentes taxones no sería una solución ya que no se busca beneficiar bandas de frecuencia especificas [42], [45]. Es por esto que se utilizan los Coeficientes cepstrales escalados linealmente (LFCC), cuyo proceso se presenta a continuación:

A partir de los segmentos obtenidos previamente en la etapa de segmentación, se calcula el vector de características para cada uno de ellos. El primer paso del proceso consiste en estimar el logaritmo de la energía del segmento $\mathbf{H} \in \mathbb{R}^{N_\mathrm{u} \times N_\mathrm{h}}$ (Ec.(5.1)), donde $N_\mathrm{u} \in \mathbb{R}$ y $N_\mathrm{h} \in \mathbb{R}$ y corresponden a la longitud del segmento en los dominios espectral y temporal respectivamente. Esta operación extrae la información acústica relevante del dominio temporal del segmento y la redistribuye por el dominio espectral de forma no lineal. El paso se realiza utilizando ventanas (*frames*) de tamaño w.

$$\mathbf{Q}_{i,m} = \log\left(\sum_{j=mw}^{m(w+1)} |\mathbf{H}_{i,j}|^2\right), \forall m = 1, \ldots, N_h/w \wedge \forall i = 1, \ldots, N_u \tag{5.1}$$

Donde $\mathbf{Q} \in \mathbb{R}^{N_\mathrm{u} \times N_\mathrm{m}}$ es una matriz con el logaritmo de las energías calculadas a partir de $\mathbf{H}$, w es el tamaño de la ventana movil (rectangular, sin solapamiento), m es el *frame* actual y $N_m = N_h/w$ es el número de logaritmos calculados en cada ventana.

Luego, se calcula la transformada de coseno discreto (DCT) de $\mathbf{Q}$ (Ec.(5.2)). El objetivo de este paso es reducir la dimensionalidad de $\mathbf{Q}$ y establecer una longitud común para el vector de características extraído para todas las vocalizaciones de animales, independientemente su duración o ancho de banda. Además, la DCT permite reconstruir una señal compleja con precisión a partir de unos pocos coeficientes.

$$\begin{aligned}\mathbf{Y}_{k,m} = N_\mathrm{p} \sum_{i=1}^{N_\mathrm{u}} \mathbf{Q}_{i,m} \cos\left(\frac{\pi}{N_\mathrm{u}}(i-1)(k-0{,}5)\right), \\ \forall k = 1, \ldots, N_k \wedge \forall m = 1, \ldots, N_\mathrm{m}\end{aligned} \tag{5.2}$$

Donde $\mathbf{Y}_{k,m}$ es una matriz que contiene los coeficientes DCT, k es el índice de la banda de frecuencias, N_k es el número de coeficientes y N_p es un factor de normalización utilizado para que la matriz transformada sea ortogonal. $N_\mathrm{p} = \sqrt{1/N_\mathrm{u}}$ para $k = 1$ y $N_\mathrm{p} = \sqrt{2/N_\mathrm{u}}$ para $2 \leq k \leq N_\mathrm{u}$.

Finalmente, concatenando 24 coeficientes con 3 características espectrales (frecuencia pico, frecuencia mínima y frecuencia máxima) se obtiene el vector de características $\mathbf{y_a} \in \mathbb{R}^{N_\mathrm{f}}$ para cada segmento $\mathbf{H}$. Cada vector es representado como un hiperpunto de dimensión 27 (N_f), que luego en la etapa de clustering se agruparan de acuerdo a su similitud.

5.2. Autoencoder Variacional

Los autocodificadores variacionales (VAE) son un método de aprendizaje profundo no supervisado para la extracción de características. Este método realiza una etapa de codificación en la se que se entrena una arquitectura de red neuronal y se extraen las características de la ultima capa oculta (previa a la capa de salida). Estas características se representan en un espacio latente de dimensionalidad reducida, manteniendo suficiente información para reconstruir las entradas originales [51]. La reconstrucción de las entradas se realiza haciendo pasar las características por una etapa de decodificación utilizando (en muchos casos) la misma arquitectura de codificación pero en sentido inverso. En teoría, si las características son representaciones correctas de los datos originales, las reconstrucciones deben ser muy similares a la entrada.

Los autocodificadores variacionales son una técnica relativamente reciente [51], por lo que se cuenta con pocos estudios en el área de bioacústica. Rowe et al. [48] usa esta técnica para el análisis de la biodiversidad enfocada a la detección de especies de aves; Ntalampiras y Potamitis [49] realizaron la implementación de esta para la detección de aves desconocidas.

El VAE es una mejora de los autocodificadores tradicionales en la que el espacio latente generado por la etapa de codificación se utiliza para aprender una distribución normal de las características y luego reconstruir las entradas originales mediante el muestreo de características. Este método proporciona un marco de referencia para el aprendizaje de modelos profundos de variables latentes y los correspondientes modelos de inferencia.

Las funciones de perdida habitualmente usadas en los autocodificadores, trabajan con la diferencia de la salida y la entrada. En el caso de los autocodificadores variacionales, se les agrega un factor de perdida llamado *KL-divergence* (Divergencia de Kullback-Leibler), este factor en vez de medir la distancia entre puntos, mide la diferencia existente entre dos distribuciones de probabilidad. Para este caso se escogió como distribución de probabilidad una distribución normal con centro en el origen y desviación estándar 1 (Ec. 5.3).

$$KL = \sum_{i=1}^{N} \sigma_i^2 + \mu_i^2 - \log(\sigma_i) - 1 \tag{5.3}$$

Donde σ_i^2 es la varianza de la función de probabilidad y μ_i^2 es la mediana al cuadrado de la misma función. Sin embargo, para este caso, si solo se calcula el valor KL como función de perdida, el autoencoder asumiría que no hay diferencia entre las especies. Para solucionar esto, se incluye una función de perdida (Ec.5.4)

$$Loss' = Loss(x, x') + KL \tag{5.4}$$

Donde x es el dato de entrada, x' es la salida, $Loss$ es la función de perdida (distancia euclidiana entre x y x') y KL es la *KL-divergence* calculada en Ec. 5.3.

La arquitectura de VAE utilizada se basa en el trabajo propuesto en [52]. Para la codificación se tiene una red neuronal compuesta por tres capas convolucionales y sus correspondientes funciones tipo paso (relu) que activa la neurona solo si se supera el umbral de cero, al ser mayor, entrega una salida idéntica al estado de la neurona. Las entradas corresponden a segmentos del espectrograma RGB de tamaño 224×224. Luego, para cada capa el tamaño de la entrada se reduce a la mitad y la cantidad de canales aumenta hasta llegar a 64. El espacio latente se generó calculando la media y la desviación estándar en una capa completamente conectada al final del codificador. La reconstrucción se realizó usando cinco capas deconvolucionales y capas relu entre ellas. Del espacio latente se obtiene un vector $\mathbf{y_b} \in \mathbb{R}^{N_L}$ donde N_L es el número de características, en este caso, 64, estas pueden usarse como características de entrada del algoritmo de agrupamiento. Las características $\mathbf{y_b}$ se evaluaron inicialmente y luego se concatenaron con las características de frecuencia (frecuencia mínima, máxima y frecuencia pico). La figura 5.1 presenta la arquitectura usada para la extracción de características para el caso de múltiples especies.

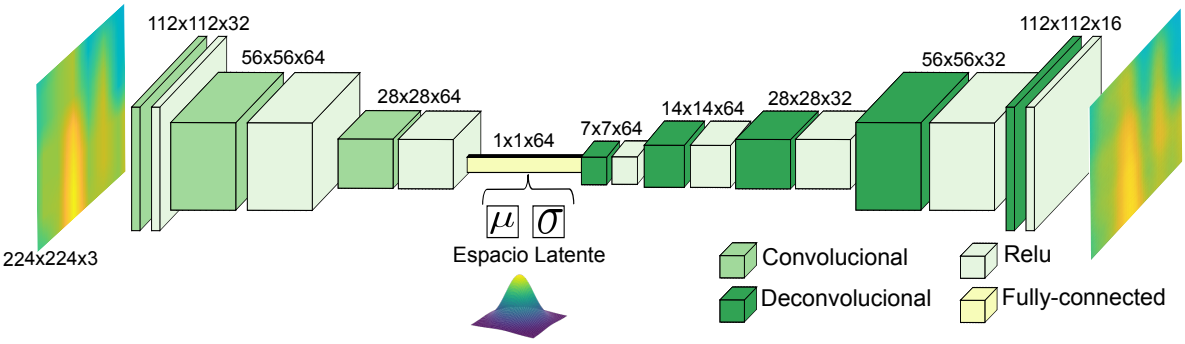

Figura 5.1: Arquitectura del autocodificador variacional utilizada para la extracción de características para la identificación de especies.

5.3. Red Neuronal Convolucional

En trabajos recientes se ha realizado la extracción de características a través de métodos de aprendizaje profundo. La mayoría de ellos utilizan redes neuronales convo-

lucionales (CNN) para ambas tareas (clasificación y extracción de características) [20]. Esta metodología presenta la ventaja de tener diferentes arquitecturas pre-entrenadas como en el caso de la VGG19 [51], ResNet50 [53], entre otras, que varían entre si en el número de capas, función de perdida y orden de los diferentes componentes en sus capas. Sin embargo, al ser un método supervisado tiene la limitación de requerir un conjunto de datos etiquetado.

Para el propósito de extracción de características, se analizó una red neuronal convolucional basada en la arquitectura KiwiNet propuesta por Bedoya y Molles [6]. Esta red fue creada para identificar individuos de una especie de ave (Great spotted kiwi/roroa) en Nueva Zelanda. Esta propuesta utiliza como núcleo la arquitectura VGG19 [51] con dos capas convolucionales adicionales que permiten reducir el número de filtros de 512 a 32 y una capa de agrupación de promedios globales para generar un espacio latente unidireccional, como se muestra en la figura 5.2.

Se decidió trabajar con la arquitectura KiwiNet por su alta precisión en la identificación de individuos de una especie de ave usando el segmento del canto de la especie.

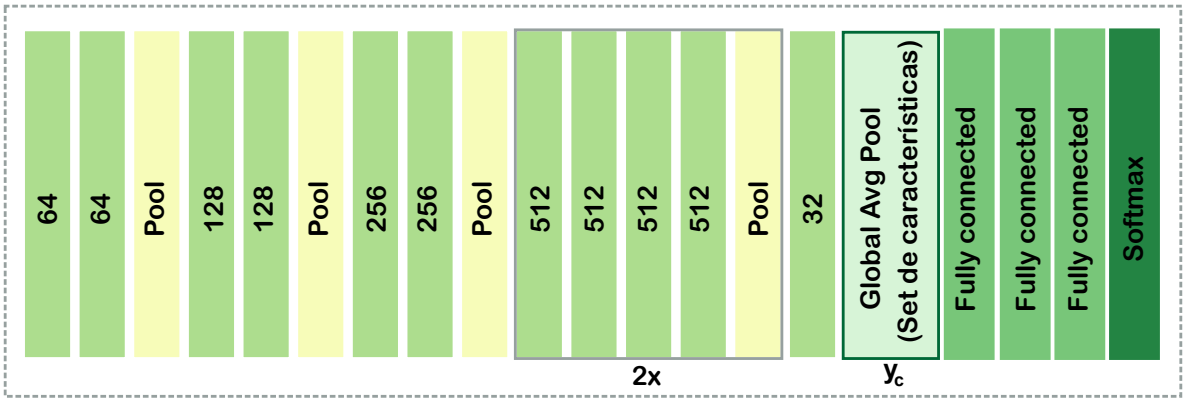

Figura 5.2: Arquitectura de CNN utilizada para la extracción de características basada en la propuesta de Bedoya y Molles [6].

