



**UNIVERSIDAD
DE ANTIOQUIA**

Reconocimiento automático de llamados de murciélagos en estado libre usando técnicas de machine learning

Autor(es)

Frank Sebastián Orozco Gallo

Kelly Johana Zapata Estrada

Universidad de Antioquia

Facultad de Ingeniería, Departamento de Ingeniería Electrónica

Medellín, Colombia

2021



Reconocimiento automático de llamados de murciélagos en estado libre usando técnicas de machine learning

Frank Sebastián Orozco Gallo

Kelly Johana Zapata Estrada

Tesis o trabajo de investigación presentada(o) como requisito parcial para optar al título de:

Ingeniero Electrónico

MEDELLIN

Asesora

Claudia Victoria Isaza Narváez, PhD

Profesora Vinculada Universidad de Antioquia

Universidad de Antioquia

Facultad de Ingeniería, Departamento de Ingeniería Electrónica

Medellín, Colombia

2021.

Resumen

En este trabajo se desarrolló una herramienta que extrae y procesa características acústicas de archivos de extensión zero-crossing o wav. La herramienta hace uso de metodologías de *machine learning* y permite el reconocimiento automático de los géneros de murciélagos *molosus*(Sonotipo 1) y *myotis*(Sonotipo 2) disminuyendo así el tiempo de procesamiento de datos al automatizar el proceso.

La herramienta desarrollada para el análisis de los 2 tipos de archivos se materializa en 2 funciones desarrolladas en el software R. Estas funciones realizan el análisis de las características extraídas de los audios mediante el uso de redes neuronales con el algoritmo de back propagation.

Estas funciones se generaron de forma que tomen todos los audios de un tipo (wav o zc) en un directorio y se le realice el análisis completo a todos los audios. El resultado generado es un archivo csv con la descripción de las características halladas y en que tiempos del audio se detectaron los sonotipos.

Palabras Claves: Clasificación, murciélagos, monitoreo acústico, zero-crossing, wav, molosus.

Contenido

Lista de figuras.....	3
Lista de tablas.....	3
Introducción.....	4
Objetivos.....	5
Objetivo General.....	5
Objetivos Específicos.....	5
Marco teórico.....	6
Redes neuronales artificiales.....	10
Materiales y Métodos.....	11
Base de datos.....	11
Metodología propuesta.....	12
Preprocesamiento.....	12
Segmentación.....	13
Archivos ZC.....	13
Archivos WAV.....	14
Extracción de Características.....	15
Consideración adicional para archivos de audio en formato ZC.....	17
Clasificación.....	19
Software implementado.....	23
Resultados y análisis.....	24
Entrenamiento.....	25
Prueba.....	27
Archivos ZC.....	27
Archivos WAV.....	29
Conclusiones.....	30
Trabajo a Futuro.....	31
Referencias.....	31

Lista de figuras

Figura 1. Etapas implementadas para la detección de sonotipos e identificación de ruido.	12
Figura 2. Ejemplo de una señal contenida en un archivo ZC.....	13
Figura 3. Ejemplo de una señal contenida en un archivo .WAV.....	14
Figura 4. Características de un llamado observadas en Anabat Insight.....	15
Figura 5. Características extraídas listadas en tabla.	16
Figura 6. Boxplot de la característica Frecuencia mínima.....	16
Figura 7. Boxplot frecuencia promedio.....	17
Figura 8. Boxplot frecuencia final.	17
Figura 9. Pendiente promedio calculada en tres partes por segmento para sonotipo 1. (a) En rojo ruido (b) En azul sonotipo 1.	18
Figura 10. Delta de la frecuencia promedio calculado en tres partes por segmento para sonotipo 2. (a) En rojo ruido (b) En azul sonotipo 2.....	19
Figura 11. Diagrama de bloques del proceso de identificación de un archivo ZC.....	19
Figura 12. Diagrama de bloques del proceso de identificación para archivos ZC.	20
Figura 13. Imagen de la red neuronal 1 usada para archivos WAV y ZC.....	21
Figura 14. RNA para diferenciar entre ruido y sonotipo 1 en archivos ZC.....	23
Figura 15. RNA para diferenciar entre ruido y sonotipo 2 en archivos ZC.....	23
Figura 16. Captura del programa para identificación de sonotipos para archivos. ZC.....	24

Lista de tablas

Tabla 1. Lista de grabaciones tratadas en el proyecto.....	11
Tabla 2. Tabla de características diferenciadoras para las redes neuronales 2 y 3.....	22
Tabla 3. Matriz de confusión de la primera red neuronal en etapa de entrenamiento para archivos ZC.....	25
Tabla 4. Matriz de confusión de la segunda red neuronal en etapa de entrenamiento para archivos ZC.....	26
Tabla 5. Matriz de confusión de la tercera red neuronal en etapa de entrenamiento para archivos ZC.....	27
Tabla 6. Matriz de confusión de la primera red neuronal en etapa de prueba archivos ZC.....	27
Tabla 7. Matriz de confusión de la segunda red neuronal en etapa de prueba para archivos ZC. ..	28
Tabla 8. Matriz de confusión de la tercera red neuronal en etapa de prueba para archivos ZC.	28
Tabla 9. Matriz de confusión total del sistema para audios en formato ZC.....	29
Tabla 10. Matriz de confusión del sistema total para el análisis de archivos WAV.....	29

Introducción

Con el fin de identificar el estado de ecosistemas, es usual monitorear el comportamiento de los murciélagos pues son indicadores de cambios en el ambiente. Para este monitoreo se hacen estudios del ultrasonido emitido por los murciélagos. Usando grabadoras como Anabat UltraMic200K y Echo Meter EM3, se puede realizar monitoreos en el espacio abierto y natural. Este tipo de monitoreo se llama monitoreo pasivo debido a que no cuenta con la interacción del hombre. El monitoreo pasivo deja como resultado una gran cantidad de grabaciones que deben ser analizadas (por lo general en formato zero-crossing o wav).

Para el análisis de los registros de audio pueden utilizarse los softwares propios de la marca de las grabadoras que permiten el estudio de algunas características como la duración de los chillidos, frecuencias fundamentales. Sin embargo, no permiten hacer una identificación automática del género de murciélagos. Esta identificación se realiza de forma manual.

En este proyecto, se propuso una herramienta que permite identificar automáticamente llamados de caza y alimentación de murciélagos analizando segmentos de audio de archivos de extensión zero-crossing o wav. Se asume que cada archivo contiene segmentos de audio relevantes. La herramienta hace uso de metodologías de *machine learning* y permite el reconocimiento automático de los géneros de murciélagos *molosus* y *myotis* disminuyendo así el tiempo de procesamiento de datos al automatizar el proceso.

Objetivos

Objetivo General

Proponer una metodología que utilice técnicas de *machine learning* y permita el reconocimiento automático de dos sonotipos asociados a 2 géneros de murciélagos a partir del análisis de una señal de audio proveniente de un archivo .wav o zero-crossing.

Objetivos Específicos

- Analizar algoritmos de reducción de ruido, segmentación, clasificación y extracción de características acústicas que funcionan correctamente en el ambiente no controlado en el que se realizan las grabaciones (hábitat natural de los murciélagos).
- Diseñar una metodología basada en algoritmos de clasificación que permita reconocer los 2 sonotipos que se diferenciarán en las diferentes grabaciones.
- Implementar la metodología propuesta en un aplicativo que permita identificar (en grabaciones que no fueron usadas en el entrenamiento) el género del murciélago que produjo el sonido.
- Realizar la validación del algoritmo entrenado con grabaciones que no se hayan usado durante el entrenamiento o pruebas y realizar el análisis de margen de error.

Marco teórico

Los murciélagos dependen de ondas de ultrasonido que emiten para sus desplazamientos, ubicar una presa, cazar, identificar objetos u otros animales. Por esta razón es natural pensar que su estudio debe hacerse con análisis acústico. Gracias al avance tecnológico, ha sido posible realizar estudios de las frecuencias que los murciélagos utilizan. Esto ha popularizado su grabación y estudio a partir de equipos de grabación como Anabat [35].

Para el estudio del comportamiento de los murciélagos se han implementado varios experimentos e investigaciones. Estas van desde mediciones de energía gastada en vuelo con ecolocalización y sin ecolocalización; [8] hasta estudios morfológicos sobre la posibilidad de que murciélagos de diferentes tamaños puedan habitar o cazar presas en medios densos [9].

Para realizar la identificación de los diferentes géneros y especies se han tomado las siguientes definiciones [10]:

- Un llamado (en inglés call) se define como un pulso vocal individual y discreto.
- Cada llamado tiene como características:
 - Un rango de frecuencia (máximo y mínimo)
 - Una duración (tiempo en milisegundos desde el inicio hasta el final)
 - Una forma (variando desde ligeramente curvilínea a distintivamente bilineal).
- Una secuencia se compone de una serie de llamados consecutivos producidos por un individuo.

En el pasado se han utilizado diferentes tipos de criterios para el reconocimiento de especies o géneros a partir del llamado que generan. El criterio más aceptado se trata del criterio cualitativo [10]. Para este criterio se utilizan ciertas características estructurales de los llamados del murciélago; principalmente las frecuencias máxima y mínima aproximadas, más los aspectos morfológicos de los llamados (linealidad y cambios en la pendiente).

Debido a la similitud que existe en las características cuantitativas para murciélagos de diferentes tamaños, el criterio cualitativo cobra relevancia. Estas similitudes se ven en características como la frecuencia mínima, máxima o intervalos entre llamados. Por esta razón se usa el criterio cualitativo en la forma de los llamados (unos más "elongados" que otros, o con una frecuencia máxima diferente) [10].

Para hacer uso de estos criterios, se extraen las características de los llamados desde las grabaciones adquiridas. Estas grabaciones se obtienen de 2 formas, capturando el espécimen y grabando sus actividades, o en un monitoreo pasivo

en el entorno natural. Como suele suceder con estos estudios existen pros y contras.

En cautiverio, un espécimen altera su comportamiento y sus patrones de vuelo; evitando las interacciones comunes de los especímenes con su entorno. Sin embargo, estas grabaciones permiten controlar los factores estudiados reduciendo la incertidumbre.

Por otro lado, utilizando grabaciones obtenidas en el ambiente natural se hace el análisis del espécimen con el ambiente que lo rodea y con los demás especímenes. Esto implica un nivel de incertidumbre en el estudio realizado, ya que la información obtenida no está apropiadamente catalogada. Debido a esto, es necesario realizar una previa identificación de la información pertinente al estudio en las grabaciones.

Para identificar la información relevante, se requiere que un experto en identificación de sonidos de murciélagos identifique la forma de un llamado. Tal labor se realiza mediante el reconocimiento de las formas (sonotipo) representadas en el espectrograma de la señal de audio; donde se visualiza en 3 dimensiones los valores de frecuencia, tiempo y potencia. La representación gráfica que se define como sonotipo se ha asociado a cada género de murciélago.

