



INSTITUTO POLITÉCNICO NACIONAL

Centro de Investigación en Computación

LABORATORIO DE SISTEMAS INTELIGENTES PARA LA AUTOMATIZACIÓN

**Modelo computacional para el análisis de patrones
acústicos subacuáticos**

T E S I S

QUE PARA OBTENER EL GRADO DE
MAESTRÍA EN CIENCIAS EN INGENIERÍA DE CÓMPUTO
CON OPCIÓN EN SISTEMAS DIGITALES

P R E S E N T A

ING. FERNANDO RUBÉN GONZÁLEZ HERNÁNDEZ

DIRECTORES DE TESIS

DR. LUIS PASTOR SÁNCHEZ FERNÁNDEZ
DR. SERGIO SUÁREZ GUERRA



MÉXICO, D.F.

DICIEMBRE 2015



INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 10:00 horas del día 24 del mes de noviembre de 2015 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

Centro de Investigación en Computación

para examinar la tesis titulada:

“Modelo computacional para el análisis de patrones acústicos subacuáticos”

Presentada por el alumno(a):

González
Apellido paterno

Hernández
Apellido materno

Fernando Rubén
Nombre(s)

Con registro:

B	1	3	0	0	7	4
---	---	---	---	---	---	---

aspirante de: **MAESTRÍA EN CIENCIAS EN INGENIERÍA DE CÓMPUTO CON OPCIÓN EN SISTEMAS DIGITALES**

Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Directores de Tesis

Dr. Luis Pastor Sánchez Fernández

Dr. Sergio Suárez Guerra

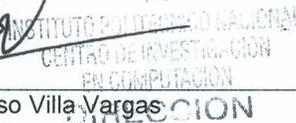
Dr. Oleksiy Pogrebnyak

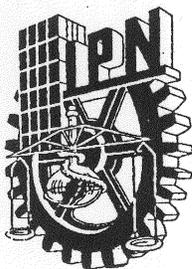
Dr. José de Jesús Medel Juárez

M. en C. Romeo Urbieta Parrazales

PRESIDENTE DEL COLEGIO DE PROFESORES

Dr. Luis Alfonso Villa Vargas



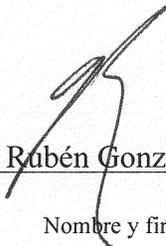


INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México, D. F. el día 1 del mes de Diciembre del año 2015, el que suscribe Fernando Rubén González Hernández alumno del Programa de Maestría en Ciencias en Ingeniería de Cómputo con número de registro B130074, adscrito al Centro de Investigación en Computación, manifiesta que es autor intelectual del presente trabajo de Tesis bajo la dirección del Dr. Luis Pastor Sánchez Fernández y el Dr. Sergio Suárez Guerra y cede los derechos del trabajo intitulado Modelo computacional para el análisis de patrones acústicos subacuáticos, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección f.er.moshi@hotmail.com. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.


Fernando Rubén González Hernández

Nombre y firma

Resumen

La acústica pasiva es muy importante en diversas áreas. Aunque viene siendo estudiada desde hace mucho tiempo, los investigadores aún siguen investigando nuevas técnicas de detección, clasificación y localización, que logren mejorar los algoritmos ya existentes, en cuanto a una mayor precisión, ahorro de energía, etc. Su utilidad depende del área donde vaya a ser usada, pero puede ser desde investigación de animales marinos, monitoreo de niveles de sonido antropogénico, monitoreo de instalaciones de petróleo hasta aplicaciones militares.

Debido a la poca investigación hecha en el país, surge la necesidad de realizar un modelo que logre clasificar señales acústicas subacuáticas en base a la firma acústica de la clase de interés, considerando algunos factores como la contaminación por ruido ambiental u otros que puedan presentarse durante el proceso de grabación. Lo que permitirá desarrollar un modelo de reconocimiento original.

Se diseñó un clasificador basado en un conjunto (“ensemble”) de redes neuronales MLP, capaz de identificar 14 clases, las cuales usan frecuencias desde 0 hasta 22,050 Hz. Para su diseño, se analizaron los diferentes tipos de sonido emitidos y se eligió la técnica de extracción que presentó mejores resultados.

Se obtuvieron buenos resultados en la clasificación de diversas señales acústicas subacuáticas, considerando el estado del arte, probando que la arquitectura propuesta puede ser usada para integrar más clases que puedan ser de interés. Estos resultados son las bases para ayudar a monitorear los océanos de México ya sea en tiempo real o diferido u otros lugares de interés.

Abstract

Passive acoustics is a very important field in so many areas. Although this field has been studied for a long time, scientists are already working on new techniques for detection, classification and localization, in order to improve the current algorithms; improving the accuracy, energy used, etc. The usefulness of this field depends on the area, but it can be implemented from marine animal research, anthropogenic levels monitoring, oil and gas facilities monitoring to military purposes.

Due to the little research made on the country, the need arises of make a model which is able to classify underwater acoustic signals based on its acoustic signature. Considering some external factors such as ambient noise or others which could be present during the recording process. This will lead to develop an original recognition model.

A classifier based on ensemble methodologies was designed, having MLP neural networks as base classifiers and being able to classify up to 14 classes, whose signals go from 0 to 22,050 Hz. The sounds involved were analyzed and the best feature extraction technique was chosen.

Good results were obtained in the classification of many underwater acoustic signals, considering the state of art, even proving that the architecture proposed can be used with more classes. These results serve as a basis to help monitoring México's seas or any other place of interest, either real-time or deferred.

Agradecimientos

Les agradezco principalmente a mis padres, Fernando González Vega e Irma Leticia Hernández González, a mi hermana Alejandra González Hernández y al resto de mi familia por todo su apoyo.

También a mis directores de tesis, Dr. Luis Pastor Sánchez Fernández y Dr. Sergio Suárez Guerra, por estar al pendiente del desarrollo de esta tesis y otros trabajos.

Al IPN, CONACYT y Fundación TELMEX por el apoyo económico recibido durante el transcurso de la maestría.

Y por último pero no menos importante a mis amigos, por haber hecho más placentera mi estadía en la maestría y en el Laboratorio.

Índice

Contenido

1.	Introducción	1
1.1.	Antecedentes	1
1.2.	Planteamiento del problema	2
1.3.	Justificación.....	3
1.4.	Hipótesis	3
1.5.	Objetivos	3
1.5.1.	Objetivo general	3
1.5.2.	Objetivos particulares.....	4
1.6.	Alcances	4
1.7.	Contribuciones.....	4
1.8.	Metodología y desarrollo de la investigación.....	5
1.9.	Estructura de la tesis.....	6
2.	Estado del arte.....	7
2.1.	Métodos de observación y monitoreo del sonido en el mar	7
2.1.1.	Sistemas acústicos de registro de datos.....	7
2.1.2.	Unidades de grabación acústica de archivamiento marino	7
2.1.3.	Sensores acústicos pasivos en tiempo real	11
2.1.4.	Etiquetas de grabación acústica pasiva	12
2.1.5.	Observatorios	13
2.2.	Efectos del ruido en mamíferos marinos.....	15
2.3.	Regulaciones internacionales	16
2.4.	Sistemas de clasificación automática	16
2.4.1.	Sistemas basados en aprendizaje automático	17
2.4.2.	Aplicaciones dedicadas a la detección y localización de cetáceos.....	19
2.5.	Conclusiones	20
3.	Marco teórico	21
3.1.	El sonido en el mar.....	21
3.1.1.	Importancia	21
3.1.2.	Ruido ambiental	23

3.1.3.	Niveles de sonido en aire y en agua	23
3.2.	Técnicas de procesamiento de voz	24
3.2.1.	Lpc Codificación predictiva lineal (LPC)	24
3.2.2.	Coeficientes cepstrales en las frecuencias de Mel (MFCC)	26
3.2.3.	Análisis de octava	29
3.3.	Redes neuronales artificiales (RNAs)	31
3.3.1.	Aprendizaje por cuantificación vectorial	32
3.3.2.	Perceptrón multicapa	34
3.4.	Clasificador mediante conjuntos	35
3.4.1.	Conjunto con estructura dependiente	36
3.4.2.	Conjunto con estructura independiente	36
3.4.3.	Métodos de combinación	38
3.4.4.	Generación de diversidad	39
3.5.	Conclusiones	40
4.	Análisis de las señales acústicas subacuáticas objeto de estudio	41
4.1.	Selección de las bases de datos	41
4.1.1.	Mamíferos marinos	41
4.1.2.	Peces	42
4.1.3.	Antropogénicos	44
4.1.4.	Naturales	44
4.2.	Clases a reconocer	44
4.3.	Emisiones de mamíferos marinos	44
4.3.1.	Odontocetos	45
4.3.2.	Misticetos	52
4.3.3.	Sirenios	54
4.4.	Emisiones de peces	55
4.5.	Emisiones de señales antropogénicas	56
4.6.	Emisiones de señales naturales	56
4.7.	Etiquetado manual de señales	57
4.8.	Base de datos generada	59
4.9.	Análisis y extracción de características	59
4.9.1.	Pre-procesamiento	59
4.9.2.	Detección de señal	59

4.9.3. Extracción de características	60
4.10. Conclusiones.....	63
5. Método de reconocimiento.....	64
5.1. Etapa de clasificación	64
5.2. Redes neuronales.....	67
5.3. Módulo de decisión	67
5.4. Conclusiones	68
6. Pruebas y resultados	69
6.1. Selección de clasificadores base	69
6.2. Elección de cantidad de coeficientes.....	71
6.3. Prueba del modelo para 11 clases	75
6.4. Escalando el modelo	76
6.5. Análisis de tiempos	82
6.6. Conclusiones	83
Conclusiones.....	84
Trabajo futuro.....	86
Referencias bibliográficas.....	87
Referencias electrónicas.....	94
Anexo A. Listado de mamíferos marinos en el Golfo de México	96
Anexo B. Peces incluidos en la Carta Nacional Pesquera	97
Anexo C. Señales acústicas usadas	100
Anexo D. Paralelización de la etapa de clasificación	103
Anexo E. Sistema de medición	104

Índice de Figuras

Fig. 1-1. Espectrograma de la superposición de señales emitidas por una cantidad desconocida de delfines.	4
Fig. 1-2. Esquema del método de investigación empleado.....	5
Fig. 2-1. Sistema de registro de datos.	7
Fig. 2-2. Esquemático de un ARP.	8
Fig. 2-3. Esquemático del sistema usado por PMEL.	9
Fig. 2-4. MARU siendo bajada al océano.	9
Fig. 2-5. EAR usado en Oahu, Hawai.....	10
Fig. 2-6. DASAR utilizado en el mar de Beaufort.	11
Fig. 2-7. Boya de detección automática de tiempo real.....	11
Fig. 2-8. Boya pequeña (esfera amarilla) conectada a sus instrumentos (esfera naranja) por medio de un cable Gumby hose.	12
Fig. 2-9. D-tag sujetado a la aleta dorsal de una ballena piloto.	13
Fig. 2-10. Vista del área de PALAOA.	14
Fig. 2-11. Zonas de influencia alrededor de una fuente de ruido.	15
Fig. 2-12. Arquitectura usada para el reconocimiento de cachalotes y barcos.	18
Fig. 3-1. Canal SOFAR.....	22
Fig. 3-2. Funcionamiento de un sonar activo.	22
Fig. 3-3. Relación entre la escala de frecuencia en Hz y la escala de Mel.....	27
Fig. 3-4. Respuesta en frecuencia del banco de filtros con amplitud normalizada.	28
Fig. 3-5. Respuesta en frecuencia del banco de filtros del análisis de octava en una escala de frecuencia lineal.	30
Fig. 3-6. Respuesta en frecuencia del banco de filtros del análisis de 1/3 de octava en una escala de frecuencia lineal.....	31
Fig. 3-7. Vectores de entrada originales.....	32
Fig. 3-8. Libro de códigos con 16 vectores.	33
Fig. 3-9. Red LVQ.....	33
Fig. 3-10. MLP con una capa oculta.....	35
Fig. 3-11. Estructura dependiente.	37
Fig. 3-12. Estructura independiente.	37
Fig. 4-1. Mapa del Golfo de México.....	41
Fig. 4-2. Clases a reconocer.	45
Fig. 4-3. Espectrograma del silbido de un T. Truncatus en el que su armónico rebasa la frecuencia máxima (1/2 de la frecuencia de muestreo).	47
Fig. 4-4. Espectrograma del silbido tipo constante (con armónicos) de un T. Truncatus....	48
Fig. 4-5. Espectrograma del silbido tipo hacia arriba de un T. Truncatus.	48
Fig. 4-6. Espectrograma del silbido tipo hacia abajo de un T. Truncatus.....	49
Fig. 4-7. Espectrograma del silbido tipo cóncavo de un T. Truncatus.....	49
Fig. 4-8. Espectrograma del silbido tipo convexo de un T. Truncatus.....	50
Fig. 4-9. Espectrograma del silbido tipo sinusoidal de un T. Truncatus.	50

Fig. 4-10. Espectrograma genérico con algunos de los diferentes parámetros de un contorno.	51
Fig. 4-11. Espectrograma de una ráfaga de pulsos de un T. Truncatus.	52
Fig. 4-12. Espectrograma de dos llamados de un B. Acutorostrata.	53
Fig. 4-13. Espectrograma de una unidad producida por una M. Novaeangliae.....	54
Fig. 4-14. Espectrograma de un chillido de un T. Manatus.	55
Fig. 4-15. Espectrograma de sonido hecho por un B. Chrysoura.	56
Fig. 4-16. Espectrograma del sonido de una embarcación pequeña.	56
Fig. 4-17. Estructura común de un registro de la base de datos.....	57
Fig. 4-18. Programa de etiquetado en funcionamiento.	58
Fig. 4-19. Comparativa de LPCs.	60
Fig. 4-20. Comparativa de MFCCs.....	61
Fig. 4-21. Comparativa de Octavas.	61
Fig. 4-22. Comparación de resolución del análisis de octava.....	62
Fig. 4-23. Programa de extracción de características.....	63
Fig. 5-1. Función tangente sigmoïdal hiperbólica.	64
Fig. 5-2. Arquitectura general del conjunto.	64
Fig. 5-3. Análisis de octava de un tipo de emisión de un O. Orca.	65
Fig. 5-4. Análisis de octava de una emisión de un M. Novaeangliae.	66
Fig. 6-1. Arquitectura del primer conjunto.....	69
Fig. 6-2. Arquitectura de la red N_1 para 11 clases.	70
Fig. 6-3. Arquitectura de la red N_2 para 11 clases.	70
Fig. 6-4. Arquitectura de la red N_3 para las 11 clases.....	71
Fig. 6-5. Arquitectura de la red N_4 para las 11 clases.....	71
Fig. 6-6. Matriz de confusión para red N_1 para 11 clases.....	73
Fig. 6-7. Matriz de confusión para red N_2	74
Fig. 6-8. Matriz de confusión de red N_3 para 11 clases.	74
Fig. 6-9. Matriz de confusión de red N_4 para 11 clases.	75
Fig. 6-10.Arquitectura del conjunto final.	76
Fig. 6-11. Arquitectura de la red N_1 para 14 clases.	77
Fig. 6-12. Arquitectura de la red N_3 para 14 clases.	77
Fig. 6-13. Arquitectura de la red N_4 para 14 clases.	78
Fig. 6-14. Matriz de confusión de red N_1 para 14 clases.	79
Fig. 6-15. Matriz de confusión de red N_3 para 14 clases.	79
Fig. 6-16. Matriz de confusión de red N_4 para 14 clases.	80
Fig. 6-17. Porcentaje de tiempo de análisis con base en la ventana.	82
Fig. D-0-1. Diagrama de paralelización de la etapa de clasificación.	103
Fig. E-0-1. Arquitectura del sistema de medición.	104

Índice de Tablas

Tabla 1. Cantidad de especies marinas por región geográfica.....	3
Tabla 2. Ejemplos de niveles de sonido submarino.	24
Tabla 3. Ejemplo de los conjuntos de entrenamiento usando bagging.	38
Tabla 4. Listado de registros usados de mamíferos marinos.	42
Tabla 5. Ficha en SAGARPA de peces del Golfo de México.	43
Tabla 6. Listado de registros usados de peces.	43
Tabla 7. Fuentes y cantidad de registros.....	58
Tabla 8. Fuentes y cantidad de patrones para entrenamiento/prueba.....	66
Tabla 9. Comparación de cantidad de coeficientes para 11 clases.....	72
Tabla 10. Resultados por red para 14 clases.	78
Tabla 11. Comparativa con el estado del arte.....	82

Glosario de Siglas

ACO	“ALOHA Cabled Observatory”
ARP	“Acoustic Recording Package”
ARU	“Acoustic Recording Unit”
AULS	“Autonomous Underwater Listening Station”
AUTEC	“Atlantic Undersea Test and Evaluation Center”
AWI	“Alfred Wegener Institut”
C	Es un lenguaje de programación
CART	“Classification And Regression Tree”
CPMM	Comité de Protección del Medio Marino
DASAR	“Directional Autonomous Seafloor Acoustic Recorder”
DAT	“Digital Audio Tape”
dB	Decibel, unidad de medida para el nivel sonoro
DOSITS	“Discovery of Sound In The Sea”
EAR	“Ecological Acoustic Recorder”
GNL	Gas Natural Licuado
ha	Hectárea, medida de superficie, que es igual a 10,000 metros cuadrados
HARP	“High frequency Acoustic Recording Package”
Hz	Hertz, unidad de frecuencia, equivale a la frecuencia de un fenómeno cuyo periodo es un segundo
INAPESCA	Instituto Nacional de Pesca
JAMSTEC	“Japan Agency for Marine-Earth Science and Technology”
JIP	“Sound and Marine Life Joint Industry Programme”
LAB	“Laboratori d'Aplicacions Bioacústiques”

LAN	“Local Area Network”
LARS	“Long Term Acoustic Recording System”
LIDO	“Listen to the Deep Ocean Environment”
LPC	“Linear Predictive Coding”
MARU	“Marine Acoustic Recording Unit”
MFCC	“Mel Frequency Cepstral Coefficients”
MLP	“Multi-Layer Perceptron”
NOAA	“National Oceanic and Atmospheric Administration”
NPS	“National Park Service”
OGP	“International Association of Oil and Gas Producers”
OMI	Organización Marítima Internacional
Pa	Pascal, unidad de presión
PALAOA	“Perennial Acoustic Observatory in the Antarctic Ocean”
PAM	“Passive Acoustic Monitoring”
PMEL	“Pacific Marine Environmental Laboratory”
POD	“Porpoise detector”
ppm	Partes por millón
PTS	“Permanent Threshold Shift”
RMS	“Root Mean Square”
RNA	Red Neuronal Artificial
SAGARPA	Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación
SNR	“Signal to Noise Ratio”
SOFAR	“Sound Fixing and Ranging”
SOSUS	“Sound Surveillance System”
TLU	“Threshold Logic Unit”

TTS	“Temporal Threshold Shift”
UE	Unión Europea
UPC	Universidad Politécnica de Cataluña
VHF	“Very High Frequency”, es la banda del espectro electromagnético que ocupa el rango de frecuencias de 30 MHz a 300 MHz
WHOAS	“Woods Hole Open Access Server”
WHOI	“Woods Hole Oceanographic Institution”

Glosario de Términos

Acústica pasiva	Es la acción de escuchar los sonidos
Aleta dorsal	Aleta presente en la región dorsal
Antropogénico	Efectos, procesos o materiales que son el resultado de actividades humanas
Aparato de fonación	Es el conjunto de los diferentes órganos que intervienen en la articulación del lenguaje
Aprendizaje automático	También conocido como “machine learning”, es una rama de la inteligencia artificial, cuyo objetivo es desarrollar técnicas que permitan a las computadoras aprender
Arrecife de coral	Comunidad marina dominada por comunidades coralinas y estructuras rocosas, con una gran diversidad de especies de algas, invertebrados y peces
Bagging	También llamado “bootstrap aggregating”, es un meta-algoritmo usado en “ensembles” para mejorar la estabilidad y precisión
Bairdiella Chrysoura	También conocido como ronco amarillo, es una especie de pez
Balaenoptera Acutorostrata	También conocida como ballena de minke común, es una especie de cetáceo misticeto
Balaenopteridae	Es la familia más abundante y diversa de cetáceos misticetos
Biomasa	Cantidad de materia acumulada en un ecosistema
Boya	Dispositivo flotante situado en el mar
Canal SOFAR	Canal “Sound Fixing and Ranging”, es una capa horizontal de agua ubicada entre los 800 y 1000 m aproximadamente, donde la velocidad del sonido es mínima, lo que le permite viajar mayores distancias antes de disiparse
Cavitación	Efecto hidrodinámico que se produce cuando el agua o un líquido pasa a gran velocidad por una arista afilada
Cetáceo	Engloba a varias familias de mamíferos marinos como las ballenas, cachalotes, zifios, delfines, orcas y marsopas

Clasificación taxonómica	Es la ordenación sistemática y jerarquizada de los grupos de animales y vegetales
Clic	Sonido realizado por algunos animales marinos
Comunicación intraespecífica	Comunicación entre animales de la misma especie
Cynoscion Nebulosus	También conocido como trucha pinta, es una especie de pez
Delphinidae	Es una familia de cetáceos odontocetos también conocidos como delfines oceánicos
Desove	Es el acto de verter los huevos y espermatozoides por los peces y anfibios en su ambiente
Ecosistema	Es el conjunto de especies de un área determinada que interactúan entre ellas y con su ambiente abiótico
Ensemble	Del inglés: conjunto; en aprendizaje automático es un conjunto de clasificadores redundantes
Espacio de hipótesis	Es el conjunto de todas las hipótesis que el algoritmo de aprendizaje es capaz de abarcar
Estero	Designa condiciones de pantano
Estridulación	Es la acción de producir sonido mediante la fricción de ciertas partes del cuerpo
Exploración sísmica	Método geofísico usado para localizar las rocas porosas que almacenan los hidrocarburos (petróleo y gas)
Fatómetro	Instrumento utilizado en navegación, para determinar la profundidad
Gumby hose	Cable elástico de comunicación entre una boya y los instrumentos de la misma
Ictiofauna	Conjunto de especies de peces que existen en una determinada región
Limnología	Ciencia que estudia los aspectos físicos y biológicos de los ecosistemas de agua dulce, especialmente los lagos
Línea lateral	Órgano sensorial que sirve para detectar movimiento y vibración en el agua circundante

Llamado	Es el nombre que reciben ciertos tipos de sonidos, hechos por animales
Megaptera Novaeangliae	También conocida como ballena jorobada, es una especie de cetáceo mysticeto
Meta-aprendizaje	También conocido como “meta-learning”, es el entendimiento de la interacción entre el mecanismo de aprendizaje y el contexto en el cual ese mecanismo es aplicable
Milla náutica	Unidad de longitud, empleada en navegación marítima y aérea
Misticeto	Son un suborden de los cetáceos, y se caracterizan por la presencia de barbas en lugar de dientes
Odontoceto	Son un suborden de los cetáceos, y se caracterizan por la presencia de dientes en lugar de barbas
Omnidireccional	En cuanto a micrófonos se refiere a que captan todos los sonidos independientemente de la dirección desde donde lleguen
Opérculo	En los peces óseos, es una aleta de hueso duro que cubre y protege las branquias
Orcinus Orca	También conocida como orca, es una especie de cetáceo odontoceto
Pez vicario	Especie de origen marino pero parcial o permanentemente adaptado a vivir en agua dulce
Physeter Macrocephalus	También conocido como cachalote, es una especie de cetáceo odontoceto
Physeteridae	Es una familia de cetáceos odontocetos
Ping	Pulso de sonido originalmente usado por los submarinos para detectar y localizar obstáculos
Plataforma de hielo	Gruesa plataforma flotante de hielo que se forma cuando un glaciar pierde parte de una capa de hielo
Pogonias Cromis	También conocido como tambor negro, es una especie de pez
Pseudorca Crassidens	También conocida como falsa orca, es una especie de cetáceo odontoceto
Radiofaro	Emisora de radio que envía de forma automática y continua

señales como ayuda a la navegación aérea y marítima

Sciaenidae	Es una familia de peces comúnmente llamados corvinas
Sensor vectorial	Dispositivo de escucha subacuática que mide tanto el movimiento de las partículas como los cambios de presión asociados con una onda de sonido
Sirenios	Son los únicos mamíferos marinos herbívoros, junto con los cetáceos, son los únicos mamíferos adaptados completamente a la vida acuática
Sistema vestibular	Sistema o aparato vestibular, está relacionado con el equilibrio y el control espacial
Sonar	Originalmente un acrónimo de “Sound Navigation and Ranging”, es una técnica que usa la propagación del sonido para navegar, comunicar o detectar objetos sobre o debajo de la superficie del agua
Sonoboya	Combinación de sonar y boya, que se despliega al impactar contra el agua, extendiendo bajo el agua un sistema de hidrófonos y una antena de radio en el aire
Stenella Frontalis	También conocido como delfín manchado del Atlántico, es una especie de cetáceo odontoceto
Tag	Dispositivo que funciona como etiqueta
Trichechus Manatus	También conocido como manatí del Caribe, es una especie de sirenio
Trichechidae	Es una familia de mamíferos placentarios
Tursiops Truncatus	También conocido como delfín nariz de botella, es una especie de cetáceo odontoceto
Vejiga natatoria	Órgano de flotación que poseen muchos peces óseos
Zifios	Son una familia de cetáceos odontocetos
Zona económica exclusiva	También conocida como mar patrimonial, es una franja de mar adyacente al mar territorial, que se extiende hasta 200 millas náuticas
Zona pelágica	Es la zona de mar abierto; algunos organismos de la zona pelágica forman cardúmenes, como las sardinas, ballenas, cachalotes y delfines

1. Introducción

1.1. Antecedentes

La bioacústica es una rama de la ciencia relacionada con diferentes disciplinas científicas que investiga la producción y recepción de sonidos biológicos, incluidos los que produce y procesa el ser humano, los mecanismos de transferencia de la información biológica por vínculos acústicos y su propagación en medios elásticos.

La bioacústica analiza también los órganos de la audición así como el aparato de fonación y los procesos fisiológicos y neurofisiológicos a través de los cuales los sonidos o señales acústicas son producidos, recibidos y finalmente procesados a nivel del sistema nervioso central. Asimismo, intenta entender las relaciones entre las características de los sonidos que produce un animal y la naturaleza del medio en el cual se utilizan, así como las funciones por las cuales están diseñados. Finalmente, incluye técnicas asociadas al sonar instrumental y biológico para su uso en el seguimiento de poblaciones, en mecanismos de identificación y comunicación, también representa un método para asesorar y controlar los efectos del ruido producido por las actividades humanas sobre las poblaciones animales.

En los océanos, además de la variedad de sonidos que ocurren en la naturaleza (el romper de las olas, rayos, terremotos, etc.) y los sonidos hechos por animales marinos, hay muchas fuentes de sonidos antropogénicos (generados por humanos).

Los sonidos producidos por animales marinos, procesos naturales, y actividades humanas llenan los océanos del mundo. Algunos sonidos submarinos se utilizan intencionalmente con una gran variedad de propósitos valiosos e importantes, como exploración petrolera, sonares militares, fatómetro, y otros sistemas comerciales de sonar. Para la comunidad científica, el sonido submarino es fundamental en la determinación de las propiedades básicas de los océanos y el estudio de los animales que viven allí (Scowcroft et al., 2012).

Además, la acústica proporciona un medio eficaz para documentar y analizar importantes procesos geológicos naturales, como terremotos, actividad volcánica y deslizamientos del fondo marino. Es fundamental utilizar el sonido para estudiar estos procesos, porque pueden tener efectos profundos en comunidades de zonas costeras e islas a nivel mundial.

También se utiliza para estudiar las distribuciones de los mamíferos marinos al escuchar los sonidos que éstos emiten (acústica pasiva). Las diferentes especies de ballenas y delfines (cetáceos) producen sonidos diferentes, incluyendo canciones, gemidos, clics, suspiros y zumbidos. Los científicos pueden escuchar estos sonidos y detectar, identificar y localizar diferentes especies de mamíferos marinos. La acústica pasiva se utiliza también para mejorar los estimados de abundancia o el tamaño poblacional de animales, demostrando ser un complemento eficaz para censos visuales tradicionales. Los

Introducción

científicos están utilizando ambas técnicas, especialmente para los cetáceos, ya que la acústica pasiva a menudo puede detectar un número mayor de animales a gran distancia y bajo el agua que lo que se obtendría con métodos visuales solamente. La acústica pasiva ha sido utilizada en los estimados de abundancia de varias especies de cetáceos como ballenas francas, ballenas minke, zifios, cachalotes, delfines jorobados y marsopas sin aletas.

La ballena franca del Atlántico Norte, en vías de extinción, se ha beneficiado del uso de la acústica pasiva. Sistemas pasivos de audición han sido desplegados a lo largo de la costa Atlántica de los Estados Unidos para monitorear continuamente la presencia de ballenas y reducir el peligro de colisión con buques, ya que los canales de entrada a los puertos comerciales pueden coincidir con el hábitat de la ballena. En relación con el desarrollo de una terminal de Gas Natural Licuado (GNL) en el puerto de Boston, Massachusetts, se desplegaron diez boyas de detección automática en las rutas de navegación del puerto. Computadoras en cada boya estimaron la similitud del sonido a una llamada de la ballena. La información sobre cuáles de las boyas detectan las vocalizaciones de la ballenas es transmitida a los buques y buques de GNL tienen el mandato de reducir su velocidad en las zonas alrededor de las boyas que han detectado las ballenas. A todos los demás barcos se les anima a comprobar las alertas y reducir la velocidad si es necesario.

1.2. Planteamiento del problema

Actualmente existe una carencia a nivel nacional en el desarrollo de modelos computacionales para el análisis de patrones acústicos subacuáticos, debido principalmente a que los centros de investigación nacionales como el CICIMAR (Centro Interdisciplinario de Ciencias Marinas), el CINVESTAV (Centro de Investigación y de Estudios Avanzados) Unidad Mérida, el Instituto de Ciencias del Mar y Limnología, el CIBNOR (Centro de Investigaciones Biológicas del Noroeste), entre otros se enfocan más a la parte biológica, y existen pocas colaboraciones por parte de la Ingeniería de Cómputo y en especial de Procesamiento de Señales en esta área.

Incluso la Armada de México no cuenta con un sistema de monitoreo de blancos de superficie y submarinos, para prevenir el tráfico de estupefacientes y de personas, así como para mantener el estado de derecho en la mar (CONACYT, 2015).

En Robinson (2015) se señala que en México no es común el uso de la hidroacústica, solo tres laboratorios utilizan esta técnica, principalmente para la estimación de biomasa y distribución de organismos marinos; y dentro de estas aplicaciones no se usan productos ni técnicas propias, se compra o usa hardware y software de otras partes del mundo, limitándose a las características y resultados para los que fueron construidos.

Aunque existen varios modelos de reconocimiento de señales acústicas subacuáticas, aquellos que son enfocados a múltiples especies de animales marinos

Introducción

presentan bajas tasas de clasificación, debido a la similitud y comportamiento de las señales a clasificar.

1.3. Justificación

En la Tabla 1 (Felder & Camp, 2009) puede verse la riqueza en especies marinas tan solo del Golfo de México; aunque Europa encabeza la lista, es 14.5 veces más grande que el Golfo pero solo tiene 1.7 veces más de especies.

Tabla 1. Cantidad de especies marinas por región geográfica.

Región geográfica	Número de especies	Tamaño geográfico
Europa	26,927	21.8 millones km ²
Hawai	6,150	2.2 millones km ²
Golfo de México	15,419	1.5 millones km ²
Sudáfrica	10,836	576,000 km ²
China	20,000+	-
Guam	6,172	-

Este trabajo establecería las bases tanto para este instituto como para México en el desarrollo de la acústica subacuática, la cual ha recibido cada vez más atención debido a sus diversos beneficios tanto con la ciencia como con la sociedad, y ha venido siendo requerido por agencias reguladoras, como una herramienta para el monitoreo y mitigación de los impactos ambientales por parte de las actividades humanas. Hoy en día el cuello de botella no se encuentra en acceder a los datos, sino a la incapacidad de procesar eficientemente, visualizar e interpretar los grandes volúmenes de datos, dentro de un sistema de gestión de datos (SABIOD, 2015).

1.4. Hipótesis

Es posible clasificar señales acústicas subacuáticas mediante un modelo computacional basado en redes neuronales con una tasa de clasificación competitiva. Se toman en cuenta un conjunto de sonidos generados por animales marinos, ruido antropogénico, natural y ambiental.

1.5. Objetivos

1.5.1. Objetivo general

- Desarrollar un modelo computacional para el reconocimiento de señales acústicas subacuáticas.

Introducción

1.5.2. Objetivos particulares

- Evaluar las bases de datos disponibles.
- Definir las clases a reconocer.
- Extraer rasgos característicos.
- Diseñar un modelo computacional basado en redes neuronales.
- Evaluar resultados del modelo de reconocimiento.
- Evaluación general y discusión de resultados.

1.6. Alcances

Este trabajo presenta un modelo computacional capaz de reconocer señales acústicas subacuáticas, para esto se utilizaron grabaciones producidas por animales marinos tanto en su hábitat natural como en cautiverio, dichas señales se obtuvieron de una base de datos abierta. En este trabajo se excluyen aquellas señales en las que existe una superposición de 2 o más sonidos emitidos por la misma especie o distintas (ver Fig. 1-1).

Se caracterizaron las fuentes de sonido a reconocer, obteniendo diferentes tipos de rasgos característicos, con los cuales se generó una base de datos con el fin de realizar el entrenamiento del modelo y decidir qué técnica y qué cantidad de rasgos era más conveniente utilizar.

