



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES ZARAGOZA

IDENTIFICACIÓN DEL CANTO DE *Turdus migratorius* (Aves) UTILIZANDO

UN MODELO ACÚSTICO ESTADÍSTICO

T E S I S

QUE PARA OBTENER EL TÍTULO DE

B I Ó L O G O

P R E S E N T A

VILLARREAL OLVERA ESAÚ TOAKI



DIRECTOR DE TESIS: DR. MEZA RUIZ IVAN VLADIMIR

ASESOR INTERNO: DR. BUENO HERNÁNDEZ ALFREDO

MÉXICO D.F., MAYO 2014



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

En México solo el 0.7 de cada mil personas son científicas. Aún no hay registro de la cantidad de biólogos dedicados al monitoreo y conservación de la biodiversidad, si se expresara numéricamente tal cantidad, ésta sería cercana a cero y más aún cuando la dividiéramos por el cien por ciento de los problemas ambientales en los que estamos envueltos. A veces la condición humana nos ciega:

“Nuestro progreso, en parte dependerá de la rapidez con la que dejemos de contar con piedras y palos, mientras sujetamos en nuestra mano el teléfono móvil más avanzado”

O bien, diría Einstein:

“No podemos resolver problemas pensando de la misma manera que cuando los creamos”

Agradecimientos

Agradezco a mi tan querida Universidad Nacional Autónoma de México a la cual orgullosamente pertenezco. Mi segunda casa, quien desde mi bachillerato en CCH-Sur me cobijó con conocimientos y valores humanísticos. A la FES-Zaragoza por brindarme los elementos para conformar un criterio objetivo en el ámbito científico, social y cultural. Mi compromiso con esta mi casa de estudios es ejercer digna y éticamente mi profesión a beneficio del pueblo de México y contribuyendo humildemente al conocimiento de la humanidad.

Antes que nada pido una disculpa por aquellas personas que me lleguen a faltar, sin embargo les garantizo que en mi persona siempre les estaré agradecido por contribuir en mi formación.

A mi amigo y director de tesis, Dr. Iván Vladimir Meza Ruiz por bríndame tu confianza, amistad y el conocimiento necesario para lograr este trabajo. La experiencia en esta gran colaboración es invaluable. Agradezco tu paciencia y la dedicación brindada a este proyecto.

A mi amigo y asesor interno, Dr. Alfredo Bueno Hernández por la amistad, confianza, conocimiento y respaldo brindado incondicionalmente en cada una de mis inquietudes para poder lograr éste y otros trabajos relacionados a la ornitología.

A mis revisores, Dr. David Nahúm Espinoza Organista, M en C Patricia Rivera García y M en C Nicté Ramírez Priego, por sus comentarios para mejorar la tesis.

A mis asesores. Dr. Caleb Rascón, M en I. Hernando Ortega Carrillo, Dr. Gibran Fuentes Pineda y a todos mis amigos y compañeros de Grupo Golem IIMAS por su apoyo, enseñanza y amistad.

A mis amigos Biol. Alejandro Gordillo, M. en C. Marco Ortiz, Noé Pacheco por sus observaciones y amistad. A Omar Paramo por tan bello artículo de divulgación.

Dedicatorias

Con todo mi amor, cariño y respeto a mi padre y amigo José Villarreal, con quien hombro a hombro he construido esta nuestra gran primer obra. Te amo y agradezco infinitamente tu paciencia, dedicación, enseñanza, apoyo y amor. “Y sabes que... esto se hace así”.

A mi hermosa y amada madre Leticia Olvera, por darme tanto amor, apoyo y enseñanza. Te amo mamá y comparto contigo este nuestro trabajo, porque a pesar de la distancia siempre has sabido estar conmigo. “Te amo mucho mi mamá”

Con mucho amor a mi querida Samantha Villarreal, sé que estas pequeña y ojalá leas algún día esto y sepas que te amo y quiero mucho.

A mis queridos hermanos, amigos y compañeros incondicionales Noé y Daniel con quien comparto éste trabajo, los amo mucho. Gracias por su confianza y apoyo “El Ciber rifa”.

A mi compañera Laura B. Silva, por tu apoyo incondicional, me da mucho gusto compartir éste logro contigo. Te amo.

Y donde quiera que te encuentres sé que lo estarás leyendo Tío Cayo, gracias logramos que este trabajo fuera “un poquito más allá de Mickey mouse” D.E.P.

Gracias y espero tengan oportunidad de ver esto donde quieran que se encuentren, los amamos mucho y los extrañamos Q.E.P.D.

Judith Villarreal

Ricardo Olvera

Sandra Zamudio

Agradezco a mis grandes y queridos amigos. Raquel M., Tania G., Sergio C., José C., Hernando O., Anabel S., Omar P., Alonso, Francisco, Karla Biroli.

Y disculpa a todos aquellos que no los alcanzo a nombrar, gracias ¡!

ÍNDICE

a) Índice de figuras.....	i
b) Índice de tablas	ii
c) Índice de diagramas	ii
d) Índice de gráficas	ii
e) Resumen.....	iii
1) Introducción.....	1
2) Antecedentes.....	3
2.1. Las Aves	
2.1.1 Diversidad de las Aves	
2.1.2 Diversidad de la Aves en México	
2.1.3 Aves amenazadas	
2.1.4 Monitoreo en Aves	
2.2. La Bioacústica	
2.2.1 La Bioacústica en las Aves	
2.2.2 Características del canto de las Aves	
2.2.3 Características acústicas en el canto de las Aves	
2.2.4 La Bioacústica en el monitoreo de Aves	
3) Planteamiento del Problema	21
3.1 Objetivo General	
3.2 Objetivos Particulares	
3.3 Justificación	
3.4 Hipótesis	
4) Método	23
4.1 Base de Audio-datos	
4.1.1 Desarrollo de Corpus	
4.2 Procesamiento Digital de la Señal	
4.2.1 Ventaneo	
4.2.2 Transformada Discreta de Fourier (DFT)	

4.3	Extracción de MFCCs	
4.3.1	Filtros Mel	
4.3.2	Transformada Discreta de Coseno (TDC)	
4.3.3	Calculo de los Delta MFCC (MFCC)	
4.4	Estadísticas: descripción de la señal	
4.5	Entrenamiento: Support Vector Machine (SVM)	
4.6	El sistema de Identificación Acústica	
4.6.1	Funcionamiento del Sistema	
5)	Planeación y Diseño de MIAA	42
5.1	Diseño Modulo de Identificación Acústica de Aves (MIAA)	
5.1.1	Características de los Componentes	
5.2	Implementación de MIAA	
6)	Diseño Experimental	56
6.1.	Evaluación	
6.1.1	Frame-Based	
6.1.2	Turn-Based	
6.1.3	Indicadores de Evaluación	
6.2	Experimentos	
6.2.1	Evaluación del Silabificador	
6.2.2	Evaluación del Clasificador Gold standard vs Gold standard	
6.2.3	Evaluación del Clasificador-Gold standard -ruido vs Gold standard	
6.2.4	Evaluación del clasificador Silabificador vs Gold standard	
6.2.5	Evaluación clasificador-Silabificador-Ruido-otras especies vs Gold standard	
6.2.6	Evaluación clasificador-Silabificador-Ruido-otras especies vs Silabificador	
6.2.7	Curva ROC	
6.2.8	Curva de aprendizaje	
7	Resultados y Análisis de Resultados	69
8	Discusión y Conclusiones	79
9	Literatura Citada	85

Índice de Figuras

1. Número de especies de aves por orden, clasificación de Sibley & Monroe	6
2. Relación masa-frecuencia de distintos grupos de animales que se comunican por sonido	11
3. Nomenclatura del canto de aves	13
4. Estructura de la siringe y músculos que la rodean	14
5. Movimiento de membrana timpaniforme	15
6. Sonido periódico con armónicos	16
7. Intensidad y distancia de la vocalización de un ave	17
8. Métricas utilizadas para caracterizar la vocalización de un ave	18
9. Etiquetación de la sílaba de <i>Turdus migratorius</i>	26
10. Representación dinámica, matemática-computacional y espectral de una señal de audio	27
11. Ventaneo de 400 valores muestra y el traslape de una de ellas con la ventana anterior	28
12. Ventana de Hamming para una muestra	29
13. DFT aplicada a una ventana	31
14. Peridiograma	31
15. Filtro Mel de 40 valores	34
16. Paso de una ventana de la DFT por un filtro Mel	35
17. Peridiograma de la unión de valores Mel en un segundo de muestra	35
18. Coeficiente logarítmico cepstral Mel	36
19. Funciones aplicadas a cada ventana	38
20. Máquina de soporte vectorial lineal	39
21. Proceso del silabificador	40
22. Funcionamiento del sistema de identificación	41
23. Diseño y composición de MIA	43
24. Ascenso y descenso del eje principal de MIAA	50
25. Posicionamiento de micrófonos de MIAA	51
26. Distribución de peso en el eje del sistema MIAA	51
27. Movimiento de micrófonos con servomotor de MIAA	53
28. Colocación de micrófono a 45° de MIAA	54
29. Cross-validation	56
30. Ejemplo Frame-base	57
31. Ejemplo Turn-base (middle point)	58
32. Representación de las curvas ROC con distintos valores guía	67
33. Módulo de identificación Acústica de Aves (MIAA)	83

Índice de Tablas

1. Especies consideradas y tiempo total de grabación.....	25
2. Descripción de recursos de audio.....	69
3. Resultados de evaluación en la etiquetación Gold standard vs silabificador.....	70
4. Resultados de evaluación del clasificador Gold standard vs Gold standard.....	71
5. Resultados de evaluación en la etiquetación silabificador vs Gold standard.....	72
6. Resultados de evaluación del clasificador silabificador-ruido-otras aves vs silabificador.....	73
7. Resultados de evaluación del clasificador silabificador-ruido-otras aves vs Gold standard.....	74
8. Resultados de evaluación del clasificador Gold standard-ruido vs Gold standard.....	75

Índice de Diagramas

1. Funcionamiento básico de MIAA.....	55
2. Comparación de etiquetación Gold standard vs etiquetas silabificador.....	61
3. Entrenamiento de modelo con silabas de Gold standard y comparación de modelo Gold standard vs Gold standard.....	62
4. Entrenamiento de modelo con silabas de Gold standard y ruido, comparación de modelo Gold standard vs Gold standard.....	63
5. Entrenamiento de modelo con silabas de silabificador, comparación de modelo Gold standard vs Gold standard.....	63
6. Entrenamiento de modelo con silabas de Gold standard y ruido, comparación de modelo Gold standard vs Gold standard.....	64
7. Entrenamiento de modelo con silabas de silabificador, ruido y otras aves, comparación de modelo vs Silabificador.....	65
8. Entrenamiento de modelo con silabas de silabificador, ruido y otras aves, comparación de modelo vs Silabificador.....	63

Índice de Gráficas

1. Curva ROC de los 3 modelos propuestos.....	77
2. Curva de acumulación.....	77

Resumen

Distintos animales utilizan diferentes frecuencias para comunicarse por sonido, esto se debe principalmente a que la evolución de los animales se ha adaptado a frecuencias particulares para dar mayor comunicación conespecífica. Por lo tanto, varios investigadores estiman el área de acción y censo de aves a partir de las vocalizaciones. Esto ha representado una ventaja sobre métodos convencionales para el monitoreo de aves. Además esto es una alternativa de seguimiento ante las problemáticas en la reducción de poblaciones silvestres. Existe la necesidad de establecer nuevas estrategias y técnicas cada vez más avanzadas, tanto en la instrumentación como en el procedimiento, con el objetivo de brindar mayores herramientas, que faciliten el estudio y monitoreo de las poblaciones silvestres. Debido a la importancia de las aves como especies indicadoras de la salud ambiental y la conservación de las mismas, existe una considerable cantidad de trabajos publicados sobre la detección automática de vocalizaciones de aves con más de una década de investigación.

En este trabajo se presenta un sistema automatizado para la identificación de la especie *Turdus migratorius* a partir de las sílabas del canto, que fueron producto de una la realización de una base de datos acústica con una duración de 59 minutos. El núcleo del sistema se basó en el pre-procesamiento digital de las señales, extracción de características de la señal de audio con los coeficientes cepstrales de los filtros Mel (MFCCs). Las MFCCs son una representación más rica en propiedades acústicas, que los análisis basados en rangos de frecuencias tradicionalmente utilizados. Estas propiedades acústicas se resumen con nueve funciones estadísticas que son utilizadas para entrenar un modelo de predicción con la herramienta de aprendizaje automático *support vector machines* (SVM).

El sistema está compuesto de un micrófono para la grabación, un silabificador, un extractor de MFCCs y un clasificador. El micrófono realiza la captura de audio que alimenta el silabificador, el cual se encarga de identificar segmentos de actividad acústica. El extractor de características transforma de estos segmentos a su representación estadística de las MFCCs, las cuales alimentan al clasificador basado en SVM quien realiza una predicción sobre el tipo de segmento. El sistema en un ambiente funciona de tal manera que cuando canta *Turdus migratorius* u otra ave, el canto es capturado por el micrófono y de los segmentos de actividad se identifican las sílabas y estas son procesadas por el extractor de características para obtener las propiedades acústicas-estadísticas de la sílaba, que por último dichas propiedades entran al clasificador el cual identifica si la actividad pertenece a *Turdus migratorius* o no.

Los resultados de la evaluación del sistema arrojaron que se recupera el 78% de las vocalizaciones, de las cuales el 83% son correctamente identificadas. Se encontró que el sistema puede identificar de manera adecuada con solo 30 minutos de entrenamiento para generar un modelo en SVM. Este sistema resultante forma parte de nuestro dispositivo robótico MIAA (Módulo de Identificación Acústica de Aves), cuyo objetivo es monitorear las aves en su hábitat, y en el futuro estimar sus poblaciones.

1

Introducción

El concepto monitoreo ha sido definido por varios autores con acepciones similares, siendo éste “un seguimiento intermitente (regular o irregular) llevado a cabo para determinar el grado de cambio en algún parámetro poblacional de acuerdo con un nivel estándar predeterminado, o determinar el grado de desviación de una norma predefinida” (Hellawell, 1991; Golsmith, 1991; Furness et al., 1993; Villaseñor, 2006). Los estudios de fauna silvestre se han realizado basados en estimaciones de métodos estadísticos, los cuales tienen una cierta variabilidad entre sí, pero que en general permiten hacer inferencias sobre las tendencias de las poblaciones (Davies & Winstead, 1987). Un parámetro muy utilizado es considerar la variabilidad en número de individuos de una población, el cual indica retrospectivamente la existencia de un cambio y salud en las poblaciones (Lack, 1954; Hutchinson, 1978; Ralph et al., 1993). Estos métodos y parámetros se han vuelto cada vez más importantes, ya que, durante la segunda mitad del siglo XX, la perturbación de ecosistemas por actividades antropogénicas ha ido en incremento, y como consecuencia se ha generado la necesidad de establecer nuevas estrategias y técnicas cada vez más avanzadas, tanto en la instrumentación tecnológica como en los procedimientos (Tirira, 1998), con el objetivo de brindar mayores herramientas, que faciliten el estudio y monitoreo de las poblaciones silvestres.

Uno de los puntos básicos y principales en el estudio de fauna silvestre es la identificación de las especies, la cual es realizada por especialistas de los distintos grupos de animales y forma parte de un proceso donde predomina la experiencia del investigador, para dar cierta confiabilidad y respaldo en la identificación. Esto mismo representa una desventaja ya que la experiencia de un investigador no se puede cuantificar y existe un gran problemática en la dispersión del conocimiento, así como el tiempo de adquisición de éste. Además, de que en muchas ocasiones el investigador especialista consume una gran cantidad de tiempo dedicado a la identificación de las especies, lo que afecta directamente el desempeño total de una investigación. Por tal motivo en este trabajo se aborda dicha problemática y se realiza una propuesta utilizando las herramientas computacionales del aprendizaje automático (en inglés, *machine learning*) para ofrecer una alternativa en la identificación de aves a partir del canto, donde específicamente se trabaja con el canto de la especie *Turdus migratorius* y se realiza una propuesta tanto en el procedimiento, como en instrumentación al presentar el dispositivo robótico “Módulo de Identificación Acústica de Aves (MIAA)”, el cual presenta una alternativa que cierra la brecha entre la investigación del monitoreo para la conservación de fauna silvestre y la tecnología, como apoyo a las labores que realizan los investigadores en campo, para el conocimiento y conservación de la fauna silvestre.

*Nuestro objetivo es preservar el pasado
usando la tecnología actual.*
Henry Ford (1863-1947)

2

Antecedentes

En esta sección se plantea el marco teórico y conceptual, el cual comprende desde la gran diversidad de aves, la amenaza a la que estas se enfrentan y como a través del tiempo se han desarrollado distintas metodologías para su monitoreo, estudio y conservación. Una de las metodologías concernientes a la bioacústica es la identificación de aves por el canto. Esta última se fundamenta en las características acústicas que presentan distintos grupos las cuales sirven como herramienta para su uso en el monitoreo. Estas metodologías, con el progreso de la tecnología han abierto un panorama más amplio para la implementación de mejores herramientas y técnicas.

2.1 Aves

La Clase Aves constituye 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 ponen huevos de cáscara dura a diferencia de los reptiles, son endotermos (regulan de forma activa su temperatura corporal). A diferencia de los mamíferos, las aves mantienen una temperatura corporal superior que es habitualmente en torno a los 40°C. Su corazón posee cuatro cavidades cardíacas, y el sistema pulmonar es extremadamente complejo, con pulmones relativamente pequeños y rígidos que están conectados a grandes cámaras corporales de dilatación conocidas como “sacos aéreos” y que en conjunto con los huesos neumáticos (llenos de aire) hacen que las aves sean muy ligeras en peso, lo que supone que una gran cantidad de aves puedan volar.

Otra característica de las aves es su pico de origen córneo y el cual también es llamado “ranfoteca”. Presentan una serie de estructuras morfológicas características de la clase, por ejemplo; su cintura escapular, la pelvis y la presencia de una fúrcula (que es la unión de sus clavículas), entre otras características. También presentan plumas, las cuales están formadas por un material córneo denominado queratina (Christiansen, 2009; Gill, 2006; Perrins, 2011).

2.1.1 Diversidad de Aves

Actualmente el número de especies de aves es muy variable y esto depende del concepto de especie utilizado. Un ejemplo de esto son los sistemas taxonómicos tradicionales que utilizan el concepto de especie biológica enfocada a la morfología (Navarro & Sánchez, 2003), los cuales reconocen alrededor de 8,200 especies (Mayr, 1942), mientras que sistemas de clasificación como el de Sibley & Monroe (1990) reconoce 9,672 (Navarro & Sánchez, 2003), siendo ésta última una de las clasificaciones más utilizadas (Colín, 2008). Sin embargo hasta el momento varios investigadores reconocen entre 9,720 y 9,845 especies de aves (Burnie, 2003; Perrins, 2011) las cuales se encuentran organizadas en 2 superórdenes, 28 órdenes, 172 familias y 2,121 géneros. El orden Passeriformes (comúnmente llamados “pájaros”) es el más diverso ya que contiene 5,900 del total de 9,845 (o, según Sibley & Monroe, 5 712 de 9672) aves del mundo, lo que corresponde aproximadamente al 60% de aves (Perrins, 2011), este orden es también uno de los más recientes evolutivamente.

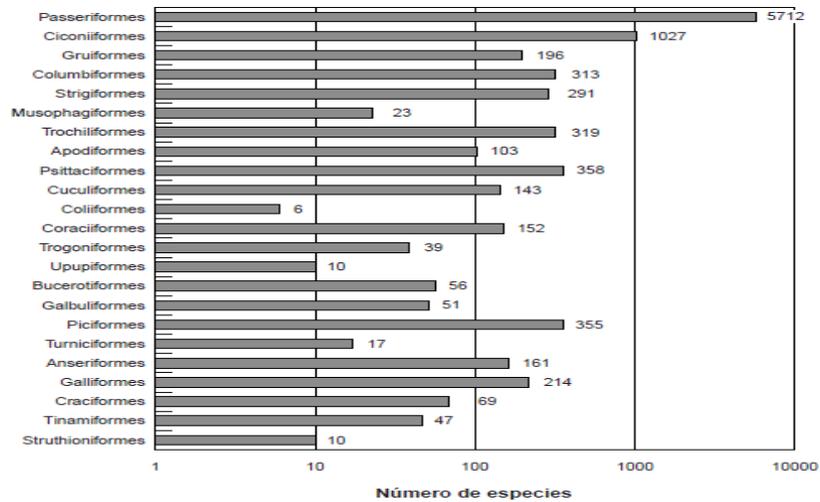


Figura 1. Número de especies de aves por orden, clasificación de Sibley & Monroe (Tomado de Navarro & Sánchez, 2003)

Se sugiere que los precursores de estas aves pueden haber existido en el Cretáceo tardío, hace unos 71 millones de años (Colín, 2008). Cuando tuvo lugar la primer gran división del orden Paseriformes; en Oscinos y Subocinos, con unas 1,151 y 4,561 especies vivas, respectivamente (Howard & Moore, 1991). No existen nombres comunes adecuados para referirse a estos dos subgrupos; los Oscinos se conocen a veces como “aves canoras” y también “aves de percha”, lo que solo sugiere algunas de sus actividades exclusivas del grupo carentes en Subocinos. La clasificación de Sibley & Monroe divide el Orden Passeriformes en dos subórdenes: Tyranni y Passeri, los cuales son equivalentes a Oscinos (Passeri) y Subocinos (Tyranni). El suborden Tyranni (Subocinos) tiene su origen casi exclusivamente en el continente americano, con la mayoría de especies confinadas en el centro y sur del continente. Por otra parte, el suborden Passeri (Oscines), aunque originarios de Europa y África (según Sibley y Alquist) ha extendido al continente americano (Perrins, 2011).

2.1.2 Diversidad de Aves en México

México ocupa entre el décimo y doceavo lugar a nivel mundial en riqueza de aves (Gonzales & Gómez, 2003), ya que alberga el 10% de la avifauna mundial, componiéndose ésta con alrededor de 1076 especies que se integran en 471 géneros, 87 familias y 22 ordenes (Navarro & Sánchez, 2003; Howard & Moore, 1991). Los componentes de la avifauna mexicana están distribuidos de manera similar a la avifauna mundial, siendo el orden Passeriformes el que mayor número de especies presenta. Las aves también se han organizado de acuerdo a su estacionalidad, ya que, pueden ser residentes temporales de invierno o verano (solo se encuentran en una época del año) y residentes permanentes, las cuales conforman el 70% de la avifauna (Navarro & Benítez, 1993) y a su vez están subdivididas en especies no-endémicas, cuasiendémicas y endémicas.