Estos sonotipos se deben analizar desde las grabaciones de forma individual para poder ser catalogados. Este proceso se realiza de forma manual con el uso de herramientas como Anlook, diseñadas por los fabricantes de herramientas de detección [35]. En consecuencia, una vez analizadas las características de los llamados se puede reconocer a que sonotipo pertenece el llamado particular. Sin embargo, hacer este análisis manual para cada grabación no permite implementar planes de monitoreo en largos periodos de tiempo pues se requiere analizar una gran cantidad de grabaciones.

Teniendo en cuenta las posibles horas de grabación, se requieren herramientas de reconocimiento automático. Dichas herramientas facilitarían el estudio de los registros evitando la labor de catalogar a que sonotipo pertenece cada llamado particular. Esto haría que el objeto de estudio comportamental se enfoque en el estudio más que en la catalogación.

El reconocimiento automático de sonidos animales es una técnica alterna a los métodos tradicionales de reconocimiento [1]. Esta técnica es además menos intrusiva al no requerir sacrificar, manipular o aproximarse al animal. De este modo se evita alterar los comportamientos y hábitats de los individuos; facilitando así responder preguntas relacionadas con el comportamiento de animales en ecosistemas específicos. Varias metodologías para el reconocimiento de sonidos en animales han sido propuestas previamente [2,3,4,5,6,7]. Estas metodologías

consisten en caracterizar el sonido del animal para luego clasificar su género o su especie. A este tipo de metodologías hacen parte de la bio-acústica.

La bio-acústica ha enfocado su trabajo en la descripción y análisis de las vocalizaciones de animales. El desarrollo de metodologías para la automatización del reconocimiento de especies a través de llamados ha sido un territorio poco explorado y menos para el caso de murciélagos. Los trabajos van orientados hacia la búsqueda de espacios acústicos y de reconocimiento [11], descripción de llamados [12], o dependencia de la vocalización con respecto a las variaciones morfológicas o geográficas [13].

El reconocimiento automático de especies basado en parámetros acústicos de los llamados de animales es un desafío. Principalmente debido a las condiciones ambientales en que se adquieren los datos y a la similitud de sonidos entre los diferentes géneros y especies. El interés en esta área ha aumentado en la última década y recientemente se han propuesto diferentes alternativas de solución.

La mayoría de las propuestas se basan en metodologías de reconocimiento del habla en humanos, aplicadas al reconocimiento de especies [2]. El procedimiento regularmente usado es el siguiente:

- a) Los llamados de la respectiva especie son grabados en su hábitat natural.
- b) Sus características acústicas son extraídas.
- c) Cada llamado es transformado en un conjunto de característica que representa los rasgos del mismo.
- d) Entrenamiento de un clasificador automático para que distinga entre diferentes conjuntos de características.
- e) Validación del clasificador introduciendo nuevos llamados para confirmar su veracidad.

Para obtener buenos resultados, se utilizan las características acústicas que presentan mayor variación entre géneros o individuos [3]. Estas características acústicas pueden ser clasificadas en dos grupos: estadísticas y no estadísticas. Las características estadísticas incluyen selección de frecuencias y cálculo de energía de las señales [7]. Los promedios de estas características acústicas han sido utilizados en algoritmos de aprendizaje; que han identificado diferentes especies de aves y ranas con un porcentaje de error aceptable [4, 14]. Las características no estadísticas como los coeficientes de predicción lineal (LPCs) [15] y coeficientes cepstrales en las frecuencias de Mel (MFCCs) [16] son comunes en el reconocimiento de voz en humanos. La aplicación de estos coeficientes a la identificación de especies ha dado buenos resultados en el reconocimiento de ranas, grillos y aves [17, 18].

Los métodos de análisis más comunes reportados en la literatura son los estadísticos. Éstos se basan en la búsqueda de picos de energía en el espectrograma de la señal; para poder utilizarlos como variables de la

vocalización del individuo [19]. Sin embargo, no es fácil encontrar puntos de referencia en el eje axial para el análisis de un llamado de murciélago. Por el contrario, la aplicación de características no estadísticas al reconocimiento de individuos no ha mostrado un desempeño alto; se entrega resultados con mayor dispersión en el desempeño de los algoritmos. No obstante, a nivel general estos modelos no estadísticos tienen mayor precisión, estabilidad y repetitividad [2].

En la práctica, cada vez es más común el uso de métodos de reconocimiento no estadísticos. Particularmente basados en el análisis de los coeficientes cepstrales en las frecuencias de Mel (MFCCs); usados para los reconocimientos automáticos de sonotipos asociados a alguna especie o género animal.

Los MFCCs son una representación del espectro de potencia de un sonido. Se basa en una transformada lineal sinusoidal de un espectro logarítmico sobre una escala no lineal de frecuencias de Mel [20]. Esta representación se basa en los cepstrum, que son el resultado de tomar la transformada de Fourier del logaritmo del espectro estimado de una señal. En la frecuencia de Mel (MFC), los cepstrum se centran en las bandas de frecuencia espaciadas según la escala de Mel. Por esta razón se aproximan más a la respuesta de los sistemas auditivos humanos en comparación con las bandas de frecuencia linealmente espaciadas. Sin embargo, la escala de mel está diseñada para el análisis de voz en humanos y está debería ser modificada por cada tipo de animal analizado en el sistema de reconocimiento de patrones [21].

Otros rasgos de las señales de audio pueden ser analizados a través de sus características acústicas. Éstas analizan los rasgos de bajo nivel en el dominio del tiempo (tales como energía, tasa de cruces por cero, tasa de silencios, entre otros); y en el dominio de la frecuencia (frecuencia, ancho de banda, entre otros) para cada llamado [7]. Estas características se extraen de grabaciones usualmente de un solo canal; en donde se pueden encontrar sonidos de varios individuos, sonidos complementarios y ruido de fondo [22].

Para poder procesar las características extraídas puede hacerse en forma no supervisada o supervisada. En aprendizaje no supervisado (agrupamiento o *clustering*) se construyen clases (grupo de datos con cierta identidad) a partir de los datos pre-procesados, desconociendo la salida esperada del agrupamiento. Por otro lado, en aprendizaje supervisado se realiza la identificación a partir de una salida esperada del procesamiento (e.g., máquinas de soporte vectorial, redes neuronales artificiales, arboles de decisión, entre otros).

Varios métodos de reconocimiento del habla en humanos se han aplicado al reconocimiento de animales (la mayoría supervisados). Estos métodos incluyen modelado sinusoidal de sílabas [23], mapas auto organizados [24], análisis

discriminante lineal (LDA) [25], redes neuronales artificiales [3], máquinas de soporte vectorial [4] y modelos ocultos de Markov [26].

A continuación, se profundiza en las redes neuronales artificiales, que es uno de los métodos más utilizados para el reconocimiento de especies animales y es el que se empleó en este proyecto.

Redes neuronales artificiales

Las RNA, redes neuronales artificiales, son modelos computacionales basados en el comportamiento neuronal. Para este método se simula el procesamiento de la información basado en el comportamiento de las neuronas.

En las RNA se manejan conjuntos de unidades denominadas neuronas artificiales. Estas neuronas se conectan entre sí para transmitir señales unas a otras, a estas conexiones se les llaman enlaces. A través de estos enlaces las señales enviadas se multiplican por un valor de peso que puede inhibir o incrementar el estado de activación de neuronas adyacentes.

A la salida de las neuronas existe un umbral o una función limitadora conocida como la función de activación, encargada de modificar el valor resultado para imponer un límite que permite la activación o no de la neurona.

Las RNA tienen dos etapas principales: entrenamiento y clasificación. En entrenamiento se utilizan datos de entrada y salida para encontrar los valores de los pesos (para cada enlace entre neuronas). A diferencia de la etapa de clasificación que hace uso de estos pesos para encontrar la señal de salida de una entrada.

El entrenamiento de las redes neuronales tiene como objetivo determinar los pesos de enlace entre neuronas. Para ello se analizan los grupos de datos pertenecientes a un mismo llamado de murciélago y se organizan en un vector de características. Este vector debe contener valores cuantitativos de las características para ser ponderadas por los pesos de cada enlace. Estos enlaces se dan entre cada neurona y cada neurona en la siguiente capa.

Las neuronas se organizan en capa de entrada, capas ocultas y capa de salida. Estas capas pueden tener un número variado de neuronas y enlaces. Los enlaces a las neuronas de la capa de entrada deben ser iguales al número de entradas en el vector de características.

Para la capa de salida, se requiere conocer la salida esperada con base en la cual se calcula el error generado. Con base a este error se cambian los pesos de los diferentes enlaces y se calcula nuevamente el error. Esto se realiza con todos los datos de entrenamiento hasta llegar a un umbral de error tolerado o un cierto

número de iteraciones. A esta etapa se la conoce como etapa de aprendizaje o entrenamiento.

Posteriormente viene la etapa de prueba donde se utilizan audios diferentes a los del entrenamiento. Es con estos datos que se realiza la validación de los pesos de la red neuronal, al comparar la salida de la red con el valor esperado de los datos. Tal comparación debe también quedar dentro del umbral que se obtiene en entrenamiento; de otra forma la red no reconocería bien audios fuera de los de entrenamiento.

Para el entrenamiento de las redes neuronales, es muy usado el método de *back propagation* [38]. Este método retroalimenta los errores encontrados durante el entrenamiento. Esto con el fin de ir ajustando los valores de los pesos de cada enlace. Este ajuste se puede dar mediante un método estocástico de gradiente descendiente u otros métodos como *extreme learning machines* [39] o redes "sin propagación" [40].

El método de *back propagation* ha sido usado previamente para el reconocimiento de especies de murciélagos [37]. En donde se utilizan RNA para la identificación de varias especies agrupadas por género. Sin embargo, estas redes se entrenaron con audios de murciélagos europeos que son menos diversos, por lo tanto, se tiene certeza en la especie a identificar. En el caso de los murciélagos que se encuentran en Colombia no se conocen antecedentes aplicativos que permitan diferenciar entre especies propias del país.

Materiales y Métodos

Base de datos

Para el desarrollo del proyecto se contó con la siguiente lista de grabaciones:

Tabla 1. Lista de grabaciones tratadas en el proyecto.

Sonotipo	Formato del archivo	Cantidad
Sonotipo 1	ZC	111
	WAV	110
Sonotipo 2	ZC	184
	WAV	179

Metodología propuesta

A continuación, presentamos la metodología usada para extraer e identificar los segmentos de los archivos proporcionados. En la Figura 1 se presentan las etapas propuestas. Son 5 etapas: grabación, segmentación, extracción de características, clasificación y por último identificación.

El desarrollo de estas etapas se realizó en unas funciones para el software R. Este es un software libre que permite realizar análisis estadísticos y uno de los más usados por la comunidad científica [42]. Este se escogió debido a su facilidad para el trato de archivos con extensión ZC permitiendo una lectura masiva de estos de manera sencilla.

