Facultad de Ingeniería y Computación

Escuela Profesional de Ingeniería de Telecomunicaciones

"Extracción de características basada en NMF para la clasificación de especies de aves usando señales de audio"

Presentado por:

Raisa Soledad Quispe Soncco

Para Optar por el Título Profesional de:

INGENIERÍA DE TELECOMUNICACIONES

Orientador: "Dr. Jimmy Diestin Ludeña Choez"

Arequipa, noviembre de 2017

PROGRAMA PROFESIONAL DE INGENIERÍA DE TELECOMUNICACIONES

Extracción de características basada en NMF para la clasificación de especies de aves usando señales de audio

Autor: Raisa Soledad Quispe Soncco

Noviembre, 2017

"Poca ciencia aleja muchas veces de Dios,
y mucha ciencia conduce siempre a Él"
Sir Francis Bacon, 1561- 1626

Índice general

Índice de figuras

Índice de cuadros

Summary

Usually for audio classification systems for example to the birds species acoustic classification, techniques of parameterization based on the Mel Frequency Cepstral Coefficient (MFCC) are used in the features extraction phase. However, it happens that although this technique provides good results, it is not quite adequate, because it was created for the Automatic Speech Recognition (ASR).

In this thesis, it is sought to improve the feature extraction process by means of a new parameterization using the method based on the Non-negative Matrix Factorization (NMF), specifically in the improvement of the conventional Mel-scale filter bank, used to obtain the cepstral coefficients. NMF has proved to be a fundamental tool for the representation of audio signals.

Experimental results have shown that learning the NMF-based auditory filter bank, compared to the Mel-scale filter bank, provides better classification rates, considering a classification scheme based on Support Vector Machine (SVM).

Resumen

Habitualmente para sistemas de clasificación de audio, por ejemplo, para la clasificación acústica de especies de aves, las técnicas de parametrización basadas en los Coeficientes Cepstrales a escala de Frecuencias Mel (MFCC) se usan en la fase de extracción de características. Sin embargo, sucede que aunque esta técnica proporciona buenos resultados, no es muy adecuada, ya que fue creada para el reconocimiento automático de la voz humana (ASR).

En esta tesis de grado, se busca mejorar el proceso de extracción de características mediante una nueva parametrización utilizando el método basado en la Factorización de Matrices No Negativas (NMF), específicamente en la mejora del banco de filtros convencional a escala Mel, utilizado para obtener los coeficientes cepstrales. NMF ha demostrado ser una herramienta fundamental para la representación de señales de audio.

Los resultados experimentales han demostrado que el aprendizaje del banco de filtros auditivo basado en la técnica NMF, en comparación con el banco de filtros a escala Mel, proporciona mejores tasas de clasificación, considerando un esquema de clasificación basado en la Máquina de Vectores de Soporte (SVM).

1 Introducción

A diario, convivimos con una infinidad de sonidos, producidos ya sean por una alarma al despertarse en las mañanas, o por mezclas de las voces de las personas, bocinas de autos, ruidos de las construcciones aledañas, entre otros; sin embargo, pocas veces nos percatamos de que existen sonidos producidos por la naturaleza desde que despertamos, para ser más precisos, por las aves, animales que no sólo nos ofrecen sus cantos a diario, sino que además, a lo largo de la historia humana, han ocupado un papel significativo en el arte, la cultura popular y mitos religiosos.

Las aves se encuentran distribuidas por todo el planeta, existiendo actualmente alrededor de 10500 especies de aves en el mundo [Navarro-Sigüenza et al., 2014], por lo que resultan ser uno de los seres vivos más extendidos sobre el planeta. El estudio de éstas, es un campo de gran interés, ya que sin ellas, el equilibrio del ecosistema no podría mantenerse y por ende la vida humana también se vería afectada. Una de las áreas de estudio de las aves es, la clasificación de éstas por sus sonidos, lo que es de gran importancia para la investigación biológica y aplicaciones de monitoreo ambiental, especialmente en la detección y localización de animales, que sirven para conocer el estado y las tendencias de la biodiversidad, identificar especies vulnerables y en riesgo de extinción, evaluar el impacto de la actividad humana sobre ecosistemas naturales, hacer seguimiento de especies invasoras, entre otros [Marsh and Trenham, 2008]. Una manera común de que los biólogos evalúen el impacto ambiental de las actividades humanas sobre los animales es detectando, localizando, identificando, y calculando la cantidad de animales en un determinado lugar.

Para poder realizar un buen estudio sobre el sonido de las aves, es también necesario, entender el origen de éstos, los que vienen dados por el funcionamiento del sistema de producción del sonido de las aves, principalmente producido por la siringe, órgano complejo en estructura y funcionamiento. Igualmente, para la producción final del sonido del ave, también actúan una diversidad de órganos dentro de su sistema respiratorio, razón por la cual, de los sonidos que encontramos en la naturaleza, pocos exhiben la diversidad y la estructura que se encuentra en el canto de las aves.

Dichos cantos son sonidos muy complejos y armónicos en estructura, pero no hay que olvidar que las aves también emiten sonidos cortos y no musicales, éstos son conocidos como llamados. Los llamados cumplen funciones tales como: de alarma, de vuelo, territoriales, etc. Generalmente los cantos son utilizados por los machos para cortejar a las hembras y poder lograr aparearse.

Tanto los cantos como los llamados están divididos jerárquicamente en frases, sílabas y elementos o notas; siendo las sílabas las estructuras base escogidas en varios estudios para el reconocimiento de cantos de aves por sus sonidos acústicos, ya que presentan la suficiente y fundamental información para poder detectar diferencias entre una u otra especie y así poder clasificarlas.

Cabe resaltar que en esta tesis de grado se hace uso del algoritmo NMF para el aprendizaje de un banco de filtros auditivo, a partir de las señales producidas por las vocalizaciones de las aves, mostrando así una mejora con respecto a uso de las características basadas en Coeficientes Cepstrales a escala de Frecuencia Mel (MFCC).

1.1. Motivación y Contexto

Existen una variedad de métodos que permiten a los investigadores en el campo de la ornitología, recolectar datos de manera ordenada y eficiente sobre las aves. Uno de ellos son los censos [Botero et al., 2005]; que consisten en determinar cuántas especies de aves hay en un área o en una región determinada, permitiendo estudiar el comportamiento de las mismas.

Dicho método se utiliza para obtener información valiosa sobre las aves, permitiendo diseñar programas adecuados para su conservación. Por ejemplo, con los censos se puede obtener información para promover la conservación y declarar en reserva a aquellas áreas o regiones donde habitan una gran cantidad de aves.

No obstante, éste método presenta diversas desventajas y requiere de personas especialistas en el campo, entre ellas podemos mencionar:

- Las personas que lo realizan deben recorrer un determinado trayecto caminando y reconocer las especies de manera visual, por lo que requiere de una cantidad considerable de tiempo.
- En caso de no identificar alguna ave, es posible que la persona deba realizar un dibujo para comparar con las ilustraciones de las guías de campo, o grabar su canto para consultar a un experto.

Como se puede notar, el método del censado requiere de tiempo y conocimiento, hecho que puede fácilmente ser enfrentado con la automatización del censado de las aves en una determinada área, ya que para poder declarar una zona en reserva y conservarla, hace falta realizar censados periódicamente, dependiendo de las características de la zona y en determinadas horas [Botero et al., 2005], existiendo áreas que resultan de difícil acceso para las personas, por lo que existe la posibilidad de la privación del censado en dichas áreas y de muchas especies de aves.

Por lo tanto, se requiere del desarrollo de nuevos sistemas de clasificación de las especies de aves, de tal manera que estos sean automáticos y sin perturbar el hábitat del ave, como por ejemplo el uso de grabaciones de audio (señales de audio).

1.2. Planteamiento del problema

Como se mencionó en la sección 1.1, uno de los mayores desafíos en la ornitología es poder prever la cantidad de especies de aves en una determinada área o región como también sus áreas de acción, y es que desde hace mucho tiempo se vienen realizando diversos estudios sobre las aves por medio de sus vocalizaciones, es decir, en los cantos y llamados que éstas emiten. Existen ciertas características que no se comparten entre una u otra especie, por lo que éstas diferencias mientras más notorias sean, de mejor manera permitirá la clasificación de especies de aves, y por ende un censado mucho más preciso. Es entonces que, existe la necesidad de establecer nuevas estrategias y técnicas cada vez más avanzadas que permitan progresar en la obtención de éstas diferencias, ya que la clasificación de las especies de aves a través de sus sonidos continúa siendo de forma manual y depende de la experiencia del investigador, lo que implica un desgaste de tiempo y limitación en cuanto a la experiencia del censador.

La técnica que es habitualmente utilizada para diferenciar dichas vocalizaciones se encuentra basada en el sistema auditivo humano, y por ende es adecuada para el tratamiento del habla humano, sin embargo, no resulta ser muy adecuada para el tratamiento de señales bioacústicas como son las de las aves; por tal motivo, el uso de la técnica de la Factorización de Matrices No-Negativas (*Non-Negative Matrix Factorization*, NMF), podría ser una solución mucho más adecuada para mejorar la diferenciación de dichas vocalizaciones, debido a que proporciona una representación a partir de los datos, acentuando la representación en el rango de frecuencias más relevante en dichas señales.

1.3. Objetivos

1.3.1. Objetivo general

Desarrollar un sistema de extracción de características, basado en la Factorización de Matrices No-Negativas (*Non-Negative Matrix Factorization*, NMF), para la clasificación de diferentes especies de aves usando señales de audio.

1.3.2. Objetivos específicos

- Diseñar un banco de filtros auditivo basado en la técnica de la Factorización de Matrices No-Negativas (NMF) para la extracción de características de las señales de audio.
- Comparar el método propuesto con la técnica convencional que son los Coeficientes Cepstrales a escala de Frecuencias Mel (*Mel Frequency Cepstral Coefficients*, MFCC).

- Realizar la etapa de la clasificación basados en el algoritmo del aprendizaje supervisado de las Máquinas de Vectores de Soporte (*Support Vector Machine*, SVM).

1.4. Metodología

1.4.1. Base de datos

La base de datos consiste de grabaciones de audio de 12 diferentes especies de aves: *Aramides cajanea(Cotara chiricote), Coereba flaveola (Platanero), Colibri thalassinus (Colibrí verdemar), Crypturellus cinereus (Tinamú sombrío), Crypturellus obsoletus (Tinamú café), Crypturellus soui (Tinamú chico), Crypturellus undulatus (Tinamú ondulado), Lathrotriccus euleri (Mosquero de euler), Piranga olivacea (Piranga escarlata), Piranga rubra rubra (Piranga roja), Rupornis magnirostris (Busardo caminero), Synallaxis azarae (Pijuí de azara)*. Dichas especies de aves habitan en Sudamérica y han sido obtenidas de la audioteca acústica: Xenocanto [Foundation, 2015].

La base datos fue dividida en dos subconjuntos; uno para entrenamiento y otro para prueba, que será descrito posteriormente. Por otra parte, las grabaciones de audio inicialmente fueron grabadas usando diferentes frecuencias de muestreo (F_S), por lo que se procedió a remuestrearlas a una única $F_S = 22050\,Hz$. Esta misma base de datos, será descrita a detalle posteriormente, en la sección 4.1.

1.4.2. Procesamiento de la señal bioacústica

- **Extracción de características**

Para la etapa de la extracción de características de las señales bioacústicas, convencionalmente se hace uso del método basado en los MFCC, que usa un banco de filtros auditivo inspirado en el sistema auditivo humano, motivo por el cual podría no ser suficientemente adecuado para señales de las vocalizaciones provenientes de las aves. En esta tesis de grado se propone el uso de NMF para obtener un banco de filtros a partir de los datos, siendo así posible que los filtros enfaticen el rango de frecuencias que contienen la mayor cantidad de la energía espectral de la señal bioacústica, lo que se traduce en una mejor representación de las señales a clasificar.

- **Entrenamiento para la clasificación**

Para esta etapa se hará uso del algoritmo SVM, el funcionamiento y teoría de ésta técnica se detallará en la sección 2.4. Investigaciones anteriores demuestran que se han obtenido buenos resultados con el uso de las SVM (*Support Vector Machine*).

Las características que representan a las señales son las entradas al clasificador, es decir, de la etapa de extracción de características se procede a la etapa del entrenamiento para la clasificación que permite la creación de modelos que representan adecuadamente a cada una de las especies.

Para el entrenamiento del clasificador, se hizo uso del conjunto de datos de entrenamiento, mientras que el conjunto de prueba fue usado para evaluar el sistema de extracción de características por medio de los resultados finales de clasificación.

1.4.3. Evaluación del rendimiento del sistema

Para la evaluación del rendimiento del sistema, se realizó una comparación entre los resultados obtenidos por medio del uso del método convencional de los Coeficientes Cepstrales a escala de Frecuencias Mel (MFCC) y los obtenidos haciendo uso de la Factorización de matrices no negativas (NMF), a los mismos que denotaremos como Coeficientes Cepstrales basados en NMF (NMF_CC).

Los resultados obtenidos están en función de tasas de clasificación (T_C) de las diferentes especies de aves. Dichas tasas de reconocimiento son los resultados en porcentaje de los datos de evaluación que fueron correctamente reconocidos del total de los datos escogidos para evaluación de la base de datos. La fórmula utilizada es la siguiente:

$$T_C = \frac{M_C}{T_M} \times 100\,\% \tag{1.1}$$

Donde:

T_C: Tasa de clasificación.

M_C: Muestras correctamente clasificadas.

T_M: Total de muestras.

2 Marco Teórico

2.1. Las aves

Las aves constituyen el grupo de animales más diverso dentro de los vertebrados terrestres, usan sus extremidades traseras para desplazarse, ya sea en tierra o en agua, mientras que las extremidades delanteras evolucionaron hasta transformarse en alas.

Estos animales poseen características muy variadas, como la presencia de plumas y pico, su reproducción ovípara, su capacidad de volar, entre otras. Aunque no todas las aves puedan volar, las que tienen dicha capacidad sufren ciertas adaptaciones, éstas presentan huesos muy porosos para pesar menos y así emplear menor energía durante el vuelo; tienen plumas que son muy ligeras y cumplen dos funciones: de abrigo y vuelo; abrigo para aquellas que vuelan a grandes alturas donde hace bastante frío y de vuelo referido a las plumas largas y rígidas que se encuentran en las alas.

Cuando las aves vuelan, realizan un gran esfuerzo muscular, sus músculos consumen bastante oxígeno y por ende el corazón debe latir hasta 300 pulsaciones por minuto, lo que ocasiona que la temperatura corporal de las aves ronde los 40°C. Además el aparato respiratorio se encuentra adaptado a una ventilación rápida debido a que poseen unos sacos aéreos que funcionan como reserva de aire, lo que les permite tener un variado repertorio de cantos y llamados dependiendo de la cantidad de sacos aéreos, éstos dan lugar a la regulación térmica del animal, para equilibrar el exceso de calor producido al volar [Lafuente, 2011].

En cuanto a la forma en que las aves adquieren su repertorio de canto, se encuentran divididas en dos grandes grupos: las suboscines y las oscines; las suboscines son aquellas aves que adquieren su repertorio de canto netamente a través de mecanismos genéticos, es decir, son 100 % innatos, en cambio, las oscines son aquellas aves que adquieren su repertorio mediante mecanismos genéticos y procesos de aprendizaje similares a la adquisición del habla en humanos [Ridgely and Guy, 1989, Ridgley and Tudor, 1994].

Tomadas en conjunto, las aves sirven para ilustrar como está distribuida la biodiversidad y son valiosas indicadoras del cambio ambiental a nivel mundial, debido a que reaccionan de forma rápida y muy visible a cualquier alteración en su medio. Por eso, contribuir a su conservación significa que también se está velando por nuestro propio bienestar y nuestro propio futuro.

2.1.1. Clasificación de las aves

En biodiversidad, se considera hasta el momento que son más de 5 millones de diferentes especies [Gandolfi, 2004] las que constituyen la población biológica del planeta, por lo que los científicos se encuentran con la necesidad de clasificar, estudiar e intercambiar información sistemática acerca de tal variedad de organismos. Para hacer esto, se debe de disponer de un sistema para nombrar a todos estos organismos y así agruparlos en forma ordenada y lógica. Pero la inclusión dentro de un grupo supone previamente la observación de las características típicas, descripción de las mismas y comparación con las de otros seres vivos conocidos. Este procedimiento se denomina "clasificación" y "taxonomía" (rama biológica que se ocupa de ordenar la diversidad biológica, formando un sistema de clasificación), tomando en cuenta no sólo la forma externa de los organismos, sino también funciones vitales, como la alimentación, la respiración, la reproducción y, hoy en día a los nuevos estudios sobre la biología molecular [Gandolfi, 2004].

La taxonomía establece normas para construir clasificaciones sobre las aves, que generalmente son jerárquicas, es decir hay categorías que se incluyen unas a otras. Como se puede ver en la Figura 2.1, las categorías que le siguen a la clase de las aves son: orden, familia, género y especie.

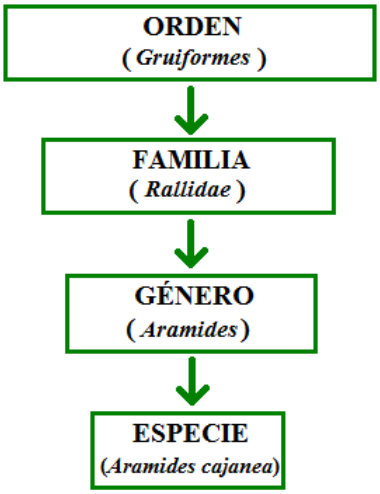

Figura 2.1: Taxonomía de las aves, con ejemplo de la especie *Aramides cajanea*.

Por otro lado, la sistemática estudia las relaciones entre los seres vivos, buscando ordenar la diversidad biológica. La sistemática filogenética estudia la formación sucesiva de las especies, es decir, cómo se originaron unas de otras y sus relaciones de parentesco. Se considera que familias cercanas sistemáticamente, también son similares en aspecto. Pero a veces ocurre un fenómeno llamado "convergencia evolutiva", haciendo posible que familias alejadas desde el punto de vista sistemático sean parecidas externamente por tener un modo de vida similar.

Las clasificaciones naturales toman muchos caracteres y así dan una mejor idea del real parentesco entre los diferentes taxones. Se basan principalmente en caracteres externos y anatómicos de las especies vivientes: paladar, huesos nasales, forma y disposición de las narinas, número de plumas de alas y cola, disposición de las escamas en los tarsos, sistema de tendones, músculos de la siringe [Bosso et al., 2009].

Finalmente, de acuerdo a la taxonomía de las aves, se presenta una distribución esquemática de las 12 especies de aves consideradas en esta tesis de grado en la Figura 2.2.

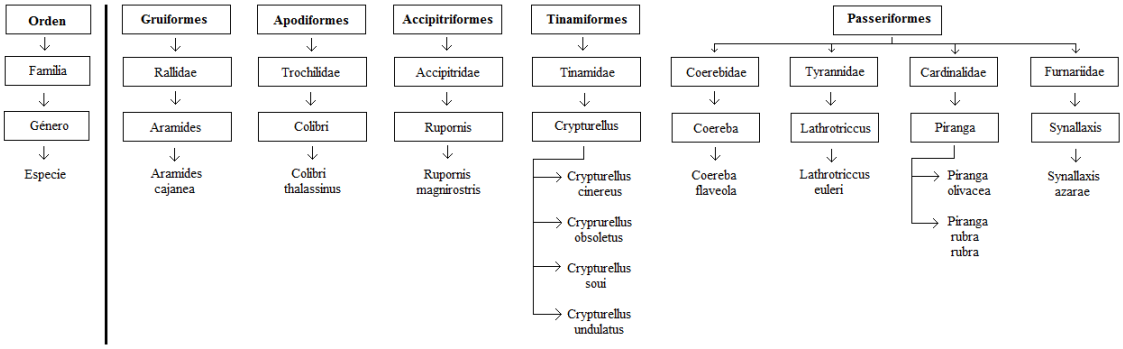

Figura 2.2: Taxonomía de las 12 especies de aves consideradas en esta tesis de grado.

2.1.1.1. Nombre científico y común

El principal objetivo de la nomenclatura científica es dar a cada especie un nombre único y universal, que permita distinguirla inmediatamente de cualquier otra. El uso del latín, al uniformizar el idioma, contribuye a evitar rivalidades localistas.

El mosquero de euler, por ejemplo, tiene varios nombres comunes en su amplia distribución geográfica: mosquerito de euler (Perú), atrapamoscas de sotobosque (Venezuela), atrapamoscas de Euler (Colombia), mosqueta de Euler (Argentina), mosqueta parda (Paraguay), mosqueta de monte (Uruguay). El nombre científico en cambio, como se dijo, es único y universal. Así, todos aquellos nombres del mosquero de euler que fueron mencionados pueden reemplazarse por uno solo, conocido en todo el mundo como: "*Lathrotriccus euleri*".

2.1.1.2. El sistema de Linneo

El creador de la clasificación y nomeclatura moderna de los animales y vegetales fue el sueco Carl von Linné (1707-1778), se le conoce con el nombre castellanizado de Linneo. En su célebre obra "Systema Naturae" (1735) estableció las bases del sistema de clasificación hoy vigente.

Según este sistema las categorías taxonómicas se ordenan jerárquicamente y reciben un nombre científico en latín. Las especies se agrupan en géneros, los géneros en familias, las familias en órdenes, los órdenes en clases, las clases en tipos y los tipos en reinos. Para el sistema de Linneo la categoría básica es la especie, cuyo nombre científico es binomial; es decir se emplean dos vocablos para su denominación. Por encima del rango de especie, todas las demás categorías taxonómicas se denominan con un solo vocablo (uninominales) [Bosso et al., 2009].

Es importante mencionar que, los individuos de una especie pueden subdividirse en grupos más pequeños denominados subespecie o raza geográfica, que exhiben pequeñas diferencias en sus caracteres y tienen denominación trinomial, es decir un nombre compuesto por tres vocablos. Cada subespecie tiene distribución geográfica diferente y los individuos que viven en zonas limítrofes de esa distribución suelen presentar caracteres intermedios [Bosso et al., 2009].

Cabe resaltar que en esta tesis de grado se realizará la clasificación de especies de aves, tomando en cuenta la definición de "especie" según el sistema de Linneo, es decir la de denominación binomial (dos vocablos), por lo que no se considerarán a las subespecies.