Las CNN se caracterizan por requerir un ajuste fino de los parámetros según la aplicación. La entrada de la red KiwiNet son segmentos de imagen de espectrograma RGB de 224×224 que contienen las vocalizaciones de las especies, esta entrada viene por defecto de la arquitectura de la VGG19. Para el entrenamiento de la red se utilizaron segmentos etiquetados para cada especie y se dividieron en ochenta por ciento para entrenamiento y veinte por ciento para validación. Las primeras siete capas se congelaron y como se presenta en [6], se modifican y se añaden dos capas antes de la capa completamente conectada como muestra la figura 5.2. El entrenamiento se realizó usando el descenso del gradiente estocástico con pasos en mini-lotes de ocho segmentos, una tasa de aprendizaje de 1×10^{-4} y un criterio de parada definido por el número de

épocas. Se utilizó el aprendizaje por transferencia aplicando la arquitectura KiwiNet con valores de la VGG19. Esta arquitectura fue entrenada previamente con el conjunto de datos de ImageNet; esto permitió acelerar el flujo de trabajo de entrenamiento y realizar el entrenamiento con un conjunto de datos pequeño. A partir de los patrones generados por las capas convolucionales se obtiene un vector $\mathbf{y_c} \in \mathbb{R}^{N_z}$ donde N_z es el número de características, en este caso, 32.

5.3.1. KiwiNet como clasificador supervisado

Debido a que las CNN se utilizan comúnmente como método de clasificación, se estimó el desempeño de la red KiwiNet entrenada para clasificar 30 especies de 4 grupos taxonómicos diferentes (aves, anfibios, mamíferos e insectos). La sección de resultados (tabla 5.1) presenta el rendimiento de la red KiwiNet como clasificador supervisado para el caso multi-especie.

5.4. Resultados

En este capitulo se analizaron tres métodos de extracción de características y se evaluó el desempeño de ellos en la identificación de múltiples especies. Cada método fue analizado de manera individual y luego en combinación con las características de frecuencia (frecuencia pico, frecuencia mínima y frecuencia máxima) de cada segmento. Se usó el algoritmo de agrupamiento LAMDA 3Π (ver capítulo 6 y se comparó también con la CNN como clasificador. La Tabla 5.1 muestra el promedio de la tasa de reconocimiento de las 30 especies que fueron etiquetadas por un experto en el subconjunto de datos C y el conjunto de datos D usando los diferentes métodos de extracción de características, el algoritmo de agrupamiento LAMDA 3Π y la CNN como clasificador.

Características	Clasificador	Desempeño
Información de frecuencia	LAMDA 3Π	84 %
LFCC con información de frecuencia	LAMDA 3Π	95 %
VAE	LAMDA 3Π	78 %
VAE con información de frecuencia	LAMDA 3Π	92 %
KiwiNet - CNN	LAMDA 3Π	89 %
KiwiNet - CNN con información de frecuencia	LAMDA 3Π	90 %
KiwiNet - CNN	CNN	88 %

Tabla 5.1: Comparación del rendimiento en la identificación de múltiples especies utilizando diferentes métodos de extracción de características.

Los resultados de la Tabla 5.1 muestran que la tasa de reconocimiento aumenta significativamente cuando la información de frecuencia se usa en combinación de los otros tres métodos probados, lo que demuestra que la información biológica es relevante para diferenciar entre especies.

La extracción de características usando LFCC en combinación con la información de frecuencia, muestra que es posible realizar la identificación de especies en diferentes bandas de frecuencia y obtener un desempeño alto. Esta combinación permite que el algoritmo de agrupación genere clústeres precisos para las especies. Sin embargo, como problema biológico, algunas vocalizaciones de las especies presentan variabilidad intra-especie debido al ruido de fondo o a las condiciones del ambiente y es posible que esto afecte la detección de algunas especies.

Por otra parte, el autocodificador variacional (VAE) es capaz de codificar y decodificar a partir de la representación de la llamada de la especie en el espectrograma, logrando una reconstrucción acertada como se muestra en la figura 5.3. Al ser un método no supervisado, presenta la ventaja de no requerir etiquetas o grandes conjuntos de datos (en este caso de vocalizaciones de especies) para reconstrucciones precisas. Sin embargo, cuando el espacio latente se usa como entrada para un algoritmo de agrupamiento, como en este caso, algunas características pueden llegar a no ser relevantes para agrupar algunas especies correctamente por lo que el promedio global de detección disminuye. Cuando las características extraídas del espacio latente se usan en combinación de la información de la frecuencia de las especies, la detección de estas mejora.

La figura 5.3 presenta uno de los resultados obtenidos en la reconstrucción de las vocalizaciones de las especies, en este caso se presenta para la especie de ave *Troglogytes aedon* y la especie de murciélago *Molosssus bondae*

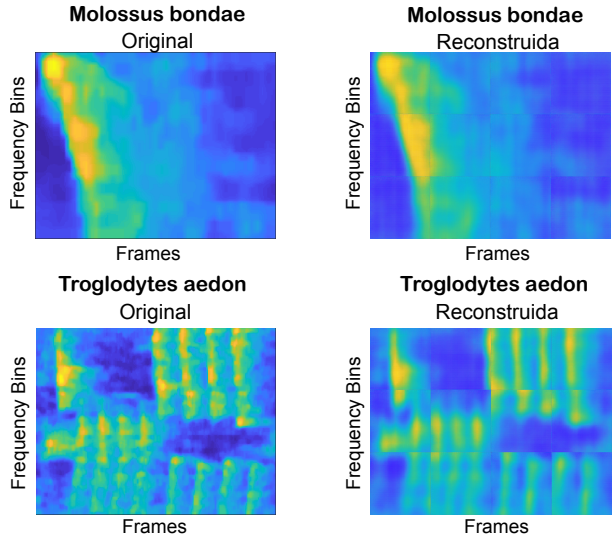

Figura 5.3: Ejemplos de llamadas originales y reconstruidas de las especies *Molosssus bondae* (murcielago) y *Troglogytes aedon*(ave) usando la arquitectura de VAE propuesta.

El desempeño de la red KiwiNet para la extracción de características fue similar a los otros métodos evaluados. La arquitectura KiwiNet se diseñó originalmente para la

extracción de características de individuos de la especie Great spotted kiwi (especie de ave de Nueva Zelanda) [6], pero al realizar un ajuste a los parámetros de esta arquitectura, se encontró que este enfoque resulta útil a la hora de hacer la identificación de multiples especies.

KiwiNet como cualquier otra arquitectura de CNN, requiere un conjunto de datos extenso y debidamente etiquetado para el entrenamiento del modelo. Debido a la diversidad de llamadas de especies, el conjunto de datos no estaba balanceado para todas las especie. Fue necesario encontrar manualmente los segmentos de canto mas representativo para cada especie y utilizarlos para el entrenamiento. Se usaron 40 vocalizaciones de cada una de las 30 especies. Esto dificultó el entrenamiento de la red.

A pesar de contar con un método supervisado como lo es la CNN, se decidió evaluar el desempeño de la red KiwiNet como clasificador debido a que recientemente, las redes neuronales convolucionales son uno de los métodos mas usados para la identificación de especies [11], [20] obteniendo un desempeño del 88 %. Esta aproximación proporcionó una linea de base para comparar la propuesta no supervisada.

La Tabla 5.2 presenta el tiempo de ejecución del proceso completo para la identificación de múltiples especies utilizando cada método de extracción de características. Es importante considerar que la etapa de entrenamiento en arquitecturas de CNN requieren recursos computacionales significativos. Este análisis se realizó en un computador de alto rendimiento (Ryzen 5 3600, 16GB RAM, Nvidia RTX 2060 super) donde la CNN es notoriamente el método que mas tiempo consume.

Características	**Tiempo de ejecución (Minutos)**
Información de frecuencia	18
LFCC con información de frecuencia	34
VAE	66
VAE con información de frecuencia	72
KiwiNet - CNN	178
KiwiNet - CNN con información de frecuencia	184

Tabla 5.2: Comparación de los tiempos de ejecución para realizar la identificación de múltiples especies utilizando diferentes métodos de extracción de características.

Finalmente, los resultados de identificación de las 30 especies analizadas se presentan en la figura 5.4. Estos resultados muestran la diferencia entre la identificación de especies con cada método de extracción de características analizado. En general, las especies de alta frecuencia como es el caso de los murciélagos (la mayoría de los mamíferos) y los ortópteros (insectos) obtuvieron un alto rendimiento, esto puede deberse a que en altas frecuencias, la presencia de ruido de fondo es baja o casi nula, lo que hace que la detección mejore.

La detección de especies en todo el espectro acústico estuvo por encima del 50 %. Esto muestra que la extracción de características usando LFCC y el uso de métodos mas recientes como los modelos de aprendizaje profundo en combinación con la información de frecuencia de las especies proporcionan información relevante sobre las vocalizaciones al algoritmo de agrupamiento. Sin embargo, el mejor desempeño (tanto en tasa de reconocimiento como en costo computacional) se obtiene con los LFCC en combinación con las características de frecuencia como entrada al algoritmo de agrupamiento.

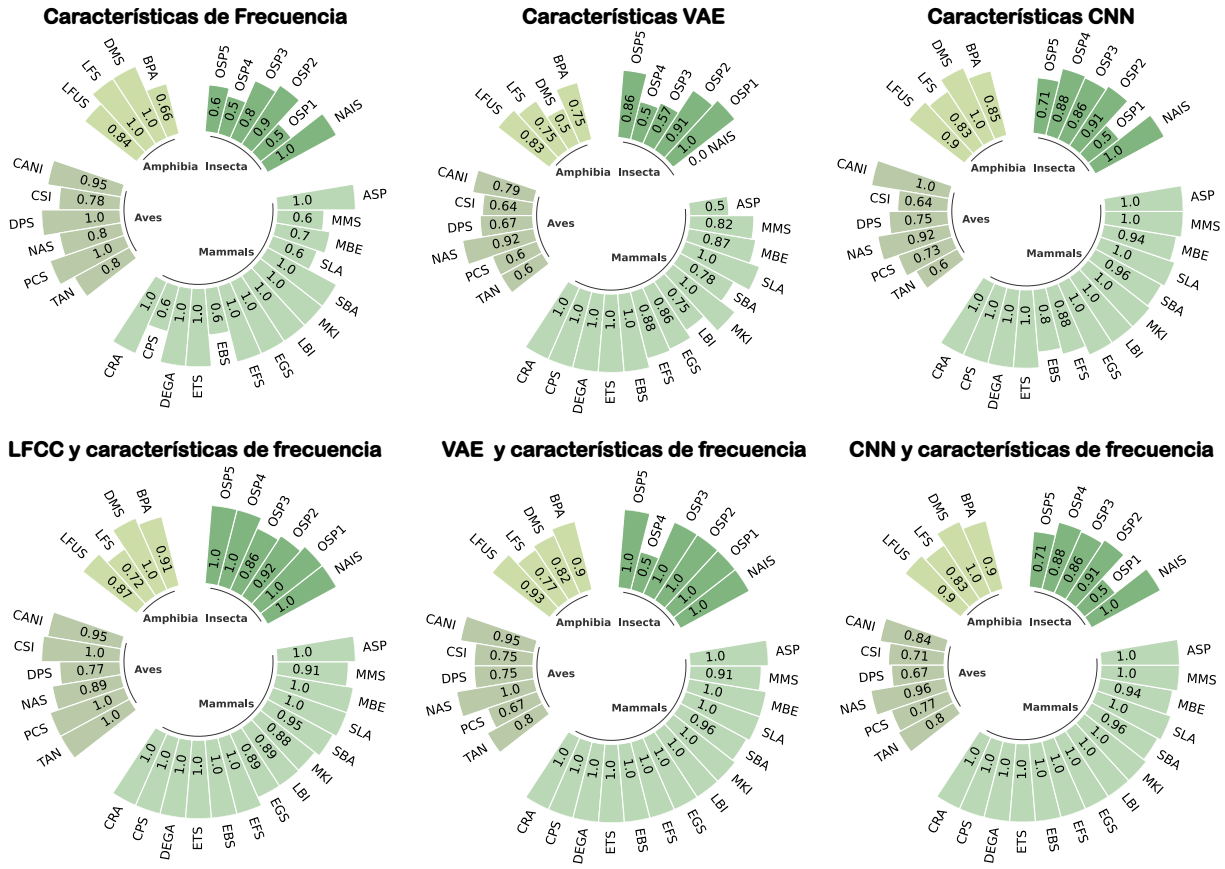

Figura 5.4: Resultados de la identificación de múltiples especies utilizando diferentes métodos de extracción de características con el algoritmo de agrupamiento. Hay seis diagramas de barras circulares, cada uno de los cuales representa los resultados de las especies utilizando cada representación de características. Cada barra representa el rendimiento de cada especie analizada. Los resultados de las especies están organizados según su grupo taxonómico (aves, anfibios, mamíferos e insectos)

Los nombres de las 30 especies fueron codificados, el nombre completo de cada especie se presenta en la Tabla 5.3.