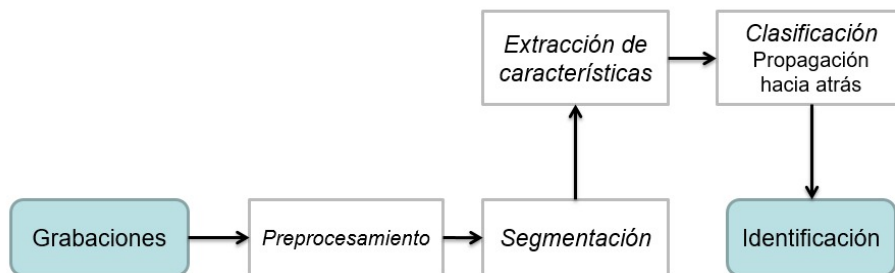


Figura 1. Etapas implementadas para la detección de sonotipos e identificación de ruido.

A continuación, se presenta la descripción detallada de cada una de las etapas.

Preprocesamiento

Esta etapa llamada preprocesamiento de datos, consiste en una etapa de reconocimiento manual de los sonotipos alojados en los archivos de los que consta nuestra base de datos. Inicialmente, se tienen las grabaciones sin catalogar, el primer paso del proceso es identificar y clasificar los segmentos de las grabaciones, aquellos que son llamados y aquellos que no. Este proceso se realizó de forma manual y nos permitió identificar la forma del sonotipo así como ciertas características prominentes que serían el insumo del algoritmo que realizaría la clasificación de los audios de forma automática.

Segmentación

Esta etapa se realiza de 2 formas diferentes para las grabaciones ZC y para WAV.

Archivos ZC

Es un formato de grabación principal para archivos bat que convierte la información espectral en una serie de puntos de tiempo frente a frecuencia como se muestra en la figura 2.

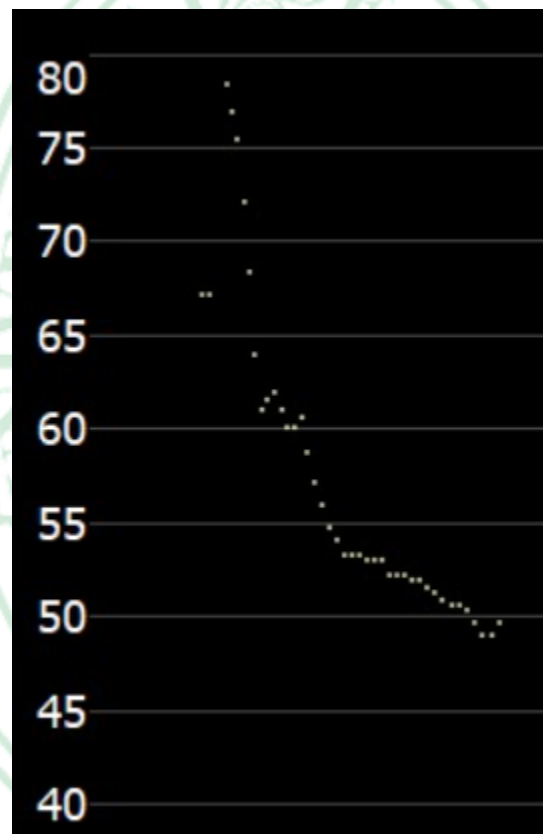


Figura 2. Ejemplo de una señal contenida en un archivo ZC.

Para las grabaciones en ZC se obtiene un vector de frecuencias y un vector de tiempos, es decir, un problema matricial.

La matriz de los audios ZC se construye mediante un muestreo en los cruces por cero de la señal de audio grabada; razón por la cual el vector de tiempos no tiene un tiempo de muestreo fijo, es decir, los tiempos aumentan de forma variable. Esto impide que se puedan establecer patrones de tiempos en los cuales realizar un análisis a la frecuencia.

Debido a lo anterior, se realiza un análisis de los audios muestreados y se encuentra que los llamados se hallan en agrupaciones de puntos del vector de frecuencia

con valor de frecuencia asignado entre los 20KHz y los 100KHz. Por este motivo se descartan valores por fuera de este rango.

Esto da como resultado un vector de frecuencias filtrado en donde se buscan agrupaciones de frecuencias consecutivas para analizar si la agrupación corresponde o no a un llamado.

Los llamados identificados en el preprocesamiento (método de detección manual) tienen una longitud mínima de 2 milisegundos aproximadamente. Esto corresponde con un mínimo de 35 posiciones de frecuencias consecutivas en el vector de frecuencia. Para los grupos de frecuencias con una cantidad menor a estos 35 se descartan y son tomadas como ruido.

Este método sirve para encontrar los segmentos de audio que correspondan a un sonotipo. Sin embargo, existen segmentos de audio que son ruido y que cumplen con el criterio de las 35 posiciones de frecuencia. Por este motivo se requiere hallar unas características que puedan diferenciar el ruido de un sonotipo, Estas características se explicarán en la siguiente sección.

Archivos WAV

Un archivo WAV o de forma de onda registra la información espectral completa dentro de un archivo de sonido, al igual que un archivo de música.

Para las grabaciones en formato WAV, se realiza el segmentado mediante el espectrograma de la señal usando la transformada corta de Fourier (STFT). Con esta información se realiza el análisis de potencia en el tiempo para determinar si se da un llamado en cierto intervalo de tiempo o si este debe ser descartado.

En la Figura 3 se observa un ejemplo de la señal vista desde el software Anabat Insight [35]. Aquí se muestra el espectrograma para un segmento de audio que contiene ruido.

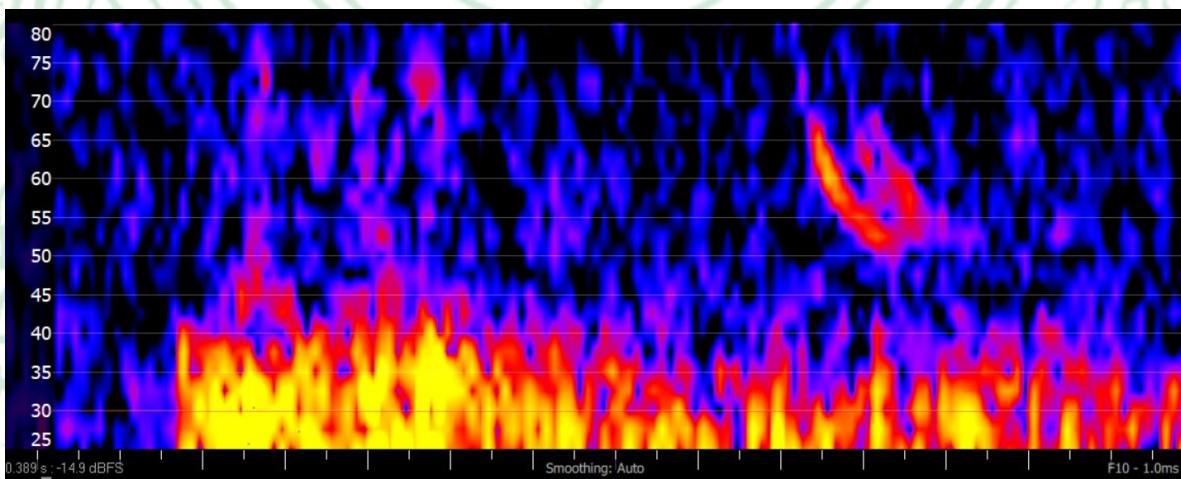


Figura 3. Ejemplo de una señal contenida en un archivo .WAV.

Extracción de Características

Luego de la segmentación de los audios se procede a la identificación de ciertas características relevantes para la construcción del algoritmo que realizará la clasificación de forma automática. Basados en algunos trabajos sobre identificación acústica de murciélagos en Europa [41] se puede apreciar que, para realizar este tipo de identificación se extraen algunos parámetros de la señal. Estos parámetros pueden ser las frecuencias de la señal entre ellas la máxima, mínima, inicial, final, etc. Para seleccionar los parámetros que diferencian los sonotipos se hizo un análisis usando la herramienta Anabat Insight [35]. Con este software se examinan los audios con la ayuda de la bióloga experta Daniela Martínez.

Las características del segmento de audio que se extraen y con las cuales se trabajan son mostradas en la Figura 4.

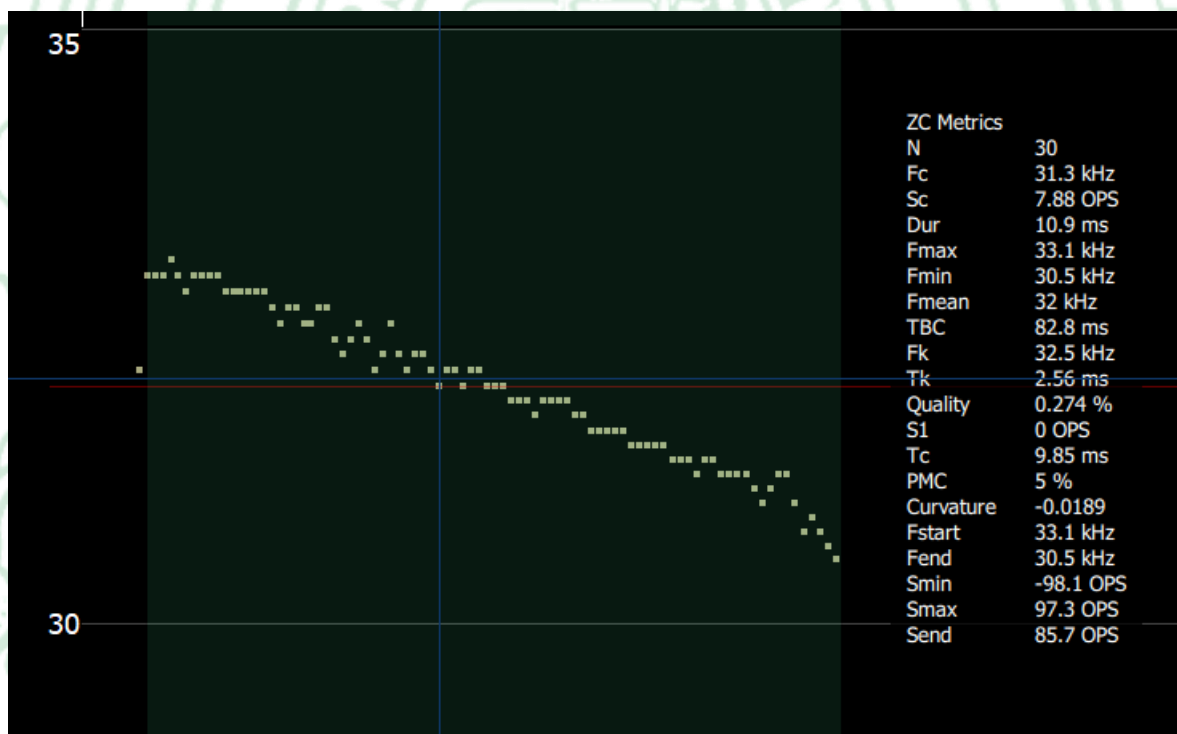


Figura 4. Características de un llamado observadas en Anabat Insight.