Sin embargo la riqueza avifaunística de México no solo radica en su gran diversidad, sino también en que un gran número de especies endémicas, es decir, exclusivas del país y éstas representan el 10% de la avifauna mexicana con 104 especies (Gonzales & Gómez, 2003). Las especies endémicas tienen un valor biológico importante, ya que la extinción de una de ellas representa una pérdida global, sin que ello determine el valor e importancia de las especies no endémicas.

2.1.3 Aves amenazadas

Existe cierta vulnerabilidad en grupos que han evolucionado aisladamente o con poblaciones muy reducidas, lo cual sugiere que un cambio drástico (extrínseco o intrínseco) los lleva fácilmente a la extinción (Iñigo, 2003). *Birdlife International* (2001) estima que a partir del año 1600 se han extinto unas 103 especies de aves en el mundo y en los últimos 200 años se han desaparecido 24 especies en México. La causa principal ha sido la reducción y fragmentación del hábitat (Ríos, 2008; Iñigo, 2003). La evaluación de *Birdlife International* en 2001 sobre el estado de conservación de aves terrestres en el mundo menciona que “una de cada diez aves está en alguna categoría de amenaza”, es decir el 12% de avifauna mundial (1,186 especies) enfrentan un grave riesgo durante los próximos 100 años (Iñigo, 2008). México ocupa el onceavo lugar a nivel mundial con un total de 40 especies amenazadas, esto según el Libro Rojo (BirdLife International 2000). Según datos de *Defenders of Wildlife*, el 24% de aves en México se encuentra en peligro de extinción, mientras que en la NORMA Oficial Mexicana NOM-059-SEMARNAT-2010 se consideran 373 especies en alguna categoría de riesgo (En peligro de extinción, Amenazada y Protección Especial) y 19 especies extintas.

2.1.4 Monitoreo en Aves

De acuerdo con la situación antes mencionada los decrementos en poblaciones de aves terrestres han hecho necesario el desarrollo de metodologías y programas específicos para el monitoreo de sus poblaciones (Ralph et al., 1996). Entre los métodos más utilizados para el monitoreo y estimación de los parámetros de las poblaciones de aves, está

la estimación de la abundancia (Ralph et al., 1995), con la captura de aves realizada con redes de niebla, siendo de las técnicas más utilizadas para la estimación de abundancia, riqueza, y demografía (Ralph et al., 1995; Rosenstock et al., 2002; Simons et al., 2007). Otra técnica muy conocida pero no tan utilizada por su alto costo económico es el uso de la telemetría, la cual ha hecho posible el seguimiento de los movimientos de las aves (Kenward 2001; McGrady et al., 2003; Jiguet & Villarubias, 2004, Cadahía et al., 2005,) y se ha convertido en una herramienta esencial y predominante en el estudio de la migración de aves (Berthold et al., 1995; Fuller et al., 1995; Hake et al., 2001), ya que permite seguir tanto las rutas migratorias, como los sitios de anidación y resguardo de éstas.

Uno de los métodos más empleados para censar aves es el conteo por puntos, el cual consiste en observar y dependiendo del investigador, oír cantos para identificar especies de aves desde puntos establecidos. Es usado por un gran número de investigadores por ser de los más eficientes y adaptables para todo tipo de terrenos y hábitats (Hamel et al., 1996). Este método permite al investigador obtener el mayor número posible de puntos de observación para cada tipo de unidad ambiental, obteniendo así un mayor número de datos independientes, que en su conjunto dan información estadística acerca del estado de la población (Ralph et al., 1996; Bibby et al., 2000).

Estos métodos permiten estudiar grandes áreas y muchos lugares en un tiempo relativamente corto, y de manera económica. Pero uno de los inconvenientes es que al tomar muestras en áreas grandes y/o remotas, se requiere de una mayor cantidad de observadores, y que el personal esté calificado para la identificación tanto visual como auditiva, además de que la calidad de los datos puede ser alterada por sesgos de cada observador

(Celis et al., 2009), ya que existen diferencias en la capacidad de los observadores para detectar, identificar, y registrar, debido a la gran variación física y acústica en las aves (Cyr 1981; Kepler & Scott, 1981; Bart, 1985; Emlen & DeJong 1992; Alldredge et al., 2007).

2.2 Bioacústica

Como se describió en el punto anterior, varios investigadores estiman el área de acción y censo a través de las vocalizaciones de aves (Flores, 1999). Sin embargo, el tema tiene todo un tras fondo científico que lleva como nombre “Bioacústica”. La Bioacústica es un campo multidisciplinario que conjuga la Biología y la Acústica, siendo está esencialmente 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 ambientes elásticos (líquidos, gaseosos, solidos).

Distintas clases de animales utilizan diferentes frecuencias (Hz) para la comunicación por sonido, esto se debe principalmente a que la evolución de los animales se ha adaptado a frecuencias particulares para dar mayor comunicación conspecífica (Fletcher, 2004). Los análisis acústicos realizados por Fletcher (2004, 2005) indican que la frecuencia óptima de comunicación es inversamente proporcional a la potencia 0.4 de la masa corporal del animal y que esto solo aplica en animales terrestres ya que la comunicación para animales acuáticos tiene sus propias particularidades, ya que la densidad del agua afecta la transmisión del sonido de manera distinta. De acuerdo con Fletcher (2004) existe una correlación frecuencia-

masa, es decir, que mientras más grande sea un animal, tendrá una frecuencia más baja (sonido grave) y un animal de talla más pequeña tendrá una frecuencia más alta (sonido agudo) como se muestra en la siguiente figura:

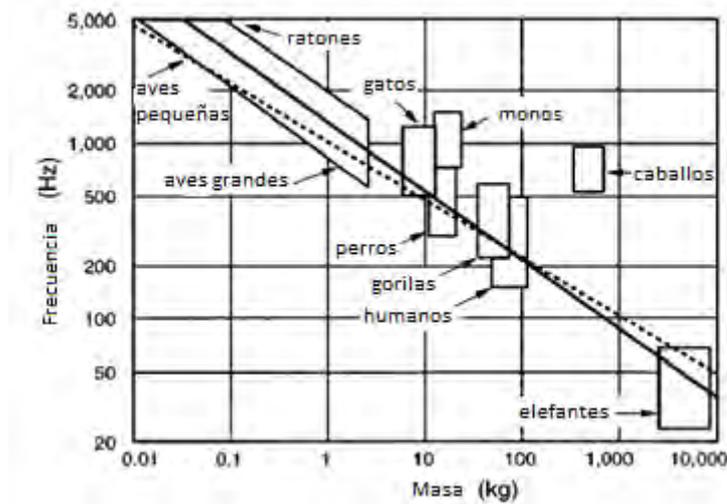


Figura 2. Relación masa-frecuencia de distintos grupos de animales que se comunican por sonido (Tomado y modificado de Fletcher, 2004).

2.2.1 La Bioacústica en las Aves

Los inicios de percepción en la variación de cantos y la asociación de éstos a distintas aves, inicia con los primeras notaciones de Barrington en 1773, quien realizó un informe para la *Royal Society* de Londres, señalando las diferencias en cantos de distintas especies de aves. En la misma época, en 1789, el naturalista inglés Gilbert White describió cómo las aves antes conocidas como reyezuelos de sauce podrían ser separadas por sus cantos en tres especies distintas. En otros trabajos como los de Craig (1943) y Marler (1952), se afirmaba la existencia de diferencias en los cantos de aves, aunque no tenían las herramientas suficientes para comprobar y cuantificar el grado de variación (Warbler, 2008). No fue sino hasta la

invención del espectrógrafo y posteriormente las primeras descripciones del canto por Thorpe (1954), que se pudo medir gráficamente el sonido y posteriormente detallar la complejidad estructural de los cantos. Esta nueva técnica también permitió hacer inferencias y relaciones entre las aves. Sibley y Ahlquist (1990) estudiaron el Orden Passeriformes (el cual comprende el 60% del total de las aves) y lo dividieron en dos grupos, Suboscines y Oscines (Howard & Moore, 1991), siendo la diferencia entre estos dos la siringe (estructura para emitir el canto) que se encuentra más desarrollada en Oscines (Tudge, 2008). A los miembros de este grupo se les conoce también simplemente como "pájaros cantores" o aves canoras. Estas aves llevan el nombre de una categoría especial de señales acústicas, es decir, "cantos" (Koodsma & Miller, 1996). Los estudios fisiológicos de Konishi (1965) con *Zonotrichia leucophrys* permitieron definir que en ciertos grupos de aves no cambia la estructura del canto en relación a la frecuencia (Hz), cuando las aves son expuestas al ensordecimiento o aislamiento, es decir, que las especies de aves tienen un rango de frecuencia fundamental característica y que dentro de la uniformidad general que caracteriza a una especie o población, los sujetos puedan diferenciarse en los matices de su canto, permitiendo así su identificación (Manning, 1985; Tubaro, 1999). En 1975 Shiovitz inició el trabajo de estandarizar los términos para describir las unidades de los componentes de los cantos de aves (Thomas, 1994) y no fue hasta 1994 cuando Thomas retoma los estudios hechos por Scholvitz y publica "A System for Describing Bird Song Units", donde define a las notas y las sílabas como unidades básicas del canto seguido de la frase y por último la canción, como se puede observar en la siguiente figura.

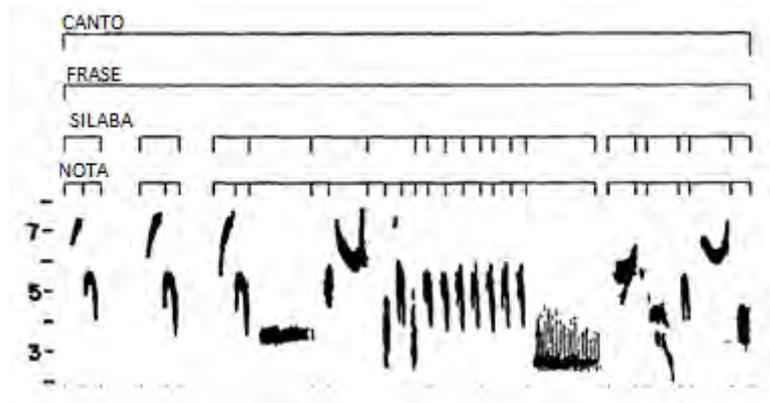


Figura 3. Nomenclatura del canto de Aves (Tomado y modificado de Thomson, 1994)

2.2.2 Características del canto las Aves

Considerando lo anterior, existen varias características del canto de las aves que lo convierte en un atractivo modelo morfo-neuro-etológico (Brenowitz, 2002), ya que, cuando un ave vocaliza, produce un sonido que sirve para comunicarse con los otros miembros de la misma especie y otras especies, debido a que la comunicación vocal es una estructura sonora especial y exclusiva, a la que se le llama “señal acústica” (Catchpole & Slater, 2008). Esta señal puede ser en dos tipos: cantos y llamados. Los cantos son producto del aprendizaje y de carácter conductual que principalmente sirven para la atracción sexual, en general tienen una estructura mucho más larga y compleja (variante en frecuencia Hz), mientras que los llamados son señales de carácter conductual que sirven para dar alarma, defensa de territorio, forrajeo, etcétera. Los llamados tienen un carácter acústico mucho más sencillo, puesto que solo corresponden a pequeñas notas simples y silabas de una frecuencia no variante (Herrera, 2013; Gill, 2006; Catchpole & Slater, 2008).

La siringe (Figura 4) es la estructura morfológica de las aves especializada para desarrollar el canto (Brenowitz, 2002; Gill, 2006; Catchpole & Slater, 2008) se encuentra conectada a los bronquios y a la tráquea. La producción del sonido es a partir de la vibración de las membranas timpaniformes (también llamado *labium*) las cuales se encuentran en la siringe, producto de la exhalación y movimiento de los músculos asociados a la siringe (Catchpole & Slater, 2008).

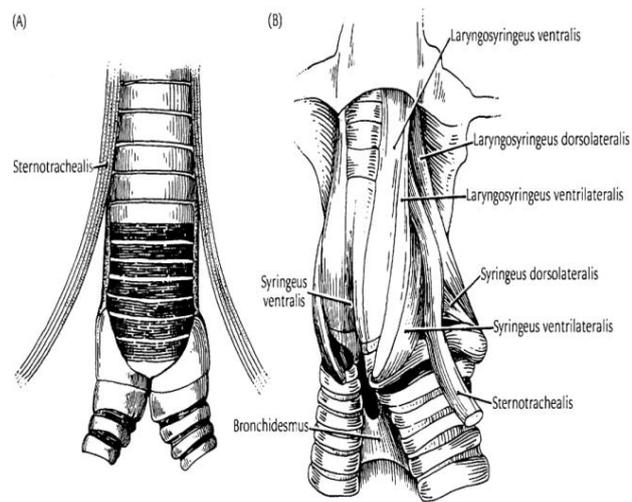


Figura 4 Estructura de la siringe y músculos que la rodean (A)muestra la conexión entre la siringe y la tráquea, (B) Muestra los diferentes músculos asociados a la siringe. (Tomado de Gill,2006),

La cantidad y complejidad de los músculos asociados a la siringe, fue uno de los factores por el cual el Sibley & More (1990) dividieron el Orden Passeriformes en dos grupos, los Oscines (Passeri) y los Subocines (Tyranni), siendo los Oscines (Passeri) el grupo que mayor cantidad de músculos asociados tiene a la siringe (más de 7) y los Subocines el que cuenta con menor cantidad de músculos (menos de 7). Con lo que se supone que al tener mayor cantidad de músculos asociados a la siringe se logra una mayor compresión y descompresión que logra modular el sonido

y da como resultado una vocalización más armoniosa. Sin embargo, existen otras diferencias entre estos dos grupos relacionada con la vocalización, una de ellas es la forma de aprendizaje de los cantos y la estructura cerebral subyacente (Catchpole & Slater, 2008).

Como se ha explicado hasta el momento, la siringe juega un papel importante en la producción de sonido al haber compresión y descompresión por los músculos asociados. Sin embargo este sistema vocal esta neumáticamente acondicionado con al menos dos componentes básicos: los pulmones muscularmente comprimidos y los sacos de aire que proporcionan un constante suministro de aire (Fletcher, 1992). Estos dos componentes logran generar que la membrana timpaniforme (Figura 5) logre vibrar y romper la tensión, lo que genera un tono puro a lo que se le denomina “Frecuencia fundamental”, mientras que el tracto vocal (tráquea, laringe, y pico) actúa como una caja de resonancia la cual atenúa armónicos y resalta la frecuencia fundamental (Podos,2001).

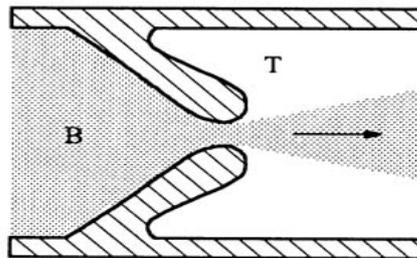


Figura 5 Movimiento de membrana timpaniforme, (B) es el flujo de aire al salir, (T) es la tensión generada por la membrana (Tomado de Fletcher, 1992)

2.2.3 Características acústicas del canto en aves

Las vocalizaciones tienen bien definidas estructuras acústicas que son características de cada especie y son fácilmente analizables (Brenowitz, 2002). El intervalo de frecuencia de las aves se encuentra aproximadamente entre los 700 Hz y 2200 Hz según Fletcher (2004), aunque Marler (2004) da un mayor rango de entre los 500 Hz y los 10,000 Hz e incluso hasta los 14 000 Hz en vencejos (Herrera, 2013). 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” (Figura 6).

Por regla general, la frecuencia fundamental es la que lleva asociada más potencia sonora (Carrión, 1998). La frecuencia fundamental y los armónicos se forman como consecuencia de la no linealidad del flujo a través del tracto vocal. Los formantes son las resonancias propias de cualquier elemento que tenga la capacidad de resonar (vibrar). Un 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 siringe (Fletcher, 1992).

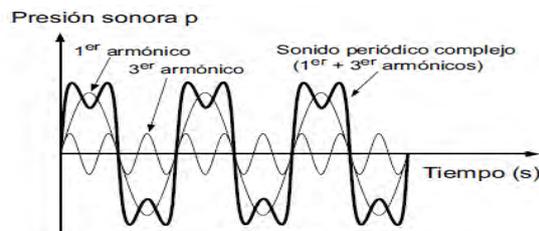


Figura 6. Sonido periódico con armónicos (Tomado de Carrión, 1998)

Por último, otro factor importante en las propiedades de la vocalización de las aves es la intensidad, magnitud ó volumen, la cual mucho depende del tamaño de la especie o individuo, intención conductual y distancia entre la emisor y el receptor. La unidad de esta propiedad es el decibelio (dB), la transmisión de la señal mucho depende de la degradación en el medio ambiente que implican factores como temperatura y humedad (Carrión, 1998). Catchpole & Slater (2008) proponen con fines ilustrativos cómo es que se va degradando el canto a distintas distancias (Figura 7).

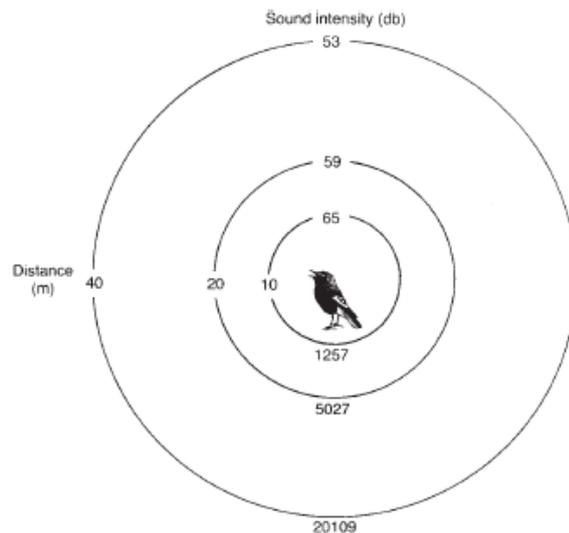


Figura 7. Intensidad y distancia de la vocalización de un ave (Tomado de Catchpole & Slater, 2008)

De acuerdo con lo anteriormente descrito, en los estudios de bioacústica se han establecido métricas como la frecuencia mínima, frecuencia máxima, duración y ancho de banda (rango o cobertura entre la frecuencia mínima y la frecuencia máxima) para caracterizar cada uno de los elementos que componen el canto de un ave en un espectrograma (Figura 8).

Dichas métricas han permitido realizar distintos estudios descriptivos y comparativos, con diferentes aplicaciones en áreas como la sistemática, la conservación y el manejo de las poblaciones de aves silvestres (Tubaro, 1999).

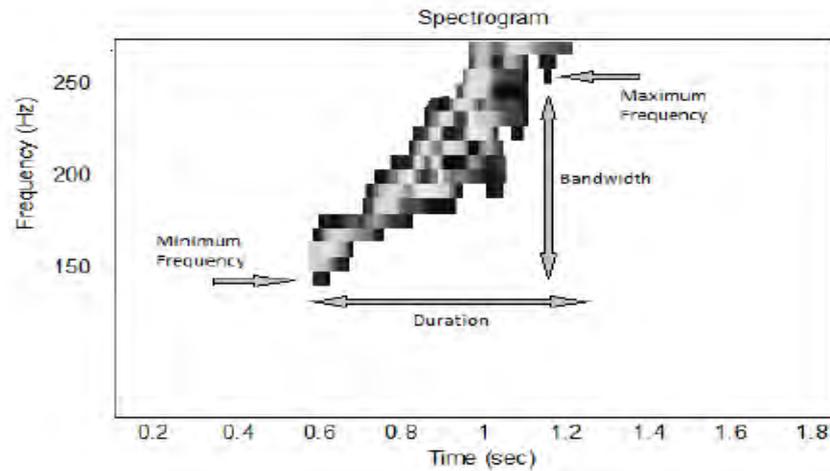


Figura 8. Métricas utilizadas para caracterizar la vocalización de un ave.
(Tomado de Abouseiman et al., 2013)

2.2.4 La Bioacústica en el monitoreo de las Aves

Los estudios antes mencionados han dado origen al desarrollo de métodos para identificar y contabilizar aves a partir del canto, cuando la observación de éstas es mínima o nula debido a que la densidad de la vegetación o los hábitos de las aves no permiten la observación (Tubaro, 1999). Dichos métodos han tenido mayor éxito en la etapa reproductiva, ya que es cuando mayor cantidad de cantos realizan las aves (Cunningham & Baker, 1987). Una de las técnicas más usadas para la captura de datos del canto es la grabación con parábolas y micrófonos direccionales, para aves en específico (Catchpole, 2004) y estudios relacionados a la taxonomía y conducta (Tubaro, 1999). Para los censos se han realizado varias evaluaciones conocidas como “sistema de grabación bioacústico”, con el

fin de estimar la riqueza y abundancia de especies. Inicialmente se realizaron grabaciones con micrófonos omnidireccionales (Farrel & Gannon, 1999; Haselmayer et al., 2000; Rempel et al., 2005; Conway & Gibbs, 2005), también se han realizado “sistemas de grabación paisaje sonoro” y grabaciones cuadrafónicas con micrófonos direccionales (Celis et al., 2009; Pieretti et. al., 2010), con micrófonos cardioides (Frommolt & Tauchert, 2012) y “unidades de grabación automática” (ARUs por sus siglas en inglés) capaces de grabar grandes cantidades de información, de acuerdo a la programación establecida por el usuario (Caycedo et al., 2013) . Estos estudios coinciden en sus resultados al tener una mayor percepción de aves acústicamente en comparación con los censos tradicionales mediante el uso de redes y el conteo por puntos, y más aún cuando se trata de estudios en áreas remotas.

Estos métodos acústicos juegan un papel prominente en los esfuerzos de monitoreo de aves debido a que muchas de ellas pueden ser escuchadas más fiablemente a distancias mucho mayores de lo que se pueden observar, sin embargo, varios factores limitantes pueden reducir la eficacia de dichas técnicas acústicas (Rosemberg, 2009). A pesar de su eficacia, la instrumentación para llevar a cabo este tipo de estudios tiene un costo económico elevado y la interpretación de datos sigue dependiendo de la experiencia del investigador, ya que la grabación y la posterior identificación con espectrogramas requieren conocimiento previo especializado. Así, la confiabilidad de los datos disminuye sensiblemente cuando el investigador no cuenta con experiencia pertinente. Otro inconveniente de los métodos mencionados acústicos es que no cumplen con el concepto de que los estudios encaminados al monitoreo y conservación de poblaciones debieran ofrecer herramientas accesibles, de fácil manejo y bajo costo para su utilización (Villaseñor, 2003).

Debido a la importancia de las aves como especies indicadoras de la salud ambiental, existe una considerable cantidad de trabajos publicados sobre la detección automática de vocalizaciones de aves basada en métodos computacionales con más de una década de investigación (Towsey et al., 2012) sin embargo existe una gran dispersión en literatura de ingeniería y ciencias biológicas (Caycedo et al., 2013), también lo que ha implicado obtener distintos enfoques.