No.	Nombre-Especie	Código	No.	Nombre-Especie	Código
1	*Boana platanera*	BPA	16	*Eptesicus furinalis*	EFS
2	*Dendropsophus microcephalus*	DMS	17	*Eumops glaucinus*	EGS
3	*Leptodactylus fragilis*	LFS	18	*Lasiurus blossevillii*	LBI
4	*Leptodactylus fuscus*	LFUS	19	*Myotis keaysi*	MKI
5	*Crotophaga ani*	CANI	20	*Saccopteryx bilineata*	SBA
6	*Crypturellus soui*	CSI	21	*Saccopteryx leptura*	SLA
7	*Dendroplex picus*	DPS	22	*Molossus bondae*	MBE
8	*Nyctidromus albicollis*	NAS	23	*Molossus molossus*	MMS
9	*Patagioenas cayennensis*	PCS	24	*Alouatta sp.*	ASP
10	*Troglodytes aedon*	TAN	25	*Neoconocephalus affinis*	NAIS
11	*cf. Rhogeessa*	CRA	26	Orthoptera sp. 1	OSP1
12	*Cynomops*	CPS	27	Orthoptera sp. 2	OSP2
13	*Dasypterus ega*	DEGA	28	Orthoptera sp. 3	OSP3
14	*Eppia truncatipennis*	ETS	29	Orthoptera sp. 4	OSP4
15	*Eptesicus brasiliensis*	EBS	30	Orthoptera sp. 5	OSP5

Tabla 5.3: Resumen del nombre de las 30 especies analizadas y el respectivo código asignado.

5.5. Conclusiones

En este capítulo se analizaron tres métodos diferentes de extracción de características: LFCC, VAE y CNN. Los autocodificadores variacionales son un método no supervisado que demostraron que es posible una detección precisa en un enfoque multi-especie en combinación con características de frecuencia. Este método presenta la ventaja de trabajar con el conjunto de datos completo debido a que no requiere etiquetas de especies. Este modelo obtuvo un bajo error de reconstrucción. Sin embargo, se utilizó el espectrograma con el segmento de llamada como una imagen, y como imagen, el espacio latente fue apropiado para la reconstrucción, pero podría mejorarse para las tareas de *clustering*.

Los métodos de aprendizaje profundo supervisado, como la CNN, mostraron un rendimiento adecuado en la tarea de identificación de múltiples especies. Sin embargo, este método depende en gran medida de un conjunto de datos robusto, incluso con aprendizaje por transferencia. Esto puede ser un problema en las aplicaciones biológicas, donde se recogen muchos datos de audio para el seguimiento de la biodiversidad y la información de las especies no siempre está disponible.

Por otra parte, se encontró que los métodos de aprendizaje profundo para la extracción de características permiten al algoritmo de agrupación realizar la identificación de especies de diferentes grupos taxonómicos. Sin embargo, los métodos basados en análisis frecuencial, como los LFCC, demostraron ser más rápidos en tiempo de ejecución sin

sacrificar el rendimiento. Además, las variables de frecuencia demostraron ser una contribución significativa para la identificación de múltiples especies. Como características de entrada al algoritmo de agrupación se eligen los coeficientes cepstrales escalados linealmente en combinación de las características de frecuencia del segmento (frecuencia maxima, minima y pico) $\mathbf{y_a}$. Este fue implementado en la metodología propuesta, mostrada en el capítulo 3.

Con este capítulo se aporta en el cumplimiento de parte del objetivo específico 1: "Identificar algoritmos de reducción de ruido, segmentación y extracción de características que permitan detectar diferencias entre especies"

Capítulo 6

Clustering e identificación de especies de animales

Es posible realizar la identificación de las vocalizaciones de las especies de manera manual utilizando herramientas de computo (por ejemplo, Raven Pro [54]) que permiten visualizar el espectrograma, segmentar la vocalización y agregarle una etiqueta. Sin embargo, esta tarea requiere habilidad y experiencia por parte del experto, demanda tiempo y está limitada para ser usada con pequeños conjuntos de datos.

Por otra parte, están los métodos de identificación automática de especies la cual suele basarse en métodos de aprendizaje supervisado. Técnicas basadas en redes neuronales convolucionales, máquinas de soporte vectorial y algoritmos de bosque aleatorio dominan esta área de investigación [9]-[11], [20]. Sin embargo, trabajar con enfoques de aprendizaje supervisado requiere datos etiquetados, lo que implica un conocimiento previo de las especies en un sitio de interés. Esto representa un problema en los biomas tropicales y los focos de biodiversidad donde la mayoría de las especies son desconocidas para la ciencia [55]-[57] como es el caso de Colombia. En estas regiones, las descripciones taxonómicas y de vocalizaciones no están disponibles para la mayoría de las especies [58] e incluso para los taxones conocidos, no se tienen suficientes muestras lo que complica el desarrollo de las técnicas de aprendizaje automático [59]. Además de esto, la mayoría de los métodos de identificación automática de especies realizan el análisis de manera especie-específico, es decir, solo buscan vocalizaciones asociadas a una especie en particular y se concentran en una banda de frecuencia específica. Para ello, se debe tener conocimiento a priori de la especie que se esté buscando y que el método esté entrenado para reconocer la especie de interés.

Debido al tipo de problema donde no siempre se cuenta con datos etiquetados y donde existe diversidad y complejidad en los cantos de las especies, se decide trabajar con un método no supervisado (*clustering*). De esta manera, los hiperpuntos representados por las características extraídas, en este caso los LFCC con información de frecuencia del segmento (ver capítulo 5), se agruparan según su cercanía en el espacio $\mathbb{R}^{N_\mathrm{f}}$. Cada

clúster agrupa los puntos correspondientes a un patrón de llamada y la variabilidad intraespecífica se preserva al tener varios clústeres que pueden asociarse a la misma especie. Por esta razón, se le da el nombre de sonotipo a cada clúster pues representa un patrón de sonido que puede estar asociado a una especie.

Dado que no se conoce las especies esperadas en el audio, es necesario utilizar un algoritmo de *clustering* que no requiera el número de clústeres como parámetro de entrada. Por ello, se elige el algoritmo LAMDA (*Learning Algorithm for Multivariate Data Analysis*)[60] con la variación 3Π [61] para realizar la tarea de agrupar los sonidos bioticos provenientes del paisaje sonoro.

6.1. LAMDA - *Learning Algorithm for Multivariate Data Analysis*

LAMDA es un método que no requiere que el usuario elija a priori el número de clases (para este caso, número de especies) como parámetro de entrada. La elección de este parámetro es uno de los principales cuellos de botella en el desarrollo de enfoques no supervisados, ya que el número de clases es posiblemente uno de los hiperparámetros mas importantes en un problema de agrupación.

Para este trabajo se implementó la versión Yager-Rivalov triple Π [61] que utiliza un operador de agregación difusa que restringe de manera natural el número de clústeres generados y ha sido validado en análisis de datos bioacústicos [6], [41]. Las características extraídas de todos los segmentos fueron normalizadas antes de utilizarlas como entrada para el algoritmo de agrupación.

El primer paso de LAMDA consiste en calcular los Grados de Adecuación Marginal (MAD) $\mathbf{M}$ (Ec.6.1), los cuales son las contribuciones de las características extraídas (coeficientes cepstrales, frecuencia mínima, frecuencia máxima, y frecuencia pico) de cada elemento (vocalización/unidad acústica) a cada uno de los clústeres existentes.

$$\mathbf{M}_{c,f} = \rho_{c,f}{}^{\widehat{y}_f}(1-\rho_{c,f})^{1-\widehat{y}_f} \tag{6.1}$$

Donde $\mathbf{M} \in \mathbb{R}^{N_\mathrm{c}\times N_\mathrm{f}}$ es la matriz con los valores de MAD extraídos del elemento analizado, $\boldsymbol{\rho} \in \mathbb{R}^{N_\mathrm{c}\times N_\mathrm{f}}$ es la matriz con los valores medios de las características N_f en cada c-th clúster, $\widehat{\mathbf{y}} \in \mathbb{R}^{N_\mathrm{f}}$ es el vector con los valores de características normalizados del elemento analizado, $f = 1, \ldots, N_\mathrm{f}$ es la característica actual, $c = 1, \ldots, N_\mathrm{c}$ es el clúster actual, N_f es el número de características y N_c es el número de clústeres existentes.

Inicialmente, el único clúster predefinido es la Clase No Informativa (NIC), que acepta todos los elementos por igual ($\rho_{0,f} = 0,5 \forall f = 1, \ldots, N_\mathrm{f}$). El primer elemento se asigna siempre a la NIC, por lo que se considera como no reconocido. Después, se crea una

nueva clase con los parámetros de la clase NIC modificados por los valores del primer elemento:

$$\rho_{1,f} = \frac{(\widehat{y}_f + \rho_{0,f})}{2} \tag{6.2}$$

Cada vez que se analiza un elemento nuevo (un vector de características de un segmento) (Ec. 6.1), los MAD obtenidos se combinan usando un operador de agregación reforzado (Ec. 6.3). El resultado de esta operación se conoce como el Grado de Adecuación Global (GAD) $\mathbf{g}_c$ de un elemento a un clúster.

$$\mathbf{g}_c = \frac{\prod_{f=1}^{N_\mathrm{f}} M_{c,f}}{\prod_{f=1}^{N_\mathrm{f}} M_{c,f} + \prod_{f=1}^{N_\mathrm{f}} (1 - M_{c,f})} \tag{6.3}$$

Donde $\mathbf{g} \in \mathbb{R}^{N_\mathrm{c}}$ se calcula utilizando los MAD de la nueva entrada. Una vez obtenidos los GAD de todos los clústeres, el elemento se clasifica en el clúster con el máximo valor de GAD. Si dicho valor maximo de GAD corresponde a la clase NIC, se crea un nuevo clúster utilizando los parámetros de la NIC actualizados con los valores del elemento (Ec. 6.2). Por otro lado, si un elemento se asigna a una clase c existente, los parámetros de la clase se actualizan con los valores del elemento entrante (Ec. 6.4).

$$\rho_{c,f}^{(k)} = \rho_{c,f}^{(k-1)} + \frac{\widehat{y}_f - \rho_{c,f}^{(k-1)}}{n_c^{(k)}} \tag{6.4}$$

Donde c es el clúster actual, $n_c^{(k)}$ es el numero de elementos clasificados en el clúster c y $\rho_{c,f}^{(k-1)}$ es el valor k $-$ 1 anterior de $\rho_{c,f}$ (antes de la actualización).