De acuerdo con este análisis se identifica que las siguientes características son relevantes: tiempo inicial y tiempo final del llamado (ms), duración del llamado (ms), frecuencia mínima, frecuencia máxima y frecuencia promedio del llamado (khz) al igual que la frecuencia inicial y la frecuencia final del llamado (khz). Estas características se escogen debido a que son las características que identifican los expertos para identificar el sonotipo.

En las funciones creadas en este proyecto se implementaron los algoritmos para estimar en R las características seleccionadas previamente.

Es importante aclarar que estas características se extrajeron independientemente del sonotipo o de la extensión del archivo analizado, es decir, las características analizadas son las mismas para los audios en ZC o WAV.

En la figura 5 se muestran los tiempos de las grabaciones donde se identifica un sonotipo, también se muestran las características extraídas para el segmento donde se visualiza el sonotipo.

Nombre de archivo	Numero	Numero de chills	Tiempo inicio	Tiempo final ch	Tiempo duracion del cl	Frecuencia minima c	Frecuencia maxima c	Frecuencia promedio c	Tiempo transcurrido al anter	Tiempo transcurrido al sig
2019-02-22 18-50-02.wav	2	1	156,463	163,457	6,994	31,250	33,898	32,702	NA	297,677
2019-02-22 18-50-02.wav	6	2	461,134	467,627	6,493	33,613	36,363	35,080	297,677	95,190
2019-02-22 18-50-02.wav	7	3	562,817	567,086	4,269	39,215	41,666	40,402	95,190	106,807
2019-02-22 18-50-02.wav	8	4	673,893	682,986	9,093	32,000	34,482	33,182	106,807	89,682
2019-02-22 18-50-02.wav	11	5	772,668	779,375	6,707	37,383	40,000	38,850	89,682	105,627
2019-02-22 18-50-02.wav	12	6	885,002	895,186	10,184	31,746	35,398	33,474	105,627	73,836
2019-02-22 18-50-02.wav	14	7	969,022	979,707	10,685	36,036	39,215	38,071	73,836	110,881
2019-02-22 18-50-02.wav	17	8	1,090,588	1,098,567	7,979	31,007	34,188	33,103	110,881	51,011
2019-02-22 18-50-02.wav	20	9	1,149,578	1,164,469	14,891	37,383	47,619	41,025	51,011	66,455
2019-02-22 18-50-02.wav	23	10	1,230,924	1,247,412	16,488	39,603	56,338	45,474	66,455	84,945
2019-02-22 18-50-02.wav	26	11	1,332,357	1,347,670	15,313	40,816	58,823	46,646	84,945	80,738
2019-02-22 18-50-02.wav	28	12	1,428,408	1,441,999	13,591	41,237	61,538	48,195	80,738	62,945
2019-02-22 18-50-02.wav	31	13	1,504,944	1,511,024	6,080	22,988	58,823	40,263	62,945	44,699
2019-02-22 18-50-02.wav	32	14	1,555,723	1,560,156	4,433	23,121	52,631	37,235	44,699	37,678
2019-02-22 18-50-02.wav	34	15	1,597,634	1,601,310	3,676	23,352	49,362	36,126	37,678	27,824
2019-02-22 18-50-02.wav	35	16	1,629,134	1,630,602	1,468	25,157	35,398	30,289	27,824	23,332
2019-02-22 18-50-02.wav	36	17	1,653,934	1,655,517	1,583	25,806	40,000	33,375	23,332	4,343
2019-02-22 18-50-02.wav	37	18	1,659,860	1,661,817	1,957	27,210	38,834	32,492	4,343	246,432
2019-02-22 18-50-02.wav	38	19	1,908,249	1,913,422	5,173	24,096	49,382	41,100	246,432	34,583
2019-02-22 18-50-02.wav	39	20	1,948,005	1,951,524	3,519	24,390	49,382	36,473	34,583	249,317

Figura 5. Características extraídas listadas en tabla.

Luego de extraer las características se procede a determinar cuáles de ellas son lo suficientemente descriptivas para clasificar. Para esto se escogen las que dan más información sobre si el llamado pertenece a el sonotipo I o al sonotipo II.

Para escoger estas características se procede a graficarlas con un diagrama de boxplots con el fin de visualizar la distribución de las variables de la manera más simplificada posible para evaluar la pertenencia a un sonotipo como se observa en las siguientes imágenes.

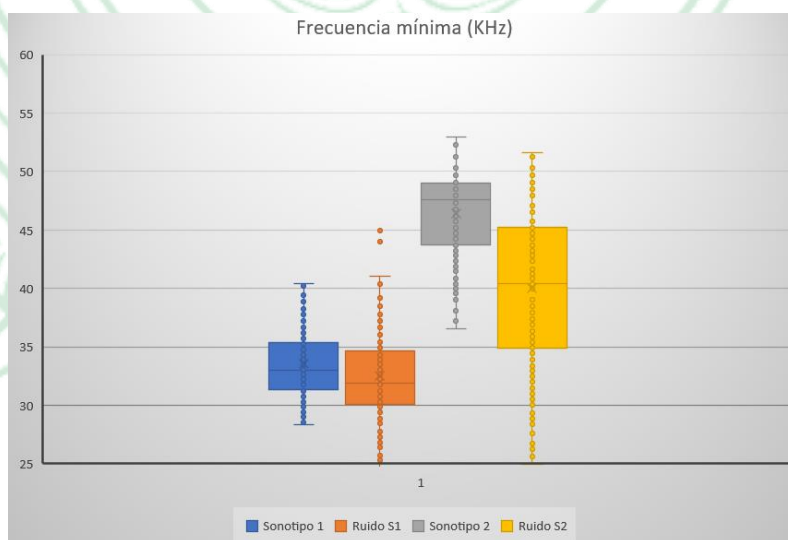


Figura 6. Boxplot de la característica Frecuencia mínima.

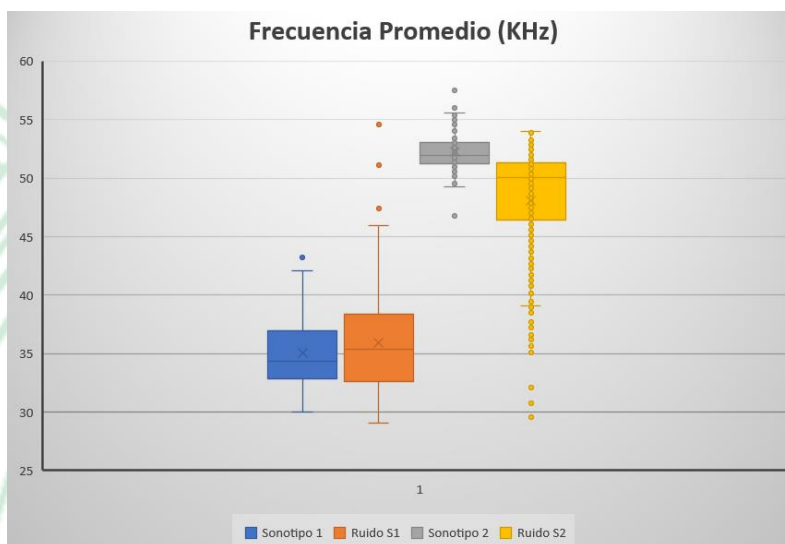


Figura 7. Boxplot frecuencia promedio.

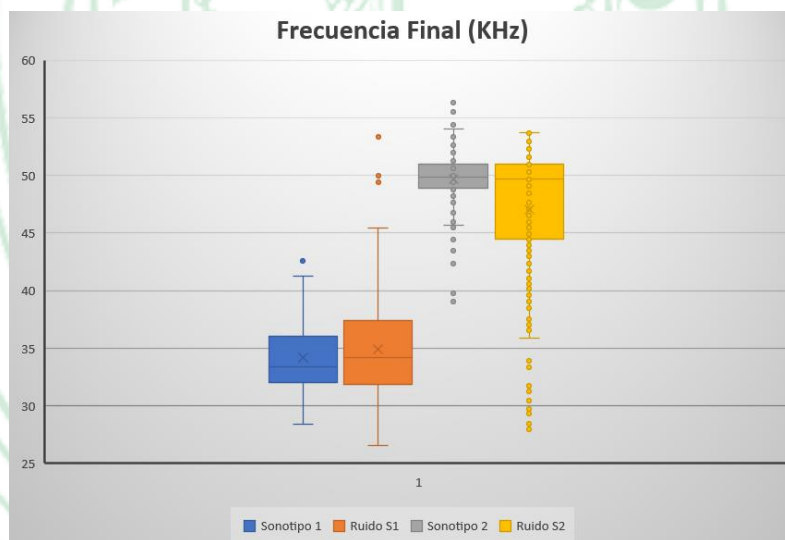


Figura 8. Boxplot frecuencia final.

Gracias a estos boxplots se determina que las características que ayudan a identificar si el llamado pertenece a sonotipo I o a sonotipo II eran la frecuencia mínima, frecuencia promedio y frecuencia final.

Consideración adicional para archivos de audio en formato ZC

Para los archivos ZC se debió realizar un paso adicional que fue escoger unas características descriptivas diferentes a las analizadas para diferenciar un sonotipo

del ruido. Para abordar este problema se recurrió a una descripción de la forma del segmento en la frecuencia.

La manera en que se describió la forma del segmento fue dividiéndolo en 3 partes y para cada una calcular la pendiente promedio entre los puntos del segmento y el promedio del delta de frecuencia entre los puntos. Estas 3 pendientes o deltas nos dan información en conjunto sobre la forma del segmento y por lo tanto un boxplot individual no basta para analizar su utilidad.

Esto nos llevó a una representación en 3 tramos de un segmento. Para el caso del sonotipo 1 se usaron las pendientes de cada tramo de segmento, mientras que para sonotipo 2 se usa el delta de frecuencia en cada tramo.

En la Figura 9 se muestran las 3 pendientes de los segmentos de sonotipo 1 en azul contrastado con las pendientes de los segmentos de ruido en rojo. Esto nos permite concluir que podemos excluir una parte del ruido con el uso de estas características mediante una RNA.