La implementación para la identificación basada en métodos computacionales generalmente se sustenta en la extracción de huellas digitales del audio correspondiente a una vocalización de un ave. Dichas características acústicas relevantes de una pieza de audio se encuentran contenidas y almacenadas en una base de datos (Cano et al., 2005), la cual posteriormente mediante con algún método de aprendizaje automático se genera un modelo el cual realiza predicciones sobre ejemplos mostrados para su identificación (Camastra & Vinciarelli, 2007).

Un enfoque común ha sido la adaptación de las herramientas de 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 (Adda-Decker, 2001; Pardo, 2002). Se ha realizado reconocimiento automático de las unidades del canto en aves con grabaciones continuas y Modelos ocultos de Markov, trabajos realizados con los pinzones *Taeniopygia guttata* y *Passerina cyanea* en laboratorio, y los resultados mostraron una gran respuesta al reconocimiento (Kogan & Morgaliash, 1997). Otra técnica utilizada es realizar experimentos para la identificación automatizada de cantos de aves con sistemas de discriminación lineal (LDA). Esta técnica permitió el reconocimiento del 87% de cantos de aves (Hashing et al., 2006) y 65-75%

(Agranat, 2009). Sin embargo una de las mayores problemáticas de la utilización de herramientas de ASR es que existe una confusión en el concepto de reconocimiento, ya que el objetivo de ASR es que una maquina (computadora) de una forma automática convierta las palabras, que son emitidas por el ser humano, a texto (Moya, 2011), por lo que no cumple con la expectativa de la identificación y por tanto ha generado una serie de fallas en el rendimiento de los programas, ya que no son capaces de identificar de forma adecuada en ambientes naturales (Towsey et al., 2012), pues al hacerlo, el algoritmo es susceptible a la clasificación errónea de una vocalización de diferentes especies con componentes similares (Agranat, 2012).

Existen otras dificultades que afectan la identificación en ambientes naturales. En primer lugar, las grabaciones hechas reciben los sonidos de todas las direcciones, esparcidas y se reflejan en los árboles, ocultas por una impredecible constelación de ruidos aleatorios como el viento, las hojas susurrantes, los aviones, los autos, otras especies de aves, ranas, insectos y mamíferos. En segundo lugar, las vocalizaciones de muchas especies son muy variadas de un individuo a otro. Por tanto cualquier algoritmo debe estar preparado para aceptar las vocalizaciones que son similares, pero no idénticas, afinado con referencias conocidas con el fin de detectar con éxito el individuo no observado previamente (Agranat, 2012). Existen distintas propuestas para grabación de aves y sistemas computacionales para la identificación a partir de grabaciones, sin embargo no han existido propuestas que combinen ambos aspectos para generar sistemas automatizados que permitan estar colectando cantos e identificando automáticamente en campo al mismo tiempo. Por lo que éste trabajo se enfoca en generar dicha propuesta.

3

Planteamiento del problema

Recapitulando, distintos animales utilizan diferentes frecuencias para comunicarse por sonido, esto se debe principalmente a que la evolución de los animales se ha adaptado a frecuencias particulares para dar mayor comunicación conespecífica. Por lo que, varios investigadores estiman el área de acción y censo de aves a partir de las vocalizaciones. Esto ha representado una ventaja sobre métodos convencionales para el monitoreo de aves. Además esto es una alternativa de seguimiento ante las problemáticas en la reducción de poblaciones silvestres. Sin embargo existe la necesidad de establecer nuevas estrategias y técnicas cada vez más avanzadas, tanto en la instrumentación como en el procedimiento, ya que el análisis e identificación de aves por canto continúa siendo de forma manual y depende de la experiencia del investigador, lo que implica un desgaste de tiempo y que dicha tarea sea restringe a la experiencia.

3.1 Objetivo General

Desarrollar un sistema de identificación acústica de *Turdus migratorius* basado en un modelo de aprendizaje automático con la extracción de características acústica-estadísticas del canto.

Diseñar e implementar un dispositivo robótico retroalimentado con el modelo de aprendizaje automático para la identificación de *Turdus migratorius* en tiempo real.

3.2 Objetivo Particular

Generar un sistema de extracción de características acústicas-estadísticas basado en los vectores de audio que conforman las sílabas del canto de *Turdus migratorius*.

Desarrollo de un modelo de aprendizaje automático para la identificación de *Turdus migratorius*.

Planeación, diseño e implementación del sistema de identificación acústica de *Turdus migratorius* en un dispositivo robótico.

3.3 Justificación

La identificación de animales basadas en parámetros acústicos y computacionales, muestra un método no invasivo que presenta una ventaja considerable, ya que se puede implementarse fácilmente y brinda una mayor cobertura en las tomas de muestra. Así mismo este tipo de sistemas ayudarían a reducir el tiempo invertido de los investigadores en la identificación manual, lo que representa un costo considerable. La principal ventaja del método es que no sería perjudicial para el bienestar animal (Fox, 2008; Agranat, 2012).

La elección para incorporación del canto de *T. migratorius* se debe a que esta especie cuenta con un status de residencia en la parte central de México y tiene niveles elevados en su abundancia, lo cual nos brinda una facilidad en la localización de dicha especie. Precizando sobre el canto de la especie, está cuenta con 220 sílabas que componen su repertorio vocal,

de las cuales entre 9-16 sílabas repite constantemente aproximadamente un 95% de las veces que esta vocaliza (Peshek& Blumstein, 2011).

Por último, los avances tecnológicos en bioacústica y monitoreo de fauna silvestre requieren de nuevas instrumentaciones las cuales permitan la recolección de datos en múltiples lugares remotos a lo largo del tiempo (Blumstein *et al*, 2011). Los sistemas actuales llamados unidades de grabación automática (ARUs por sus siglas en inglés) solo permiten realizar grabaciones, sin embargo no logran tener un proceso automatizado de identificación.

3.4 Hipótesis

Si las Aves tienen un patrón acústico característico en el canto, entonces posible crear un modelo computarizado, automático para la identificación de *Turdus migratorius*, el cual esté basado en comparaciones estadísticas y acústicas.

4

Método

En esta sección se muestra el proceso para generar un sistema automático para la identificación acústica de aves como eje central. Esta serie de actividades para alcanzar nuestro objetivo fue realizada en el sistema operativo Ubuntu, con el lenguaje de programación Python, este proceso llevó una serie de actividades paralelas:

- La primera actividad consistió en la generación de una base de referencia (*corpus*) que contara con grabaciones de los cantos de aves, y la segmentación de las grabaciones para separar los cantos del ruido.
- La segunda fue el procesamiento digital de los cantos para generar un modelo acústico basado en la extracción de los MFCCs (*Mel-frequency cepstral coefficients*). Posteriormente los valores resultantes de la extracción se utilizaron para generar un modelo estadístico. Estos dos últimos se consideraron para alimentar un sistema de aprendizaje con la herramienta de aprendizaje automático SVM (*Support Vector Machines*), que de acuerdo a los valores ingresados al sistema de aprendizaje para cada especie, dio como resultante un peso de referencia para cada especie.
- La actividad final constó de un diseño experimental para evaluar el modelo obtenido de SVM. Ésta comprendió de una evaluación con grabaciones y otra con un sistema de identificación en tiempo real, para lo que se construyó un Módulo de Identificación Acústica de Aves (MIAA, ver Planeación y diseño de MIAA).

4.1 Base de Audio-datos

Para poder generar un sistema que identifique especies de aves, fue necesario crear una serie de bases de datos de audio, por lo que se realizó una colecta de audio con cantos y llamados para 6 especies pertenecientes al Orden Passeriformes (Tabla 1), considerando que los cantos de cada especie contara con características acústicas particulares y distintos grados de complejidad en la composición del repertorio vocal. La selección de los cantos se realizó de acuerdo a la calidad y claridad del audio. Dichos audios fueron obtenidos de búsquedas en las audiotecas bioacústicas: Xenocanto, Maculay Library The Cornell Lab Ornithology, Biblioteca de Sonidos de Aves del Museo de Zoología de la Facultad de Ciencias, UNAM y de biblioteca privada (Autor). Debido a la gran variedad de formatos en los que se encontraban los audios, se realizó una estandarización con la edición de las grabaciones utilizando el software Cubase 5 Steinberg (*Advanced Music Production System*), por lo que se estandarizó al formato de audio .wave.

Especies	Tiempo
Turdus migratorius (Linnaeus, 1766)	60 min
Turdus rufopalliatus (Lafresnaye, 1840)	16 min
Myadestes occidentalis (Stejneger, 1882)	36 min
Thryomanes bewickii (Audubon, 1827)	24 min
Toxostoma curvirostre (Swainson, 1827)	36 min
Cardinalis cardinalis (Linnaeus, 1758)	21 min.
	Total 4.1 hrs

Tabla 1. Especies consideradas y tiempo total de grabación.

4.1.1 Desarrollo de Corpus

En tecnologías de identificación de voz, se genera un *corpus* para crear un modelo acústico a partir de una base de audio-datos. En el presente trabajo se utilizó la base de audio-datos antes mencionada, siendo el primer paso la etiquetación de las sílabas del canto de aves [1], con el software Praat versión 5.3. La finalidad es la segmentación manual de las sílabas y el sonido de fondo (Figura 9), para generar una etiquetación llamada (*gold standard*) basada en la experiencia de un experto y, así tener una referencia más precisa de las sílabas.

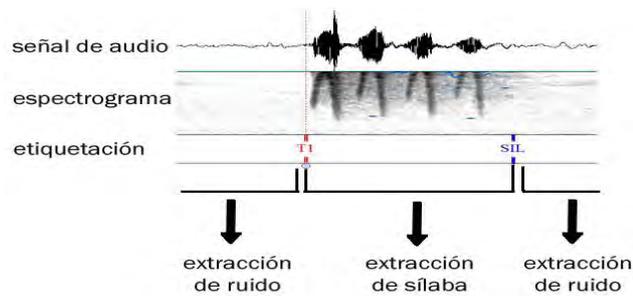


Figura 9. Etiquetación de sílaba de *Turdus migratorius*. La terminación *TI* señala el inicio de la sílaba y *SIL* la finalización de la sílaba y el inicio de un silencio.

[1] Debido a que las notas del canto tienen varias dificultades en la identificación entre las fronteras de ellas, ya que no siempre es notable su distinción. Se optó por utilizar sílabas que de igual manera tienen una consistencia acústica. Este mismo concepto también se ha utilizado en trabajos de reconocimiento e identificación de voz humana como el de Oropeza & Suarez (2006).

4.2 Procesamiento Digital de la Señal

Una grabación digital de un canto de ave es una “señal aperiódica y discreta” (Mitra, 2007; Cuevas et al., 2010), y puede estar representada de dos maneras, de forma dinámica y espectral (Figura 10). En la forma dinámica se representa un vector $[n]$ que contiene valores/muestra de las variaciones de volumen con respecto al tiempo y en la forma espectral se representan las frecuencias que componen el canto de un ave. La cantidad de valores/muestras depende de la actividad obtenida y de la duración de la señal grabada. Debido a que generalmente la cantidad de valores/muestra son muy grandes y variantes, es muy difícil crear un modelo que represente todas las silabas de un repertorio vocal (variedad de cantos de un ave). Para esto se estableció la extracción de las características acústicas, con el objetivo así reducir el tamaño de los vectores que representan el canto de un ave y encontrar patrones representativos.

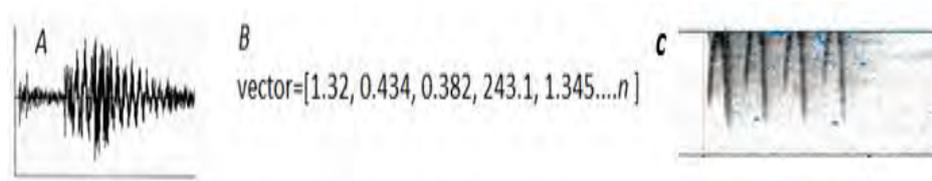


Figura 10. (A) Representación dinámica de una señal de audio. (B) Representación matemática-computacional de una señal de audio. (C) Representación espectral.

4.2.1 Ventaneo

Como se mencionó anteriormente, la señal de audio no es periódica, por lo cual no es posible calcular sus componentes en frecuencias únicas. Debido a esto se hace de manera discreta, es decir por ventaneo. Cada ventana tiene una longitud de 400 valores-muestra y un traslape de 160 valores-muestra entre ventanas (Figura 11).

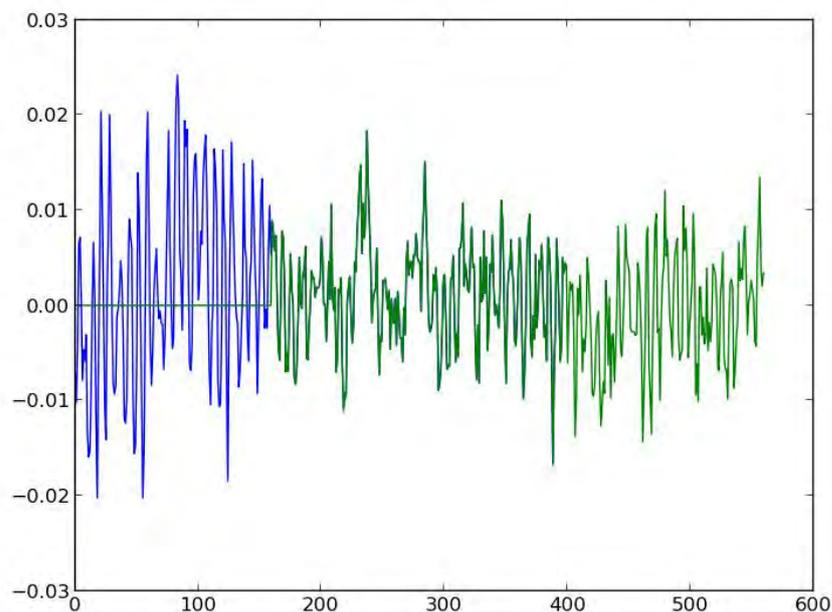


Figura 11. Se muestra el ventaneo de 400 valores-muestra y el traslape de una de ellas con la ventana anterior.

El objetivo del ventaneo fue realizar suficientes muestras de tal manera que se cubra la riqueza de frecuencias. Sin embargo el corte entre ellas es muy abrupta por lo que puede generar ruido (Sanjit, 2007), para evitar dicho ruido se implementó la ventana Hamming (Fórmula 1). La implementación de la ventana Hamming consistió en realizar una disminución gradual de los lóbulos laterales, con un incremento correspondiente en el ancho del

lóbulo principal, por último, esto permitió unir las ventanas de tal manera que se convirtió en un filtro de frecuencias espurias o ruido (Figura 12).

$$\text{Hamming: } w(n) = 0.54 + 0.46 \cos\left(\frac{2\pi n}{2M + 1}\right)$$

Fórmula 1. Implementación de ventana Hamming.

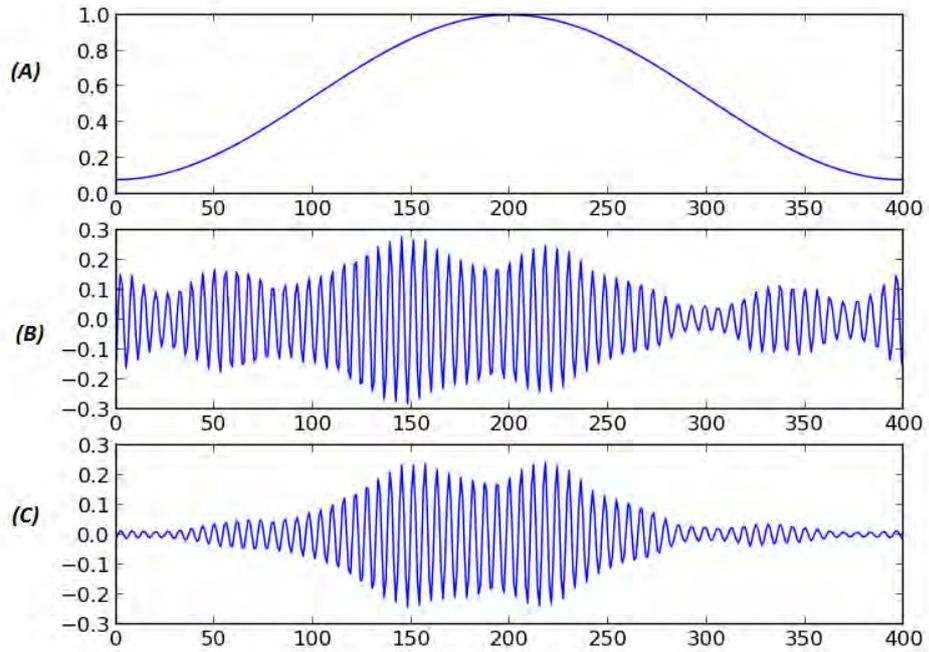


Figura 12. (A) Ventana de Hamming para una muestra. (B) Muestra sin ventana de Hamming. (C) Ventana de Hamming aplicada a la muestra

4.2.2 Método de Transformada Discreta de Fourier (DFT)

El análisis frecuencial sobre señales continuas se realiza básicamente a través de la Transformada de Fourier (Proakis & Manolakis, 1996). La importancia de estos métodos radica en la descomposición de la señal en frecuencia, lo cual es muy útil para la identificación de las frecuencias con mayor energía y dominantes. Se han desarrollado algoritmos que han logrado hacer que su cálculo sea rápido “*Fast Fourier Transform*” (Proakis & Manolakis, 1996; Snajit, 2007). Sin embargo, las grabaciones obtenidas de las aves son digitales, por lo que son señales aperiódicas y discretas (como ya se había mencionado anteriormente), por lo que existe una manera de analizar este tipo de señales con una variante llamada “Transformada Discreta de Fourier” (DFT, véase Fórmula 2).

La transformada discreta de Fourier al igual que la Transformada de Fourier posee un modelo para la transformación de una señal discreta a su correspondiente en el dominio de frecuencia, de igual manera permite a través de otro modelo reconstruir la señal (Cuevas et al., 2010).

$$F(N) = \frac{1}{\sqrt{M}} \sum_{n=0}^{M-1} f(n) \cdot e^{-j \frac{2\pi Nn}{M}}$$

Fórmula 2. DFT para una señal discreta $f(n)$ de longitud M .

Se calculó la DFT para cada ventana obtenida (Figura. 13) y se identificaron las frecuencias dominantes con respecto a la energía.

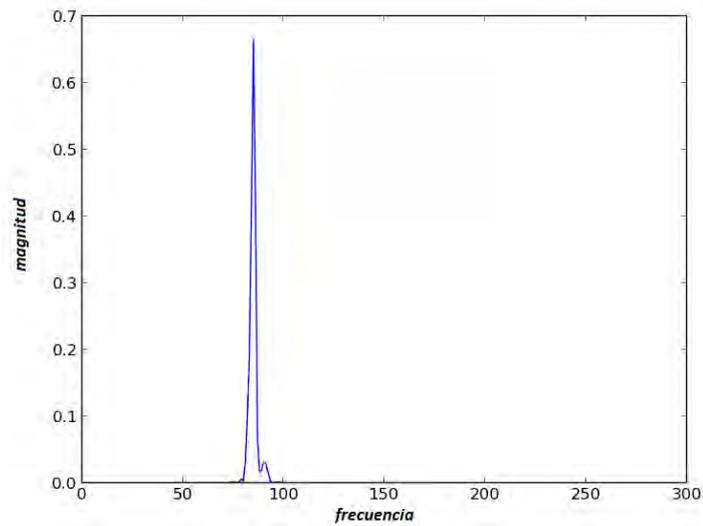


Figura 13. DFT aplicada a una ventana

Con los valores resultantes de la aplicación de la DFT a cada una de las ventanas se calculó el *peridiograma* para representar la energía por cada valor de frecuencia con respecto al tiempo (Figura 14).

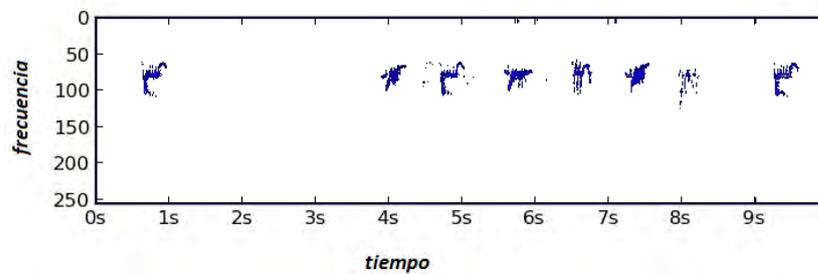


Figura 14. Peridiograma

4.3 Extracción de Frecuencias Mel de los coeficientes cepstrales (MFCCs)

La extracción de una buena representación paramétrica de señales acústicas es una tarea importante para el rendimiento de reconocimiento e

identificación (Lindasalwa et al., 2010). Existe una amplia variedad de parámetros que se pueden utilizar para representar las partes más sobresalientes de una señal como: los coeficientes de la DFT de la señal, los coeficientes cepstrum, los coeficientes cepstrales de la escala Mel, la energía de la señal, entre otros. Todos ellos se han empleado de forma individual o combinada en sistemas de reconocimiento de voz y de características musicales (Salcedo, 2009). Sin embargo, para generar un sistema más versátil que permita la identificación de una amplia variedad de características acústicas, se precisó una parametrización lo más general posible que extrajera toda la información valiosa de la señal. Por ello, los parámetros escogidos para representar la información de la señal son los “Coeficientes Cepstrum de la Escala Mel” (MFCCs), los cuales se basan en las percepciones auditivas humanas (según la escala Mel). La escala Mel es una aproximación a la escala perceptual humana. Viene dada por la siguiente expresión:

$$Mel(f) = 2595 \cdot \log\left(1 + \frac{f}{700}\right)$$

Fórmula 3. Mel. (Tomado de O'Shaughnessy, 2000)

4.3.1 Filtros Mel

Como se había mencionado anteriormente en el punto 4.2.2, la Transformada Discreta de Fourier permitió identificar las frecuencias que componen la señal, sin embargo la cantidad de frecuencias en una grabación ambiental donde se encuentra un ave es muy grande; por lo que se diseñó un filtro Mel. Para el cálculo de los MFCCs se utilizan una serie

de filtros triangulares paso-banda (en un determinado rango de frecuencias) con un solapamiento entre filtros. La idea principal de este parámetro es la obtención de vectores de coeficientes cepstrum en los cuales el espaciamiento en frecuencias no es lineal, si no que se distribuye de manera logarítmica simulando la percepción humana (Salcedo, 2009).

Se realizó un diseño de un filtro de 40 valores mediante la siguiente ecuación:

$$c_i = \sqrt{\frac{2}{P}} \sum_{j=1}^P m_j \cdot \cos\left(\frac{\pi i}{P}(j - 0.5)\right)$$

Fórmula 4. Cálculo del número de filtros Mel. Donde P es el número de filtros aplicados a la señal y m , es el parámetro resultante del filtrado Mel en la banda j .