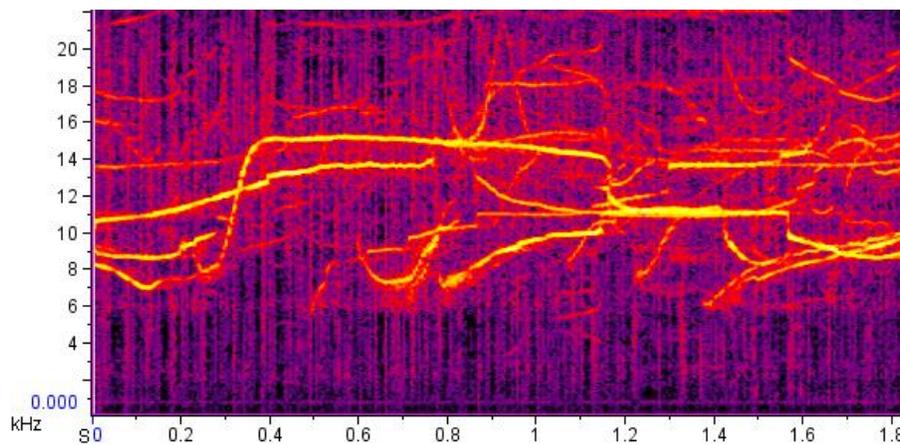


Fig. 1-1. Espectrograma de la superposición de señales emitidas por una cantidad desconocida de delfines.

1.7. Contribuciones

Un modelo computacional competitivo con las siguientes características:

Introducción

- 1) Clasifica señales acústicas subacuáticas de mamíferos marinos, peces, sonidos naturales, barcos y sonidos desconocidos (comunes durante este tipo de grabaciones). En total 14 clases.
- 2) Puede ser implementado en un sistema de tiempo real debido al bajo tiempo que requiere para analizar y clasificar las señales.
- 3) Con base en su arquitectura puede ser paralelizado y fácilmente modificado, aumentando las clases a reconocer y manteniendo altos porcentajes de clasificación.

1.8. Metodología y desarrollo de la investigación

El desarrollo de este trabajo se realizó utilizando la siguiente metodología:

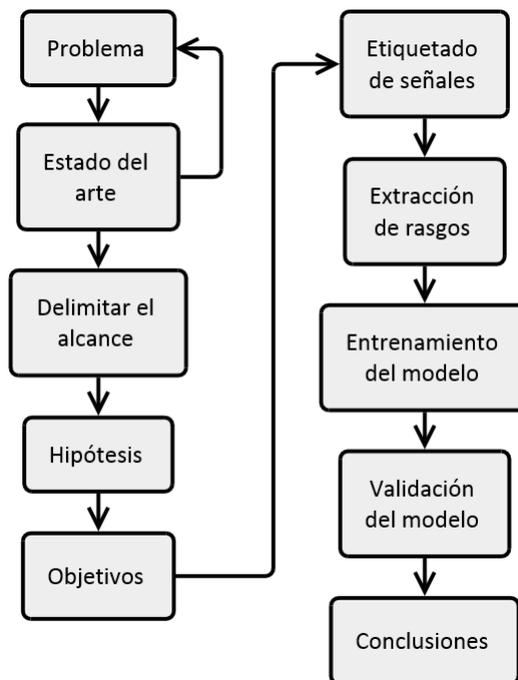


Fig. 1-2. Esquema del método de investigación empleado.

- Establecer los problemas a resolver
- Analizar el estado del arte para conocer qué es lo que se ha hecho nacional e internacionalmente, cómo y qué se puede hacer al respecto
- Buscar las bases de datos que contenían las señales más apropiadas para este trabajo
- Realizar el etiquetado manual de las señales recopiladas

Introducción

- Realizar la extracción de rasgos característicos con diferentes técnicas de procesamiento y con diferentes cantidades de rasgos
- Analizar los espectrogramas de las señales etiquetadas, para diseñar un modelo de clasificación basado en redes neuronales artificiales
- Realizar pruebas con los distintos rasgos extraídos, para determinar cuál arroja mejores resultados
- Implementar un módulo de decisión que combina las salidas de las redes neuronales, el cual determina a qué clase corresponde
- Realizar pruebas para comprobar el funcionamiento del sistema completo

1.9. Estructura de la tesis

Capítulo 2. Se presenta el estado del arte, considerando las técnicas y modelos más usados, así como los que actualmente han presentado mejores resultados en la clasificación principalmente de animales marinos.

Capítulo 3. Se tratan los fundamentos teóricos, que se usaron para el desarrollo de este trabajo.

Capítulo 4. Se presentan las clases a reconocer, analizando los sonidos que pueden generar cada una y sus características extraídas.

Capítulo 5. Se describe el método de reconocimiento, el cual incluye el modelo neuronal y el módulo de decisión.

Capítulo 6. Se presentan los experimentos y resultados, basados en el modelo computacional propuesto para la clasificación de señales acústicas subacuáticas.

Conclusiones. Se presentan las conclusiones con base en los resultados obtenidos y en los objetivos planteados.

Trabajo futuro. Basándose en los resultados obtenidos, se presenta una propuesta de posibles mejoras a los algoritmos y al modelo probado.

Anexo A. Se presenta el listado de las especies de mamíferos marinos que habitan el Golfo de México.

Anexo B. Se presenta el listado de las especies de peces que habitan el Golfo de México, en base a la ficha de la SAGARPA.

Anexo C. Se citan todas las grabaciones utilizadas en este trabajo para el entrenamiento del modelo y validación del mismo.

Anexo D. Se presenta un esquema de la paralelización del modelo propuesto.

Anexo E. Se menciona el equipo disponible con el que cuenta el laboratorio, para el desarrollo de grabaciones de campo (subacuáticas).

2. Estado del arte

2.1. Métodos de observación y monitoreo del sonido en el mar

A continuación se describen algunos métodos comunes de observación y monitoreo usados en el mar.

2.1.1. Sistemas acústicos de registro de datos

Los estudios con hidrófonos móviles han sido usados para grabar peces y localizar sus áreas de desove. Los científicos sumergen en el agua un hidrófono, conectado a un sistema de grabación, por un tiempo específico (usualmente algunos minutos), y esto lo hacen en diferentes puntos de grabación. Este tipo de estudios son bastante útiles, ya que se cubre un área geográfica bastante grande; sin embargo, requieren de mucho esfuerzo y pueden estar sujetas a las condiciones climáticas.

Hoy en día existen sistemas de registro de datos, los cuales pueden permanecer grandes periodos de tiempo bajo el agua y almacenar sonidos en intervalos de tiempo programados. Algunos ejemplos de estos sistemas son los sistemas de grabación acústica a largo plazo (LARS, por sus siglas en inglés) y la estación autónoma de escucha subacuática (AULS, por sus siglas en inglés).



Fig. 2-1. Sistema de registro de datos.

2.1.2. Unidades de grabación acústica de archivamiento marino

El uso del monitoreo acústico pasivo (PAM, por sus siglas en inglés) se ha incrementado recientemente con la disponibilidad de dispositivos de adquisición de datos para observar y monitorear animales marinos. Un tipo de sensor acústico pasivo es la unidad de grabación acústica de archivamiento marino (ARU, por sus siglas en inglés).

Existen diferentes tipos de ARU, los cuales pueden ser usados en diferentes áreas marinas; en general pueden permanecer funcionando hasta por un año, cuentan con una fuente de alimentación propia, almacenan los datos internamente, y deben ser recuperados para descargar la información recopilada y que ésta pueda ser analizada.

Estado del arte

En ocasiones suelen usarse varias unidades al mismo tiempo, para incrementar la precisión de los datos adquiridos y lograr detectar, localizar y rastrear la vocalización de peces y mamíferos marinos.

Los paquetes de grabación acústica (ARPs, por sus siglas en inglés) se han usado para monitorear mamíferos marinos y estudiar el ruido ambiental por largos periodos. Estos paquetes graban sonidos hasta de 1 kHz para identificar las vocalizaciones de baja frecuencia hechas por los mysticetos. Estos paquetes han sido de gran utilidad en las épocas del año cuando no es posible hacer reconocimiento visual.

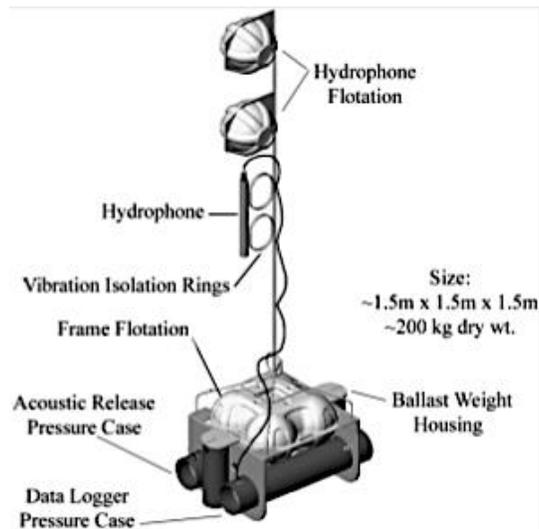


Fig. 2-2. Esquemático de un ARP.

Los paquetes de grabación acústica de altas frecuencias (HARPs, por sus siglas en inglés) son muy similares a los ARP, solo que los HARP graban sonidos de 10 Hz hasta 100 kHz; se usan principalmente para estudiar odontocetos, aunque también se ha usado para investigar y monitorear varias especies de cetáceos.

El laboratorio PMEL de la Administración Nacional Oceánica y Atmosférica (NOAA, por sus siglas en inglés) de Estados Unidos de América, ha desarrollado un sistema anclado que cuenta con hidrófonos en el canal de sonido profundo (canal SOFAR), lo cual permite detectar señales de fuentes muy distantes. Los paquetes graban de 1 Hz hasta 20 kHz. Del mismo modo estos paquetes han sido usados para estudiar varias especies de cetáceos.

Estado del arte

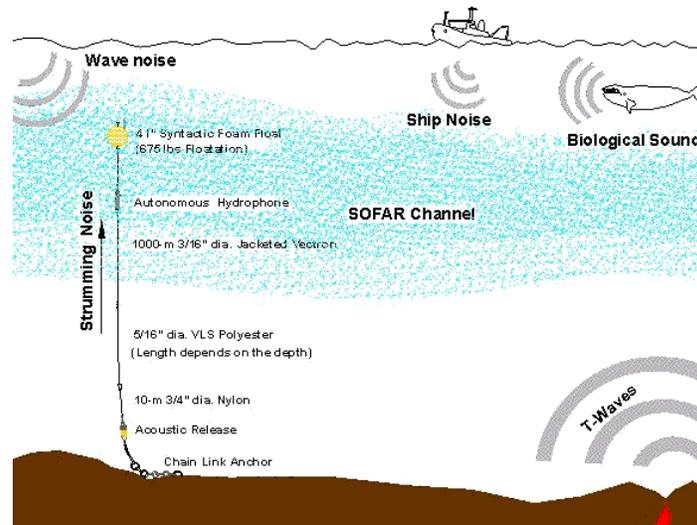


Fig. 2-3. Esquemático del sistema usado por PMEL.

La Universidad de Cornell ha desarrollado unos sistemas conocidos como unidades de grabación acústica marinas (MARUs, por sus siglas en inglés) o boyas "pop-up". Estos sistemas difieren de otros sistemas ARU en que son más pequeñas y portables. Están empaquetadas en esferas de vidrio de 17" de diámetro, que cuentan con cascos de seguridad amarillos para su protección. Se han estado usando para monitorear varias especies de cetáceos, medición del ruido ambiental y estudiar a la ballena franca, la cual está en peligro de extinción.



Fig. 2-4. MARU siendo bajada al océano.

La grabadora acústica ecológica (EAR, por sus siglas en inglés) es usada para monitorear el sonido ambiental de los arrecifes de coral y otros ecosistemas. Graba sonidos hasta de 30 kHz. Estos sistemas son una herramienta muy efectiva para documentar patrones temporales de una amplia variedad de tipos de señales acústicas en

Estado del arte

los arrecifes coralinos, y pueden detectar y monitorear la aparición de sonidos antropogénicos como el de los motores de los barcos.



Fig. 2-5. EAR usado en Oahu, Hawaii.

Algunos sistemas ARU además de almacenar los datos para su análisis posterior, procesan los datos internamente. Tales como los detectores de marsopas (PODs, por sus siglas en inglés), los cuales son completamente automatizados. Son sistemas de monitoreo acústico pasivo que detectan odontocetos por reconocer sus clics de ecolocalización.

Dentro de los PODs existen varias versiones, los T-PODs se usan para monitorear la actividad y comportamiento de la marsopa común, así como para evaluar la reacción de los cetáceos a la construcción de aerogeneradores. Los C-PODs detectan sobre una banda más ancha los clics y adquieren un rango más amplio de datos para la identificación de especies. La versión de aguas profundas (DeepC-POD) puede trabajar en profundidades de hasta 3 km, suficiente para cubrir el rango completo de profundidad de inmersión de especies como los zifios.

Otros sistemas ARU incluyen la componente direccional, un arreglo de grabadoras acústicas autónomas direccionales para el fondo marino (DASARs, por sus siglas en inglés) se usan para analizar la ubicación de ballenas boreales y evaluar la relación entre los sonidos asociados con plataformas petroleras o exploraciones sísmicas y la distribución de los “llamados” de las ballenas. Estos sistemas están diseñados para permanecer en el fondo del océano y grabar sonidos hasta por 116 días. Esto permite su uso por mucho tiempo en áreas donde el hielo, actividades en la superficie u otros factores impiden el uso de sonoboyas.

Estado del arte

Cada DASAR incluye un hidrófono omnidireccional para grabar los niveles de presión sonora y dos sensores vectoriales los cuales determinan la dirección de la fuente de sonido; y puede grabar sonidos hasta de 1 kHz.

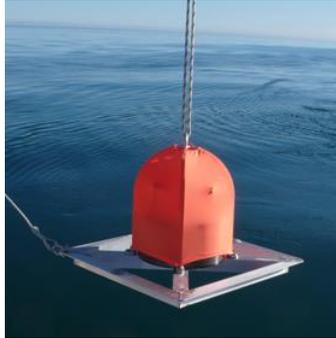


Fig. 2-6. DASAR utilizado en el mar de Beaufort.

2.1.3. Sensores acústicos pasivos en tiempo real

Estos sistemas graban y analizan los “llamados” casi en tiempo real y transmiten los datos a la costa. La Universidad de Cornell y la Institución Oceanográfica “Woods Hole” (WHOI, por sus siglas en inglés) han desarrollado boyas de detección automática para cumplir con un mandato de cumplimiento ambiental que requiere el puerto de aguas profundas “Northeast Gateway” para gas natural licuado (GNL), en la bahía de Massachusetts. Debido a que la ruta hacia la terminal de GNL atraviesa el hábitat de la ballena franca del norte, la cual se encuentra en peligro de extinción, debido principalmente al impacto con las embarcaciones. Una medida tomada para evitar estos choques, fue el arreglo de las boyas “inteligentes”.



Fig. 2-7. Boya de detección automática de tiempo real.

Diez boyas crean un sistema de detección de ballenas a lo largo de la vía de navegación principal al puerto de Boston. Cada boya cuenta con un hidrófono a 60-120' debajo de la superficie. Los sonidos grabados son llevados a la boya que se encuentra en la superficie por un cable llamado "Gumby hose", el cual es un cable elástico que puede estirarse al menos dos veces su longitud original lo cual reduce la tensión en la línea de anclaje, y permite que el sistema resista el fuerte oleaje y que la boya se mantenga en la superficie. Dentro de la boya se encuentra una computadora que continuamente analiza los sonidos subacuáticos para detectar posibles llamados de la ballena franca. Cuando se detecta algo se manda vía teléfono celular o satelital a un servidor en el centro de mando y control, donde analistas validan los sonidos; los analistas determinan si los sonidos han sido detectados dentro de un rango de 5 millas náuticas de cada boya, si es así, se alerta a los buques de GNL entrantes dónde fue escuchada la ballena. Cuando los buques se acercan a la línea de navegación, los analistas reportan cada 20 minutos cualquier cambio.



Fig. 2-8. Boya pequeña (esfera amarilla) conectada a sus instrumentos (esfera naranja) por medio de un cable Gumby hose.

2.1.4. Etiquetas de grabación acústica pasiva

Las etiquetas ("tags" en inglés) permiten a los científicos acompañar a un animal marino mientras se sumerge, las grabaciones y otros datos recopilados son críticos para entender el comportamiento de las especies marinas, tanto su comportamiento natural como su respuesta a los sonidos subacuáticos. Las etiquetas deben ser pequeñas, ligeras, tolerantes a la presión (especialmente si serán usadas en aguas profundas), y deben tener suficiente tiempo de grabación.

Pueden ser sujetos al animal por ventosas, disparados con una pistola de aire o una ballesta (para ser incrustados en su piel o en su capa de grasa) o pegados directamente al cuerpo del animal. El WHOI ha desarrollado una etiqueta (D-tag) que además de contar con una grabadora, tiene una brújula digital, sensores de temperatura, de presión y un acelerómetro triaxial. Después de un tiempo preprogramado la etiqueta se desprende del animal (si es que seguía sujeto) y sale a flote, y gracias a un pequeño radiofaro VHF, los investigadores pueden rastrear la señal de radio y recuperar el D-tag.

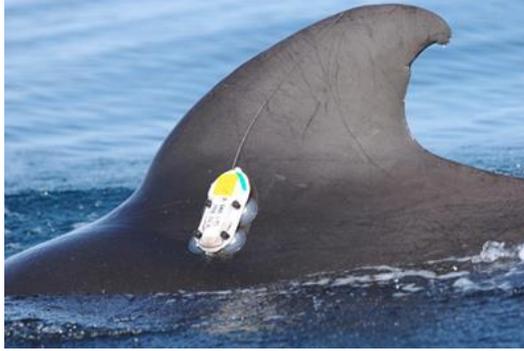


Fig. 2-9. D-tag sujetado a la aleta dorsal de una ballena piloto.

Otras etiquetas como la Acousonde, también son usados como grabadoras autónomas, suspendidas desde un cable, colocadas en el fondo marino o en un vehículo robótico u operado remotamente.

2.1.5. Observatorios

Los sistemas cableados de hidrófonos también proveen de datos continuos casi en tiempo real. Aquí los hidrófonos se sitúan de manera permanente o casi permanente, y son ubicados en zonas pelágicas, donde es raro tener contacto visual con animales marinos. Algunos ejemplos de estos sistemas son: el Sistema de Vigilancia Sónica (SOSUS, por sus siglas en inglés) y el arreglo usado para pruebas de alcance en las Bahamas por el Centro de Pruebas y Evaluación Submarina del Atlántico (AUTEK, por sus siglas en inglés), ambos sistemas de la armada estadounidense. Por lo que este tipo de sistemas son de acceso restringido debido a que son de uso militar o por el tipo de datos que se manejan, además usualmente están restringidos a señales de baja frecuencia, ya que así fueron diseñados. Otros observatorios oceánicos incluyen grandes redes de sensores oceanográficos que miden variables físicas, químicas, geológicas y biológicas, incluyendo acústica subacuática en el océano y en el fondo marino; estos observatorios sofisticados proveen de comunicación en ambos sentidos entre la costa y los diversos instrumentos oceanográficos, lo que permite hacer investigación científica continuamente. Estos sistemas permitirán monitoreo y ubicación en tiempo real de animales marinos, grabación continua en un ancho de banda más amplio y no tienen restricción en la cantidad de datos que se pueden almacenar, acceso y fuente de alimentación.

Dentro de los observatorios cableados más importantes, se encuentran el ALOHA (ACO, por sus siglas en inglés) que provee de observaciones oceanográficas en tiempo real a una profundidad de unos 4.8 km, los datos son compartidos en tiempo real con la comunidad oceanográfica y el público en general; otro observatorio importante es el “NEPTUNE Ocean Observatory”, el cual forma parte de la iniciativa “Ocean Networks

Estado del arte

Canada”, de la Universidad de Victoria, los datos son accesibles a científicos y público en general en casi tiempo real, sin embargo, dado a que comparte los datos con las Fuerzas Armadas Canadienses, éstas pueden evitar la publicación de ciertos datos.

El Observatorio Acústico Perenne en el Océano Antártico (PALAOA, por sus siglas en inglés) es una estación de escucha autónoma, sostenida por el Instituto Alfred Wegener (AWI, por sus siglas en alemán). La estación trabaja con energía eólica y solar, y se encuentra en la plataforma de hielo Ekström; la estación graba continuamente sonidos de 10 Hz hasta 20-70 kHz. Además de haber grabado varias especies marinas, suele grabar el crujir del hielo de los icebergs al chocar o desprenderse. Los datos son transmitidos vía LAN inalámbrica desde PALAOA hasta la base alemana Neumayer (ubicada en la misma plataforma) y de ahí hasta el AWI, y debido al ancho de banda limitado, la conexión puede perderse temporalmente.

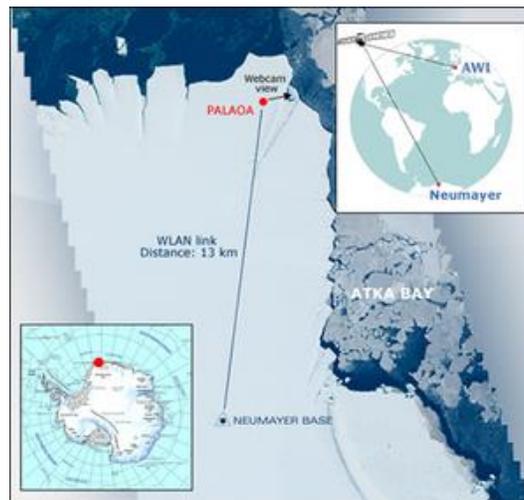


Fig. 2-10. Vista del área de PALAOA.

El proyecto LIDO (“Listen to the Deep Ocean Environment”) es un proyecto internacional de monitoreo acústico pasivo, liderado por el Laboratorio de Aplicaciones Bioacústicas (LAB, por sus siglas en catalán) de la Universidad Politécnica de Cataluña (UPC). El proyecto procesa datos en tiempo real de plataformas submarinas ancladas y equipadas con cableado a tierra; detecta, clasifica y localiza sonidos naturales, biológicos y antropogénicos. Está asociado con otros observatorios como el NEPTUNE, el JAMSTEC (de Japón), entre otros.

2.2.Efectos del ruido en mamíferos marinos

Los efectos del ruido y los rangos sobre los cuales ocurren dependiendo de las características acústicas del ruido (nivel, duración, ciclo de trabajo, etc.), el entorno y el animal bajo consideración.

En la Fig. 2-11 se representan algunas de las posibles zonas de influencia alrededor de una fuente de ruido:

- En el rango audible el ruido ha decaído debido a pérdidas por la propagación, además de que puede estar por debajo de la capacidad auditiva del animal o debajo del ruido ambiental.
- En la siguiente zona se presentan cambios de comportamiento, algunos indicadores incluyen cambios en la dirección y velocidad de nado, duración de las inmersiones, duración en la superficie, velocidad en la respiración, etc.
- En el enmascaramiento, el ruido antropogénico interfiere con sonidos importantes para los mamíferos marinos como señales de ecolocalización, sonidos de presas o depredadores.
- La fatiga auditiva se produce cuando a cierto nivel y duración, las células ciliadas del oído interno se fatigan, lo cual genera un incremento en el umbral auditivo (cambio temporal del umbral, o TTS por sus siglas en inglés). Después de cierto tiempo (de minutos a días), la audición regresa a la normalidad.
- En la zona de lesión, en lugar de generarse un cambio temporal del umbral, la audición no regresa del todo, un umbral se queda permanentemente (cambio permanente del umbral, o PTS por sus siglas en inglés). Sin embargo, los efectos fisiológicos no solo se dan en la audición, otros lesiones potenciales son en el sistema vestibular, el sistema nervioso, el reproductivo, en tejidos, entre otros.

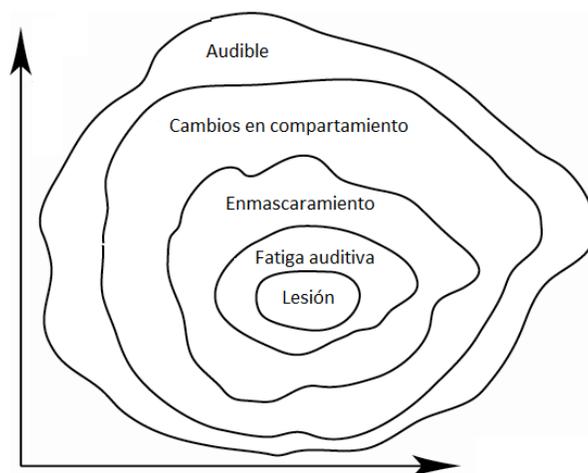


Fig. 2-11. Zonas de influencia alrededor de una fuente de ruido.

2.3.Regulaciones internacionales

En Europa los océanos Atlántico y Ártico, el Mar Báltico, el Mar del Norte, el Mar Mediterráneo y el Mar Negro, bañan sus 70,000 km de costa. El 90% del comercio exterior y el 40 % del comercio interior de la UE se desarrollan por mar (EUR-Lex El acceso al Derecho de la Unión Europea, 2015).

Por lo que desde el 2002, la Comisión Europea se comprometió a definir una estrategia con el objetivo general de promover la utilización sostenible de los mares y proteger los ecosistemas marinos. Así en 2008 se publicó la Directiva 2008/56/CE del Parlamento Europeo y del Consejo (Council Directive 2008/56/EC), por la que se establece un marco de acción comunitaria para la política del medio marino (Directiva marco sobre la estrategia marina). Ésta pretende establecer un marco en el que los Estados miembros deberán adoptar las medidas necesarias para conseguir un buen estado medioambiental del medio marino en el año 2020.

El Descriptor 11 lleva como título: “La introducción de energía, incluido el ruido subacuático, se sitúa en niveles que no afectan de manera adversa al medio marino”, en él se establece:

- La distribución temporal y espacial de los ruidos impulsivos de alta, baja y media frecuencia: Proporción de días y su distribución a lo largo de un año natural en zonas de una determinada superficie, así como su distribución espacial, en los que las fuentes sonoras antropogénicas superen niveles que puedan producir en los animales marinos un significativo impacto, medidos en la banda de frecuencias de 10 Hz a 10 kHz como nivel de exposición sonora (en dB re 1 $\mu\text{Pa}^2\text{s}$) o como nivel de presión acústica de pico (en dB re 1 $\mu\text{Pa}_{\text{pico}}$) a un metro.
- El ruido continuo de baja frecuencia: Evolución del nivel de ruido ambiental en las bandas de 1/3 de octava 63 y 125 Hz (como frecuencia central) (re 1 μPa RMS; nivel de ruido medio en estas bandas de octavas a lo largo de un año), medido por estaciones de observación o, si procede, haciendo uso de algún modelo.

En México existe el Comité de Protección del Medio Marino (CPMM), el cual está constituido por todos los Estados miembros, está facultado para examinar asuntos de la competencia de la Organización Marítima Internacional (OMI), que se relacionen con la prevención de la contaminación del mar; sin embargo, solo a la ocasionada por buques (contaminación por hidrocarburos, naufragios, etc.).

2.4.Sistemas de clasificación automática

En la última década se han estudiado varios métodos para la clasificación automática de vocalizaciones de mamíferos marinos, el interés en este campo de investigación ha crecido debido a la necesidad de describir objetivamente las vocalizaciones y ser capaz de hacerlo de forma rápida y efectiva (Pace et al., 2010).

Sin embargo, el repertorio vocal de los mamíferos marinos es extremadamente variable de especie a especie, y en algunos cetáceos, los “llamados” intraespecíficos son muy complejos. Además es bien sabido que las señales subacuáticas tienen generalmente una estructura compleja, varias componentes no lineales en tiempo y frecuencia, mezcladas frecuentemente en ruido no estacionario arbitrario (ver sección 4.3).

El primer paso para una caracterización, es identificar el segmento de tiempo en el que ocurre un evento de interés. Esto puede hacerse manualmente, teniendo una persona para escuchar el sonido y analizar su espectrograma, e identificar el inicio y fin del segmento; automáticamente, con una computadora; o semiautomáticamente, con una computadora identificando segmentos “candidatos” y una persona manualmente verifica cada uno para determinar si se trata de un evento o no.

Algunos sistemas de clasificación tras haber identificado el segmento de tiempo, se calcula un espectrograma del segmento, y se marca una caja de anotación; la caja de anotación consiste de una caja que delimita la región en tiempo y frecuencia, la cual contiene el sonido de interés. Esta caja también puede ser elegida de forma manual o automática (Mellinger & Bradbury, 2007).

La mayoría de los trabajos sobre clasificación automática del ruido generado por embarcaciones, se han realizado en el dominio de la frecuencia; usando el espectro de potencia y correlación cruzada (Soares-Filho et al., 2002; Howell et al., 2003; Kang et al., 2004; Chung et al., 2011), modelos autorregresivos (Eom et al., 1997; Huang et al., 1997) y transformada wavelet (Eom et al., 1997; Chen et al., 1998).

Mientras que para la detección de especies, las primeras técnicas automáticas de detección usaban filtros adaptados (“matched filters”), modelos ocultos de Markov, y correlación cruzada del espectrograma (Clark et al., 1987).

Existen trabajos (Oppenheim et al., 1976; Roch et al., 2007; Brown & Smaragdis, 2009) en los que se usan coeficientes cepstrales para caracterizar en su mayoría la información del tono de diferentes vocalizaciones. También se han usado características de percepción auditiva, contornos en frecuencia, entre otras (Gillespie, 2004; Brown & Miller, 2007; Urazghildiiev et al., 2008; Ma & Chen, 2008; Urazghildiiev et al., 2009).

2.4.1. Sistemas basados en aprendizaje automático

Los métodos anteriores fueron mejorados por medio del aprendizaje automático (“machine learning”), usando clasificadores basados en redes “feedforward” (Mellinger & Clark, 1993; Potter et al., 1994; Deecke et al., 1999; Mellinger, 2004; Mazhar et al., 2007; Pourhomayoun et al., 2013). También se han implementado otros algoritmos, como árboles de clasificación y regresión (CART, por sus siglas en inglés) para el reconocimiento de vocalizaciones de la ballena franca (Dugan et al., 2010a; Dugan et al., 2010b); en estos se suelen usar técnicas avanzadas, en las que se combinan varios métodos de reconocimiento corriendo en paralelo (Pourhomayoun et al., 2013).

Estado del arte

Las ballenas son las especies más estudiadas, principalmente el cachalote; esto se debe a sus capacidades únicas para comunicarse. Es muy común el uso técnicas desarrolladas para el análisis de voz, la utilidad de estas técnicas radica en la similitud entre la voz humana y los cantos de las ballenas: ambas ocupan una banda de frecuencia similar, y ambas presentan segmentos tonales y segmentos de banda ancha.

Abousleiman et al. (2013) desarrollaron un algoritmo para preprocesar el sonido antes de pasar a un clasificador jerárquico basado en árboles, su objetivo es determinar la presencia de la ballena franca. El clasificador identifica un solo tipo de vocalización hecho por la ballena, dicha vocalización es mejor conocida como “contact call” o “up call”, este método logra un 85% de clasificación.

Zaugg et al. (2010) desarrollaron 4 módulos autónomos, para la clasificación en tiempo real de clics provenientes del cachalote y de sonidos impulsivos producidos por barcos (ver Fig. 2-12). Los 2 primeros módulos detectan los segmentos que contienen los sonidos impulsivos dentro de bandas de frecuencia específicas (1-5 y 5-20 kHz), por cada impulso los clasificadores calculan la probabilidad estimada de que el impulso fuera producido por un cachalote, y devuelve las posiciones (dentro del segmento) de los impulsos.

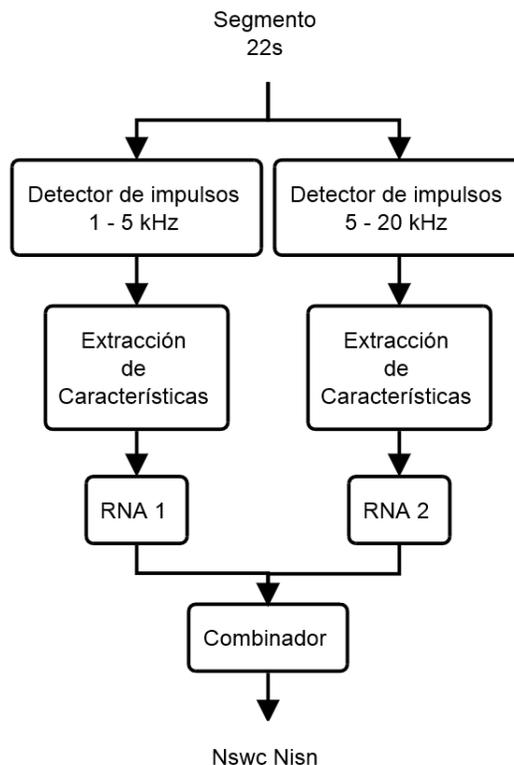


Fig. 2-12. Arquitectura usada para el reconocimiento de cachalotes y barcos.

Los 2 módulos restantes clasifican los impulsos detectados como clics del cachalote o impulsos del barco, extrayendo características como la media, la desviación estándar, curtosis, etc., de la señal en tiempo y frecuencia. Y usando 2 redes neuronales de base radial (una por cada banda), las salidas se combinaban y se escogía el máximo de las 2 redes (ecuaciones (1) y (2)).

$$N = \underset{i}{\text{mediana}}(P_i) \sum_{i:P_i>0.5} P_i \quad (1)$$

donde P_i representa la probabilidad estimada de que fuera un clic de cachalote; y la probabilidad de que fuera un sonido del barco se obtenía de $1 - P_i$.

$$N_{swc} = \max(N_1, N_2) \quad (2)$$

La ecuación (2) determina que es suficiente encontrar clics de cachalote en cualquiera de las 2 bandas, como para asegurar que el segmento de 22 s contiene clics. Para los impulsos del barco, solo se combinaban los valores de $1 - P_i$.

André et al. (2011) hicieron básicamente el mismo procedimiento, pero en lugar de clics de cachalote, clasificaban clics ultrasónicos de cetáceos. Los 2 trabajos obtuvieron buenos porcentajes de reconocimiento con 2 grandes bloques de prueba: André et al. (2011) obtuvieron un 97-100% y Zaugg et al. (2010) un 94-98%, con cada bloque de prueba respectivamente.