2.1.2. Diversidad de las aves

Existen aves en casi cualquier parte del mundo, éstas existen prácticamente en todos los hábitats, desde los desiertos más bajos hasta las montañas más altas. Los patrones de diversidad de aves son dirigidos por factores biogeográficos fundamentales, en donde los países tropicales, sobre todo en Sudamérica, son los que albergan la mayor riqueza de especies, tal como se muestra en la Figura 2.3.

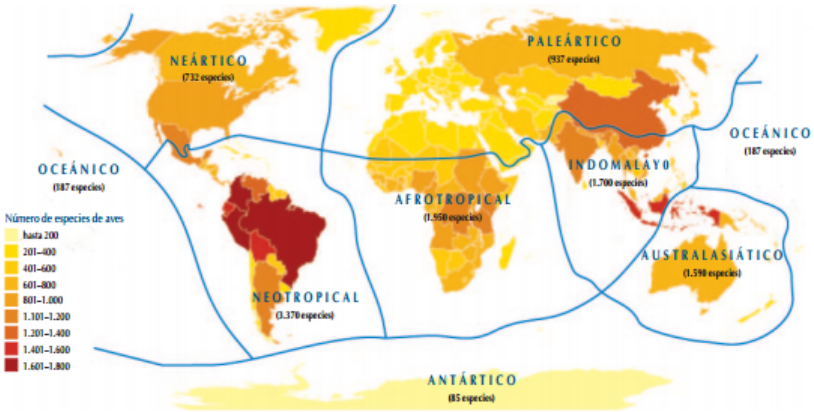

Figura 2.3: Distribución de aves del mundo según el ámbito geográfico y el país [Alison Stattersfield, 2008].

Actualmente las aves comprenden casi 10500 especies [Navarro-Sigüenza et al., 2014], lo que las hace piezas clave de la biodiversidad, aunque se pueden encontrar en cual-

quier lugar, cada especie es única en cuanto a su ecología y distribución. Muchas tienen pequeñas áreas de distribución y la mayoría están restringidas a ciertos tipos de hábitats.

América Latina y el Caribe son las regiones con mayor diversidad biológica en el planeta y alberga varios de los países megadiversos del mundo. La región en conjunto posee el 41 % de las aves y los niveles de endemismo[1]son muy altos. De los 10 países megadiversos; Colombia, Perú y Brasil cuentan con 1860, 1835 y 1822 especies de aves respectivamente [FRANCO et al., 2009, Plenge, 2010, Pedro F. Develey, 2009]. En consecuencia, son los 3 países con mayor cantidad de especies de aves en el mundo.

Inclusive, se ha demostrado que las áreas más importantes para las aves en todo el planeta identificadas por BirdLife (organización dedicada a la contribución del conocimiento científico y a la conservación de las aves), y que son conocidas como áreas importantes para la conservación de las aves (IBA, *Important Bird and Biodiversity Areas*), cubren hasta el 80 % del resto de la biodiversidad mundial.

En el caso específico de Perú, su diversidad geográfica y topográfica lo ha ubicado como el segundo país con mayor biodiversidad y densidad de aves en el mundo. Cuenta con 128 de las áreas más importantes para la conservación de las aves (IBAs). Asimismo, cuenta con 115 especies endémicas de aves [SERNANP, 2012].

La distribución geográfica de las 12 especies de aves en el Perú, consideradas en esta tesis de grado, se muestran la Figura 2.4. De acuerdo a la categoría taxonómica de género; la Figura 2.4 (a) muestra a las especies correspondientes al género *Crypturellus*, que son: *Crypturellus cinereus, Crypturellus soui, Crypturellus undulatus* y *Crypturellus obsoletus,* como se puede ver su hábitat se encuentra al este del Perú, ya que generalmente prefieren vivir en bosques tropicales o pantanosos.

La Figura 2.4 (b) muestra a las especies: *Aramides cajanea, Synallaxis azarae* y *Rupornis magnirostris,* se puede ver que el hábitat de la especie *Aramides cajanea,* que pertenece al género *Aramides,* se encuentra al este del Perú, generalmente prefiere vivir en bosques pantanosos, en cambio, se puede ver que la especie *Synallaxis azarae,* que pertenece al género *Synallaxis,* habita en la región sierra del Perú, debido a que mayormente prefiere los bosques de montaña, por último la especie *Rupornis magnirostris,* que pertenece al género *Rupornis,* también se encuentra al este del Perú, donde abundan los bosques densos preferidos por esta especie.

Seguidamente la Figura 2.4 (c), muestra la distribución geográfica en el Perú de las especies: *Coereba flaveola, Colibri thalassinus* y *Lathrotriccus euleri,* se puede ver que el hábitat de la especie *Coereba flaveola,* que pertenece al género *Coereba,* se encuentra divido por un lado al este y por el otro al oeste del Perú, donde habita en áreas abiertas y bordes de los bosques. La especie *Colibri thalassinus,* que pertenece al género *Colibri,* habita los bosques montañosos de la sierra del Perú. Y la especie *Lathrotriccus euleri,* que pertenece al género *Lathrotriccus,* habita al este del Perú donde abundan los bosques densos preferidos por esta especie.

[1]Endémico: Propio y exclusivo de determinadas localidades o regiones.

Finalmente la Figura 2.4 (d), muestra a dos especies correspondientes al género *Piranga, que son la Piranga olivacea* y *la Piranga rubra rubra,* como se puede ver su hábitat cubre las regiones sierra y selva del Perú, estas especies de aves prefieren los bosques que les proporcionen sombra, como los bosques tropicales y mixtos [Kaufman, 2001].

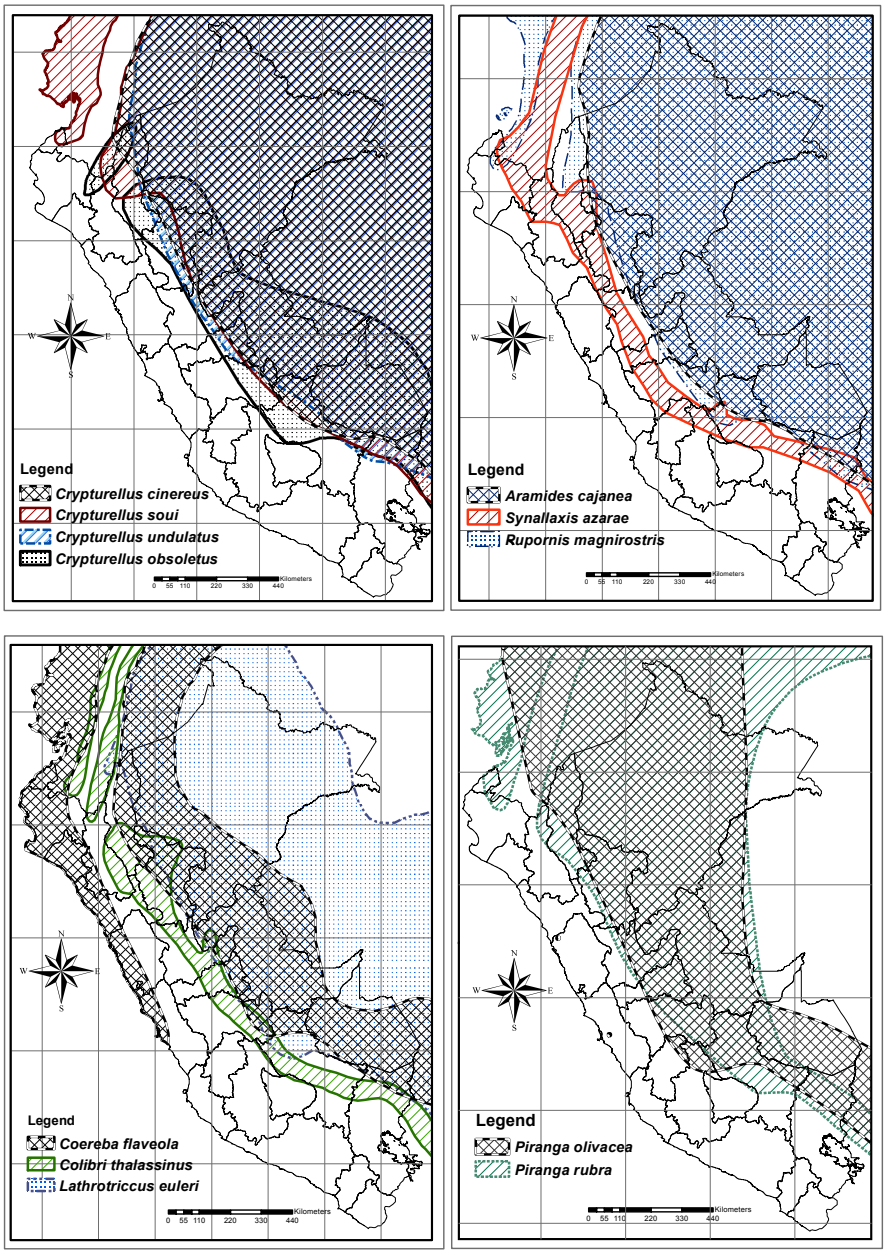

Figura 2.4: Distribución geográfica de 12 especies de aves en el Perú [BirdLife, 2015].

2.1.3. Monitoreo biológico de las aves

El monitoreo biológico de las aves consiste en determinar la variación de las poblaciones de aves a lo largo del tiempo, uno de los métodos para poder llevar a cabo este objetivo es el censo de las aves, no solo por el estudio de las variaciones en su población sino porque también permiten establecer el parámetro más directo para determinar la probabilidad de extinción de una especie, ya que cuanto más escaso es un taxón, más probabilidad tiene de desaparecer, lo que indicaría que en los lugares donde se determina un incremento de la desaparición de una o más especies probablemente existan factores que amenazan la conservación de éstas y seguidamente podrían ser tomadas acciones para detener las extinciones de especies de aves. Por otra parte, el monitoreo también permite conocer la relación que existe entre las aves y su ambiente. De esta manera, el monitoreo enfocado en la conservación y el conocimiento de las aves es fundamental para el buen funcionamiento de los ecosistemas y el bienestar social de la población humana.

2.1.3.1. Método del censado

Los censos o conteos se utilizan para conocer cuántas especies de aves hay en un área o en una región. Pueden utilizarse diferentes técnicas, según el tiempo disponible o las características de la zona. La persona que realiza el censo debe reconocer las especies de la zona, con base a sus formas, colores o cantos. En caso de no identificar alguna ave, es posible que la persona realice un dibujo para posteriormente compararlo con las ilustraciones de las guías de campo, o grabar su canto para consultar a un experto. Durante el censo, el observador cuenta todas las aves que ve o escucha en un período de tiempo determinado y preferiblemente antes de las 10 am cuando las aves están más activas, debido a que los machos prefieren emitir cantos a tempranas horas. Según la técnica elegida, es recomendable realizar varios puntos o transectos en la zona de estudio.

- **Censos a lo largo de transectos:** en esta técnica el observador camina a velocidad constante a lo largo de una línea que cruza la zona de interés. Esa línea, llamada transecto, puede ser un camino que atraviese un área. Su longitud puede estar entre los 100 a 500 m y puede tener ancho fijo o variable. En los transectos de ancho fijo sólo se registran las especies vistas a una distancia específica (por ejemplo 25 m) y en los de ancho variable se cuentan las aves observadas a cualquier distancia del transecto.
- **Censos desde puntos de radio fijo:** aquí el observador se sitúa en el centro de un círculo imaginario de 25 m de radio y realiza el conteo durante 10 minutos. Es importante que el observador se asegure que entre los centros de los puntos haya una distancia mínima de 150 m como se puede observar en la Figura 2.5.

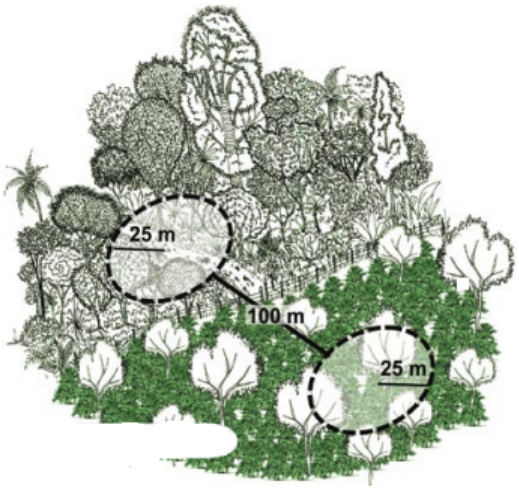

Figura 2.5: Censos desde puntos de radio fijo [Botero et al., 2005].

- **Censos de búsqueda intensiva:** este método consiste en que el observador recorra un área determinada sin seguir una trayectoria fija para localizar, contar e identificar aves. Debe efectuar una serie de tres censos de 20 minutos cada uno, en tres áreas distintas de 3 hectáreas cada una si se encuentra en zona de bosque y 10 hectáreas o más en zonas abiertas. Las áreas pueden ser colindantes o estar completamente separadas. Además, se deben censar las mismas áreas a lo largo de todo el año.

- **Censos de mapeo de parcelas:** este otro método, está basado en la conducta territorial de las aves y consiste en marcar sobre un plano la posición de las aves observadas en visitas consecutivas a la parcela de estudio a lo largo de la temporada reproductora. La finalidad es determinar el número de territorios y estimar la densidad de machos reproductores por especie en el área. Uno o dos observadores visitarán parcelas determinadas como mínimo 8 veces durante la temporada reproductora. El tiempo necesario dependerá del tamaño y de las características del terreno en el área de estudio, así como de la densidad de aves [Ralph et al., 1996].

De los cuatro métodos descritos, el método de los censos desde puntos de radio fijo suele ser el más apropiado en la mayoría de los casos y ha sido adoptado como método estándar de monitoreo [Ralph et al., 1995]. En suma, con los censos se obtiene información valiosa acerca de las especies presentes en un lugar, sobre el hábitat que ellas prefieren o si son comunes o escasas. También sirven para realizar comparaciones entre las aves presentes en diferentes lugares o en diferentes épocas.

2.2. Bioacústica en las aves

Como se explicó en el apartado anterior, censar aves de la manera tradicional es un trabajo muy laborioso e incluso subjetivo, debido a que depende del observador identificar a una determinada ave y posteriormente clasificarla dentro de una especie. Es por tal motivo, que existe un área de investigación científica mediante la cual se puede estimar el área de acción y censo de las aves a través de sus vocalizaciones, dicha área lleva por nombre "Bioacústica". La Bioacústica es un campo multidisciplinario que conjuga la Biología y la Acústica, dedicada esencialmente a la investigación de la producción y recepción del sonido biológico, así como los mecanismos de transmisión de información biológica (señales) por vínculos acústicos y la propagación de ésta en los diferentes ambientes: sólidos, líquidos y gaseosos.

2.2.1. Mecanismo de producción del sonido del ave

El sonido de las aves se encuentra caracterizado por dos tipos de vocalizaciones: los cantos y los llamados; los cantos son producto del aprendizaje y del carácter conductual que principalmente sirven para la atracción sexual y están asociados a los machos, en general tienen una estructura mucho más larga y compleja, mientras que los llamados son señales de carácter conductual que tienen un carácter acústico más sencillo y sirven de vuelo, para dar alarma, de defensa de territorio, entre otros. A continuación se relatará sobre cómo es que se producen dichos cantos y llamados, así como también la forma en que se encuentran organizados.

Para la explicación del mecanismo de producción del sonido del ave, se partirá por explicar el sistema que origina dichos sonidos: el sistema fonador del ave, que se encuentra compuesto principalmente por su sistema respiratorio y el tracto vocal del ave. Aunque el tracto vocal de las aves es diferente al humano, existen similitudes en cuanto a su anatomía se refiere. La diferencia fundamental recae sobre el órgano que produce los sonidos en las aves, que es la siringe, mientras que en los humanos es la laringe, pero lo común en ambos, es el paso del aire por dichos órganos donde se generan vibraciones ya sea tanto en las membranas de la siringe en las aves, o en las cuerdas vocales en el caso de los humanos, y dichas vibraciones dan lugar a los diferentes sonidos en las aves y a la voz en los seres humanos. Sin embargo, los sistemas de producción de sonido y del habla, tanto en aves como en humanos, respectivamente, son mucho más complejos y es que en complejidad, el sistema de producción de sonido de las aves lo es más, razón por la cual los diferentes sonidos de las aves, ya sean cantos y llamados resultan ser variados y duraderos en tiempo, lo que para el humano es difícil de realizar, es decir las aves pueden emitir un sonido mucho más duradero que el del habla humano, cabe mencionar, que las aves emiten sonidos constantes en amplitud en el tiempo, mientras que el habla del humano decae en amplitud a través del tiempo. Veremos a continuación con mayor detalle cómo funciona el sistema de producción del sonido del ave.

Las principales partes en el mecanismo de la producción del sonido de las aves son: los pulmones, la siringe, la tráquea, la laringe, la boca y el pico. El aire que proviene desde los pulmones al igual que en los humanos, en el momento de la exhalación, se propaga a través de los bronquios hacia la siringe, donde se produce la principal fuente de sonido. A continuación el sonido producido en la siringe es modulado por el tracto vocal del ave, donde se encuentran la tráquea, la laringe, la boca y el pico. En la Figura 2.6 se muestran los órganos que participan en la producción del canto en el ave. Las dimensiones y partes de ésta figura, varían considerablemente dependiendo de las especies, sin embargo, la organización que presenta el esquema es muy uniforme a la mayoría de las especies de aves [Fagerlund, 2004].

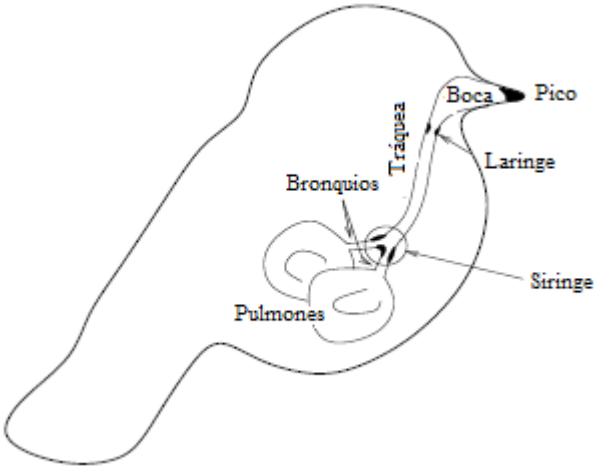

Figura 2.6: Partes y organización del mecanismo de la producción del sonido aviar [Fagerlund, 2004].

2.2.1.1. La Siringe

Es el órgano principal donde se genera el sonido, por lo que resulta de mucha importancia conocerlo a detalle. El anatomista alemán Müller, clasificó a las especies de aves de acuerdo a su anatomía siringeal limitando sus estudios a las Paseriformes [Müller, 1878]. Posteriormente Beddard tomó en cuenta un rango más amplio de especies de aves en sus estudios [Beddard, 1898]. Muchos estudios realizados después de los estudios de Müller y Beddard, confirmaron sus clasificaciones.

Se han podido encontrar tres diferentes tipos de siringes: la traqueo-bronquial, la traqueal y la bronquial, esta clasificación se debe a los distintos elementos que conforman la tráquea y el bronquio, y a la posición en que se encuentra el principal mecanismo de producción del sonido. Cuando dicho mecanismo se encuentra en el bronquio, puede localizarse en diferentes posiciones entre los dos bronquios. Los elementos de la tráquea que consisten en anillos de cartílago, normalmente se encuentran completos a lo largo de la tráquea, mientras que en el caso de los bronquios sus

elementos están emparejados de forma incompleta con anillos de cartílago en forma de C, con terminaciones abiertas, una frente a la otra [Fagerlund, 2004].

Las aves del orden Paseriformes y suborden Passeri, constituyen el mayor grupo de las aves que cubren alrededor de 4000 a más de 9000 especies de aves del total [Catchpole and Slater, 1995]. Dentro de éste grupo de aves, la siringe donde se producen sus cantos, es muy compleja en estructura, sin embargo, también es muy uniforme para las aves dentro de éste grupo [King, 1989], por lo que fue tomada como un prototipo que se muestra en la Figura 2.7.

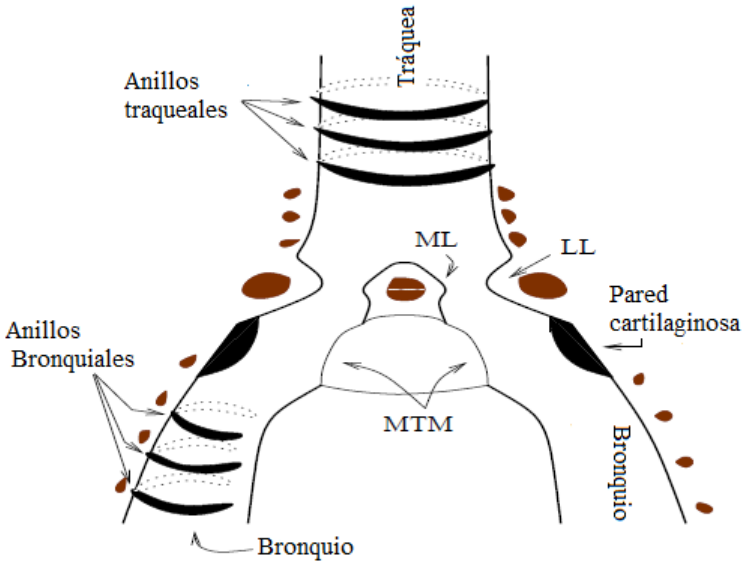

Figura 2.7: Vista esquemática de la siringe de las aves. [Fagerlund, 2004]

La siringe que está localizada en la unión entre la tráquea y los dos bronquios, se encuentra clasificada como siringe traqueo-bronquial. Sobre este tipo de siringe, existen dos teorías sobre cómo se produce el sonido de las aves, una está relacionada con la Membrana Timpaniforme Medial (*Medial Tympaniform Membrane*, MTM), mientras que la otra, se relaciona con los Labios Laterales (LL, Lateral Labia) y Labios Mediales (*Medial Labia*, ML), parecidos a las cuerdas vocales humanas.

La primera teoría explica que cuando el ave está cantando, el flujo del aire que proviene desde los pulmones, ocasiona que la MTM siringeal vibre en cada bronquio a través del efecto de Bernoulli [Fletcher, 1992], dichas vibraciones no lineales de la membrana son opuestas a la pared cartilaginosa. El sonido y el movimiento de la membrana están controladas por un par simétrico de músculos alrededor de la siringe, las membranas pueden vibrar independientemente una de la otra con diferentes modos y frecuencias fundamentales.