Como se puede apreciar en la figura 15, el diseño del filtro Mel concentra en mayor cantidad las frecuencias bajas y con mayor amplitud (energía) ya que estas son más informativas, mientras que los filtros ubicados en altas frecuencias y con menor amplitud contienen menor información, sin embargo estas complementan la información acústica del canto de un ave.

Cada filtro se utilizó para concentrar la amplitud en distintas bandas (rango de frecuencia) en un solo valor, por lo que el resultado se resume en un vector de 40 valores.

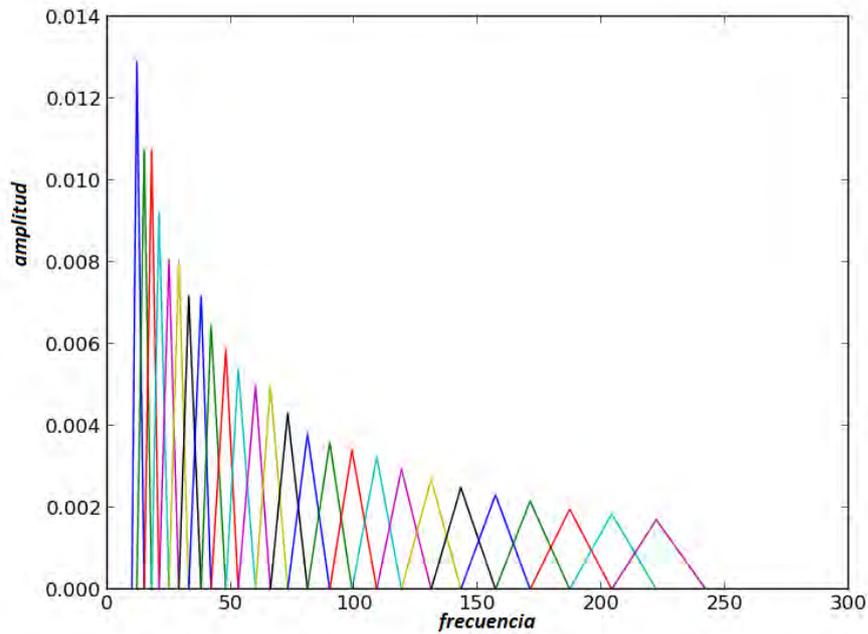


Figura. 15 Como se puede apreciar el filtro Mel de 40 valores concentra las frecuencias bajas en gran cantidad y disminuye la cantidad mientras aumenta de frecuencia.

La energía de una señal se define como la suma de valores cuadráticos de la misma y debido a que los valores de energía generalmente son muy grandes, se modificó la escala de la energía por una escala logarítmica para facilitar el manejo de los datos y su representación, por tanto un segmento de N muestras tiene la siguiente expresión:

$$E_N = \log \sum_{n=1}^N x^2(n)$$

Fórmula 5. Energía con su expresión logarítmica

Como se muestra en la figura 16 las ventanas resultantes de un coeficiente Mel y la figura 17 representa el peridiograma de un conjunto de coeficientes Mel.

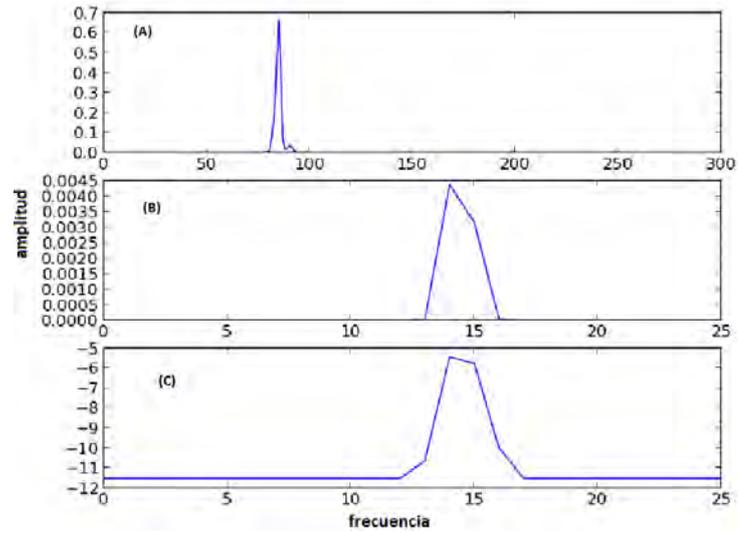


Figura 16. (A) Paso de una ventana de la DFT por un filtro Mel. (B) Valores obtenidos dentro de un filtro Mel. (C) Valores logarítmicos del filtro Mel.

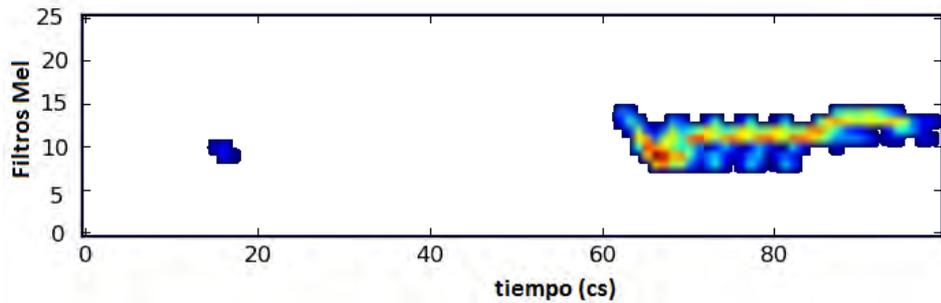


Figura 17. Periodograma de la unión de valores Mel en un segundo de muestra.

4.3.2 Transformada Discreta de Coseno (TDC)

Es posible recomprimir la información contenida de los 40 valores Mel incorporando la TDC para reducir los valores, lo cual nos permitió comprimir la señal de 40 coeficientes por ventana a solo 16 coeficientes cepstrales, lo que nos permitió que la señal obtuviera elementos informativos con menor número de valores como se muestra en la figura 18.

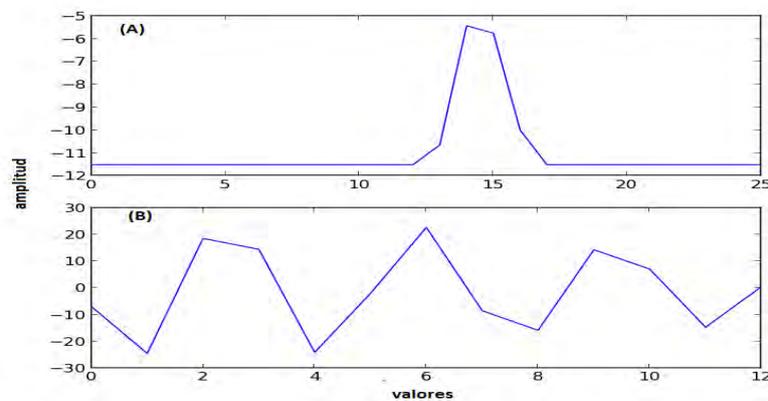


Figura. 18 (A) Coeficiente logarítmico cepstral Mel. (B) Coeficiente descriptivo espectral resultado de la TDC.

4.3.3 Calculo de los Delta MFCC (Δ MFCC)

A diferencia de las características MFCC las cuales solo describen el envolvente espectral de potencia de una ventana, autores como Davis (1980) y Huang et al., (2001) señalan que también estas tienen información en su dinámica, es decir, nos señalan cuáles son las trayectorias diferenciales y de aceleración con respecto al tiempo de cada ventana, y sugieren que este tipo de expresiones pueden aumentar el rendimiento en las tareas de identificación y reconocimiento. Por lo que

en este también estudio se consideró hacer una comparativa de rendimiento entre los MFCC y los Δ MCFF, para encontrar un mejor rendimiento en la tarea de identificación del canto de *T. migratorius*.

Por lo que se calcularon los coeficientes delta, con la siguiente fórmula:

$$d_t = \frac{\sum_{n=1}^N n(c_{t+n} - c_{t-n})}{2 \sum_{n=1}^N n^2}$$

Fórmula 6. Calculo de la dinámica de los MFCC

4.3.4 Estadísticas: descripción de la señal

De acuerdo a lo anterior el vector resultante con 16 valores por ventana representan solo un fragmento de una silaba y no toda en su conjunto. Lo cual indica que hasta este punto solo se ha representado una sección de la señal de un canto de *T. migratorius*. Además se desconoce el número de ventanas por silaba, tarea que dificulta la retroalimentación del sistema automático de aprendizaje. Por lo que se describieron estadísticamente en conjunto la serie de ventanas y se agruparon en un solo vector, de tal manera que cada silaba de *T. migratorius* queda representada por un solo vector. Este tipo de implementación para la descripción estadística se ha realizado en los trabajos como el Bodgan et al., (2007) y Björn et al., (2011). Donde se utilizaron nueve funciones estadísticas, de tal manera que se redujo el tamaño de la señal de una silaba completa y queda resumida a un solo vector de 144 valores, como se ejemplifica en la figura 19.

Ventana 1	...	Ventana n	Total
*	...	*	Media
*	...	*	Desviación Estándar
*	...	*	Cuartil 1
*	...	*	Cuartil 2
*	...	*	Cuartil 3
*	...	*	Mínimo
*	...	*	Máximo
*	...	*	Asimetría
*	...	*	Curtosis

Figura 19. Funciones aplicadas a cada ventana.

4.5 Entrenamiento: *Support Vector Machine* (SVM)

La máquina de soporte vectorial o *support vector machine* (SVM) fue propuesta por primera vez por Vapnik (1982). Esta base matemática permite que SVM se pueda utilizar para la clasificación de patrones y ha llegado a ser popular debido a sus características atractivas y robustez en el campo del aprendizaje automático. La idea central de SVM es construir un hiperplano de tal manera que se genere un margen maximizado en la separación entre los ejemplos positivos y negativos como se muestra en la figura 20 (Kamruzzaman, 2010). Considerando lo anterior el entrenamiento de SVM consistió en ingresar la serie vectores resultantes de las

extracciones de características acústicas y estadísticas tanto de las sílabas de *Turdus migratorius* (positivos) y ruido “No *Turdus migratorius*” (negativos) donde se incluían las otras 5 especies más los intervalos de “ruido” entre sílabas.

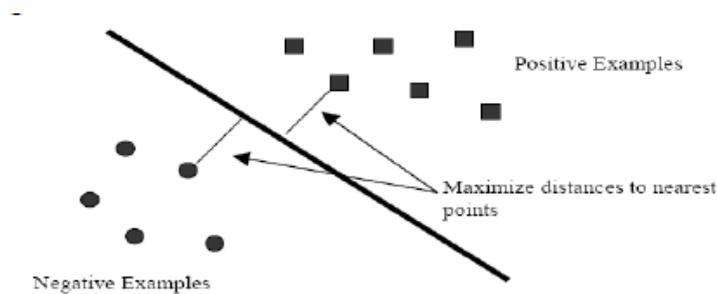


Figura 20. Máquina de soporte vectorial lineal (tomado de Kamruzzaman, 2010).

De esta manera, SVM se entrena y se encarga de encontrar la diferencia entre el canto de “*Turdus migratorius*” y “No *Turdus migratorius*” (ruido). Este análisis da como resultado un modelo aprendido que puede ser utilizado para predecir la entrada de nuevos ejemplos desconocidos y de esta manera SVM toma una decisión respecto a lo aprendido.

4.6 El sistema de identificación

El sistema está compuesto de una serie de procesos para que a partir de grabaciones se logre la identificación del canto de *Turdus migratorius*. Sin embargo retomando el objetivo de esta tesis y consideraciones para generar un sistema orientado al monitoreo, este debiera recabar una serie de datos y realizar la identificación en tiempo real con el prototipo MIAA (Ver capítulo 4). Por lo que, se diseñó un medio de captura de audio llamado

“silabificador” como se muestra en la figura 21. El silabificador se encarga de realizar la captura en tiempo real basado en actividad acústica con un proceso que se encuentra dividido en dos fases: a) detector de actividad general (DAG) y b) detector de actividad específica (DAE). El detector de actividad general está basado en el umbral de energía por actividad de acústica (Milanovic, 1999) la cual va retomando las secuencias de actividad por muestras y a su vez considera la energía encontrada en cada muestra anterior, para poder así extraer un segmento de audio y recabarlo. Entonces, el segmento recabado es pasado por el detector de actividad específico el cual se encarga de analizar la señal de tal manera que cuando entra un canto este es segmentado en silabas (Pikrakis, 2008).

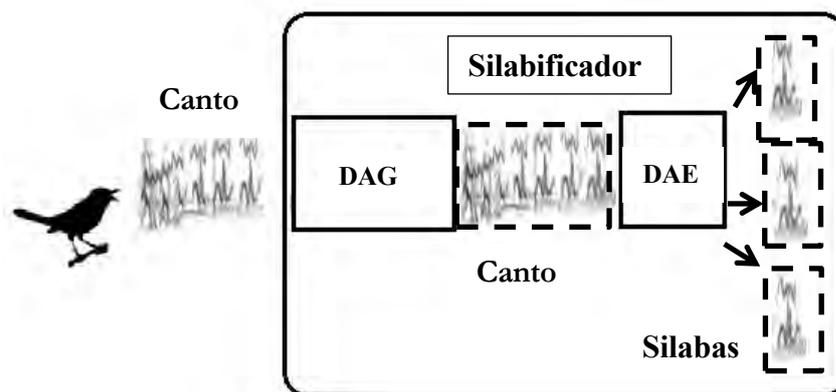


Figura 21. Proceso del silabificador. (DAG) Detector acústico general. (DAE) Detector acústico específico.

4.6.1 Funcionamiento del sistema de identificación

En la figura 22 se muestra como el sistema está compuesto de un micrófono para la grabación, un silabificador, un extractor de MFCCs y un clasificador. El micrófono realiza la captura de audio que alimenta el silabificador, el cual se encarga de identificar segmentos de actividad acústica. El extractor de características transforma de estos segmentos a su representación estadística de las MFCCs, las cuales alimentan al clasificador basado en SVM quien realiza una predicción sobre el tipo de segmento. El sistema en un ambiente funciona de tal manera que cuando canta *Turdus migratorius* u otro (puede ser otra ave o ruido), el canto es capturado por el micrófono y de los segmentos de actividad se identifican las silabas y estas son procesadas por el extractor de características para obtener las propiedades acústicas-estadísticas de la silaba, que por último dichas propiedades entran al clasificador el cual identifica si la actividad pertenece a *Turdus migratorius* o no.

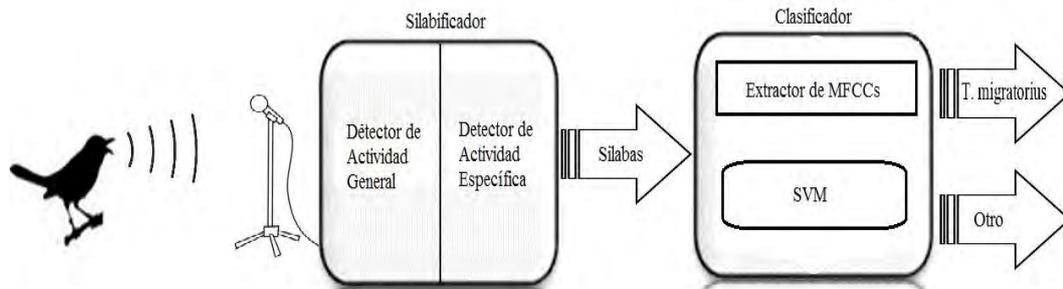


Figura 22. Funcionamiento del sistema de identificación.

Planeación y Diseño de MIAA

En esta sección se describen los pasos que se siguieron para la elaboración de MIIA (Modulo de Identificación Acústica de Aves) basado en las necesidades que actualmente se requieren en los estudios relacionados al monitoreo de aves, tales como auto sustentabilidad energética, automatización y toma de variables ambientales que brinden mayor información en los estudios pertinentes. También se incluyen las consideraciones de la implementación mecánica y electrónica del prototipo, así como las opciones que se obtuvieron en el mercado.

5.1 Diseño del Módulo de Identificación Acústica de Aves

El diseño del prototipo está basado en un tripié retráctil (Figura 23) el cual lleva como nombre MIIA, consta de una serie de elementos mecánicos y electrónicos con la finalidad de que MIIA tenga autonomía suficiente, considerando que en un futuro puedan existir varias unidades de MIIA, lo que implicaría que en el trabajo de campo se obtuviera un mayor número de datos en menor tiempo y mayor cobertura en distintas áreas.

MIIA, a pesar de ser modular (contener varios elementos), está orientado principalmente en la captura de datos acústicos para la identificación de aves, así como de otros elementos (GPS, temperatura y humedad relativa) que refuerzan el registro de información en campo, siendo de interés en los estudios relacionados a la bioacústica y fauna silvestre. También se consideró el hecho de que dicho prototipo debiera ser accesible al

transporte y sobre todo que el sistema se alimentara de energía renovable, sin la necesidad de depender de pilas desechables.

Al ser MIIA un módulo mecánico-electrónico para comunicar y manejar cada uno de los elementos, fue necesario implementar una unidad central o cerebro. Por tanto, se consideró como herramienta de comunicación, manejo y lectura de datos la tarjeta micro-controlador Arduino UNO, por su fácil utilización, capacidad de controlar varios elementos electrónicos (digitales y análogos) y que esté es de código libre (open source).

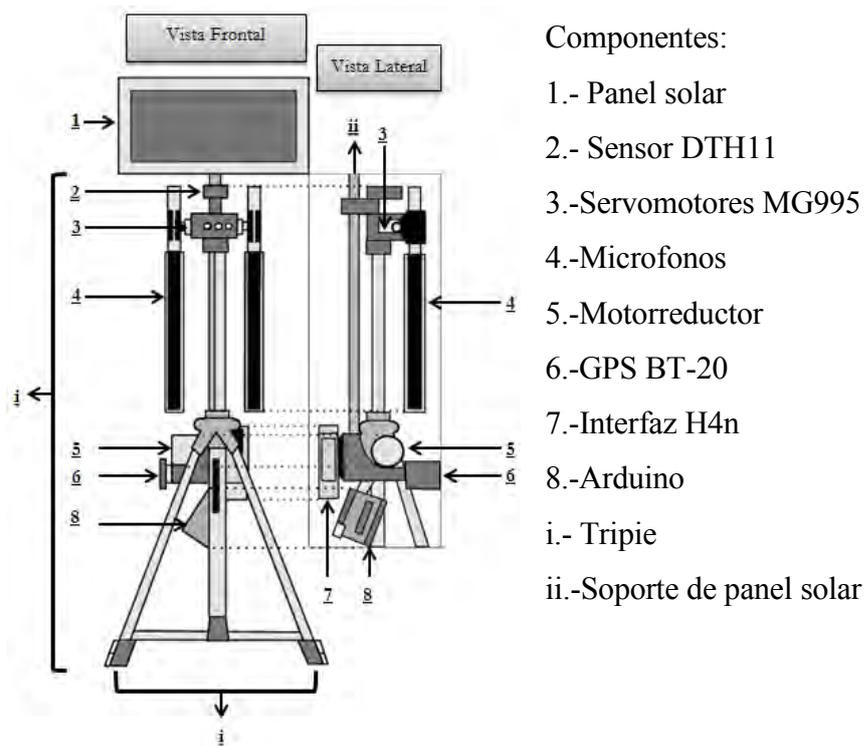


Figura 23. Diseño y composición de MIIA

5.1.1 Características de los Componentes

1. Panel Solar.

Un aspecto importante en la sustentabilidad eléctrica para el mantenimiento del prototipo fue la implementación de un panel solar, el cual permitió no depender de energías convencionales como las pilas, ya que estas tienen un tiempo de vida limitado y son un factor importante en la contaminación. La implementación del panel solar fue una solución idónea, ya que nos ofreció, además de la retroalimentación con energía solar (autosustentable), una mayor cantidad de energía disponible y la no cooperación en la contaminación del medio ambiente. El panel solar que se utilizó es de tipo Amorfo y cuenta con las siguientes características:

- Máximo Voltaje 15.6 v DC,
- Potencia Total 7.5 watts y
- Corriente Total 500 mA.
- Pila de retroalimentación con salidas de 12V, 12watts y 2.6Ah.

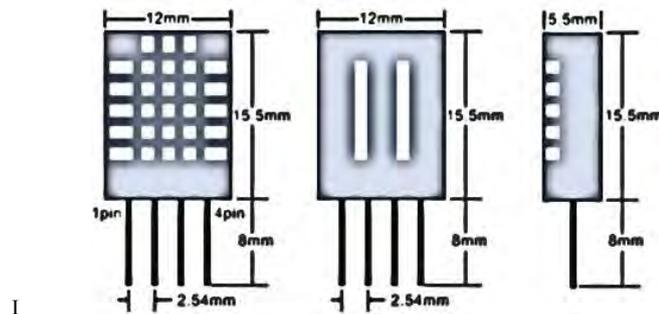
2. DTH11: Sensor de temperatura y Humedad.

Dentro de las formas de registro de información en campo uno de los apartados considerados es el relacionado a factores climáticos (Romero *et al*, 2007), ya que estudios como el de Botero (2009) apoyan la hipótesis de que el canto de las aves, así como, la complejidad de éste, puede verse afectada por patrones climáticos, que alteran tanto la distribución de las especies como la tasa de canto para la atracción, impulsada por la selección sexual y factores conductuales. Por este motivo se consideró la implementación del sensor DHT11 que utiliza una comunicación y

sincronización que da como producto lectura de datos relacionados a la temperatura y humedad relativa. La obtención de los datos se realizó mediante comunicación serial (PC-Arduino-DHT11).

Las características del sensor son las siguientes:

- Voltaje de operación 3.5-5.5V Dc.
- Consumo máximo 1mA.
- Mide de 20-90% RH (humedad relativa) con un margen de error de +/-5%.
- Mide de 0-50°C con un margen de error de +/- 2°C



MAGEN DE SENSOR DTH11(Tomado de Datasheet [2] DTH11)

3. Servomotores MG995.

Una de las características consideradas fue la movilidad en el prototipo, ya que puede dar resguardo y posicionamiento de los micrófonos, el cual está basado en la dependencia del movimiento de servomotores. Los servomotores realizan el movimiento de posicionamiento dentro de un rango de operación el cual es de solo 180° y da soporte a los micrófonos, permitiendo así el posicionamiento adecuado y resguardo.

[2] Datasheet es el documento en el cual se detallan las características, especificaciones, funciones, aplicaciones, etc., de un componente electrónico.

Se utilizaron los servomotores MG995, que tienen las siguientes características:

- Trabajan a un voltaje de 4.8V –7.2 V.
- Tienen un torque (fuerza de rotación) de 13 Kg/cm con 4.8V y 15 Kg/cm con 6V.
- Cuentan con una velocidad de operación de 0.17 seg/60 grados a 4.8V y 0.13 seg/ 60 grados a 6 V.
- La lectura y escritura de la posición se hace mediante comunicación digital, la movilidad de estos es solo de 180°.