Cada clúster resultante representa un patrón sonoro especifico denominado sonotipo y cada grupo, según el sonotipo representado, se asocia a la vocalización de una especie animal.

El experto asocia los sonotipos que considera representativos con las especies de interés. Luego, con esta asociación, es posible identificar la presencia de la especie (o las especies) de interés en nuevos conjuntos de datos.

En el capítulo 7 se presentan los resultados de la metodología para identificación de vocalizaciones de múltiples especies propuesta.

Capítulo 7

Resultados de la metodología para identificación de vocalizaciones de múltiples especies

A continuación, se presentan los resultados obtenidos con la metodoogía propuesta en el capítulo 3 con las consideraciones explicadas en los capítulos 4, 5 y 6. Para determinar el desempeño de esta propuesta, se usaron las bases de datos, casos de estudio y métricas descritas en el capítulo 2.

7.1. Caso 1: Reconocimiento de vocalizaciones de múltiples especies de animales

En este caso de estudio se puso a prueba la metodología para reconocimiento de vocalizaciones de múltiples especies utilizando el conjunto de datos A. Se usó esta propuesta para caracterizar toda la actividad acústica en un sitio con gran biodiversidad biológica. Utilizando la metodología propuesta en el capítulo 3, se encontraron los clústeres automáticamente y los correspondientes sonotipos asociados. A continuación, un experto en las especies de la zona asoció los sonotipos de los clústeres resultantes a las respectivas vocalizaciones de las especies.

Se identificó la presencia/ausencia de 39 especies de la zona. La tabla 7.1 muestra el desempeño en la detección de cada grupo taxonómico y la figura 7.1 muestra el desempeño de cada una de las especies identificadas. De todas las especies analizadas, solo cinco de ellas fueron detectadas con un desempeño inferior al 60 % (Fig 7.1). Se presentaron varios casos en los que una especie no tenía asignado el género y epíteto específico, esto es común en países tropicales como Colombia, el cual es considerado como un foco de biodiversidad y muchas vocalizaciones emitidas por las especies presentes en el paisaje sonoro son aún desconocidas. Sin embargo, se pudo identificar claramente su grupo taxonómico (por ejemplo, Avian sp.1, Orthoptera sp.1,..., etc.)

Taxón	Frecuencias	Sensibilidad	Especificidad
Amphibia	2 – 5 kHz	93 %	91 %
Aves	2 – 10 kHz	85 %	90 %
Insecta	4 – 7 kHz	90 %	95 %

Tabla 7.1: Desempeño de la metodología propuesta en la detección de especies de diferentes grupos taxonómicos.

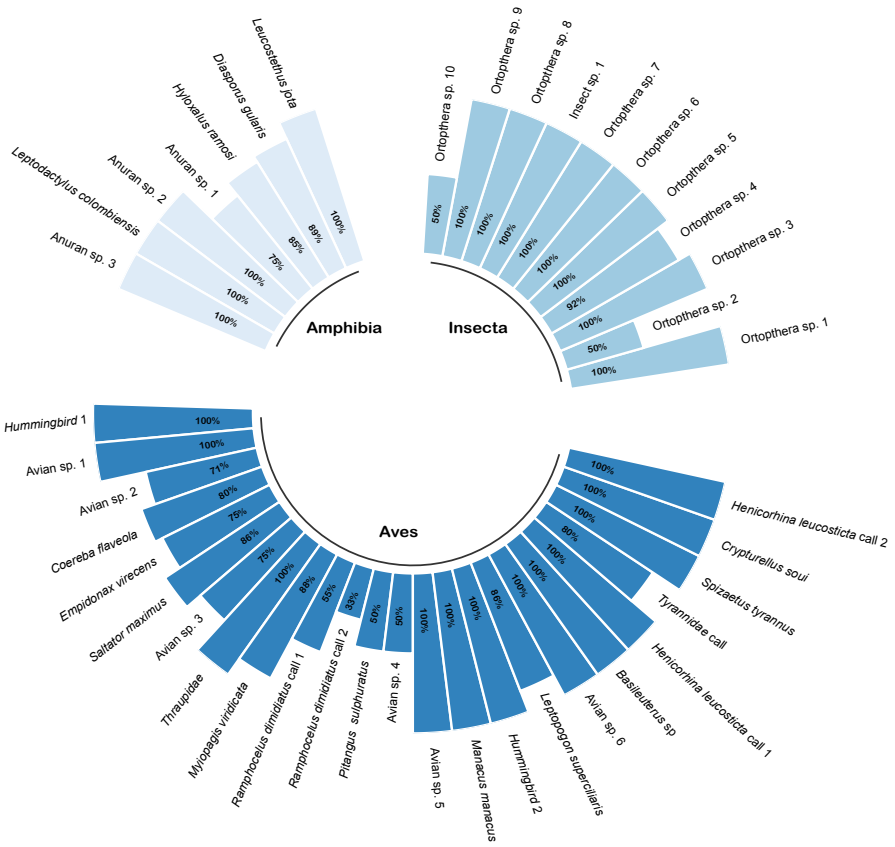

Figura 7.1: Resultados de agrupación del método propuesto para 41 especies. Cada barra representa la tasa de acierto para cada especie. Los resultados están representados de acuerdo al grupo taxonómico (Aves, Amphibia, Insecta).

Este método diferenció dos especies de anuros *Leucostethus jota* y *Hyloxalus ramosi* que vocalizan en rangos de frecuencia similares como se observa en la figura 7.2. Las vocalizaciones de estas especies se encuentran cerca a los 4.5 kHz y tienen un patrón de llamada similar. Este método identificó los llamados como sonotipos diferentes dejando las llamadas de *Leucostethus jota* en un grupo (rectángulos rojos) e *Hyloxalus ramosi* en otro (rectángulos azules), lo que permitió diferenciar entre las dos especies con lla-

madas similares.

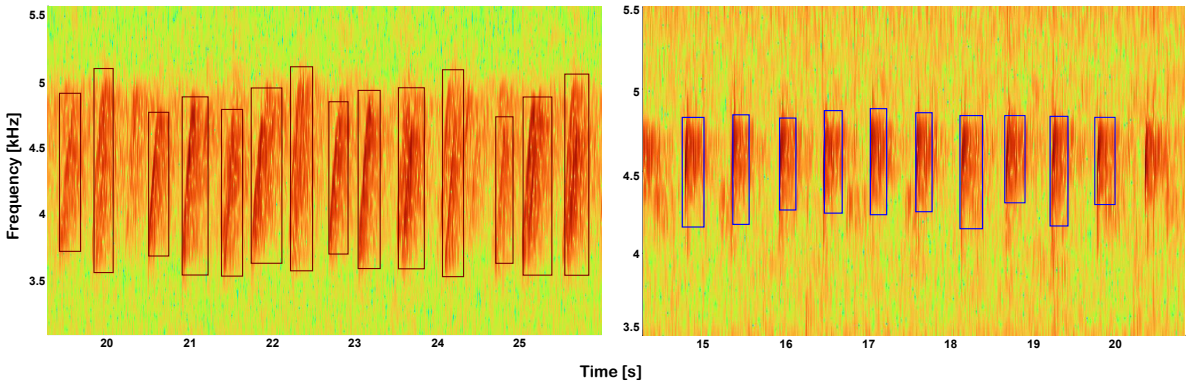

Figura 7.2: Vocalizaciones de especies de anuros. Los segmentos rojos están asociados a la espeice*Leucostethus jota* y los segmentos azules a *Hyloxalus ramosi*. La metodología propuesta fue capaz de diferencias las llamadas y separarlas en diferentes grupos.

Especies con alta complejidad vocal (es decir, llamadas de larga duración con notas en múltiples bandas de frecuencia) se identificaron en la mayoría de los casos utilizando la llamada completa o parte significativa de esta, como en el caso de la especie de ave *Basileuterus* sp. (figura 7.3). La llamada de esta especie de ave oscila en un rango de frecuencias entre 6 - 10 kHz y tiene diferentes notas. Esta propuesta permite detectar las variaciones que pueden estar presentes en los cantos de las especies al contar con un algoritmo de *clustering* que no busca un patrón especifico de cada especie, sino la similitud con el patrón representado por los clústeres.

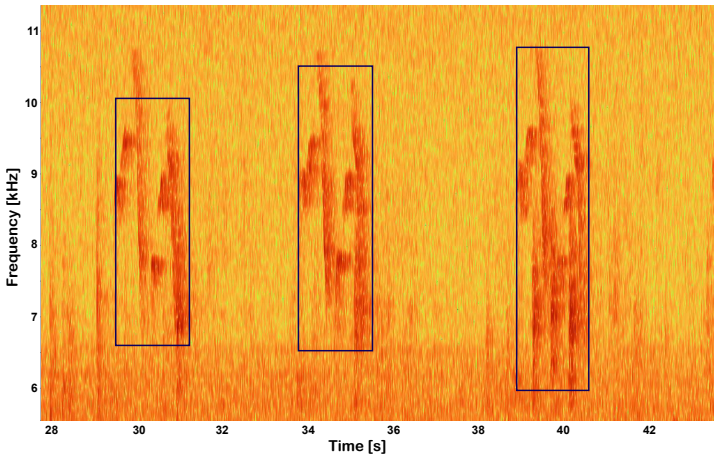

Figura 7.3: Vocalización compleja asociada a la especie de ave*Basileuterus* sp. Cada rectángulo azul es un segmento de la llamada de la especie detectado por la metodología propuesta.

Para las especies que presentaron bajo desempeño (*Ramphocelus dimidiatus, Pitangus sulphuratus*, Avian sp. 5, Otrhoptera sp. 2, Orthoptera sp. 10), se tiene la hipótesis de que esto se debió al número reducido de ejemplos de llamadas disponibles en las grabaciones, con las cuales no se pudo generar una agrupación propia para las vocalizaciones.

7.2. Caso 2: Comparación con otros algoritmos de identificación de especies

En este caso de estudio, se comparó el desempeño de la metodología propuesta con otras herramientas computacionales como Autodetec del paquete WarbleR [26], MonitoR [27], y el software Kaleidoscope Pro [28]. Estas herramientas permiten la detección de especies a partir de grabaciones de audio. Aunque los métodos propuestos en los paquetes de R (WarbleR y MonitoR) no incluyen una etapa de *clustering*, se decidió incluirlo como referencia para comparar, ya que son ampliamente utilizados por los biólogos para la identificación de especies.

Se utilizó el conjunto de datos B que contiene mas grabaciones de audio respecto al conjunto A. Este conjunto de datos tiene cinco especies diferentes, etiquetadas manualmente y contiene algunas grabaciones con ruido de fondo fuerte.

Para la detección de especies con los paquetes de R, se utilizaron 200 grabaciones seleccionadas al azar para cada especie: *Leucostethus jota, Diasporus gularis, Espadarana prosoblepon, Boana boans*, y *Crypturellus soui*. Los parámetros de cada herramienta se definieron en función de las características acústicas de cada especie, como las bandas de frecuencia, la amplitud y la duración de la llamada. En el caso de WarbleR, existe un umbral de amplitud que ayuda a diferenciar la señal de interés del ruido de fondo. Para definir este parámetro, realizamos pruebas con tres valores diferentes: 10 %, 15 % y 20 %.

Para el análisis con Kaleidoscope Pro, se tomaron 100 audios de cada especie para la etapa de entrenamiento y 100 para la prueba. En los parámetros de la señal, el rango de frecuencias, la duración de la vocalización y el intervalo de tiempo máximo entre cada vocalización se modificaron según cada especie. Todos los parámetros para el análisis de cluster se dejaron por defecto, excepto la ventana FFT, en la cual se usó la opción 5,33 ms (128 @0-12kHz, 256 @13-24kHz, 512 @25-48kHz, 1024 @49-96kHz).