2.4.2. Aplicaciones dedicadas a la detección y localización de cetáceos

Existen varias aplicaciones fáciles de usar y brindan buenos resultados, estos están escritos en Fortran, C o Matlab. A continuación se mencionan los más importantes:

- “PIMC Track” (Glotin, 2015): Este software permite detectar clics del cachalote y localizarlos simultáneamente en tiempo real (Giraudet & Glotin, 2006).
- “Chrisar-Lab / Bio” (CHRISAR, 2015): Funciona para varias aplicaciones acústicas subacuáticas y uno de sus componentes está dedicado al análisis de sonidos biológicos, especialmente sonidos emitidos por cetáceos. Este software contiene una herramienta de análisis bioacústico (el usuario puede emplear filtros digitales y correlación cruzada), y puede ser usada para grabaciones con múltiples canales, ya sea en tiempo real o en tiempo diferido.
- “RainbowClick” (Marine conservation research, 2015): Está dedicado a la detección automática en tiempo real de clics del cachalote; es posible clasificar clics de otras especies, como la marsopa común. Puede ser usado para rastrear cachalotes en tiempo real o diferido.

- “XBat” (XBAT, 2015): Desarrollado en Matlab, funciona para la detección de sonido, basándose en filtros adaptados. Se puede usar correlación sobre los espectrogramas y pueden aplicarse filtros digitales antes de analizar la señal grabada.
- “PAMGuard” (PAMGUARD, 2015): En el 2005, un consorcio de laboratorios científicos y compañías petroleras, crearon este software dedicado al PAM de cetáceos. PAMGuard es soportado por JIP (“Sound and Marine Life Joint Industry Programme”) y OGP (“International Association of Oil and Gas Producers”). PAMGuard cuenta con una versión “open source”, que detecta, clasifica y localiza los sonidos de cetáceos; para la detección se hace una extracción de características del espectrograma, el filtro adaptado se basa en la correlación de los espectrogramas (Gillespie et al., 2009). En la versión open source, los investigadores pueden proponer añadir sus propias funciones.
- “Avisoft” (Avisoft Bioacoustics, 2015): Es conocido por el análisis de sonidos biológicos y fue usado originalmente por biólogos para las observaciones de aves. El software puede ser usado para el análisis de sonido (espectrograma en tiempo real, filtrado digital), análisis de tren de pulsos, editar, clasificar (correlación cruzada del espectrograma), y cuenta con un sintetizador para generar cantos y llamados artificiales.
- “Raven” (The Cornell Lab of Ornithology/Raven, 2015): Raven fue desarrollado por el programa de investigación bioacústica del laboratorio de ornitología de la Universidad de Cornell. Este software puede ser usado para la adquisición de datos, realizar mediciones, análisis de sonidos y visualización del espectrograma.
- “ROCCA”: Oswald et al. (2011) desarrollaron ROCCA, el cual está incorporado a PAMGuard; es un paquete o módulo que permite medir 54 características del contorno de los silbidos y es capaz de clasificar silbidos de 7 especies (*Globicephala Macrorhynchus*, *Pseudorca Crassidens*, *Steno Bredanensis*, *Stenella Attenuata*, *Stenella Coeruleoalba*, *Stenella Longirostris*, y *Tursiops Truncatus*) y 1 género (*Delphinus*). Su clasificador está basado en análisis por Random Forest (bosques aleatorios); entrenado con las 54 características dio como resultado un porcentaje de clasificación del 62%.

2.5.Conclusiones

En este capítulo se presentaron los diversos sistemas que se usan en la actualidad para el monitoreo acústico pasivo en los mares. Se mencionaron los efectos negativos de los ruidos antropogénicos sobre los animales marinos, lo que ha hecho que se establezcan regulaciones y observatorios para el control de este tipo de emisiones. Se revisaron los trabajos más importantes, con el fin de conocer qué técnicas se han implementado y cuáles son los objetivos actuales en este campo. También se mencionaron aplicaciones que emplean algoritmos computacionales para el reconocimiento principalmente de cetáceos. Si bien es un campo que ya tiene mucho tiempo, aún queda mucho por hacer, especialmente en México.

3. Marco teórico

3.1. El sonido en el mar

El sonido consiste de vibraciones mecánicas que viajan como una onda, el sonido no se propagaría en el vacío, ya que requiere de algún medio para poder transmitirse. El medio de propagación puede ser tanto un sólido como un fluido (Au & Hastings, 2008).

El sonido viaja aproximadamente 1,500 m/s en agua de mar, mientras que en el aire, a unos 340 m/s. Sin embargo, la velocidad del sonido en agua de mar no es un valor constante, y aunque las variaciones en la velocidad del sonido no son altas, tienen efectos importantes sobre cómo el sonido viaja en el océano. La velocidad del sonido c [m/s] en el mar depende de la temperatura T [°C], salinidad S [ppm] y profundidad D [m], pudiendo ser expresada como sigue (Mackenzie, 1981).

$$c = 1448.96 + 4.591T - 5.304 \times 10^{-2}T^2 + 2.374 \times 10^{-4}T^3 + 1.34(S - 35) + \dots \quad (3)$$
$$1.63 \times 10^{-2}D + 1.675 \times 10^{-7}D^2 - 1.025 \times 10^{-2}T(S - 35) - 7.139 \times 10^{-13}TD^3$$

Esta ecuación es válida para los siguientes rangos:

$$0 < T < 30 \text{ } ^\circ\text{C}$$

$$30 < S < 40 \text{ ppm}$$

$$0 < D < 8,000 \text{ m}$$

En latitudes medias en las profundidades del océano, la velocidad de sonido más baja se produce a una profundidad de aproximadamente 1 km. El sonido se curva o refracta hacia la región donde la velocidad del sonido es más baja creando un canal de sonido en el que las ondas sonoras pueden viajar largas distancias. Este se conoce como el canal de sonido profundo o canal de sonido de fijación y oscilación sonora (SOFAR, por sus siglas en inglés), ver Fig. 3-1.

3.1.1. Importancia

El sonido recorre distancias mucho mayores que la luz bajo el agua. La luz viaja solo unos cientos de metros en el océano antes de que se absorba o disperse. Incluso donde hay luz es más difícil ver a lo lejos bajo el agua comparado con el aire, limitando la visión en el ambiente marino. Además de la visión, muchos animales terrestres dependen en gran medida de las señales químicas y el sentido del olfato para las funciones importantes de la vida (como marcar límites territoriales). Algunos peces utilizan el olfato para detectar arrecifes cercanos, pero en general, el olfato es menos importante para muchas especies marinas que para los mamíferos terrestres. Muchos animales marinos dependen del sonido para la supervivencia y dependen de adaptaciones que les permiten percibir su entorno acústicamente, comunicarse, buscar comida, y protegerse bajo el agua.

Marco teórico

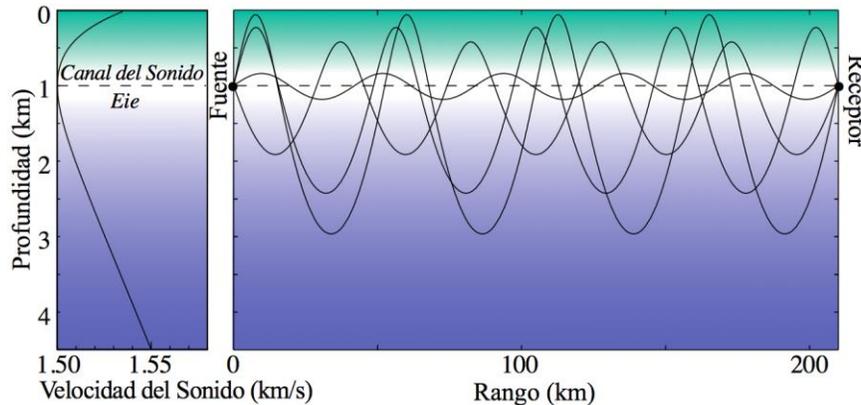


Fig. 3-1. Canal SOFAR.

En el agua, las ondas de sonido son medidas típicamente por su presión, usando un transductor llamado hidrófono. El sonar (“sound navigation and ranging” o SONAR, por sus siglas en inglés) es una tecnología que utiliza ondas sonoras para identificar los objetos y sus ubicaciones en el océano. Existen dos tipos de sonar: activo y pasivo. El sonar pasivo utiliza equipo que solo escucha o recibe el sonido. Las ondas sonoras producidas por una fuente de sonido se reciben y puede utilizarse para identificar o localizar el objeto que emite el sonido. El sonar activo, por otro lado, envía ondas de sonido y luego escucha (recibe) el eco de retorno. El eco de retorno puede utilizarse para identificar el tipo de objeto (ballena, roca, barco) que refleja el sonido así como su alcance y velocidad.



Fig. 3-2. Funcionamiento de un sonar activo.

Las temperaturas del océano pueden medirse utilizando el sonido. La mayoría de los cambios en la velocidad del sonido en el océano abierto son debidos a los cambios de temperatura. El sonido viaja más rápido en aguas más cálidas y más lento en aguas frías.

Marco teórico

Para medirla, un pulso de sonido es enviado desde una fuente de sonido subacuática y escuchado por un dispositivo de audición (hidrófono) en el agua a una distancia conocida (hasta miles de kilómetros). Midiendo con precisión el tiempo que tarda el sonido en viajar desde la fuente al receptor, puede calcularse su velocidad y con esto relacionarla con la temperatura del agua entre la fuente y el hidrófono. La medida de temperatura de agua de esta manera es muy eficaz y proporciona datos útiles para la comprensión de las corrientes oceánicas y estudiar cambios climáticos.

3.1.2. Ruido ambiental

El sonido total es, por el principio de la superposición, una combinación lineal de campos de presión originados por fuentes individuales. Una o más de esas fuentes pueden constituir una señal de interés, donde el resto es ruido no deseado. La distinción entre señal y ruido depende de la aplicación, uno primero debe decidir las características de las fuentes que producen ruido (Carey & Evans, 2011).

El sonido submarino es generado por una variedad de fuentes naturales, como el romper de las olas, la lluvia y la vida marina. También se produce por una variedad de fuentes hechas por el hombre, como barcos y sonares militares. El sonido de fondo en el océano se llama ruido ambiental.

Las principales fuentes de ruido ambiental pueden clasificarse por la frecuencia del sonido. En el rango de frecuencia de 20-500 Hz, el ruido ambiental en el hemisferio norte es principalmente debido al ruido generado por barcos cargueros que se encuentran a lo lejos. En el rango de frecuencia de 500-100,000 Hz, el ruido ambiental es principalmente debido a burbujas y rocío asociado al rompimiento de las olas (Scowcroft et al., 2012).

3.1.3. Niveles de sonido en aire y en agua

Las ondas de sonido en el agua y las ondas de sonido en el aire son fundamentalmente similares; sin embargo, la forma en que se reportan los niveles de sonido es muy diferente. La intensidad relativa de sonido en el agua en decibeles (dB) no puede ser comparada directamente a la intensidad relativa del sonido en dB en el aire.

Para evitar confusiones, debemos especificar que los sonidos en el agua, un medio más denso, se midieron relativos (re) a 1 microPascal (μPa) y que los sonidos en el aire se midieron re 20 μPa . Para hacer la distinción se suele usar “dB submarino” para sonidos submarinos.

Existe una relación entre sonoridad y decibeles (dB): un aumento de 10 dB de intensidad es percibido por las personas como el doble de la sonoridad de un sonido.

Según el sonido viaja lejos de la fuente, la intensidad disminuye porque las ondas de sonido se dispersan (extendiendo la pérdida) y porque parte del sonido es absorbido por el agua de mar. Por ejemplo, la intensidad de un rayo que mide 260 dB submarinos a

Marco teórico

un metro, disminuiría a 220 dB submarinos a una distancia de 100 m. Los sonidos de alta frecuencia no viajan tan lejos a través del océano como los sonidos de baja frecuencia porque los de alta frecuencia son absorbidos rápidamente.

Tabla 2. Ejemplos de niveles de sonido submarino.

Fuente	Nivel de sonido (dB submarino a 1 m)
Rayo	260
Ballena jorobada	144-174
Cangrejo pistola	183-189 (tope a tope)
Silbido de delfín	125-173
Clics de cachalote	236
Buque-tanque grande	186
Sonar naval de frecuencia media	235
Exploración de petróleo y gas utilizando armas de aire comprimido	259 (tope)

3.2. Técnicas de procesamiento de voz

En la extracción de características se buscan rasgos cuyos valores sean similares para objetos de una clase similar, y diferentes de los valores de cualquier otra clase (o del “background”). Existen diferentes técnicas utilizadas para la extracción de características, a continuación se mencionan algunas técnicas comúnmente empleadas en procesamiento de voz, pero con las cuales se han obtenido buenos resultados en la clasificación de otros tipos de señales acústicas (Márquez-Molina et al., 2014; Sánchez-Pérez et al., 2013).

3.2.1. Lpc Codificación predictiva lineal (LPC)

El modelo de filtro “todo polos” es el más utilizado y el análisis de auto recursividad de la señal es la herramienta para hallar los coeficientes del filtro en cuestión. La funcionalidad matemática que da como resultado los coeficientes del filtro, es una relación matricial que se obtiene al realizar la predicción lineal de las muestras de las señales, con lo cual se obtienen los coeficientes de predicción lineal (LPC, por sus siglas en inglés), también llamados parámetros LPC, que no son otra cosa que los coeficientes de un filtro de “todo polos”. Para obtenerlos se realiza un procedimiento en el dominio del tiempo, y está basado en la consideración de que el término n ésimo $s(n)$ de una secuencia de valores en el tiempo, puede ser estimado mediante una combinación lineal a partir de los p términos anteriores (Atal & Hanauer, 1971; Makhoul, 1976).

$$\hat{s}(n|S_{n-1}) = \sum_{k=1}^p a_k s(n-k) \quad (4)$$

El cálculo del error de predicción $e(n)$ está dado por

$$e(n) = s(n) - \hat{s}(n|S_{n-1}) \quad (5)$$

Para obtener los coeficientes LPC $\{a_k\}$ se definen los segmentos de señal y de error en un segmento de datos, y los tiempos en función de n ; lo que se busca es minimizar la señal de error mínimo cuadrado (E) en el intervalo. Cabe destacar que el error disminuye al aumentar el número de datos considerados en la predicción (k).

$$E = \sum_n \left(s(n) - \sum_{k=1}^p a_k s(n-k) \right)^2 \quad (6)$$

Para minimizar E , se deriva respecto a cada una de las variables a_k , por lo que

$$\frac{\partial E}{\partial a_i} = 0 \quad (7)$$

para $i = 1, 2, \dots, p$.

$$\frac{\partial}{\partial a_i} \sum_n \left(s(n) - \sum_{k=1}^p a_k s(n-k) \right)^2 = 0 \quad (8)$$

$$\sum_{k=1}^p a_k \sum_n s(n-i)s(n-k) = \sum_n s(n)s(n-i) \quad (9)$$

El método de autocorrelación sugiere una manera simple y eficiente de definir los límites de las sumas al asumir que los segmentos de voz $s(n)$ son nulos fuera del intervalo $0 \leq n \leq N - 1$. La función de autocorrelación permite medir matemáticamente el parecido entre una señal y una versión retrasada en el tiempo de la misma señal. El pico mayor de la función de autocorrelación está localizado en la posición cero de desplazamiento. La localización del siguiente pico, máximo global de la función tras el primer cruce por cero, da una estimación del periodo, y la altura una indicación de la periodicidad de la señal. La ecuación (9) puede ser escrita en términos de la autocorrelación

$$\sum_{k=1}^p a_k R(i-k) = R(i) \quad (10)$$

donde $i = 1, 2, \dots, p$.

La ecuación (10) se puede expresar como

$$\begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(0) & \dots & R(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ R(p-1) & R(p-2) & \dots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(p) \end{bmatrix} \quad (11)$$

Esta matriz $p \times p$ valores de autocorrelación es una matriz tipo Toeplitz, donde p es el número de coeficientes o parámetros LPC que se quieren calcular; la cual se puede resolver mediante el método de Levinson-Durbin (Durbin, 1960; Levinson, 1947), obteniendo los parámetros LPC, los cuales permiten aproximar la señal original minimizando el error cuadrático medio.

3.2.2. Coeficientes cepstrales en las frecuencias de Mel (MFCC)

Los coeficientes MFCC se utilizan ampliamente en el reconocimiento automático de voz, se derivan de la magnitud del espectro de la transformada de Fourier, mediante la aplicación de un banco de filtros espaciados en la escala de frecuencia de Mel. Se calcula el logaritmo de la energía en cada filtro y se suma esta energía antes de aplicar la transformada discreta de coseno (DCT) para obtener el vector de características MFCC (Shannon et al., 2003).

La escala de frecuencia utilizada para la separación de los filtros es la escala de Mel, esta escala es una escala perceptual que fue propuesta por Stevens et al. (1937). La escala se diseñó a través de experimentos de la percepción humana, donde se pedía a los sujetos ajustar un tono de estímulo perceptualmente a la mitad de separación de un tono de referencia. La escala resultante fue una en la que 1 Mel representa una milésima parte del tono de 1 kHz (Huang et al., 2001) y una duplicación de Mel's produce una duplicación de percepción del tono (Zheng et al., 2001). El cálculo de la MFCC incluye los siguientes pasos:

- 1) Se aplica la transformada discreta de Fourier al segmento de la señal ventaneada, para obtener su respuesta en el dominio de la frecuencia, obteniendo el espectro de potencia $P(f)$.

$$P_k(f) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi k/n} \quad (12)$$

- 2) El espectro $P(f)$ es mapeado a lo largo del eje de frecuencia f [Hz] hacia el eje de frecuencia de Mel $P(B)$, usando la ecuación (13) (Picone, 1993; Schroeder, 1977), donde B es la frecuencia de Mel.

$$B(f) = 1125 \log_{10} \left(1 + \frac{f}{700} \right) \quad (13)$$

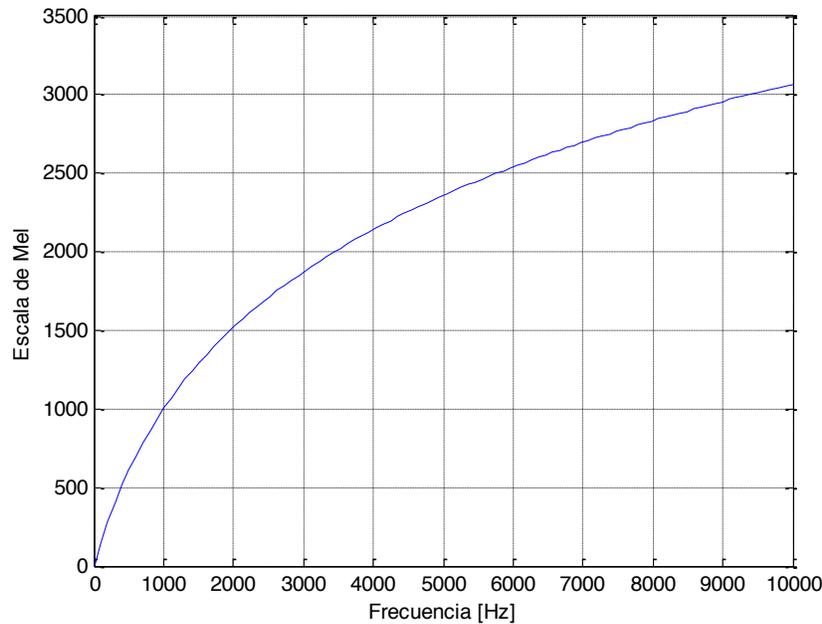


Fig. 3-3. Relación entre la escala de frecuencia en Hz y la escala de Mel.

En la Fig. 3-3 se puede ver el comportamiento de la ecuación (13), el cual es aproximadamente el reflejo de la percepción del oído humano.

- 3) El espectro de potencia resultante se convoluciona con un filtro triangular pasa banda $P(B)$ en $\theta(B)$. La convolución con las curvas de enmascaramiento, ecuación (14), reduce significativamente la resolución espectral de $\theta(B)$ en comparación con la original $P(f)$, que permite la disminución de resolución de $\theta(B)$. La Fig. 3-4 muestra un ejemplo de la respuesta en frecuencia del banco de filtros triangulares con amplitud normalizada, definidos por la ecuación (14).

Marco teórico

$$H_m(k) = \begin{cases} 0, & k < f(m-1) \\ \frac{k - f(m-1)}{f(m) - f(m-1)}, & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)}, & f(m) \leq k \leq f(m+1) \\ 0, & k > f(m+1) \end{cases} \quad (14)$$

donde $1 \leq k \leq f_s/2$ y $1 \leq m \leq M$, la cantidad de filtros es representada por M .

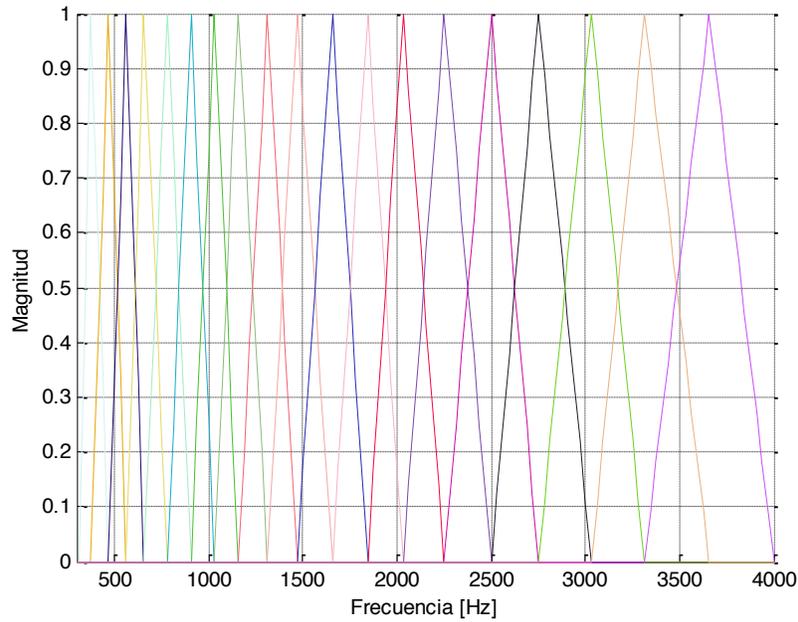


Fig. 3-4. Respuesta en frecuencia del banco de filtros con amplitud normalizada.

- 4) Mediante la convolución discreta de H_m con $\theta(B)$ se obtienen muestras del espectro de potencia para cada banda crítica, y posteriormente se obtiene una salida logarítmica de energía, dada por la ecuación (15).

$$X(m) = \log_{10} \left(\sum_{k=1}^{N-1} |\theta(k)|^2 H_m(k) \right) \quad (15)$$

donde $i = 1, 2, \dots, M$.

- 5) Los coeficientes MFCC son obtenidos mediante la transformada coseno de Fourier, (16).

$$c_i = \sum_{i=1}^M X_i \cos\left(j\left(i - \frac{1}{2}\right)\frac{\pi}{M}\right) \quad (16)$$

donde $j = 1, 2, \dots, J$, M es el número de filtros y J es el número de coeficientes que se desean calcular.

3.2.3. Análisis de octava

Una octava es el intervalo entre dos frecuencias que tienen una proporción de 2:1. En el análisis de octava, la señal a ser analizada es dividida en señales parciales, para después poder obtener el nivel de sonido por bandas de frecuencia y no del espectro completo, reduciendo así la resolución del espectro pero reduciendo significativamente la cantidad de valores que funcionarán como rasgos característicos. Para lograr ello, se usa un banco de filtros con varios filtros conectados en paralelo.

Cada filtro se caracteriza por tener:

- Frecuencia de corte inferior (f_l)
- Frecuencia de corte superior (f_u), la cual al ser un filtro de octava será el doble de la frecuencia de corte inferior
- Ancho de banda (Δf), que es la diferencia entre f_l y f_u ($\Delta f = f_l - f_u$)
- Frecuencia central (f_c) que se extiende de f_l a f_u y que divide a la octava en dos intervalos iguales en la escala logarítmica. La f_c está determinada por la ecuación (17).

$$f_c = \sqrt{f_l f_u} \quad (17)$$

Dado que se conocen las características del filtro $f_u = 2f_l$, entonces $f_c = \sqrt{2}f_l$ y el ancho de banda del filtro de una octava es

$$\Delta f = 2f_l - f_l = f_l \quad (18)$$

$$\Delta f = \frac{f_c}{\sqrt{2}} \quad (19)$$

Cuando se necesita información más detallada del espectro que la obtenida con un análisis de octava “normal”, se hace un análisis de $1/n$ de octava, lo cual subdividirá cada banda de octava en n partes. Por ejemplo para un tercio de octava (el cual es un tipo de análisis de octava ampliamente usado en análisis acústico y procesamiento de señales de audio), la frecuencia de corte superior del filtro pasa banda es $f_u = \sqrt[3]{2}f_l$, la frecuencia

Marco teórico

central sería $f_c = \sqrt[6]{2}f_l = 1.12f_l$, de esta manera el ancho de banda de cada filtro será $\Delta f = 0.26f_l$.

Las bandas de frecuencia son equidistantes en una escala de frecuencia logarítmica, por el otro lado en una escala de frecuencia lineal, el ancho de banda incrementará logarítmicamente.

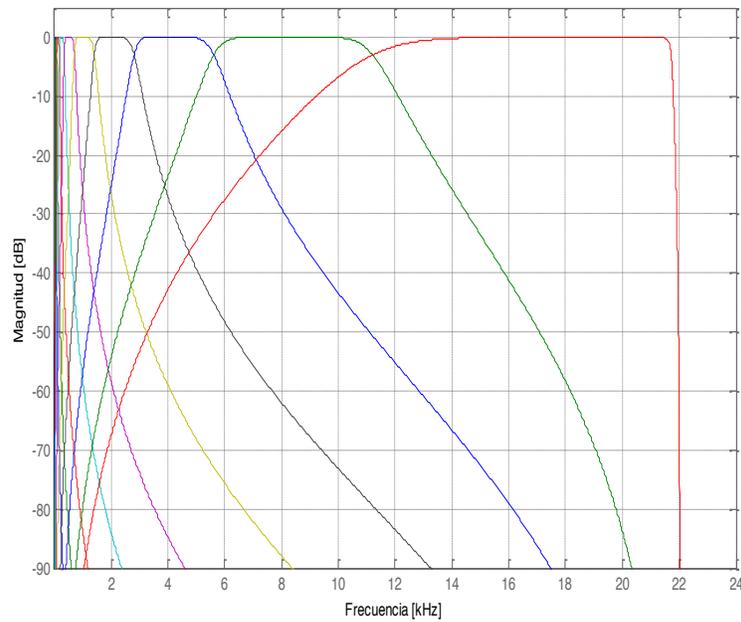


Fig. 3-5. Respuesta en frecuencia del banco de filtros del análisis de octava en una escala de frecuencia lineal.

Las frecuencias de corte inferior y superior de los filtros están estandarizadas por normas internacionales como ANSI S1.11-2004 e IEC 1260:1995 (ANSI, 2004; IEC, 1995).

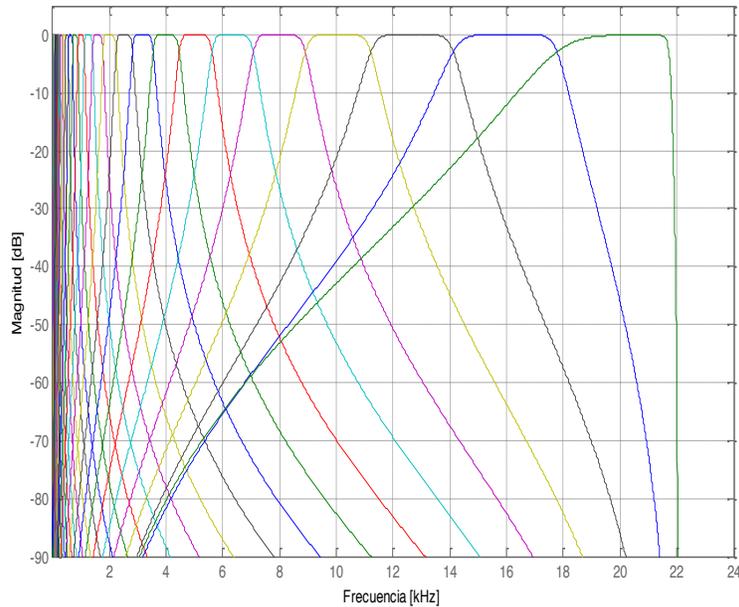


Fig. 3-6. Respuesta en frecuencia del banco de filtros del análisis de 1/3 de octava en una escala de frecuencia lineal.

3.3.Redes neuronales artificiales (RNAs)

Las redes neuronales artificiales (RNAs, por sus siglas en inglés) están basadas en el sistema nervioso humano; las señales de los sensores en el cuerpo transmiten la información de ver, escuchar, oler, temperatura, etc. Las neuronas son células autónomas especializadas en transmitir, procesar y almacenar esta información, la cual es esencial para nosotros, para responder de forma adecuada a estímulos externos e internos. El cerebro está compuesto por una gran cantidad de neuronas trabajando en paralelo. Un modelo matemático de la neurona (llamado perceptrón [McCulloch & Pitts, 1943]) ha sido usado para probar e imitar el funcionamiento del cerebro, en particular su peculiaridad de procesamiento en paralelo, con el fin de emular su capacidad de reconocimiento de patrones.

Las RNAs pueden ser usadas en una amplia variedad de tareas de clasificación, como reconocimiento de carácter, reconocimiento de voz, detección de fraude, diagnósticos médicos, etc. La neurona artificial es una unidad lógica umbralada (TLU, por sus siglas en inglés) que acepta múltiples entradas, las cuales son ponderadas y sumadas; si la suma excede un umbral determinado, la perceptrón se dispara y un valor de salida (una probabilidad) es transmitido a la siguiente unidad (Dougherty, 2013).

3.3.1. Aprendizaje por cuantificación vectorial

El aprendizaje por cuantificación vectorial (LVQ, por sus siglas en inglés) es una colección de algoritmos de aprendizaje supervisado que entrenan un libro de códigos, estos libros consisten de una pequeña cantidad de vectores que corresponden a cada clase. En la Fig. 3-7 se muestran 157 vectores de entrada originales, los cuales se reducen a un libro de códigos de solo 16 vectores (ver Fig. 3-8).

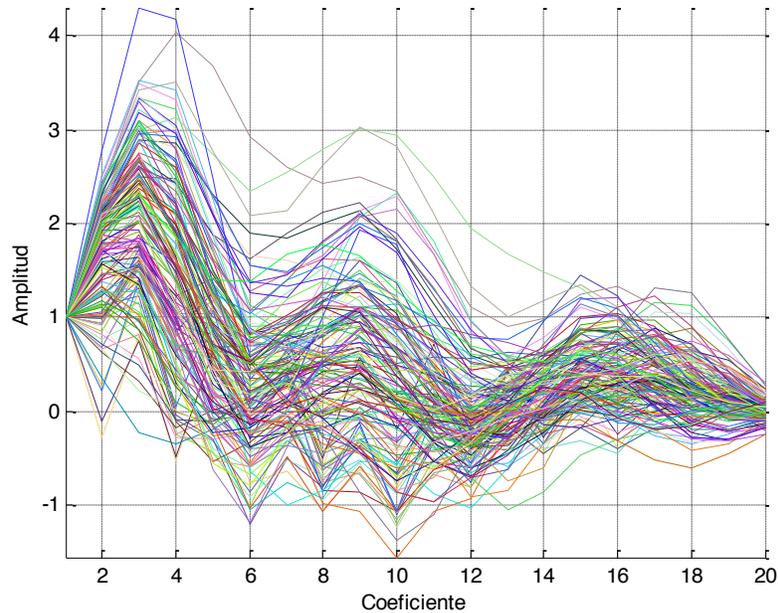


Fig. 3-7. Vectores de entrada originales.

Un vector pertenecerá a la clase en la cual se encuentre el vecino más cercano. Por lo que la ubicación de los vectores del libro, determinarán el éxito de la clasificación. Los algoritmos están diseñados para encontrar las ubicaciones óptimas de los vectores del libro.

Marco teórico

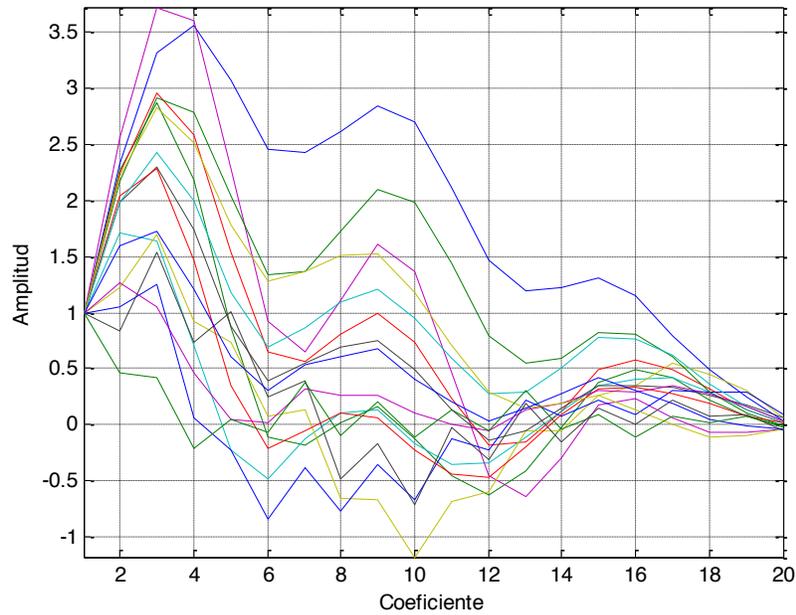


Fig. 3-8. Libro de códigos con 16 vectores.

Una red LVQ se compone de 2 capas, la primera es una capa competitiva, y la segunda es una capa lineal (ver Fig. 3-9). El número de neuronas en la primera capa (S^1) debe ser mayor o al menos igual que el número de neuronas de la segunda capa (S^2).

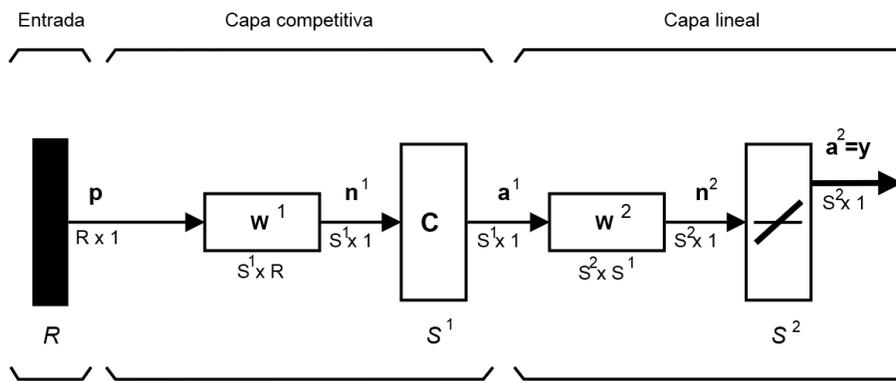


Fig. 3-9. Red LVQ.

donde R es el número de elementos en el vector de entrada.