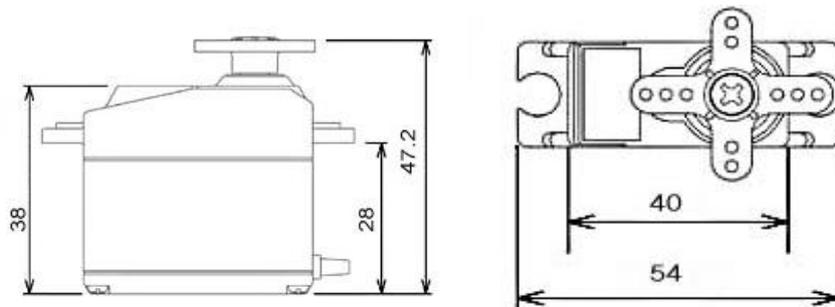


IMAGEN DE SERVO MOTOR MG995 (Tomado de Datasheet [2] MG995)

4. Micrófonos Audio-technica AT-8015 y AT-835a

Se consideró que a pesar de que el canto de las aves tiene una gran diversificación , el ancho de banda es relativamente pequeño ya que solo es de aproximadamente 2 kHz, es decir que el rango de frecuencias es de entre los 3kHz y 5 kHz. Por ello se consideró que los micrófonos debieran responder a dichas frecuencias y se optó por los micrófonos direccionales AT-8015 y AT-835^a, que cuentan con las siguientes características:

- Operan con fuente de energía phantom de 11V a 52V DC o con una batería de 1.5V

- Tienen un ángulo de aceptación estrecho adecuado para el registro de sonido a larga distancia, además de un excelente rechazo del sonido procedente de los lados y la parte posterior del micrófono. La posición de atenuación reduce el registro de ruido ambiental de frecuencias graves (como el tráfico o los sistemas de aire acondicionado) y las vibraciones mecánicas.

El micrófono AT-8015 tiene una frecuencia de respuesta 40-20 kHz, respuesta señal/ruido 72 dB, 1 kHz a 1 Pa (Pascal).

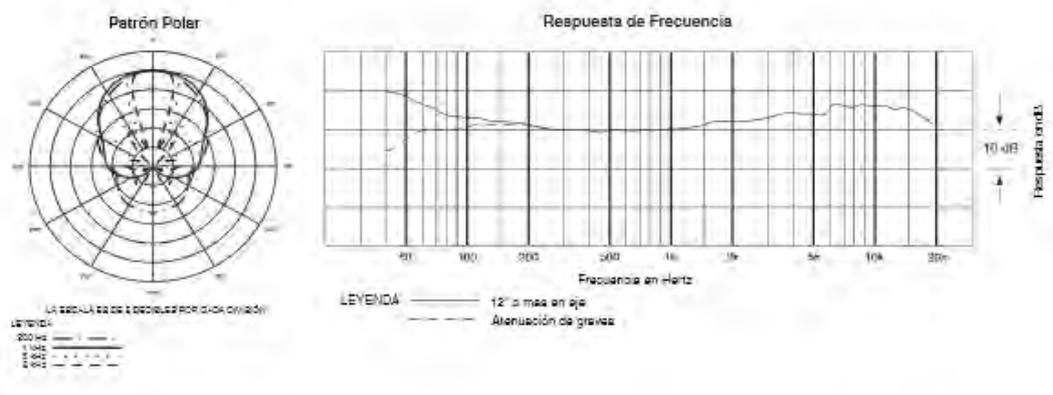


IMAGEN DE PATRON POLAR Y RESPUESTA DE FRECUENCIA DE MICROFONO AT-8015(Tomado de Datasheet [2] AT-8015)

El micrófono AT-835a tiene una frecuencia de respuesta 40-20 kHz, respuesta señal/ruido 70 dB, 1 kHz a 1 Pa (Pascal).

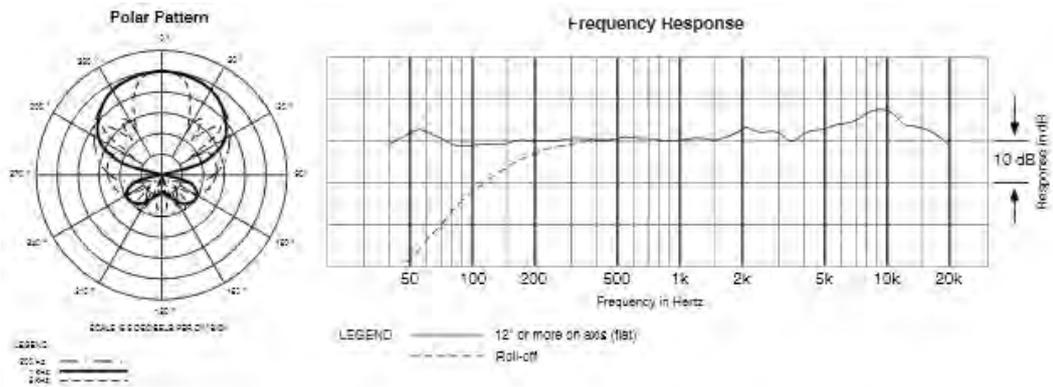


IMAGEN DE PATRON POLAR Y RESPUESTA DE FRECUENCIA DE MICROFONO AT-835a (Tomado de Datasheet [2] AT-835a)

5. Motorreductor DC (Corriente Directa)

El movimiento del prototipo, también consta de un motorreductor el cual soporta el peso de los servomotores y micrófonos. Un motorreductor es un motor de corriente directa que cuenta con engranes para reducir la velocidad y aumentar el torque (fuerza de rotación). El motorreductor utilizado opera a 12v.

6. GPS BT-20 (Sistema de Posicionamiento Global)

Uno de los datos importantes que se toman en estudios de campo es la localización a nivel geográfico, que permite evaluar patrones geográficos sobre distintas características poblacionales. Por lo que se consideró necesaria la implementación de un GPS.

El GPS BT-20 es un dispositivo que opera a 3.3V y se comunica vía bluetooth/PC con lecturas de comunicación serial. La comunicación está establecida bajo el protocolo de comunicación NMEA (National Marine Electronics Association) con lecturas en código ASCII (American Standard Code for Information Interchange).

7. Interfaz H4n Zoom

Las Interfaz de Audio H4n es un dispositivo que permite la captura de audio en la PC con dos micrófonos independientes. La función de esta interfaz es para captura de los sonidos ambientales donde se incluyen los cantos de aves y con ayuda de micrófonos. Este dispositivo opera a 5V o con dos pilas AA de 1.5 V.

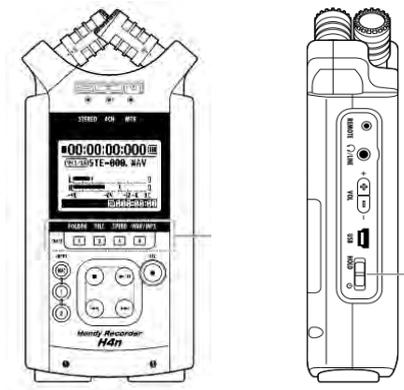


IMAGEN DE INTERFAZ Hn4 Zoom

(Tomado de Datasheet [2] H4n)

8. Arduino UNO

Arduino es una plataforma de cómputo libre [open-source] basado en una tarjeta I/O [entradas y salidas] programada en un ambiente con un lenguaje similar a C llamado *Processing/Wiring*. El objetivo de este micro-

controlador es obtener comunicación y control independiente de cada una de los componentes de MIAA.

Características:

- Micro-controlador ATmega328
- Voltaje de entrada: 7 a 12V
- 14 pines digitales I/O (6 salidas PWM)
- 6 entradas analógicas
- 32k de memoria flash
- Velocidad de reloj 16MHz

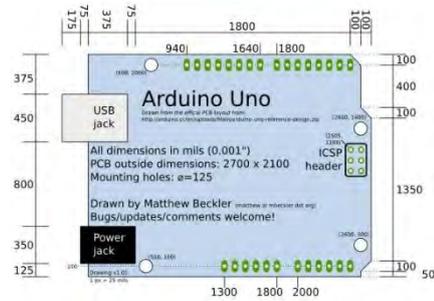


Imagen tomada de www.bricogeek.com

5.2 Construcción de Prototipo

Una vez obtenido el diseño que tomaría el prototipo, se consideró que el control sería más eficaz al ser menos voluminoso, por lo que, el prototipo solo tendría dos movimientos:

- Ascenso y descenso del eje principal (Figura 24)
- Posicionamiento de los micrófonos. (Figura 25)

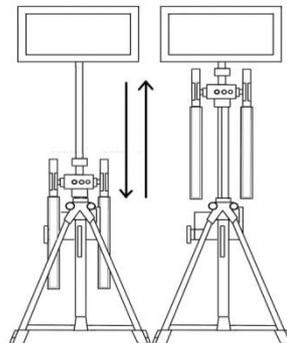


Figura 24. Ascenso y descenso del eje principal de MIAA

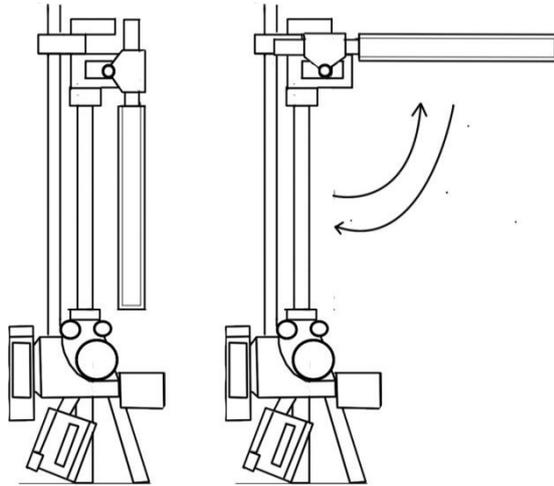


Figura 25. Posicionamiento de micrófonos de MIAA

Se realizaron los cálculos correspondientes para saber qué torque debieran tener como mínimo el motorreductor (la fuerza para la elevación de eje principal) y los servomotores (para el posicionamiento de los micrófonos).

Primero se realizó el análisis de todo el ensamble para saber la distribución de las fuerzas en el eje principal (Figura 26) .

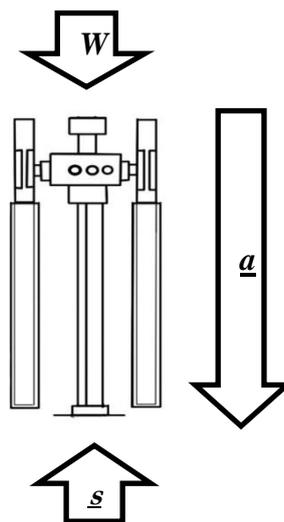


Figura 26. Distribución de peso en el eje del sistema de MIAA

Se pesó la parte soportada por el eje principal y se obtuvo un peso de 0.5404Kg , Se calculó la peso soportado sobre el eje:

$$P = \frac{W}{s}$$

$$P = \frac{0.5404Kg}{1} = 0.5404 Kg$$

Se calculó fuerza ejercida en el eje:

$$F = a * P$$

$$F = 9.81 \frac{m}{s^2} * 0.5404 kg$$

Por lo tanto la fuerza a vencer por el motorreductor es de:

$$F = 5.301324 kg.m/s^2$$

Dónde:

- **P** es el peso soportado (kilogramos fuerza)
- **F** es la fuerza N (Newton)
- **W** es el peso (Kg) que se ejerce,
- **s** es el número de ejes que soportan
- **a** es la aceleración de caída gravitacional (9.81 m/s²).

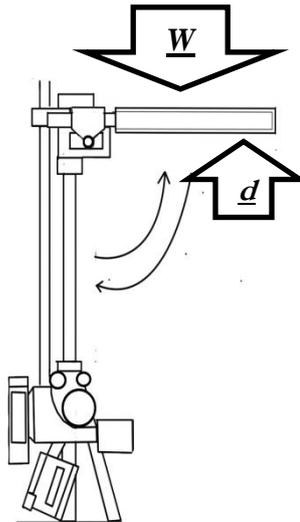


Figura 27. Movimiento de micrófonos con servomotor de MIAA

El cálculo de torque de los servomotores se realizó con la siguiente ecuación:

$$A = W(D)$$

$$A = 0.235Kg (37cm)$$

El torque mínimo de los servomotores es de :

$$A = 8.695 Kg/cm$$

Dónde:

- **A** es el torque mínimo
- **W** el peso (Kg) del micrófono
- **D** es la distancia (cm) del micrófono

También se consideró el cálculo del ángulo de orientación de los micrófonos para obtener una mayor cobertura, ya que la actividad de las aves corresponde según sus hábitos de nidificación, alimentación y cortejo.

La estratificación de la vegetación resulta en una estratificación de los recursos y de las comunidades animales (Smith, 1973). La vegetación puede variar en tres estratos principales, herbáceos, arbustivos y arbóreos, los cuales tienen distintas alturas, siendo éstas desde unos cuantos centímetros como la zona de herbáceas, entre 1 y 5 m los arbustivos y hasta 50 m la zona arbórea (Llata, 2003). Por tanto se tomó en cuenta el grado de ángulo del patrón polar aproximado a los 60° de los micrófonos para encontrar una mayor cobertura en los distintos estratos, siendo 45° en relación al eje horizontal del prototipo la posición adecuada (Figura 3.5).

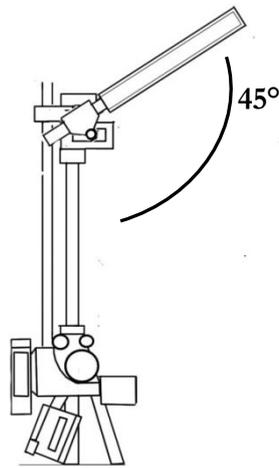


Figura 28. Colocación de micrófono a 45° de MIAA

5.3 Implementación de MIAA

Una vez ensamblado el sistema físicamente se continuo con la implementación, la cual tuvo como objetivo programar cada elemento electrónico para el manejo, colocación y la captura de datos correspondientes. La idea básica del funcionamiento del sistema MIAA operando en campo, se basa en que el investigador lleve a MIAA a campo y la coloque, siendo el siguiente paso el manejo a control remoto con un ordenador (computadora) en el tiempo en el que el investigador requiera llevar a cabo la identificación.

Para concretar las acciones básicas del funcionamiento de MIAA, fue necesario generar un diagrama de flujo con las distintas posibilidades y opciones en su posicionamiento el cual se muestra a continuación.

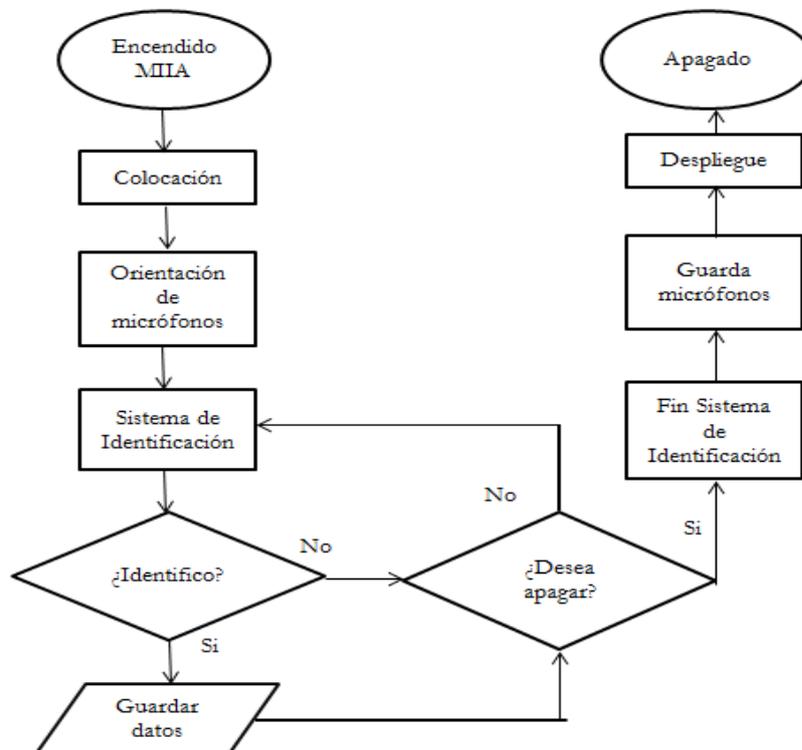


Diagrama 1. Funcionamiento básico de MIAA

6.1. Evaluación

La evaluación del sistema consistió en sacar una serie de indicadores los cuales marcan el desempeño del sistema. Dado que el corpus es reducido y solo consiste con 59 minutos de grabación y el mínimo sugerido aproximadamente es de 250 minutos para poder generar un buen desempeño (Irtza & Hussain, 2013; Pineda et al., 2009), fue necesario establecer la técnica para la evaluación llamada *cross-validation*. Esta técnica se basó en dividir el corpus (las grabaciones) en 10 segmentos iguales aleatoriamente, de los cuales se entrenaron 9 segmentos y el restante se utiliza para comparar el predictor del modelo generado con respecto al *corpus gold estándar*, de tal manera que la iteración se va recorriendo para evaluar cada uno de los 10 segmentos tal y como se muestra en la Figura 29.

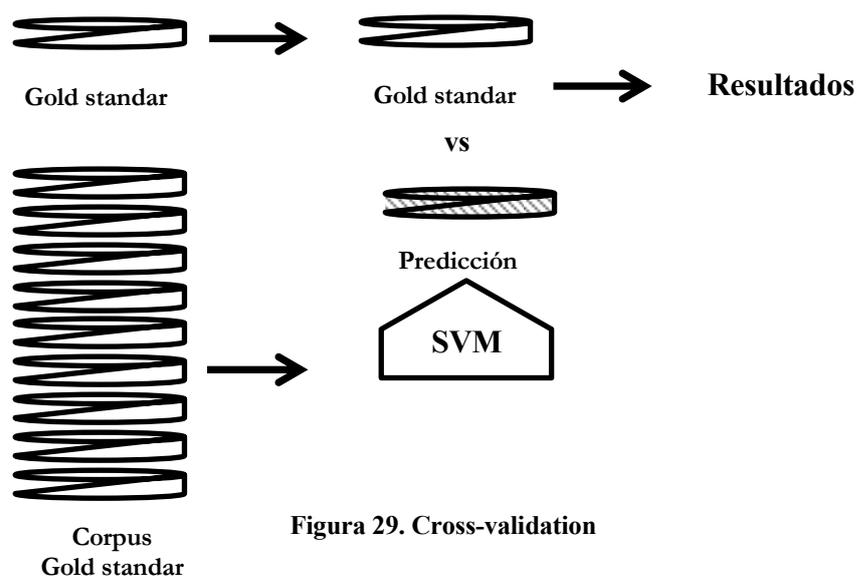


Figura 29. Cross-validation

Cada una de las comparaciones fue hecha con dos técnicas diferentes propuestas por Eyben et al., (2010).

6. 1. 1. Frame-Based

Esta es una comparación entre la etiquetación *gold standard* (A) y la predicción, la cual es segmentada en pequeñas ventanas (B) para verificar los aciertos y fallos en la predicción por cada una de las ventanas. Se consideran las siguientes variantes:

- Verdaderos Positivos (VP): Cuando coinciden las ventanas entre *gold standard* y la predicción.
- Falsos Positivo (FP): Cuando la predicción no coincide con la ventana de *gold standard* .
- Falso Negativo (FN): Cuando no predice lo que está en *gold standar*.
- Verdadero Negativo (FN): Cuando coinciden en que no hay actividad la predicción y *gold standard* .

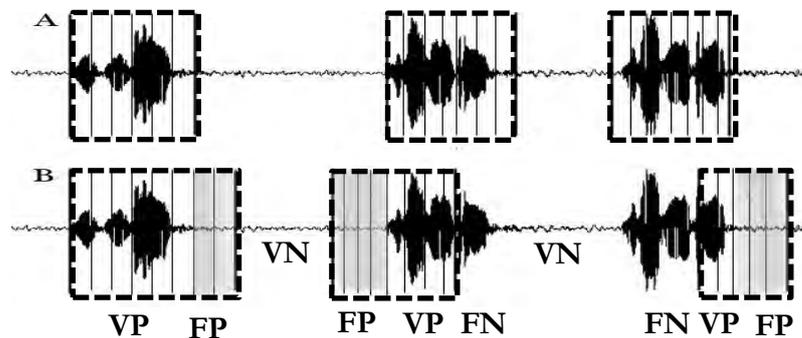


Figura 30. Ejemplo Frame-base

6.1.2. Turn-Based (*middle point* o punto medio).

A diferencia de *Frame-base*, en esta comparación se toma el punto medio de las etiquetas *gold standard* para comparar si coincide el punto medio con alguna parte de la etiqueta en la predicción y se consideran los siguientes casos:

- Verdadero Positivo (VP): Cuando coincide el punto medio de *gold standard* con alguna parte de la predicción.
- Falso Negativo (FN): Cuando el punto medio de *gold standard* no coincide con la predicción y hay actividad.
- Verdadero Negativo (VN): Cuando no hay etiqueta y no predice.
- Falso Positivo (FP): Cuando la predicción no coincide con la ventana de *gold standard*.

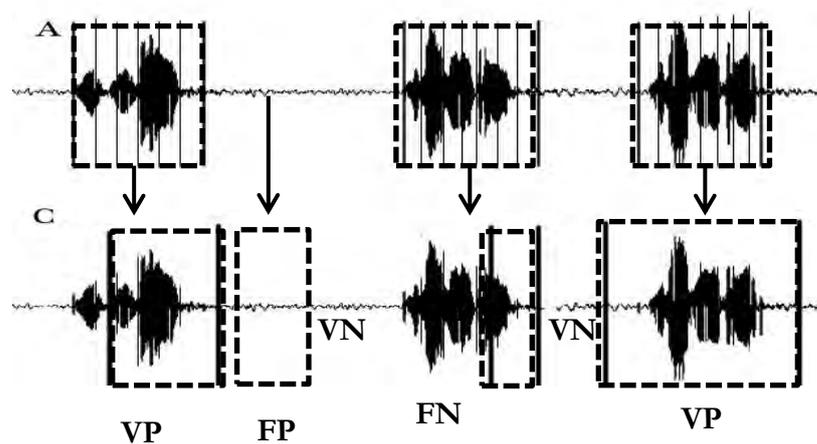


Figura 31. Ejemplo Turn-base (middle point)

6.1.3 Indicadores de Evaluación

Los indicadores para la evaluación y cuantificación del desempeño del sistema basado en probabilidades, fueron escogidos de acuerdo a la propuesta de Sokolova et al., (2006), Powers (2011) y Towsey (2012), siendo este último un estudio de identificación acústica de animales, donde se utilizaron las siguientes definiciones estándar para evaluar el aprendizaje automático (*Machine Learning*):

$$\mathbf{Recall} = \frac{VP}{VP+FN}$$

Recall (Sensibilidad): Es la proporción de casos recuperados que se predijeron correctamente o también se define como la sensibilidad cuya medida es cobertura de los casos reales. Esta medida tiende a no ser muy valorada, ya que en la mayoría de casos se centra en el grado de confianza de un clasificador. Sin embargo, en este trabajo la métrica es de gran relevancia, puesto que para los sistemas de monitoreo es importante el número de especies identificadas para poder calcular el tamaño de la comunidad o población.

$$\mathbf{Precisión} = \frac{VP}{VP+FP}$$

Precisión (Especificidad): Es la confianza que denota la proporción de predicción acertada en la identificación del canto de *Turdus migratorius* o también se define como la medida de exactitud de los aspectos predichos en contraste con la tasa de descubrimiento real.

$$\mathbf{F-Score} = 2 * \frac{\text{precision*recall}}{\text{precision+recall}}$$

F-Score: Es el rendimiento del sistema o también es conocido como la puntuación del sistema. *F-score* permite mezclar recall y precisión para poder comparar el rendimiento contra otros sistemas de aprendizaje automático.