Para la metodología propuesta en este trabajo, se utilizaron los mismos conjuntos de datos de entrenamiento y prueba utilizados con Kaleidoscope Pro. Cada especie fue entrenada con el conjunto de entrenamiento (asociación de sonotipo a especie) y luego reconocida con los 100 audios de la partición de prueba. En este caso, no es necesario que el usuario establezca los parámetros para las especies.

En la tabla 7.2 se compara el enfoque propuesto con tres herramientas informáticas ampliamente utilizadas para el reconocimiento de especies. La mejor detección de especies se realizó con la metodología propuesta en este trabajo con un desempeño del 75 %, utilizando un método completamente no supervisado.

Especies	Software					
	ATH10	ATH15	ATH20	MonitoR	KP	Metodología propuesta
Leucostethus jota	0.55	0.47	0.94	0.69	0.62	0.74
Diasporus gularis	0.49	0.51	0.48	1.00	0.68	0.71
Esparadana prosoblepon	0.14	0.21	0.07	0.30	0.35	0.82
Boana boans	0.77	0.92	0.92	0.46	0.38	0.46
Cryptorellus soui	0.50	0.50	0.00	0.50	0.50	1.00
Detección promedio	0.49	0.52	0.48	0.52	0.51	**0.75**

Tabla 7.2: Resultados de detección de presencia-ausencia para cada especie analizadas usando las propuestas de librerías en R y el software disponible: WarbleR-Autodetec con un umbral de 10 % (ATH10), WarbleR-Autodetec con un umbral de 15 % (ATH15), WarbleR-Autodetec con un umbral de 20 % (ATH20), MonitoR, Kaleidoscope Pro (KP) y la metodología propuesta.

Para obtener un desempeño adecuado en los paquetes de R y en el software Kaleidoscope Pro, fue necesario ajustar los parámetros de entrada para cada una de las especies. Para ello, fue necesario conocer de antemano las características de la señal, como la banda de frecuencia de cada especie, la amplitud y la duración de la llamada. Además, en el caso de Kaleidoscope pro, fue necesario seleccionar los parámetros para el análisis de clúster, teniendo que elegir el tamaño de la ventana de Fourier, el número de clústeres y la distancia de los clústeres al centro. El enfoque propuesto en este trabajo, en comparación, no requiere de esta parametrización.

En este caso de estudio, la baja detección de las especies se debió a actividad de alta ganancia presente en múltiples bandas de frecuencia. Se analizó la relación señal a ruido y se encontró, como se presenta en la figura 7.4, una relación señal a ruido baja, indicando presencia de ruido en múltiples bandas de frecuencia. Además, se escucharon los audios y se encontraron varias grabaciones con presencia de lluvia. Por tanto, se puede suponer que el comportamiento asociado a la lluvia enmascaró la señal y la metodología propuesta fue capaz de detectar algunas de las vocalizaciones enmascaradas por el fuerte ruido de fondo como se muestra en la figura 7.5 donde se puede observar el espectrograma con actividad en múltiples bandas de frecuencia y la detección de los

cantos de las especies de anuros *Diasporus gularis* (3 kHz) y *Esparadana prosoblepon* (6 kHz).

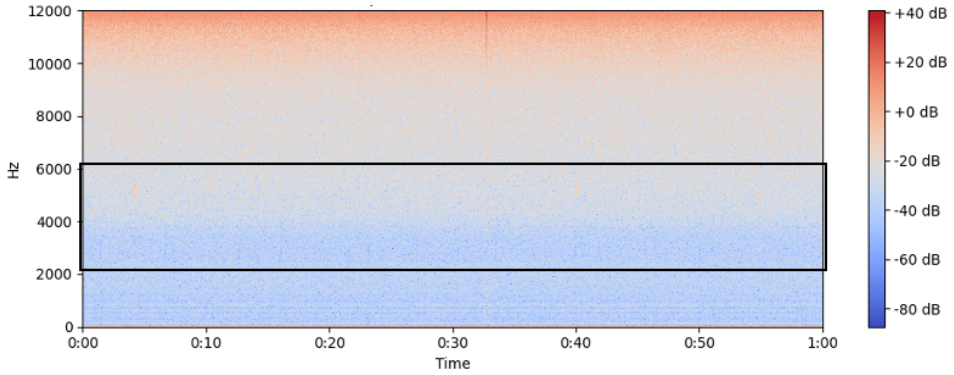

Figura 7.4: Análisis de relación señal a ruido de audio con posible ruido en múltiples bandas de frecuencia. Se encuentra una relación señal a ruido baja (en azul) para las frecuencias bajas, mostrando presencia de ruido.

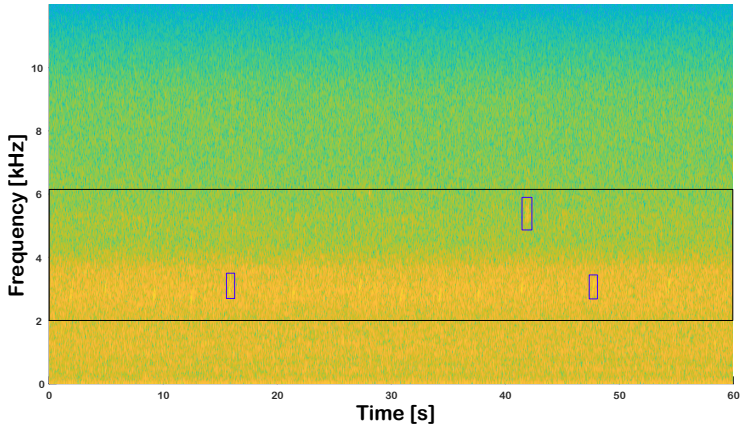

Figura 7.5: Grabación de audio con ruido de fondo asociado a múltiples bandas de frecuencia. Dos especies de anuros fueron detectadas *Diasporus gularis* (3kHz) y *Esparadana prosoblepon* (5.5kHz).

7.3. Caso 3: Validación del método con datos diferentes y aplicación para evaluación de la biodiversidad

Para evaluar el desempeño de la identificación de especies de la propuesta no supervisada, un grupo de expertos etiquetó manualmente un subconjunto de 321 grabaciones de audio seleccionadas al azar. Los audios incluyen llamadas de aves, anuros y una

especie de primate. Las llamadas de todas las grabaciones (del conjunto de datos C) fueron analizadas con la metodología propuesta. La tabla 7.3 muestra el desempeño en la detección de cada grupo taxonómico y la figura 7.6 muestra el desempeño de las 11 especies identificadas.

Taxón	Frecuencias	Sensibilidad	Especificidad
Amphibia	2 – 7 kHz	78 %	80 %
Aves	400 Hz – 15 kHz	93 %	84 %
Mammals	400 Hz	100 %	86 %

Tabla 7.3: Desempeño de la metodología propuesta en la detección de especies de diferentes grupos taxonómicos en una zona geográfica diferente.

Figura 7.6: Resultados de agrupación del método propuesto para 11 especies. Las barras indican las tasas de aciertos para cada especie en todas las grabaciones del conjunto de datos C. Los resultados están representados de acuerdo con su grupo taxonómico(Aves, Amphibia y Mammals).

En este caso de estudio se encontraron casos en los que la metodología propuesta, detectó vocalizaciones con baja intensidad en el espectrograma (relación señal a ruido alta, figura 7.7), como en el caso de la especie de ave *Nyctidromus albicolis* (Fig. 7.8). Además, fue posible encontrar especies cantando en la misma banda de frecuencias en el mismo audio, como ocurre con la especie de ave *Nyctidromus albicolis* y la especie de anuro *Leptodactylus fragilis* (Fig. 7.9). En este caso, el algoritmo de agrupación puede fallar y la tarea del experto aquí, es escoger el clúster (sonotipo) que mejor describa a cada especie.

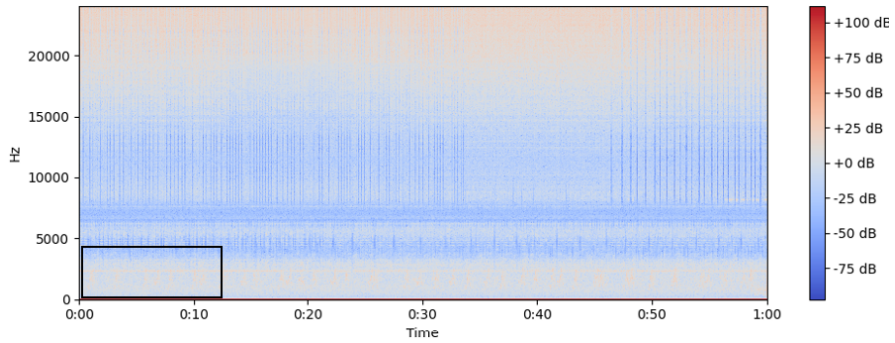

Figura 7.7: Relación señal a ruido alta en audio con detección de vocalizaciones con baja intensidad indicando ausencia de ruido significativo en la banda de frecuencia de interés en el audio.

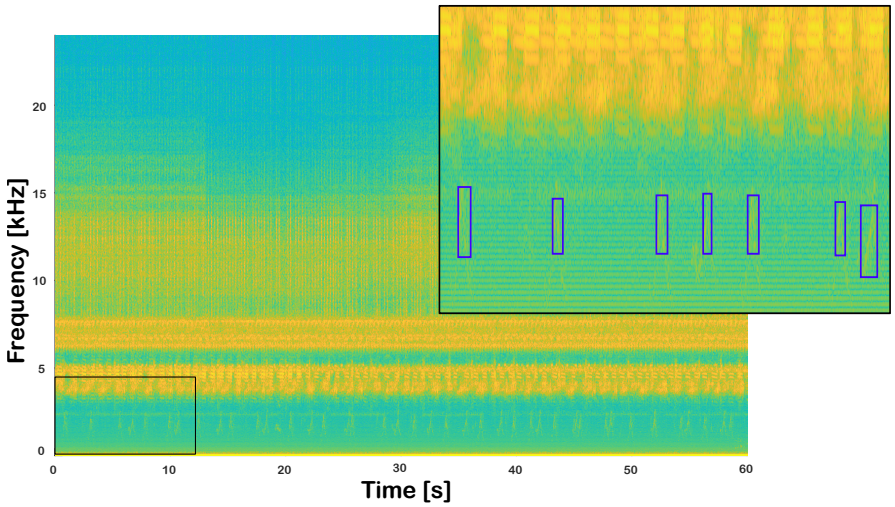

Figura 7.8: Zoom de la llamada de la especie de ave *Nyctidromus albicollis*. El espectrograma de la especie contiene baja ganancia (la especie se ubica entre 900 Hz and 4.5 kHz). Sin embargo, las llamadas de la especie fueron segmentadas por el método propuesto.