La entrada de la primera capa se expresa con la ecuación (20), es aquí donde se calcula la distancia directamente entre la entrada y el vector de pesos. La salida de la

primera capa viene dado por la ecuación (21), lo que hace que la neurona cuyo vector de pesos esté cercano al vector de entrada tenga salida 1 y las otras neuronas, tengan salida 0. La salida de la red viene dada por la ecuación (22), donde la matriz de pesos W^2 combina las subclases dentro de una sola, las columnas de esta matriz representan las subclases, y las filas representan las clases.

$$n_i^1 = -\| \mathbf{w}_i^1 - \mathbf{p} \| \quad (20)$$

$$\mathbf{a}^1 = \text{compet}(\mathbf{n}^1) \quad (21)$$

$$\mathbf{a}^2 = W^2 \mathbf{a}^1 \quad (22)$$

Una propiedad importante de esta red, es que el proceso de combinar subclases para formar clases, permite a la red LVQ crear clases más complejas. Una capa competitiva estándar tiene la limitación de que puede crear solo regiones de decisión convexas; la red LVQ soluciona esta limitación.

3.3.2. Perceptrón multicapa

En 1958, Frank Rosenblatt propuso el modelo del perceptrón, definiendo pesos numéricos y un patrón espacial de interconexión donde el aprendizaje toma lugar adaptando los pesos de la red mediante un algoritmo numérico. La red perceptrón multicapa (MLP, por sus siglas en inglés) consiste de un conjunto de neuronas organizadas por capas: una capa de entrada, una o más capas ocultas y una capa de salida (ver Fig. 3-10). A partir de múltiples entradas, el perceptrón calcula una salida única, formando una combinación lineal con sus pesos de entrada y estableciendo la salida con cierta función de activación, como se expresa en la ecuación (23) (Haykin, 1998).

$$y = f \left(\sum_{i=1}^n w_i x_i + b \right) \quad (23)$$

Antes de usar una red MLP, ésta se debe de entrenar para que una entrada en particular conduzca a un objetivo de salida específico. Para esto, se aplica un algoritmo de entrenamiento que adapta los valores de los pesos y polarizaciones, con base en un vector de entrada y su error asociado. El objetivo del proceso es hallar nuevos pesos y polarizaciones repetidamente hasta que el error se reduzca a un mínimo deseado. Las redes MLP se pueden entrenar usando el algoritmo de retro-propagación (Rumelhart et al., 1986).

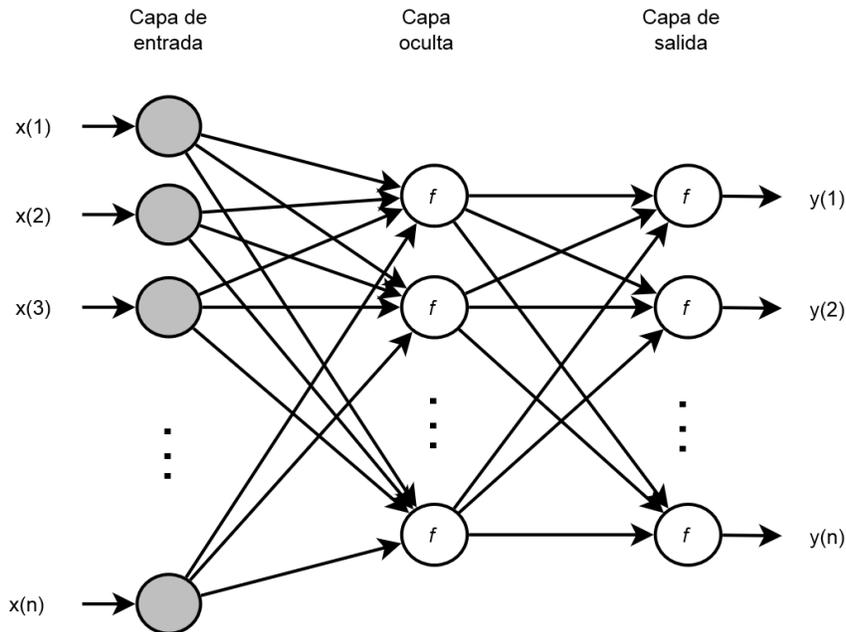


Fig. 3-10. MLP con una capa oculta.

3.4. Clasificador mediante conjuntos

La idea principal de una metodología en conjunto, también conocida como decisión por comité, es la de ponderar varios clasificadores individuales y combinarlos, para obtener un clasificador que supere a cada uno de los clasificadores por separado. La idea del conjunto en un aprendizaje supervisado ha sido investigada desde finales de los setenta; Tukey (1977) sugirió combinar 2 modelos de regresión lineal, dos años después, Dasarathy & Sheela (1979) sugirieron particionar el espacio de entrada usando 2 o más clasificadores. El mayor avance se dio en los noventa, cuando Hansen & Salamon (1990) sugirieron un conjunto de RNAs (configuradas de forma similar) para mejorar el desempeño predictivo de una sola RNA.

Un conjunto típico para clasificación está compuesto por:

- Conjunto de entrenamiento. Usamos la notación A para denotar el conjunto de atributos de entrada que contienen n atributos: $A = \{a_1, a_2, \dots, a_n\}$ y y para representar la clase o el target.
- Inductor base. El inductor es un algoritmo de inducción que recibe un conjunto de entrenamiento y forma un clasificador, que representa una relación generalizada entre los atributos de entrada y los atributos target. Sea I un inductor, $M = I(A)$ representará un clasificador M que fue inducido por I sobre un conjunto de entrenamiento A .
- Generador de diversidad. Es responsable de generar los clasificadores.

- **Combinador.** También conocido como módulo de decisión, es responsable de combinar las salidas de varios clasificadores.

Un conjunto puede tener 2 tipos de estructuras: dependientes o independientes; según su estructura será cómo se construirán los conjuntos.

3.4.1. Conjunto con estructura dependiente

En un conjunto dependiente, la salida de un clasificador es usada en la construcción del siguiente clasificador. La ventaja de esto es que se usa el conocimiento generado en las iteraciones previas para el aprendizaje de las siguientes iteraciones.

Existen 2 métodos dentro de un conjunto dependiente (Provost & Kolluri, 1997):

- 1) “Incremental Batch Learning”. El algoritmo de aprendizaje usa el conjunto de entrenamiento actual junto al clasificador anterior para construir el siguiente. El clasificador construido en la última iteración es elegido como el clasificador final.
- 2) “Model-guided Instance Selection”. En este algoritmo, los clasificadores que fueron construidos en las iteraciones previas son usados para manipular el conjunto de entrenamiento para la siguiente iteración (ver Fig. 3-11). El método más conocido es el “boosting” (también conocido como “arcing” o “adaptive resampling and combining”), el cual es un método general para mejorar el desempeño de un aprendizaje débil (como en los árboles de decisión o las reglas de clasificación).

3.4.2. Conjunto con estructura independiente

En un conjunto independiente (ver Fig. 3-12), cada clasificador es construido de forma independiente y sus salidas son combinadas de alguna manera.

El conjunto de datos original es transformado en varios conjuntos con los cuales varios clasificadores son entrenados. Estos nuevos conjuntos pueden ser mutuamente excluyentes o pueden traslaparse. Un módulo de decisión arroja la clasificación final; como el módulo de decisión es usualmente independiente de los algoritmos de inducción, puede ser usado con diferentes inductores con sus respectivos conjuntos de datos, además de que se puede paralelizar fácilmente. Por lo que no solo mejora el desempeño de clasificación sino también reduce el tiempo total de ejecución.

El método más conocido dentro de una estructura independiente es el “bagging” (“bootstrap aggregating”). El método busca incrementar la precisión, creando un mejorado clasificador compuesto I^* , amalgamando las salidas de los clasificadores en una clasificación única.

Marco teórico

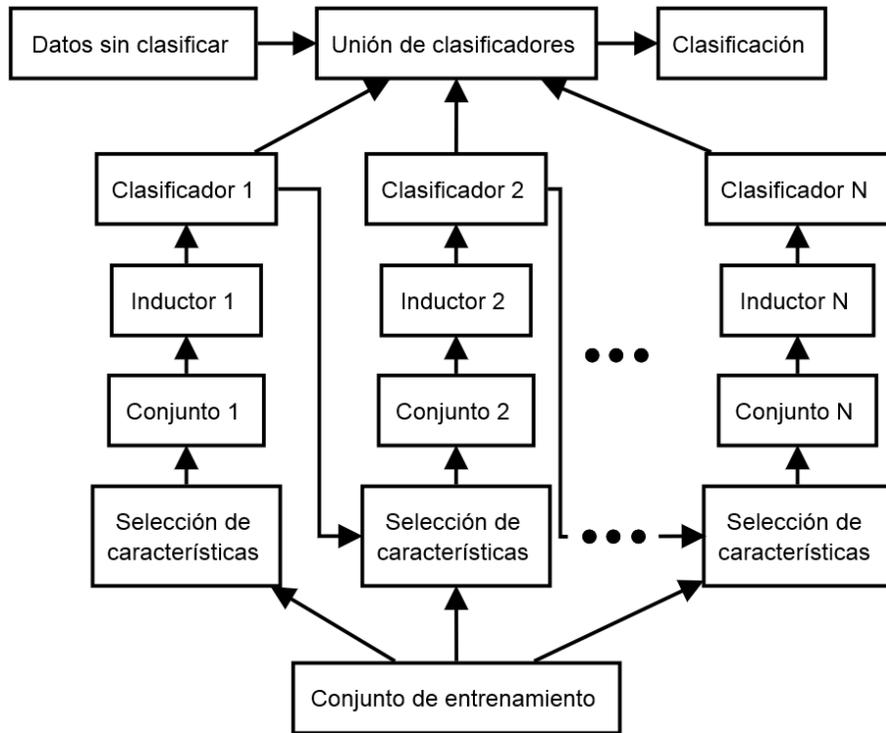


Fig. 3-11. Estructura dependiente.

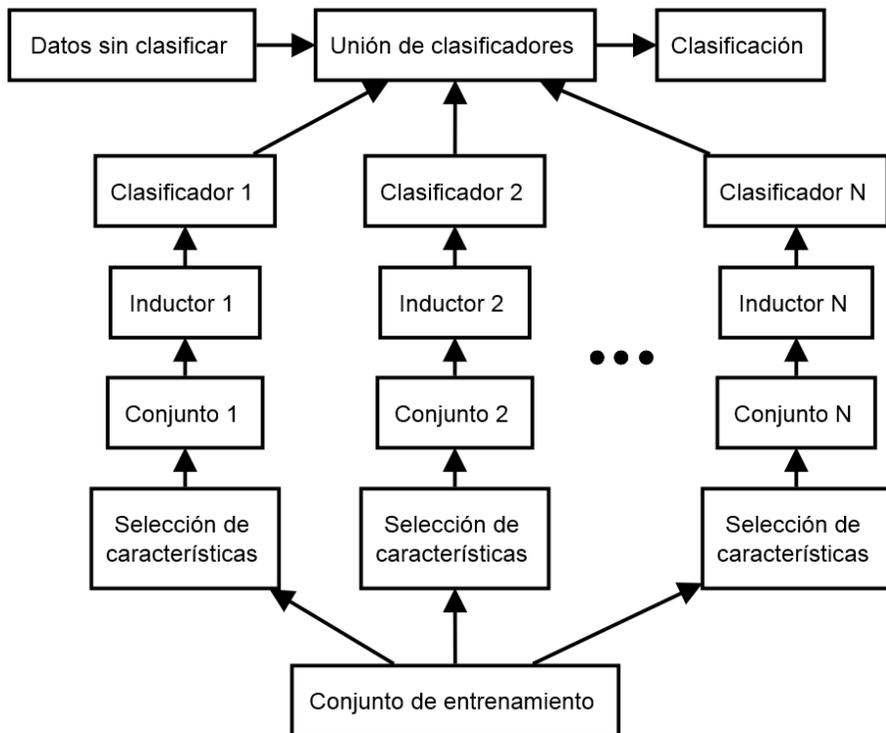


Fig. 3-12. Estructura independiente.

Marco teórico

Cabe mencionar que bagging entrena cada clasificador con una redistribución aleatoria del conjunto de entrenamiento (ver Tabla 3). Al igual que boosting, bagging es una técnica que mejora la precisión de un clasificador al generar un modelo compuesto que combina múltiples clasificadores (derivados del mismo inductor). Ambos métodos siguen un método por votos, con el fin de combinar las salidas de los diferentes clasificadores.

Además de bagging existen otros métodos como:

- “Wagging”, que es una variante de bagging en la que cada clasificador es entrenado con el conjunto de entrenamiento completo, pero a cada instancia se le asigna estocásticamente un peso.
- Bosques aleatorios, el cual usa una gran cantidad de árboles de decisión sin podar.
- Comités de validación cruzada, donde se generan k clasificadores al dividir el conjunto de entrenamiento en k subconjuntos y entrenándolos.

Tabla 3. Ejemplo de los conjuntos de entrenamiento usando bagging.

Ejemplo de un solo clasificador sobre un conjunto imaginario	
	Conjunto de entrenamiento (original)
Conjunto de entrenamiento 1	{1, 2, 3, 4, 5, 6, 7, 8}

Ejemplo de bagging sobre el mismo conjunto	
	Conjunto de entrenamiento (redistribuido)
Conjunto de entrenamiento 1	{2, 7, 8, 3, 7, 6, 3, 1}
Conjunto de entrenamiento 2	{7, 8, 5, 6, 4, 2, 7, 1}
Conjunto de entrenamiento 3	{3, 6, 2, 7, 5, 6, 2, 2}
Conjunto de entrenamiento 4	{4, 5, 1, 4, 6, 4, 3, 8}

3.4.3. Métodos de combinación

Existen 2 métodos principales para combinar las salidas de los clasificadores “base” (con los que está compuesto el conjunto): métodos ponderados y métodos por meta-aprendizaje. Los métodos ponderados son útiles si los clasificadores “base” desarrollan la misma tarea y tienen un desempeño comparable, por otro lado, los métodos por meta-aprendizaje van más orientados a casos en los que ciertos clasificadores constantemente clasifican correctamente, o constantemente clasifican erróneamente, ciertas instancias.

En los métodos ponderados, la clasificación de un clasificador tiene una fuerza proporcional a su peso asignado. Estos pesos pueden ser fijos o determinados dinámicamente para la instancia específica a ser clasificada. Dentro de los métodos ponderados existen varias formas de establecer los pesos; el método básico de los

conjuntos es la votación por mayoría, en la que la clasificación se lleva de acuerdo a la clase que obtiene el mayor número de votos, matemáticamente puede ser escrito como

$$clase(x) = \arg \max_{c_i \in \text{dom}(y)} \left(\sum_k g(y_k(x), c_i) \right) \quad (24)$$

donde $y_k(x)$ es la clasificación del k -ésimo clasificador y $g(y, c)$ es definido como

$$g(y, c) = \begin{cases} 1, & y = c \\ 0, & y \neq c \end{cases} \quad (25)$$

Para conocer más acerca de otros métodos ponderados o los métodos por meta-aprendizaje, consultar Rokach (2010).

3.4.4. Generación de diversidad

La diversidad es crucial para obtener conjuntos precisos (Tumer & Ghosh, 1996; Krogh & Vedelsby, 1995; Kuncheva, 2005; Kuncheva & Whitaker, 2003). De acuerdo a Hu (2001), el uso de diversos clasificadores da como resultado clasificaciones no correlacionables, lo cual mejora la precisión de la clasificación. Algunos de los métodos empleados para este propósito son:

- Manipular los conjuntos de entrenamiento, cada clasificador es entrenado con una variación diferente o subconjunto del original (bagging y boosting pertenecen a esta categoría).
- Manipular el inductor, existen varios métodos dentro de esta categoría, básicamente se busca manipular la búsqueda dentro del espacio de hipótesis; por ejemplo la forma más simple de esta categoría es iniciando la búsqueda en el espacio de hipótesis desde diferentes puntos de inicio, en redes neuronales se podría asignar diferentes pesos iniciales (hoy en día se considera que no es suficiente para conseguir una buena diversidad [Brown et al., 2005]).
- Cambiar la representación de los targets, en esta categoría, en lugar de inducir un único clasificador complicado, varios clasificadores con diferentes y usualmente representaciones más simples son inducidos.
- Particionar el espacio de búsqueda, la idea es que cada miembro del conjunto explora una parte diferente del espacio de búsqueda. Cada subespacio es considerado independientemente y el modelo "total" es una unión de modelos más simples.
- Hibridación, usando varios inductores base o estrategias.

3.5.Conclusiones

En este capítulo se presentaron algunas técnicas de extracción de características usadas en voz, las cuales se piensa que tienen un buen desempeño con los animales marinos, debido a la semejanza de algunos sonidos producidos. Se eligieron redes neuronales como clasificadores base, debido a los buenos resultados que ha presentado en este campo de investigación y en muchos otros.

Se toma en cuenta el uso de metodologías de conjuntos, las cuales trabajan bajo la idea de crear varios clasificadores base que sean sencillos, en lugar de crear uno solo que intente resolver un gran problema, al final cada clasificador base se tomará en cuenta para que vote por una o más clases.

4. Análisis de las señales acústicas subacuáticas objeto de estudio

Las señales acústicas subacuáticas cubren un amplio rango de sonidos, por lo que primero el trabajo se había acotado a sonidos que pudieran generarse en los mares del país. Sin embargo, tan solo México posee más de 11,500 km de líneas de costa (en las que existen aproximadamente 1,500,000 ha de lagunas costeras y esteros, 5,083 km² de islas marinas y su zona económica exclusiva ocupa 3,149,920 km² [UNAM/Instituto de Ciencias del Mar y Limnología, 2015]); por lo cual se redujo la zona a el Golfo de México (ver Fig. 4-1).



Fig. 4-1. Mapa del Golfo de México.

4.1. Selección de las bases de datos

Considerando la zona de interés, se realizó una búsqueda de bases de datos de los sonidos más comunes que se generan en esta zona. Encontrando 4 grupos: mamíferos marinos, peces, sonidos naturales y antropogénicos. A continuación se muestra a detalle el resultado de las búsquedas por grupo.

4.1.1. Mamíferos marinos

Los mamíferos marinos son un grupo variado de aproximadamente 130 especies de mamíferos que se han adaptado a la vida en el mar o dependen de él para su alimentación. En este grupo se incluyen los cetáceos (ballenas, delfines y marsopas), los sirenios (manatíes y dugongos), los pinnípedos (focas verdaderas, otarios y morsas) y algunas nutrias (nutria marina y gato de mar). El oso polar, aunque no es acuático, también se suele agrupar en esta categoría, debido a que vive en los hielos marinos

Análisis de las señales acústicas subacuáticas objeto de estudio

durante todo o la mayor parte del año y a su alto grado de adaptación a la vida en el mar (Stirling, 1988).

Se sabe que la zona del Golfo de México es hogar de 29 especies conocidas (ver Anexo A), incluyendo el manatí del oeste de India y 28 especies de ballenas barbadas, dentadas y delfines (Greenpeace, 2015).

Sin embargo, después de verificar las 29 especies en diversas bases de datos (DOSITS, 2015; Museum für naturkunde berlin, 2015; NPS, 2015; The Cornell Lab of Ornithology/ML, 2015; WHOAS, 2015; WHOI, 2015; entre otras), en la Tabla 4 se muestran los mejores registros encontrados.

Tabla 4. Listado de registros usados de mamíferos marinos.

Id	Base de Datos	Registros encontrados
2	Macaulay Library	48
7	Macaulay Library	380
12	Macaulay Library	139
14	Macaulay Library	38
18	Macaulay Library	39
21	Macaulay Library	87
24	Macaulay Library	99
29	Macaulay Library	11

Las razones por las que las otras especies no se incluyeron, es debido a que: no existían registros en las bases de datos, eran pocos registros, la SNR era muy baja, interferencia, etc. Aunque existen más bases de datos, Macaulay Library tiene un amplio repertorio y en sus registros se suele detallar información importante: como filtros aplicados, detalles de algunos sonidos que se escuchan, etc.

4.1.2. Peces

El término “pez” no se encuentra en ninguna clasificación taxonómica, pero este grupo de organismos incluye a todos los vertebrados acuáticos con respiración branquial, línea lateral, escamas y opérculos que habitan los cuerpos de agua del mundo.

A nivel mundial, los peces son el grupo de vertebrados más abundante. Sin embargo, constituyen el grupo más amenazado, los peces ocupan todos los hábitats acuáticos posibles, pueden vivir en el Tíbet a elevaciones de más de 5,200 m sobre el nivel del mar, donde algunos peces de la familia Balitoridae son dominantes y exitosos; incluso en nuestro país se pueden ver nadar los mexcalpiques (*Girardinichthys multiradiatus*) a

Análisis de las señales acústicas subacuáticas objeto de estudio

2,790 m sobre el nivel del mar en las Lagunas de Zempoala. De la misma forma se pueden encontrar peces en las fosas oceánicas hasta 7,000 m de profundidad.

México es reconocido como un país mega-diverso donde los peces constituyen el grupo de vertebrados con mayor riqueza de especies, ya que en la actualidad se conocen un total de 2,763 especies en el país, y éstas representan aproximadamente el 9.8% del total de los peces conocidos en el mundo. El panorama actual de la ictiofauna mexicana, que incluye 505 peces dulceacuícolas, 2,224 marinos y 563 estuarinos y vicarios; caracterizados como especies de distribución de las regiones neártica, neotropical o de transición en la parte continental y por 7 provincias marinas.

Al igual que con los mamíferos marinos, se consideraron especies que habitan en la zona del Golfo de México; considerando aquellos que están dentro de la Carta Nacional Pesquera (INAPESCA, 2015) (ver Tabla 5).

Tabla 5. Ficha en SAGARPA de peces del Golfo de México.

Ficha en SAGARPA
Bagre marino
Huachinango y pargo
Jurel y Cojinuda
Lisa y Lebrancha
Mero, negrilla y abadejo
Raya
Robalo y Chucumite
Sábalo
Sardina
Sierra y peto
Trucha de mar

Tabla 6. Listado de registros usados de peces.

Id	Base de Datos	Registros encontrados
79	Macaulay Library	22
80	Macaulay Library	88
83	Macaulay Library	30

Los 11 grupos mencionados en la carta, implican 94 especies de peces (ver Anexo B). De los cuales solo los incluidos en la Tabla 6 fueron los mejores registros encontrados; los motivos por los cuales el resto de las especies no fueron tomadas en cuenta fueron los mismos que con los mamíferos marinos.

Análisis de las señales acústicas subacuáticas objeto de estudio

4.1.3. Antropogénicos

Hay muchas fuentes de sonidos antropogénicos (generados por humanos) tanto en los océanos, lagos, etc. En bases de datos se encontraron registros de motores (DOSITS, NPS), refiriéndose a motores de embarcaciones; sonidos de construcción (DOSITS), generados por construcciones cerca de la costa (hincado de pilotes, etc.); y otros (DOSITS, WHOI), haciendo referencia a turbinas, sonares, etc.

Sin embargo, tras analizar los registros de los mamíferos marinos, se encontraron una gran cantidad de sonidos antropogénicos (debidos principalmente a motores), por lo que se consideraron estos registros de motores, para mantener registros con características similares (condiciones de grabación, SNR, etc.)

4.1.4. Naturales

De los naturales se encontraron 2 fuentes principales: debido a lluvia (DOSITS, NPS, WHOI) y a olas (DOSITS, NPS, WHOI). Pero al igual que con los sonidos antropogénicos, los registros de mamíferos marinos contenían este tipo de sonidos, por lo que se tomaron en cuenta los que se llegaron a registrar.

4.2. Clases a reconocer

En la Fig. 4-2 se muestran las clases a reconocer, en ella se agruparon los animales marinos por familias, ya que los sonidos dentro de una familia llegan a tener características similares (ver las siguientes secciones). Al final el modelo será capaz de reconocer entre 14 clases: 7 cetáceos (5 odontocetos y 2 misticetos), 1 sirenio, 3 peces, sonidos naturales, barcos y sonidos desconocidos (los cuales fueron registrados en las grabaciones pero no se determinó exactamente el origen de dichos sonidos).

4.3. Emisiones de mamíferos marinos

Los animales marinos utilizan sonido para percibir su entorno, comunicarse, buscar comida y protegerse bajo el agua. Generan sonidos para atraer parejas, defender sus territorios y coordinar actividades en grupo. Los mamíferos marinos usan el sonido para mantener el contacto entre la madre y las crías, para la reproducción y mostrar agresividad (Scowcroft et al., 2012).

Este trabajo se limitará a los sonidos que son producidos internamente por los animales y no se considerarán aquellos producidos por un animal golpeando un objeto o la superficie del agua con alguna parte de su cuerpo. A continuación se detallan los diferentes tipos de sonidos producidos por odontocetos y misticetos, los cuales son 2 subórdenes de mamíferos marinos.

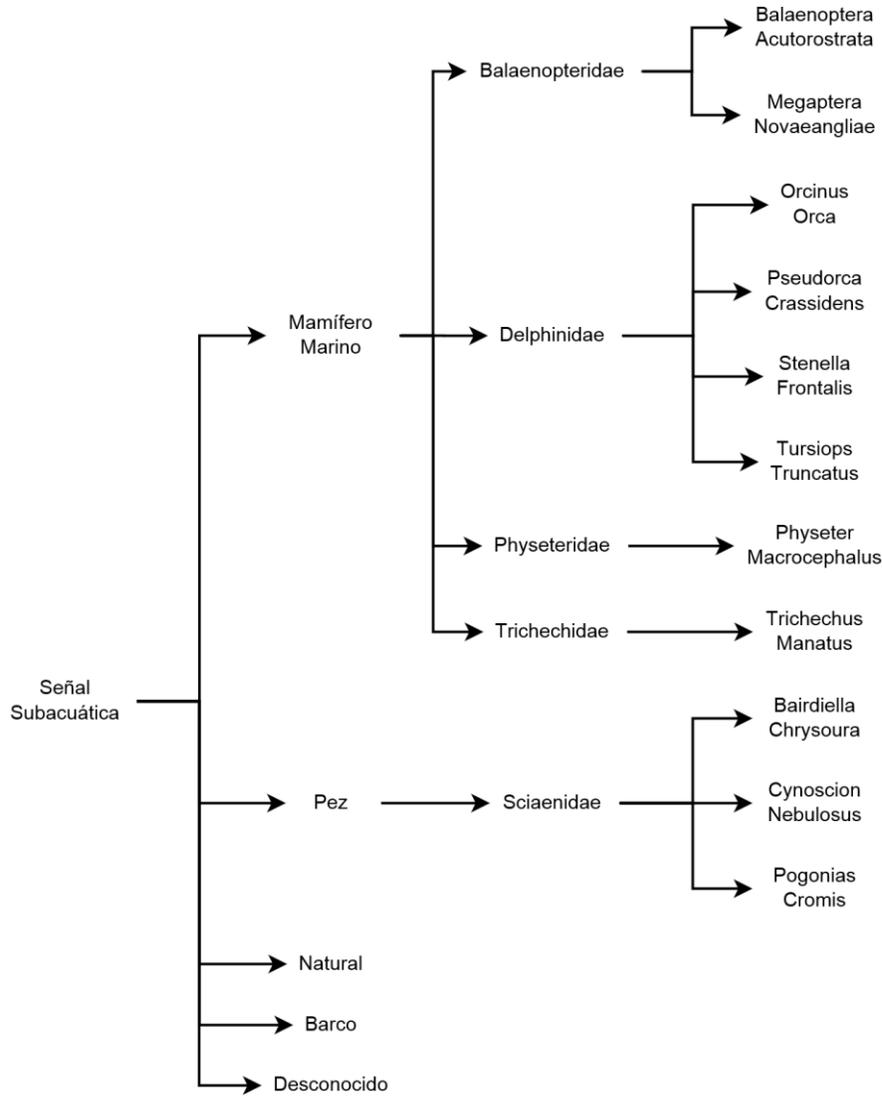


Fig. 4-2. Clases a reconocer.

4.3.1. Odontocetos

Los sonidos producidos por los odontocetos (ballenas dentadas y delfines) pueden ser clasificados en dos amplias categorías: tonos continuos de frecuencia variante (conocidos como silbidos) y clics de banda ancha (Evans, 1967), incluyendo ráfagas de impulsos. Los silbidos y las ráfagas de impulsos pueden ser categorizados como sonidos sociales. Los silbidos parecen ser usados para comunicaciones intraespecíficas (entre individuos de una sola especie), (Herman & Tavolga, 1980). La mayoría de la investigación sobre sonidos sociales se ha llevado a cabo con silbidos, existen pocos trabajos con las ráfagas de impulsos; tal vez la principal razón sea que la frecuencia fundamental de los

Análisis de las señales acústicas subacuáticas objeto de estudio

silbidos se encuentra principalmente en el rango sónico o audible (debajo de los 20 kHz) y por lo tanto son más fáciles de grabar por las grabadoras convencionales.

Los sonidos sociales son aquellos que son usados por los odontocetos en un contexto social y podrían ser, pero no necesariamente, usados para comunicarse. En la práctica general se refieren a estos sonidos como vocalizaciones; sin embargo, este término significa dar voz, y se refiere al hecho de que los sonidos son producidos por las cuerdas vocales en la laringe de los humanos y de otros animales, y las investigaciones en odontocetos apuntan al sistema nasal como el lugar donde se producen los sonidos, por lo que no es adecuado referirse a estas emisiones como vocalizaciones.

Silbidos

Los silbidos son tonos modulados en frecuencia y en amplitud, que típicamente duran entre 50 ms y 3 s o más. Aunque muchos odontocetos producen silbidos, hay pocos que no lo hacen. Algunas de estas especies que nunca o rara vez silban son:

- *Phocoena phocoena* (Marsopa común)
- *Phocoenoides dalli* (Marsopa de Dall)
- *Cephalorhynchus commersonii* (Delfín pío)
- *Cephalorhynchus hectori* (Delfín de Héctor)
- *Neophocaena phocaenoides* (Marsopa sin aleta)
- *Kogia breviceps* (Cachalote pigmeo)
- *Physeter catodon* (Cachalote)
- Otras marsopas en la familia Phocoenidae, delfines en el género *Cephalorhynchus* y otros pocos delfines.

Sin embargo, todos los odontocetos estudiados hasta la fecha, emiten clics o ráfagas de impulsos.

Los silbidos han sido grabados tradicionalmente con grabadoras comerciales de cinta, grabadoras de casete, grabadoras DAT, y en la pista de audio de grabadoras de video. Todos estos instrumentos tienden a tener anchos de banda limitados, con una frecuencia máxima de entre 15 y 22 kHz. Aunque la componente fundamental de los silbidos a menudo cae entre 5 y 15 kHz (Herman & Tavolga, 1980), algunas especies como el *Lagenorhynchus albirostris* producen silbidos que se pueden extender más allá de los 20 kHz.

Aun cuando la componente fundamental de los silbidos se encuentre dentro del rango sónico, estos pueden tener armónicos de mayor orden, que se pueden extender considerablemente hacia el rango ultrasónico (Lammers & Au, 1996). Estos armónicos son una parte integral de un silbido, y su presencia indica que el silbido se desvía de una onda

Análisis de las señales acústicas subacuáticas objeto de estudio

senoidal pura. Los ingenieros eléctricos tradicionalmente usan las amplitudes de los armónicos de más alto orden para medir la cantidad de distorsión en una señal cuando se es comparada con una onda senoidal pura. Mientras el ancho de banda sea más limitado, la apariencia de la onda será más sinusoidal, dicha diferencia sería perceptible por un odontoceto.

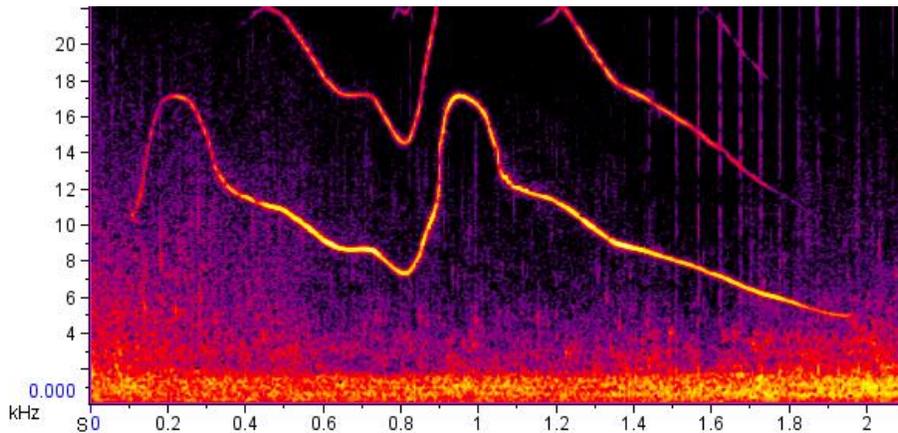


Fig. 4-3. Espectrograma del silbido de un T. Truncatus en el que su armónico rebasa la frecuencia máxima (1/2 de la frecuencia de muestreo).

Hay una gran variedad de silbidos, tanto de animales en cautiverio como salvajes. El método tradicional (que sigue siendo usado por muchos para establecer las diferentes categorías y determinar el repertorio de silbidos) consiste en inspección visual de los espectrogramas con un énfasis particular en la forma del contorno del silbido. Esto hace que la técnica sea muy subjetiva y ha producido resultados muy diferentes entre investigadores. Por lo tanto, la investigación en clasificación de “llamados” está plagada con dificultades en la nomenclatura y falta de alguna estandarización en las técnicas y clasificación de categorías. Algunas de las llamadas pueden ser muy complejas y puede incluir combinaciones de diferentes estructuras individuales.