Curva de Características Operativas del Receptor (ROC)

La curva ROC es un gráfico en el que se observa la comparación de dos características operativas de los resultados observados (verdaderos positivos y falsos positivos), de tal manera que la curva ROC nos indica qué tan estricto es el clasificador SVM y las dos tendencias que se pueden esperar del sistema (Hanley & McNeil, 1982; Martínez, 2007). :

- Un sistema muy estricto, el cual tiene mucha precisión en identificación, sin embargo existen hay pérdida de información puesto que hay muchos falsos positivos.
- Un sistema permisivo, indica una buena cobertura (recuperación de datos) con menos falsos positivos, sin embargo la precisión tiende a ser baja.

6.2 Experimentos

Como se vio anteriormente en las evaluaciones del sistema, existe una serie de indicadores los cuales señalan el desempeño del mismo, por lo que se generaron experimentos para probar el sistema en sus distintas etapas y en diferentes condiciones.

6.2.1 Evaluación del silabificador:

El objetivo de esta evaluación fue cuantificar el rendimiento del silabificador, ya que es la primer etapa por la que se captura el canto de *Turdus migratorius*, el cual es segmentado en silabas. La evaluación consistió en dividir el corpus con *cross-validation* y comparar las etiquetas del *gold standard* con las etiquetas predichas por el silabificador, con el fin de comparar el número de coincidencias en la predicción de silabas. Estas comparaciones de coincidencias se realizaron con *frame-base* y *turn-base*.

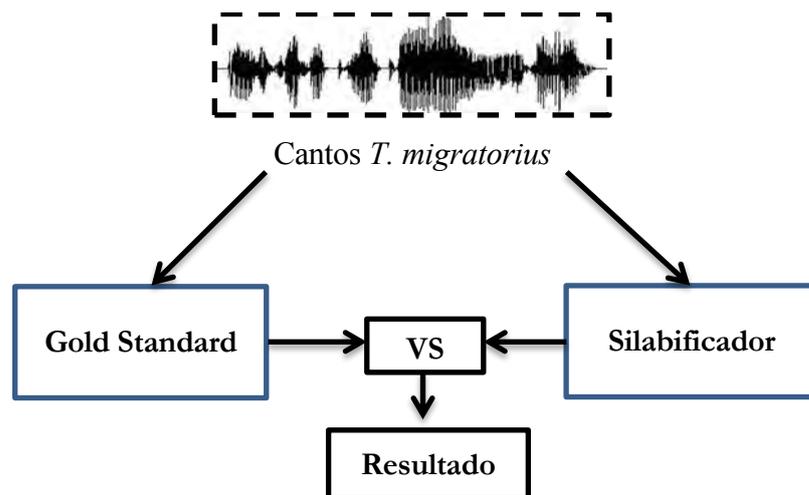


Diagrama 2. Comparación de etiquetación *Gold standard* vs etiquetas silabificador

6.2.2 Evaluación del Clasificador *gold standard* vs *gold standard*

El objetivo de esta evaluación fue cuantificar el rendimiento de un modelo realizado con silabas de *T. migratorius* de la etiquetación *gold standard* y evaluar con silabas de *T. migratorius* de la etiquetación *gold standard*. En la evaluación al corpus se le aplicó cross-validation y se compararon las etiquetas del *gold standard* con las etiquetas *gold standard*, con el fin de comparar el número de coincidencias en la predicción de sílabas. Estas comparaciones de coincidencias se realizaron con *frame-base* y *turn-base*.

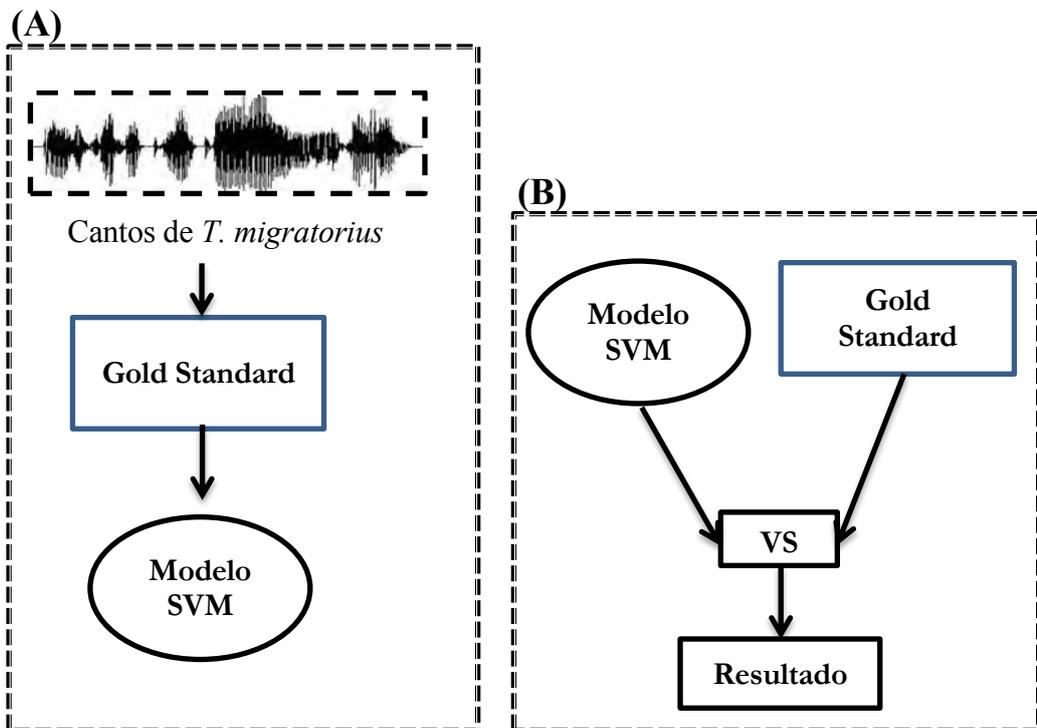


Diagrama 3. (A)Entrenamiento de modelo con silabas de *Gold standard*.
(B)Comparación de modelo *Gold standard* vs *Gold standard*.

6.2.3 Evaluación del Clasificador *gold standard*-ruido vs *gold standard*

El objetivo de esta evaluación fue cuantificar el rendimiento del sistema cuando es entrenado con silabas de *T. migratorius* y ruido de la etiquetación *gold standard*. El modelo resultante es comparado con silabas de *T. migratorius*. La comparación se realizó haciendo *cross-validation* del corpus y se evaluaron las coincidencias con *frame-based* y *turn-based*.

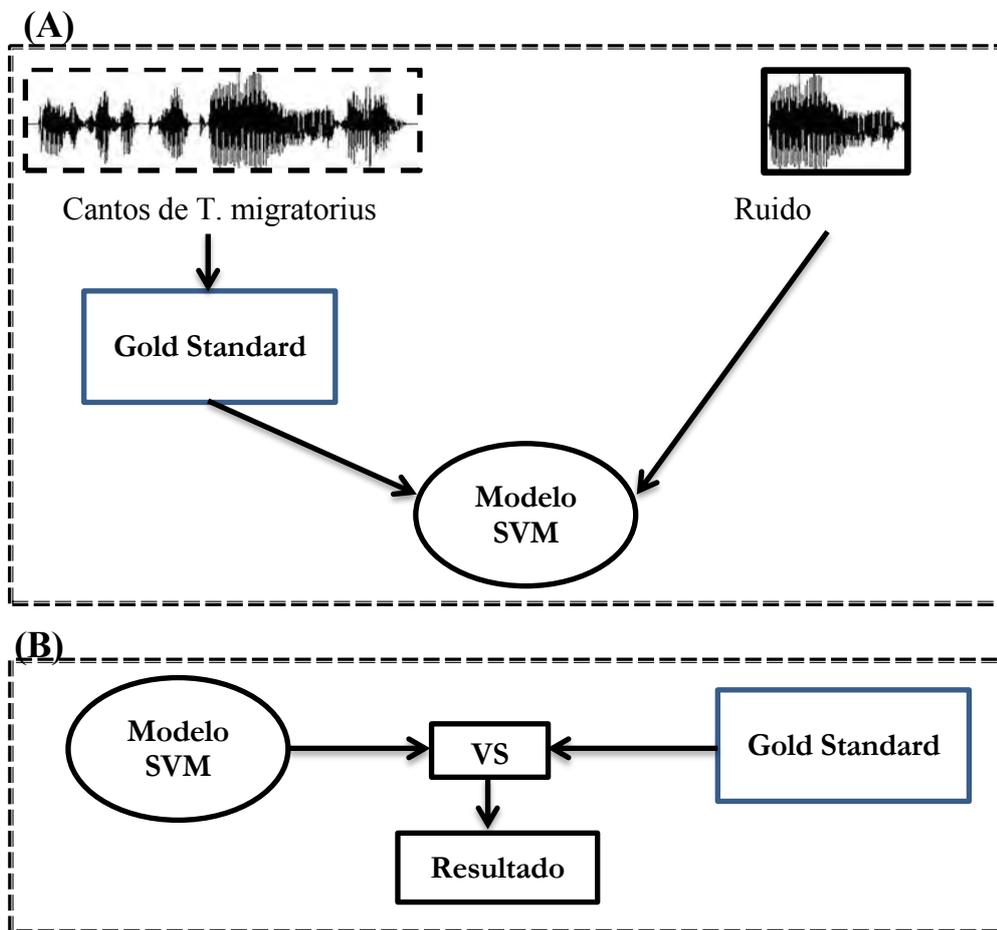


Diagrama 4. (A)Entrenamiento de modelo con silabas de *Gold standard* y ruido.
(B)Comparación de modelo *Gold standard* vs *Gold standard*.

6.2.4 Evaluación del clasificador Silabificador vs *Gold standard*

El objetivo de esta evaluación fue cuantificar el rendimiento del sistema cuando solo es entrenado con sílabas y es comparado con ejemplos positivo. Esta consistió en generar un modelo de clasificación con SVM, incorporando las sílabas obtenidas del corpus con la etiquetación *gold standard* de *Turdus migratorius*. Se compararon las predicciones del silabificador contra las predicciones del clasificador con *cross-validation* y se evaluaron las coincidencias con *frame-based* y *turn-based*.

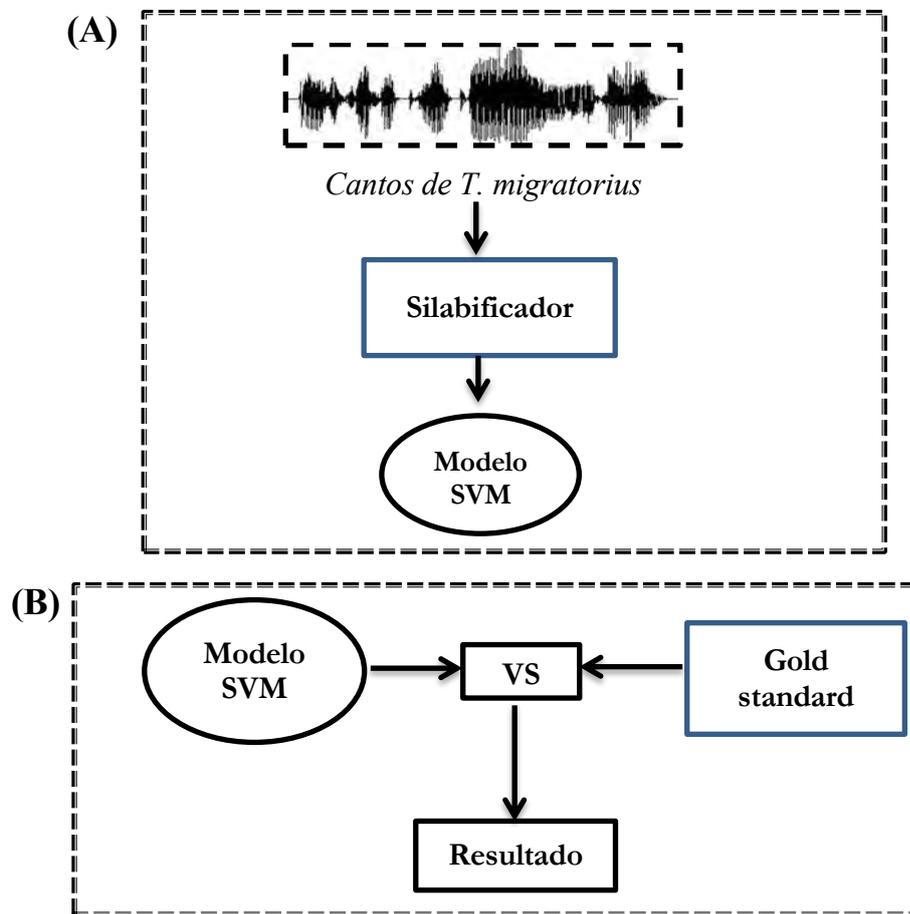


Diagrama 5. (A)Entrenamiento de modelo con sílabas de Silabificador. (B)Comparación de modelo *Gold standard* vs Silabificador.

6.2.5 Evaluación clasificador-Silabificador-Ruido-otras especies vs *gold standard*

El objetivo de esta evaluación fue identificar el rendimiento del sistema, cuando se incorporan silabas de *T. migratorius* y ruido o “no *T. migratorius*” (donde se incluyeron 5 especies). En esta evaluación se generó un modelo de clasificación con las silabas obtenidas del silabificador y los segmentos con ruido. Se compararon las predicciones del clasificador contra las predicciones del *gold standard* con *cross-validation* y se evaluaron las coincidencias con *frame-based* y *turn-based*.

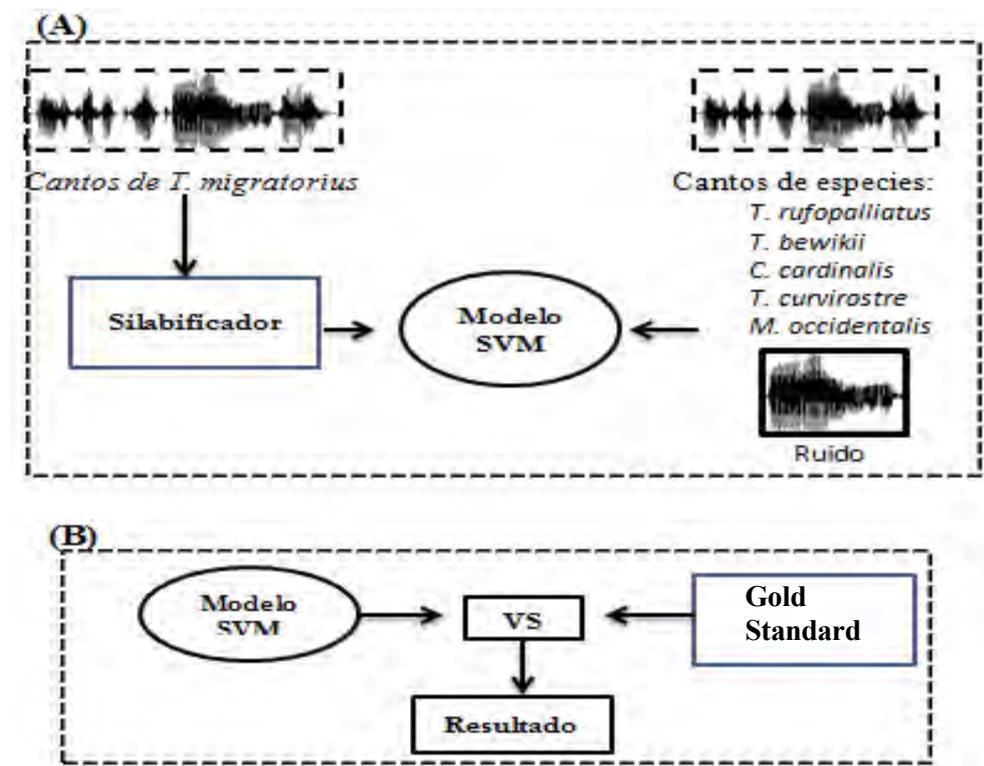


Diagrama 6. (A)Entrenamiento de modelo con silabas de Silabificador, ruido y otras especies. (B)Comparación de modelo vs Silabificador.

6.2.6 Evaluación clasificador-Silabificador-Ruido-otras especies vs Silabificador

El objetivo de esta evaluación fue identificar el rendimiento del sistema, cuando se incorporan sílabas de *T. migratorius* y ruido o “no *T. migratorius*” (donde se incluyeron 5 especies). En esta evaluación se generó un modelo de clasificación con las sílabas obtenidas del silabificador y los segmentos con ruido. Se compararon las predicciones del clasificador contra las predicciones del silabificador con *cross-validation* y se evaluaron las coincidencias con *frame-based* y *turn-based*.

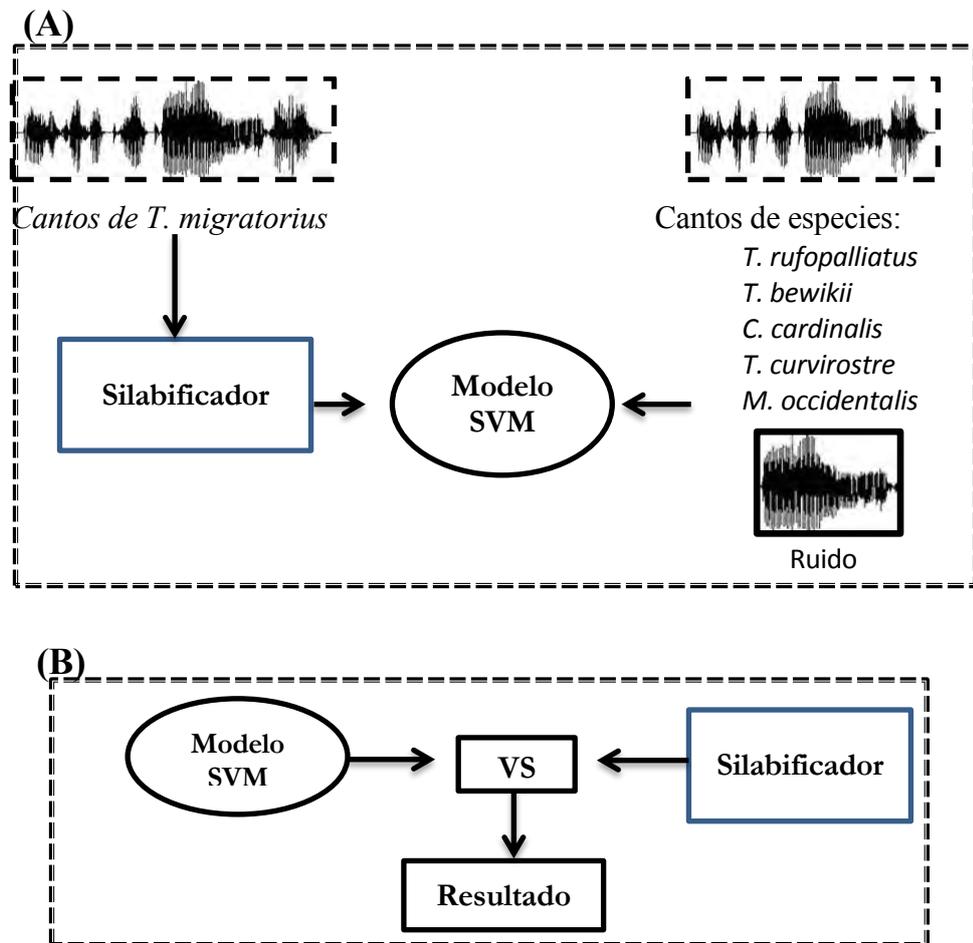


Diagrama 7. (A)Entrenamiento de modelo con sílabas de Silabificador, ruido y otras especies. (B)Comparación de modelo vs Silabificador.

6.2.7 Curva ROC

Con esta prueba diagnóstica de desempeño, el objetivo fue determinar qué composición de características, entre tres propuestas, discrimina mejor a los ejemplos expuestos de *T. migratorius* y No *T. migratorius*. Debido a la necesidad de señalar qué tan “permisivo” o “estricto” es el sistema, fue necesario comparar los resultados con rangos propuestos del área bajo la curva (ABC). Se entiende como Área Bajo la Curva (ABC) a la probabilidad de clasificar correctamente entre ejemplos de “*T. migratorius*” y “No *T. migratorius*” seleccionados al azar. Los valores del ABC de la curva ROC van entre 0.5 (igual o al azar), siendo el valor máximo 1 (caso ideal). Sin embargo, suele ser aceptable un valor de discriminación cuando supera el 0.7, como se muestra en la siguiente gráfica (Hanley & McNeil, 1982; Martínez, 2007).

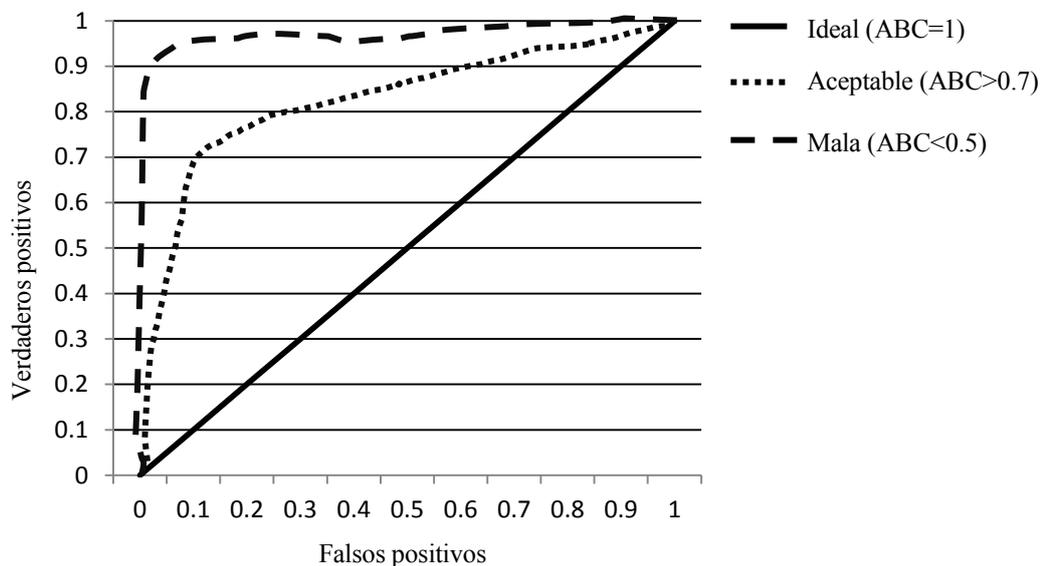


Figura 32. Representación de las curvas ROC con los distintos valores guía.