En cambio, la segunda teoría demostró por medio de una imagen endoscópica que la MTM podría no ser la principal fuente de sonido [Goller and Larsen, 1997b], Goller

sugiere que el sonido se produce por dos tejidos blandos, los ML y LL, similares a las cuerdas vocales humanas. El sonido se produce por el paso del flujo del aire a través de dichos tejidos vibratorios. Por otra parte, un estudio más reciente que consistió en remover quirúrgicamente la MTM [Goller and Larsen, 2002], demostró que después de dicha eliminación, las aves que ya no contaban con la MTM, podían emitir sonidos y casi cantar con normalidad. Aun así, fueron encontrados pequeños cambios en la estructura del canto, lo que indica que la MTM si cumple una función en la producción del sonido. Pero también, es posible que las aves puedan ser capaces de compensar la pérdida de la MTM [Fagerlund, 2004].

Como la siringe del ave tiene una estructura compleja y diversa, es posible que la primera teoría sobre la MTM sea correcta para algunas especies, debido a que [Goller and Larsen, 2002], limitó sus estudios a dos especies de aves: el cardenal rojo (*Cardinalis cardinalis*) y el diamante mandarín (*Taeniopygia guttata*). Por otro lado, [Gaunt et al., 1982], realizó un estudio sobre palomas, a manera de evidenciar la teoría de las MTM. Además en [Goller and Larsen, 1997a], se encontró que la principal fuente de sonido en pichones y palomas, es la membrana timpaniforme (MTM), no obstante, dicha membrana está localizada en la tráquea y no en el bronquio [Fagerlund, 2004].

2.2.1.2. Tráquea

La tráquea en las aves es parecida a un tubo que ese encuentra entre la siringe y la laringe, funciona como un resonador para el sonido producido por la siringe. Entre sus elementos, se encuentran los anillos cartilaginosos, que están normalmente completos [McLELLAND, 1989]. Como entre las especies de aves pueden encontrarse aves muy grandes como muy pequeñas, presentando anatomías diferentes entre ellas, la cantidad de anillos cartilaginosos dentro de la tráquea de las aves depende del tamaño de sus cuellos, que aproximadamente, están entre 30 anillos para pequeños paserines y alrededor de 350 anillos a lo largo del cuello de los flamencos y grullas. Ahora bien, en muchas de las especies, la tráquea es un arreglo de bucles o bobinas, por lo que, el tamaño de la tráquea resulta mucho mayor que el tamaño del cuello. Por otro lado, se ha discutido sobre si los bucles de la tráquea del ave, mejoran la función de transferencia, ya que la tráquea tiene muchos modos diferentes de vibración [Gaunt et al., 1987]. En algunas especies, la tráquea se encuentra unida a lo sacos aéreos o expansiones bulbosas de las aves. En algunos pingüinos (*Spheniscidae*) y petreles (*Procellariidae*) la tráquea está fragmentada en dos canales, por lo que gracias a dicha característica, dichas especies tienen sonidos característicos [Fagerlund, 2004].

2.2.1.3. Laringe, boca y pico

Como sabemos, la laringe es el órgano principal para la producción del habla en los humanos, ya que en su interior se encuentran las cuerdas vocales. Por el contrario,

en el caso de las aves la laringe no contiene dichas cuerdas vocales, se tienen pocos estudios sobre la función de la laringe en las aves, por lo que su función en la producción del sonido es aún controversial, ya que al parecer la laringe parece solo jugar un pequeño rol o casi nada en la producción del sonido. Por otro lado, la cavidad oral en el caso de los humanos, realiza un trabajo de resonador y como filtro al sonido producido por la laringe; y en el caso de las aves, su boca también opera como una cavidad resonadora, pero es algo menos flexible. Las aves pueden controlar el área de sección transversal de la boca [Fletcher and Tarnopolsky, 1999] con la lengua, a pesar de ello, solo pocas especies principalmente los loros, pueden usar la lengua para producir sonidos como los humanos [Patterson and Pepperberg, 1994] ya que en la mayoría de las aves, la lengua es muy rígida.

En el caso del pico del ave, el análisis de su comportamiento acústico resulta muy difícil, debido a que la forma del pico de las aves es muy compleja. La apertura y cierre del pico, cambian las propiedades acústicas del mismo. Estudios recientes han demostrado la importancia del pico en la producción del sonido [Hoese et al., 2000], dicho estudio muestra que los cambios de la apertura y cierre del pico afectan en la longitud efectiva del tracto vocal, pero dichos efectos resonadores en el tracto vocal son no lineales, cabe resaltar también que otra forma de modificar el tamaño de su tracto vocal es por medio de movimientos de su cabeza [Westneat et al., 1993].

2.2.2. Estructura del sonido del ave

Como se ha mencionado, existe una diversidad de sonidos que las aves pueden emitir, normalmente éstos están divididos en dos categorías: cantos y llamados. Los cantos se encuentran limitado a las aves canoras[2], cubriendo aproximadamente la mitad del total de las aves. Aquellas aves que no cantan utilizan sus sonidos solo para fines comunicativos, por lo que generalmente son las aves canoras las que poseen mayor complejidad en la producción de sus sonidos como también diversidad en sus repertorios, esto debido a que la habilidad de poder controlar la producción de sus sonidos es mucho mejor [Gaunt et al., 1987].

Las principales características del sonido de las aves son: la frecuencia fundamental y sus armónicos. Los sonidos vocales de las aves se encuentran muy relacionados a los sonidos vocales humanos tanto en estructura como en la forma que son producidos. No obstante, el control sobre el tracto vocal de las aves es menos complejo que en los humanos. En los sonidos vocales de las aves la frecuencia fundamental recae entre 100 Hz y 1 kHz para diferentes especies [Fagerlund, 2004]. Las aves pueden enfatizar intensidades de los diferentes armónicos con propiedades de filtrado del tracto vocal.

Las aves pueden también producir tonos puros que no incluyen ningún armónico. Tanto los sonidos vocales como los tonos puros pueden ser modulados en frecuencia y amplitud. Las modulaciones en amplitud casi siempre son producidas por la siringe,

[2]Canora: Dicho de un ave de canto grato y melodioso.

pero las diferentes intensidades entre los armónicos están basados en las propiedades del tracto vocal. Las modulaciones en frecuencia se dividen en dos categorías: modulaciones de frecuencia continuas y saltos de frecuencia abruptos. Además los sonidos de las aves pueden también ser ruidosos, de banda ancha, o confusos en estructura [Fletcher and Tarnopolsky, 1999].

2.2.2.1. Teoría del doble sonido del ave

Debido a que las aves poseen dos membranas vibratorias independientes en la siringe, éstas pueden producir dos ondas portadoras totalmente independientes. Las diferentes especies de aves hacen uso de las fuentes de doble sonido de diferentes maneras, por ejemplo la especie de los canarios (*Serinus canarius*) utilizan una sola fuente siringeal para la producción del sonido, mientras que la especie del carbonero de capucha negra (*Parus atricapillus*) produce sonidos complejos mediante el uso de ambas fuentes [Fagerlund, 2004].

Existen tres métodos diferentes por los que se producen los sonidos: el sonido producido por cualquiera de las dos membranas, es decir solo una; el sonido producido por ambas membranas juntas o por intercambio de la fuente de sonido por una membrana a otra [Fletcher, 1992]. Cuando las aves hacen uso de ambas membranas pueden generar los mismos o diferentes sonidos, resulta muy común que algunas especies usen los tres métodos en la producción del sonido.

2.2.2.2. El canto de las aves

Como se estuvo mencionando, los cantos de las aves son vocalizaciones largas y complejas, producidas espontáneamente por los machos. En algunas especies las hembras también producen cantos e incluso realizan dúos de canto con los machos, estos cantos de las hembras tienden a ser mucho más simples que los producidos por los machos. Algunas especies sólo cantan en determinados periodos de tiempo del año, la mejor temporada para observar aves cantando es cuando se está en la época de cría de primavera. Cuando se está en la temporada de cría, los cantos de los machos tienen dos funciones: una es atraer a las hembras y la otra es repeler a los machos rivales, incluso en algunas especies los cantos para atraer hembras es más largo y complejo que cuando el canto se da con fines de defensa territorial.

En el amanecer es cuando el canto de los machos tiene un mayor desenvolvimiento, debido a que las condiciones de alimentación son mejores después del amanecer y por lo tanto las aves tienen mucho más tiempo para cantar, otra razón es debido a que las condiciones del ambiente son favorables para la transmisión del canto, ya que el viento y las turbulencias de aire son reducidas. También resulta ser el mejor momento del día para tomar posesión de territorios libres. Por otro lado, las hembras son mucho más fértiles en el amanecer, por lo que resulta ser el mejor momento para copular.

En ambientes densos, es decir bosques densos y selvas tropicales, el problema de la transmisión del sonido de las aves sufre de dos principales complicaciones: la atenuación y la degradación. Las propiedades de propagación del sonido son diferentes dependiendo del ambiente y altura donde se encuentren las aves. Sin embargo, los sonidos emitidos por las aves pueden adaptarse a las condiciones ambientales, por lo que pueden ser transmitidos y recibidos óptimamente, lo que no siempre significa una transmisión a gran distancia ya que depende de la función por la que fue emitido el sonido.

Por otra parte, los cantos de las aves tienen niveles jerárquicos, los cuales son: frases, sílabas y elementos o notas [Catchpole and Slater, 1995] como se pueden observar en la Figura 2.8. El elemento es la unidad básica del canto, siendo la componente más pequeña en el espectrograma, la unión de uno más elementos forman una sílaba, la estructura de una sílaba varía bastante ya que el número de elementos varían en una sílaba. Asimismo, el conjunto de sílabas que siguen un patrón forman una frase, las sílabas en una frase son típicamente similares una a la otra. De igual manera, una canción está constituida por una serie de frases, se dice que cuando un ave cambia el orden o tipo de frases en sus cantos se producen diferentes tipos de cantos, el repertorio de las aves puede variar desde unos pocos a centenares tipos de cantos en diferentes especies.

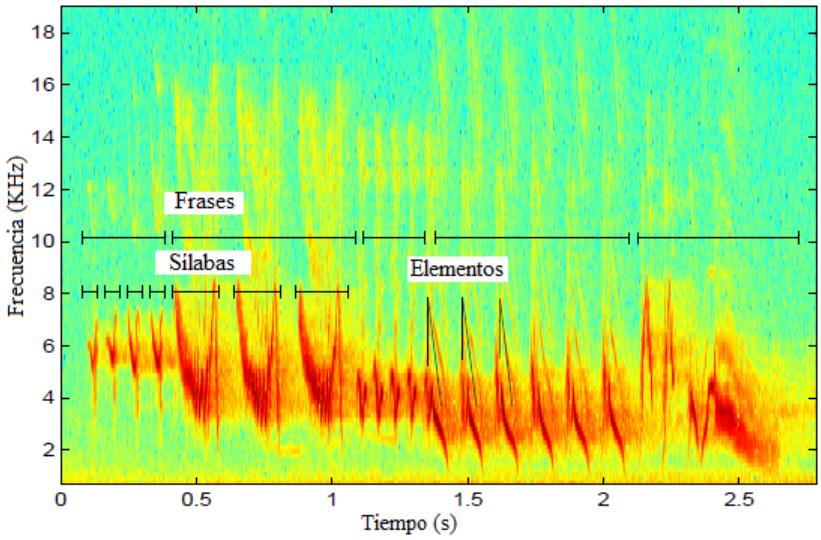

Figura 2.8: Niveles jerárquicos del canto del pinzón común (*Fringilla coelebs*) [Fagerlund, 2004].

2.2.2.3. El llamado de las aves

Normalmente los llamados de las aves son cortos y simples, pero pueden también ser complejos y hasta a veces ser confundidos con cantos, especialmente cuando se

producen una serie de llamados. Los llamados típicamente ocurren por una función específica y pueden ser emitidos tanto por machos como por hembras a lo largo de todo el año. Tienen al menos 10 diferentes categorías, las principales son:

- **De contacto:** para saber la ubicación de un ave (importante en especies que se desplazan en grupos entre la vegetación) o para juntar a los pichones (aún antes de salir del huevo).
- **De alarma:** para alertar a la bandada, utilizada por especies que viven en grupos o gregarias.
- **De reclamo:** sonido fuerte emitido por los pichones para solicitar su comida.

Incluso algunas aves tienen más de un llamado para una categoría y a veces hacen uso de llamados muy similares para significados diferentes. Los sonidos de llamados son importantes para aquellas aves canoras que generalmente tienen un mayor repertorio de llamados que aquellas aves que solo emiten llamados, es decir las aves que no emiten cantos.

2.2.3. Características acústicas del canto en aves

Las características acústicas de las aves, se pueden describir gracias a que existen en las vocalizaciones de las aves, estructuras acústicas que son características de cada especie y son fácilmente analizables.

En general, el intervalo de frecuencia de las aves se encuentra aproximadamente entre los 700 Hz y 2200 Hz [Fletcher and Tarnopolsky, 1999], pero según otras investigaciones existe un mayor rango entre los 500 Hz y los 10 000 Hz e incluso hasta los 14 000 Hz en vencejos (*Apus apus*) [Olvera, 2014]. La estructura acústica de la vocalización está caracterizada por una frecuencia de origen, denominada "frecuencia fundamental" o primer armónico y un conjunto finito (en ocasiones infinito) de frecuencias múltiplos de ésta, a los cuales se les denomina "armónicos".

Por regla general, la frecuencia fundamental es la que lleva asociada más potencia sonora. La frecuencia fundamental y los armónicos se forman como consecuencia de las modificaciones del flujo de aire a través del tracto vocal. Las formantes son las resonancias propias de cualquier elemento que tenga la capacidad de resonar. Una formante es el pico de intensidad en el espectro de un sonido y es la concentración de energía que se da en una determinada frecuencia. Estos son generados por las resonancias de la tráquea, y las oscilaciones que se generan a partir de los grupos de los armónicos de la oscilación en la siringe [Fletcher, 1992].

Otro factor importante en las vocalizaciones de las aves es la intensidad, magnitud o volumen, la que depende mucho del tamaño de la especie o individuo, intención conductual, y distancia entre el emisor-receptor. Dicha distancia, depende a su vez de los hábitats en los cuales viven y se comunican las aves, que suelen ser ecosistemas muy complejos. A menudo se encuentran ambientes muy heterogéneos, con gran

variedad en la composición vegetal, pudiendo encontrarse densas capas de vegetación en algunas zonas y espacios más abiertos en otras, también existen diferencias en la orografía del terreno, donde son los terrenos planos los que facilitan la propagación del sonido de las aves; otro punto es la alternancia en los factores climatológicos, como la temperatura del aire y su humedad, los cuales contribuyen a cambiar los patrones de absorción y reflexión de la energía y distintos niveles de contaminación acústica.

Para evitar la degradación, la atenuación y el solapamiento de sus cantos, tanto con el ruido ambiental como con el canto producido por los sonidos emitidos por las aves vecinas y de esta manera aumentar el espacio activo del sonido, las aves han desarrollado estrategias de adaptación de sus cantos en función de las condiciones locales que presentan los ambientes en los que habitan [Pacual García, 2012].

2.2.4. Monitoreo bioacústico de las aves

La necesidad de implementar un sistema de monitoreo bioacústico automatizado nació en la década de los noventa, con estudios que buscaron evaluar la composición de especies de aves terrestres migratorias que hacen sus viajes durante la noche, desde los sitios donde se reproducen en el hemisferio norte hacia el hemisferio sur donde se resguardan de la época de invierno. Muchas de estas especies emiten cortos llamados mientras vuelan. Estos llamados son específicos a las especies, por lo que se cree que son utilizados por la dificultad de la comunicación visual; así, gracias a ello la bandada mantiene la ubicación espacial con el fin de evitar colisiones, mantiene también la conexión durante el vuelo y puede estimar la dirección del viento.

Los posteriores estudios de bioacústica realizados sobre las aves, permitieron que la identificación y censado de aves se pueda dar gracias al desarrollo de métodos que permiten lograr dicho objetivo. De la misma manera, estos estudios han podido hacer frente al problema que para los métodos tradicionales del estudio de aves han sido grandes retos, estos se dan cuando la vegetación donde habitan las aves es muy densa, como también cuando el acceso de los observadores (censadores de aves) es muy dificultoso, es decir en lugares remotos o cuando los hábitos de las aves no permiten la observación. Cabe resaltar, que los estudios realizados mediante la bioacústica coinciden en sus resultados comparados a los obtenidos por medio de métodos tradicionales, esto se debe a que se tiene una mayor percepción de aves acústicamente que visualmente.

Un enfoque común sobre los métodos utilizados en la bioacústica, ha sido la adaptación de las herramientas del reconocimiento del habla automático (ASR*, Automatic Speech Recognition*), ya que puede vincularse con el substancial progreso de la tecnología y los algoritmos computacionales para el reconocimiento de patrones acústicos, que inicialmente se empleó para el estudio del habla en humanos [Olvera, 2014].

Sin embargo, existen adversidades en el reconocimiento de señales bioacústicas, algunos de los desafíos son: los sonidos de animales se graban en ambientes altamente

ruidosos, se pueden presentar distintas vocalizaciones al mismo tiempo (traslapadas), los individuos emisores de sonido pueden estar en movimiento (esto causa variaciones en intensidad); en el caso del reconocimiento automatizado en aves la dificultad depende, también, de la forma como las especies obtienen sus cantos. Debido a que existen aves que adquieren su repertorio de canto netamente a través de mecanismos genéticos (suboscines) y aves que adquieren su repertorio mediante mecanismos genéticos y procesos de aprendizaje (oscines) [Ridgely and Guy, 1989, Ridgley and Tudor, 1994]. Como consecuencia, la variabilidad en los cantos en individuos y poblaciones de una misma especie es alta, y es común que en estas especies se presenten dialectos y variaciones del canto debido a aislamientos geográficos.

2.3. Métodos para la representación de señales acústicas

La representación de señales o parametrización es la extracción de características de una determinada señal de audio, permitiendo así describirla mediante su información más significativa por medio de un conjunto de parámetros, conocidos como características de la señal.

En general, las características de la señal deben de proporcionar una adecuada representación de la información contenida en la señal a tratar. Por ejemplo, en el caso específico del reconocimiento de especies de aves; se busca que la representación de características de las señales (sonidos) emitidas por éstas, sean específicas de cada especie con la mayor precisión posible. Para poder lograrlo, el conjunto de parámetros de sus señales debe contener cualidades específicas, tales como: la sensibilidad a la particularidad fisiológica por especie (tracto vocal), y la habilidad de tomar en cuenta el estilo de como emite el sonido cada diferente especie.

A continuación, se describen dos métodos utilizados para la representación de señales acústicas.

2.3.1. Coeficientes Cepstrales en la escala de Frecuencias Mel (MFCC)

Los Coeficientes Cepstrales a escala de Frecuencias Mel (MFCC), es una técnica de parametrización muy utilizada en el estudio del reconocimiento del habla humano, sin embargo, también ha demostrado tener un gran potencial sobre el análisis bioacústico. Trabajos como los de [Somervuo et al., 2006, Fox et al., 2008, Cheng et al., 2010], entre otros, han hecho uso de esta técnica para sus procesos de parametrización en tratamientos de señales bioacústicas, obteniendo buenos resultados.

Davis y Mermelstein (DyM) introdujeron el término "Mel Frecuency Cepstral Coefficient" en 1980, que significan coeficientes cepstrales a escala de frecuencias mel, cuando combinaron filtros triangulares perceptualmente distribuidos con la transformada discreta del coseno del logaritmo de las energías de salida de los filtros.

El espaciado entre dichos filtros sobre el eje de frecuencia, imita el comportamiento frecuencial del oído humano, que consiste de un espaciado lineal a frecuencias por debajo de 1000 Hz y un espaciado logarítmico por encima de 1000 Hz.

MFCC es un tipo particular de coeficientes cepstrales derivados de la aplicación del cepstrum sobre una ventana de tiempo de una señal de audio. El cepstrum por su parte se define como la transformada de fourier inversa del espectro logarítmico de la señal (ecuación 2.1).

$$cepstrum(x[n]) = \hat{x} = F^{-1}\{log|\mathrm{F(x[n])}|\} \tag{2.1}$$

Por otro lado, analizar el cepstrum desde un punto de vista matemático, se puede concluir que se trata de un operador que transforma una convolución en el tiempo en una suma en el dominio cepstral. De esta forma se consigue separar las dos componentes de información de la señal de audio: la excitación y el tracto vocal, como se muestra en las ecuaciones 2.2 y 2.3.

$$x[n] = e[n] * h[n] \tag{2.2}$$

$$\hat{x}[n] = \hat{e}[n] + \hat{h}[n] \tag{2.3}$$

Donde $x[n]$ representa la convolución entre la excitación ($e[n]$) y el tracto vocal ($h[n]$), después de la aplicación del cepstrum en (2.2), la covolución se transforma en una suma en el dominio cepstral, lo que se conoce como deconvolución homomórfica.

Aunque muchas técnicas de parametrización hacen uso del cepstrum, con la técnica MFCC se mejora la eficiencia del uso conjunto del cepstrum y la simulación del comportamiento frecuencial del oído humano. El sistema auditivo humano procesa la señal de voz en el dominio espectral, caracterizándose por tener mayores resoluciones en bajas frecuencias, lo que precisamente se consigue mediante el uso de la escala mel, asignando mayor relevancia a las bajas frecuencias.

Para la obtención de los coeficientes MFCC se siguen varias etapas: Enventanado, Transformanda discreta de Fourier (*Discrete Fourier Transform*, DFT) , Banco de filtros mel, y el logaritmo de la Transformada del coseno discreto (*Discrete Cosine Transform*, DCT) , ilustradas en la Figura 2.9.

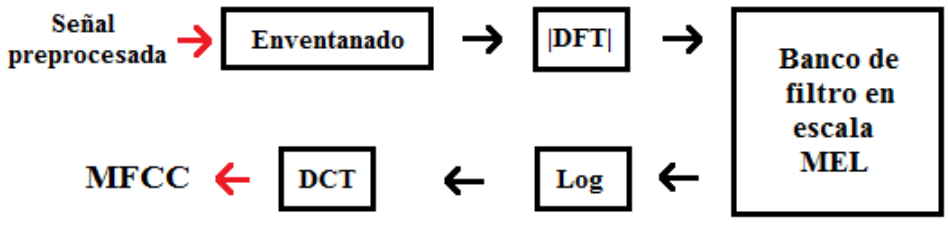

Figura 2.9: Diagrama de bloques para la obtención de los coeficientes MFCC.