En los primeros estudios se le daba mucho énfasis a las posibilidades lingüísticas de estos sonidos, por lo que varios investigadores hicieron distinciones muy finas entre contornos similares. Caldwell & Caldwell (1965) observaron 5 diferentes contornos de más de 1,400 silbidos de 4 delfines comunes. Mientras que Dreher & Evans (1964) identificaron 17 diferentes contornos de 9 *Tursiops Truncatus* que se encontraban en cautiverio; 16 en un grupo de 5 *Tursiops gilli*, 19 en un grupo de delfines comunes salvajes, y 7 en una manada de unas 12 ballenas piloto de aleta corta. Dreher & Evans (1964) también reportaron 23 contornos diferentes para 3 jóvenes *Tursiops*, aunque hacían mención que algunas clasificaciones únicas tenían variaciones mínimas de alguna otra clasificación. Uno puede ser muy estricto en aceptar contornos dentro de una categoría

Análisis de las señales acústicas subacuáticas objeto de estudio

establecida y por lo tanto tener más tipos de contornos que si uno hubiera sido laxo. No obstante, si se toman categorías más generales, los silbidos pueden agruparse en 6-7 diferentes contornos (Bazúa-Durán, 1997), el resto de las categorías suponen variaciones mínimas dentro de las 6 principales:

1. Frecuencia constante: señales donde la frecuencia cambió un 25% o menos sobre la duración total de la señal. En esta categoría, la frecuencia del silbido no permanece realmente constante a lo largo del tiempo, pero el cambio de frecuencia es mínimo. Taruski (1979) fue el primero en usar esta categoría para describir los silbidos de ballenas piloto.

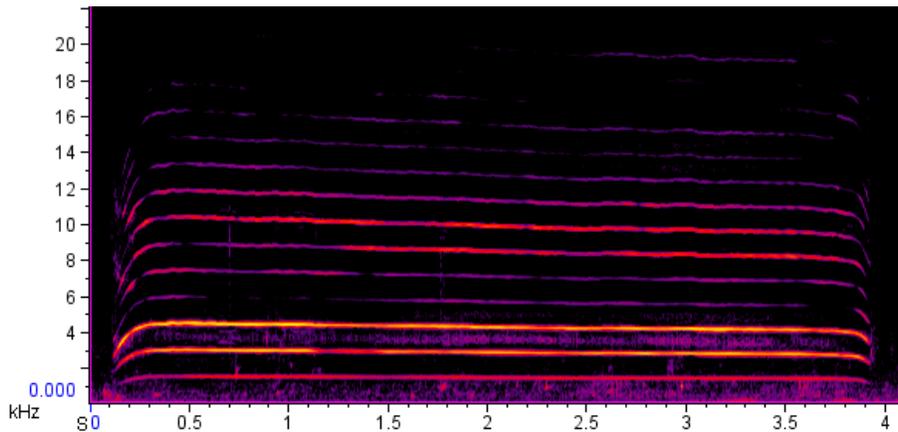


Fig. 4-4. Espectrograma del silbido tipo constante (con armónicos) de un T. Truncatus.

2. Hacia arriba (“upsweep”): silbidos modulados en frecuencia, con una frecuencia que va aumentando conforme pasa el tiempo y no tienen importantes puntos de inflexión.

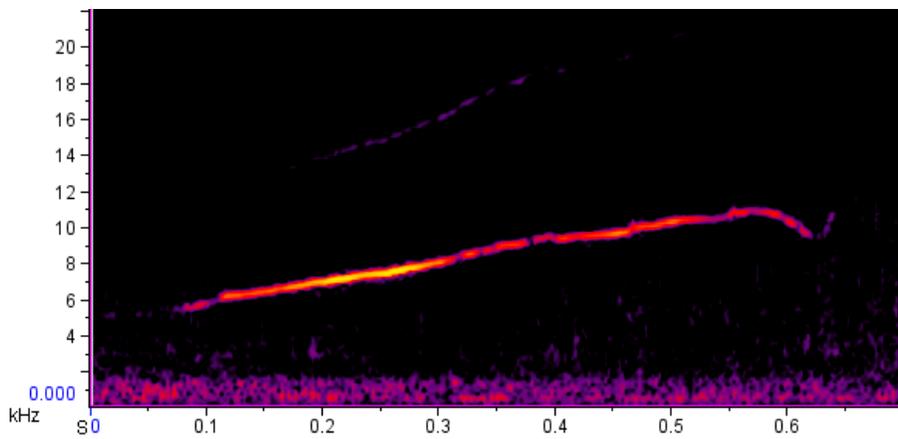


Fig. 4-5. Espectrograma del silbido tipo hacia arriba de un T. Truncatus.

Análisis de las señales acústicas subacuáticas objeto de estudio

3. Hacia abajo (“downsweep”): silbidos modulados en frecuencia, con una frecuencia que va disminuyendo conforme pasa el tiempo y no tienen importantes puntos de inflexión.

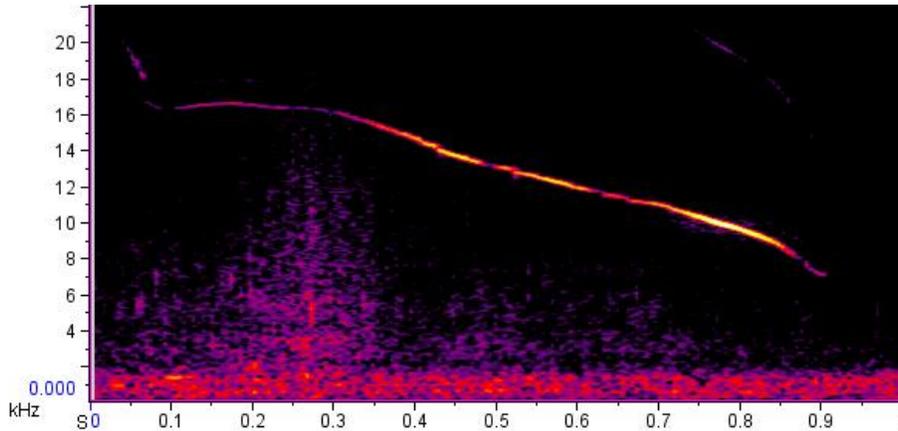


Fig. 4-6. Espectrograma del silbido tipo hacia abajo de un T. Truncatus.

4. Cóncavo o cresta: silbidos modulados en frecuencia, con una frecuencia que inicialmente iba aumentando con el tiempo, seguido de una porción final en la que la frecuencia disminuye con el tiempo.

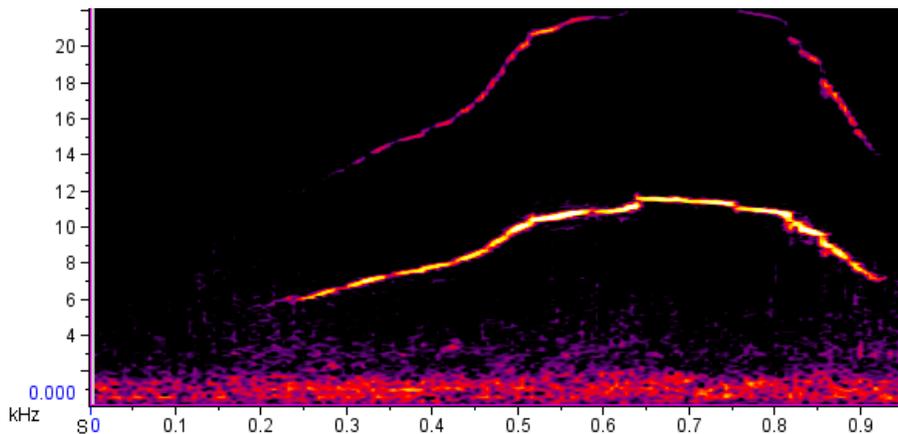


Fig. 4-7. Espectrograma del silbido tipo cóncavo de un T. Truncatus.

5. Convexo o valle: silbidos modulados en frecuencia, con una frecuencia que inicialmente iba disminuyendo con el tiempo, seguido de una porción final en la que la frecuencia aumenta con el tiempo.

Análisis de las señales acústicas subacuáticas objeto de estudio

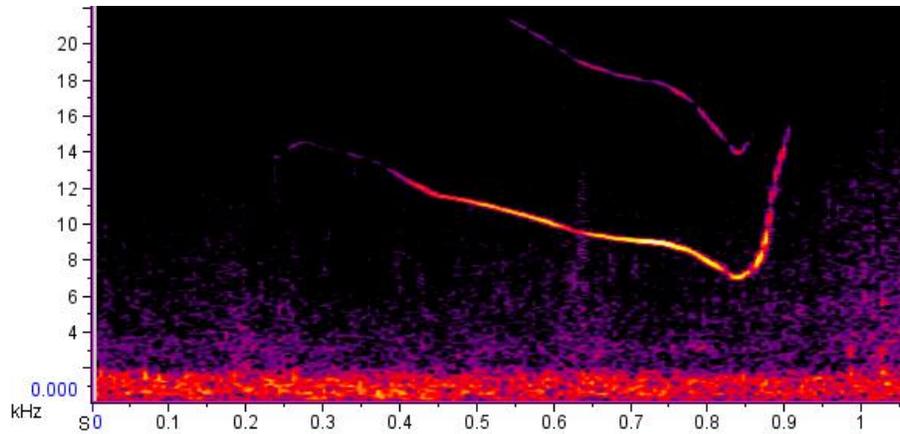


Fig. 4-8. Espectrograma del silbido tipo convexo de un T. Truncatus.

6. Sinusoidal o múltiple: silbidos modulados en frecuencia con más de una repetición de una cresta o valle, y el contorno pareciera ser una señal sinusoidal con al menos dos puntos de inflexión. Las categorías 2-6 están basadas en la pendiente del silbido y el número de puntos de inflexión, donde un punto de inflexión es definido como un punto en el que la pendiente del contorno cambia de dirección.

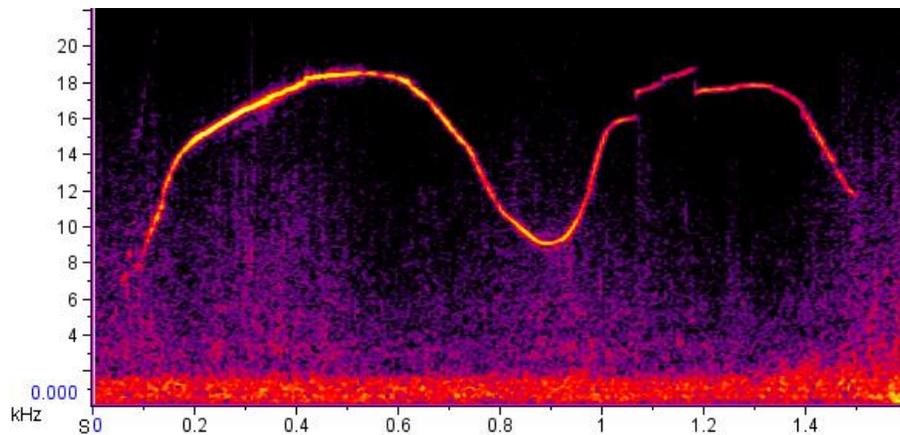


Fig. 4-9. Espectrograma del silbido tipo sinusoidal de un T. Truncatus.

Driscoll (1995) agregó otra categoría basada en el tiempo y la presencia de modulación de frecuencia. Ella se refiere a esta categoría como chirrido (“chirp”), las señales en esta categoría tienen una corta duración (menor o igual a 300 ms) y están moduladas en frecuencia. La forma del contorno puede parecerse a contornos de otras categorías, pero la principal diferencia es que las señales son relativamente cortas. Caldwell & Caldwell (1970) fueron los primeros en describir chirridos en silbidos producidos por *Tursiops Truncatus*.

Análisis de las señales acústicas subacuáticas objeto de estudio

Un método más cuantitativo de categorizar los silbidos, involucra la determinación de varios parámetros que pueden caracterizarlos (ver Fig. 4-10).

Algunos de los diferentes parámetros asociados incluyen:

1. Frecuencia inicial
2. Frecuencia final
3. Frecuencia mínima
4. Frecuencia máxima
5. Número de puntos de inflexión
6. Frecuencia de los puntos de inflexión
7. Duración
8. Interrupción en el contorno o número de pasos
9. Presencia de armónicos

Estos parámetros han sido usados como entrada para diferentes programas estadísticos para su análisis. Los resultados de este método generalmente clasifican los silbidos dentro de las 6 categorías previamente mencionadas. Los contornos pueden ser usados directamente en clasificar los silbidos, sin embargo, se debe de tener cuidado por variaciones en la duración del sonido emitido. Los silbidos pueden tener las mismas frecuencias inicial, final, mínima y máxima; pero una ligera variación en la duración puede dar como resultado una clasificación incorrecta.

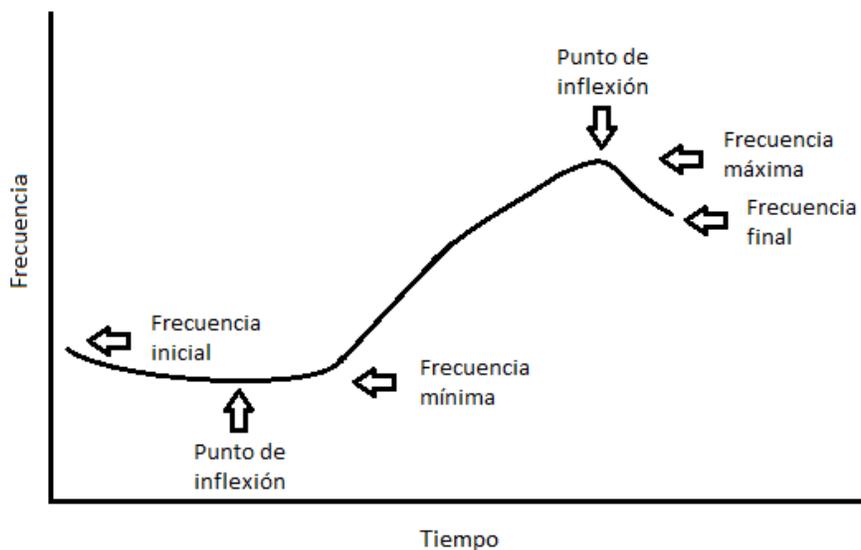


Fig. 4-10. Espectrograma genérico con algunos de los diferentes parámetros de un contorno.

Análisis de las señales acústicas subacuáticas objeto de estudio

Ráfagas de pulsos

A diferencia de los silbidos, todos los odontocetos producen este tipo de sonidos. En los delfines y ballenas pequeñas, se caracterizan por una tasa alta de repetición (mayor a 300 pulsos por segundo) o pequeños intervalos entre pulsos (menor a 3 ms). Para el oído humano, este tipo de sonidos asemejan diferentes tipos de sonidos debido al intervalo entre pulsos o a la tasa de repetición. Este tipo de sonidos no han sido estudiados tan a fondo como los silbidos; debido principalmente al ancho de banda que comprenden las componentes fundamentales de un silbido (las cuales se encuentran dentro del rango auditivo humano), en comparación a las de un pulso (las cuales se pueden extender más allá de 100 kHz [Lammers & Au, 1996], inclusive teniendo solo componentes de frecuencia en el rango ultrasónico).

La diferencia entre los clics de ecolocalización y las ráfagas de pulsos radica en los intervalos entre clics y la intensidad. Sin embargo, determinar la intensidad de la fuente o la presión de un sonido emitido requiere un previo conocimiento de la distancia entre el hidrófono y el emisor, lo cual no es tan simple en la naturaleza, o en un tanque en cautiverio donde el animal puede estar nadando rápido mientras emite los pulsos. Por lo tanto, el intervalo entre clics es probablemente el mejor método para determinar si un tren de pulsos es un clic de ecolocalización o una ráfaga de pulsos.

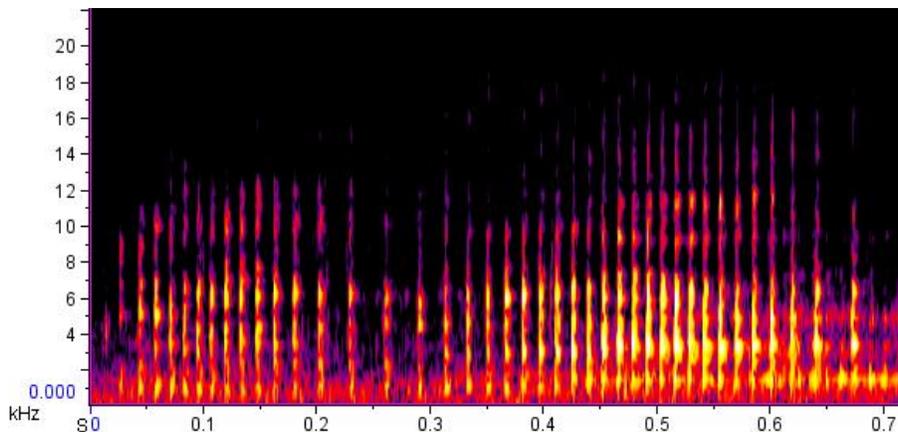


Fig. 4-11. Espectrograma de una ráfaga de pulsos de un T. Truncatus.

4.3.2. Mistictetos

Hay 11 especies de mistictetos (cetáceos barbados), y a excepción de la ballena franca pigmea, todos han llegado a ser grabados (Clark, 1990). En el caso de los mistictetos, es apropiado y conveniente referirse a los sonidos emitidos como vocalizaciones, ya que no hay evidencia de cómo son producidos y bien pueden ser producidos dentro de la laringe. Las vocalizaciones de los cetáceos pueden ser divididas en 2 categorías generales: “llamados” y “cantos” (Clark, 1990).

Análisis de las señales acústicas subacuáticas objeto de estudio

Llamados

Los llamados a su vez pueden ser subdivididos en: 1) llamados simples, 2) llamados complejos y 3) clics, pulsos, golpes y gruñidos (Clark, 1990).

Los llamados simples a menudo son de baja frecuencia y son modulados tanto en frecuencia como en amplitud, asimismo suelen contener armónicos con la mayoría de la energía debajo de 1 kHz. Debido a su baja frecuencia necesitan ser editados para poder ser escuchados por una persona.

Los llamados complejos son pulsos de banda ancha con una mezcla de ruido de amplitud modulada y/o una frecuencia fundamental modulada. Este tipo de llamados genera una amplia variedad de sonidos, que pueden ser escuchados como gritos, gruñidos o rugidos. La mayoría de la energía está comprendida entre los 500 y 5,000 Hz.

Los clics, pulsos, golpes y gruñidos son de señales de muy corta duración (menores a 0.1 s), con poca o nula modulación de frecuencia. Clark (1990) reportó que los clics y pulsos son las señales más cortas (menores a 2 ms) con frecuencias entre 3 y 31 kHz, mientras que los gruñidos y golpes duran más (entre 50 y 100 ms) y se encuentran en el rango de 100 a 1,000 Hz.

Los llamados tienen el mismo problema que los silbidos de los odontocetos, hace falta una nomenclatura estandarizada. Varios llamados similares a menudo son catalogados de manera diferente por los investigadores.

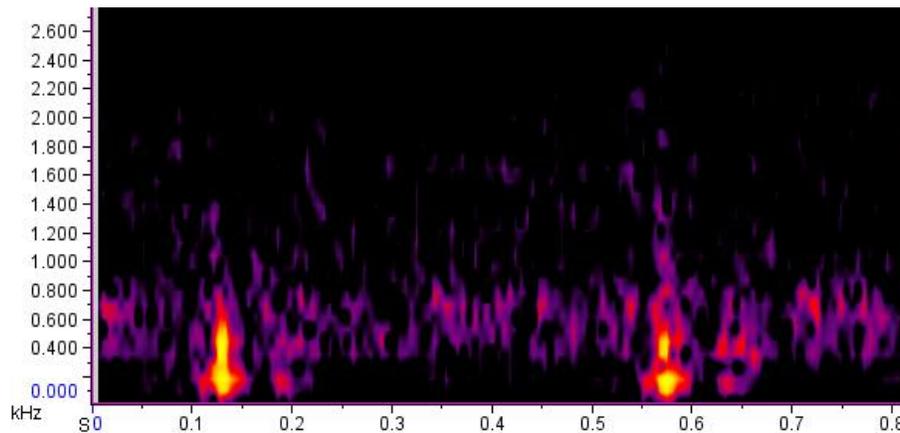


Fig. 4-12. Espectrograma de dos llamados de un B. Acutorostrata.

Cantos

Los cantos son definidos como secuencias de notas que ocurren en una secuencia regular y en un patrón de tiempo (Clark, 1990). Solo se conocen 4 especies de mysticetos

Análisis de las señales acústicas subacuáticas objeto de estudio

que producen cantos: *Balaenoptera Musculus* (Ballena azul), *Balaenoptera Physalus* (Rorcual común), *Balaena Mysticetus* (Ballena de Groenlandia) y la *Megaptera Novaeangliae* (Ballena jorobada). Siendo esta última la que ha recibido más atención por parte de los investigadores, debido en parte a la facilidad de encontrar a esta ballena en el verano, especialmente en Hawai (Océano Pacífico) y en Puerto Rico (Océano Atlántico). Cabe mencionar que los cantos de años consecutivos son muy similares, pero los cantos a través de años no consecutivos tendrán pocas similitudes. A continuación se explica cómo se forman los cantos:

Los sonidos conocidos como unidades, son los sonidos más cortos que parecen ser continuos para el oído humano, y son repetidos varias veces con una gran redundancia. Una frase consta de unidades repetidas o secuencias de unidades repetidas, y a menudo están subdivididas en subfrases. Las frases pueden ser repetidas de 20 a 30 veces por un solo animal. A menudo, una secuencia de frases repetidas es seguida de otra secuencia de una frase diferente, a esto se le llama tema. Los cantos consisten de 5 a 7 temas producidos en un orden estereotipado. La duración de los cantos no es constante, pueden durar algunos minutos, varias horas y algunas han llegado a durar hasta 24 horas (Winn & Winn, 1978).

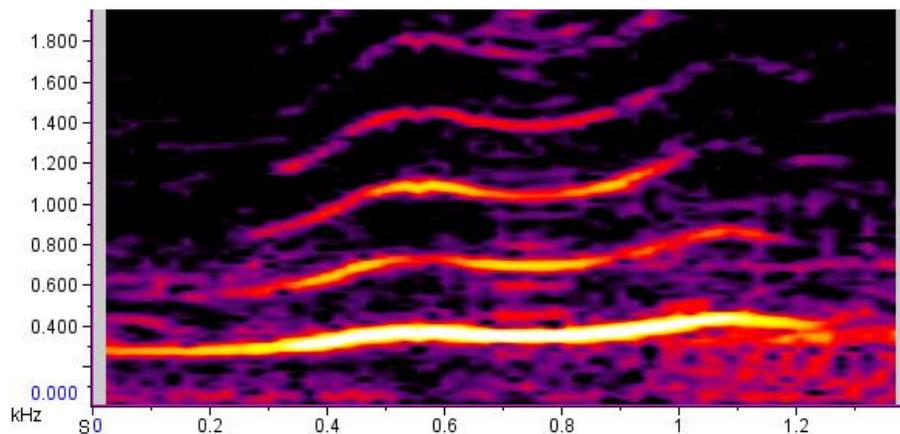


Fig. 4-13. Espectrograma de una unidad producida por una *M. Novaeangliae*.

4.3.3. Sirenios

Se ha hecho muy poca investigación sobre vocalizaciones de manatíes y dugongos, uno de los motivos es que algunas de estas especies suelen ser muy calladas y solo emiten sonidos en circunstancias de miedo, irritación, molestia o cuando los machos están excitados; aunque los manatíes pasan su vida entera en el agua, se sabe muy poco de sus vocalizaciones subacuáticas. Hartman (1979) reportó que el *Trichechus Manatus* (Manatí del Caribe) produce chillidos de tono alto, chirridos y gritos. Por su parte Schevill & Watkins (1965b) registraron llamados del *T. Manatus* con una frecuencia fundamental entre 2.5 y 5 kHz, aunque algunos llegaban a los 0.6 kHz. Los sonidos son relativamente

Análisis de las señales acústicas subacuáticas objeto de estudio

débiles, unos 10 a 12 dB a 3 o 4 m, por encima del ruido ambiental; y con duraciones relativamente cortas de 0.15 a 0.5 s.

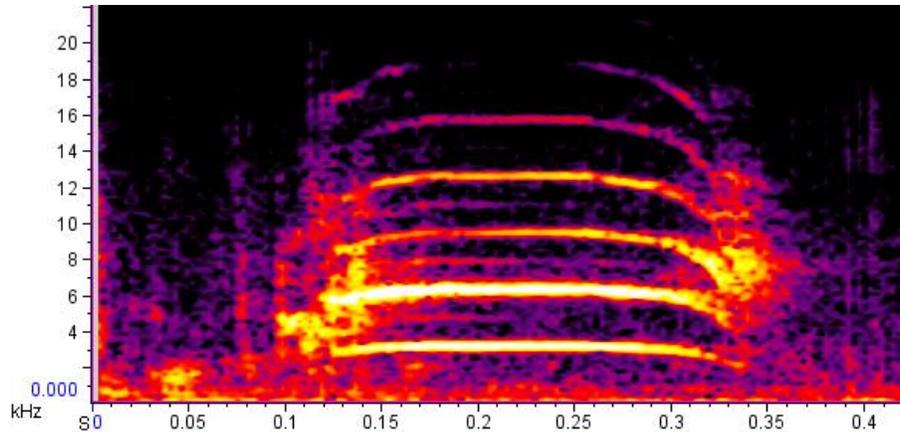


Fig. 4-14. Espectrograma de un chillido de un T. Manatus.

Los dugongos también producen vocalizaciones que han sido descritas como silbidos, chirridos y chirridos-chillidos (Nishiwaki & Marsh, 1985). Sus sonidos se encuentran en un rango de 1 a 8 kHz y parecen ser muy similares a los producidos por los manatíes.

4.4. Emisiones de peces

La complejidad del sonido producido por peces, radica en que producen sonido de muchas maneras, ninguna de ellas involucra una laringe o alguna estructura similar usada por los vertebrados terrestres. Existen 2 métodos en los que producen sonido: por estridulación y por manipulación de músculos sónicos que se encuentran cerca o sobre la vejiga natatoria (Zelick et al., 1999). Los sonidos generados por estridulación son típicamente del tipo pulso (ver Fig. 4-15), de banda ancha y de corta duración (10-50 ms). Los pulsos producidos se asemejan a los clics de ecolocalización producidos por los delfines, excepto que duran más y son de menor frecuencia.

Por otro lado, los sonidos producidos por los músculos sónicos, tienden a ser pulsos tonales. Estos sonidos duran más que los sonidos producidos por estridulación. Los sonidos de peces con músculos sónicos, nunca han sido estudiados a fondo.

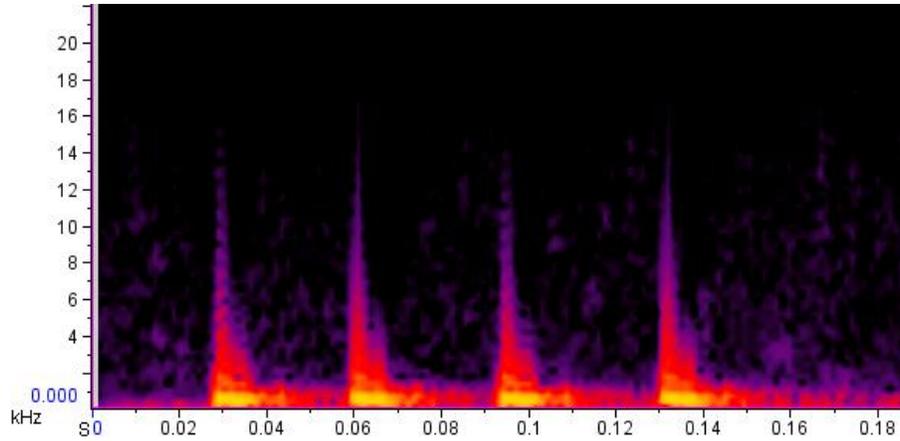


Fig. 4-15. Espectrograma de sonido hecho por un B. Chrysoira.

4.5. Emisiones de señales antropogénicas

En cuanto a los sonidos generados por humanos, existen una amplia variedad de sonidos, siendo de particular interés los generados por las embarcaciones (ver Fig. 4-16), debido principalmente por la cavitación generada por la propela. Es importante conocer el sonido producido por las embarcaciones, debido a que la gran mayoría de las grabaciones realizadas en campo para estudios de animales, contienen ruido generado por la embarcación desde donde se realizaba la grabación o de embarcaciones cercanas.

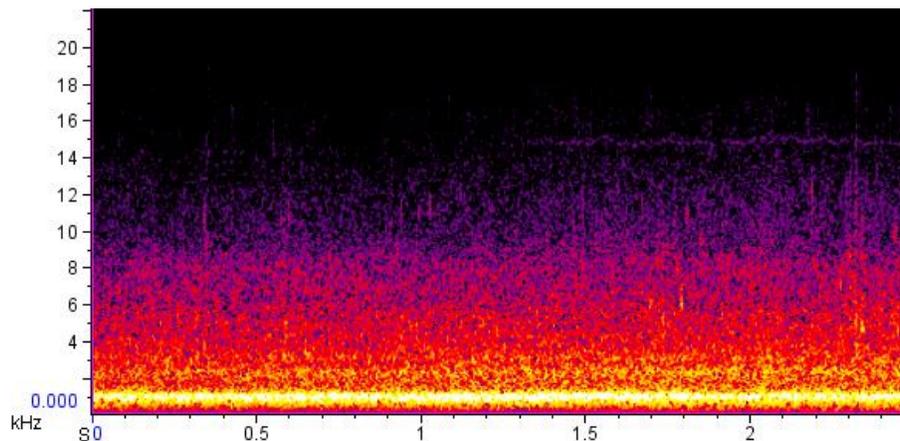


Fig. 4-16. Espectrograma del sonido de una embarcación pequeña.

4.6. Emisiones de señales naturales

Las principales fuentes de ruido ambiental pueden clasificarse por la frecuencia del sonido. En el rango de frecuencias de 20-500 Hz, el ruido ambiental en el hemisferio norte es principalmente debido al ruido generado por entregas distantes. En el rango de

Análisis de las señales acústicas subacuáticas objeto de estudio

frecuencias de 500-100,000 Hz, el ruido ambiental es principalmente debido a lluvia, así como burbujas y rocío asociado al rompimiento de las olas.

4.7. Etiquetado manual de señales

Las grabaciones obtenidas de la base de datos no son grabaciones puras de las fuentes de interés. Son grabaciones hechas en campo, las cuales tienen una estructura similar a la de la Fig. 4-17, donde al inicio de la grabación hay un bloque de voz (detallando el número de registro de la grabación, la fuente del sonido, fecha de grabación, etc.), después viene el bloque de grabación de la fuente de interés (en la que puede haber más bloques de voz, cortes en la señal, tonos de calibración, etc.). La duración de las grabaciones varía aproximadamente de 5 minutos a poco más de 1 hora. Otra limitante es que en muchas grabaciones existe la superposición de 2 o más señales, lo cual por ahora queda fuera de los alcances de este trabajo.

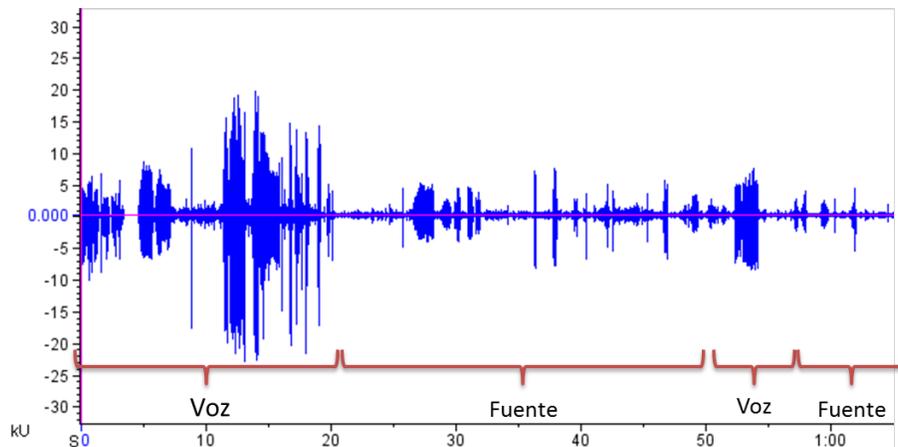


Fig. 4-17. Estructura común de un registro de la base de datos.

En voz para detectar dónde comienza y termina una señal de interés, se suele ir haciendo un ventaneo de la señal, de la cual se extrae la energía y la cantidad de cruces por cero; si dichos valores rebasan cierto umbral se puede considerar que hay un segmento de voz. Sin embargo, en este tipo de grabaciones es muy difícil hacer esto desde un inicio, porque se generan una gran cantidad de sonidos que en realidad serían falsos-positivos.

Por esta razón y con el fin de obtener los mejores resultados, se debe de hacer un recorte manual de las señales basado en el espectrograma y el sonido. Al final se tendrá una base de datos "limpia" de cada fuente, de la cual se podrán extraer las características de una forma más fácil y con mejores resultados. Para hacer el recortado, se creó un programa (ver Fig. 4-18) que ayuda a ir analizando la señal por segmentos de tiempo

Análisis de las señales acústicas subacuáticas objeto de estudio

establecidos por el usuario, permite el uso de filtros pasa altas, pasa bajas o pasa banda, selección de canal, diezmado, etc., para un mejor análisis.