En este caso de estudio, se aplicó el método propuesto como indicador de biodiversidad, asociando el número de sonotipos propuestos (clústeres) a la riqueza acústica

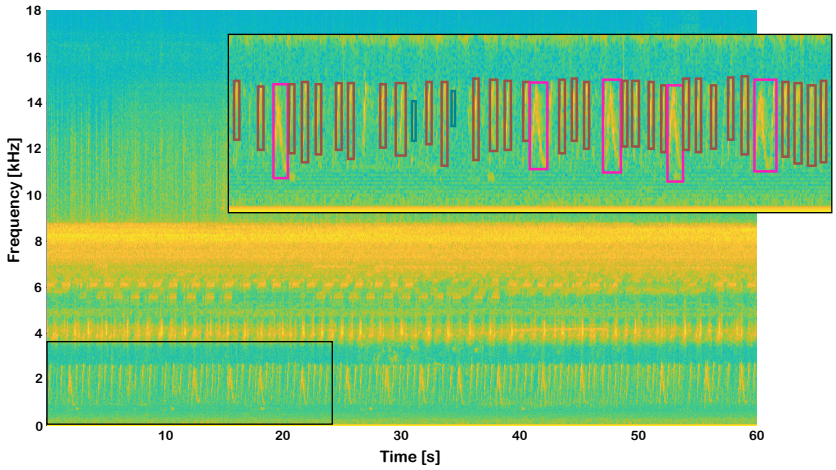

Figura 7.9: Zoom de las llamadas de las especies *Nyctidromus albicollis* y *Leptodactylus fragilis* vocalizando en la misma banda de frecuencia. Aquí es necesario elegir el clúster adecuado y eliminar segmentos ruidosos con el fin de prevenir una posible confusión en el algoritmo.

presente en el sitio. Para ello, se utilizaron 2.317 grabaciones de audio correspondientes a dos puntos geográficos diferentes del lugar de estudio (KLSA13 y KLSA14) registradas durante el mes de marzo de 2021. Los puntos geográficos fueron elegidos debido a la variedad de la biodiversidad presente en el lugar, donde diez especies fueron seleccionadas y etiquetadas por expertos. Se analizaron las grabaciones de cada lugar y la metodología sugiere sonotipos. Luego, se tiene en cuenta el número de sonotipos diferentes en cada grabación y se calcula la media del número de sonotipos por hora, generando de esta forma un patrón acústico de 24 horas. Se usaron los resultados del enfoque propuesto para corroborar si la estimación de la biodiversidad que se hace con esta propuesta, corresponde a lo encontrado con los índices acústicos asociados a la biodiversidad (ACI, BI, NP, SO). Para cada índice acústico y los sonotipos encontrados, se tomó el valor máximo para normalizar cada uno de los valores.

Las figuras 7.10A y 7.10B muestran el patrón de 24 horas para cada índice acústico normalizado y para los sonotipos encontrados con la metodología propuesta para el sitio KLSA13 y KLSA14 respectivamente. Al analizar los audios grabados durante una temporada con el enfoque propuesto, si se cuenta el número de sonotipos encontrados para cada hora (línea negra en la figura 7.10), es posible obtener una tendencia similar a la que proporcionan los índices acústicos, indicando las variaciones de la biofonía a lo largo del día.

En este caso, tal y como se presenta en la figura 7.10, es posible ver que el método propuesto en este trabajo genera estimaciones del componente biofónico de forma similar a los índices acústicos. Sin embargo, este enfoque permite ir un paso mas allá, ya que es posible realizar la identificación automática de animales presentes en cada sitio

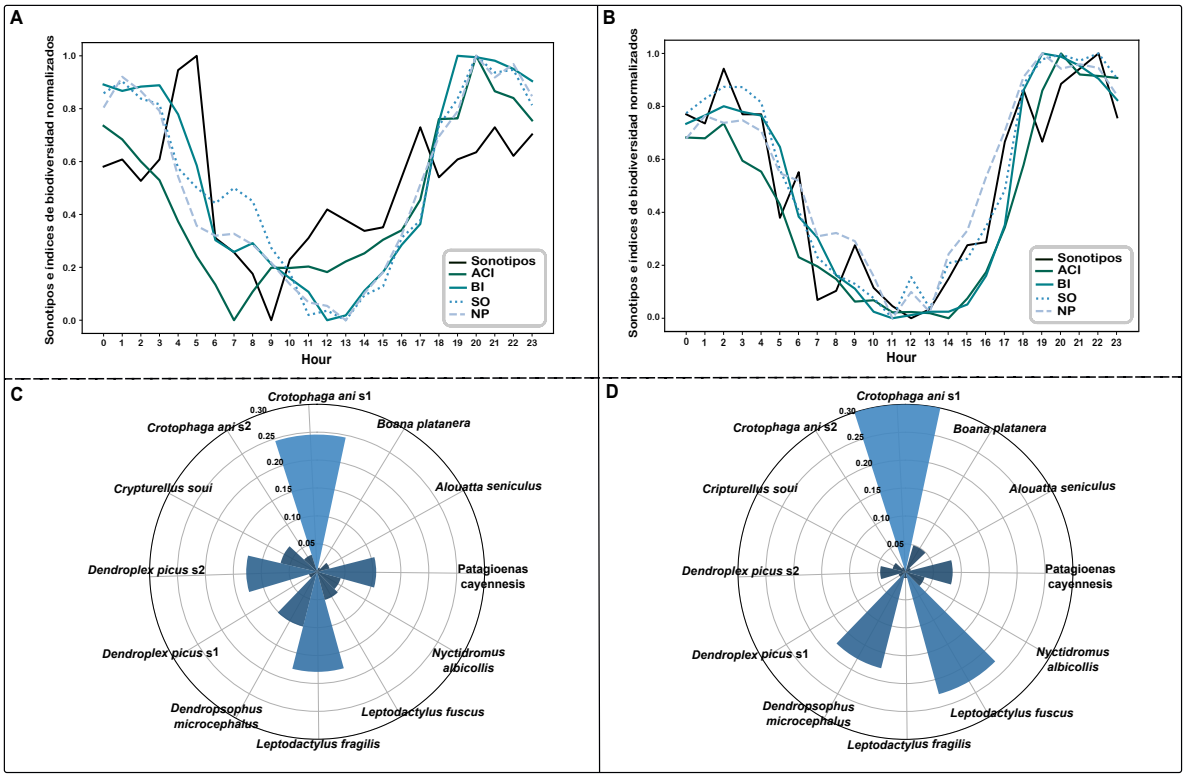

Figura 7.10: Patrón de actividad acústico generado usando el método propuesto para los sitios KLSA13 (**A**) y KLSA14 (**B**) en marzo de 2021 comparado con los cuatro índices acústicos comúnmente utilizados para la estimación de la riqueza de especies. ACI (índice de complejidad acústica), BI (índice bioacústico), NP (número de picos) y SO (ocupación espectral). El eje y corresponde al valor normalizado de cada índice acústico. En el caso de esta propuesta, corresponde al valor normalizado del número medio de sonotipos encontrados para cada hora del día. Figuras **C** y **D** corresponden a la estructura acústica del sitio de acuerdo a las especies identificadas. El porcentaje de detección para cada zona está presente.

y a partir de eso, conocer la estructura acústica del sitio según la especie seleccionada (7.10C-D) sirviendo como indicador de la biodiversidad.

Es evidente que ambos sitios presentan un patrón de biofonía similar (Fig 7.10A-B). Sin embargo, al identificar las especies que hacen parte del paisaje sonoro, es posible obtener la estructura acústica de cada sitio y a partir de esto observar que en este caso, KLSA13 y KLSA14 (Fig. 7.10C-D) presentan una estructura diferente.

Al identificar las especies usando la metodología propuesta, es posible además, obtener el patrón de comportamiento acústico de dichas especies. La figura 7.11 muestra a tendencia de los patrones de comportamiento acústico de las especies analizadas para este caso de estudio en el sitio KLSA13. En esta gráfica se observa la contribución

de la biofonía (por parte de las especies identificadas) durante el día y la noche en la temporada de muestreo.

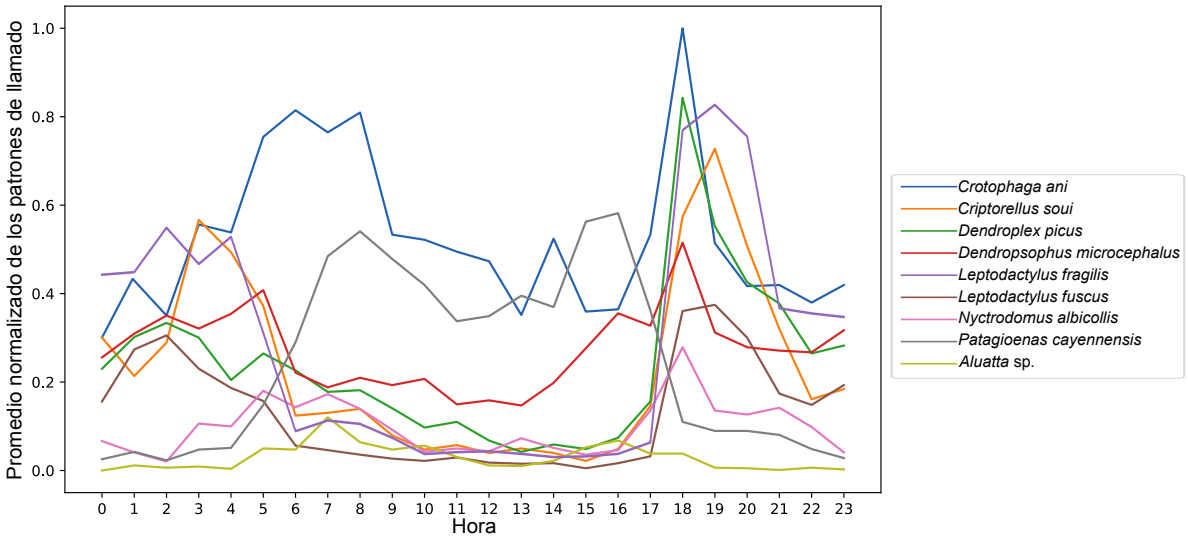

Figura 7.11: Tendencia de los patrones de comportamiento acústico de múltiples especies. Cada línea de color corresponde a las especies identificadas usando la metodología propuesta, estas líneas describen el patrón de comportamiento acústico de cada especie en el sitio KLSA13 durante la temporada de grabación de marzo de 2021.

7.4. Caso 4: Reconocimiento de llamadas de especies en ultrasonido

La metodología propuesta permite hacer la detección de especies en el espectro ultrasónico ($>$ 20 kHz) sin la necesidad de realizar un ajuste de parámetros. Esto fue probado utilizando grabaciones de audio adquiridas con una grabadora que permite capturar datos en ultrasonido. El conjunto de datos incluía 13 especies de murciélagos y 6 especies de ortópteros, todos ellos encontrados en altas frecuencias y etiquetados por expertos en bioacústica.

La tabla 7.4 muestra el desempeño en la detección de cada grupo taxonómico en el espectro ultrasónico y la figura 7.12 muestra el desempeño de cada especie identificada. En este caso, no se hizo ninguna modificación a los parámetros para obtener este rendimiento.

Taxón	Frecuencias	Sensibilidad	Especificidad
Insecta	20 kHz – 30 kHz	96 %	82 %
Mammals	20 kHz - 70 kHz	96 %	84 %

Tabla 7.4: Desempeño de la metodología propuesta en la detección de especies en el espectro ultrasónico.

Figura 7.12: Identificación de especies en espectro ultrasónico de manera no supervisada. Los resultados se organizan de acuerdo al grupo taxonómico de cada especie Insecta, Mammalia). Las barras indican las tasas de aciertos para cada especie en todas las grabaciones.

7.5. Conclusiones

En este capítulo se presenta el enfoque no supervisado propuesto para la metodología multi-especie. Con este enfoque no supervisado fue posible detectar la presencia-ausencia de especies asociadas a diferentes grupos taxonómicos logrando un rendimiento entre el 75-96 %. Este rendimiento también se ha conseguido con metodologías super-

visadas como las redes neuronales convolucionales [9], [11], [14] y otros métodos de aprendizaje automático de tipo supervisado [10], [45], [62] para casos especie-específico en donde se requieren datos etiquetados de la especie objetivo.

En comparación con otras metodologías propuestas para la identificación de especies como librerías de R o software licenciado, se encontraron detecciones notables con la metodología propuesta a pesar de la actividad en múltiples bandas de frecuencia que enmascaraban algunas de las llamadas de las especies. Además, en el enfoque propuesto no es necesario elegir parámetros específicos para cada especie, solo requiere que un experto asocie los clústeres generados por el algoritmo (sonotipo) a una especie.