- **Enventanado**

Debido a que las grabaciones de audio en campo son señales estadísticamente no estacionarias, sus características espectrales varían con el tiempo, motivo por el cuál es necesario segmentar la señal en ventanas, para así poder obtener un proceso cuasiestacionario por cada una de las ventanas y obtener sus características. El tamaño de la ventana debe ser lo suficientemente largo para poder extraer información de la señal, pero lo suficientemente corto para se pueda considerar como estacionario.

En reconocimiento de voz y procesamiento de audio, se usa comúnmente una ventana de entre 10 y 50 ms. Como se observó que las vocalizaciones más cortas están alrededor de 60 ms, se decidió calcular las características con ventanas de un tercio de este mínimo, 20 ms. Para mantener la continuidad de la información de la señal, se realiza el enventanado con bloques de muestras solapados entre sí, de tal manera que no se pierde información en la transición entre ventanas. Generalmente el solapamiento se lleva a cabo con un desplazamiento de 10 ms, obteniéndose coeficientes MFCC cada 10 ms [Gupta et al., 2013].

Aunque existen varios tipos de ventanas (Hammning, Hanning, Rectangular, etc.), generalmente se hace uso de la ventana Hamming con el fin de poder suavizar discontinuidades y así minimizar la distorsión de la señal.

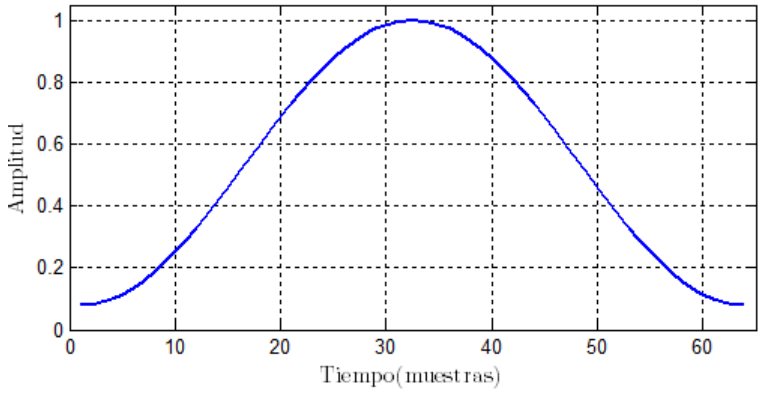

Figura 2.10: Ventana Hamming usada en la etapa del enventanado de los MFCC.

Por otra parte, se define a la ventana como $W_m(n)$, $0 \leq m \leq N_m - 1$, donde N_m representa la cantidad de muestras dentro de cada trama, y la salida después del enventanado de la señal quedando representada de la siguiente manera:

$$Y_m = X_m.W_n(m) \tag{2.4}$$

Donde Y_m representa la señal de salida, después de multiplicar la señal de entrada X_m con la ventana Hamming $W_n(m)$. Matemáticamente la ventana Hamming se representada como:

$$W_n(m) = 0{,}54 - 0{,}46cos\left(\frac{2\pi m}{N_m - 1}\right) \tag{2.5}$$

- **Transformada discreta de Fourier (DFT)**

Gracias a la etapa del enventanado, la señal es cuasiestacionaria, es decir, se asemeja a un sistema lineal e invariante en el tiempo, que cumple ciertas propiedades permitiendo realizar una representación y análisis del contenido en frecuencia por cada intervalo de la señal dividida, haciendo uso de la Transformada Discreta de Fourier (DFT).

Una de las propiedades es que la respuesta a secuencias sinusoidales es también sinusoidal, de igual frecuencia y con amplitud y fase determinadas por el sistema. Esta propiedad hace que las representaciones de las señales mediante sinusoides o exponenciales complejas (es decir, las representaciones de Fourier) sean muy útiles.

Dada una señal en tiempo discreto $x(n)$ con N muestras, su transformada $X(k)$ está dada por:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi kn}{N}} \tag{2.6}$$

En la práctica, el costo computacional del cálculo de la DFT se reduce utilizando la transformada rápida de Fourier (*Fast Fourier Transform,* FFT).

- **Banco de filtros a escala de frecuencias MEL**

Una vez que se obtiene la señal en el dominio de la frecuencia, ésta pasa por un banco de filtros paso banda triangulares, que tienen sus frecuencias centrales y anchos de banda seleccionados de acuerdo a la escala mel [Ittichaichareon et al., 2012], esto con el propósito de emular el comportamiento frecuencial del oído humano al percibir un sonido.

La escala mel proporciona la manera de cómo deben estar espaciados los filtros y que tan amplios estos deberían ser, por ello con esta escala, los filtros son mucho más amplios en altas frecuencias y angostos en bajas frecuencias. Para el diseño de la escala mel, realizaron un mapeo entre la escala de frecuencia real (Hz) y la escala de frecuencias percibidas por varias personas (mel). Durante dicho mapeo la escala mel mostró un comportamiento lineal para valores antes de 1KHz y logarítmico para valores después de dicha frecuencia, como se muestra en la Figura 2.11.

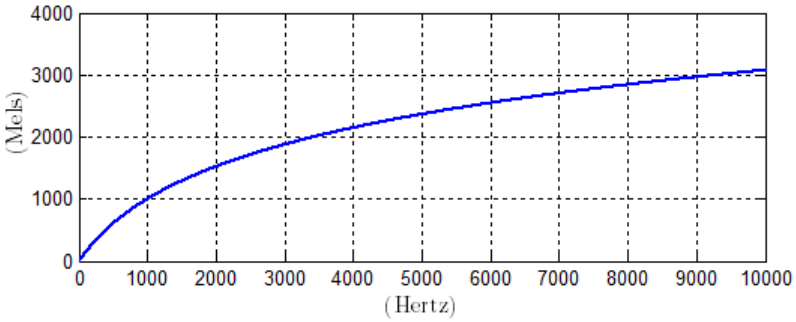

Figura 2.11: Escala Mel.

Sobre la fórmula para la conversión de la frecuencia de Hz (f_{lin}) a mels (f_{mel}), es necesario mencionar que se realizaron bastantes estudios sobre la percepción frecuencial del sistema auditivo humano. Una de las primeras aproximaciones fue la de Koenig en 1949 [Koenig, 1949], conocida como la escala Koenig, la que es exactamente lineal por debajo de los 1000 Hz y logarítmica para valores superiores a los 1000 Hz.

Por otra parte, en ese mismo año Fant realizó una aproximación más precisa, dada por la siguiente ecuación:

$$f_{mel} = k_{const}.log_n\left(1 + \frac{f_{lin}}{F_b}\right) \tag{2.7}$$

Posteriormente en 1973, para F_b= 1000, Fant encontró que ésta aproximación era más cercana a la escala mel comparada con la de Koenig, sólo para el rango de frecuencias de 0 a 5 KHz.

$$f_{mel} = \frac{1000}{log_n 2}.log_n\left(1 + \frac{f_{lin}}{1000}\right) \tag{2.8}$$

Una de las ventajas de la ecuación 2.8 es que los valores de f_{mel} no cambian al momento de elegir la base n del logaritmo.

Se realizaron también otras aproximaciones derivadas de la ecuación 2.7, que hacen uso del logaritmo decimal y natural, y que además especifican un valor determinado para la constante k_{const}. Las aproximaciones son las siguientes:

$$f_{mel} = 2595.log_{10}\left(1 + \frac{f_{lin}}{700}\right) \tag{2.9}$$

$$f_{mel} = 1127.ln\left(1 + \frac{f_{lin}}{700}\right) \tag{2.10}$$

Siendo éstas dos últimas ecuaciones 2.9 y 2.10 las más utilizadas para el cálculo de los MFCC. Cabe resaltar también que dichas ecuaciones proporcionan una mejor aproximación de la escala mel a frecuencias menores de 1000 Hz, sacrificando precisión para frecuencias mayores a 1000 Hz [Ganchev, 2005].

En la Figura 2.12, se muestra un banco de filtros triangulares presentados por DyM en 1980.

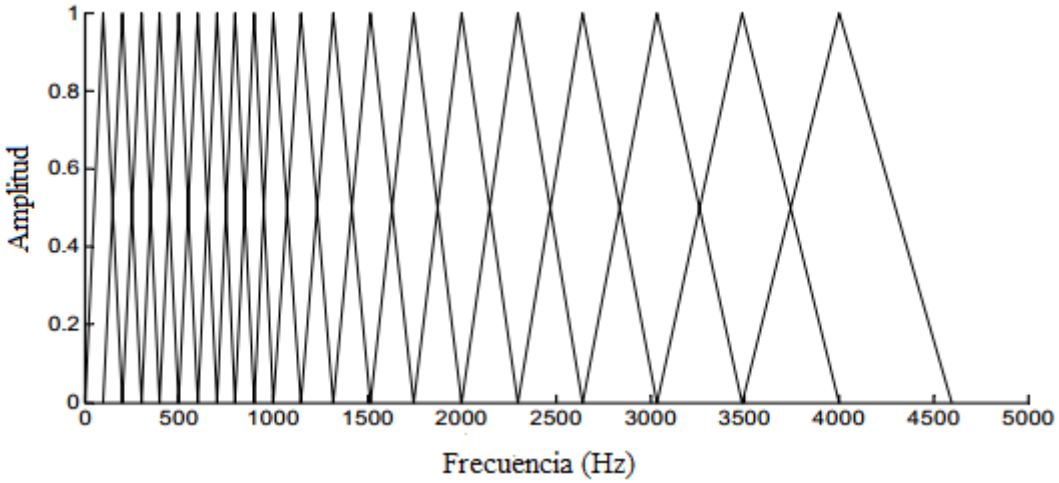

Figura 2.12: Banco de filtros en la escala mel usado por DyM. Las Frecuencias centrales de los primeros diez filtros están linealmente espaciados y las demás diez tienen un espaciamiento logarítmico de sus frecuencias centrales [Ganchev, 2005].

- **Logarítmo de la transformada del coseno discreto (DCT)**

Una vez obtenida la señal a la salida de los filtros, se calcula la energía correspondiente en cada uno de ellos y luego se le aplica el logaritmo, pasando al dominio de la potencia espectral logarítmica. Trabajar en dicho dominio provoca que en las bandas adyacentes de filtros exista un alto grado de correlación, lo que a su vez ocasionaría la obtención de coeficientes espectrales estadísticamente muy dependientes entre ellos, por lo tanto, para eliminar dicha dependencia se hace uso de la Transformada discreta del coseno (DCT), además que la DCT generalmente se usa para la compresión de datos, es decir después de aplicarle la DCT a una señal se tiene más información concentrada en un pequeño número de coeficientes, por lo que se requiere menor almacenamiento para representar el espectro mel en un pequeño número de coeficientes. Las salida después de la DCT se conoce como los coeficientes cepstrales en la escala de frecuencias mel o más comúnmente como los MFCC. Donde los coeficientes están representados por la siguiente ecuación:

$$C_{MFCC[c]} = \sum_{k=0}^{N-1} (logX_k) cos \left[c \left(k - \frac{1}{2} \right) \frac{\pi}{N} \right]; \quad c = 0...F. \tag{2.11}$$

Donde $C_{MFCC[m]}$ representa a los MFCC, X_k la energía de cada salida por filtro, N el número máximo de canales de los filtros mel, y c el número de coeficientes por trama.

2.3.2. Factorización de Matrices No-negativas(NMF)

La factorización de matrices no negativas (NMF) es una técnica útil para la representación de datos no negativos. Fue inicialmente introducida por Pattero y Taper en el año de 1994 [Paatero and Tapper, 1994], sin embargo, no fue sino hasta el año de 1999, que estudios y trabajos realizados por Lee y Seung [Lee and Seung, 2001] la hicieron mucho más conocida.

El problema básico en NMF se establece de la siguiente manera: dada una matriz $V \epsilon \mathbb{R}_+^{N \times M}$ (con $v_{nm} \geq 0$), y un grado R, tal que $R \leq min(N, M)$. Se busca encontrar dos matrices no negativas $W = [w_1, w_2, ..., w_R] \epsilon \mathbb{R}_+^{N \times R}$ y $H = [h_1, h_2, ..., h_R] \epsilon \mathbb{R}_+^{R \times M}$, que factoricen a V tanto como sea posible.

$$V \approx WH \tag{2.12}$$

La matriz W contiene en sus columnas a los vectores base representativos de los datos, que combinados de forma adecuada a través de los coeficientes de combinación lineal (contenidos en la matriz H) logran reconstruir la matriz V, que contiene los datos a ser representados. Por lo tanto, si los vectores base lograsen encontrar la estructura oculta en los datos de la matriz V , se estaría obteniendo una mejor aproximación de la ecuación 2.12 [Lee and Seung, 2001].

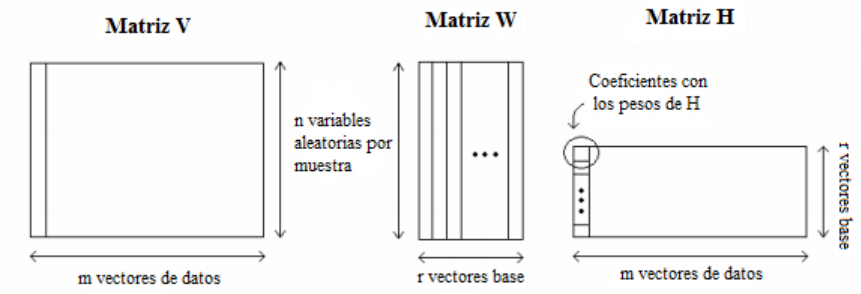

Figura 2.13: Representación esquemática de las matrices V,W y H del modelo NMF [Núñez Martínez, 2005].

Ya que el objetivo de NMF es poder aproximar lo mejor posible V a WH, es decir minimizar su diferencia, se hace uso de funciones de coste que permiten cuantificar la calidad de la aproximación entre V y WH. Existen dos funciones de coste muy utilizadas para este problema, una tiene que ver con la distancia entre dos matrices no negativas V y WH, conocida como la distancia Euclideana; y la otra es una

medida conocida como la divergencia de Kullback-Leibler, ambas representadas en las ecuaciones 2.13 y 2.14 respectivamente.

$$||\text{V-WH}||^2 = \sum_{ij}(V_{ij} - [WH]_{ij})^2 \tag{2.13}$$

$$D(V||WH) = \sum_{ij}\left(V_{ij}log\frac{V_{ij}}{[WH]_{ij}} - V_{ij} + [WH]_{ij}\right) \tag{2.14}$$

La divergencia de Kullback-Leibler (2.14) ha demostrado tener buenos resultados en tareas del procesamiento del habla, tales como: la mejora de la señal del habla [Ludeña-Choez and Gallardo-Antolín, 2012], extracción de características en tareas relacionadas al procesamiento del audio [Schuller et al., 2010], clasificación de eventos acústicos [Ludeña-Choez and Gallardo-Antolín, 2015], entre otros. Por tal motivo, ha sido considerado su uso en esta tesis de grado. Con la finalidad de encontrar un valor óptimo local para la divergencia de Kullback-Leibler entre V y WH, puede usarse un esquema iterativo con reglas de actualización multiplicativas como se propone en [Lee and Seung, 1999], indicadas en las ecuaciones 2.15 y 2.16 :

$$W \leftarrow W \otimes \frac{\frac{V}{WH}H^T}{1H^T} \tag{2.15}$$

$$H \leftarrow H \otimes \frac{W^T\frac{V}{WH}}{W^T1} \tag{2.16}$$

Donde 1, es una matriz de tamaño V , cuyos elementos son todos unos. NMF produce una representación específica de los datos, reduciendo la redundancia.

2.4. Algoritmo para la clasificación de señales acústicas

En esta sección, se presentará un algoritmo para la clasificación de una colección de datos representativos, obtenidos de la etapa de extracción de características de una señal acústica. La técnica está basada en algoritmos de aprendizaje supervisado, es decir las etiquetas de los datos de clasificación se encuentran predefinidas. Dicha técnica es conocida como: Máquina de Vectores de Soporte (SVM, *Support Vector Machine*).

2.4.1. Máquina de Vectores de Soporte (SVM)

La teoría de las máquinas de vectores de soporte (SVM*, Support Vector Machines*) fue desarrollada por Vladimir Vapnik en el año 1979. Es una técnica de clasificación

que proporciona un hiperplano clasificador óptimo, debido a que pueden existir una variedad de hiperplanos que pueden separar a los datos, pero sólo existe uno que puede maximizar la distancia entre el hiperplano y el punto más cercano de los datos (margen). Éste es el hiperplano clasificador óptimo, mostrado en la Figura 2.14.

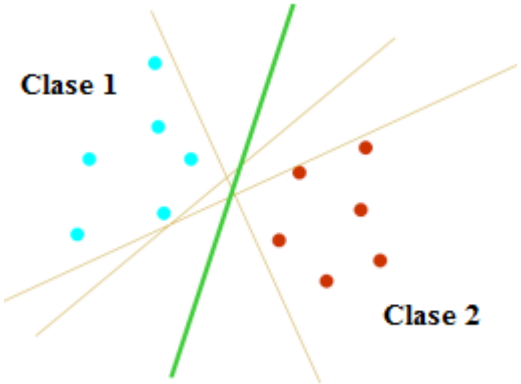

Figura 2.14: Hiperplano clasificador óptimo [Gunn et al., 1998]

Para poder clasificar los datos de entrada, el hiperplano clasificador óptimo se encuentra en un espacio de alta dimensión (denominado espacio de características, $\mathcal{H}$), generado por medio de una función de transformación llamada kernel $\phi(\cdot) : R^d \rightarrow \mathcal{H}$ [Burges, 1998], como se puede ver en la Figura 2.15.

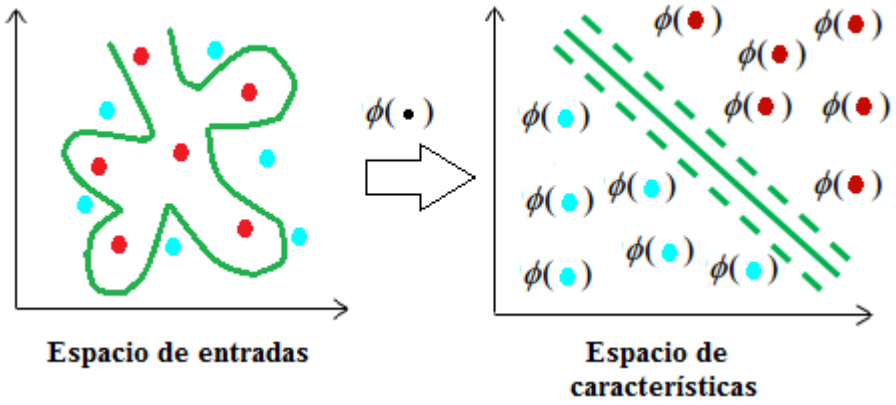

Figura 2.15: Transformación de espacios por la función kernel.

La siguiente formulación a la solución de la SVM se ha tomado de [Mayhua López et al., 2013]. De manera general, la solución de la SVM se representa mediante la siguiente función discriminante:

$$f(x) = w^T \Phi(x) + b \tag{2.17}$$

donde w y b determinan el hiperplano separador en el espacio de características. Con la finalidad de maximizar el margen, anteriormente mencionado, w y b se pueden obtener como la solución del siguiente problema de optimización [Schölkopf and Smola, 2002]:

$$\underset{w,b,\xi}{mín} \frac{1}{2} \parallel w \parallel_2^2 + C \sum_{l=1}^{L} \xi^{(l)} \tag{2.18}$$

$$s.t.\ y^{(l)} \left[w^T \Phi(x^{(l)}) + b \right] \geq 1 - \xi^{(l)}; \quad l = 1, ..., L$$
$$\xi^{(l)} \geq 0; \quad l = 1, ..., L$$

donde C es un parámetro positivo que controla el compromiso entre la sencillez del modelo y el error de clasificación , y $\left\{ \xi^{(l)} \right\}_{l=1}^{L}$ es el conjunto de las variables de holgura que se introducen para permitir que algunas muestras estén dentro del margen o mal clasificadas (Figura 2.16). El funcional de este problema de optimización es convexo, lo que garantiza la unicidad de la solución.

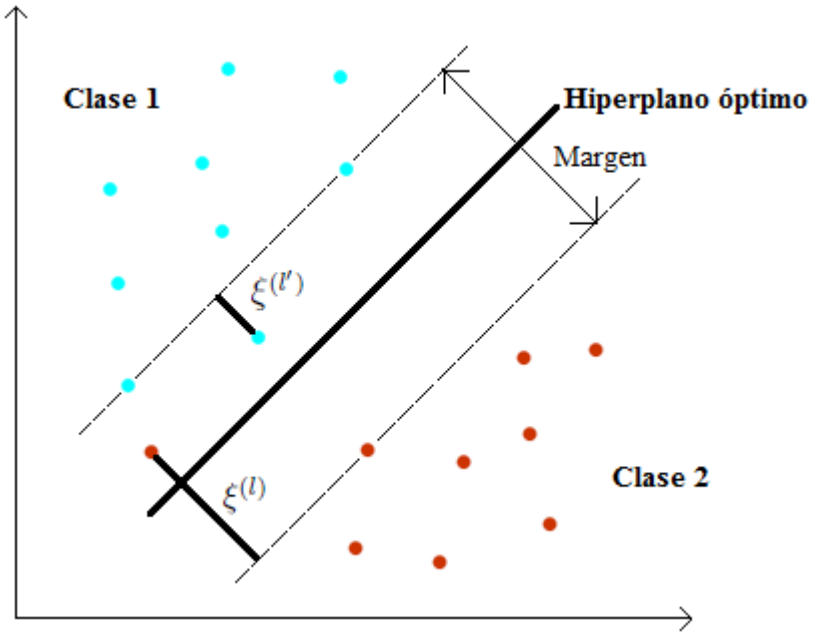

Figura 2.16: Caso no separable en un problema de dos dimensiones.