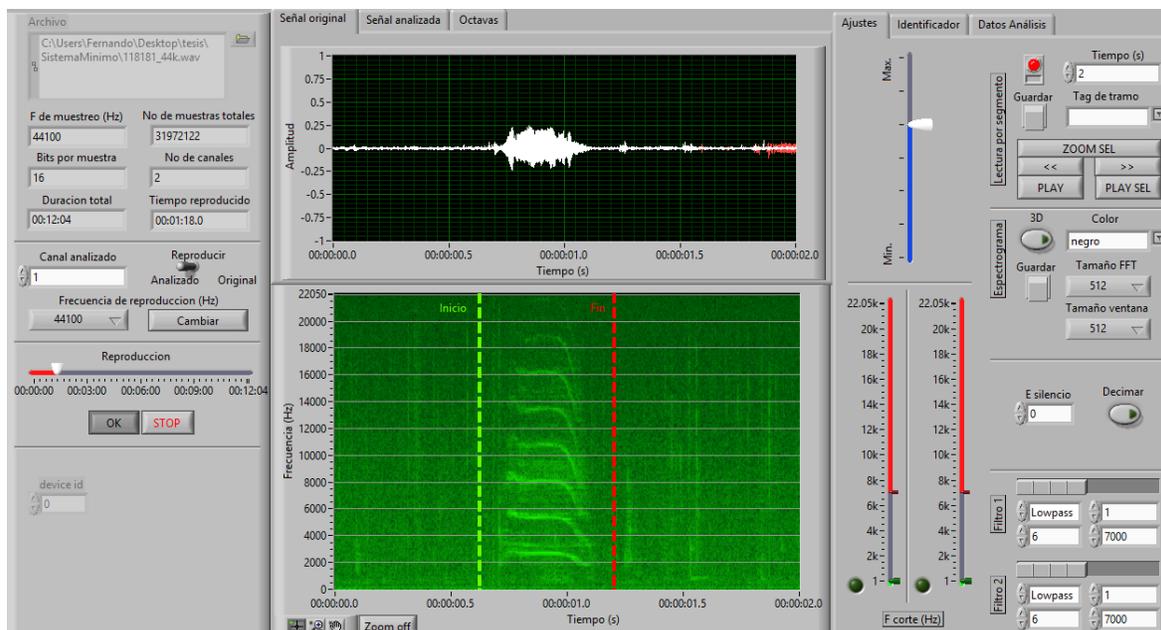


Fig. 4-18. Programa de etiquetado en funcionamiento.

Tabla 7. Fuentes y cantidad de registros.

Fuente	Número de registros
B. Acutorostrata	178
M. Novaeangliae	506
O. Orca	644
P. Crassidens	396
S. Frontalis	296
T. Truncatus	171
P. Macrocephalus	128
T. Manatus	378
B. Chryssoura	116
C. Nebulosus	332
P. Cromis	475
Natural	87
Barco	31
Desconocido	230
Total	3,968

Análisis de las señales acústicas subacuáticas objeto de estudio

4.8. Base de datos generada

Después de haber hecho el recortado manual de las grabaciones mostradas en el Anexo C, se generó una base de datos por fuente (ver Tabla 7).

Con esta base de datos es más fácil generar una segunda base, que contenga los rasgos característicos usando diferentes técnicas de extracción.

4.9. Análisis y extracción de características

4.9.1. Pre-procesamiento

En las secciones 4.3-4.6 se analizó el comportamiento de las señales a analizar; del análisis se puede ver que las señales utilizan todo el espectro (0 - 22.05 kHz), por lo cual es imposible aplicar a todas las señales un filtro pasa bajas, pasa altas o pasa banda, o técnicas para disminuir la frecuencia de muestreo, ya que esto eliminaría información importante. Por lo que solo se aplica un pre-énfasis al bloque de señal leída por el sistema, para aumentar los niveles de las altas frecuencias, que es donde se encuentran la mayoría de los sonidos emitidos por las clases.

4.9.2. Detección de señal

Si bien podría crearse un sistema que intentará clasificar todos los segmentos de la señal que fuera leyendo, no solo consumiría tiempo de procesamiento, sino energía y espacio de almacenamiento, los cuales son muy importantes de cuidar; considerando que este tipo de sistemas pueden terminar operando en sistemas autónomos con recursos limitados (batería y espacio de almacenamiento). Por lo que conviene crear módulos de detección para que todos los eventos que simplemente sean ruido ambiental sean descartados.

En el estado del arte se han usado una gran cantidad de técnicas para la detección de las señales de interés, incluso creando algoritmos de detección que consumen más tiempo que los algoritmos de clasificación. En este trabajo se implementó una técnica sencilla que se usa en procesamiento de voz: para detectar dónde se encuentran los segmentos de voz.

En este caso se hace un ventaneo de la señal leída (23 ms con un traslape del 50%) en tiempo real o diferido, de dicha ventana se extrae la energía (ver ecuación (26)). Si rebasa el umbral del 50% de la máxima energía, se considera que es el inicio de la señal; y ésta empieza a ser almacenada en un búfer, el proceso de almacenamiento en búfer termina cuando la energía de la ventana leída no rebasa el mismo umbral.

$$E = \sum_n |s(n)|^2 \quad (26)$$

4.9.3. Extracción de características

En un principio se consideraron los coeficientes LPC, para esto se realizaron varias pruebas de clasificación con redes MLP provisionales (hasta con 47 coeficientes), el error más grande se notaba en la clase Delphinidae. En la Fig. 4-19 se muestra la extracción de 30 coeficientes LPC de un silbido de una O. Orca, el silbido fue recortado de diferente forma (con solo unos milisegundos de diferencia), por eso se muestran coeficientes rojos y azules. Puede verse que los primeros coeficientes cambian considerablemente, no solo en magnitud sino también en dirección, esto es un problema debido al comportamiento en ocasiones errático de los sonidos producidos por los animales.

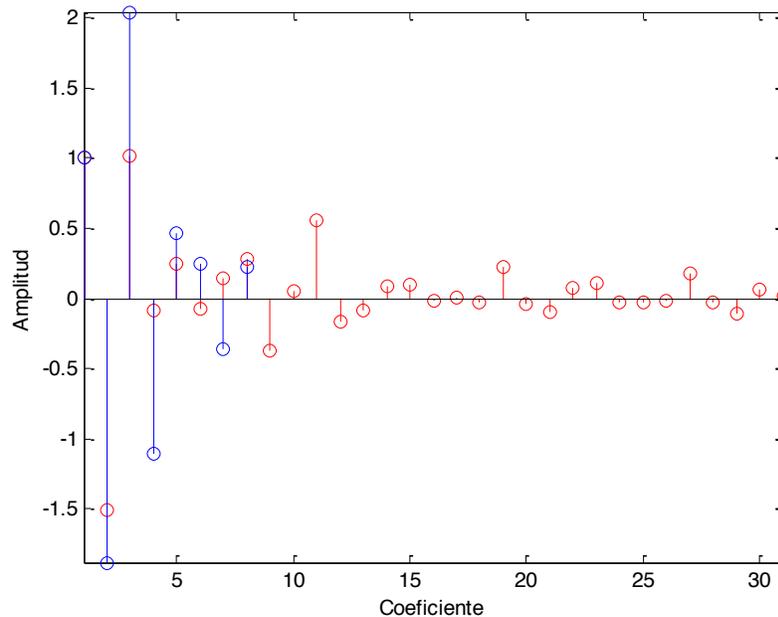


Fig. 4-19. Comparativa de LPCs.

Por otro lado, en la Fig. 4-20 se muestra la extracción del mismo silbido (recortado de forma distinta) pero con coeficientes MFCC, al ser niveles de energía por bandas de frecuencia puede verse que la similitud entre los rojos y azules mejora notablemente; también los porcentajes de clasificación, de las redes neuronales de prueba mejoraron, pero nunca superaron el 85% para el conjunto de prueba.

Por el contrario, con el análisis de octavas no solo se solucionaba el problema que se presentaba con los LPC (ver Fig. 4-21), desde las pruebas iniciales las octavas presentaron buenos porcentajes de clasificación (rara vez bajaban del 90%) y un rápido entrenamiento. Debido a los altos porcentajes de clasificación se decidió solo usar esta técnica de extracción y no mezclar técnicas, como se ha hecho en algunos trabajos de reconocimiento de aeronaves (Márquez-Molina et al., 2014; Sánchez et al., 2012).

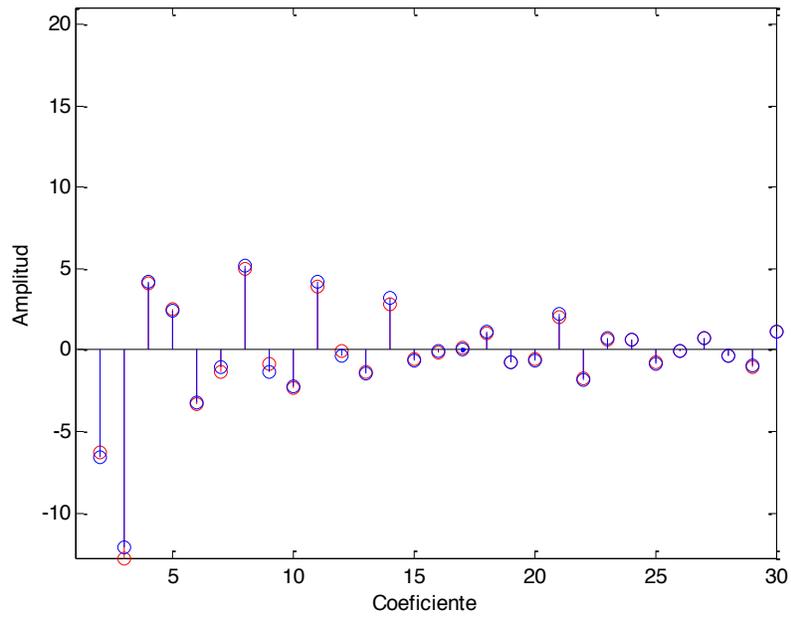


Fig. 4-20. Comparativa de MFCCs.

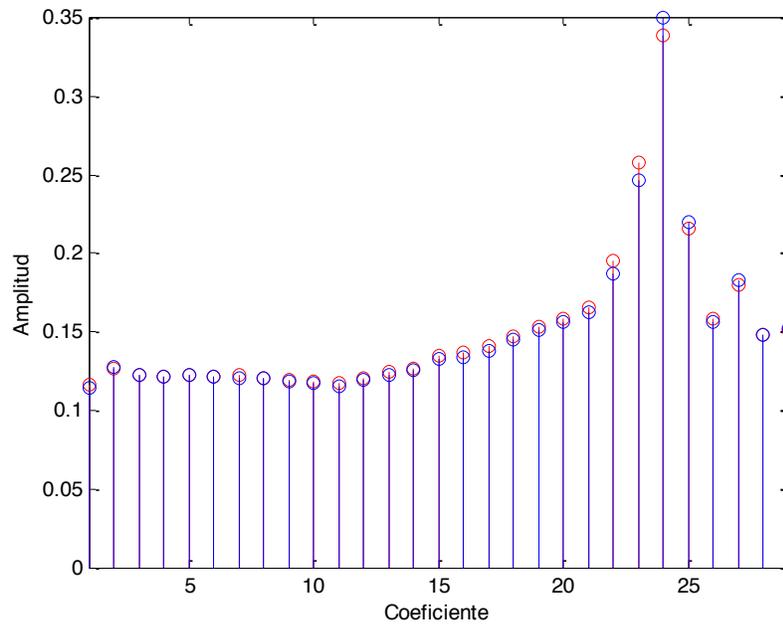


Fig. 4-21. Comparativa de Octavas.

Análisis de las señales acústicas subacuáticas objeto de estudio

Como se mencionó en la sección 3.2.3, cuando se hace un análisis $1/N$ de octava se tiene mayor resolución del espectro conforme aumenta el valor de N (ver Fig. 4-22). Teniendo la desventaja de tener que trabajar con un mayor número de rasgos que representan a un sonido. En este trabajo se hicieron experimentos con $1/3$, $1/6$ y $1/12$, los resultados de los experimentos se muestran en la sección 6.2.

Como las señales obtenidas no fueron realizadas con el mismo equipo de grabación, condiciones ambientales, etc. Es importante normalizar los coeficientes después de haber sido extraídos, ya que no es conveniente ingresar los coeficientes considerando la dispersión de los valores. Para esto se hace uso de una normalización que ha sido aplicada al análisis de octavas (Sánchez et al., 2012); primero se usa la ecuación (27) básicamente para eliminar el signo de los coeficientes, con la ecuación (28) se normalizan según el valor máximo y la ecuación (29) impide que el valor máximo (que siempre tendrá valor de 1, debido a la ecuación (28)) atenúe al resto de los valores.

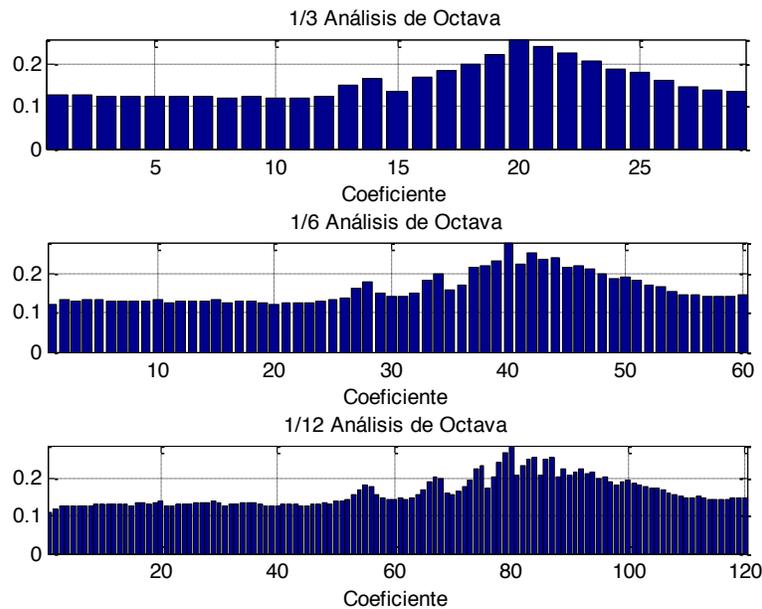


Fig. 4-22. Comparación de resolución del análisis de octava.

$$O(f_i) = \left| \frac{1}{\text{Bandpower}(f_i)} \right| \quad (27)$$

$$H_{norm}(w) = \frac{H(w)}{\max(H(w))} \quad (28)$$

Análisis de las señales acústicas subacuáticas objeto de estudio

$$H_{norm2}(w) = \frac{H_{norm}(w)}{\alpha(\mu + \sigma)} \quad (29)$$

donde $\alpha = 5$.

Para facilitar la extracción de características de las bases de datos, se realizó una aplicación (ver Fig. 4-23); la cual permite ingresar la ruta a donde se encuentran los archivos “.wav” y con solo un botón hacer la extracción automática utilizando diferentes técnicas como LPC, MFCC y octavas, pudiendo establecer la cantidad de rasgos que se desean y creando también los libros de códigos utilizando VQ. Los rasgos extraídos se almacenan en una base de datos, para un manejo más fácil de los mismos; en este estudio se utilizó SQL Server.

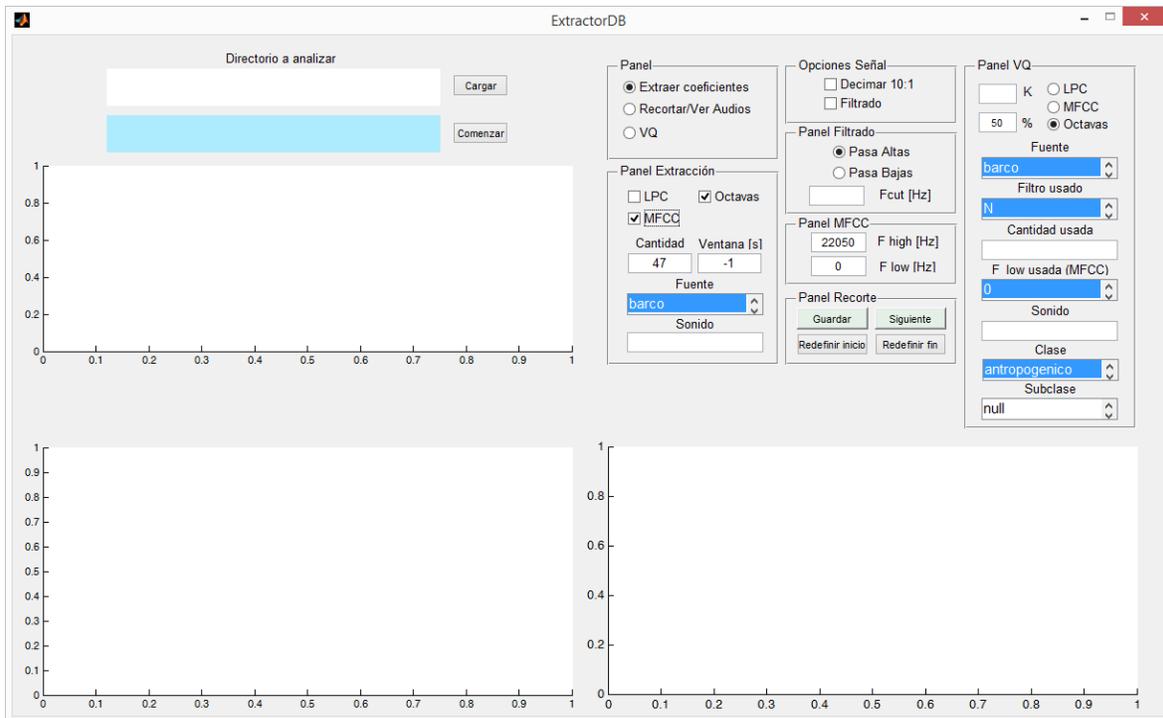


Fig. 4-23. Programa de extracción de características.

4.10. Conclusiones

Se analizaron las señales a clasificar en frecuencia y tiempo, con el fin de saber cómo pre-procesar, detectar y clasificarlas. Se crearon 2 programas, uno para el análisis y etiquetado de las señales y otro programa que permite realizar la extracción de características de las señales, utilizando varias técnicas, con el fin de experimentar con ellas y elegir la mejor.

5. Método de reconocimiento

5.1. Etapa de clasificación

Para realizar la clasificación se hizo un conjunto independiente, teniendo como clasificadores base 4 redes neuronales (su selección se muestra en la sección 6.1) MLP y un combinador o módulo de decisión, basado en votación. Donde cada red se enfoca en un subconjunto diferente de patrones, las 4 redes tienen una función de activación tangente sigmoide hiperbólica (ver Fig. 5-1 y ecuación (30)).

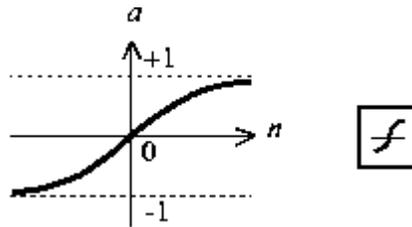


Fig. 5-1. Función tangente sigmoide hiperbólica.

$$a = \frac{e^n - e^{-n}}{e^n + e^{-n}} \quad (30)$$

Los clasificadores base reciben 23, 41 y 60 rasgos, teniendo una arquitectura como se muestra en la Fig. 5-2.

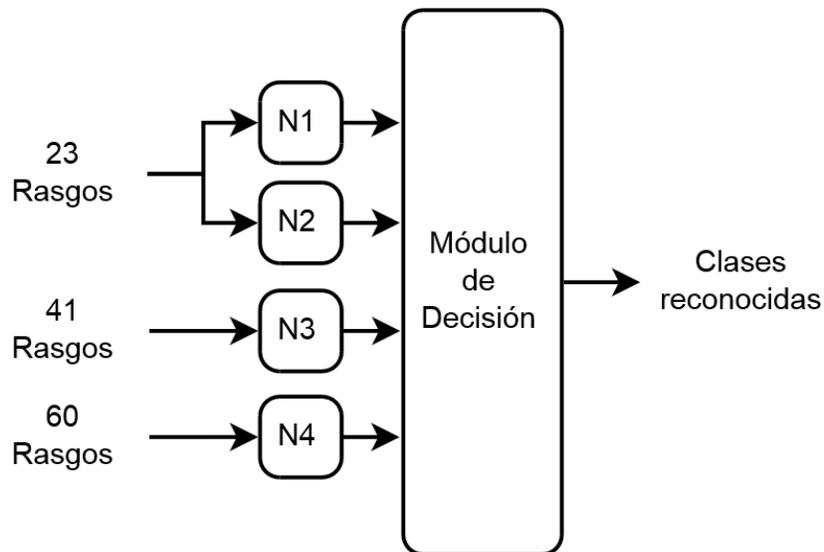


Fig. 5-2. Arquitectura general del conjunto.

Método de reconocimiento

Esta selección de características se hizo después de analizar cuáles rasgos contenían la mayor parte de la información; si recordamos el análisis de octavas es una representación de la energía contenida en bandas de frecuencia, y como se detalló en la sección 4.3, la mayoría de las fuentes emiten en ciertas bandas, lo que hace que la información más importante se agrupe en éstas. En la Fig. 5-3 se muestra el análisis de 1/6 de octava de un silbido de un odontoceto, en ella se puede ver que los últimos coeficientes contienen rasgos más importantes en comparación con los primeros coeficientes. Por otro lado, en la Fig. 5-4 se muestra el llamado de un mysticeto, en el que los primeros coeficientes son los que contienen los rasgos más importantes.

Después de revisar las señales a clasificar, considerando un análisis de 1/6 de octava, se agruparon básicamente en 3 conjuntos (ver ecuaciones (31), (32) y (33)):

1. Altas frecuencias, el cual cubre los últimos 23 coeficientes (cubriendo desde 1.5 kHz)
2. Bajas frecuencias, el cual cubre los primeros 41 coeficientes (cubriendo hasta 2.1 kHz)
3. Todo el espectro, o sea los 60 coeficientes

$$F_1 = \{f(k) \mid k = 38, 39, \dots 60\} \quad (31)$$

$$F_2 = \{f(k) \mid k = 1, 2, \dots 41\} \quad (32)$$

$$F_3 = \{f(k) \mid k = 1, 2, \dots 60\} \quad (33)$$

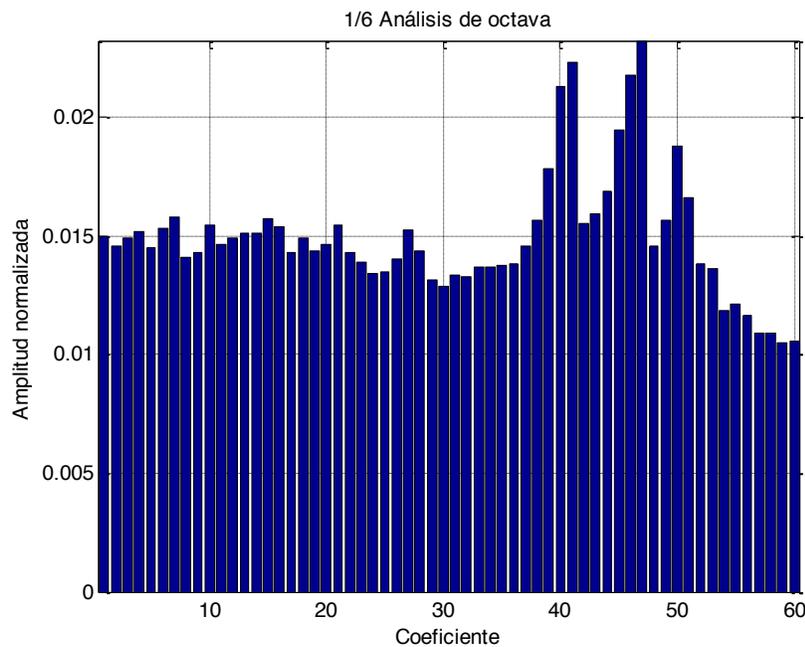


Fig. 5-3. Análisis de octava de un tipo de emisión de un O. Orca.

Método de reconocimiento

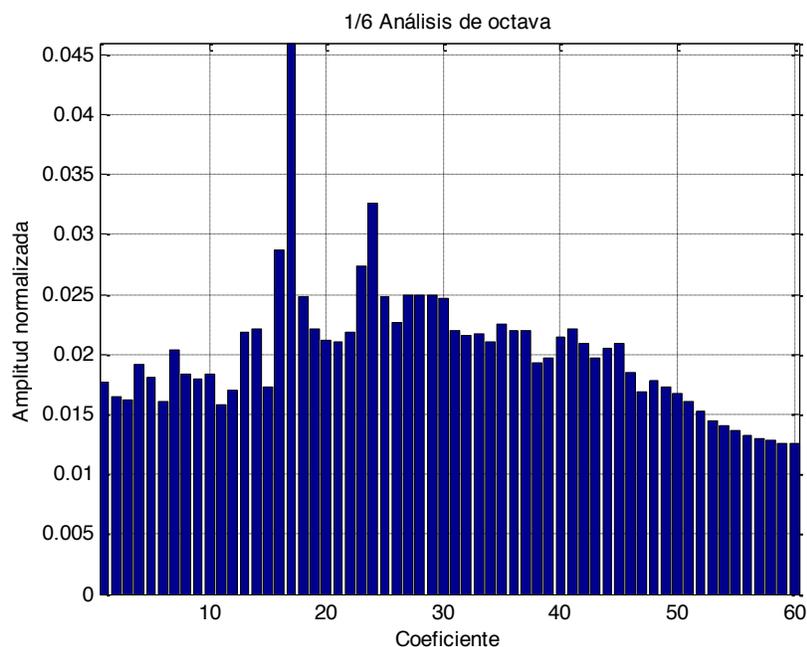


Fig. 5-4. Análisis de octava de una emisión de un *M. Novaeangliae*.

Tabla 8. Fuentes y cantidad de patrones para entrenamiento/prueba.

Fuente	Conjunto de patrones de entrenamiento/prueba
B. Acutorostrata	125/53
M. Novaeangliae	354/152
O. Orca	451/193
P. Crassidens	277/119
S. Frontalis	207/89
T. Truncatus	120/51
P. Macrocephalus	90/38
T. Manatus	265/113
B. Chrysoura	81/35
C. Nebulosus	232/100
P. Cromis	333/142
Natural	61/26
Barco	22/9
Desconocido	161/69
Total	2,779/1,189

5.2. Redes neuronales

Para el entrenamiento de las redes se utilizó un 70% de los patrones para entrenar y un 30% para prueba (ver Tabla 8). En pruebas se experimentó con redes neuronales MLP y LVQ, la razón de usar LVQ era para que al crear un libro de códigos (dada la razonable cantidad de muestras) obtuviera las regiones que mejor representaban a dichos rasgos característicos, y aunque en el entrenamiento la red LVQ obtenía siempre porcentajes de clasificación arriba del 98%, cuando se probaba con el conjunto de prueba, el porcentaje rondaba entre 60 y 65%.

Para determinar cuántos clasificadores base se iban a usar, primero se agruparon por red aquellas clases que utilizaban el mismo conjunto de características (F_1 , F_2 o F_3), después se procedió a realizar el entrenamiento por separado de las redes; cuando se presentaba un bajo porcentaje de clasificación en alguna red, ésta se simplificaba creando más redes o agrupando las clases de diferente forma (ver sección 6.1).

5.3. Módulo de decisión

El módulo de decisión está basado en una decisión por votación, sea la ecuación (34) la salida de todas las redes neuronales.

$$N = \begin{bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{bmatrix} \quad (34)$$

El primer paso es asignarle a cada salida de las 4 redes un peso al momento de hacer la votación, para esto se asignaron pesos equivalentes. En la ecuación (35) se muestra un ejemplo; en él se expresa que para determinar los votos recibidos para la especie O. Orca, implica tomar en cuenta las M salidas de las 4 redes (solo aquellas que se encargan de decir si se trata de esta especie) y a cada salida asignarle un peso de $1/M$.

$$OOrca = \frac{1}{M} salida_1 + \frac{1}{M} salida_2 + \dots + \frac{1}{M} salida_M \quad (35)$$

Por lo cual se genera una matriz de pesos (W), en la que cada fila es una clase y cada columna es una salida de las 4 redes. El último paso en la votación es contar los votos recibidos para cada clase y determinar al ganador, para determinar al ganador simplemente se saca el valor máximo (ver ecuación (36)).

$$C = \underset{i}{\operatorname{argmax}}(WN) \quad (36)$$

5.4.Conclusiones

En este capítulo se analizó el comportamiento de las octavas como rasgos característicos, notando que la carencia de filtros en la etapa de pre-procesamiento puede ser compensada con la elección de diferentes bandas de frecuencia. Por lo que se eligieron los subconjuntos que en análisis de octava representan completamente las señales a clasificar.

También se presentó la arquitectura del modelo de reconocimiento, el cual usa redes neuronales MLP en paralelo, donde cada red se enfoca en un subconjunto diferente del conjunto original de rasgos característicos; reduciendo la cantidad de rasgos con las que cada red se entrenaría y en cierta forma considerando solo los rasgos más importantes para cada clase a reconocer.

Se presentó la forma en la que el modelo realizará la decisión final, considerando las opiniones de cada salida de las redes que componen el conjunto.

6. Pruebas y resultados

6.1. Selección de clasificadores base

En un inicio se planteó la idea de reconocer solo 11 clases (la clase Pez/Sciaenidae no estaba considerada), debido a que se desconocía qué tan bueno sería este tipo de arquitecturas para reconocer una gran variedad de señales acústicas. Por lo cual la primera arquitectura no consideraba las especies B. Chrysoura, C. Nebulosus y P. Cromis. Dejando un conjunto que reconocía solo 11 clases (ver Fig. 6-1).

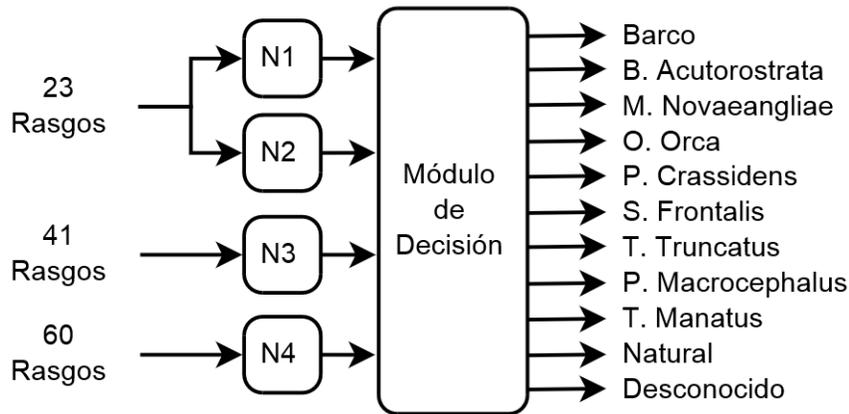


Fig. 6-1. Arquitectura del primer conjunto.

Considerando el método mencionado en la sección 5.2, se hicieron varios experimentos (usando 6 bandas por octava); en el primero se usaron 5 redes neuronales, cada red clasificaba una clase general: Balaenopteridae, Delphinidae, Physeteridae, Trichechidae y Otros (sonidos naturales, antropogénicos y desconocidos). Aunque el porcentaje de clasificación por red era alto, los mismos resultados se obtenían con 4 redes. Sin embargo, cuando se usaban 3 redes o menos, el porcentaje disminuía significativamente. Por lo que se decidió dejar solo 4 redes neuronales MLP.

La primera red (N_1 , ver Fig. 6-2) tiene 3 salidas dedicadas a las clases Delphinidae (en general), Physeteridae (P. Macrocephalus) y Trichechidae (T. Manatus) y una salida que engloba al resto de las clases; para esto la red recibe el subconjunto F_1 .

La segunda red (N_2 , ver Fig. 6-3) tiene 4 salidas dedicadas exclusivamente a la clase Delphinidae, clasificando a las 4 especies: O. Orca, P. Crassidens, S. Frontalis y T. Truncatus; usando el mismo subconjunto que N_1 .

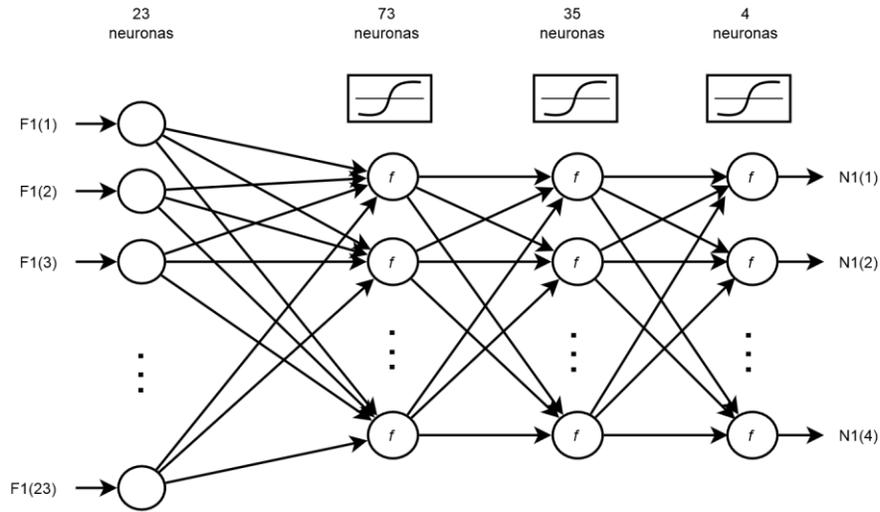


Fig. 6-2. Arquitectura de la red N_1 para 11 clases.

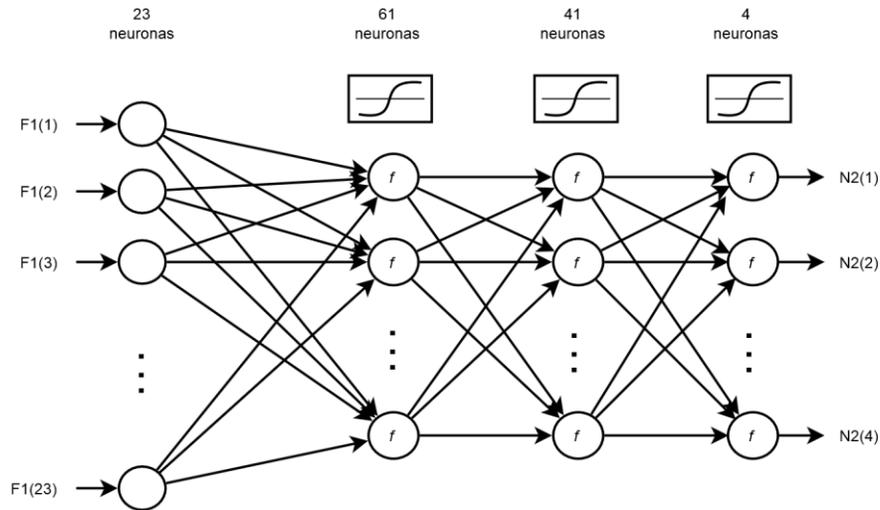


Fig. 6-3. Arquitectura de la red N_2 para 11 clases.

La tercera red (N_3 , ver Fig. 6-4) tiene 2 salidas dedicadas a clasificar entre B. Acutorostrata y M. Novaeangliae, y una salida que engloba al resto de los animales marinos; usando el subconjunto F_2 .

La cuarta red (N_4 , ver Fig. 6-5) tiene 4 salidas dedicadas a las clases Barco, Mamífero marino, Natural y Desconocido; usando el subconjunto F_3 .