El método de agrupación LAMDA 3Π no requiere el número de clases como parámetro de entrada. Por lo tanto, el número de clústeres generados por hora dentro del periodo de tiempo analizado, informa sobre la biofonía presente en el paisaje sonoro y es posible además, identificar las especies presentes en el área de estudio. Este algoritmo de agrupación puede diferenciar entre las llamadas de las especies y permite encontrar sonotipos de especies que no se esperaban a priori ya que crea clústeres de forma natural a partir de las diferencias encontradas entre los segmentos del paisaje sonoro en las frecuencias audibles y ultrasónicas.

Al comparar la estimación de la biofonía de los índices acústicos con la aplicación de la metodología propuesta, se evidencia que los clústeres propuestos (sonotipos) responden a la biofonía del sitio de estudio. Esto muestra que la etapa de segmentación si da respuesta a la vocalización de los animales y las características extraídas permiten diferenciar entre especies. Con la metodología propuesta en este trabajo no solo se puede identificar la tendencia de la biofonía sino también los patrones de actividad acústica por especie.

Con este capítulo se cumple con los objetivos 2, 3 y 4 “Analizar técnicas de aprendizaje no supervisado para identificar cual es el modelo que más se ajusta a la clasificación de las especies. Diseñar una metodología para identificar los patrones de actividad acústica de múltiples especies. Validar la metodología propuesta a través de pruebas con un número alto de grabaciones de bosque seco tropical y bosque húmedo colombiano”

Capítulo 8

Conclusiones

En este trabajo de investigación se presenta una metodología no supervisada para la identificación de vocalizaciones de especies de animales en paisajes sonoros tropicales utilizando un segmentador basado en análisis de imagen, coeficientes cepstrales escalados linealmente e información de frecuencia como características y un algoritmo de agrupamiento con el que se obtuvo resultados de identificación de especies pertenecientes a diferentes grupos taxonómicos con un alto desempeño. Esta propuesta permite la detección de especies en audios con biofonía presente en múltiples bandas de frecuencia, en altas frecuencias y llamadas con baja intensidad en espectrogramas sin ruido de fondo significativo. Con este enfoque es posible analizar datos registrados en diferentes paisajes y con diferentes tipos de grabadoras.

Es posible utilizar los resultados de esta metodología para realizar evaluaciones de biodiversidad similar a los índices acústicos con la ventaja de poder identificar vocalizaciones de especies de aves, anuros, murciélagos, insectos y primates presentes en múltiples bandas de frecuencia en el paisaje sonoro y obtener la contribución de cada especie a la estructura acústica de cada sitio analizado. También se pueden obtener los patrones de actividad acústica durante una temporada de grabación de las especies.

La metodología propuesta es relativamente robusta al ruido, pero en algunos casos, este puede afectar los resultados de la agrupación. Entre sus limitaciones se encuentra que requiere grabaciones donde se cuente con la presencia de segmentos (bien definidos) de la especie para formar clústeres confiables. Las diferencias en las llamadas de las especies debido al ruido de fondo u otras condiciones del entorno que afectan la intensidad de las llamadas en el espectrograma, pueden generar un número alto de clústeres que reflejan la variabilidad que se presenta en estos casos. Esta limitación puede resolverse utilizando índices de validación de clústeres o las propiedades de lógica difusa del algoritmo LAMDA, el cual es uno de los trabajos futuros.

8.1. Publicaciones

Los resultados presentes en este documento se encuentran publicados en un artículo que fue aceptado en la revista Methods in Ecology and Evolution para publicación como resultado de este trabajo de investigación.

Guerrero, M. J., Bedoya, C. L., López, J. D., Daza, J. M., Isaza, C. (2023). Acoustic animal identification using unsupervised learning. Methods in Ecology and Evolution, 00, 1–15. https://doi.org/10.1111/2041210X.14103

Además, fue presentado en la *17th Ibero-American Conference on Artificial Intelligence* - IBERAMIA 2022, cuya publicación se encuentra en Springer-Lectures in computer Science como proceedings.

Guerrero, M. J., Restrepo, J., Nieto-Mora, D.A., Daza, J.M., Isaza, C. (2022). Insights from Deep Learning in Feature Extraction for Non-supervised Multi-species Identification in Soundscapes. In: Advances in Artificial Intelligence – IBERAMIA 2022. Lecture Notes in Computer Science, vol 13788. Springer, Cham. https://doi.org/10.1007/978-3-031-22419-5_19

Bibliografía

[1] S. L. Pimm, S. Alibhai, R. Bergl et al., «Emerging Technologies to Conserve Biodiversity,» *Trends in Ecology and Evolution*, vol. 30, n.º 11, págs. 685-696, 2015, ISSN: 01695347. DOI: 10.1016/j.tree.2015.08.008. dirección: http://dx.doi.org/10.1016/j.tree.2015.08.008.

[2] S. L. Dumyahn y B. C. Pijanowski, «Soundscape conservation,» *Landscape Ecology*, vol. 26, n.º 9, págs. 1327-1344, 2011, ISSN: 09212973. DOI: 10.1007/s10980-011-9635-x.

[3] J. Sueur y A. Farina, «Ecoacoustics: the Ecological Investigation and Interpretation of Environmental Sound,» *Biosemiotics*, vol. 8, n.º 3, págs. 493-502, 2015, ISSN: 18751350. DOI: 10.1007/s12304-015-9248-x.

[4] T. M. Aide, A. Hern y M. Campos-cerqueira, «Species Richness (of Insects) Drives the Use of Acoustic Space in the Tropics,» *Remote Sensing in Ecology and Conservation*, págs. 1-12, 2017. DOI: 10.3390/rs9111096.

[5] J. S. Ulloa, T. Aubin, D. Llusia, C. Bouveyron y J. Sueur, «Estimating animal acoustic diversity in tropical environments using unsupervised multiresolution analysis,» *Ecological Indicators*, vol. 90, n.º March, págs. 346-355, 2018, ISSN: 1470160X. DOI: 10.1016/j.ecolind.2018.03.026.

[6] C. L. Bedoya y L. E. Molles, «Acoustic censusing and individual identification of birds in the wild,» n.º 19, 2021.

[7] D. A. Yip, C. L. Mahon, A. G. MacPhail y E. M. Bayne, «Automated classification of avian vocal activity using acoustic indices in regional and heterogeneous datasets,» *Methods in Ecology and Evolution*, vol. 2021, n.º December 2020, págs. 1-13, 2021, ISSN: 2041210X. DOI: 10.1111/2041-210X.13548.

[8] S. R. Ross, N. R. Friedman, K. L. Dudley, M. Yoshimura, T. Yoshida y E. P. Economo, «Listening to ecosystems: data-rich acoustic monitoring through landscape-scale sensor networks,» *Ecological Research*, vol. 33, n.º 1, págs. 135-147, 2018, ISSN: 14401703. DOI: 10.1007/s11284-017-1509-5.

[9] Z. J. Ruff, D. B. Lesmeister, L. S. Duchac, B. K. Padmaraju y C. M. Sullivan, «Automated identification of avian vocalizations with deep convolutional neural networks,» *Remote Sensing in Ecology and Conservation*, vol. 6, n.º 1, págs. 79-92, 2020, ISSN: 20563485. DOI: 10.1002/rse2.125.

[10] J. Xie, K. Indraswari, L. Schwarzkopf, M. Towsey, J. Zhang y P. Roe, «Acoustic classification of frog within-species and species-specific calls,» *Applied Acoustics*, vol. 131, n.º April 2017, págs. 79-86, 2018, ISSN: 1872910X. DOI: 10.1016/j.apacoust.2017.10.024.

[11] J. LeBien, M. Zhong, M. Campos-Cerqueira et al., «A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network,» *Ecological Informatics*, vol. 59, n.º April, pág. 101 113, 2020, ISSN: 15749541. DOI: 10.1016/j.ecoinf.2020.101113. dirección: https://doi.org/10.1016/j.ecoinf.2020.101113.

[12] J. Xie, M. Towsey, J. Zhang y P. Roe, «Image processing and classification procedure for the analysis of Australian frog vocalisations,» *EMR 2015 - Proceedings of the 2015 ACM International Workshop on Environmental Multimedia Retrieval*, n.º June, págs. 15-20, 2015. DOI: 10.1145/2764873.2764878.

[13] M. A. Roch, M. S. Soldevilla, J. C. Burtenshaw, E. E. Henderson y J. A. Hildebrand, «Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California,» *The Journal of the Acoustical Society of America*, vol. 121, n.º 3, págs. 1737-1748, 2007, ISSN: 0001-4966. DOI: 10.1121/1.2400663.

[14] Z. J. Ruff, D. B. Lesmeister, C. L. Appel y C. M. Sullivan, «Workflow and convolutional neural network for automated identification of animal sounds,» *Ecological Indicators*, vol. 124, n.º July 2020, pág. 107 419, 2021, ISSN: 1470160X. DOI: 10.1016/j.ecolind.2021.107419. dirección: https://doi.org/10.1016/j.ecolind.2021.107419.

[15] M. Premoli, D. Baggi, M. Bianchetti et al., «Automatic classification of mice vocalizations using Machine Learning techniques and Convolutional Neural Networks,» *PLoS ONE*, vol. 16, n.º 1 January, págs. 1-16, 2021, ISSN: 19326203. DOI: 10.1371/journal.pone.0244636. dirección: http://dx.doi.org/10.1371/journal.pone.0244636.

[16] O. Ovaskainen, U. Moliterno de Camargo y P. Somervuo, «Animal Sound Identifier (ASI): software for automated identification of vocal animals,» *Ecology Letters*, vol. 21, n.º 8, págs. 1244-1254, 2018, ISSN: 14610248. DOI: 10.1111/ele.13092.

[17] H. Gan, J. Zhang, M. Towsey et al., «Data selection in frog chorusing recognition with acoustic indices,» *Ecological Informatics*, vol. 60, pág. 101 160, 2020, ISSN: 15749541. DOI: 10.1016/j.ecoinf.2020.101160. dirección: https://doi.org/10.1016/j.ecoinf.2020.101160.

[18] I. Potamitis, «Unsupervised dictionary extraction of bird vocalisations and new tools on assessing and visualising bird activity,» *Ecological Informatics*, vol. 26, n.º P3, págs. 6-17, 2015, ISSN: 15749541. DOI: 10.1016/j.ecoinf.2015.01.002. dirección: http://dx.doi.org/10.1016/j.ecoinf.2015.01.002.

[19] J. Xie, K. Hu, M. Zhu e Y. Guo, «Bioacoustic signal classification in continuous recordings: Syllable-segmentation vs sliding-window,» *Expert Systems with Applications*, vol. 152, pág. 113 390, 2020, ISSN: 09574174. DOI: 10.1016/j.eswa.2020.113390. dirección: https://doi.org/10.1016/j.eswa.2020.113390.

[20] D. Stowell, «Computational bioacoustics with deep learning: a review and roadmap,» *PeerJ*, vol. 10, e13152, 2022, ISSN: 21678359. DOI: 10.7717/peerj.13152.

[21] D. Stowell, «Computational Bioacoustic Scene Analysis,» en *Computational Analysis of Sound Scenes and Events*, 2018, págs. 303-333, ISBN: 9783319634500. DOI: 10.1007/978-3-319-63450-0.

[22] M. Towsey, J. Wimmer, I. Williamson y P. Roe, «The Use of Acoustic Indices to Determine Avian Species Richness in Audio-recordings of the Environment,» *Ecological Informatics*, vol. 21, págs. 110-119, 2013. DOI: 10.1016/j.ecoinf.2013.11.007.