Para la solución de la ecuación 2.18, se hace uso de los multiplicadores de Lagrange $\left\{ a^{(l)} \right\}_{l=1}^{L}$ y las correspondientes condiciones de Karush-Kuhn-Tucker (KKT) [Karush, 1939, Kuhn and Tucker, 1951], transformando el problema de optimización en el de programación cuadrática (QP, "*Quadratic Programming*").

$$\max_{a} -\frac{1}{2}\sum_{l=1}^{L}\sum_{l'=1}^{L} a^{(l)}a^{(l')}y^{(l)}y^{(l')}K(x^{(l)},x^{(l')})+\sum_{l=1}^{L} a^{(l)} \tag{2.19}$$

$$s.t.\ \sum_{l=1}^{L} a^{(l)}y^{(l)} = 0$$
$$C \geq a^{(l)} \geq 0; \quad l = 1, ..., L$$

donde $K(\cdot,\cdot)$ es el núcleo asociado a $\Phi(\cdot)$. Existen tres funciones núcleo típicas, RBF (Gaussiana), polinómica y lineal; mostradas a continuación en el mismo orden:

$$K(x^{(l)},x^{(l')}) = exp(-\parallel x^{(l)} - x^{(l')} \parallel^2 /2\sigma^2),$$

$$K(x^{(l)},x^{(l')}) = (x^{(l)}.x^{(l')} + 1)^p,$$

$$K(x^{(l)},x^{(l')}) = (x^{(l)}.x^{(l')}).$$

El clasificador SVM no lineal toma la forma:

$$f(x) = \sum_{l=1}^{L} a^{(l)}y^{(l)}K(x^{(l)},x) + b \tag{2.20}$$

donde $a^{(l)}$son constantes con valores reales y positivos que se corresponden con la solución del problema QP. Una de las ventajas de esta formulación radica en que sólo aquellos datos con $a^{(l)} > 0$ forman parte de la solución. Estos puntos se denominan Vectores Soporte (SVs, "*Support Vectors*"), y corresponden a muestras de entrenamiento mal clasificadas o a muestras que se encuentran ubicadas en o dentro del margen.

Posteriormente, para determinar el valor de b se consideran los SVs que satisfacen la condición:

$$y^{(l)}(\sum_{l'\epsilon S} a^{(l')}y^{(l')}K(x^{(l)},x^{(l')}) + b) = 1 \tag{2.21}$$

donde S indica el conjunto de índices de los SVs. El valor de b se calcula como:

$$b = \frac{1}{N_{S'}}\sum_{l\epsilon S'}\left(y^{(l)} - \sum_{l'\epsilon S} a^{(l')}y^{(l')}K(x^{(l)},x^{(l')})\right) \tag{2.22}$$

donde S' indica el conjunto de índices de muestras que tienen $0 < a^{(l)} < C$.

Por último, el valor de los coeficientes $a^{(l)}$define las siguientes situaciones:

1. Si $a^{(l)} = 0$, indica que $x^{(l)}$ está bien clasificado y se ubica fuera del margen.
2. Cuando $0 < a^{(l)} < C$, se tiene que $x^{(l)}$ está estrictamente sobre el margen y por tanto bien clasificado.
3. Si $a(l) = C$, implica que $x^{(l)}$ se ubica dentro del margen y puede estar mal clasificado o correctamente clasificado.

2.4.1.1. Métodos SVM multiclase

La clasificación basada en SVM descrita anteriormente es una clasificación binaria ($z = 2$). No obstante, los problemas del mundo real generalmente requieren la discriminación de más de dos clases. Es por ello, que surgió la necesidad de implementar métodos que pudieran resolver problemas multiclase ($z > 2$). Existen dos esquemas representativos para resolver problemas SVM multiclase: uno contra el resto (*one-against-all*) y uno contra uno (*one-against-one*). Ambos esquemas, son casos especiales de los Códigos de corrección de errores (ECOC, *Error Correcting Output Codes*), que descomponen el problema multiclase en un conjunto predefinido de problemas binarios [Dietterich and Bakiri, 1995].

- **Esquema uno contra el resto:** este esquema construye z clasificadores que definen otros tantos hiperplanos que separan la clase i de los $z - 1$ restantes. Durante la etapa de prueba, la etiqueta de la clase se determina mediante un clasificador binario que da el valor de salida máximo. El mayor problema de este esquema es el desbalance de los datos del conjunto de entrenamiento, es decir, que existan muchas más muestras para unas clases que para otra [Vapnik and Vapnik, 1998].
- **Esquema uno contra uno:** este esquema, también conocido como clasificación en parejas, evalúa todos los clasificadores en pareja e induce $\frac{Z(Z-1)}{2}$ clasificadores, uno para cada par de clases posible, enfrentando así a todas las clases una a una. La aplicación de cada clasificador a un ejemplo de prueba daría un voto a la clase ganadora. Un ejemplo de prueba se etiqueta a la clase con más votos. El tamaño de los clasificadores creados por el esquema uno contra uno es mucho mayor que el del esquema de uno contra el resto. Además, en comparación con el esquema de uno contra el resto, el esquema uno contra uno es más simétrico [Schölkopf et al., 1999]. Cabe resaltar que en esta tesis de grado se hace uso de este esquema para la clasificación final de las 12 especies de aves.

3 Estado del Arte

A continuación, se presentará el estado del arte del estudio sobre clasificación de aves por medio de sus vocalizaciones.

3.1. Métodos realizados en estudios iniciales

Muchos de los estudios tradicionales sobre los sonidos de las aves están basados en la inspección visual de los espectrogramas o sonogramas de los sonidos. Realizar continuamente una identificación de los espectrogramas a lo largo del conjunto de los sonidos de las aves de forma manual, resulta ser una tarea extremadamente laboriosa y demandante de tiempo. Por lo que, el reconocimiento automático a través de los sonidos de las aves resulta ser la solución a dicho problema [Lee et al., 2008].

Uno de los primeros intentos en reconocer a las aves automáticamente por sus sonidos fueron realizados por [Anderson et al., 1996] donde utilizaron el alineamiento temporal dinámico (DTW*, Dynamic Time Warping*), para el análisis automático de grabaciones continuas de cantos de aves. Ellos realizaron la comparación directa de los espectrogramas de las señales, e identificaron constituyentes como también límites constituyentes, que permitieron la identificación de una amplia gama de señales y componentes de la señal. Los vectores de características se obtuvieron de las magnitudes logarítmicas de la Transformada Rápida de Fourier (FFT*, Fast Fourier Transform*) entre los 0.5 a 10 KHz. Realizaron la evaluación de su método sobre las vocalizaciones de las aves: indigo buntings (*Passerina cyanea*) y zebra finches (*Taeniopygia guttata*). Los datos fueron recolectados de un ambiente con bajo ruido. Las modelos representativos (sílabas), fueron etiquetados manualmente y clasificados en cantos y llamados con más del 97% de precisión.

Seguidamente en [Kogan and Margoliash, 1998], compararon dos técnicas: la DTW y los modelos ocultos de Markov (HMM, *Hidden Markov Model*), para el reconocimiento automático del canto de las aves, la experimentación se realizó sobre las especies de aves: zebra finches (*Taeniopygia guttata*) e indigo buntings (*Passerina cyanea*) , (las mismas especies de aves de [Anderson et al., 1996]), en este caso las sílabas fueron representadas por espectrogramas, se compararon seis tipos de parámetros de características para los HMM, entre ellos, los coeficientes de predicción lineal (LPC*, Linear Prediction Coefficients*), los MFCC, los coeficientes de bancos de filtros mel logarítmicos, y los coeficientes de banco de filtros mel lineales, de los cuales los mejores resultados se obtuvieron con los MFCC y la etapa de la clasificación se

realizó comparando las características de los datos de prueba con prototipos predefinidos (aprendidos de la etapa de entrenamiento). Los resultados demostraron que las técnicas basadas en DTW tenían un desempeño satisfactorio, sin embargo, con la DTW se requirió de una cuidadosa selección de los modelos que pueden necesitar de un conocimiento más experimentado para grabaciones ruidosas o presencia de llamadas confusas de corta duración, por lo que en muchos experimentos los HMM demostraron un mejor desempeño que la DTW. Cabe resaltar que una desventaja de los HMM es la clasificación errónea para vocalizaciones de corta duración o unidades de canto muy variables en su estructura.

3.2. Métodos recientes

Los primeros estudios realizados sobre los modelos de mezclas gausianas (GMM*, Gaussian Mixture Models*) aplicadas al estudio de la vocalización de la aves, fueron decritos por [Cheng et al., 2010]. En este estudio se hizo uso de la técnica de los coeficientes cepstrales a escala de frecuencia mel (MFCC, *Mel Frequency Cepstral Coefficients*) para la parametrización y los GMM para la clasificación, basó su estudio sobre aves ubicadas en la china: Timalí pekinés (*Rhopophilus pekinensis*), Mosquitero de Hume (*Phylloscopus humei*), Mosquitero de Gansu (*Phylloscopus kansuensis*) y el Bubul Chino (*Pycnonotus sinensis*). Los datos fueron de tamaños específicos, los cuales se obtuvieron a partir de recortes de las grabaciones. Las grabaciones fueron tomadas en el ambiente libre y con ruido de fondo. Se logró el reconocimiento de entre 89.1 % y el 92.5 % .

En el caso de [Stattner et al., 2013], se buscaba realizar una metodología eficiente para automatizar la recolección de datos, para la parametrización de las señales se utilizaron las técnicas MFCC y los Coeficientes Cepstrales de Predicción Lineal (LPCC*, Linear Prediction Coefficients Cepstrals*), y para la clasificación se utilizaron métodos como: DTW, árboles de decisión C 4.5, los árboles de decisión de bosques aleatorios (RF*, Random Forest*), técnicas bayesianas (NB*, Naive Bayes*), perceptrones multicapa (MLP*, Multi-layer Perceptron*) junto a las redes neuronales artificiales (ANN*, Artificial Neural Network*). Los mejores resultados se obtuvieron usando la parametrización MFCC y como clasificador a NB con una tasa del 98.52% de precisión.

A diferencia de los estudios previos, una nueva representación paramétrica para los sonidos de las aves es propuesta en [Fagerlund and Laine, 2014], éste estudio es en el dominio temporal y consiste en los estadísticos de los patrones temporales a corto plazo de las señales de las vocalizaciones de las aves, es decir, los patrones temporales son representados por medio de la permutación de ordenar los valores de la amplitud de la señal en ventanas cortas, por lo que cada ventana fue representada por un índice de código de permutación y los índices de ventanas seguidas formaban una secuencia simbólica de códigos de permutación . Finalmente, a partir de la secuencia de código de permutación se construye una matriz de frecuencia de par de permutación (PPF,

Permutation Pair Frequency), que es la representación paramétrica para eventos de audio y se utiliza como modelo estadístico de las señales para la clasificación. En este estudio, se hizo uso del método de los k-vecinos más cercanos (k-NN, *k-Nearest Neighbour*) para la etapa de la clasificación, los resultados mostraron altas tasas de porcentaje en el reconocimiento en comparación a las obtenidas por medio de los MFCC, además los resultados permitieron concluir que con ventanas cortas de tiempo se puede encontrar información esencial para una clasificación precisa.

Con la finalidad de mejorar la precisión del reconocimiento de los sonidos de los animales en diversos entornos de ruido con baja relación señal a ruido (SNR, *Signal to Noise Ratio*), en [Li and Wu, 2015] se presenta, un método de reconocimiento del sonido animal basado en una característica doble del espectrograma. Lo que realizan es extraer la función de proyección y la característica de varianza de patrón binario local (LBPV, *Local Binary Pattern Variance*) del espectrograma, para generar la característica doble. Es decir, la característica de función de proyección, consiste en la reducción de dimensiones de los vectores de la trama del espectrograma de la señal a través la descomposición de los eigenvalores con la finalidad de que sean adecuados para la clasificación, por otra parte, la característica LBVP se forma acumulando las varianzas correspondientes de todos los píxeles para cada patrón binario local uniforme (ULBP, *Uniform Local Binary Pattern)* en el espectrograma, éste caracteriza la estructura espacial de la textura de la imagen, y la varianza describe la información de contraste de la textura de la imagen. Finalmente, hicieron uso de los bosques aleatorios como método de clasificación, su base de datos consistió de 40 sonidos animales que contienen sonidos de aves, sonidos de mamíferos, y sonidos de insectos. En la parte experimental, simularon un ambiente real bajo diferentes entornos de ruido con diferentes SNRs. Utilizaron el ruido del viento, el ruido del tráfico y el ruido de la lluvia, probaron que las tasas promedio de precisión de característica doble es 37,86% más alta que MFCC, 16,58% más alta que la característica LBPV y 5,71% superior a la función de proyección, por lo que se concluyó que la combinación de dos características puede mejorar efectivamente el rendimiento del reconocimiento.

El siguiente trabajo [Debnath et al., 2016], presenta una técnica de procesamiento de datos de audio con procesamiento de imágenes. En primera instancia lo que hacen es obtener el espectrograma de los archivos de audio después de aplicar la Transformada de Fourier de Tiempo Corto (STFT, *Short Time Fourier Transform*) usando ventanas Hamming de tamaño de 512 muestras y 75 % de solape entre ventana y ventana. Recortaron el espectrograma dejando sólo la parte que contenía la información de la vocalización del ave y, también convirtieron el espectrograma en una imagen a escala de grises para luego reducir el ruido de fondo asignando valores de 1 ó 0 a cada pixel de acuerdo a un umbral, por lo que obtuvieron un espectrograma binario, al cual le aplicaron un operador morfológico de cierre para aislar los límites de los diferentes componentes de la imagen, asimismo, le aplicaron un operador morfológico de dilatación para obtener una mejor forma de los patrones. Luego le aplicaron un filtro de mediana basado en el número de pixeles conectados

en un mismo componente, quitando así los pixeles más aislados. Seguidamente, se etiquetaron y segmentaron todos aquellos pixeles conectados en una misma área y se volvió a aplicar un filtrado basado en umbrales dentro de cada segmento, para poder retirar el ruido restante. Finalmente, se escogen los segmentos que serán las características de la vocalización de las aves mediante una técnica de pareo de plantillas. En este estudio, se experimentaron con tres diferentes métodos para la clasificación; los de predicción: (RDT, *Random Decision Tree*), (ETR, *Extra Tree Regressor*) y la máquina de vectores de soporte (SVM, *Support Vector Machine)*, obteniendo mejores resultados con el método de la regresión forestal aleatoria con 0.962 como área bajo la curva Característica Operativa del Receptor (ROC, *Receiver Operating Characteristic*).

4 Marco experimental

En el presente capítulo, se muestra el marco experimental de ésta tesis de grado. Iniciando con la descripción de la base de datos usada, en la sección 4.1. El esquema de parametrización para la clasificación de especies de aves en la sección 4.2, que contiene a su vez; el preprocesamiento realizado sobre la base de datos, la descripción de la parametrización propuesta de la señales de audio de las aves y el sistema de clasificación usado. Por último, en la sección 4.3 se presentan experimentos preliminares, que permitieron establecer la duración óptima de las tramas de audio para la clasificación basada tanto en segmentos como sílabas, esto con la finalidad de fijar los mencionados parámetros para la etapa de resultados experimentales.

4.1. Base de datos

La base de datos está compuesta de 1316 archivos de audio con una duración total de 15:33:24 horas, distribuida entre las 12 especies de aves como se muestra en el cuadro 4.1.

Etiqueta de clase	Especie	Duración (hh:mm:ss)	Cantidad de archivos
1	Aramides cajanea	00:55:40	64
2	Coereba flaveola	01:39:40	201
3	Colibri thalassinus	01:51:40	131
4	Crypturellus cinereus	01:52:23	94
5	Crypturellus obsoletus	01:03:34	87
6	Crypturellus soui	01:07:43	100
7	Crypturellus undulatus	00:57:52	68
8	Lathrotriccus euleri	00:58:02	120
9	Piranga olivacea	00:55:42	64
10	Piranga rubra rubra	00:50:28	72
11	Rupornis magnirostris	02:08:33	181
12	Synallaxis azarae	01:12:02	134

Cuadro 4.1: Distribución de tiempo de grabación y cantidad de archivos de audio por especie de ave.

Originalmente los archivos de audio no tenían la misma frecuencia de muestreo, por lo que se procedió a normalizar a una única frecuencia de muestreo establecida en F_S= 22050 Hz.

Se analizaron los espectrogramas de cada uno de los archivos de audio de las 12 especies de aves, obteniéndose así una frecuencia máxima de 10335 Hz, que pertenece a la especie de ave, *Colibri thalassinus*. De acuerdo al teorema de Nyquist & Shannon, que define a la frecuencia de muestreo óptima, como el doble de la frecuencia máxima, es decir, $F_S \geq 2F_{max}$, es que se estableció F_S= 22050 Hz.

Debido a que la base de datos es reducida, resultaría difícil extraer conclusiones lo suficientemente solventes de los experimentos. Para solucionar este problema, la base de datos ha sido extendida artificialmente usando un esquema de validación cruzada 6-fold, que consiste básicamente en dividir la base de datos en 6 grupos, 5/6 entrenamiento y 1/6 para prueba. Finalmente, los resultados finales resultan del promedio de los 6 sub experimentos. Con esta técnica se proporciona una mayor consistencia a las pruebas y una mayor fiabilidad a los resultados obtenidos [Ludeña-Choez and Gallardo-Antolín, 2015].

4.2. Esquema de parametrización para la clasificación de especies de aves

En la Figura 4.1 se muestra el diagrama de bloques del esquema de parametrización para la clasificación de especies de aves usando señales de audio, este diagrama se compone básicamente de 3 módulos: preprocesamiento, parametrización y clasificación. El módulo de preprocesamiento de la señal consiste básicamente en la limpieza de las señales de audio, el mismo que será descrito en la subsección 4.2.1. El punto principal de este proyecto recae sobre el módulo de la parametrización, aquí se aplican las técnicas MFCC y NMF para buscar la mejor caracterización posible para los archivos de audio, dicho módulo será descrito en la subsección 4.2.2. El módulo de clasificación, descrito en la subsección 4.2.3, está compuesto del clasificador en sí, y de los modelos que se generan mediante las SVM, para cada una de las 12 clases de audio de las especies de aves consideraras en esta tesis de grado.

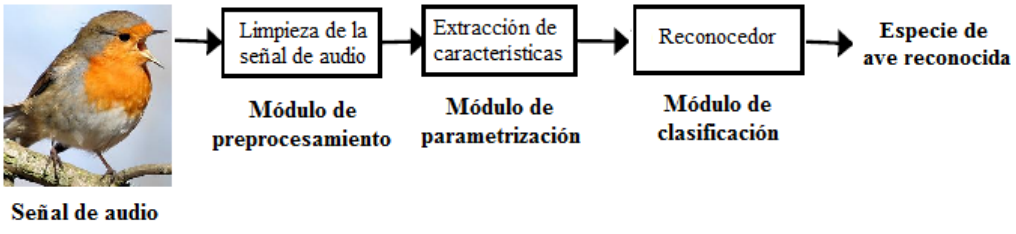

Figura 4.1: Diagrama de bloques del esquema de de parametrización para la clasificación.

4.2.1. Módulo de preprocesamiento

Básicamente, el módulo de preprocesamiento consiste en la limpieza de las señales de audio de las aves. Es decir, la eliminación de frecuencias que no pertenecen a las vocalizaciones de las aves, como el ruido ambiental, voz humana, vocalizaciones de otras clases de animales, etc.

Para este propósito, se utilizó un filtro paso banda Butterworth de orden 12, con frecuencias de corte de 200 Hz a 10500 Hz. Para la determinación de las frecuencias de corte, se realizó la visualización de los espectrogramas de cada uno de los archivos de audio de las 12 especies de aves, siendo la frecuencia mínima igual a 200 Hz y la mayor igual a 10500 Hz, por lo que para las experimentaciones, sólo fueron consideradas aquellas frecuencias contenidas en esta banda de frecuencia que contiene la región espectral de las vocalizaciones de las aves.

En la Figura 4.2, se muestra una comparación de los espectrogramas obtenidos de las vocalizaciones de la especie *Colibri thalassinus*, en la Figura 4.2 (a) se puede ver en el espectrograma que la energía espectral de la vocalización del *Colibri thalassinus* se encuentra contenida en el rango de frecuencias de 2500 Hz a 10500 Hz aproximadamente, mientras que en la Figura 4.2 (b) se puede ver además de la energía espectral de la vocalización del *Colibri thalassinus,* una energía espectral correspondiente al ruido ambiental que se encuentra contenida en el rango de frecuencias de 0 Hz a 1500 Hz aproximadamente, gran parte del ruido ambiental es extraído gracias a la aplicación del filtro paso banda Butterworth de orden 12, no obstante, parte del ruido ambiental queda sin poder ser eliminado ya que en esas mismas bandas de frecuencia se encuentran vocalizaciones de las otras 11 especies de aves, esta es una de las razones que ocasionan una mala clasificación.

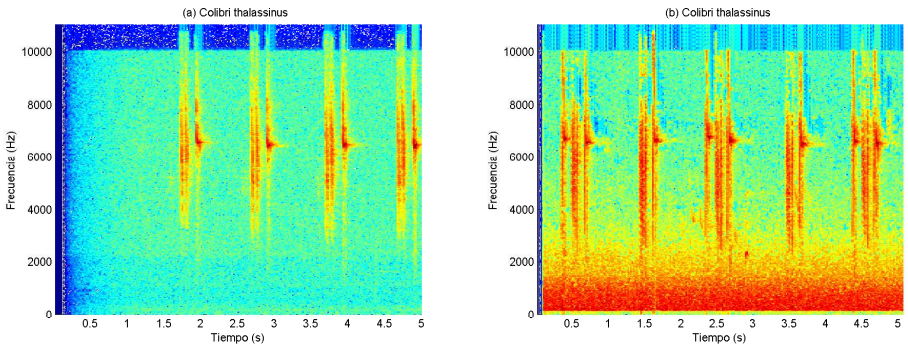

Figura 4.2: Comparación de espectrogramas de las vocalizaciones del *Colobri thalassinus*: (a) Espectrograma sin ruido ambiental, (b) Espectrograma con ruido ambiental.

La Figura 4.3, muestra 12 espectrogramas obtenidos de las vocalizaciones de las 12 especies de aves consideradas en esta tesis de grado. Las Figuras 4.3 (d), (e), (f), (g) y (h) corresponden al orden *Passeriformes* de acuerdo a la taxonomía de las aves,

asimismo, las Figuras 4.3 (i), (j), (k) y (l), pertenecen al orden *Tinamiformes* de acuerdo a la taxonomía de las aves.