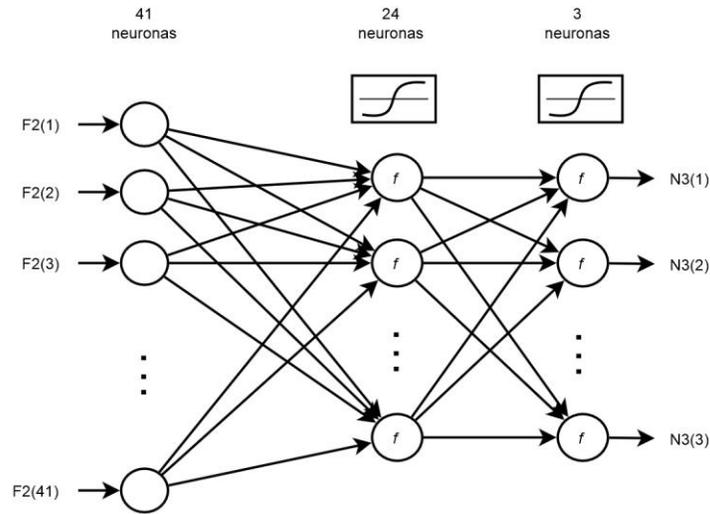


Fig. 6-4. Arquitectura de la red N₃ para las 11 clases.

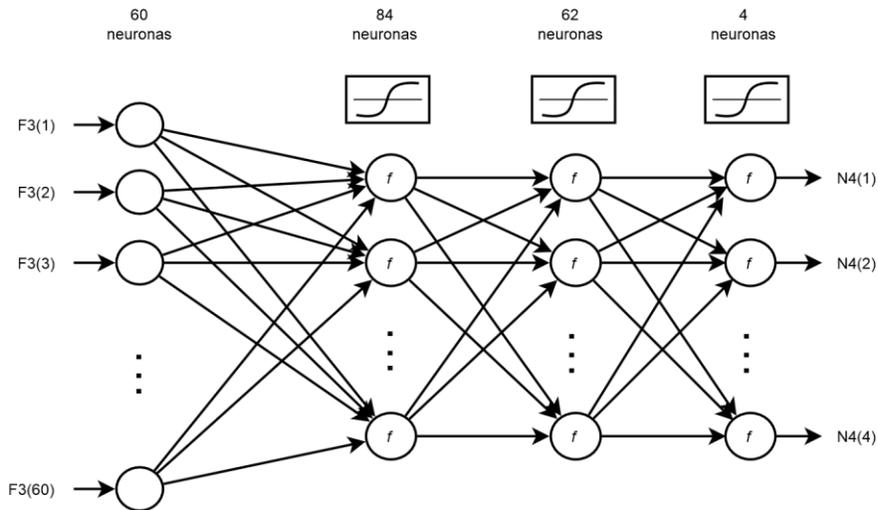


Fig. 6-5. Arquitectura de la red N₄ para las 11 clases.

Con las 4 redes neuronales, usando un análisis de 1/6 de octava se obtuvieron porcentajes por red entre el 93% y 99% para el conjunto de prueba.

6.2. Elección de cantidad de coeficientes

Tras haber probado las redes con 6 bandas por octava, se decidió modificar la arquitectura de las 4 redes para hacer el entrenamiento con 3 y 12 bandas por octava (menor y mayor resolución espectral respectivamente). Cabe destacar que al modificar el

Pruebas y resultados

análisis a 3 y 12 bandas por octava, implicaban una cantidad menor y mayor de coeficientes. Por lo que los conjuntos F_1 , F_2 y F_3 para 3 bandas por octava se convirtieron en las ecuaciones (37), (38) y (39)

$$F_1 = \{f(k) \mid k = 19, 20, \dots 29\} \quad (37)$$

$$F_2 = \{f(k) \mid k = 1, 2, \dots 20\} \quad (38)$$

$$F_3 = \{f(k) \mid k = 1, 2, \dots 29\} \quad (39)$$

Lo mismo ocurrió para 12 bandas por octava

$$F_1 = \{f(k) \mid k = 75, 76, \dots 120\} \quad (40)$$

$$F_2 = \{f(k) \mid k = 1, 2, \dots 82\} \quad (41)$$

$$F_3 = \{f(k) \mid k = 1, 2, \dots 120\} \quad (42)$$

En ambos casos se mantuvo el concepto de analizar bajas frecuencias, altas frecuencias y todo el espectro. Manteniendo las frecuencias centrales de los bancos de filtros lo más parecido posible. Así se obtuvieron los resultados presentados en la Tabla 9.

Tabla 9. Comparación de cantidad de coeficientes para 11 clases.

Mejor entrenamiento	Análisis de octava		
	1/3	1/6	1/12
N_1	91%	95%	92%
N_2	89%	94%	91%
N_3	99%	99%	99%
N_4	96%	96%	96%

En la Tabla 9 se puede ver que el análisis óptimo es con 6 bandas por octava, debido a que el análisis con 3 bandas por octava no es tan bueno para discernir entre las subclases de Delphinidae (O. Orca, P. Crassidens, S. Frontalis y T. Truncatus), y aunque el análisis con 12 bandas da más resolución en frecuencia, extrae el doble de características que el de 6 bandas por octava, lo cual consume más tiempo de procesamiento al extraer las características, al entrenar las redes neuronales y al probar el sistema. Además de que vemos que los porcentajes de clasificación no mejoraron; por esto se elige una extracción de características empleando 6 bandas por octava.

Pruebas y resultados

En las Fig. 6-6 a Fig. 6-9, se muestran las matrices de confusión, en ellas podemos ver el desempeño de los 4 clasificadores.

En el primer clasificador se puede ver que la mayor confusión se presenta en la clase 3 (T. Manatus), la cual llega a ser confundida con la clase 1 (Delphinidae); lo cual era de esperarse debido a la similitud entre los sonidos emitidos.

En el segundo clasificador la mayor confusión ocurre en la clase 4 (T. Truncatus), la cual llega a ser confundida con la clase 1 (O. Orca).

En el tercer clasificador la mayor confusión ocurre en la clase 1 (B. Acutorostrata), la cual tiene una muy ligera confusión con la clase 3 (la cual consta de un agrupamiento de varias clases).

El último clasificador tiene la mayor confusión en la clase 2 (Natural), la cual llega a ser confundida con la clase 1 (Barco).

Salida	1	447 55.3%	5 0.6%	22 2.7%	4 0.5%	93.5% 6.5%		
	2	2 0.2%	33 4.1%	1 0.1%	0 0.0%	91.7% 8.3%		
	3	3 0.4%	0 0.0%	90 11.1%	1 0.1%	95.7% 4.3%		
	4	0 0.0%	0 0.0%	0 0.0%	200 24.8%	100% 0.0%		
		98.9% 1.1%	86.8% 13.2%	79.6% 20.4%	97.6% 2.4%	95.3% 4.7%		
			Target	1	2	3	4	

Fig. 6-6. Matriz de confusión para red N₁ para 11 clases.

Pruebas y resultados

Salida	1	193 42.7%	2 0.4%	1 0.2%	7 1.5%	95.1% 4.9%
	2	0 0.0%	115 25.4%	5 1.1%	5 1.1%	92.0% 8.0%
	3	0 0.0%	0 0.0%	81 17.9%	1 0.2%	98.8% 1.2%
	4	0 0.0%	2 0.4%	2 0.4%	38 8.4%	90.5% 9.5%
		100% 0.0%	96.6% 3.4%	91.0% 9.0%	74.5% 25.5%	94.5% 5.5%
		1	2	3	4	
		Target				

Fig. 6-7. Matriz de confusión para red N₂.

Salida	1	147 18.2%	0 0.0%	0 0.0%	100% 0.0%
	2	0 0.0%	53 6.6%	0 0.0%	100% 0.0%
	3	5 0.6%	0 0.0%	603 74.6%	99.2% 0.8%
		96.7% 3.3%	100% 0.0%	100% 0.0%	99.4% 0.6%
		1	2	3	
		Target			

Fig. 6-8. Matriz de confusión de red N₃ para 11 clases.

Pruebas y resultados

Salida	1	797 87.4%	11 1.2%	1 0.1%	7 0.8%	97.7% 2.3%
	2	7 0.8%	15 1.6%	0 0.0%	0 0.0%	68.2% 31.8%
	3	0 0.0%	0 0.0%	8 0.9%	0 0.0%	100% 0.0%
	4	4 0.4%	0 0.0%	0 0.0%	62 6.8%	93.9% 6.1%
		98.6% 1.4%	57.7% 42.3%	88.9% 11.1%	89.9% 10.1%	96.7% 3.3%
	1	2	3	4		
			Target			

Fig. 6-9. Matriz de confusión de red N₄ para 11 clases.

6.3. Prueba del modelo para 11 clases

Considerando pesos equivalentes para las 15 salidas de las 4 redes y usando los pesos mostrados a continuación, se obtuvo un porcentaje de clasificación del 91.8% para el conjunto de prueba.

$$W = \begin{bmatrix}
 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1/3 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 \\
 1/3 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 \\
 1/4 & 0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 0 & 0 & 0 & 1/4 & 0 & 0 & 0 \\
 1/4 & 0 & 0 & 0 & 0 & 0 & 1/4 & 0 & 1/4 & 0 & 0 & 1/4 & 0 & 0 & 0 \\
 1/4 & 0 & 0 & 0 & 0 & 0 & 1/4 & 0 & 0 & 1/4 & 0 & 1/4 & 0 & 0 & 0 \\
 1/4 & 0 & 0 & 0 & 0 & 0 & 1/4 & 0 & 0 & 0 & 1/4 & 1/4 & 0 & 0 & 0 \\
 1/3 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 \\
 1/3 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\
 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
 \end{bmatrix} \quad (43)$$

6.4. Escalando el modelo

Debido a los buenos resultados del modelo de 11 clases, se decidió escalar la arquitectura para que aceptara las 14 clases (ver Fig. 6-10); analizando las señales de la clase Pez, se decidió que no había necesidad de agregar más redes neuronales, bastaba con modificar 3 de ellas y agregar pesos al módulo de decisión.

Como las emisiones de la clase Pez son de baja frecuencia se optó por agruparla como se hizo con la clase Balaenopteridae, ya que también son de baja frecuencia. Por esta misma razón la red N_2 no se tuvo que modificar, porque esta red solo toma en cuenta las altas frecuencias de la clase Delphinidae. Tras haber modificado la arquitectura de las otras 3 redes, éstas quedaron como se muestra en las Fig. 6-11, Fig. 6-12 y Fig. 6-13.

La red N_1 (ver Fig. 6-11) se modificó porque se agregaron los rasgos de las especies B. Chrysoira, C. Nebulosus y P. Cromis, a la clase que engloba al resto de las clases.

La red N_3 (ver Fig. 6-12) ahora no solo clasifica entre B. Acutorostrata, M. Novaeangliae y otros animales marinos; en esta red se incluyó la clasificación de las especies B. Chrysoira, C. Nebulosus y P. Cromis; por lo que ahora en lugar de tener solo 3 salidas, tiene 6. En cuanto a arquitectura a esta red se le tuvo que agregar una capa más.

En la red N_4 (ver Fig. 6-13) se intentó tener 5 salidas (en lugar de 4), dedicadas a las clases Barco, Mamífero marino, Pez, Natural y Desconocido; sin embargo, el porcentaje de clasificación no era tan bueno como el de la red para 11 clases. Por lo que se decidió probar con 4 salidas, dedicadas a las clases Barco, Animal marino, Natural y Desconocido; obviamente la clase Animal marino contenía a Mamífero marino y Pez con el mismo target. Esta arquitectura logró equipararse a la red para 11 clases.

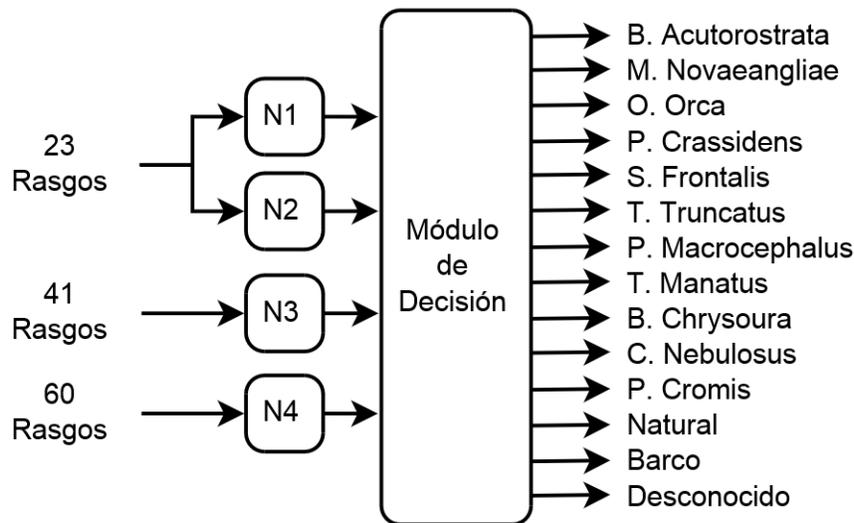


Fig. 6-10. Arquitectura del conjunto final.

Pruebas y resultados

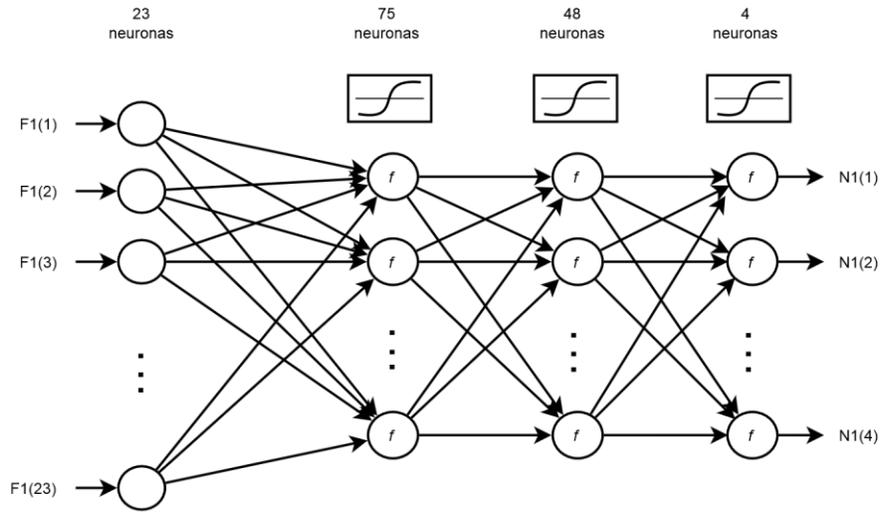


Fig. 6-11. Arquitectura de la red N_1 para 14 clases.

Los resultados de las nuevas redes se muestran en la Tabla 10, los cuales no varían mucho con los de 11 clases.

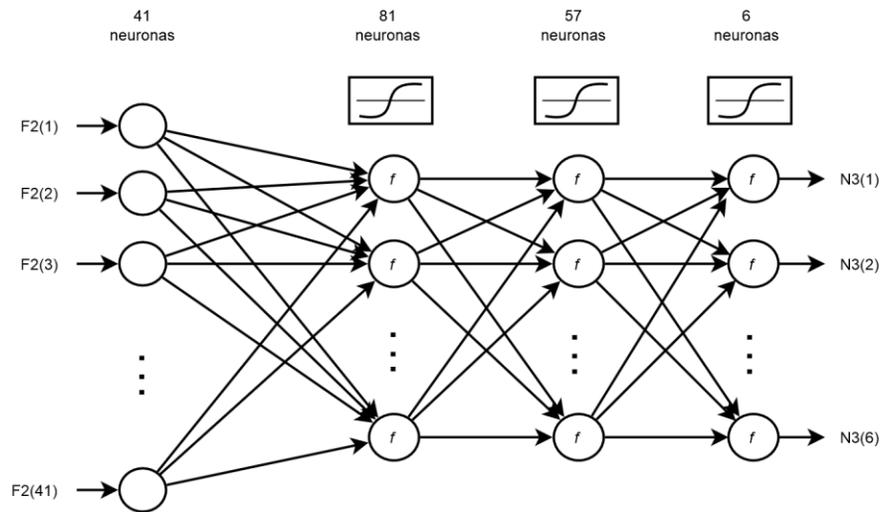


Fig. 6-12. Arquitectura de la red N_3 para 14 clases.

Pruebas y resultados

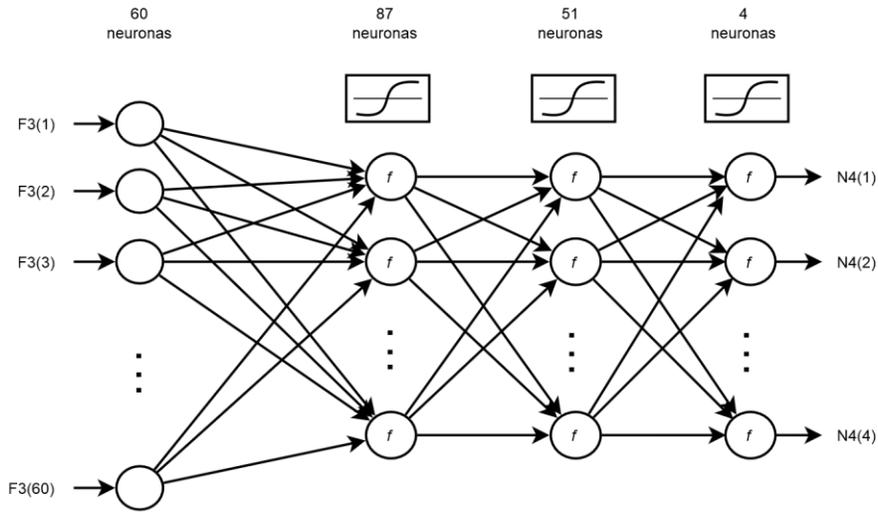


Fig. 6-13. Arquitectura de la red N_4 para 14 clases.

Tabla 10. Resultados por red para 14 clases.

Mejor entrenamiento	Porcentaje de clasificación
N_1	95%
N_2	94%
N_3	99%
N_4	97%

En las Fig. 6-14 a Fig. 6-16, se muestran las matrices de confusión pero para los clasificadores de 14 clases.

En el primer clasificador de nuevo se puede ver que la mayor confusión se presenta en la clase 3 (T. Manatus), con la clase 1 (Delphinidae).

Ahora el tercer clasificador clasifica más clases, la mayor confusión se presenta en la clase 1 (M. Novaeangliae), la cual es confundida con la clase 4 (C. Nebulosus).

El último clasificador, presenta la mayor confusión en la clase 2 (Natural), la cual es confundida con la clase 1 (la cual agrupa mamíferos marinos y peces).

Pruebas y resultados

Salida	1	443 40.8%	6 0.6%	23 2.1%	2 0.2%	93.5% 6.5%
	2	1 0.1%	31 2.9%	1 0.1%	0 0.0%	93.9% 6.1%
	3	6 0.6%	0 0.0%	87 8.0%	0 0.0%	93.5% 6.5%
	4	2 0.2%	1 0.1%	2 0.2%	480 44.2%	99.0% 1.0%
		98.0% 2.0%	81.6% 18.4%	77.0% 23.0%	99.6% 0.4%	95.9% 4.1%
	1	2	3	4	Target	

Fig. 6-14. Matriz de confusión de red N₁ para 14 clases.

Salida	1	148 13.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 0.2%	98.7% 1.3%
	2	0 0.0%	53 4.9%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	3	0 0.0%	0 0.0%	35 3.2%	0 0.0%	0 0.0%	1 0.1%	97.2% 2.8%
	4	2 0.2%	0 0.0%	0 0.0%	100 9.2%	0 0.0%	0 0.0%	98.0% 2.0%
	5	1 0.1%	0 0.0%	0 0.0%	0 0.0%	141 13.0%	0 0.0%	99.3% 0.7%
	6	1 0.1%	0 0.0%	0 0.0%	0 0.0%	1 0.1%	600 55.3%	99.7% 0.3%
		97.4% 2.6%	100% 0.0%	100% 0.0%	100% 0.0%	99.3% 0.7%	99.5% 0.5%	99.3% 0.7%
	1	2	3	4	5	6	Target	

Fig. 6-15. Matriz de confusión de red N₃ para 14 clases.

Pruebas y resultados

Salida	1	1078 90.7%	11 0.9%	3 0.3%	11 0.9%	97.7% 2.3%
	2	2 0.2%	15 1.3%	0 0.0%	1 0.1%	83.3% 16.7%
	3	2 0.2%	0 0.0%	6 0.5%	2 0.2%	60.0% 40.0%
	4	3 0.3%	0 0.0%	0 0.0%	55 4.6%	94.8% 5.2%
		99.4% 0.6%	57.7% 42.3%	66.7% 33.3%	79.7% 20.3%	97.1% 2.9%
	1	2	3	4	Target	

Fig. 6-16. Matriz de confusión de red N_4 para 14 clases.

Considerando de nuevo pesos equivalentes, ahora para las 18 salidas de las 4 redes y usando los pesos mostrados a continuación, se obtuvo un porcentaje de clasificación del 92.5% para el conjunto de prueba.

Pruebas y resultados

Tabla 11. Comparativa con el estado del arte.

Método empleado	Fuentes clasificadas	Porcentaje de clasificación
Rasgos temporales y espectrales, y clasificador jerárquico (Abousleiman et al., 2013)	1	85%
Rasgos temporales y espectrales, y redes de base radial (André et al., 2011)	2	97% - 100%
Rasgos temporales y espectrales, y redes de base radial (Zaugg et al., 2010)	2	94% - 98%
Rasgos temportales y espectrales, y bosques aleatorios (Oswald et al., 2011)	8	62%
Análisis de octava, y conjunto de MLPs (Este trabajo)	11	91%
	14	92%

6.5. Análisis de tiempos

Como se mencionó en el estado del arte, este tipo de aplicaciones deben de ser algoritmos rápidos, debido al tamaño de datos que deben de procesar y a que deben ahorrar la mayor cantidad de energía posible. Debido a que ambos modelos (11 y 14 clases) tienen arquitecturas similares, ambos se comportan del mismo modo en cuanto a tiempos de ejecución se refiere.

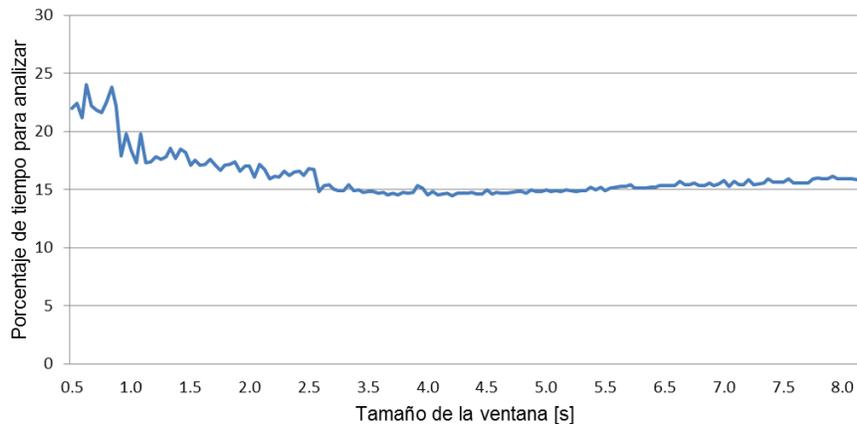


Fig. 6-17. Porcentaje de tiempo de análisis con base en la ventana.

Pruebas y resultados

Se probó el modelo con diferentes tamaños de ventana a 44.1 kHz, considerando que la señal a leer puede ser desde un pequeño silbido o ráfaga de pulsos, hasta un silbido o canto que dure varios segundos. Los resultados se muestran en la Fig. 6-17, en ella se puede ver que a partir de que el tamaño de la ventana es mayor que 3.5 s, el algoritmo se estabiliza y consigue hacer la clasificación de la señal en el 15% del tiempo del tamaño de la ventana. Aun cuando el tamaño de la ventana es corto, el algoritmo no tomaba más del 25% para completar el análisis.

6.6. Conclusiones

Se probaron los rasgos extraídos con redes provisionales, obteniendo los mejores resultados con las octavas, utilizando los subconjuntos considerados en el capítulo anterior. Se probaron varias bandas por octava, llegando a la conclusión que el mejor resultado se obtenía con el análisis de 1/6 de octava.

Se modificó la cantidad de redes neuronales para determinar cuántos clasificadores eran necesarios en el conjunto, se probó que con 4 redes se obtenía el mejor desempeño.

En este capítulo se experimentó con 2 arquitecturas, una que clasificaba 11 clases y otra que clasificaba 14. Se hicieron pruebas experimentales para evaluar el desempeño de las redes por separado, obteniendo porcentajes de clasificación por arriba del 90% con las 2 arquitecturas. Asimismo se presentaron las pruebas de las redes como conjunto, obteniendo un porcentaje de 91% para la arquitectura de 11 clases y 92% para la de 14 clases.

Se determinaron los tiempos requeridos de análisis, para saber si la arquitectura propuesta podría ser implementada en un sistema de tiempo real, en el que debe de cumplir con cotas de tiempo. Se encontró que el modelo nunca rebasaba un 20% de la duración de la señal analizada.

Conclusiones

Conclusiones

Con referencia al objetivo general y los objetivos particulares, se presentan las siguientes conclusiones:

- Si bien las señales acústicas subacuáticas pueden generarse en múltiples lugares como lagos, mares, piscinas, etc. No existen bases de datos de todos los sonidos, debido a que algunos son de uso comercial, militar, etc. Por lo que se analizaron diferentes bases de datos en línea, así como algunos enlaces en descarga continua (“streaming”) a observatorios oceánicos. Estos últimos generan una cantidad inmensa de datos, sin embargo, al tratarse de grabaciones “crudas” puede darse el caso de que después de 3 horas de datos almacenados, no se genere ningún evento, y en caso de generarse no se conoce exactamente qué fue lo que produjo dicho evento, lo cual complica demasiado el uso de este tipo de fuente de datos. Lo más complejo de evaluar las bases de datos fue distinguir los sonidos que se generaban, ya que por lo general el etiquetado de sonidos lo realiza una persona que conoce las firmas acústicas de ciertos animales, barcos, etc. Y mucha gente asocia que los animales generan un solo tipo de sonido; es muy fácil para una persona distinguir en una grabación la presencia de un ladrido, un maullido, el canto de un ave, etc. Pero los animales marinos producen sonidos que no son comunes y una especie puede producir varios sonidos que no se parecen uno del otro.
- Este trabajo se enfocó en sonidos generados en mares debido a la riqueza natural con la que cuenta México, por lo cual más de la mitad de las clases a reconocer son animales marinos, sin embargo, debido a tal riqueza, el trabajo tuvo que acotarse solo a la zona del Golfo de México, aun así faltaron muchas clases por agregar al modelo, ya sea porque no generan sonidos o porque no se lograron encontrar registros que pudieran ser usados en este trabajo.
- De las técnicas de extracción de características probadas, el análisis de octavas demostró un mejor desempeño con las 14 clases utilizadas, en particular el análisis de 1/6 de octava. Gracias al uso de este análisis, la señal leída no se tiene que centrar; ya que solo considera los niveles de energía por banda, no importando en qué momento ocurrieron estos. Sin embargo, esto se convierte en una limitante, ya que los rasgos no dan características a detalle de las señales, por ejemplo sería muy difícil determinar de qué tipo de silbido se trataba (cóncavo, convexo, etc.), a menos que se hiciera un barrido de la señal detectada y se fuera analizando la evolución en el tiempo de la energía por bandas; lo cual consumiría más tiempo

Conclusiones

del que se usó en el método descrito. Lo que podría funcionar sería el estimado de individuos, pero para caracterizar el sistema se tendrían que conocer más detalles de cómo fueron realizadas las grabaciones (distancias, profundidades, equipo de grabación, etc.).

- Se desarrolló un modelo computacional basado en 4 redes MLP trabajando en paralelo, capaz de reconocer señales acústicas subacuáticas. Obteniendo buenos resultados en la clasificación de 11 y 14 clases de fuentes, las cuales incluyen mamíferos marinos, peces, sonidos naturales, barcos y sonidos desconocidos. Se probó que el uso de una metodología en conjunto permite que un problema complejo pueda dividirse en varios problemas mucho más sencillos.
- Se obtuvieron porcentajes de clasificación del 91-92%, para 11 y 14 clases respectivamente (probados con 1,189 registros); los cuales son buenos resultados, considerando la cantidad de clases a reconocer y las dificultades que representan los sonidos generados por algunos animales marinos.
- Se comprobó que la arquitectura del modelo de reconocimiento mantiene buenos porcentajes de clasificación, incluso si se agregan más clases. Debido a su misma arquitectura tiene la posibilidad de que sea paralelizado (ver Anexo D), lo cual aceleraría el proceso de reconocimiento; sin embargo, se probó que trabajando en un solo núcleo, el proceso completo de clasificación se logra en un 20% del tiempo de la duración la señal.

En resumen, la hipótesis fue comprobada y los objetivos planteados se cumplieron satisfactoriamente.

Trabajo futuro

Como trabajo a futuro se consideran las siguientes oportunidades de estudio y mejoras a los algoritmos y al modelo empleado en este trabajo:

- Dependiendo del área de interés donde este tipo de modelo quiera ser empleado, se debería de considerar la extracción de otro tipo de rasgos característicos, no tanto para mejorar la clasificación, sino para darle al usuario final más información sobre el sonido clasificado, algo que solo usando octavas es muy difícil. De la misma manera, dependiendo del área, se debería de trabajar en un sistema de detección de eventos mucho más eficiente; en algunos sistemas automáticos, lo que se hace es determinar cuántos eventos por bloque leído se generaban, si no se rebasaba cierta cantidad, el bloque no pasaba a la etapa de análisis.
- Debido al buen desempeño de la arquitectura empleada, ésta debería de ser evaluada en pruebas de campo, tal vez con menos y diferentes clases; para esto el Laboratorio de Sistemas Inteligentes para la Automatización de este centro de investigación, cuenta con el equipo necesario para realizar mediciones de campo (ver Anexo E), debido a cuestiones de tiempo no se pudieron realizar este tipo de mediciones, pero a futuro permitiría ampliar las clases a reconocer, generando una base de datos propia, con firmas acústicas de cualquier tipo de señal acústica subacuática; la desventaja en cuanto a las mediciones de campo radica en que se debe de estar completamente seguro de qué fue lo que emitió el sonido, no es lo mismo conseguir firmas acústicas de barcos por ejemplo, los cuales se ven claramente en la superficie a conseguir sonidos de animales marinos, los cuales pueden estar mezclados y a varios metros de profundidad.
- En caso de querer mejorar el porcentaje de clasificación del modelo propuesto, la opción que parece más recomendable sería tomar en cuenta los rasgos característicos en espacio y tiempo, similares a los que se mostraron en la Fig. 4-10. Estos rasgos dan mucho más detalle de la evolución de la señal en tiempo, los cuales cambian bastante entre familias de animales. Otra posible solución es el uso de cadenas de Markov (tal vez en paralelo), debido al comportamiento de las señales producidas por varios mamíferos marinos.

Referencias bibliográficas

Referencias bibliográficas

- Abousleiman, R., Qu, G., & Rawashdeh, O. (2013). North Atlantic right whale contact call detection. *ICML 2013, Proceedings: Workshop on Machine Learning for Bioacoustics*, Atlanta, GA.
- André, M., van der Schaar, M., Zaugg, S., Houégnigan, L., Sánchez, A.M., & Castell, J.V. (2011). Listening to the Deep: Live monitoring of ocean noise and cetacean acoustic signals. *Marine Pollution Bulletin*, 63, 18-26. doi:10.1016/j.marpolbul.2011.04.038
- ANSI. (2004). Standard S1.11-2004: Specification for octave band and fractional octave band analog and digital filters: American National Standards Institute.
- Atal, B. S., & Hanauer, S. (1971). Speech analysis and synthesis by linear prediction of the speech wave. *Journal of the Acoustical Society of America*, 50, 637-655. doi:10.1121/1.1912679
- Au, W. W. L., & Hastings, M. C. (2008). Principles of marine bioacoustics. New York, NY: Springer.
- Bazúa-Durán, M. C. (1997). Comparisons of whistles among different groups dolphins (*Tursiops truncatus*) from the Gulf of Mexico, MS Thesis, Universidad Nacional Autónoma de México, México, D.F.
- Brown, G., Wyatt, J., Harris, R., & Yao, X. (2005). Diversity creation methods: a survey and categorisation. *Journal of Information Fusion*, 6(1), 5-20.
- Brown, J. C., & Miller, P. J. O. (2007). Automatic classification of killer whale vocalizations using dynamic time warping. *Journal of the Acoustical Society of America*, 122(2), 1201-1207. doi:10.1121/1.2747198
- Brown, J. C., & Smaragdis, P. (2009). Hidden Markov and Gaussian mixture models for automatic call classification. *Journal of the Acoustical Society of America, JASA Express Letters*, 125(6), 221-224. doi:10.1121/1.3124659
- Caldwell, M. C. & Caldwell, D. K. (1965). Individualized whistle contours in bottlenosed dolphins (*Tursiops truncatus*). *Science*, 207, 434-435.
- Caldwell, D. K., & Caldwell, M. C. (1970). Etiology of the chirp sounds emitted by the Atlantic bottlenose dolphin: a controversial issue. *Underwater Naturalist*, 6, 6-8.
- Carey, W. M., & Evans R.B. (2011). Ocean Ambient Noise Measurement and Theory. New York, NY: Springer.