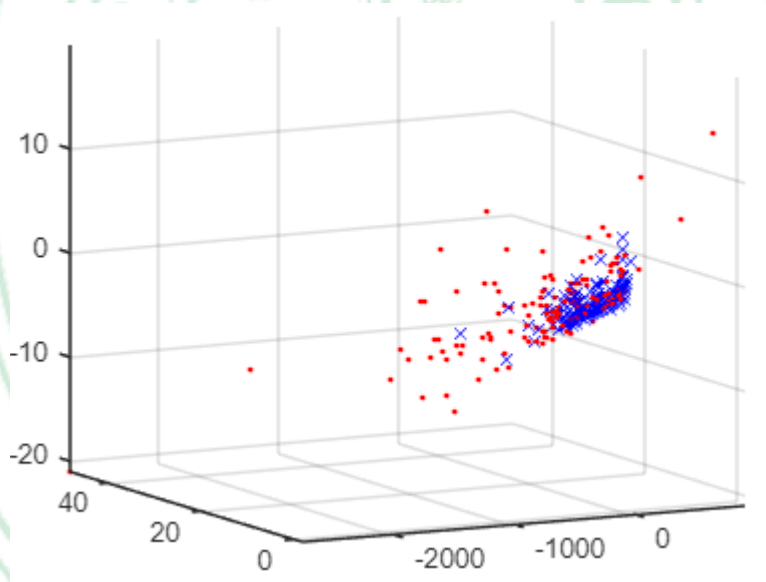


Figura 9. Pendiente promedio calculada en tres partes por segmento para sonotipo 1. (a) En rojo ruido (b) En azul sonotipo 1.

De forma similar, en la Figura 10 se muestran los 3 deltas de frecuencia en los segmentos de sonotipo 2 en azul comparados con los deltas de los segmentos de ruido en rojo. De esta comparación podemos inferir que se puede minimizar una parte del ruido que se encuentra fuera de espacio donde se encuentran los sonotipo 2.

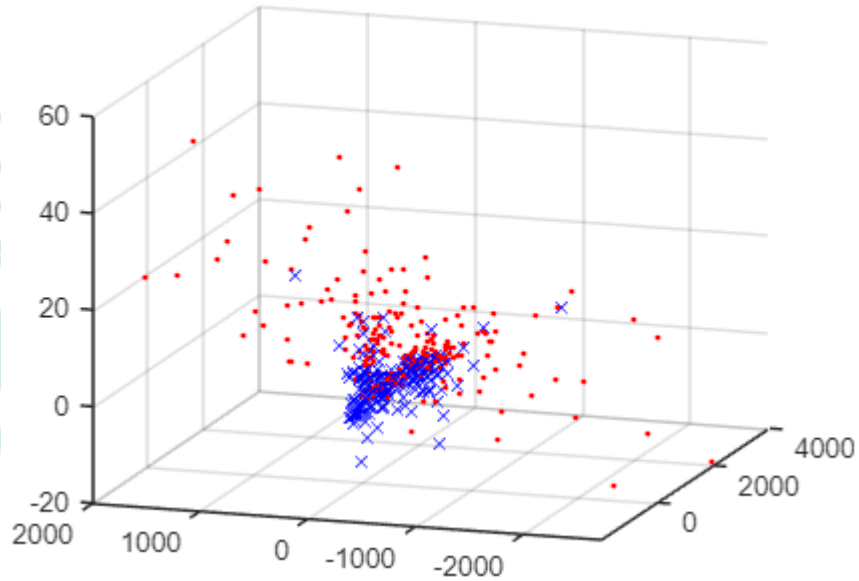


Figura 10. Delta de la frecuencia promedio calculado en tres partes por segmento para sonotipo 2. (a) En rojo ruido (b) En azul sonotipo 2.

Clasificación

Para el proceso de clasificación se propuso una RNA que permite la clasificación entre los 2 sonotipos analizados como se observa en la Figura 11.

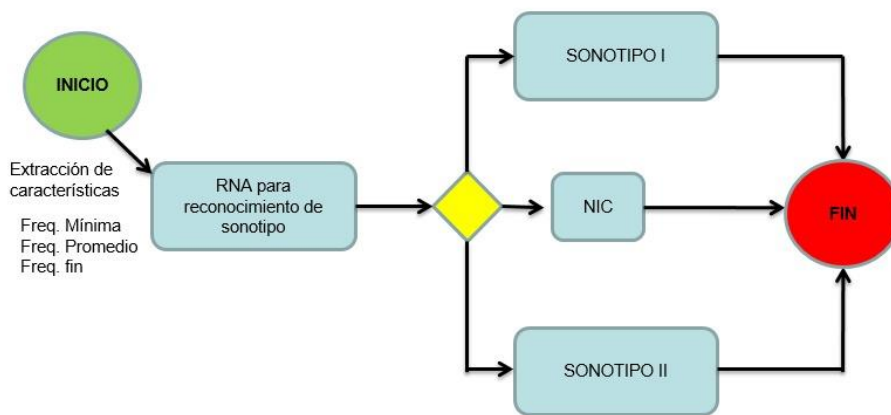


Figura 11. Diagrama de bloques del proceso de identificación de un archivo ZC.

Sin embargo, dada la descripción del audio en los archivos ZC, no es posible discernir con otra característica como la energía o la potencia de la señal si un segmento del audio es relevante para analizar. Por este motivo resultan muchos segmentos que son ruido dentro de la información a analizar, por lo cual se propuso un paso adicional para los archivos ZC el cual es mostrado en la figura 12.

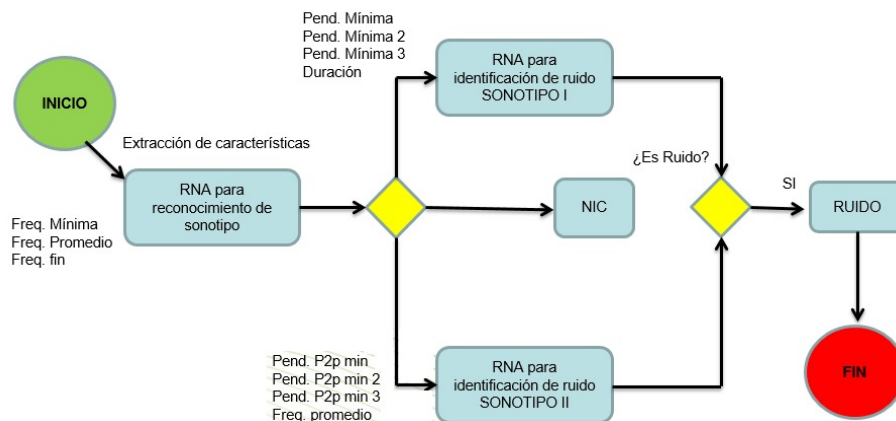


Figura 12. Diagrama de bloques del proceso de identificación para archivos ZC.

Para los 2 tipos de archivos se utiliza la misma red neuronal entrenada la cual se presenta en la Figura 13. Esto debido a que como se menciona previamente las características utilizadas para ambos archivos son discriminantes y la extracción de las características para WAV nos dan un valor muy cercano al trabajado con los archivos en ZC.

La idea del sistema propuesto para ZC es minimizar el error causado por la segmentación. Para ello se realizó un paso adicional a la clasificación de sonotipo. En este paso, una vez clasificado como Sonotipo 1 o Sonotipo 2, se entrenó una red neuronal que nos permita a través de la forma del segmento clasificarlo como ruido y no como un sonotipo. Este proceso se realizó por separado para cada sonotipo.

Por esta razón, el sistema propuesto para los archivos ZC, se compone de 3 RNA en cascada. La primera RNA que se comparte con los archivos WAV identifica el sonotipo como 1 o 2. Posteriormente, se utiliza una RNA para cada sonotipo, donde se clasifican los segmentos como ruido o como un sonotipo tal como se presenta en la figura 12.

Reconocimiento de sonotipo – primera red neuronal

La primera red neuronal del sistema propuesto realiza el reconocimiento del sonotipo. Se encarga de diferenciar aquellos segmentos de audio que pertenecen a el sonotipo I, sonotipo II o a la Clase de No Identificación (NIC) con la ayuda de las características diferenciadoras la frecuencia mínima, promedio y final.

Es importante mencionar que los segmentos de audio no siempre van pertenecer a sonotipo I o sonotipo II; es posible que los segmentos de audio no pertenezcan a ninguno de estos o que sean ruido. Por este motivo la función de activación de la

red neuronal se configuró como lineal y apuntando a 1 como salida para sonotipo I y a 2 como salida para sonotipo II.

Como umbral para la detección se determinó que 1 ± 0.35 y 2 ± 0.35 serían salidas aceptables para sonotipo I o II respectivamente. De este modo, todo lo que este por fuera de estos rangos, se considera entonces como NIC.

Por último, la red recibe las características de entrada frecuencia mínima, promedio y final a una única capa de 5 neuronas como se muestra en la figura 13.

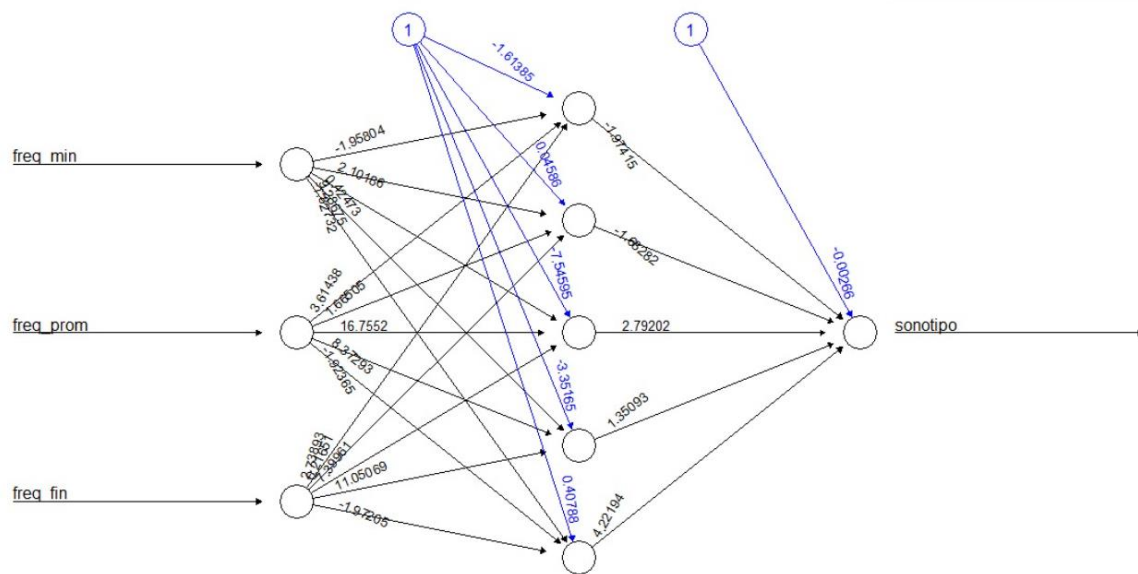


Figura 13. Imagen de la red neuronal 1 usada para archivos WAV y ZC.

A continuación, se explican las diferentes RNAs adicionales que integran al sistema para la aplicación de archivos ZC.

RNA para identificación de ruido

Debido a que hay una gran cantidad de segmentos de ruido para segmentos de audio ZC se propuso un paso adicional. Este paso consiste en adicionar las redes neuronales dos y tres como método de detección de ruido.