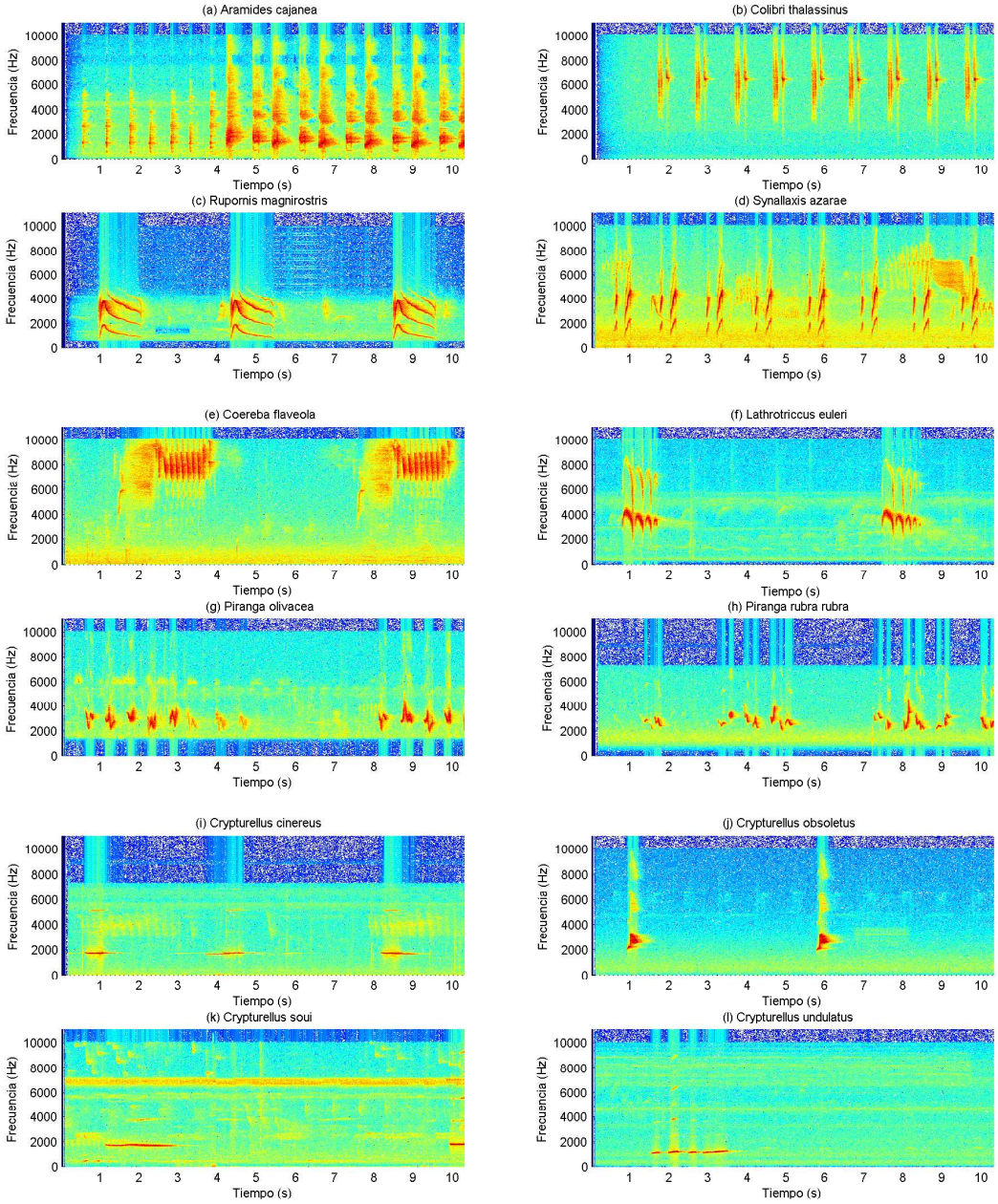

Figura 4.3: Espectrogramas obtenidos de las vocalizaciones de las 12 especies de aves: (a) *Aramides cajanea*, (b) *Colibri thalassinus*, (c) *Rupornis magnirostris*, (d) *Synallaxis Azarae*, (e) *Coereba flaveola*, (f) *Lathrotriccus euleri*, (g) *Piranga olivacea*, (h) *Piranga rubra rubra*, (i) *Crypturellus cinereus*, (j) *Crypturellus obsoletus*, (k) *Crypturellus soui*, (l) *Crypturellus undulatus*.

4.2.2. Módulo de parametrización

Este módulo permite convertir las señales de audio en un conjunto de características que las representen, características que sean lo más discriminantes posibles entre clases o especies, para poder así conseguir una mejor clasificación de éstas, por tal motivo lo que se busca es conseguir el mejor conjunto de características posible.

Como se explicó en la subsección 2.3.1, para poder obtener los parámetros MFCC de las señales de audio a clasificar, es decir, los coeficientes cepstrales a escala de frecuencias mel, la señal preprocesada debe pasar por una serie de etapas: enventanado, aplicación de la DFT, aplicación de un banco de filtros a escala a frecuencias mel, y finalmente la aplicación del logaritmo y la DCT. Como se muestra en la Figura 4.4, el punto de trabajo de esta tesis recae específicamente sobre la etapa del banco de filtros, siendo estos aprendidos mediante la técnica de la factorización de matrices no negativas (NMF), es decir que, a partir de las mismas señales de audio en el dominio de la frecuencia se pueda obtener el escalamiento y forma de los filtros, siendo ésta la base sobre la cual se espera mejorar las tasas de clasificación finales en comparación a las obtenidas por la técnica habitualmente utilizada MFCC. En pocas palabras, la intención es poder experimentar con parámetros MFCC como línea base y los obtenidos mediante el uso de la técnica NMF, denotados en este trabajo como Coeficientes Cepstrales basados en NMF (NMF_CC).

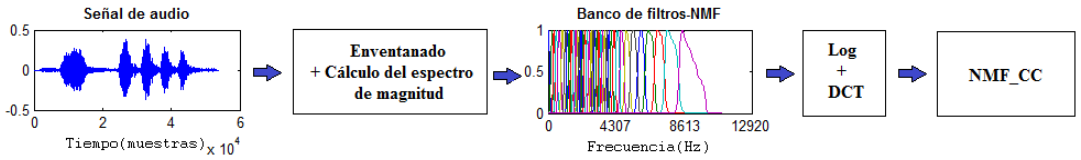

Figura 4.4: Diagrama de bloques del proceso de extracción de características mediante NMF.

Para la división de la señal en tramas, se utilizaron ventanas Hamming de 10 ms de duración con 5 ms de solapamiento para la clasificación en base a segmentos, y ventanas de 20 ms de duración con 10 ms de solapamiento para la clasificación en base a sílabas. Luego se aplica la transformada rápida de Fourier (FFT) para obtener el espectro de cada trama, después se realiza el cálculo de la log-energía en cada banda usando un banco de 40 filtros a escala de frecuencia mel (MFCC) o 40 filtros aprendidos mediante NMF (NMF_CC), según sea el caso, seguidamente se les aplica la transformada del coseno discreto (DCT), obteniéndose así los 12 primeros coeficientes cepstrales.

Finalmente, las características obtenidas por medio del banco de filtros, ya sean MFCC o NMF_CC, llamadas también características a corto plazo, son sintetizadas mediante la integración temporal de características, que resultan del cálculo de parámetros estadísticos (media, desviación estándar y simetría) de las característi-

cas a corto plazo, contenidas en cada segmento y sílaba, los cuales son la entrada al clasificador SVM.

4.2.2.1. Entrenamiento del banco de filtros basado en NMF

Para entrenar el banco de filtros basado en NMF, se hizo uso de las muestras escogidas para entrenamiento de la base de datos, específicamente los espectros de amplitud de las diferentes señales (V), usando una duración de trama de 10 ms con solape de 5 ms, además para la limpieza de los datos de entrenamiento se hizo uso de un filtro Butterworth de orden 12 con frecuencias de corte de 200 Hz a 10500 Hz, debido a que en dicho rango de frecuencias se encuentra la información de las vocalizaciones de las 12 especies de aves consideradas en esta tesis de grado.

La construcción del banco de filtros está basada en la teoría explicada en la subsección 2.3.2, donde se establece a dos matrices $W \epsilon \mathbb{R}_{+}^{N \times R}$ y $H \epsilon \mathbb{R}_{+}^{R \times M}$, éstas matrices corresponden a la factorización de la matriz $V \epsilon \mathbb{R}_{+}^{N \times M}$, que está formada por la magnitud de los espectrogramas de las diferentes vocalizaciones de las aves, las filas de esta matriz representadas por N es el número de bins de frecuencia, mientras que las columnas representadas por M representan el número total de tramas en el conjunto de entrenamiento.

Las matrices W y H, se obtienen usando las reglas de aprendizaje de las ecuaciones 2.15 y 2.16 (subsección 2.3.2). La matriz W contiene en sus columnas el número de Vectores Base Espectrales (*Spectral Basis Vectors*, SBVs) que representan las bases del espectro de magnitud de las señales acústicas de las aves, y que pueden interpretarse como las respuestas en frecuencia de los filtros, del banco de filtros auditivo de la parametrización propuesta NMF_CC, es decir, el número de filtros del banco de filtros a ser aprendido, que en este caso se ha establecido en 40 filtros ($R = 40$). Por otra parte la matriz H contiene a los coeficientes de activación o ganancia que combinados con los SBV´s devuelven la matriz V.

En la Figura 4.5 se muestran los 40 filtros, del banco de filtros convencional utilizado en la parametrización MFCC y del banco de filtros propuesto, hallado mediante NMF. Se puede observar que en la Figura 4.5 (a), existe una mayor concentración de filtros en las bajas frecuencias para el caso MFCC, mientras que dicha concentración de filtros se presenta en frecuencias medias para el caso NMF (Figura 4.5 (b)), específicamente en el rango de frecuencias de 1000 Hz a 4000 Hz, sugiriendo que esta banda de frecuencias es mucho más relevante y por lo tanto esencial para el sistema auditivo de las aves. Para frecuencias mayores de 4000 Hz, la distribución de los filtros es similar al banco de filtros triangular y para el rango de frecuencias menores que 1000 Hz se tiene menos filtros con mayores anchos de banda.

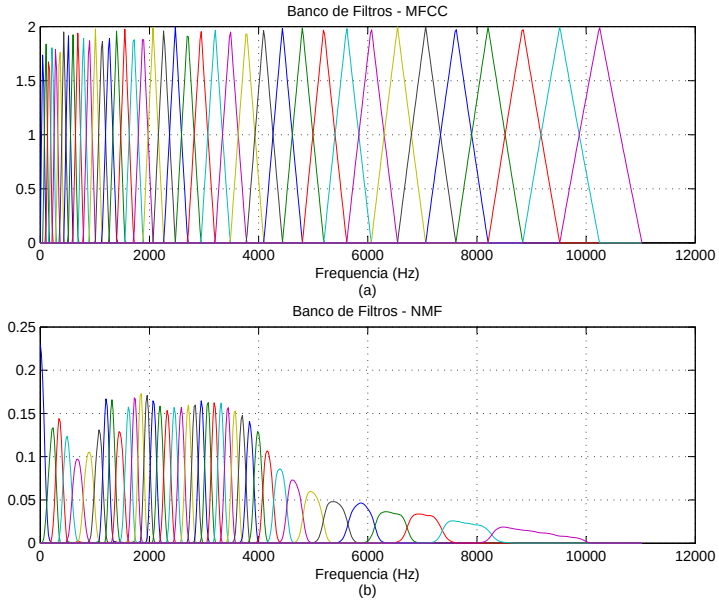

Figura 4.5: Comparación de banco de filtros, (a) Banco de filtros MFCC, (b) Banco de filtros NMF.

4.2.3. Módulo de clasificación

Al clasificador ingresan características segmentales obtenidas mediante el uso de la técnica de la integración temporal de características, dichas características segmentales pueden ser segmentos o sílabas de la señal de audio de las aves, el clasificador diseñado aquí se encuentra basado en la técnica de la Máquina de Vectores de Soporte (SVM).

Como se explicó a detalle en la subsección 2.4.1, para resolver el problema de la clasificación, dicha técnica se basa en la búsqueda de un hiperplano óptimo de separación, hallado por medio de las características de diferentes clases que se encuentren lo más próximas a dicho hiperplano, siendo éstas las llamadas vectores de soporte. Sin embargo, la dificultad recae cuando dichas características no son linealmente separables (lo que ocurre en nuestro caso), ocasionando que no se pueda encontrar dicho hiperplano, por tal motivo se hace necesario el uso de una función kernel que permita hallar una dimensión mayor donde las características sean separables de manera lineal y por ende poder encontrar el hiperplano óptimo de separación.

La función utilizada es la kernel RBF (*Radial Basis Function*), sus parámetros de ajuste óptimos fueron obtenidos mediante una validación cruzada 6-fold en cada uno de los sub-experimentos en que fue dividido el conjunto de entrenamiento utilizado, explicado en la sección 4.1.

En la etapa de entrenamiento, para encontrar los modelos acústicos de cada una de las especies de aves se hizo uso de la configuración SVM de uno contra uno, y para la

etapa de prueba cada archivo de prueba es comparado con cada modelo entrenado de la etapa de entrenamiento. Para la decisión final de cada especie de ave se utilizó un esquema de mayoría de votos, donde la clase mayormente identificada por el clasificador es clasificada como la clase final.

4.3. Experimentos preliminares

Como se mencionó, estos experimentos se realizaron con la finalidad de poder establecer la trama óptima a utilizar para la clasificación basada tanto en segmentos como en sílabas. Los experimentos se realizaron haciendo uso de los 12 MFCC más la log-energía y los resultados son mostrados como tasas de porcentaje de clasificación.

4.3.1. Análisis de trama para segmentos

Para poder determinar la duración óptima de la trama para la clasificación en base a segmentos, se realizaron 3 experimentos que permitieron obtener resultados de las tasas de clasificación haciendo uso de 3 diferentes duraciones de trama: 6 ms, 10 ms y 20 ms; los experimentos se realizaron haciendo uso de los 12 MFCC más la log-energía de los mismos.

Duración de trama (ms)	Tasa de clasificación promedio de segmentos (%)	Tasa de clasificación promedio de ficheros (%)
6	56.07	69.64
10	55.57	70.52
20	54.44	69.48

Cuadro 4.2: Promedio de la las tasas de clasificación (T_C) basada en segmentos para cada una de las duraciones de trama.

En el Cuadro 4.2 se muestra los resultados de los 3 experimentos realizados para cada una de las 3 duraciones de trama, que están expresados como promedio de las tasas de clasificación de las 12 especies de aves. La tasa de clasificación promedio de ficheros hace referencia a las grabaciones de audio que fueron correctamente reconocidas debido a que contenían una mayoría de segmentos correctamente reconocidos.

De los resultados se puede observar que en base a sólo segmentos la duración de trama óptima podría ser la de 6 ms, sin embargo, en base a ficheros la duración de trama óptima de trabajo sería la de 10 ms, la decisión final aquí fue optar por la duración de trama que en promedio de segmentos y ficheros, proporcione la mayor tasa de clasificación, que en este caso fue la trama de duración de 10 ms.

La Figura 4.6 muestra resultados de las tasas de clasificación basada en segmentos para cada una de las 12 especies de acuerdo a cada una de las duraciones de trama.

Se puede observar también que con la trama de 10 ms se obtienen mejores resultados para la mayoría de las especies.

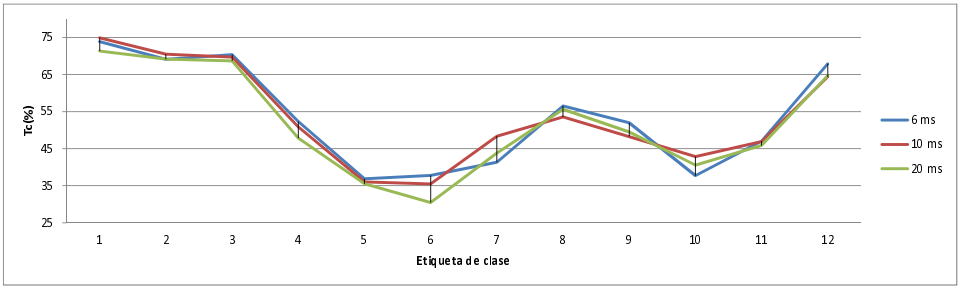

Figura 4.6: Evolución de las tasas de clasificación (T_C) basada en segmentos, para cada especie de acuerdo a la duración de trama utilizada.

4.3.2. Análisis de trama para sílabas

Para poder determinar la duración óptima de la trama para la clasificación en base a sílabas, se realizaron 3 experimentos que permitieron obtener resultados de las tasas de clasificación haciendo uso de 3 diferentes duraciones de trama: 6 ms, 10 ms y 20 ms; los experimentos se realizaron haciendo uso de los 12 MFCC más la log-energía de los mismos.

Duración de trama (ms)	Tasa de clasificación promedio de sílabas (%)	Tasa de clasificación promedio de ficheros (%)
6	55.67	69.40
10	58.28	70.20
20	60.84	69.56

Cuadro 4.3: Promedio de la las tasas de clasificación (T_C) basada en sílabas para cada una de las duraciones de trama.

En el Cuadro 4.3 se muestra los resultados de los 3 experimentos realizados para cada una de las 3 duraciones de trama, que están expresados como promedio de las tasas de clasificación de las 12 especies de aves. La tasa de clasificación promedio de ficheros hace referencia a las grabaciones de audio que fueron correctamente reconocidas debido a que contenían una mayoría de sílabas correctamente reconocidas.

De los resultados se puede observar que en base a sólo sílabas, la duración de trama óptima es la de 20 ms, sin embargo, en base a ficheros la duración de trama óptima es la de 10 ms, la decisión final aquí fue optar por la trama que en promedio de segmentos y ficheros, proporcione la mayor tasa de clasificación la cuál fue la trama de duración de 20 ms, por lo tanto 20 ms es la trama elegida para los siguientes experimentos basados en la clasificación por sílabas.

La Figura 4.7 muestra resultados de las tasas de clasificación basada en sílabas para cada una de las 12 especies de acuerdo a cada una de las duraciones de trama. Se puede observar también que con la trama de 20 ms se obtienen mejores resultados para la mayoría de las especies.

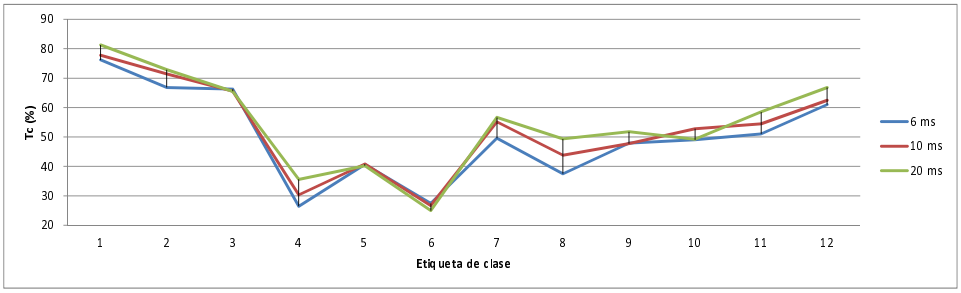

Figura 4.7: Evolución de las tasas de clasificación (T_C) basada en sílabas, para cada especie de acuerdo a la duración de trama utilizada.

5 Resultados experimentales

En este capítulo se muestran los resultados obtenidos mediante la realización de la parte experimental. La intención principal es poder experimentar con parámetros MFCC y con parámetros obtenidos mediante el uso de la técnica NMF, para finalmente poder comparar los resultados de clasificación obtenidos mediante dichas técnicas. Del análisis de éstos experimentos se podrán obtener conclusiones sobre cuál de los dos tipos de parámetros resulta ser más apropiado para una buena clasificación de los archivos de audio correspondientes a diferentes especies de aves, que es la finalidad última de esta tesis de grado.

La configuración final usada en los experimentos, esta descrita de la siguiente manera:

- Para la extracción de las características a corto plazo; se analizaron las señales de audio cada 5 ms usando una ventana Hamming de 10 ms para la clasificación basada en segmentos, asimismo, se analizaron las señales de audio cada 10 ms usando una ventana Hamming de 20 ms para la clasificación basada en sílabas.
- La parametrización convencional (MFCC) y la propuesta (NMF), constan de 12 coeficientes cepstrales, más la log-energía de cada trama y las primeras derivadas correspondientes (cuando se indique).
- En ambos casos, los bancos de filtros MFCC y NMF_CC, consisten de 40 filtros.
- Con respecto a la aplicación de NMF para el diseño del banco de filtros, para cada grupo de entrenamiento y prueba, de los 6 grupos en que fue extendida la base de datos mediante el uso del esquema de validación cruzada 6-fold (descrita en la sección 4.1), se obtuvo un banco de filtros usando NMF aplicando la teoría descrita en la subsección 2.3.2, sobre el correspondiente grupo de entrenamiento. En tal caso, para el cálculo del espectograma (V), el tamaño usado de la ventana fue de 10 ms con superposición del 50 %.
- En ambos casos (MFCC y NMF_CC), se hizo uso de la técnica de la integración temporal de características basada en los estadísticos (media, desviación estándar y simetría) de las características a corto plazo: la primera integración temporal se realizó en base a segmentos con duración de 2000 ms con solapes de 1000 ms; mientras que la segunda se realizó en base a la duración de las sílabas que presentaban cada una de las especies. Es decir, los experimentos muestran las tasas de clasificación basadas en segmentos y sílabas.

Los experimentos realizados, para cada uno de los dos tipos de clasificación, basado en segmentos y basado en sílabas, se muestran en el siguiente orden:

- Experimentos con parámetros MFCC más log-energía.
- Experimentos con parámetros NMF_CC más log-energía.
- Experimentos con parámetros MFCC más log-energía y parámetros delta.
- Experimentos con parámetros NMF_CC más log-energía y parámetros delta.

La log- energía se calcula por cada trama, esta considera la energía de la señal bioacústica y el logarítmo sirve para ponderar el amplio rango dinámico que presenta dicha señal. De la misma forma, los parámetros delta, son considerados debido a que permiten tomar en cuenta las características dinámicas (evolución de los coeficientes a lo largo del tiempo) tanto de los MFCC como de los NMF_CC.

Para terminar, los experimentos muestran resultados basados en segmentos, sílabas y ficheros, este último hace referencia a las grabaciones de audio que fueron correctamente reconocidas debido a que contenían una mayoría de segmentos o sílabas correctamente reconocidas, es decir, los ficheros son grabaciones de audio que se clasifican de acuerdo a los segmentos o sílabas que sean correctamente reconocidas.