Con esta evaluación se diagnosticó el desempeño de clasificación de los tres modelos propuestos con distintas características y combinación de éstas, para encontrar el más eficiente en la tarea de clasificación. En los siguientes puntos se muestran los tres modelos propuestos:

- 1) Modelo generado con extracción de características MFCCs
- 2) Modelo generado con extracción de características Δ MFCCs
- 3) Modelo generado con extracción de características MFCCs y Δ MFCCs (o también llamado Both).

Se compararon las predicciones del clasificador contra las predicciones del silabificador con *cross-validation* y la evaluación de coincidencias se realizó con *turn-based*.

6.2.8 Curva de aprendizaje

El objetivo de la curva de aprendizaje es cuantificar el número de minutos/sílabas necesarios para generar un rendimiento aceptable (*f1-score*) basado en la acumulación de tiempo, para predecir mejoras de productividad en el sistema de aprendizaje automático propuesto para el dispositivo robótico MIAA.

De acuerdo al contexto anteriormente explicado, se agregaron 5 minutos/sílaba de grabaciones de *T. migratorius* (incluyendo ruido y otras aves) con el silabificador para generar un modelo en cada iteración. Se comparó cada iteración con las etiquetas del silabificador y se evaluaron las predicciones con *turn-based*, cuantificando el rendimiento de cada iteración con *f1-score*.

Resultados y Análisis de Resultados

7. Recursos para el corpus

En este trabajo se contaron con recursos de grabaciones focales de campo obtenidas de Cornell Laboratory Of Ornithology Maculay Library, con las que se conformó el corpus de *Turdus migratorius*, *Turdus rufopalliatus*, *Myadestes occidentalis*, *Thryomanes bewickii*, *Toxostoma curvirostre* y *Cardinalis cardinalis*.

Especie	Tiempo	Grabaciones(.wav)
<i>Turdus migratorius</i>	60 min	59
<i>Turdus rufopalliatus</i>	16 min	11
<i>Myadestes occidentalis</i>	36 min	4
<i>Thryomanes bewickii</i>	24 min	3
<i>Toxostoma curvirostre</i>	36 min	5
<i>Cardinalis cardinalis</i>	21 min	4
Total	4 h 1 min	86

Tabla 2. Descripción de recursos de audio

En el caso de las grabaciones de *T. migratorius*, sirvieron para generar una base y a partir de ahí el modelo acústico estadístico para su identificación. Las demás grabaciones de las especies restantes sirvieron para complementar y robustecer el modelo de identificación, el cual se verifica en las experimentaciones 7.4 y 7.5 (ver más adelante).

7.1 Gold standard vs Silabificador

El corpus de *T. migratorius* estuvo compuesto por 59 grabaciones, del cual se obtuvo un total 2,671 sílabas con la etiquetación hecha por un experto (en inglés *gold standard*), es decir, que en promedio de obtuvieron 45 sílabas etiquetadas manualmente por grabación. Cabe mencionar que son aproximados los números de sílabas por grabación, ya que no se obtuvo control sobre el canto de cada individuo de *T. migratorius*, puesto que el canto es consecuencia de actividad conductual. En el procesamiento del corpus con el silabificador se obtuvieron 3,060 sílabas, de las cuales hay 389 sílabas de más que representan el 11.45% por encima del total *gold standard*. Sin embargo la evaluación de coincidencias explica de manera más objetiva esta diferencia, como se muestra en la siguiente tabla:

	Turn-based	Frame-based
Precisión	87.49 %	52.47%
Recall	75.15%	77.32%
F1	78.30%	60.97%

Tabla. 3. Resultados de evaluación en la etiquetación *Gold standard* vs Silabificador.

Como se puede apreciar en los resultados de precisión, existe una gran diferencia entre la evaluación *turn-based* y *frame-based*, esto se debe a que en *turn-based* es a nivel silábico y existe una mayor probabilidad de coincidir con la predicción, para contar los verdaderos positivos, independientemente de la longitud y/o colocación de la etiquetación en la predicción, mientras que *frame-based* se comportó como una evaluación más rigurosa al nivel de señal y de precisión, ya que tienen que coincidir

cada una de las ventanas para poder tener verdaderos positivos y así tener una buena precisión. Visto de otra manera, en la evaluación *frame-based* existe una mayor cantidad de falsos positivos y/o falsos negativos. Por otra parte en ambas evaluaciones aproximadamente se recuperan (Recall) la misma cantidad de sílabas y el rendimiento del *recall* depende más de la estructura y composición del algoritmo silabificador. Los resultados de F1-score favorecen la evaluación *turn-based*, principalmente por la explicación antes mencionada, la cual se encuentra relacionada con la precisión al nivel sílaba. Sin embargo esto no significa que *turn-based* es mejor que *frame-based*, ya que ambas evaluaciones reflejan distintas perspectivas del comportamiento del silabificador.

7.2 Evaluación del Clasificador-*gold standard* vs *Gold standard*

El objetivo de esta evaluación fue cuantificar el rendimiento de un modelo realizado con sílabas de *T. migratorius* de la etiquetación *gold standard* y evaluar con sílabas de *T. migratorius* de la etiquetación *gold standard*. En esta evaluación se entrenó un modelo de clasificación con las sílabas *gold standard*, los resultados son los siguientes:

	Turn-based	Frame-based
Precisión	100 %	100%
Recall	100%	100%
F1	100%	100%

Tabla 4. Resultados de evaluación del clasificador- *Gold standard* vs *Gold standard*

Los resultados de la tabla 7.2 muestran el comportamiento de ambas evaluaciones de manera ideal (caso trivial), ya que supondría un ambiente en el que únicamente canta *T. migratorius*. Por otra parte en términos de requerimientos las máquinas de soporte de vectores (SVM), basan su entrenamiento en la maximización del margen existente entre el hiperplano separador de las instancias de dos o más clases (Deco, 2012). Sin embargo estos resultados a pesar de ser triviales, resguardan una gran importancia, ya que en primera instancia nos indican que el clasificador SVM logra detectar el canto de *T. migratorius* sin ninguna otra variable.

7.3 Evaluación del clasificador Silabificador vs Gold standard

El objetivo de esta evaluación fue cuantificar el rendimiento del sistema cuando es entrenado con sílabas de *T. migratorius* y ruido de la etiquetación *gold standard*, con el modelo resultante se evaluaron las coincidencias con la etiquetación *gold standard*. Los resultados son los siguientes:

	Turn-based	Frame-based
Precisión	73.94 %	45.94%
Recall	64.64%	68.34%
F1	67.34%	53.71%

Tabla 5. Resultados de evaluación en la etiquetación Silabificador vs Gold standard

El desempeño del sistema en *f1-score* corresponde directamente al tipo de entrenamiento realizado con el silabificador, por lo tanto las predicciones fueron realizadas basadas en el silabificador (donde se pierde certeza), es decir, la comparación de coincidencias entre el clasificador-silabificador contra el *gold standard* presentan la misma problemática de la evaluación del silabificador vs *gold standard*, donde el tamaño de la de la etiqueta de predicción no corresponde directamente a la etiquetación *gold standard* y hay una gran cantidad de falsos positivos y/o falsos negativos.

7.4 Evaluación clasificador-Silabificador-Ruido-otras especies vs Silabificador.

El objetivo de esta evaluación fue cuantificar el rendimiento del sistema cuando solo es entrenado con sílabas y es comparado con ejemplos positivos. En esta evaluación se entrenó con un modelo de clasificación incorporando el corpus con el silabificador, segmentos de ruido y los cantos de otras aves. Se comparó y se evaluaron las coincidencias con la etiquetación del silabificador, los resultados son los siguientes:

	Turn-based	Frame-based
Precisión	83.26 %	87.30%
Recall	83.26 %	87.30%
F1	83.26 %	87.30%

Tabla. 6. Resultados de evaluación del clasificador Silabificador-ruido-otras aves vs Silabificador

Estos resultados nos indican que el modelo identifica 8 de cada 10 silabas correctamente con una precisión del 83%. Existe una consistencia en los resultados de *f1-score* ya que las etiquetas generadas y comparadas pertenecen al silabificador (son constantes positivos), por lo que no existe variabilidad en el tamaño de las etiquetas y en la comparación coinciden completamente, sin embargo a pesar de que en ambos resultados de *f1-score* se encuentran por encima del 80% y retomando el resultado de la evaluación “silabificador vs *gold standard*”, recordaremos que el silabificador no tiene una precisión de 100%. Por lo que al 83% de esta evolución se le tendría que restar el valor obtenido en la prueba *gold standard* vs silabificador (78.30% de precisión).

7.5 Evaluación clasificador-Silabificador-Ruido-otras especies vs *gold standard*

El objetivo de esta evaluación es identificar el rendimiento del sistema, cuando se incorporan silabas de *T. migratorius* y ruido (donde se incluyen las especies restantes del *corpus*).

En esta evaluación se entrenó con un modelo de clasificación incorporando el corpus con el silabificador, segmentos de ruido y los cantos de otras aves. Se comparó y se evaluaron las coincidencias con la etiquetación *gold standard*, los resultados son los siguientes:

	Turn-based	Frame-based
Precisión	87.18 %	53.87%
Recall	75.63%	81.92%
F1	78.74%	63.41%

Tabla. 7. Resultados de evaluación del clasificador Silabificador-ruido-otras aves vs *gold standard*

El desempeño del sistema en *f1-score* muestra una mejora en comparación con las evaluaciones pasadas, ya que la incorporación de ruido y los cantos de otras aves delimitan de mejor manera las clases “*Turdus migratorius*” y “No *Turdus migratorius*” en el modelo de SVM y se aproxima de mejor manera a la dinámica acústica de una ambiente natural. Sin embargo el rendimiento general sigue siendo afectado como se mencionó anteriormente por las diferencias entre la etiquetación del silabificador y *gold standard*, donde el tamaño de las etiquetas no coinciden totalmente.

7.6 Evaluación del Clasificador-gold standard -ruido vs Gold standard

El objetivo de esta evaluación es identificar el rendimiento del sistema, cuando se incorporan silabas hechas por el experto de *T. migratorius* y ruido (donde se incluyen las especies restantes del *corpus*). De tal manera que este sistema supondría el máximo desempeño a tener, los resultados son los siguientes:

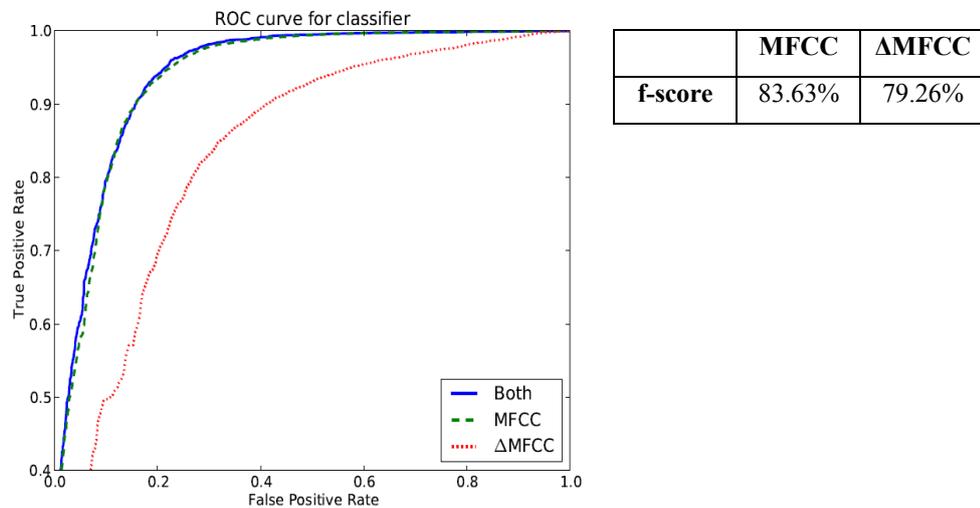
	Turn-based	Frame-based
Precisión	91.54 %	89.07%
Recall	95.29%	94.37%
F1	92.85%	90.91%

Tabla. 8. Resultados de la evaluación del clasificador- Gold standard -ruido vs *Gold standard*.

Observamos que los resultados de ambas métricas son muy similares ya que la comparación entre silabas es consistente por la etiquetación *gold standard* del corpus para el entrenamiento y la evaluación, sin embargo la caída del sistema del 7.15% y 9.09% en *f1-score*, refleja que la incorporación de ruido afecta directamente el desempeño del sistema, ya que el ruido no se puede definir, debido a que este es infinitamente variable. Por otra parte estos resultados reflejan directamente el máximo desempeño que podría tener el sistema, suponiendo que el silabificador se comportara de manera ideal teniendo un rendimiento como el que es realizado por un experto. También sugiere que el entrenamiento del sistema con SVM tiene un límite de aprendizaje, y que en un futuro es necesario generar nuevas estrategias para tener un mejor desempeño.

7.7 Curva ROC

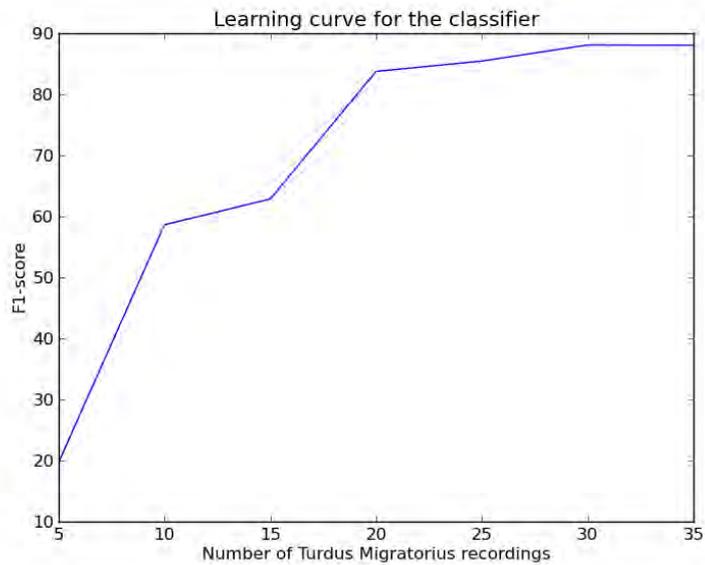
Los resultados de esta evaluación diagnóstica muestran un claro comportamiento en la clasificación de las muestras azarosas de audio de *T. migratorius* y No *T. migratorius*. Como se puede apreciar en la gráfica 6.7, el desempeño de la curva ROC tanto en el modelo de MFCC, como en el modelo *both* (combinación de MFCC y Δ MFCC) muestran un desempeño muy similar, teniendo estos valores del ABC por encima del 0.7 y siendo mucho más estrictos en comparación con la curva ROC del modelo Δ MFCC, siendo este último el más permisivo, ya que tiene un valor del ABC muy cercano del mínimo sugerido. Cabe señalar que el modelo propuesto con las características MFCC es el que tiene mayor estabilidad y relevancia, por lo que esté se utilizó para alimentar el sistema MIAA. Estos resultados concuerdan con el desempeño del sistema, utilizando la métrica f-score antes propuesta donde el modelo MFCC tiene un mayor rendimiento.



Gráfica. 1. A) Curvas ROC de los 3 modelos propuestos, B) f-score de cada modelo.

7.8 Curva de aprendizaje

La curva de aprendizaje muestra que el mínimo necesario para poder entrenar el sistema con un f-score del 83.67%, es de aproximadamente 40 minutos, que son equivalentes a aprox. 1810 sílabas. Merece mencionarse que estos resultados solo son aplicables para la complejidad y estructura acústica del canto de *T. migratorius*, el cual cuenta con 220 sílabas según Peshek & Blumstein (2011). Por lo antes mencionado y para trabajos a futuro, cabe señalar que la cantidad de minutos para poder obtener el mismo rendimiento dependerá tanto de la cantidad de sílabas, así como de la complejidad acústica del canto de cualquier especie.



Gráfica. 2 Curva de acumulación.

En términos generales con las evaluaciones aquí presentadas, se buscó la mejor metodología para diseñar un modelo capaz de identificar a *Turdus migratorius* en su hábitat natural, por lo que se consideraron varios escenarios desde los más “triviales” donde solo se tiene el canto de *T. migratorius* hasta los que asemejan acústicamente la gran variedad y composición de cantos y ruidos de un hábitat natural.

Discusión y Conclusiones

Hasta este punto y de acuerdo a los resultados, se corrobora en general que es posible generar un sistema computacional capaz de identificar de manera automática las sílabas pertenecientes al canto de *T. migratorius* y cualquier otra especie que cumpla con patrones de repetición en su canto. Además cabe señalar que el rendimiento del sistema no es posible compararlo con otros sistemas de manera directa y objetiva, ya que en la presentación de resultados en casi la totalidad de trabajos se reportan solo los valores de precisión (accuracy), como en los trabajos de Kogan y Margolias (1997), Kigan (1997) y Lee et al., (2006), solo por mencionar algunos, los cuales si bien nos indican la certeza que se tiene en la identificación de un fragmento del canto (notas y/o sílabas), no demuestran la cantidad de ellos recuperados (*recall*). En el mismo sentido, con el trabajo de Towsey et al. (2012), donde se utilizó la misma métrica utilizada en este trabajo, no es posible comparar los resultados ya que sugerimos que el desempeño del sistema dependerá de complejidad, estructura y características que presente el canto.

Retomando la idea principal con fines conservacionistas y el objetivo que aborda este trabajo, resaltamos la importancia de las métricas aquí utilizadas de precisión y recuperación (*recall*), ya que estas últimas dos tienen un papel importante en la futura investigación para el apoyo de la conservación de la biodiversidad, puesto que precisamos cuántas sílabas se logran recuperar y establecemos la precisión de cada una de ellas, lo que sugiere plantear en un futuro no solo la identificación del canto de un ave sino también a saber cuántos individuos hay.

Por lo antes mencionado nuestra discusión se centra en el desempeño general del sistema, así como en sus eficiencias y deficiencias que presenta, y de cómo dar una posible solución a estas últimas.

8.1 Discusión

Los resultados del sistema claramente se ven afectados por el desempeño del silabificador, el cual no logra hacer la captura y una segmentación adecuada de cada sílaba, ya que en muchos casos la segmentación resultaba desfasada (siendo más grande o más chica). Por lo tanto este aspecto nos obliga a establecer nuevas estrategias de mayor precisión en el manejo de captura en señales digitales.

Por otra parte, el desempeño del clasificador SVM nos sugiere que podemos incrementar la cantidad de características acústicas para representar cada una de la sílaba del canto y así poder representarlas y describirlas acústicamente, situación alentadora, ya que podría existir una mejora en el desempeño del modelador incluyendo algunas otras métricas en trabajos como los de Umberto et al., (1982), Deecke et al., (1999), Bardeli et al., (2010) y Neal et al., (2011) quienes realizan representaciones del tiempo-frecuencia, características básicas como frecuencia específica, duración y pitch del canto, también conocido como frecuencia fundamental (Potanmintis y Ganchev, 2008) son algunas de las sugerencias que podrían implementarse para tener un mejor desempeño. Existen otras metodologías en la representación de la señal de los cantos, las cuales implican la utilización gráfica de la señal de audio. Dichas metodologías como las de Anderson et al., (1996) y Lee et al, (2013) quienes a partir de representaciones gráficas, logran realizar correlaciones de los espectrogramas, sugieren una excelente metodología híbrida, ya que existiría una complementación acústicas y gráfica para corroborar la identidad de los individuos.

Prototipo MIAA

En el presente trabajo se concluyó la construcción del dispositivo robótico MIAA, el cual funcionó satisfactoriamente en las tareas básicas, tanto de identificación, así como en el despliegue de movimientos para su colocación. Sin embargo las complicaciones que se tuvieron en el proceso de prueba fueron relacionadas al consumo de energía, ya que el panel solar no alcanzaba a cubrir todas las necesidades para un funcionamiento correcto, por lo que se establecieron otras estrategias para abastecer de manera adecuada la energía. Otra observación realizada en el periodo de pruebas, fue la capacidad de captura de audio de los micrófonos, la cual evidentemente por el tipo de micrófonos (direccionales) solo logra capturar el sonido cuando se apunta directamente a la fuente emisora de sonido, lo que implica que si una ave se encuentra cantando fuera del área de captura, podría haber una pérdida de información. Ante lo mencionado, las complicaciones no representan un factor limitante, ya que ante la gran variedad de sistemas abastecedores de energía en el mercado, es fácil solucionar dicha problemática. Con respecto a los micrófonos, un estudio realizado por Rempel et al., (2005) propone la utilización de micrófonos omnidireccionales para la grabación de aves en su ecosistema, ya que los micrófonos omnidireccionales tiene una mayor cobertura de captura con respecto a los direccionales.

Conclusión

En el trabajo actual se ha desarrollado un sistema capaz de identificar al ave *Turdus migratorius* a partir de las sílabas de su canto, obteniendo 8 de cada 10 sílabas correctamente con una precisión del 83%. Así mismo se diseñó y construyó un dispositivo robótico (MIAA) el cual pretende acercar las tareas de monitoreo faunístico al avance tecnológico para facilitar, tener mayor cobertura y precisión en dichas tareas, como apoyo a la investigación y conservación de fauna. El dispositivo MIAA puede llegar a cubrir distintos objetivos en campo como ayuda a la investigación, de tal manera y aventurándonos a trabajos a futuros, éste tipo de dispositivos podrían utilizarse y desplegarse en el terreno formando transectos y/o cuadrantes para detectar la actividad y flujo de especies de aves en tiempo real, guardando información valiosa para generar modelos de la dinámica de las poblaciones de aves en tiempo y espacio. Por lo que con este tipo de dispositivos se podrían saber aspectos de la biología de las aves con mayor precisión y así establecer manejos para dar prioridad a áreas de anidación, reproducción y resguardo de aves en su hábitat natural. Cabe mencionar que este tipo de dispositivos no solo se restringe al grupo de aves, puesto que varios grupo de animales vertebrados terrestres e insectos (anuros, quirópteros, ortópteros, etc.) tienen cantos y llamados característicos de las especies, que se podrían incluirse en la detección para su monitoreo y dándole un valor agregado muy importante a este tipo de dispositivos. Por último cabe señalar que este tipo de dispositivos no pretenden suplantar el trabajo del biólogo, si no que ayudarían a complementar las tareas e investigaciones para el uso y conocimiento de la biodiversidad, ya que la cantidad de recursos humanos con los que se cuenta actualmente no cubren todas las necesidades y espacios por estudiar. Lo anterior se ha reflejado en las siguientes contribuciones.