Estas redes neuronales se utilizan luego del primer reconocimiento para una segunda identificación de los llamados al igual que una clasificación de estos. En estas redes presentadas en la Figura 14 para sonotipo 1 y Figura 15 para sonotipo 2 se toman como características diferenciadoras variables diferentes a las que corresponden a la entrada de la primera red neuronal. Esto con el fin de tener una

segunda clasificación de los segmentos de audio y una mayor fiabilidad de la diferenciación

Estas características son las siguientes:

Tabla 2. Tabla de características diferenciadoras para las redes neuronales 2 y 3.

RNA	Característica diferenciadora
Sonotipo 1	Delta de frecuencia promedio calculada en el primer tramo del segmento de audio
	Delta de frecuencia promedio calculada en el segundo tramo del segmento de audio
	Delta de frecuencia promedio calculada en el tercer tramo del segmento de audio
	Duración
Sonotipo 2	Pendiente promedio punto a punto calculada en el primer tramo del segmento de audio
	Pendiente promedio punto a punto calculada en el segundo tramo del segmento de audio
	Pendiente promedio punto a punto calculada en el tercer tramo del segmento de audio
	Frecuencia promedio

Con la ayuda de estas redes neuronales la clasificación de ruido y sonotipo se realiza de una forma más consistente lo que permite obtener resultados más concluyentes y disminuir el porcentaje de error en una cantidad considerable.

Para estas redes, se construyeron se optó por una configuración de 2 capas de red con 3 y 5 neuronas respectivamente como se muestra en las figuras 13 y 14.

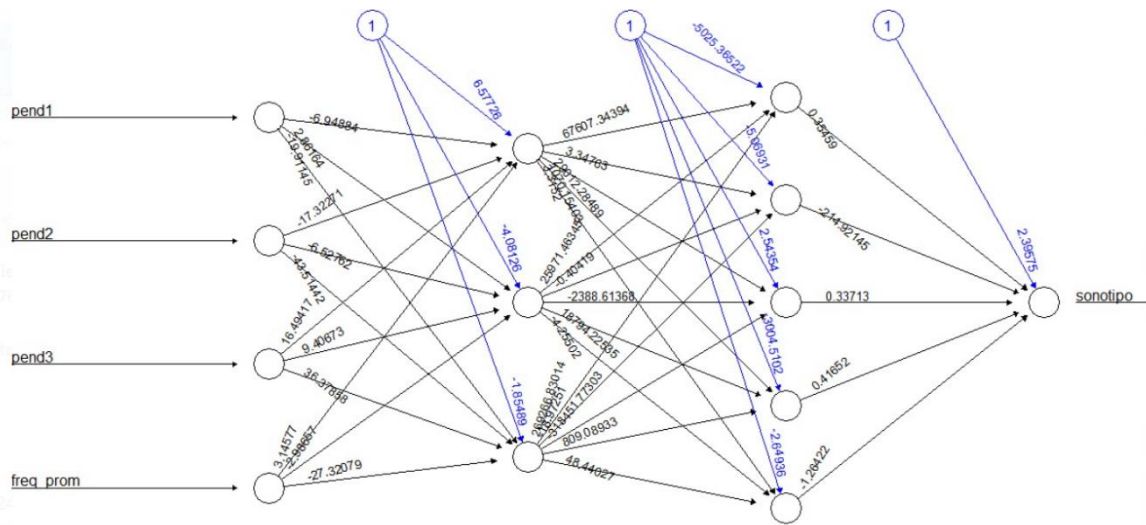


Figura 14. RNA para diferenciar entre ruido y sonotipo 1 en archivos ZC.

Cabe aclarar que la identificación de estas características se realiza de manera similar a las características diferenciadoras de la primera red, haciendo uso de los diagramas de cajas y bigotes (boxplots).

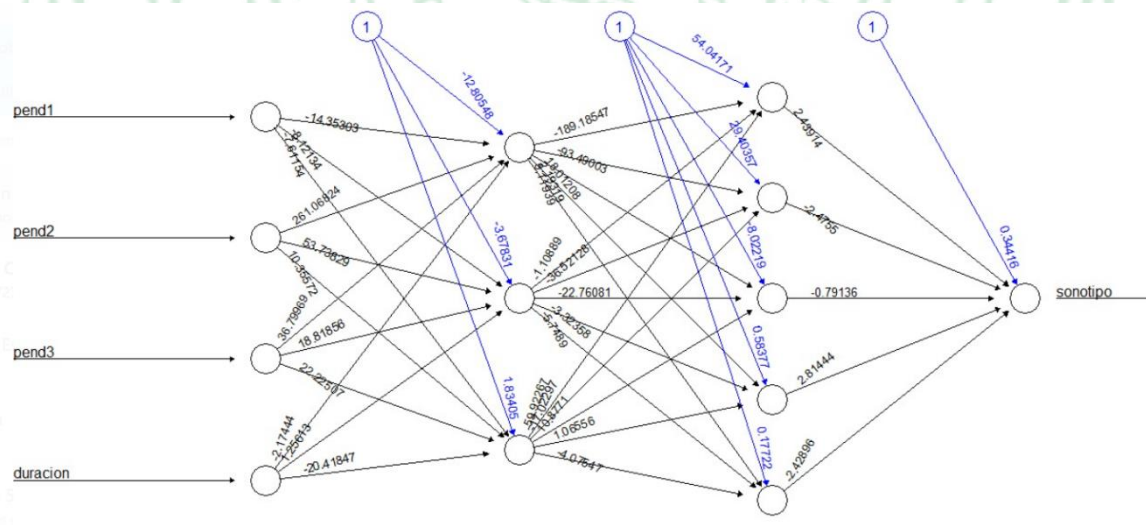


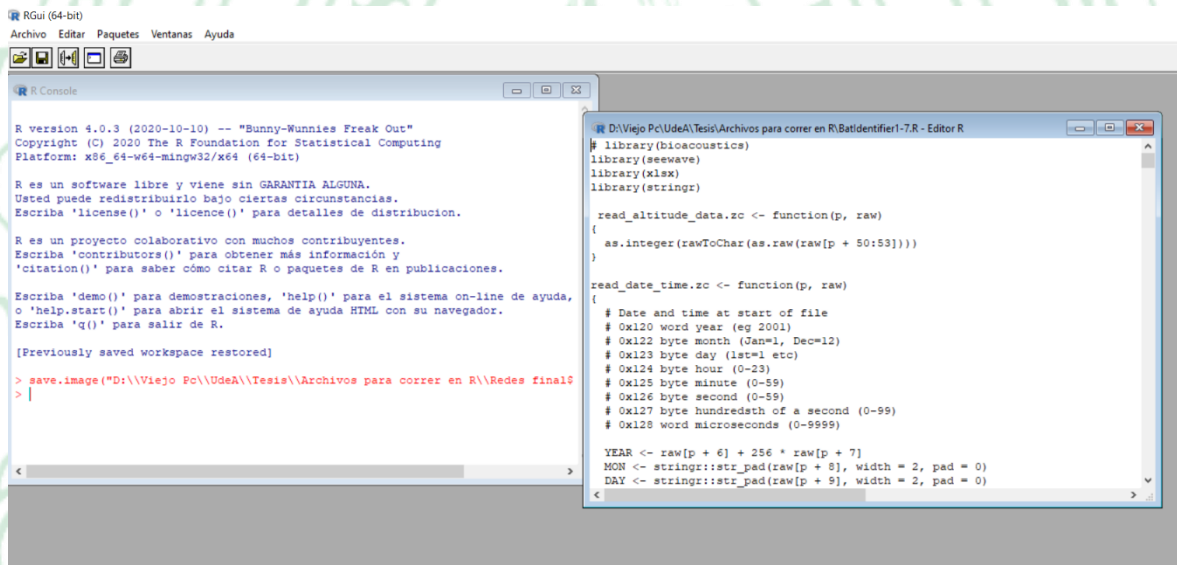
Figura 15. RNA para diferenciar entre ruido y sonotipo 2 en archivos ZC.

Software implementado

Los algoritmos se implementaron como una herramienta para el análisis de archivos de audio que contienen llamados de murciélagos, su objetivo es el de identificar y clasificar los llamados que se encuentran dentro de los archivos ZC o WAV. Con la

metodología descrita anteriormente el código permite la identificación de los sonotipos de dos géneros de murciélagos que se encuentran en las grabaciones que son insumo para el programa.

Nuestro trabajo da como resultado dos funciones desarrolladas para R [42] las cuales al ser ejecutadas requieren de la importación de otras librerías para su correcto funcionamiento. Luego de que el archivo realiza esta importación requiere que el usuario ingrese la ruta de su computador donde se encuentra la lista de los archivos en ZC o en WAV que requiere analizar. Al ingresar la ruta, los programas agrupan todos los archivos contenidos en la ruta ingresada y proceden a realizar el procesamiento y caracterización de estos. Luego de un tiempo prudente de catalogación y el cual depende de la extensión del archivo analizado (ZC o WAV) los programas agrupan los resultados finales en un archivo de extensión CSV ubicado en la misma ruta que se ingresó para su análisis.



```
RGui (64-bit)
Archivo  Editar  Paquetes  Ventanas  Ayuda

R Console
R version 4.0.3 (2020-10-10) -- "Bunny-Wunnies Freak Out"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribución.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

[Previously saved workspace restored]

> save.image("D:\\Viejo Pc\\UdeA\\Tesis\\Archivos para correr en R\\Redes final
> |

D:\Viejo Pc\UdeA\Tesis\Archivos para correr en R\BatIdentif1-7.R - Editor R
# library(bioacoustics)
library(seewave)
library(xlsx)
library(stringr)

read_altitude_data.zc <- function(p, raw)
{
  as.integer(rawToChar(as.raw(raw[p + 50:53])))
}

read_date_time.zc <- function(p, raw)
{
  # Date and time at start of file
  # 0x120 word year (eg 2001)
  # 0x122 byte month (Jan=1, Dec=12)
  # 0x123 byte day (1st=1 etc)
  # 0x124 byte hour (0-23)
  # 0x125 byte minute (0-59)
  # 0x126 byte second (0-59)
  # 0x127 byte hundredth of a second (0-99)
  # 0x128 word microseconds (0-9999)

  YEAR <- raw[p + 6] + 256 * raw[p + 7]
  MON <- stringr::str_pad(raw[p + 8], width = 2, pad = 0)
  DAY <- stringr::str_pad(raw[p + 9], width = 2, pad = 0)
}
```

Figura 16. Captura del programa para identificación de sonotipos para archivos. ZC.

El resultado de los análisis por cada archivo nos dio un promedio de 85,5 segundos de procesamiento por archivo. En este trabajo el tiempo de duración de cada audio fue de 1 minuto.

Resultados y análisis

A continuación, se presentan los resultados obtenidos en la implementación de los algoritmos con los datos de prueba y los de entrenamiento para ambos formatos de archivo (WAV y ZC). Se explica la cantidad de archivos analizados para cada red de cada sistema en su etapa de prueba y de entrenamiento.