[23] M. Depraetere, S. Pavoine, F. Jiguet, A. Gasc, S. Duvail y J. Sueur, «Monitoring animal diversity using acoustic indices: Implementation in a temperate woodland,» *Ecological Indicators*, vol. 13, págs. 46-54, 2012. DOI: 10.1016/j.ecolind.2011.05.006.

[24] C. Sánchez-Giraldo, C. L. Bedoya, R. A. Morán-Vásquez, C. V. Isaza y J. M. Daza, «Ecoacoustics in the rain: understanding acoustic indices under the most common geophonic source in tropical rainforests,» *Remote Sensing in Ecology and Conservation*, vol. 6, n.º 3, págs. 248-261, 2020, ISSN: 20563485. DOI: 10.1002/rse2.162.

[25] A. Restrepo, C. Molina-Zuluaga, J. P. Hurtado, C. M. Marín y J. M. Daza, «Amphibians and reptiles from two localities in the northern Andes of Colombia,» *Check List*, vol. 13, págs. 203-237, 2017. DOI: 10.15560/13.4.203.

[26] M. Araya-Salas y G. Smith-Vidaurre, *Methods Ecol Evol - 2016 - Araya-Salas - warbleR an r package to streamline analysis of animal acoustic signals.pdf*, 2017.

[27] J. Katz, S. D. Hafner y T. Donovan, «Tools for automated acoustic monitoring within the R package monitoR,» *Bioacoustics*, vol. 25, n.º 2, págs. 197-210, 2016, ISSN: 21650586. DOI: 10.1080/09524622.2016.1138415. dirección: http://dx.doi.org/10.1080/09524622.2016.1138415.

[28] A. Wildlife, «Kaleidoscope Pro 5 User Guide,» *Wildlife Acoustics, Inc.: Maynard, MA, USA*, 2020.

[29] T. Ullmann, C. Hennig y A. L. Boulesteix, «Validation of cluster analysis results on validation data: A systematic framework,» *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, n.º November 2021, págs. 1-19, 2021, ISSN: 19424795. DOI: 10.1002/widm.1444.

[30] N. Pieretti, A. Farina y D. Morri, «A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI),» *Ecological Indicators*, vol. 11, págs. 868-873, 2010. DOI: 10.1016/j.ecolind.2010.11.005.

[31] N. Boelman, G. Asner, P. Hart y R. Martin, «Multi-trophic invasion resistance in Hawaii: Bioacoustics, field surveys, and airborne remote sensing,» *Ecological applications : a publication of the Ecological Society of America*, vol. 17, 2008. DOI: 10.1890/07-0004.1.

[32] A. Gasc, J. Sueur, S. Pavoine, R. Pellens y P. Grandcolas, «Biodiversity Sampling Using a Global Acoustic Approach: Contrasting Sites with Microendemics in New Caledonia,» *PLoS ONE*, vol. 8, 2013. DOI: 10.1371/journal.pone.0065311.

[33] E. C. Rojas, C. S. Giraldo, C. Bedoya y J. M. D. Rojas, «Habitat and acoustic spectrum as determinant factors of the occupation of neotropical anurans,» *Biota Colombiana*, vol. 23, n.º 1, 2022, ISSN: 2539200X. DOI: 10.21068/2539200X.910.

[34] J. Xue, Z. Feng y P. Zhang, «Spectrum Occupancy Measurements and Analysis in Beijing,» *IERI Procedia*, vol. 4, págs. 295-302, 2013, ISSN: 22126678. DOI: 10.1016/j.ieri.2013.11.042. dirección: http://dx.doi.org/10.1016/j.ieri.2013.11.042.

[35] C. Bedoya, C. Isaza, J. M. Daza y J. D. López, «Automatic identification of rainfall in acoustic recordings,» *Ecological Indicators*, vol. 75, págs. 95-100, 2017, ISSN: 1470-160X. DOI: 10.1016/j.ecolind.2016.12.018. dirección: http://dx.doi.org/10.1016/j.ecolind.2016.12.018.

[36] J. Xie, J. G. Colonna y J. Zhang, «Bioacoustic signal denoising: a review,» *Artificial Intelligence Review*, vol. 54, n.º 5, págs. 3575-3597, 2021, ISSN: 15737462. DOI: 10.1007/s10462-020-09932-4. dirección: https://doi.org/10.1007/s10462-020-09932-4.

[37] J. Xie, M. Towsey, M. Zhu, J. Zhang y P. Roe, «An intelligent system for estimating frog community calling activity and species richness,» *Ecological Indicators*, vol. 82, n.º November 2016, págs. 13-22, 2017, ISSN: 1470160X. DOI: 10.1016/j.ecolind.2017.06.015. dirección: http://dx.doi.org/10.1016/j.ecolind.2017.06.015.

[38] R. R. Kvsn, J. Montgomery, S. Garg y M. Charleston, «Bioacoustics Data Analysis-A Taxonomy, Survey and Open Challenges,» *IEEE Access*, vol. 8, págs. 57 684-57 708, 2020, ISSN: 21693536. DOI: 10.1109/ACCESS.2020.2978547.

[39] M. Ducrettet, P.-M. Forget, J. Ulloa et al., «Acoustic monitoring of the White-throated Toucan (Ramphastos tucanus) in disturbed tropical landscapes,» *Biological Conservation*, vol. 245, 2020.

[40] T. M. Aide, C. Corrada Bravo, M. Campos Cerqueira, C. Milan, G. Vega y R. Alvarez, «Real-time bioacoustics monitoring and automated species identification,» *PeerJ*, vol. 1, e103, jul. de 2013. DOI: 10.7717/peerj.103.

[41] C. Bedoya, C. Isaza, J. Daza y J. D. López, «Automatic recognition of anuran species based on syllable identification,» *Ecological Informatics*, vol. 24, págs. 200-209, 2014.

[42] J. Xie, T. Michael, J. Zhang y P. Roe, «Detecting frog calling activity based on acoustic event detection and multi-label learning,» *Procedia Computer Science*, vol. 80, págs. 627-638, 2016, ISSN: 18770509. DOI: 10.1016/j.procs.2016.05.352. dirección: http://dx.doi.org/10.1016/j.procs.2016.05.352.

[43] N. Otsu, «A Threshold Selection Method from Gray-Level Histograms,» *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, págs. 62-66, 1979.

[44] I. Potamitis, S. Ntalampiras, O. Jahn y K. Riede, «Automatic bird sound detection in long real-field recordings: Applications and tools,» *Applied Acoustics*, vol. 80, págs. 1-9, 2014, ISSN: 0003682X. DOI: 10.1016/j.apacoust.2014.01.001. dirección: http://dx.doi.org/10.1016/j.apacoust.2014.01.001.

[45] C. M. T.-G. Juan J. Noda David Sánchez-Rodríguez, «We are IntechOpen , the world ' s leading publisher of Open Access books Built by scientists , for scientists TOP 1 %,» *Intech*, vol. 32, n.º tourism, págs. 137-144, 2018, ISSN: 00664804. dirección: https://www.intechopen.com/books/advanced-biometric-technologies/liveness-detection-in-biometrics.

[46] G. Sharma, K. Umapathy y S. Krishnan, «Trends in audio signal feature extraction methods,» *Applied Acoustics*, vol. 158, pág. 107 020, ene. de 2020. DOI: 10.1016/j.apacoust.2019.107020.

[47] S. M. Nirosha Priyadarshani e I. Castro, «Automated birdsong recognition in complex acoustic environments: a review,» *Avian Biology*, 2018. DOI: 10.1111/jav.01447.

[48] B. Rowe, P. Eichinski, J. Zhang y P. Roe, «Acoustic auto-encoders for biodiversity assessment,» *Ecological Informatics*, vol. 62, n.º January, pág. 101 237, 2021, ISSN: 15749541. DOI: 10.1016/j.ecoinf.2021.101237. dirección: https://doi.org/10.1016/j.ecoinf.2021.101237.

[49] S. Ntalampiras e I. Potamitis, «Acoustic detection of unknown bird species and individuals,» *CAAI Transactions on Intelligence Technology*, n.º October 2020, 2021, ISSN: 24682322. DOI: 10.1049/cit2.12007.

[50] J. Xie, K. hu, Y. Guo, Q. Zhu y J. Yu, «On loss functions and CNNs for improved bioacoustic signal classification,» *Ecological Informatics*, vol. 64, pág. 101 331, mayo de 2021. DOI: 10.1016/j.ecoinf.2021.101331.

[51] C. Dong, T. Xue y C. Wang, «The feature representation ability of variational autoencoder,» *Proceedings - 2018 IEEE 3rd International Conference on Data Science in Cyberspace, DSC 2018*, págs. 680-684, 2018. DOI: 10.1109/DSC.2018.00108.

[52] T. Fukumoto, *Anomaly detection using Variational Autoencoder (VAE)*. dirección: https://github.com/mathworks/Anomaly-detection-using-Variational-Autoencoder-VAE-/releases/tag/1.0.1.

[53] K. He, X. Zhang, S. Ren y J. Sun, «Deep Residual Learning for Image Recognition,» jun. de 2016, págs. 770-778. DOI: 10.1109/CVPR.2016.90.

[54] K. L. Y. C. for Conservation Bioacoustics at the Cornell Lab of Ornithology., *Raven Pro: Interactive Sound Analysis Software (Version 1.6.3) [Computer software]*. dirección: `https://ravensoundsoftware.com/.`.

[55] B. Scheffers, L. Joppa, S. Pimm y W. Laurance, «What we know and don't know about Earth's missing biodiversity,» *Trends in ecology evolution*, vol. 27, págs. 501-510, 2012. DOI: `10.1016/j.tree.2012.05.008`.

[56] X. Giam, B. Scheffers, N. Sodhi, D. Wilcove, G. Ceballos y P. Ehrlich, «Reservoirs of richness: Least disturbed tropical forests are centres of undescribed species diversity,» *Proceedings. Biological sciences / The Royal Society*, vol. 279, págs. 67-76, 2011. DOI: `10.1098/rspb.2011.0433`.

[57] L. Joppa, D. Roberts, N. Myers y S. Pimm, «Biodiversity hotspots House most undiscovered plant species,» *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, 2011. DOI: `10.1073/pnas.1109389108`.

[58] G. Brehm, K. Fiedler, C. Häuser y H. Dalitz, «Methodological Challenges of a Megadiverse Ecosystem,» vol. 198, 2008. DOI: `10.1007/978-3-540-73526-7_5`.

[59] M. Acconcjaioco y S. Ntalampiras, «One-shot learning for acoustic identification of bird species in non-stationary environments,» *Proceedings - International Conference on Pattern Recognition*, págs. 755-762, 2020, ISSN: 10514651. DOI: `10.1109/ICPR48806.2021.9412005`.

[60] J. Aguilar-Martin y R. López-de-Mantarás, «The process of classification and learning the meaning of linguistic descriptors or concepts,» *Aproximate Reasoning in Desicion Analysis*, págs. 165-175, 1982.

[61] C. Bedoya, J. Waissman y C. Isaza, «Yager– Rybalov Triple pi Operator as a Means of Reducing the Number of Generated Clusters in Unsupervised Anuran Vocalization Recognition,» en *Nature-Inspired Computation and Machine Learning*, ép. Lecture Notes in Computer Science, vol. 8857, Springer International Publishing, 2014, págs. 382-391.

[62] S. Brodie, S. Allen-Ankins, M. Towsey, P. Roe y L. Schwarzkopf, «Automated species identification of frog choruses in environmental recordings using acoustic indices,» *Ecological Indicators*, vol. 119, n.º August, pág. 106 852, 2020, ISSN: 1470160X. DOI: `10.1016/j.ecolind.2020.106852`. dirección: `https://doi.org/10.1016/j.ecolind.2020.106852`.