5.1. Resultados experimentales basado en segmentos

En este apartado se muestran los resultados experimentales basados en segmentos, obtenidos haciendo uso de las parametrizaciones MFCC y NMF. La duración de trama a utilizar es la determinada en la subsección 4.3.1, que es de 10 ms con solapes de 5 ms. Los resultados se muestran en forma de matrices de confusión, que contienen tasas de clasificación para las 12 especies de aves: *Aramides cajanea, Coereba flaveola, Colibri thalassinus, Crypturellus cinereus, Crypturellus obsoletus, Crypturellus soui, Crypturellus undulatus, Lathrotricus euleri, Piranga olivacea, Piranga rubra rubra, Rupornis magnirostris, Synallaxis azarae*, representadas por sus respectivas etiquetas de clase (1 al 12).

5.1.1. Experimentos MFCC y NMF con parámetros: CC+logE

En este apartado se muestran los resultados de dos experimentos realizados en base a las técnicas de parametrización: MFCC y NMF_CC, haciendo uso de 12 parámetros CC (Coeficientes Cepstrales) más la log-energía.

La Figura 5.1, contiene dos matrices de confusión que muestran las tasas de clasificación para cada una de las especies, resaltadas en sus respectivas diagonales principales, de la misma manera se muestran los porcentajes de confusión entre especies en las celdas restantes. Las columnas corresponden a la clase correcta, mientras que las filas son la hipótesis, es decir, la posible clase correcta.

	1	2	3	4	5	6	7	8	9	10	11	12
1	83.33	0.52	0.79	1.11	0.00	0.00	1.54	0.00	0.00	0.00	0.56	0.00
2	3.33	91.10	22.22	6.67	8.14	9.47	1.54	4.39	3.33	0.00	9.04	4.00
3	0.00	3.66	72.22	0.00	0.00	0.00	0.00	0.00	3.33	1.52	1.69	1.60
4	0.00	1.57	0.79	63.33	6.98	22.11	7.69	2.63	3.33	0.00	3.39	2.40
5	1.67	0.00	0.79	3.33	51.16	11.58	3.08	2.63	0.00	6.06	2.82	0.00
6	0.00	0.00	0.00	21.11	20.93	38.95	9.23	4.39	0.00	3.03	2.82	0.00
7	5.00	0.52	0.00	2.22	1.16	4.21	70.77	0.88	0.00	1.52	1.13	0.00
8	0.00	0.00	0.79	1.11	3.49	3.16	1.54	74.56	5.00	6.06	4.52	0.80
9	0.00	0.00	0.79	0.00	2.33	0.00	0.00	1.75	76.67	16.67	1.13	0.00
10	1.67	0.00	0.00	0.00	1.16	0.00	1.54	1.75	5.00	57.58	1.13	0.80
11	5.00	2.62	0.79	1.11	4.65	9.47	3.08	5.26	3.33	6.06	68.93	14.40
12	0.00	0.00	0.79	0.00	0.00	1.05	0.00	1.75	0.00	1.52	2.82	76.00

(a)

	1	2	3	4	5	6	7	8	9	10	11	12
1	83.33	0.00	0.00	1.11	0.00	1.05	6.15	0.88	0.00	1.52	1.69	0.00
2	6.67	85.86	24.60	6.67	9.30	10.53	4.62	6.14	3.33	3.03	8.47	3.20
3	0.00	7.33	69.84	0.00	0.00	1.05	0.00	0.00	5.00	1.52	0.56	2.40
4	0.00	0.52	0.79	62.22	6.98	13.68	6.15	1.75	0.00	0.00	2.82	1.60
5	1.67	0.52	0.00	2.22	52.33	10.53	6.15	1.75	1.67	9.09	2.26	0.80
6	1.67	0.00	0.00	17.78	17.44	47.37	9.23	2.63	1.67	1.52	4.52	0.00
7	0.00	1.05	0.00	3.33	2.33	5.26	66.15	0.00	1.67	0.00	1.69	0.00
8	0.00	0.52	0.00	0.00	3.49	3.16	0.00	79.82	5.00	9.09	3.39	0.00
9	0.00	0.00	1.59	0.00	0.00	0.00	0.00	1.75	76.67	10.61	0.00	0.00
10	1.67	0.52	0.00	0.00	1.16	0.00	0.00	0.00	3.33	59.09	1.13	0.80
11	5.00	3.14	1.59	6.67	6.98	6.32	1.54	4.39	1.67	4.55	72.32	8.80
12	0.00	0.52	1.59	0.00	0.00	1.05	0.00	0.88	0.00	0.00	1.13	82.40

(b)

Figura 5.1: Matrices de confusión [%] a nivel de ficheros para la parametrización CC+logE , (a) Basada en MFCC; (b) Basada en NMF_CC.

La matriz de la Figura 5.1 (a), muestra resultados obtenidos mediante la parametrización MFCC y su porcentaje promedio de clasificación es de 70.52 %, se puede observar también que las tasas de clasificación más altas se presentan para las especies: *Aramides cajanea, Coereba flaveola y Piranga olivacea* (1, 2 y 9), que son de 83.33 %, 91.10 % y 76.67 % respectivamente, mientras que las tasas de clasificación más bajas se presentan para las especies: *Crypturellus obsoletus, Crypturellus soui y Piranga rubra rubra* (5, 6 y 10), que son de 51.16 %, 38.95 % y 57.58 % respectivamente; por otra parte la matriz de la Figura 5.1 (b), muestra resultados obtenidos mediante la parametrización NMF_CC con un porcentaje de clasificación promedio de 71.55 %, las tasas de clasificación más altas se presentan para las especies: *Aramides cajanea, Coereba flaveola y Synallaxis azarae* (1, 2 y 12), que son de 83.33 %, 85.86 % y 82.40 % respectivamente, mientras que las tasas de clasificación más bajas se presentan para las especies: *Crypturellus obsoletus, Crypturellus soui y Piranga rubra rubra* (5, 6 y 10), que son de 52.33 %, 47.37 % y 59.09 % respectivamente. En particular, la especie con mayor porcentaje de confusión es el *Colibri thalassinus*

(3) reconocida como la especie *Coereba flaveola* (2), con un porcentaje de 22.22 % y 24.60 % para MFCC y NMF_CC respectivamente.

Como puede observarse las tasas de clasificación más bajas se presentan para las mismas especies en ambos casos (MFCC y NMF_CC), concluyendo que las obtenidas mediante NMF_CC, superan a sus correspondientes obtenidas mediante MFCC, asimismo la tasa de clasificación promedio de NMF_CC supera también a la tasa de clasificación promedio de MFCC.

5.1.2. Experimentos MFCC y NMF con parámetros: CC+logE+△

En este apartado se muestran los resultados de dos experimentos realizados en base a las técnicas de parametrización: MFCC y NMF_CC, haciendo uso de 12 parámetros CC (*Coeficientes Cepstrales*) más la log-energía y la primera derivada ($\triangle MFCC$ y $\triangle NMF\_CC$, respectivamente).

	1	2	3	4	5	6	7	8	9	10	11	12
1	85.00	0.00	0.00	1.11	0.00	0.00	3.08	0.00	0.00	0.00	0.56	0.00
2	3.33	91.62	11.11	0.00	3.49	7.37	1.54	3.51	5.00	3.03	6.78	3.20
3	0.00	4.71	84.92	1.11	0.00	0.00	0.00	1.75	0.00	3.03	1.13	4.00
4	0.00	0.00	0.79	64.44	6.98	21.05	9.23	0.88	1.67	1.52	3.95	1.60
5	5.00	0.00	0.00	2.22	56.98	5.26	3.08	0.88	0.00	1.52	1.69	0.00
6	0.00	0.52	0.79	22.22	16.28	47.37	4.62	2.63	1.67	0.00	3.39	0.00
7	3.33	1.05	0.00	2.22	1.16	3.16	72.31	0.00	0.00	0.00	1.13	0.00
8	0.00	0.00	0.00	1.11	4.65	6.32	1.54	77.19	3.33	3.03	5.08	0.80
9	0.00	0.00	0.79	0.00	1.16	0.00	0.00	1.75	78.33	10.61	1.13	0.80
10	0.00	0.00	0.00	0.00	1.16	0.00	0.00	1.75	3.33	66.67	0.00	2.40
11	3.33	2.09	0.79	4.44	8.14	8.42	4.62	9.65	6.67	7.58	72.32	8.00
12	0.00	0.00	0.79	1.11	0.00	1.05	0.00	0.00	0.00	3.03	2.82	79.20

(a)

	1	2	3	4	5	6	7	8	9	10	11	12
1	86.67	0.00	0.00	2.22	0.00	0.00	4.62	0.88	0.00	0.00	1.13	0.00
2	5.00	86.39	14.29	2.22	5.81	8.42	3.08	7.89	3.33	4.55	6.78	3.20
3	0.00	6.28	80.16	1.11	1.16	0.00	0.00	1.75	1.67	3.03	0.56	1.60
4	0.00	0.52	0.00	61.11	8.14	17.89	3.08	0.88	1.67	0.00	2.26	0.00
5	1.67	0.00	0.79	2.22	52.33	5.26	7.69	0.88	0.00	1.52	3.39	1.60
6	1.67	1.05	0.79	23.33	17.44	56.84	3.08	1.75	0.00	1.52	3.39	0.80
7	0.00	1.05	0.00	3.33	0.00	3.16	70.77	0.00	0.00	0.00	1.13	0.80
8	0.00	0.00	0.00	1.11	3.49	4.21	1.54	79.82	5.00	1.52	3.39	0.00
9	0.00	0.00	0.79	0.00	1.16	0.00	0.00	0.88	78.33	10.61	0.00	0.00
10	0.00	0.00	0.00	0.00	2.33	0.00	0.00	1.75	1.67	71.21	0.56	0.80
11	3.33	4.71	0.79	3.33	6.98	3.16	4.62	3.51	8.33	6.06	76.84	6.40
12	1.67	0.00	2.38	0.00	1.16	1.05	1.54	0.00	0.00	0.00	0.56	84.80

(b)

Figura 5.2: Matrices de confusión [%] a nivel de ficheros para la parametrización CC+logE+△ , (a) Basada en MFCC; (b) Basada en NMF_CC.

La Figura 5.2, contiene dos matrices de confusión que muestran las tasas de clasifi-

cación para cada una de las especies, resaltadas en sus respectivas diagonales principales, de la misma manera se muestran los porcentajes de confusión entre especies en las celdas restantes. Las columnas corresponden a la clase correcta, mientras que las filas son la hipótesis, es decir, la posible clase correcta.

La matriz de la Figura 5.2 (a) muestra resultados obtenidos mediante la parametrización MFCC y su porcentaje promedio de clasificación es de 74.74 %, se puede observar también que las tasas de clasificación más altas se presentan para las especies: *Aramides cajanea, Coereba flaveola y Colibri thalassinus* (1,2 y 3) que son de 85.00 %, 91.62 % y 84.92 % respectivamente, mientras que las tasas de clasificación más bajas se presentan para las especies: *Crypturellus obsoletus y Crypturellus soui* (5 y 6) que son de 56.98 % y 47.37 % respectivamente; por otra parte la matriz de la Figura 5.2 (b) muestra resultados obtenidos mediante la parametrización NMF_CC con un porcentaje de clasificación promedio de 75.30 %, las tasas de clasificación más altas se presentan para las especies: *Aramides cajanea, Coereba flaveola y Synallaxis azarae* (1, 2 y 12), que son de 86.67 %, 86.39 % y 84.80 % respectivamente, mientras que las tasas de clasificación más bajas se presentan para las especies: *Crypturellus obsoletus y Crypturellus soui* (5 y 6) que son de 52.33 % y 56.84 % respectivamente. En particular, la especie con mayor porcentaje de confusión es el *Crypturellus cinereus* (4) reconocida como la especie *Crypturellus soui* (6), con un porcentaje de 22.22 % y 23.33 % para MFCC y NMF_CC respectivamente.

Como puede observarse la tasa de clasificación más baja obtenida mediante MFCC para la especie *Crypturellus obsoletus* (5) es moderadamente superior a la obtenida mediante NMF_CC, mientras que la tasa de clasificación más baja para la especie *Crypturellus soui* (6) obtenida mediante NMF_CC supera por mucho a su correspondiente obtenida mediante MFCC, asimismo la tasa de clasificación promedio de NMF_CC supera también a la tasa de clasificación promedio de MFCC. Cabe resaltar también, que adicionar el uso de la primera derivada ($\triangle$) sobre estos resultados experimentales, elevó las tasas de clasificación con respecto a los resultados obtenidos en el apartado 5.1.1, donde sólo se hizo uso de los 12 CC más la log-energía.

5.2. Resultados experimentales basado en sílabas

En este apartado se muestran los resultados experimentales basados en sílabas, obtenidos haciendo uso de las parametrizaciones MFCC y NMF. La duración de trama a utilizar es la determinada en la subsección 4.3.2, que es de 20 ms con solapes de 10 ms. Los resultados se muestran en matrices de confusión que contienen tasas de clasificación para las 12 especies de aves: *Aramides cajanea, Coereba flaveola, Colibri thalassinus, Crypturellus cinereus, Crypturellus obsoletus, Crypturellus soui, Crypturellus undulatus, Lathrotricus euleri, Piranga olivacea, Piranga rubra rubra, Rupornis magnirostris, Synallaxis azarae*, representadas por sus respectivas etiquetas de clase (1 al 12).

5.2.1. Experimentos MFCC y NMF con parámetros: CC+logE

En este apartado se muestran los resultados de dos experimentos realizados en base a las técnicas de parametrización: MFCC y NMF_CC, haciendo uso de 12 parámetros CC (*Coeficientes Cepstrales*) más la log-energía.

	1	2	3	4	5	6	7	8	9	10	11	12
1	85.00	2.62	3.17	5.56	6.98	6.32	9.23	3.51	0.00	1.52	1.69	2.40
2	5.00	89.01	18.25	14.44	13.95	16.84	3.08	10.53	3.33	0.00	8.47	4.00
3	0.00	7.85	76.19	3.33	2.33	2.11	1.54	0.88	1.67	0.00	2.26	3.20
4	1.67	0.00	0.00	53.33	5.81	17.89	3.08	0.88	0.00	0.00	0.56	0.80
5	1.67	0.00	0.00	3.33	47.67	7.37	3.08	0.88	0.00	4.55	1.69	0.00
6	0.00	0.00	0.00	6.67	5.81	33.68	1.54	3.51	0.00	0.00	1.69	0.00
7	1.67	0.00	0.00	3.33	1.16	2.11	73.85	0.88	1.67	0.00	0.00	0.00
8	1.67	0.00	0.79	2.22	3.49	2.11	0.00	64.91	3.33	6.06	2.82	0.00
9	0.00	0.00	1.59	0.00	2.33	0.00	0.00	3.51	73.33	13.64	1.13	1.60
10	1.67	0.00	0.00	0.00	2.33	1.05	3.08	1.75	15.00	66.67	1.13	2.40
11	0.00	0.00	0.00	4.44	5.81	4.21	0.00	4.39	1.67	6.06	69.49	4.00
12	1.67	0.52	0.00	3.33	2.33	6.32	1.54	4.39	0.00	1.52	9.04	81.60

(a)

	1	2	3	4	5	6	7	8	9	10	11	12
1	83.33	3.14	3.17	4.44	9.30	7.37	18.46	4.39	0.00	0.00	1.13	4.80
2	6.67	89.53	23.02	16.67	9.30	17.89	6.15	13.16	3.33	6.06	7.91	4.00
3	0.00	5.76	72.22	4.44	3.49	6.32	1.54	0.00	3.33	1.52	1.13	3.20
4	0.00	0.00	0.00	55.56	5.81	21.05	0.00	0.00	0.00	0.00	1.69	0.80
5	1.67	0.00	0.00	2.22	53.49	9.47	1.54	0.00	0.00	3.03	3.39	0.00
6	0.00	0.00	0.00	7.78	6.98	29.47	0.00	0.00	0.00	0.00	0.56	0.80
7	1.67	0.00	0.00	1.11	1.16	1.05	69.23	0.88	0.00	0.00	0.00	0.00
8	3.33	1.05	0.79	3.33	2.33	2.11	0.00	71.05	3.33	1.52	1.69	0.80
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.63	80.00	10.61	2.82	0.80
10	3.33	0.00	0.79	0.00	2.33	2.11	0.00	4.39	8.33	77.27	1.13	0.80
11	0.00	0.52	0.00	3.33	3.49	2.11	0.00	0.88	1.67	0.00	75.14	2.40
12	0.00	0.00	0.00	1.11	2.33	1.05	3.08	2.63	0.00	0.00	3.39	81.60

(b)

Figura 5.3: Matrices de confusión [%] a nivel de ficheros para la parametrización CC+logE , (a) Basada en MFCC; (b) Basada en NMF_CC.

La Figura 5.3, contiene dos matrices de confusión que muestran las tasas de clasificación para cada una de las especies, resaltadas en sus respectivas diagonales principales, de la misma manera se muestran los porcentajes de confusión entre especies en las celdas restantes. Las columnas corresponden a la clase correcta, mientras que las filas son la hipótesis, es decir, la posible clase correcta.

La matriz de la Figura 5.3 (a), muestra resultados obtenidos mediante la parametrización MFCC y su porcentaje promedio de clasificación es de 69.56 %, se puede observar también que las tasas de clasificación más altas se presentan para las especies: *Aramides cajanea, Coereba flaveola y Synallaxis azarae* (1, 2 y 12), que son de 85.00 %, 89.01 % y 81.60 % respectivamente, mientras que las tasas de clasifica-

ción más bajas se presentan para las especies: *Crypturellus cinereus, Crypturellus obsoletus y Crypturellus soui* (4, 5 y 6) que son de 53.33 %, 47.67 % y 33.68 % respectivamente; por otra parte la matriz de la Figura 5.3 (b), muestra resultados obtenidos mediante la parametrización NMF_CC con un porcentaje de clasificación promedio de 71.39 %, las tasas de clasificación más altas se presentan para las especies: *Aramides cajanea, Coereba flaveola y Synallaxis azarae* (1, 2 y 12), que son de 83.33 %, 89.53 % y 81.60 % respectivamente, mientras que las tasas de clasificación más bajas también se presentan para las especies: *Crypturellus cinereus, Crypturellus obsoletus y Crypturellus soui* (4, 5 y 6), que son de 55.56 %, 53.49 % y 29.47 % respectivamente. En particular, la especie con mayor porcentaje de confusión es el *Colibri thalassinus* (3) reconocida como la especie *Coereba flaveola* (2), con un porcentaje de 18.25 % y 23.02 % para MFCC y NMF_CC respectivamente.

Se puede observar que para las especies: *Crypturellus cinereus y Crypturellus obsoletus* (4 y 5) las tasas de clasificación más bajas, son mayores obtenidas mediante NMF_CC que con MFCC, y para la especie *Crypturellus soui* (6) la tasa de clasificación obtenida mediante MFCC es moderadamente mayor a la obtenida con NMF_CC, aun así la tasa de clasificación promedio de NMF_CC supera a la tasa de clasificación promedio de MFCC.

5.2.2. Experimentos MFCC y NMF con parámetros: CC+logE+△

En este apartado se muestran los resultados de dos experimentos realizados en base a las técnicas de parametrización: MFCC y NMF_CC, haciendo uso de 12 parámetros CC (*Coeficientes Cepstrales*) más la log-energía y la primera derivada ($\triangle MFCC$ y $\triangle NMF\_CC$, respectivamente).

La Figura 5.4, contiene dos matrices de confusión que muestran las tasas de clasificación para cada una de las especies, resaltadas en sus respectivas diagonales principales, de la misma manera se muestran los porcentajes de confusión entre especies en las celdas restantes. Las columnas corresponden a la clase correcta, mientras que las filas son la hipótesis, es decir, la posible clase correcta.

La matriz de la Figura 5.4 (a), muestra resultados obtenidos mediante la parametrización MFCC y su porcentaje promedio de clasificación es de 75 %, donde se puede ver también, que las tasas de clasificación más altas se presentan para las especies: *Aramides cajanea, Coereba flaveola y Synallaxis azarae* (1,2 y 12) que son de 88.89 %, 93.72 % y 85.25 % respectivamente, mientras que las tasas de clasificación más bajas se presentan para las especies: *Crypturellus cinereus, Crypturellus obsoletus y Crypturellus soui* (4, 5 y 6) que son de 56.67 %, 54.65 % y 44.21 % respectivamente; por otra parte la matriz de la Figura 5.4 (b), muestra resultados obtenidos mediante una parametrización NMF_CC con un porcentaje de clasificación promedio de 76.36 %, las tasas de clasificación más altas se presentan para las especies: *Aramides cajanea, Coereba flaveola y Synallaxis azarae* (1, 2 y 12), que son de 90.48 %, 91.62 % y 86.89 % respectivamente, igualmente las tasas de clasificación más bajas también

	1	2	3	4	5	6	7	8	9	10	11	12
1	88.89	1.57	1.60	5.56	5.81	5.26	13.85	2.63	0.00	3.13	1.69	3.28
2	3.17	93.72	12.00	12.22	11.63	11.58	3.08	6.14	5.00	3.13	7.91	2.46
3	1.59	3.66	83.20	2.22	3.49	2.11	0.00	0.00	1.67	1.56	1.69	4.10
4	0.00	0.00	0.00	56.67	4.65	17.89	1.54	0.88	0.00	0.00	1.13	0.82
5	0.00	0.00	0.80	2.22	54.65	6.32	1.54	1.75	0.00	1.56	1.69	0.00
6	0.00	0.00	0.00	11.11	4.65	44.21	1.54	1.75	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	2.22	0.00	1.05	73.85	0.00	0.00	0.00	0.00	0.00
8	3.17	0.52	0.80	3.33	5.81	3.16	0.00	72.81	1.67	4.69	7.34	0.82
9	0.00	0.00	0.80	0.00	0.00	0.00	1.54	2.63	83.33	9.38	0.00	0.00
10	1.59	0.00	0.00	0.00	2.33	0.00	0.00	2.63	6.67	68.75	1.13	1.64
11	0.00	0.00	0.80	3.33	4.65	6.32	1.54	3.51	1.67	4.69	74.01	1.64
12	1.59	0.52	0.00	1.11	2.33	2.11	1.54	5.26	0.00	3.13	3.39	85.25

(a)

	1	2	3	4	5	6	7	8	9	10	11	12
1	90.48	2.09	0.80	4.44	8.14	7.37	12.31	2.63	0.00	0.00	1.13	4.10
2	3.17	91.62	14.40	13.33	10.47	17.89	3.08	7.02	5.00	1.56	6.21	1.64
3	0.00	4.71	80.00	1.11	2.33	1.05	0.00	0.00	0.00	3.13	1.13	2.46
4	0.00	0.00	0.00	58.89	9.30	12.63	0.00	0.88	0.00	0.00	1.69	0.82
5	0.00	0.00	0.00	3.33	54.65	9.47	3.08	0.88	0.00	1.56	2.26	0.00
6	0.00	0.52	0.00	12.22	3.49	42.11	0.00	0.88	0.00	0.00	0.56	0.00
7	0.00	0.00	0.00	1.11	0.00	1.05	78.46	0.00	0.00	0.00	0.56	0.82
8	1.59	1.05	0.80	1.11	3.49	1.05	0.00	78.07	1.67	1.56	3.39	0.00
9	0.00	0.00	0.00	0.00	2.33	0.00	0.00	1.75	85.00	6.25	0.00	0.00
10	3.17	0.00	0.80	0.00	1.16	1.05	0.00	3.51	8.33	82.81	2.26	0.82
11	0.00	0.00	0.00	2.22	3.49	3.16	0.00	0.88	0.00	0.00	75.71	2.46
12	1.59	0.00	3.20	2.22	1.16	3.16	3.08	3.51	0.00	3.13	5.08	86.89

(b)

Figura 5.4: Matrices de confusión [%] a nivel de ficheros para la parametrización CC+logE+△ , (a) Basada en MFCC; (b) Basada en NMF_CC.

se presentan para las especies: *Crypturellus cinereus, Crypturellus obsoletus y Crypturellus soui* (4, 5 y 6) que son de 58.89 %, 54.65 % y 42.11 % respectivamente. En particular, la especie con mayor porcentaje de confusión es el *Crypturellus soui* (6) reconocida como la especie *Crypturellus cinereus* (4), con un porcentaje de 17.89 % para MFCC y reconocida como la especie *Coereba flaveola* (2), con un porcentaje de 17.89 % para NMF_CC.