Entrenamiento

En esta sección se presentan los resultados obtenidos al entrenar las redes neuronales para el sistema implementado.

Primera red neuronal:

Esta primera red es válida para los archivos de tipo ZC y WAV, ya que las características que se extraen son las mismas y se maneja la misma normalización.

En la Tabla 3 se muestra la matriz de confusión de la etapa de entrenamiento para la primera red neuronal que fue entrenada con audios en formato ZC. En esta etapa y para esta red se analizaron 288 segmentos de audio en el algoritmo para clasificar sonotipos.

En esta ejecución se reconocieron 137 segmentos pertenecientes a sonotipo 1 y 151 archivos pertenecientes a sonotipo 2 usando la clasificación supervisada. Como se visualiza en la Tabla 3 donde la fila 2 y la columna 2 indican los segmentos correctamente reconocidos como sonotipo 1, la fila 3 y la columna 3 indican los segmentos correctamente reconocidos como sonotipo 2.

La Tabla 3 nos indica los valores de la precisión (valores en verde) para cada sonotipo. Esto indica que tenemos una precisión del 100% en la detección de sonotipo 1 y sonotipo 2.

Adicionalmente la sensibilidad es del 100% de los segmentos de sonotipo 1 identificados como sonotipo 1 y del 100% de los segmentos de sonotipo 2 identificados como sonotipo 2.

Tabla 3. Matriz de confusión de la primera red neuronal en etapa de entrenamiento para archivos ZC.

Matriz de confusión Primer red				
	Sonotipo 1	Sonotipo 2	NIC	Sensitividad
Sonotipo 1	137	0	0	1
Sonotipo 2	0	151	0	1
	1	1		

Estos resultados se obtuvieron con los datos de entrenamiento para los cuales solo se utilizaron los datos de segmentos que identificados como sonotipos y ninguno de los que se identificaron como ruido.

Aunque el desempeño de los datos de entrenamiento fue del 100% se debe verificar el desempeño a clasificar nuevos datos (de prueba) no usados en entrenamiento. Este resultado se presenta más adelante.

Segunda red neuronal:

En esta etapa se analizaron 479 segmentos de audio en el algoritmo para clasificar segmentos ZC. En esta ejecución se reconocieron correctamente 196 segmentos pertenecientes a sonotipo 1 y 206 segmentos pertenecientes a ruido usando la clasificación supervisada. También se detectaron incorrectamente 60 segmentos como ruido cuando pertenecían al sonotipo 1 y 17 como ruido cuando pertenecían al sonotipo 1.

La Tabla 4 nos indica los valores de la precisión (valores en verde) para cada sonotipo. Esto señala que tenemos una precisión del 76,56% en la detección de sonotipo 1 y de un 92% de la detección de ruido de forma precisa.

Adicionalmente se detectaron 196 segmentos de sonotipo 1 del total de 213 correctamente, esto equivale al 92,02% este valor corresponde a la sensibilidad para sonotipo 1. De igual manera se detectaron 206 segmentos de ruido del total de 266 correctamente equivalentes a una sensibilidad del 77.44%.

Tabla 4. Matriz de confusión de la segunda red neuronal en etapa de entrenamiento para archivos ZC.

Matriz de confusión Filtro Ruido S1			
NN S1 Entrenamiento	Sonotipo 1	Ruido	Sensitividad
Sonotipo 1	196	17	0,920187793
Ruido	60	206	0,77443609
	0,765625	0,923766816	

Tercera red neuronal:

En la Tabla 5 se muestra la matriz de confusión de la etapa de entrenamiento para la tercera red neuronal. Se analizaron 589 segmentos de audio en el algoritmo para clasificar segmentos ZC para esta red. En la tabla 5 podemos observar en la fila 2 columna 2 que se reconocieron correctamente 158 segmentos pertenecientes al sonotipo 2 y en la fila 3 columna 3 se observan 320 segmentos reconocidos como ruido.

La Tabla 5 también nos indica los valores de la precisión (valores en verde) para cada sonotipo. Esto señala que tenemos una precisión del 74,52% en la detección de sonotipo 2 y de un 84% de la detección de ruido de forma precisa.

Adicionalmente se detectaron 158 segmentos de sonotipo 2 correctamente de un total de 215 esto equivale al 73.48%. Este valor corresponde a la sensibilidad para sonotipo 2. De igual forma se observa que se detectaron 320 segmentos de ruido correctamente de un total de 374 representando un valor de sensibilidad de 85.55%.

Tabla 5. Matriz de confusión de la tercera red neuronal en etapa de entrenamiento para archivos ZC.

Matriz de confusión Filtro Ruido S2			
NN S2 Entrenamiento	Sonotipo 2	Ruido	Sensitividad
Sonotipo 2	158	57	0.734883721
Ruido	54	320	0.855614973
	0.745283019	0.848806366	

Prueba

En esta sección se presentan los resultados obtenidos al utilizar las redes neuronales para el sistema implementado con datos diferentes a los de entrenamiento.

Archivos ZC

Primera red neuronal

En esta etapa se analizaron 464 segmentos de audio en el algoritmo para clasificar segmentos ZC.

De la Tabla 6 se observa como 175 segmentos de audio fueron correctamente identificados como sonotipo 1 y 44 erróneamente identificados como sonotipo 2. 226 segmentos se reconocieron correctamente como pertenecientes a sonotipo 2 y erróneamente se identificaron 4 como sonotipo 1.

Tabla 6. Matriz de confusión de la primera red neuronal en etapa de prueba archivos ZC.

Matriz de confusión Primer red				
NN1 Pruebas	Sonotipo 1	Sonotipo 2	NIC	Sensitividad
Sonotipo 1	175	4	3	0,961538462
Sonotipo 2	44	226	12	0,80141844
	0,799086758	0,982608696		

La Tabla 6 también nos indica los valores de la precisión (valores en verde) para cada sonotipo. Esto señala que tenemos una precisión del 79,90% en la detección de sonotipo 1 y de un 98.26% de la detección de sonotipo 2 de forma precisa.

Adicionalmente se detectaron 175 segmentos de sonotipo 1 correctamente de un total de 179 equivalentes al 96.15%. Este valor corresponde a la sensitividad para sonotipo 1. De igual forma se observa que se detectaron 226 segmentos de sonotipo 2 correctamente de un total de 270 representando un valor de sensitividad de 80.14%.

Segunda red neuronal:

Para esta red se analizaron 120 segmentos de audio en el algoritmo para clasificar segmentos ZC, de los cuales 56 segmentos fueron correctamente reconocidos como sonotipo 1, 38 correctamente como ruido. 17 segmentos fueron erróneamente reconocidos como sonotipo 1 siendo estos ruidos y 9 erróneamente como ruido siendo estos sonotipo 1.

Tabla 7. Matriz de confusión de la segunda red neuronal en etapa de prueba para archivos ZC.

Matriz de confusión Filtro Ruido S1			
NN S1 Pruebas	Sonotipo 1	Ruido	Sensitividad
Sonotipo 1	56	9	0,861538462
Ruido	17	38	0,690909091
	0,767123288	0,808510638	

La Tabla 7 también nos indica los valores de la precisión (valores en verde) para cada sonotipo. Esto señala que tenemos una precisión del 76,71% en la detección de sonotipo 1 y de un 80,85% de la detección de ruido de forma precisa.

Adicionalmente se detectaron 56 segmentos de sonotipo 1 correctamente de un total de 65 esto equivale al 86,15%. Este valor corresponde a la sensitividad para sonotipo 1. De igual manera se observa que se detectaron 38 segmentos de ruido correctamente de un total de 55 representando un valor de sensitividad de 69,09%.

Tercera red neuronal:

Se analizaron 147 segmentos de audio en el algoritmo para clasificar segmentos ZC. Se reconocieron correctamente 40 segmentos pertenecientes al sonotipo 2 y se clasificaron correctamente 70 segmentos como ruido.

Tabla 8. Matriz de confusión de la tercera red neuronal en etapa de prueba para archivos ZC.

Matriz de confusión Filtro Ruido S2			
NN S2 Pruebas	Sonotipo 2	Ruido	Sensitividad
Sonotipo 2	40	25	0,615384615
Ruido	12	70	0,853658537
	0,769230769	0,736842105	

La Tabla 8 también nos indica los valores de la precisión (valores en verde) para cada sonotipo. Esto señala que tenemos una precisión del 76,92% en la detección de sonotipo 2 y de un 73,68% de la detección de ruido de forma precisa.

Adicionalmente se detectaron 40 segmentos de sonotipo 2 correctamente de un total de 65 esto equivale al 61,53%. Este valor corresponde a la sensitividad para sonotipo 2. De igual manera se observa que se detectaron 70 segmentos de ruido correctamente de un total de 82 representando un valor de sensitividad de 85,36%.

Finalmente, para el sistema total integrado por las 3 RNA se obtienen los resultados mostrados en la Tabla 9.

Se observa entonces que:

- Para sonotipo 1 se reconocen correctamente 244 de los 279 segmentos identificados como sonotipo 1 lo que equivale a una sensibilidad del 87,45%.
- Para sonotipo 2 se identifican 171 de 284 segmentos correctamente lo que equivale a una sensibilidad del 60,21%.
- Para el ruido se identifican 586 de 799 segmentos de forma correcta, lo que equivale a una sensibilidad del 73,34%.

Tabla 9. Matriz de confusión total del sistema para audios en formato ZC.

Total	Sonotipo 1	Sonotipo 2	Ruido	NIC	Sensitividad
Sonotipo 1	244	0	31	4	0.874551971
Sonotipo 2	1	171	111	1	0.602112676
Ruido	77	114	586	22	0.733416771
	0.757763975	0.6	0.804945055		

La precisión del sistema total se ve en las casillas en verde de la Tabla 9. Aquí se observa que 244 de los 322 segmentos clasificados como Sonotipo 1 están debidamente clasificados, lo que equivale a una precisión del 75,77%. Para Sonotipo 2 se tienen 171 de 285 segmentos debidamente clasificados lo que da una precisión del 60%. Por último, para el ruido se clasifican 586 de 728 segmentos de forma correcta lo que da una precisión del 80,49%.

Archivos WAV

Para los archivos WAV no fue necesario reentrenar la red 1. Los resultados permitieron trabajar solo con esta red. Se analizaron 575 segmentos de audio en el programa para clasificar segmentos de audios WAV. Se reconocieron correctamente los segmentos pertenecientes al sonotipo 1 y al sonotipo 2 con las características utilizadas.

Tabla 10. Matriz de confusión del sistema total para el análisis de archivos WAV.