Contribuciones

Artículo

- Villarreal O. E.T., Rascón C., Meza R. I. V. (2013). Emotion based features of bird singing for *Turdus migratorius* identification. In: Castro F., Gelbukh A., González M. (Eds.) *Advances in Soft Computing and Its Applications. MICAI 2013*, Springer-Verlag Berlin Heidelberg. Part II, pp.521-530.

Abstract

- Villarreal, T., Meza, I. and Rascón, C. (2013). Identificación del canto del *Turdus migratorius* utilizando un modelo acústico, estadístico y computacional. XII Congreso para el estudio y conservación de las aves en México (CECAM), pp. 80.

Presentaciones

- Villarreal Olvera E.T., Meza Ruiz I. y Rascón Estabané C., “Identificación del canto de *Turdus migratorius* utilizando un modelo acústico estadístico y computacional”. *XII Congreso para el estudio y conservación de las aves en México*, 2013.
- Villarreal Olvera E.T., Meza Ruiz I. y Rascón Estabané C., “Emotion based features of bird singing for *Turdus migratorius* identification”. *12th Mexican International Conference, MICAI 2013*.

Hardware

- Dispositivo robótico MIAA



Figura 32. Módulo de Identificación Acústica de Aves (MIAA). Izquierda: vista frontal; Centro: Vista lateral y Derecha: Vista lateral

Software

Código abierto

- <http://golem.iimas.unam.mx/git?p=golemaudio.git;a=summary>

Adda-Decker M. (2001). Towards multilingual interoperability in automatic speech recognition. *Speech communication* 35: 5-20

Abousleiman R., Qu G., Rawashdeh O. (2013). North Atlantic Right whale contact call detection. In: Glotin H., Clark C., LeCun Y., Dugan P., Halkias X., Sueur J., (Eds.). *The 1st International Workshop on Machine Learning for Bioacoustics*. Atlanta, USA. Pp. 27-32

Agranat I. (2012). Classification algorithms for species identification in noisy environments. *Bioacoustics: The International Journal of Animal Sound and its Recording*. 21:1, 25.

Alldredge, M. W., T. R. Simson, and K. H. Pollock (2007). Factors affecting aural detections of songbirds. *Ecological Applications* 17: 948–955.

Anderson, S.E., Amish S. D., Margoliash D. (1996). Template-based automatic recognition of birdsong syllables from continuous recording. *Acoustical Society of America* 100 (2), Pt. 1.

Bapista, L. F. Y D. E. Kroodsma (2001). Avian Bioacoustics. A Tribute to Luis Baptista. Pp. 11-52 In del Hoyo, J., Elliot, A. & Sargarta, J. (Eds). *Handbook of the Birds of the World. Vol 6. Mousebirds to Hornbills*. Lynx Edicions, Barcelona.

Bardeli R., D. Wolff, F. Kurth, M. Koch, K. H. Tauchert, and K.-H. Frommolt (2010). “Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring,” *Pattern Recognition Letters*, vol. 31, no. 12, pp. 1524–1534. 174, 175, 182

Bart, J. (1985). Causes of recording errors in singing bird surveys. *Wilson Bulletin* 97: 161–172.

Berthold, P., Nowak, E. & Querner, U. (1995). Satelliten-Telemetrie eines Zugvogels von Mitteleuropa bin in das südafrikanische Winterquartier eine Fallstudie am Weis-storch (*Ciconia ciconia*). *J.Orn.* 136: 73–76.

Berlanga, H., Brambila, J., Castillejos, E., Cruz, M., Díaz, D., Escobar, M., Garza, A., Gómez de Silva, H., Grosselet, M., Hinojosa, O., Oliveras, A., Mellink, E., MacKinnon, B., Montejo, J., Ortiz-Pulido, R., Pérez, M.,

Riojas, M., Rodríguez-Contreras, V., Sánchez-Gonzalez, L. (2007). Taller: “Revisión y recategorización de las AICAS de México” Mineral del Chico, Hidalgo. En: Pagina de la red de Conocimientos sobre las Aves de México (AVESMX). CONABIO/NABCI, Birdlife Intl.

Bibby, C.J., N.D. Burgess, D.A. Hill and S.H. Mustoe. (2000). *Bird Census Techniques*. 2nd Ed. Academic Press, London, UK.

Bogdan Vlasenko, Bjrn Schuller, Andreas Wendemuth, and Gerhard Rigoll (2007). Frame vs. turn-level: Emotion recognition from speech considering static and dynamic processing. In AnaC.R. Paiva, Rui Prada, and RosalindW. Pi- card, editors, *Active Computing and Intelligent Interaction, volume 4738 of Lecture Notes in Computer Science*, pages 139-147.

Böjrn Schuller (2011). Voice and speech analysis in search of states and traits. In: Theo Gevers Albert Ali Salah, editor, *Computer Analysis of Human Behavior*, pages 227:253.

Blumstein, D. J. Mennill, P. Clemins, L. Girod, K. Yao, G. Patricelli, J. L. Deppe, A. H. Krakauer, C. Clark, K. A. Cortopassi, S. F. Hanser, B. McCowan, A. M. Ali, and A. N. G. Kirschel (2011). Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus,” *Journal of Applied Ecology*, vol. 48, no. 3, pp. 758–767.

Brenowitz E.A. (2002). Birdsong: Integrating physics, physiology, and behavior. *Journal Comp Physiol. Springer-Verlag*. 188: 827-828.

Cadahía, L., Urios, V. & Negro, J.J. (2005). Survival and movements of satellite tracked Bonelli’s Eagles during their first winter. *Ibis* 147: 415–419.

Camstra F., Vinciarelli A. (2007). Machine learning for audio, image and video analysis. *Springer*.

Cano P., Batlle E., Kalker T. & Haitsma J. (2005). A Review of Audio Fingerprinting. *Journal of VLSI Signal Processing* 41, 271–284.

Carrión I. A., (1998). *Diseño acústico de espacios arquitectónicos*. Edición de la Universidad Politécnica de Cataluña. Barcelona, España.

Catchpole, C.K. (2004). Syllable repertoire and the size of the song control system in captive canaries (*Serinus canaria*). *J. Neurobiol.* 60: 21-27

Catchpole C.K., Slater P.J.B. (2008). *Bird Song: Biological themes and variations. Second edition.* Cambridge University Press. New York USA.

Caycedo R., Ruiz M., Orozco A. (2013). Reconocimiento automatizado de señales bioacústicas: Una revisión de métodos y aplicaciones. *Ingeniería y ciencia*. Vol. 9, no. 18, pp. 171-195.

Cuevas E., Zaldivar D., Cisneros M. (2010). *Procesamiento Digital de Imágenes usando Matlab & Simulink*. Grupo Alfaomega. México.

Botero, A., Boogert J., Vehrencamp S., Lovette I. (2009). Climatic patterns predict the elaboration of song displays in songbirds. *Current Biology*, Vol. 19, no. 13, pp. 1151-1155.

Burnie, D. (2003). *Animal*. Dorling Kindersley. Londres.

Celis-Murillo A., Deppe, J. L. and M. F. Allen. (2009). Using soundscape recordings to estimate species abundance, richness and composition. *J. Field Ornithol.* Vol. 80, no.1. pp. 64–78.

Cunningham Michel A. & Baker Mayron C. (1987). Microgeographic song variation in the nuttalla's white-Crowned sparrow. Department of Zoology, Colorado State University. *The Cooper Ornithological Society, The Condor*. Vol. 89, pp. 261-275.

Christiansen Per. (2009). *Birds*. Kieron Connolly Madrid. España. pp. 8-13

Cyr, A. 1981. Limitation and variability in hearing ability in censusing birds. *Studies in Avian Biology*. Vol. 6, pp. 327–333.

Davis, S., Mermelstein P. (1980). Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. In *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 28, no. 4, pp. 357-366.

Davis, D. E. & R. L. Winstead. (1987). Estimación de tamaños de poblaciones de vida silvestre. En: T. R. Rodríguez (Eds.). *Manual de Técnicas de gestión de vida silvestre*. The Wildlife Society, Inc. Washington, D.C. Pp. 233- 258.

Deco Sampredro J. (2012). Estudio y aplicación de técnicas de aprendizaje automático orientadas al ámbito médico: estimación y explicación de predicciones individuales. *Trabajo de Fin de Máster presentado para la obtención del título de Máster en Ingeniería Informática y de Telecomunicaciones (major en inteligencia computacional)*. Universidad Autónoma de Madrid. Escuela politécnica Superior, Departamento de ingeniería informática. Madrid España. p. 103.

Emlen, J. T., & M. J. Dejong. 1992. Counting birds: the problem of variable hearing abilities. *Journal of Field Ornithology* 63: 26–31.

Eyben F., Wöllmer M., Graves A., Schuller B., Douglas C. E., Cowie R. (2010). On-line emotion recognition in a 3-D activation-valence-time continuum using acoustic and linguistic cues. *Journal multimodal user interfaces*. Vol. 3, pp. 7-9.

Fletcher Neville H., (2004). A simple frequency-sealling rule for animal communication. *Journal Acoustics Society*. Vol. 115, no. 5, p. 1.

Fox J. S., (2008). A new perspective on acoustic individual recognition in animal swith limited call sharing or changing repertoires. *Animal Behaviour*. Vol. 75, pp. 1187-1194.

Fuller, M.R., Seegar, W.S. & Howey, P.W. (1995). The use of satellite telemetry systems for the study of bird migration. *Israel J. Zool*. 41: 243–252.

Furnes, R.W., J.J.D. Greenwood & P.J. Jarvis. (1993). Can birds be used to monitor the environment? In: R.W Furness & J.J.D. Greenwood (Eds.) *Birds as monitors of environmental change*. Champman & Hall. Londres, Reino Unido. pp. 1-41.

Gill F. B. (2006). “Ornithology” third editions. Audubon society. *Editorial Freeman and Company*. New York. USA.

Goldsmith, F.B. (1991). Monitoring for conservation and ecology. 1º edición. *Chapman & Hall*. Londres, Reino Unido.

Gonzales G. F. & Gómez de Silva H. (2003). Amenazas, estrategias e instrumentos para la conservación de las aves. In: *Conservación de aves*:

experiencias en México, H. Gómez de Silva y A. Oliveras de Ita. (Eds.). CIPAMEX/CONABIO/NFWF, México, D. F. p. 86–119.

Hake, M., Kjellén, N. & Alerstam, T. (2003). Age-dependent migration strategy in honey buzzards *Pernis apivorus* tracked by satellite. *Oikos*. Vol. 103, pp. 385–396.

Hamel, P. B., Smith, W. P., Twedt, D. J., Woher, J. R., Morris, E., Hamilton, R. B., and R. J. Cooper. (1996). A land managers guide to point counts of birds in the South east. United States Department of Agriculture. Forest. *Southern Research Station. General Technical Report SO*. p.102.

Hanley, James A., MacNeil, Barbara J., (1982). The meaning an use of the area under a receiver operating characteristics (ROC) curve. *Radiology* Vol. 143, pp. 29-36.

Haselmayer, J. and J. S. Quinn. (2000). A comparison of point counts and sound recording as bird survey methods in Amazonian Southeast Peru. *The Condor* Vol. 102, pp. 887–893.

Hellawel, J.M. (1991). Development of a rationale for monitoring. In: F.B. Goldsmith (Ed.). *Monitoring for conservation and ecology*. Chapman & Hall. Londres, Reino Unido. pp. 1-14.

Herrera J. M. I., (2013). Uso de la Bioacústica como herramienta para determinar la riqueza de aves del interior de la selva alta perennifolia en la zona núcleo de la Sierra de Santa Martha, Reserva de la Biosfera de los Tuxtlas, Veracruz. *Tesis para la obtención de título de licenciatura. Facultad de Ciencias UNAM*. México D.F.

Howell, S. N. G. & S. Webb. (1995). *A guide to the birds of Mexico and northern Central America*. Oxford University. Oxford, Reino Unido.

Huang, A. Acero, and H. Hon. (2001). *Spoken Language Processing: A guide to theory, algorithm, and system development*. Prentice Hall.

Hutchinson, G.E. (1978). *An introduction to population ecology*. NewHaven, CT: Yale University Press.

Iñigo E. Elias, Enkeirlin Hoeflich E., (2003). Amenazas, estrategias e instrumentos para la conservación de las aves. In: *Conservación de aves*:

experiencias en México, H. Gómez de Silva y A. Oliveras de Ita. (Eds.). CIPAMEX/CONABIO/NFWF, México, D. F. pp. 86–119.

Irtza S., Hussain S., (2013). Minimally balanced corpus for speech recognition. *Communications Signal Processing, and their Applications (ICCSPA)*, 2013 1st International Conference.

Jiguet, F. & Villarubias, S. (2004). Satellite tracking of breeding black storks *Ciconia nigra*: new incomes for spatial conservation issues. *Biological Conservation Vol.* 120. pp. 153–160.

Kamruzzaman, A. N. M. Rezaul Karim, Md. Saiful Islam, Md. Emdadul Haque. (2010). Speaker Identification using MFCC-Domain Support Vector Machine. *International Journal of Electrical and Power Engineering*, Vol. 1, no. 3, pp. 274-278

Konishi, M., (1965). The role of auditory feedback in the control vocalization in the white-crowned sparrow. *Z. Tierpsychol.*, Vol.22. pp. 270-83.

Kenward, R. E. (2001). *A Manual for Wildlife Radio Tagging*. London: Academic Press.

Kepler, C. B., AND J. M. Scott. (1981). Reducing bird count variability by training observers. *Studies in Avian Biology* Vol. 6. pp. 366–371.

Kogan J., Margolias D., (1997). Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study. *Journal of the Acoustical Society of America*. Vol. 103. Pp. 2185-2196.

Kroodsma DE, Miller EH (1996). *Ecology and evolution of acoustic communication in birds*. Ithaca, N Y: Cornell UP.

Lack, David. (1954). *The Natural regulation of animal numbers*. London: Oxford University Press.

Lack, David. (1966). *Population studies of birds*. Oxford: Clarendon Press.

Lee, S.-B. Hsu, J.-L. Shih, and C.-H. Chou. (2013). “Continuous Birdsong Recognition Using Gaussian Mixture Modeling of Image Shape Features,” *IEEE Transactions on Multimedia*. Vol. 15. pp. 454–464.

Lee, C.H., Lee K.Y., Huang Z. R. (2006). Automatic Recognition of Birds songs using cepstral coefficients. *Journal of information technology and applications*. Vol. 1 No. 1, pp. 17-23.

Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi. (2010). Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. *Journal of computing*. Vol. 2 no. 3.

Llata L. M. (2003). Ecología y medio ambiente. Editorial progreso. México D.F. p. 74-75.

Manning Aubrey, (1985). Introducción a la conducta animal., Editorial Alianza. Madrid España. 3º Edición. pp. 125-153.

Marler, P. (2004). Science and birdsong: the good old days. In: Marler P. y Slabbekoorn H. (Eds.), Nature's Music. *The Science of Birdsong*. Elsevier. USA. pp. 1-19.

Martínez, Camblor P. (2007). Comparación de pruebas diagnósticas desde la curva ROC. *Revista Colombiana de Estadística*. Vol. 30. no. 2. pp. 163-176.

Mayr, E. 1946. The number of species of birds. *Auk*. Vol. 63. pp. 64-69.

McGrady, M.J., Ueta, M., Potapov, E., Utekhina, I., Masterov, V., Ladyguine, A., Zykov, V., Cibor, J., Fuller, M. & Seegar, W.S. (2003). Movements by juvenile and immature Steller's Sea Eagles *Haliaeetus pelagicus* tracked by satellite. *Ibis* 145: 318–328.

Milanovic S, Z. Lukac, and A. Domazetovic. (1999). One solution of speech activity detection. In *Conference on Telecommunications ETRAN*,

Neal, F. Briggs, R. Raich, and X. Z. Fern. (2011). Time-frequency segmentation of bird song in noisy acoustic environments. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011*. pp. 2012 –2015.

Potamitis and T. Ganchev. (2008). Generalized Recognition of Sound Events: Approaches and Applications, *Studies in Computational Intelligence (SCI)* Vol. 120. Pp. 41–79.

Salcedo C. F. J. (2009). Modelos Ocultos de Markov: Del reconocimiento de voz a la música. Editor: Salcedo Campos Javier Francisco. Málaga. España.

Sanjit K. Mitra. (2007). Procesamiento de señales digitales. Un enfoque en computadora. 3 Ed. *Mc Graw Hill*. University of California. USA.

Navarro-Sigüenza., Sánchez-González L. (2003). La diversidad de Ave. In: *Conservación de Aves, Experiencias en México*. H. Gómez de Silva y A. Oliveras de Ita. (Eds.). CIPAMEX/CONABIO/NFWF, México, D. F. pp. 24-56.

Oropeza R., Suárez G., (2006). Algoritmos y métodos para el reconocimiento de voz en español mediante sílabas. *Computación y Sistemas* Vol. 9 Núm. 3, pp. 270-286.

O'Shaughnessy D. (2000). *Speech Communications: Human and Machine*. Institute of Electrical and Electronics Engineers. Editor John Wiley & Sons Inc; Edición: 2nd Revised edition.

Perrins C. (2011). *The new Encyclopedia of Birds*. Editorial LIBSA. Madrid.

Peshek R.K. & Blumstein D.T. (2001). Can rarefaction be used to estimate song repertoire size in birds?. The Rocky Mountain Biological Laboratory, Box 519, Crested Butte, CO 81224, USA. *Current Zoology* Vol.57 no.3. pp. 300–306.

Pieretti, N., Farina, A. & Morri, F. D. (2010) A new methodology to infer the singing activity of an avian community: the Acoustic Complexity Index (ACI). *Ecological Indicators*, Vol.11. pp. 868-873.

Pineda, L. A., Castellanos, H., Cuétara, J., Galescu, L., Juárez, J., Llisterra, J., Pérez, P., and Villaseñor, L. (2009). The corpus Dimex100: Transcription and evaluation. *Language Resources and Evaluation*.

Pikrakis A., Theodoros Giannakopoulos, and Sergios Theodoridis (2008). An overview of speech/music discrimination techniques in the context of audio recordings. In George A. Tsihrantzis and Lakhmi C. Jain, editors, *Multimedia Services in Intelligent Environments*. Vol. 120. p 81.

Power, D.M.W. (2011). Evaluation: from precision, recall and f-measure to roc, informedness, markedness & correlation. *Journal of machine learnig technologies*, Vol. 2. no. 1. pp. 37-63.

Proakis, J. G., & Manolakis G. D. (1996). *Digital Signal Processing. Principles, Algorithms, and Applications*. Third Edition. Prentice-Hall International, inc. USA.

Ralph, C. J., Geupel G. R., P. Pyle, T.E. Martin, D.F. DeSante And B. Milan. (1996). Manual de métodos para el monitoreo de aves terrestres. General Technical. Rep. PSW-GTR-159. Albany, CA: *Pacific Aouthwest Research Stattion*, Forest Service, U.S. Departament of Agriculture. pp. 45.

Rempel, R. S, Hobson K. A., Holborn G., Van Wilgerburg S. L., Elliot J., (2005). Bioacoustic monitoring of forest songbirds: Interpreter variability and effects of configuration and digital processing methods in the laboratory. *Journal of field ornithology*. Vol.76. no.1. pp. 1-11.

Romero, A. M., Hernández S., Estrada G., Owen D. (2007). Mamíferos pequeños: Manual de técnicas de captura, preparación preservación y estudios. *La presa de ciencias*, Facultad de Ciencias UNAM. México D.F.

Tirira, Diego S. (1998). *Biología, sistemática y conservación de los mamíferos del Ecuador*, Museo de Zoología, Centro de Biodiversidad y Ambiente, Pontificia Universidad Católica del Ecuador. Publicación especial. Vol. 1. Pp. 93-125.

Tubaro, Pablo L., 1999. *Bioacústica aplicada a la sistemática, conservación y manejo de poblaciones naturales de aves*. Laboratorio de Biología del Comportamiento, Instituto de Biología y Medicina Experimental.Vta. de Obligado 2490 - 1428 Buenos Aires, Argentina. *Etología*, 7:19-32

Towsey, M., Birgit P., Alfredo Nantes, Jason Wimmer & Paul Roe (2012): A toolbox for animal call recognition, Bioacoustics: *The International Journal of Animal Sound and its Recording*, Vol. 21. No.2. 107-125.

Tudge, C. (2008). The Secrets life of birds. *Penguin books*. Londres, Inglaterra. 187-194.

Thomson, S., LeDoux K., Moody K. (1994). A System for describing birds song units. Bioacoustics: *The International Journal of Animal Sound and its Recording*. Vol.5. no.4 pp. 267-279

Thorpe, W.H. (1958). The learning of song patterns by birds, with especial reference to the song of the Chaffinch, *Fringilla coelebs*. *Ibis*. Vol.100. pp. 535- 570.

Sanabria, A., C. Pedraza, J. Vitola. (2010). Algoritmo rápido para la búsqueda de audio por contenido sobre GPUs. *Intekhnia*. Vol.6 no.1 pp.35-43

Sanchez, C.F., Tarango A., (2007). “Métodos de estimación de poblaciones de Fauna Silvestre, principios y prácticas”, *Reproducción y manejo de Fauna silvestre.*, México D.F., UAM-Iztapalapa. Vol.3, p. 58

Sokolova, M., Japkowicz N., Szpakowicz S., (2006). Beyond accuracy, f-score and ROC: a family of discriminant measures for performance evaluation. Published in Proceeding AI'06 *Proceeding of the Australian joint conference on Artificial Intelligence: advanced in Artificial Intelligence*. Pp. 1015-1021.

Smith, A. P. (1973). Stratification of temperate and tropical forests. *American Naturalist* Vol. 107. Pp. 671-683.

Umberto, Filibeck , Marcello Nicoli , Patrizia Rossi & Giorgio Boscagli (1982). Detection by frequency analyzer of individual wolves howling in a chorus: A preliminary report, *Bolletino di zoologia*, Vol. 49. no. 2. 151-154

Vapnik, V. (1982). *Estimation of Dependences Based on Empirical Data*, Springer-Verlag.

Villaseñor Gómez J., Santana E. C., (2003). El monitoreo de poblaciones herramienta necesaria para la conservación de aves en México. *Conservación de aves: experiencias en México*, H. Gómez de Silva y A. Oliveras de Ita. (Eds.). CIPAMEX/CONABIO/NFWF, México, D. F.

Sibley, C. G. & B. L. Monroe Jr. (1990). *Distribution and taxonomy of the birds of the world*. Yale University Press. New Haven, Connecticut.

Sibley, C. G. & J. E. Ahlquist. (1991). *Phylogeny and classification of birds*. Yale University Press. New Haven, Connecticut.