Se puede observar que la tasa de clasificación más baja, obtenida mediante NMF_CC para la especie *Crypturellus cinereus* (4) es superior a la obtenida mediante MFCC, mientras que la tasa de clasificación más baja, para la especie *Crypturellus obsoletus* (5) obtenida mediante NMF_CC se mantiene a la obtenida mediante MFCC, y la tasa de clasificación más baja, obtenida mediante MFCC para la especie *Crypturellus soui* (6) es moderadamente mayor a la obtenida mediante NMF_CC. Aun así la tasa de clasificación promedio de NMF_CC supera a la tasa de clasificación promedio de MFCC. Cabe resaltar también que, adicionar el uso de la primera derivada (△) sobre estos resultados experimentales, elevó las tasas de clasificación con respecto a

los resultados obtenidos en el apartado 5.2.1, donde sólo se hizo uso de los 12 CC más la log-energía.

5.3. Comparación de resultados

A manera de resumen, en este apartado se presenta un cuadro comparativo de los experimentos presentados en las secciones anteriores, los mismos que se encuentran basados en la clasificación de segmentos como en la clasificación de sílabas, de la misma manera, también se muestran los resultados obtenidos mediante el método habitual de parametrización MFCC frente a la parametrización propuesta NMF_CC.

Características a corto plazo	Parametrización MFCC		Parametrización NMF_CC	
	Experimentos basados en segmentos			
	Segmento [%]	Fichero [%]	Segmento [%]	Fichero [%]
12 CC+logE	55.57± 0.84	70.52 ± 2.52	56.13 ± 0.84	71.55 ± 2.5
12 CC+logE+△	57.88 ± 0.84	74.74 ± 2.40	59.28 ± 0.84	75.30 ± 2.39
	Experimentos basados en sílabas			
	Sílaba [%]	Fichero [%]	Sílaba [%]	Fichero [%]
12 CC+logE	60.84 ± 0.59	69.56 ± 2.55	61.93 ± 0.59	71.39 ± 2.5
12 CC+logE+△	63.93 ± 0.59	75.00 ± 2.4	64.94± 0.59	76.36 ± 2.35

Cuadro 5.1: Tasa de clasificación [%] para los métodos de parametrización MFCC y NMF_CC, basados en la clasificación de segmentos y en sílabas.

Como se puede observar en el Cuadro 5.1, las tasas de clasificación mediante la parametrización NMF_CC mejoran con respecto a las obtenidas mediante la parametrización MFCC, en todos los casos el uso de los parámetros delta (△) o primera derivada aumentan las tasas de clasificación de manera significativa, también se puede observar que la clasificación basada en sílabas es mucho mejor que la basada en segmentos, esto debido a que las sílabas representan mejor las estructuras fonéticas de las vocalizaciones de las aves. Cabe mencionar también que, los resultados experimentales haciendo uso de la segunda derivada (△△), no supone una mejora con respecto a las tasas de clasificación obtenidas con sólo la primera derivada (△), razón por la cual no se consideran en este cuadro.

6 Conclusiones y trabajos futuros

En este capítulo se repasa todo lo que se ha observado durante el desarrollo de este trabajo de tesis, y sobre todo se extraen conclusiones a partir de los resultados experimentales que se han ido mostrando a lo largo del capítulo anterior. También se mostrarán algunas líneas de investigación que podrían servir para trabajar sobre ellas en un futuro.

6.1. Conclusiones

- Se realizó el diseño del banco de filtros auditivo basado en la técnica de la Factorización de Matrices No-Negativas (NMF), con el cuál se obtuvo una mayor concentración de filtros en el rango de frecuencias medias, justamente donde la energía espectral de las vocalizaciones de las aves se hace presente, debido a que con este diseño se calcula el banco de filtros a partir del comportamiento y naturaleza de los datos, es decir, las vocalizaciones de las aves. Por lo tanto se corrobora que un banco de filtros hallado mediante la técnica NMF resulta ser mucho más adecuado para el tratamiento de señales bioacústicas de las aves, que el hallado mediante MFCC, ya que este último presenta una mayor concentración de filtros en las bajas frecuencias y fue diseñado para el tratamiento de señales de la voz humana.
- Se efectuó la comparación del método de parametrización propuesto (NMF_CC) contra el método de parametrización convencional (MFCC), por medio de las tasas de clasificación obtenidas mediante el uso del clasificador SVM, lográndose alcanzar mayores tasas de clasificación mediante el uso de la parametrización NMF_CC. Esto quiere decir que la técnica propuesta ha logrado superar a la técnica que habitualmente se utiliza para la extracción de características de las señales de audio, MFCC.
- Considerar duraciones de trama cortas como 10 ms, proporciona buenos resultados cuando se clasifica en base a segmentos, mientras que cuando se clasifica en base a sílabas las duraciones de tramas mucho más grandes del orden de los 20 ms resultan ser óptimas ya que se estaría considerando por trama a una sílaba completa, que contiene información acústica del ave.
- De las experimentaciones presentadas se pudo observar que incluir la primera derivada a los 12 coeficientes más la log-energía, resultó en un incremento sobre las tasas de clasificación. Esto quiere decir que, aparte de considerar

la representación del tracto vocal mediante los 12 coeficientes, considerar la información temporal de dichos coeficientes junto a la log-energía, influye de manera positiva al momento de extraer las características de la señal y por ende mejora las tasas de clasificación.

- De las experimentaciones se puede observar que la clasificación basada en sílabas presenta tasas de reconocimiento mayores a las obtenidas basadas en segmentos. Esto comprueba que las sílabas representan expresiones fonéticas de cada especie de ave lo que proporciona una mejor caracterización individual por especie, explicándose así la razón por la cual clasificar en base a sílabas produce mejores tasas de clasificación.

6.2. Líneas futuras de investigación

La clasificación de audio a partir de vocalizaciones de aves, es un campo de investigación en el que aún queda mucho por hacer, especialmente en ambientes reales con presencia de mucho ruido. Es una aplicación de creciente importancia debido a la necesidad de poder proteger a las especies de aves y aportar en la investigación sobre éstas.

Se señalan a continuación posibles investigaciones que se podrían llevar a cabo, que guardan cierta relación con este proyecto:

- Trabajar con archivos de audio que no contengan la técnica de compresión de la señal de audio, ya que trabajar con formatos de audio sin compresión podría suponer mejoras en los resultados finales de clasificación. Lo anterior podría ser resuelto con la obtención de las grabaciones de audio en un ambiente de laboratorio, asimismo también, mediante la grabación de los audios de las aves en su hábitat mismo, de manera personal.
- Considerar una base de datos mucho más extensa y lo más limpia posible, que permita poder realizar mejores entrenamientos de los modelos con los que se vaya a clasificar a las especies de aves.

Bibliografía

[Alison Stattersfield, 2008] Alison Stattersfield, Leon Bennun, M. J. (2008). El estado de conservación de las aves del mundo: Indicadores en tiempos de cambio. BirdLife International.

[Anderson et al., 1996] Anderson, S. E., Dave, A. S., and Margoliash, D. (1996). Template-based automatic recognition of birdsong syllables from continuous recordings. *The Journal of the Acoustical Society of America*, 100(2):1209–1219.

[Beddard, 1898] Beddard, F. E. (1898). *The structure and classification of birds*. Longmans, Green, and Company.

[BirdLife, 2015] BirdLife, I. (2015). Bird species distribution maps of the world. *BirdLife International, Cambridge, UK and NatureServe, Arlington, USA*.

[Bosso et al., 2009] Bosso, A., Carman, R., Claver, J., Ferrari, C., Haene, E., Leiberman, J., López, H., Montaldo, N., Nardini, C., Narosky, T., et al. (2009). Observación de aves silvestres en libertad: una actividad apasionante al alcance de todos (curso de iniciación).

[Botero et al., 2005] Botero, J., Arbeláez, D., and Lentijo, G. (2005). Métodos para estudiar las aves. *Cenicafe*, 8.

[Burges, 1998] Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167.

[Catchpole and Slater, 1995] Catchpole, C. and Slater, P. (1995). Bird song: Biological themes and variations. *Cambridge University, Cambridge*.

[Cheng et al., 2010] Cheng, J., Sun, Y., and Ji, L. (2010). A call-independent and automatic acoustic system for the individual recognition of animals: A novel model using four passerines. *Pattern Recognition*, 43(11):3846–3852.

[Debnath et al., 2016] Debnath, S., Roy, P. P., Ali, A. A., and Amin, M. A. (2016). Identification of bird species from their singing. In *Informatics, Electronics and Vision (ICIEV), 2016 5th International Conference on*, pages 182–186. IEEE.

[Dietterich and Bakiri, 1995] Dietterich, T. G. and Bakiri, G. (1995). Solving multiclass learning problems via error-correcting output codes. *Journal of artificial intelligence research*, 2:263–286.

[Fagerlund, 2004] Fagerlund, S. (2004). *Automatic recognition of bird species by their sounds*. PhD thesis, Helsinki University of technology.

[Fagerlund and Laine, 2014] Fagerlund, S. and Laine, U. K. (2014). New parametric representations of bird sounds for automatic classification. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 8247–8251. IEEE.

[Fletcher, 1992] Fletcher, N. H. (1992). *Acoustic systems in biology*. Oxford University Press.

[Fletcher and Tarnopolsky, 1999] Fletcher, N. H. and Tarnopolsky, A. (1999). Acoustics of the avian vocal tract. *The Journal of the Acoustical Society of America*, 105(1):35–49.

[Foundation, 2015] Foundation, X.-C. (2015). Compartiendo cantos de aves de todo el mundo. *http://www.xeno-canto.org/*.

[Fox et al., 2008] Fox, E. J., Roberts, J. D., and Bennamoun, M. (2008). Call-independent individual identification in birds. *Bioacoustics*, 18(1):51–67.

[FRANCO et al., 2009] FRANCO, A., DEVENISH, C., BARRERO, M., and ROMERO, M. (2009). Colombia: Areas importantes para la conservación de las aves américa.

[Ganchev, 2005] Ganchev, T. D. (2005). *Speaker recognition*. PhD thesis, University of Patras.

[Gandolfi, 2004] Gandolfi, M. C. (2004). Biodiversidad y biotecnología: reflexiones en bioética.

[Gaunt et al., 1982] Gaunt, A. S., Gaunt, S. L., and Casey, R. M. (1982). Syringeal mechanics reassessed: evidence from streptopelia. *The Auk*, pages 474–494.

[Gaunt et al., 1987] Gaunt, A. S., Gaunt, S. L., Prange, H. D., and Wasser, J. S. (1987). The effects of tracheal coiling on the vocalizations of cranes (aves; gruidae). *Journal of Comparative Physiology A*, 161(1):43–58.

[Goller and Larsen, 2002] Goller, F. and Larsen, O. (2002). New perspectives on mechanisms of sound generation in songbirds. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 188(11):841–850.

[Goller and Larsen, 1997a] Goller, F. and Larsen, O. N. (1997a). In situ biomechanics of the syrinx and sound generation in pigeons. *Journal of Experimental Biology*, 200(16):2165–2176.

[Goller and Larsen, 1997b] Goller, F. and Larsen, O. N. (1997b). A new mechanism of sound generation in songbirds. *Proceedings of the National Academy of Sciences*, 94(26):14787–14791.

[Gunn et al., 1998] Gunn, S. R. et al. (1998). Support vector machines for classification and regression. *ISIS technical report*, 14:85–86.

[Gupta et al., 2013] Gupta, S., Jaafar, J., Ahmad, W. F. W., and Bansal, A. (2013). Feature extraction using mfcc. *Signal & Image Processing*, 4(4):101.

[Hoese et al., 2000] Hoese, W. J., Podos, J., Boetticher, N. C., and Nowicki, S. (2000). Vocal tract function in birdsong production: experimental manipulation of beak movements. *Journal of Experimental Biology*, 203(12):1845–1855.

[Ittichaichareon et al., 2012] Ittichaichareon, C., Suksri, S., and Yingthawornsuk, T. (2012). Speech recognition using mfcc. In *International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012) July*, pages 28–29.

[Karush, 1939] Karush, W. (1939). Minima of functions of several variables with inequalities as side conditions. *Master thesis, University of Chicago*.

[Kaufman, 2001] Kaufman, K. (2001). Lives of north american birds. *Recuperado el 23/09/2017, de http://www.audubon.org/bird-guide*.

[King, 1989] King, A. (1989). Functional anatomy of the syrinx. *Form and function in birds*, 4:105–192.

[Koenig, 1949] Koenig, W. (1949). A new frequency scale for acoustic measurements. *Bell Telephone Laboratory Record*, 27:299–301.

[Kogan and Margoliash, 1998] Kogan, J. A. and Margoliash, D. (1998). Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden markov models: A comparative study. *The Journal of the Acoustical Society of America*, 103(4):2185–2196.

[Kuhn and Tucker, 1951] Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492, Berkeley, Calif. University of California Press.

[Lafuente, 2011] Lafuente, G. A. (2011). I jornadas cientificas sobre patologia, biologia y manejo en animales exoticos y salvajes. *Revista Complutense de Ciencias Veterinarias*, 5(2):154–184.

[Lee et al., 2008] Lee, C.-H., Han, C.-C., and Chuang, C.-C. (2008). Automatic classification of bird species from their sounds using two-dimensional cepstral coefficients. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(8):1541–1550.

[Lee and Seung, 1999] Lee, D. D. and Seung, H. S. (1999). Algorithms for non-negative matrix factorization. *Nature*, 401(6755):788–791.

[Lee and Seung, 2001] Lee, D. D. and Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pages 556–562.

[Li and Wu, 2015] Li, Y. and Wu, Z. (2015). Animal sound recognition based on double feature of spectrogram in real environment. In *Wireless Communications & Signal Processing (WCSP), 2015 International Conference on*, pages 1–5. IEEE.

[Ludeña-Choez and Gallardo-Antolín, 2012] Ludeña-Choez, J. and Gallardo-Antolín, A. (2012). Speech denoising using non-negative matrix factorization

with kullback-leibler divergence and sparseness constraints. In *Advances in Speech and Language Technologies for Iberian Languages*, pages 207–216. Springer.

[Ludeña-Choez and Gallardo-Antolín, 2015] Ludeña-Choez, J. and Gallardo-Antolín, A. (2015). Feature extraction based on the high-pass filtering of audio signals for acoustic event classification. *Computer Speech & Language*, 30(1):32–42.

[Marsh and Trenham, 2008] Marsh, D. M. and Trenham, P. C. (2008). Current trends in plant and animal population monitoring. *Conservation Biology*, 22(3):647–655.

[Mayhua López et al., 2013] Mayhua López, E. T. et al. (2013). Elementos locales en conjuntos de clasificadores diseñados por"boosting".

[McLELLAND, 1989] McLELLAND, J. (1989). Larynx and trachea. *Form and function in birds*, 4:69–103.

[Müller, 1878] Müller, J. P. (1878). *On certain variations in the vocal organs of the Passeres that have hitherto escaped notice*. Clarendon Press.

[Navarro-Sigüenza et al., 2014] Navarro-Sigüenza, A. G., Rebón-Gallardo, M. F., Gordillo-Martínez, A., Peterson, A. T., Berlanga-García, H., and Sánchez-González, L. A. (2014). Biodiversidad de aves en méxico. *Revista mexicana de biodiversidad*, 85:476–495.

[Núñez Martínez, 2005] Núñez Martínez, J. (2005). Estudio de nuevos algoritmos de descomposicion lineal de observaciones en componentes. *Universidad de Sevilla*.

[Olvera, 2014] Olvera, E. T. V. (2014). Identificación del canto de turdus migratorius (aves) utilizando un modelo acústico estadístico. Master's thesis, Universidad nacional autónoma de México.

[Paatero and Tapper, 1994] Paatero, P. and Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2):111–126.

[Pacual García, 2012] Pacual García, J. (2012). Adaptación al ruido urbano: Estrategias comunicativas de una comunidad de aves en los alrededores del aeropuerto de barajas.

[Patterson and Pepperberg, 1994] Patterson, D. K. and Pepperberg, I. M. (1994). A comparative study of human and parrot phonation: Acoustic and articulatory correlates of vowels. *The Journal of the Acoustical Society of America*, 96(2):634–648.

[Pedro F. Develey, 2009] Pedro F. Develey, J. M. G. (2009). Important bird areas americas : Brazil. BirdLife INTERNATIONAL.

[Plenge, 2010] Plenge, M. A. (2010). List of birds of peru. sernanp. perú.

[Ralph et al., 1995] Ralph, C. J., Droege, S., and Sauer, J. R. (1995). Managing and monitoring birds using point counts: standards and applications.

[Ralph et al., 1996] Ralph, C. J., Geupel, G. R., Pyle, P., Martin, T. E., DeSante, D. F., and Milá, B. (1996). Manual de métodos de campo para el monitoreo de aves terrestres.

[Ridgely and Guy, 1989] Ridgely, R. S. and Guy, T. (1989). *The birds of South America: Volume 1: the oscine passerines*, volume 1. University of Texas Press.

[Ridgley and Tudor, 1994] Ridgley, R. and Tudor, G. (1994). The birds of south america. the suboscine passerines vol. 2. *University of Texas, Austin, Texas*.

[Schölkopf et al., 1999] Schölkopf, B., Burges, C. J., and Smola, A. J. (1999). *Advances in kernel methods: support vector learning*.

[Schölkopf and Smola, 2002] Schölkopf, B. and Smola, A. J. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

[Schuller et al., 2010] Schuller, B., Weninger, F., Wöllmer, M., Sun, Y., and Rigoll, G. (2010). Non-negative matrix factorization as noise-robust feature extractor for speech recognition. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 4562–4565. IEEE.

[SERNANP, 2012] SERNANP (2012). Perú: País megadiverso. *Recuperado el 15/09/2017, de http://www.sernanp.gob.pe/documents/10181/88081/Peru-+Pais+Megadiverso.pdf*.

[Somervuo et al., 2006] Somervuo, P., Harma, A., and Fagerlund, S. (2006). Parametric representations of bird sounds for automatic species recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(6):2252–2263.

[Stattner et al., 2013] Stattner, E., Segretier, W., Collard, M., Hunel, P., and Vidot, N. (2013). Song-based classification techniques for endangered bird conservation. *arXiv preprint arXiv:1306.5349*.

[Vapnik and Vapnik, 1998] Vapnik, V. N. and Vapnik, V. (1998). *Statistical learning theory*, volume 1. Wiley New York.

[Westneat et al., 1993] Westneat, M. W., Long, J., Hoese, W., and Nowicki, S. (1993). Kinematics of birdsong: functional correlation of cranial movements and acoustic features in sparrows. *Journal of Experimental Biology*, 182(1):147–171.

Nomenclatura

ASR	Reconocimiento Automático de Voz
CC	Coeficientes Cepstrales
DCT	Transformada del Coseno Discreto
DFT	Transformada Discreta de Fourier
DTW	Distorsión Dinámica Temporal
ECOC	Códigos de Corrección de Errores
ETR	Arbol de Regresión Extra
FFT	Transformada Rápida de Fourier
GMM	Modelos de Mezclas Gaussianas
HMM	Modelos Ocultos de Markov
IBA	Areas Importantes de Aves y Biodiversidad
k-NN	k- Vecinos más Cercanos
KKT	Karush-Kuhn-Tucker
LBPV	Varianza de Patrón Binario Local
LL	Labios Laterales
LPC	Codificación de coeficientes Predictiva Lineal
LPCC	Coeficientes Cepstrales de Predicción Lineal
MFCC	Coeficientes Cepstrales a escala de Frecuencia Mel
ML	Labios Mediales
MTM	Membranas Timpaniformes Mediales
NB	Técnicas Bayesianas

NMF	Factorización de Matrices No negativas
NMF CC	Coeficientes Cepstrales basados en NMF
PPF	Frecuencia de Par de Permutación
QP	Programación Cuadrática
RBF	Función de Base Radial
RDT	Arbol de Desición Aleatoria
RF	Bosques Aleatorios
ROC	Característica Operativa del Receptor
SBV	Vectores Base Espectrales
SNR	Relación Señal a Ruido
STFT	Transformada de Fourier de Tiempo Corto
SV	Vectores Soporte
SVM	Máquina de Vectores de Soporte
ULBP	Patrón Binario Local Uniforme