Total	Sonotipo 1	Sonotipo 2	NIC	Total	Sensitividad
Sonotipo 1	138	0	0	138	1
Sonotipo 2	0	130	0	130	1
Ruido	26	57	224	307	0.729641694
	0.841463	0.695187166	1		

No todos los segmentos que se identificaron a partir de los umbrales de potencia de la señal correspondían a sonotipos. Por este motivo se añadió una fila para indicar los segmentos de audio que no hacen parte de los sonotipos identificados para las grabaciones y que se consideran ruido.

Para este ruido que ingresa al sistema, se observa que baja la precisión de los sonotipos teniendo en cuenta que se detectan algunos de los segmentos de ruido como sonotipo 1 o sonotipo 2.

Aun así, se observa que la precisión del sistema es del 84% para sonotipo 1 y del 69% para sonotipo 2. Mientras que la sensibilidad de los sonotipos 1 y 2 es del 100% y la sensibilidad del ruido es del 77,96%.

Para terminar, se realizó un análisis del desempeño de las funciones para WAV y ZC donde se trabajó con la base de datos de grabaciones de un máximo de 15 segundos.

El resultado de los análisis para archivos ZC nos arrojó un resultado promedio de 30,6 segundos de procesamiento por cada archivo ZC. Esto quiere decir que el tiempo de análisis por cada audio es de aproximadamente 0,5 min para audios de máximo 15 segundos.

Para los archivos WAV, se realiza un procesamiento computacional más pesado, ya que se debe analizar el espectrograma de la señal. Por este motivo el análisis de rendimiento es de mayor tiempo de ejecución que en los archivos ZC.

El resultado de los análisis nos arrojó un resultado promedio de 85,5 segundos de procesamiento por cada archivo WAV. Esto quiere decir que el tiempo de análisis por cada audio es de aproximadamente 1,5 min para audios de máximo 15 segundos.

Conclusiones

- En este trabajo se propuso una forma de clasificar automáticamente los llamados de murciélagos utilizando patrones de sonido, realizando un análisis no invasivo y ayudando a un monitoreo más sencillo para los expertos.
- Gracias al método de clasificación manual se pudo estimar la similitud entre llamados. Esto ofrece una ventaja en cuanto a la detección de sonotipos.
- Los resultados del clasificador se pueden ver afectados debido a la presencia de ruido en los audios o grabaciones ya que este se encuentra presente en las grabaciones que se utilizan para las clasificaciones debido a que las grabaciones son tomadas al aire libre no se encuentran ausentes del ruido del ambiente natural o de otras especies en el ambiente monitoreado.
- Aunque el método es rápido para la clasificación de llamados en archivos ZC, este es más lento para los archivos WAV debido a la duración extensa de los archivos y a que en estos deben analizarse las características de potencia contenida en este tipo de archivos.

Trabajo a Futuro

Como trabajo a futuro se propone:

- Realizar una mejora en la etapa de segmentación, esta etapa resulta de gran importancia pues de aquí se parte para una buena selección de características y clasificación.
- Disminuir ruido de fondo de la señal. Si es posible debe intentarse grabar a los murciélagos con el menor ruido de fondo o realizar un pre tratamiento de la señal para disminuir el ruido logrando que este no sea un factor que opaque la clasificación de los sonotipos.
- Migrar el software a una plataforma con menos costo computacional y más amigable para el usuario.
- Modularizar el código construido para un mejor entendimiento de sus particularidades para futuras implementaciones.
- Implementar la programación paralela en el código construido para incrementar el tiempo de análisis del sistema usado para analizar los archivos con extensión .WAV ofreciendo una disminución considerable en el tiempo de clasificación de estos.

Referencias

- [1] W.P. Chen, et al., "Automatic recognition of frog calls using a multi-stage average spectrum" *Computers & Mathematics with Applications*, Volume 64, Issue 5, 2012.
- [2] J. Cheng, Y. Sun and L. Ji, "A call-independent and automatic acoustic system for the individual recognition of animals: A novel model using four passerines", *Pattern Recognition*, Volume 43, Issue 11, 2010.
- [3] E.J.S. Fox, "A new perspective on acoustic individual recognition in animals with limited call sharing or changing repertoires", *Animal Behaviour* 75, 2008.
- [4] M.A. Acevedo, et al., "Automated classification of bird and amphibian calls using machine learning: a comparison of methods", *Ecological Informatics* 4, 2009.
- [5] E.J.S. Fox, J.D. Roberts and M. Bennamoun, "Text-independent speaker identification in birds", *Proceedings of the Inter speech and Ninth International Conference on Spoken Language Processing*, vols.1-5, 2006.
- [6] M.A. Roch, et al., "Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California", *Journal of the Acoustical Society of America*, vol. 121, 2007.

[7] C. Molnar, et al., "Classification of dog barks: a machine learning approach", *Animal Cognition* 11, 2008.

[8] J. R. Speakman & P. A. Racey, "No cost of echolocation for bats in flight", Department of Zoology, University of Aberdeen. *Nature*, vol. 350, 1995.

[9] Darren J. H. Sleep and R. Mark Brigham, "An experimental test of clutter tolerance un bats", Department of Biology, University of Regina, *Journey of Mammalogy*, vol. 84, 2003.

[10] Michael J. O'Farrell, Bruce W. Miller and William L. Gannon, "Qualitative identification of free-flying bats using the Anabat detector", *Journal of Mammalogy*, vol. 80, 1999.

[11] Bernal, X., Guarizo, C., Luddecke, H., "Geographic variation in advertisement call and genetic structure of *colostethus palmatus* (anura, dendrobatidae) from the colombian andes". *herpetologica*, 61 (4), 395–408, 2005.

[12] Jiménez-Rivillas, C., Vargas, L., Fang, J., Di Filippo, J., Daza, J., "Advertisement Call of *Diasporus anthrax* (Lynch, 2001) (Anura: Eleutherodactylidae) with Comparisons to Calls from Congeneric Species", *South American Journal of Herpetology*, 8(1), 1–4, 2013.

[13] Bernal, X., Guarizo, C., Luddecke, H., "Geographic variation in advertisement call and genetic structure of *colostethus palmatus* (anura, dendrobatidae) from the colombian andes". *herpetologica*, 61 (4), 395–408, 2005.

[14] C.J. Huang, et al., "Frog classification using machine learning techniques", *Expert Systems with Applications*, vol 36, 2009.

[15] L. Rabiner and B.H. Juang, "Fundamentals of Speech Recognition", Prentice Hall, Englewood Cliffs, NJ, 1993.

[16] S. Nakagawa, L. Wang and S. Ohtsuka, "Speaker Identification and Verification by Combining MFCC and Phase Information", *IEEE Transactions on Audio, Speech, and Language Processing*, Volume: 20, Issue: 4, 2012.

[17] C.H. Lee et al., "Automatic recognition of animal vocalizations using averaged MFCC and linear discriminant analysis", *Pattern Recognition Letters* 27, 2006.

[18] V.M. Trifa, et al., "Automated species recognition of ant birds in a Mexican rainforest using hidden Markov models", *Journal of the Acoustical Society of America*, vol. 123, 2008.

[19] P. Somervuo, A. Harma and S. Fagerlund, "Parametric representations of bird sounds for automatic species recognition", *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, 2006.

[20] P. Mermelstein "Distance measures for speech recognition, psychological and instrumental" *Pattern Recognition and Artificial Intelligence*, 1976.

[21] Md. Sahidullah and G. Saha, "Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition". *Speech Communication*, vol. 54, 2012.

[22] Y. Wang, Z. Liu, and J.C. Huang. "Multimedia content analysis using both audio and visual cues". *IEEE Signal Processing Magazine*, 2000.

[23] A.Hama, "Automatic identification of bird species based on sinusoidal modeling of syllables", *Proceeding of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol.5, 2003.

[24] E.E. Vallejo, M.L. Cody and C.E. Taylor, "Unsupervised acoustic classification of bird species using hierarchical self-organizing maps", *Progress in Artificial Life, Proceedings*, 2007.

[25] C.H. Lee et al., "Automatic recognition of animal vocalizations using averaged MFCC and linear discriminant analysis, *Pattern Recognition Letters* 27, 2006.

[26] V.M. Trifa, et al., "Automated species recognition of ant birds in a Mexican rainforest using hidden Markov models", *Journal of the Acoustical Society of America*, vol. 123, 2008.

[27] C. Isaza, et al., "Contribution of fuzzy classification for the diagnosis of complex systems" *7th IFAC International Symposium of Fault Detection, Supervision and Safety of Technical Processes SAFEPROCESS*, 2009.

[28] H. Meinedo and J. Neto, "Audio segmentation, classification and clustering in a broadcast news task", *proceeding of IEEE International Conference on: Acoustics, Speech, and Signal Processing (ICASSP 03)*, Volume 2, 2003.

[29] K.L. Wu, "Analysis of parameter selections for fuzzy c-means", *Pattern Recognition*, Volume 45, Issue 1, January, 2012.

[30] W.M.B.W. Mohd, et al., "An Improved Parameter less Data Clustering Technique based on Maximum Distance of Data and Lloyd k-means Algorithm", *Procedia Technology*, Volume 1, 2012.

[31] G. Grekousis and H. Thomas, "Comparison of two fuzzy algorithms", *geodemographic segmentation analysis: The Fuzzy C-Means and Gustafson–Kessel methods*, Volume 34, 2012.

[32] J. Aguilar and R. Lopez De Mantaras, "The process of classification and learning the meaning of linguistic descriptors of concepts". *Approximate reasoning in decision analysis*, North Holland, 1982.

[33] Robert M. R. Barclay, "Bats are not birds – A cautionary note on using echolocation calls to identify bats: a comment", *Department of Biological Sciences, University of Calgary, Journal of Mammalogy*, vol. 80, 1999.

[34] Carol Bedoya, Isaza Claudia, Daza Juan M. and Lopez Jose D., "Automatic recognition of anuran species based on syllable identification". Ecological Informatics 24, 2014.

[35] Titley Scientific, "Anabat Systems". <https://www.titley-scientific.com/us/products/anabat-systems>. Mayo 5, 2020.

[36] Manrique Lina M. Registran nuevas especies de murciélagos para Colombia. El espectador. Publicado en <https://www.elespectador.com/>, 28 de junio de 2019.

[37] Charlotte L. Walters Robin. Freeman Alanna Collen. Christian Dietz M. Brock Fenton Gareth Jones Martin K. Obrist Sébastien J. Puechmaille. Thomas Sattler Björn M. Siemers Stuart Parsons Kate E. Jones; et al. "A continental-scale tool for acoustic identification of European bats". British Ecological Society, 2012.

[38] Rumelhart, D., Hinton, G. & Williams, R. Learning representations by back-propagating errors. Nature, 1986.

[39] Huang, Guang-Bin; Zhu, Qin-Yu; Siew, Chee-Kheong. "Extreme learning machine: theory and applications". Neurocomputing, 2006.

[40] Widrow, Bernard; et al. "The no-prop algorithm: A new learning algorithm for multilayer neural networks". Neural Networks, 2013.

[41] The R Foundation. "The R Project for Statistical Computing". The R developer's team.