Referencias bibliográficas

- Chen, C., Lee, J., & Lin, M. (1998). Classification of underwater signals using wavelet transforms and neural networks. *Mathematical and Computer Modelling*, 27(2), 47–60. doi:10.1016/S0895-7177(97)00259-8
- Chung, K. W., Sutin, A., Sedunov, A., & Bruno, M. (2011). DEMON Acoustic Ship Signature Measurements in an Urban Harbor. *Advances in Acoustics and Vibration*, 2011. doi:10.1155/2011/952798
- Clark, C. W. (1990). Acoustic behavior of mysticete whales, in Sensory Abilities of Cetaceans, J. Thomas and R. Kastelein, eds. (Plenum, New York), 571–583.
- Clark, C. W., Marler, P., & Beeman, K. (1987). Quantitative analysis of animal vocal phonology: an application to swamp sparrow song. *Ethology*, 76(2), 101-115. doi:10.1111/j.1439-0310.1987.tb00676.
- Council Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive). OJ L 164/19.*
- Dasarathy, B. V., & Sheela, B. V. (1979). Composite classifier system design: concepts and methodology. *Proceedings of the IEEE*, 67(5), 708-713. doi:10.1109/PROC.1979.11321
- Deecke, V. B., Ford, J. K. B., & Spong, P. (1999). Quantifying complex patterns of bioacoustic variation: Use of a neural network to compare killer whale (*Orcinus orca*) dialects. *Journal of the Acoustical Society of America*, 105(4), 2499-2507. doi:10.1121/1.426853
- Dougherty, G. (2013). Pattern Recognition and Classification, An introduction. New York, NY: Springer. doi:10.1007/978-1-4614-5323-9
- Dreher, J. L. & Evans, (1964). Cetacean Communication, in Marine Bio-Acoustics, N. Tavolga, ed. (Pergamon, Oxford, England), 373–393.
- Driscoll, A. D. (1995). The Whistles of Hawaiian Spinner Dolphins, MS Thesis, University of California. Santa Cruz, CA.
- Dugan, P. J., Rice, A. N., Urazghildiiev, I. R., & Clark, C. W. (2010a). North Atlantic right whale acoustic signal processing: Part I. Comparison of machine learning recognition algorithms. IEEE Proceedings of the 2010 Long Island Systems, Applications and Technology Conference, 1-6, Farmingdale, NY. doi:10.1109/LISAT.2010.5478268
- Dugan, P. J., Rice, A. N., Urazghildiiev, I. R., & Clark, C. W. (2010b). North Atlantic right whale acoustic signal processing: Part II. Improved decision architecture for auto-detection using multi-classifier combination methodology. IEEE Proceedings of the

Referencias bibliográficas

- 2010 Long Island Systems, Applications and Technology Conference, 1–6, Farmingdale, NY. doi:10.1109/LISAT.2010.5478287
- Durbin, J. (1960). The fitting of time series models. *Revue de l'Institut International de Statistique*, 28(3), 233-244. doi:10.2307/1401322
- Eom, K., Wellman, M., Srour, N., Hillis, D., & Chellappa, R. (1997). Acoustic target classification using multiscale methods. *Proceedings of 1997: Sensors and Electron Devices Symposium*, College Park, MD.
- Evans, W. E. (1967). Vocalization among marine animals, in *Marine Bioacoustics Vol 2*, W. N. Tavolga, ed. (Pergamon, New York), 159–186.
- Felder, D. L., & Camp, D.K. (2009). *Gulf of Mexico: Origin, Waters and Biota, Volume I: Biodiversity*. Texas A&M University Press.
- Gillespie, D. (2004). Detection and classification of right whale calls using an edge detector operating on a smoothed spectrogram. *Journal of Canadian Acoustics*, 32(2), 39–47.
- Gillespie, D., Gordon, J., McHugh, R., McLaren, D., Mellinger, D. K., Redmond, P., Thode, A., Trinder, P. & Deng X. Y. (2009). PAMGUARD: Semiautomated, open-source software for real-time acoustic detection and localization of cetaceans. *Journal of the Acoustical Society of America*, 125(4), 2547. doi:10.1121/1.4808713
- Giraudet, P., & Glotin, H. (2006). Real-time 3D tracking of whales by precise and echo-robust TDOAs of clicks extracted from 5 bottom-mounted hydrophones records of AUTECH, *Applied Acoustics*, 67, 1106-1117.
- Hansen, L. K., & Salamon, P. (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 12(10), 993-1001. doi:10.1109/34.58871
- Hartman, D. S. (1979). Ecology and behavior of the manatee (*Trichechus manatus*) in Florida, *American Society of Mammalogists*, 5.
- Haykin, S. (1998). *Neural networks: A comprehensive foundation*. Upper Saddle River, NJ: Prentice-Hall.
- Herman, L. M. & Tavolga, W. N. (1980). The Communication Systems of Cetaceans, in *Cetacean Behavior: Mechanisms and Function*, L. H. Herman, ed. (Wiley-Interscience, New York), 149–209.
- Howell, B. P., Wood, S., & Koksall, S. (2003). Passive sonar recognition and analysis using hybrid neural networks. *IEEE proceedings of Oceans 2003*, 4, 1917–1924, San Diego, CA. doi:10.1109/OCEANS.2003.178182
- Hu, X. (2001). Using rough sets theory and database operations to construct a good ensemble of classifiers for data mining applications. *Proceedings of the IEEE*:

Referencias bibliográficas

- International Conference on Data Mining*, 233-240. San Jose, CA. doi:10.1109/ICDM.2001.989524
- Huang, J., Zhao, J., & Xie, Y. (1997). Source classification using pole method of AR model. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1, 567–570, Munich. doi:10.1109/ICASSP.1997.599701
- Huang, X., Acero, A., & Hon, H. (2001). Spoken language processing: A guide to theory, algorithm, and system development. Upper Saddle River, NJ: Prentice Hall.
- IEC. (1995). Standard IEC1260: Electroacoustics – Octave-band and fractional-octave-band filters: International Electrotech Commission.
- Kang, C., Zhang, X., Zhang, A., & Lin, H. (2004). Underwater acoustic targets classification using welch spectrum estimation and neural networks. In F. Yin, J. Wang & C. Guo (Eds.), *Advances in Neural Networks – ISNN 2004* Vol. 3173 (pp. 930–935). Berlin: Springer. doi:10.1007/978-3-540-28647-9_153
- Krogh, A., & Vedelsby, J. (1995). Neural networks ensembles, cross validation and active learning. *Advances in neural information processing systems*, 7, 231-238. doi:10.1016/j.neucom.2014.08.100
- Kuncheva, L. (2005). Combining pattern classifiers. New York, NY: Wiley Press,
- Kuncheva, L., & Whitaker, C. (2003). Measures of diversity in classifier ensembles and their relationship with ensemble accuracy. *Machine Learning*, 51(2), 181-207. doi:10.1023/A:1022859003006
- Lammers, M. O. & Au, W. W. L. (1996). Broadband recording of social acoustic signals of the Hawaiian Spinner and spotted dolphins. *Journal of the Acoustical Society of America*, 100, 2609. doi:10.1121/1.417642
- Levinson, N. (1947). The Wiener RMS error criterion in filter design and prediction. *Journal of Mathematics and Physics*, 25(1), 261-278.
- Ma, Y., & Chen, K. (2008). A time-frequency perceptual feature for classification of marine mammal sounds. *9th International Conference on Signal Processing*, 2820–2823, Beijing. doi:10.1109/ICOSP.2008.4697734
- Mackenzie, K.V. (1981). Nine-term equation for sound speed in the oceans. *Journal of the Acoustical Society of America*, 70, 807-812.
- Makhoul, J. (1976). Correction to linear prediction: A tutorial review. *Proceedings of the IEEE*, 64(2), 285. doi:10.1109/PROC.1976.10102
- Márquez-Molina, M., Sánchez-Pérez, L. A., Sánchez-Fernández, L. P., & Suárez-Guerra, S. (2014). Aircraft take-off noises classification based on human auditory's matched

Referencias bibliográficas

- features extraction. *Applied Acoustics*, 84, 83-90. doi:10.1016/j.apacoust.2013.12.003
- Mazhar, S., Ura, T., & Bahl, R. (2007). Vocalization based Individual Classification of Humpback Whales using Support Vector Machine. *IEEE proceedings of OCEANS 2007*, 1-9, Vancouver, BC. doi:10.1109/OCEANS.2007.4449356
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.
- Mellinger, D. K. (2004). A comparison of methods for detecting right whale calls. *Journal of Canadian Acoustics*, 32(2), 55–65.
- Mellinger, D. K., & Bradbury, J. W. (2007). Acoustic measurement of marine mammal sounds in noisy environments. *Proceedings of Second International Conference on Underwater Acoustic Measurements: Technologies and results*, Harklion, Greece, 273-280.
- Mellinger, D. K., & Clark, C. W. (1993). A method for filtering bioacoustic transients by spectrogram image convolution. *IEEE proceedings of OCEANS 1993*, 3, 122–127, Victoria, BC. doi:10.1109/OCEANS.1993.326171
- Nishiwaki, M. & Marsh, H. (1985). Dugong (Dugong dugon) (Müller, 1776), in *Handbook of Marine Mammals*, Vol. 3, S. H. Ridgway and R. Harrison, eds. (Academic Press, London), 1–31.
- Oppenheim, A. V., Kopec, G. E., & Tribolet, J. M. (1976). Signal analysis by homomorphic prediction. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(4), 327–332. doi:10.1109/TASSP.1976.1162828
- Oswald, M., Oswald, J. N., Lammers, M. O., Rankin, S., & Au, W. W. L. (2011). Integration of real-time odontocete call classification algorithm into PAMGUARD signal processing software. *Journal of the Acoustical Society of America*, 129, 2639. doi:10.1121/1.3588787
- Pace, F., Benard, F., Glotin, H., Adam, O., & White, P. (2010). Subunit definition and analysis for humpback whale call classification. *Applied Acoustics*, 71, 1107-1112. doi:10.1016/j.apacoust.2010.05.016
- Picone, J. W. (1993). Signal modeling techniques in speech recognition. *Proceedings of the IEEE*, 81(9), 1215-1247. doi: 10.1109/5.237532
- Potter, J. R., Mellinger, D. K., & Clark, C. W. (1994). Marine mammal call discrimination using artificial neural networks. *Journal of the Acoustical Society of America*, 96(3), 1255–1262.

Referencias bibliográficas

- Pourhomayoun, M., Dugan, P., Popescu, M., & Clark, C. (2013). Bioacoustic signal classification based on continuous region processing, grid masking and artificial neural network. *ICML 2013, Proceedings: Workshop on Machine Learning for Bioacoustics*, 61-66, Atlanta, GA.
- Provost, F. J., & Kolluri, V. (1999). A survey of methods for scaling up inductive learning algorithms. *Data Mining and Knowledge Discovery*, 3(2), 131-169. doi:10.1023/A:1009876119989
- Robinson, C. (2015). *Información*. [robmen@unam.mx].
- Roch, M. A., Soldevilla, M. S., Burtenshaw, J. C., Henderson, E. E., & Hildebrand, J. A. (2007). Gaussian mixture model classification of odontocetes in the southern California bight and the Gulf of California. *Journal of the Acoustical Society of America, JASA Express Letters*, 121(3), 1737-1748. doi:10.1121/1.2400663
- Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1-2), 1-39. doi:10.1007/s10462-009-9124-7
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536.
- Sánchez, L. P., Sánchez, L. A., Carbajal, J. J., & Rojo, A. (2012). Aircraft classification and acoustic impact estimation based on real-time take-off noise measurements. *Neural Processing Letters*, 38(2), 239-259. doi:10.1007/s11063-012-9258-5
- Sánchez-Pérez, L. A., Sánchez-Fernández, L. P., Suárez-Guerra, S., & Carbajal-Hernández, J. J. (2013). Aircraft class identification based on take-off noise signal segmentation in time. *Expert Systems with Applications*, 40(13), 5148-5159. doi:10.1016/j.eswa.2013.03.017
- Schevill, W. E. & Watkins, W. A. (1965b). Underwater calls of Trichechus Manatee). *Nature*, 205, 373-374.
- Schroeder, M. R. (1977). Recognition of complex acoustic signals. *Life sciences research report*, Berlin, 5, 323-328.
- Scowcroft, G., Vigness-Raposa, K., Knowlton, C., & Morin, H. (2012). Discovery of Sound in the Sea. University of Rhode Island.
- Shannon, B. J., & Paliwal, K. K. (2003). A comparative study of filter bank spacing for speech recognition. *Microelectronic engineering research conference*.
- Soares-Filho, W., Seixas, J. M., & Caloba, L. P. (2002). Enlarging neural class detection capacity in passive sonar systems. *IEEE International Symposium on Circuits and Systems*, 3, 105-108, Scottsdale, AZ. doi:10.1109/ISCAS.2002.1010171

Referencias bibliográficas

- Stevens, S. S., Volkman, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude of pitch. *Journal of the Acoustical Society of America*, 8(3), 185-190. doi:10.1121/1.1915893
- Stirling, I. (1988). *Polar Bears*. University of Michigan Press, Ann Arbor.
- Taruski, A. G. (1979). The whistle repertoire of the North Atlantic pilot whale (*Globicephala melaena*) and its relationship to behavior and environment, in *Behavior of Marine Animal: Current Perspectives in Research*, H. E. Winn and B. L. Olla, eds. (Plenum Press, New York), 345–368.
- Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley, Reading.
- Tumer, K., & Ghosh, J. (1996). Error correlation and error reduction in ensemble classifiers. *Connection science, special issue on combining artificial neural networks: ensemble approaches*. 8(3-4), 385-404.
- Urazghildiiev, I. R., Clark, C. W., Krein, T. P., & Parks, S. E. (2009). Detection and recognition of North Atlantic right whale contact calls in the presence of ambient noise. *IEEE Journal of Oceanic Engineering*, 34(3), 358–368. doi:10.1109/JOE.2009.2014931
- Urazghildiiev, I., Clark, C. W., & Krein, T. (2008). Acoustic detection and recognition of fin whale and North Atlantic right whale sounds. *Proceedings of the IEEE Workshop on New Trends for Environmental Monitoring using Passive Systems*, 1–6, Hyeres, French Riviera. doi:10.1109/PASSIVE.2008.4786994
- Winn, H. E. & Winn, L. K. (1978). The song of the humpback whale *Megaptera novaeangliae* in the West Indies. *Marine Biology*, 47, 97–114. doi:10.1007/BF00395631
- Zaugg, S., van der Schaar, M., Houégnigan, L., Gervaise, C., & André, M. (2010). Real-time acoustic classification of sperm whale clicks and shipping impulses from deep-sea observatories. *Applied Acoustics*, 71(11), 1011-1019. doi:10.1016/j.apacoust.2010.05.005
- Zelick, R., Mann D. A., & Popper, A. N. (1999). Acoustic communication in fishes and frogs. In R. R. Fay & A. N. Popper (Eds.), *Comparative hearing: Fishes and Amphibians – Vol. 11* (363–411). New York, NY: Springer. doi:10.1007/978-1-4612-0533-3_9
- Zheng, F., Zhang, G., & Song, Z. (2001). Comparison of different implementations of MFCC. *Journal of Computer Science & Technology*, 16(6), 582–9.

Referencias electrónicas

Avisoft Bioacoustics. (2015). Hardware and Software for Investigating Animal Acoustic Communication. Recuperado de <http://www.avisoft.com/>

CHRISAR (2015). CHRISAR Software Technologies. Recuperado de <http://www.chrisar.fr/>

CONACYT. (2015). Fondo sectorial de investigación y desarrollo en ciencias navales, Demandas específicas, convocatoria SEMAR-CONACYT 2013-01. Recuperado de <http://www.conacyt.mx/index.php/el-conacyt/convocatorias-y-resultados-conacyt/convocatorias-fondos-sectoriales-constituidos/convocatoria-semar-conacyt/convocatorias-abiertas-semar-conacyt/convocatoria-2013-01-2/594-demandas-especificas-2013/file>

DOSITS. (2015). *Discovery of Sound in the Sea*. Recuperado de <http://www.dosits.org/>

EUR-Lex El acceso al Derecho de la Unión Europea. (2015). Asuntos marítimos y pesca. Recuperado de http://eur-lex.europa.eu/summary/chapter/maritime_affairs_and_fisheries.html?locale=es&root_default=SUM_1_CODED=02

Glotin. (2015). ONCET. Recuperado de <http://glotin.univ-tln.fr/oncet/>

Greenpeace. (2015). Biodiversidad del Golfo de México en riesgo por derrame petrolero. Recuperado de <http://www.greenpeace.org/mexico/es/Noticias/2010/Mayo/biodiversidad-del-golfo-de-mex/>

INAPESCA. (2015). Carta Nacional Pesquera. Recuperado de <http://www.inapesca.gob.mx/portal/documentos/publicaciones/carta-nacional-pesquera/Carta-Nacional-Pesquera-2012.pdf>

Marine conservation research. (2015). Research software. Recuperado de <http://www.marineconservationresearch.co.uk/downloads/logger-2000-rainbowclick-software-downloads/>

Museum für naturkunde berlin. (2015). *Webseite des Tierstimmenarchivs*. Recuperado de <http://www.tierstimmenarchiv.de/>

NPS. (2015). *National Park Service*. Recuperado de <http://www.nps.gov/index.htm>

PAMGUARD. (2015). Open source software for passive acoustic monitoring. Recuperado de <http://www.pamguard.org/>

SABIOD. (2015). ICML 2013, Workshop on Machine Learning for Bioacoustics. Recuperado de <http://sabiod.univ-tln.fr/icml2013>

Referencias electrónicas

The Cornell Lab of Ornithology/ML. (2015). *The Cornell Lab of Ornithology/Macaulay Library*. Recuperado de <http://macaulaylibrary.org/>

The Cornell Lab of Ornithology/Raven. (2015). RAVEN, Interactive sound analysis software. Recuperado de <http://www.birds.cornell.edu/brp/raven/RavenOverview.html>

UNAM/Instituto de Ciencias del Mar y Limnología. (2015). El instituto. Recuperado de <http://www.icmyl.unam.mx/?q=node/6>

WHOAS. (2015). *WHOAS: Woods Hole Open Access Server, a repository for the Woods Hole scientific community*. Recuperado de <https://darchive.mblwhoilibrary.org>

WHOI. (2015). *Woods Hole Oceanographic Institution*. Recuperado de <http://www.whoi.edu/>

XBAT. (2015). Extensible bioacoustic tool. Recuperado de <https://dl.dropboxusercontent.com/u/4142063/build/home.html>

Anexo A. Listado de mamíferos marinos en el Golfo de México

Anexo A. Listado de mamíferos marinos en el Golfo de México

Id	Orden	Suborden	Familia	Nombre científico
1	Cetacea	Mysticeti	Balaenidae	Eubalaena glacialis
2	Cetacea	Mysticeti	Balaenopteridae	Balaenoptera acutorostrata
3	Cetacea	Mysticeti	Balaenopteridae	Balaenoptera borealis
4	Cetacea	Mysticeti	Balaenopteridae	Balaenoptera edeni
5	Cetacea	Mysticeti	Balaenopteridae	Balaenoptera musculus
6	Cetacea	Mysticeti	Balaenopteridae	Balaenoptera physalus
7	Cetacea	Mysticeti	Balaenopteridae	Megaptera novaeangliae
8	Cetacea	Odontoceti	Delphinidae	Feresa attenuata
9	Cetacea	Odontoceti	Delphinidae	Globicephala macrorhynchus
10	Cetacea	Odontoceti	Delphinidae	Grampus griseus
11	Cetacea	Odontoceti	Delphinidae	Lagenodelphis hosei
12	Cetacea	Odontoceti	Delphinidae	Orcinus orca
13	Cetacea	Odontoceti	Delphinidae	Peponocephala electra
14	Cetacea	Odontoceti	Delphinidae	Pseudorca crassidens
15	Cetacea	Odontoceti	Delphinidae	Stenella attenuata
16	Cetacea	Odontoceti	Delphinidae	Stenella clymene
17	Cetacea	Odontoceti	Delphinidae	Stenella coeruleoalba
18	Cetacea	Odontoceti	Delphinidae	Stenella frontalis
19	Cetacea	Odontoceti	Delphinidae	Stenella longirostris
20	Cetacea	Odontoceti	Delphinidae	Steno bredanensis
21	Cetacea	Odontoceti	Delphinidae	Tursiops truncatus
22	Cetacea	Odontoceti	Kogiidae	Kogia breviceps
23	Cetacea	Odontoceti	Kogiidae	Kogia simus
24	Cetacea	Odontoceti	Physeteridae	Physeter macrocephalus
25	Cetacea	Odontoceti	Ziphiidae	Mesoplodon bidens
26	Cetacea	Odontoceti	Ziphiidae	Mesoplodon densirostris
27	Cetacea	Odontoceti	Ziphiidae	Mesoplodon europaeus
28	Cetacea	Odontoceti	Ziphiidae	Ziphius cavirostris
29	Sirenia	-	Trichechidae	Trichechus manatus

Anexo B. Peces incluidos en la Carta Nacional Pesquera

Anexo B. Peces incluidos en la Carta Nacional Pesquera

Id	Familia	Nombre común	Nombre científico
1	Ariidae	Bagre, bosh, curuco	Ariopsis felis
2	Ariidae	Bagre bandera	Bagre marinus
3	Carangidae	Cojinuda, cojinúa	Caranx crysos
4	Carangidae	Jurel amarillo, común, vaca	Caranx hippos
5	Carangidae	Jurel blanco, jurel ojón	Caranx latus
6	Carangidae	Jurel negro	Caranx lugubris
7	Centropomus	Chucumite	Centropomus parallelus
8	Centropomus	Robalo prieto	Centropomus poeyi
9	Centropomus	Robalo blanco	Centropomus undecimalis
10	Clupeidae	Sardina de escama fina, lacha	Brevoortia gunteri
11	Clupeidae	Sardina lacha	Brevoortia patronus
12	Clupeidae	Sardina carapachona	Harengula clupeola
13	Clupeidae	Sardina escamuda	Harengula jaguana
14	Clupeidae	Sardina vivita de hebra	Opisthonema oglinum
15	Dasyatidae	Raya látigo o balá	Dasyatis americana
16	Dasyatidae	Raya o balá	Dasyatis sabina
17	Dasyatidae	Raya mariposa	Gymnura micrura
18	Lutjanidae	Huachinango de seda	Etelis oculatus
19	Lutjanidae	Pargo criollo, pargo lunajero	Lutjanus analis
20	Lutjanidae	Pargo	Lutjanus apodus
21	Lutjanidae	Huachinango de aleta negra	Lutjanus buccanella
22	Lutjanidae	Huachinango de castilla	Lutjanus campechanus
23	Lutjanidae	Cubera	Lutjanus cyanopterus
24	Lutjanidae	Pargo muleta, parguete, gris	Lutjanus griseus
25	Lutjanidae	Pargo perro, caballera	Lutjanus jocu
26	Lutjanidae	Pargo rojo	Lutjanus purpureus
27	Lutjanidae	Rubia, villajaiba	Lutjanus synagris
28	Lutjanidae	Huachinango ojo amarillo	Lutjanus vivanus
29	Lutjanidae	Canané, rubio, rubia	Ocyurus chrysurus
30	Lutjanidae	Besugo	Rhomboplites aurorubens
31	Mugilidae	Lisa	Mugil cephalus
32	Mugilidae	Lebrancha	Mugil curema
33	Myliobatidae	Chucho, raya pintada	Aetobatus narinari
34	Myliobatidae	Manta cubanita o chucha	Rhinoptera bonasus
35	Sciaenidae	Trucha blanca	Cynoscion arenarius
36	Sciaenidae	Trucha	Cynoscion nothus
37	Sciaenidae	Croca	Leiostomus xanthurus
38	Scombridae	Peto, carito	Scomberomorus cavalla
39	Scombridae	Sierra	Scomberomorus
40	Scombridae	Sierra	Scomberomorus regalis

Anexo B. Peces incluidos en la Carta Nacional Pesquera

41	Scorpaenidae	Rascacio	Neomerinthe hemingwayi
42	Scorpaenidae	Cabrilla roja	Cephalopholis fulva
43	Serranidae	Cabrilla, payaso verde	Epinephelus adscensionis
44	Serranidae	Lenteja, abadejo	Epinephelus drummondhayi
45	Serranidae	Mero extraviado	Epinephelus flavolimbatus
46	Serranidae	Payaso rojo	Epinephelus guttatus
47	Serranidae	Cherna, mero guasa	Epinephelus itajara
48	Serranidae	Mero	Epinephelus morio
49	Serranidae	Mero negro, fiat	Epinephelus nigritus
50	Serranidae	Cherna pinta	Epinephelus niveatus
51	Serranidae	Mero del Caribe	Epinephelus striatus
52	Serranidae	Negrillo	Mycteroperca bonaci
53	Serranidae	Gallina*	Mycteroperca
54	Acanthuridae	Cirujano pardo	Acanthurus bahianus
55	Acanthuridae	Cirujano rayado	Acanthurus chirurgus
56	Acanthuridae	Cirujano azul	Acanthurus coeruleus
57	Carangidae	Pámpano de hebra	Alectis ciliaris
58	Carangidae	Cojinuda	Carangoides bartholomaei
59	Carangidae	Zapatero	Oligoplites saurus
60	Carangidae	Jorobado, papelillo, corcovado	Selene vomer
61	Carangidae	Jorobado, papelillo	Selene setapinnis
62	Carangidae	Esmedregal, medregal	Seriola dumeril
63	Carangidae	Coronado	Seriola zonata
64	Gerreidae	Mojarra trompetera	Gerres cinereus
65	Haemulidae	Ronco amarillo, canario	Conodon nobilis
66	Haemulidae	Ronco jeníguaro	Haemulon aurolineatum
67	Haemulidae	Ronco condensado	Haemulon flavolineatum
68	Haemulidae	Burro	Haemulon macrostomum
69	Haemulidae	Boquilla, chac-chí	Haemulon plumieri
70	Haemulidae	Ronco carité	Haemulon sciurus
71	Haemulidae	Armado	Orthopristis chrysoptera
72	Kyphosidae	Chopa negra	Kyphosus sectator
73	Labridae	Boquinete, pargo lobo, gallo	Lachnolaimus maximus
74	Monacanthidae	Pez puerco, cochino, bota, lija	Aluterus scriptus
75	Ostraciidae	Torito	Acanthostracion quadricornis
76	Pomacanthidae	Gallineta café	Pomacanthus arcuatus
77	Pomacanthidae	Gallineta negra	Pomacanthus paru
78	Sciaenidae	Ronco rayado	Bairdiella ronchus
79	Sciaenidae	Postá, pollito	Bairdiella chrysoura
80	Sciaenidae	Trucha pinta, corvina pinta	Cynoscion nebulosus
81	Sciaenidae	Berrugato	Menticirrhus saxatilis
82	Sciaenidae	Gurrubata, tambor	Micropogonias undulatus
83	Sciaenidae	Tambor negro	Pogonias cromis
84	Sciaenidae	Corvina ocelada	Sciaenops ocellatus

Anexo B. Peces incluidos en la Carta Nacional Pesquera

85	Serranidae	Cabrilla	Cephalopholis cruentata
86	Serranidae	Serrano arenero, bolo	Diplectrum formosum
87	Sparidae	Sargo, chopa, Posthá	Archosargus
88	Sparidae	Mojarrón	Calamus bajonado
89	Sparidae	Cachipluma	Calamus calamus
90	Sphyraenidae	Barracuda, picuda	Sphyraena barracuda
91	Stromateidae	Palometa pámpano	Peprilus paru
92	Tetraodontidae	Conejo blanco	Lagocephalus
93	Tetraodontidae	Botete globo	Sphoeroides
94	Triglidae	Rubio volador	Prionotus evolans

Anexo C. Señales acústicas usadas

A continuación se muestran las grabaciones utilizadas en este trabajo, todas pueden ser encontrados en: Macaulay Library/Cornell Lab of Ornithology. Se incluye el nombre de la persona que realizó la grabación y el número del catálogo en Macaulay Library.

Balaenoptera Acutorostrata – Ballena de minke común

- C. Levenson, ML 118107
- C. Levenson, ML 118108
- C. Levenson, ML 118114
- C. Levenson, ML 118117
- P. J. Perkins, ML 117714

Megaptera Novaeangliae – Ballena jorobada

- P. J. Perkins, ML 118163
- P. J. Perkins, ML 118179
- R. Nelson, ML 193432
- W. W. Steiner, ML 118147

Orcinus Orca – Orca

- D. K. Ljungblad, ML 128216
- D. K. Ljungblad, ML 128217
- D. K. Ljungblad, ML 128218
- D. K. Ljungblad, ML 128219
- D. K. Ljungblad, ML 128223

Physeter Catodon – Cachalote

- D. K. Ljungblad, ML 128271
- D. K. Ljungblad, ML 128272
- P. J. Perkins, ML 110897
- P. J. Perkins, ML 112936
- P. J. Perkins, ML 117702

Anexo C. Señales acústicas usadas

Pseudorca Crassidens – Falsa orca

- J. Barlow, ML 120530
- K. S. Norris, ML 114720
- P. J. Perkins, ML 112994
- P. J. Perkins, ML 112995
- W. V. Ward, ML 125030

Stenella Frontalis - Delfín manchado del Atlántico

- P. J. Perkins, ML 110863
- P. J. Perkins, ML 111976
- P. J. Perkins, ML 111978
- P. J. Perkins, ML 112329
- W. W. Steiner, ML 122502

Tursiops Truncatus – Delfín nariz de botella

- J. Barlow, ML 120779
- J. Barlow, ML 120781
- R. C. Stein, ML 125036
- R. C. Stein, ML 125037
- W. V. Ward, ML 125038

Trichechus Manatus – Manatí del Caribe

- T. J. O'Shea, ML 118181
- T. J. O'Shea, ML 118182
- T. J. O'Shea, ML 118183
- T. J. O'Shea, ML 118184
- T. J. O'Shea, ML 118186

Bairdiella Chrysoura – Ronco amarillo

- M. P. Fish, ML 116455
- D. M. Baltz, ML 124522
- D. M. Baltz, ML 124523

Cynoscion Nebulosus – Trucha pinta

- D. M. Baltz, ML 124555

Anexo C. Señales acústicas usadas

Pogonias Cromis – Tambor negro

- M. P. Fish, ML 111910
- M. P. Fish, ML 111911
- M. P. Fish, ML 111912
- M. P. Fish, ML 111913
- M. P. Fish, ML 111914
- D. M. Baltz, ML 126246
- D. M. Baltz, ML 126247

Anexo D. Paralelización de la etapa de clasificación

Considerando el uso de un lenguaje de programación visual gráfico y la arquitectura planteada, la etapa de clasificación puede paralelizarse de modo que cada clasificador base sea implementado en un núcleo diferente (ver Fig. D-0-1), obviamente considerando un equipo con al menos 4 núcleos.

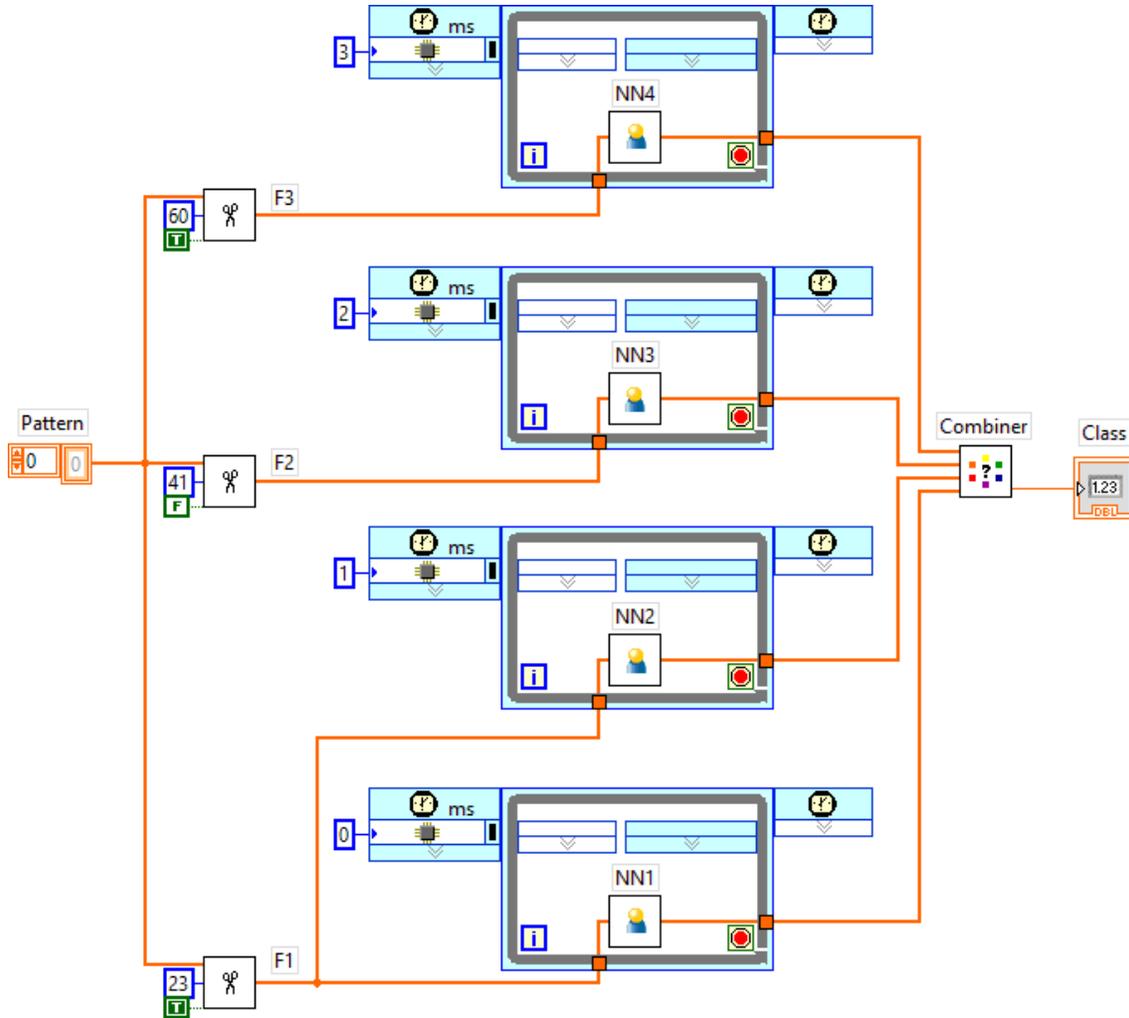


Fig. D-0-1. Diagrama de paralelización de la etapa de clasificación.

Anexo E. Sistema de medición

El laboratorio cuenta con hidrófonos omnidireccionales con un rango usable de 1 a 170 kHz (con acoplamiento a cable BNC), preamplificadores de voltaje (resistentes al agua), módulos de batería para los preamplificadores (resistentes al agua) y tarjetas de adquisición de datos (con múltiples entradas) para un muestreo máximo de 51.2 kS/s.



Fig. E-0-1. Arquitectura del sistema de medición.