RIC REVISTA DE INICIACIÓN CIENTÍFICA
Portal de revistas: http://revistas.utp.ac.pa

Clasificador automático de clase (adulto-cría) mediante características distintivas en vocalizaciones de manatíes

Automatic classifier (adult-calf) using distinctive characteristics in manatee vocalizations

Jaira M. Morales Magallón[1,*], *Hazel Pacheco*[1], *Fernando Merchan Spiegel*[1]

[1] Facultad de Ingeniería Eléctrica, Universidad Tecnológica de Panamá

***Autor de correspondencia:** jaira.morales@utp.ac.pa

Resumen. En este documento se desarrollan algunos métodos específicos para llevar a cabo un clasificador que sea capaz de identificar manatíes adultos y cría de distintos audios. La duración de las vocalizaciones se evaluó con un método de eliminación de ruido basado en eliminación de silencios, para luego aprovechar esta señal y evaluar uno de los principales parámetros, la frecuencia fundamental la cual se calculó con dos métodos (cepstrum y THD) y de esta manera se compararon los valores obtenidos y se evaluaron en un clasificador de mínima distancia.

Palabras clave. Cepstrum, frecuencia fundamental, manatíes, THD, tiempo.

Abstract. In this document, some specific methods are developed to carry out a classifier that is capable of identifying adult and calf manatees from different audios. The duration of the vocalizations will be evaluated with a noise elimination method based on the elimination of silences, then take advantage of this signal and evaluate one of the main parameters, the fundamental frequency which will be calculated in two methods (cepstrum and THD); In this way, compare the obtained values and evaluate them in a minimum distance classifier.

Keywords. Cepstrum, fundamental frequency, manatees, THD, time.

1. Introducción

El manatí antillano es una especie que está amenazada en su entorno de repartición. Para fomentar su conservación es fundamental ubicar y contar individuos [1]. La estimación de la población por medio de sus vocalizaciones es una alternativa confiable, de bajo precio, no invasiva y nueva en la zona centroamericana. En este contexto de estudio de población una información de gran interés es determinar el rango de edad y sexo de los individuos de la población. El procesamiento de los audios en esta metodología acústica requiere normalmente para luego poder detectar y extraer la señal de interés. Para la clasificación de edad y género se requerirá estudiar características de las señales.

Según investigaciones realizadas, se suele trabajar con 6 variables acústicas y se han encontrado que los individuos varían significativamente en frecuencia fundamental, banda enfatizada, rango de frecuencia y contorno de llamada (el patrón general de complejidad en la modulación de frecuencia), pero uno de los más relevantes es la frecuencia fundamental [2].

En [3], se estudiaron los rangos de valores de algunos de estos parámetros de las señales que propios de machos adultos, hembras adultas, machos y hembras jóvenes. El estudio se concentró en parámetros tales como la frecuencia fundamental, la duración y el número de armónicos de la vocalización.

Existen varios métodos para estimar la frecuencia fundamental, entre ellos se pueden mencionar: la autocorrelación, la correlación cruzada y el método del cepstrum [4].

En repetidos estudios se ha demostrado que el método cepstrum tiene la capacidad de realizar una estimación más precisa de la frecuencia fundamental, en la señal de voz

https://doi.org/10.33412/rev-ric.v8.1.3507 Vol. 8 - N.° 1 · enero - junio · 2022

enmascarada por ruido blanco, que los otros métodos utilizados para estimar este criterio como lo es la Distorsión Armónica Total (THD) [4].

Los clasificadores son esenciales ya que con ellos se determinan finalmente si los criterios utilizados son efectivos para el desarrollo, en la medida que se obtengan las firmas de referencia.

Existen varios criterios para clasificar, uno de ellos es el clasificador de mínima distancia (vecino más cercano) es el criterio más sencillo, y consiste en asignar cada muestra a la clase más cercana. Para ello se mide la distancia euclidiana entre el vector de características y de esta manera se obtiene la matriz de confusión que es aquella que ayuda finalmente a clasificar los casos correctos o no.

Para esta investigación se busca clasificar en rangos de edad la especie a través de la determinación de frecuencia fundamental, utilizando el método de cepstrum y distorsión armónica total (THD) para luego someter este parámetro a un clasificador automático que determine si la vocalización se trata de una cría o un adulto.

2. Materiales y métodos/metodología

Se tiene por objetivo clasificar de manera automática un grupo específico, Adultos y crías, utilizando métodos como cepstrum complejo y distorsión armónica total (THD), para identificación de la frecuencia fundamental. La frecuencia fundamental es un criterio importante porque indica la onda sonora más simple de frecuencias más baja

. La mayoría de la data utilizada para realizar este estudio, son audios de conversaciones entre madres y crías en donde se pudo observar que las crías presentan frecuencias más altas con mayor número de armónicos que las madres y menor duración en su señal de voz.

El esquema del sistema de clasificación de señales consta de una etapa de preprocesamiento o detección de vocalización es previo al análisis del parámetro de interés que es la frecuencia fundamental.

2.1 Detección de Vocalizaciones en las grabaciones

En esta etapa de detección de señales se busca ubicar en que secciones de las grabaciones se encuentran las vocalizaciones de manatíes. Esto se realiza en dos pasos: reducción de ruido y detección de señal basado en energía y con criterio de duración.

2.1.1 Reducción de ruido

Los audios provenientes de hidrófonos normalmente se reciben con ruido de fondo y sonidos de otras especies, lo cual complica calcular la ubicación exacta de una vocalización, para resolver esta problemática es necesario aplicar filtros digitales.

Un filtro es una herramienta útil y ayuda a extraer una señal más clara. En este caso un filtro pasa altas y pasa bajas tipo Butterwoth de orden 6 se aplica a la señal para enfatizar las frecuencias altas de los formantes y de esta manera las frecuencias menores a 1.5kHz y mayores a 16kHz se descarten.

Aquellos sonidos que se podrían confundir con vocalizaciones de la especie en estudio no son los únicos presentes, también se toman en cuenta las perturbaciones ambientales que necesita métodos más complejos. Para ello se aplica el método de substracción espectral que es un método de supresión de ruido ampliamente utilizado en el contexto de voz humana [5]. Este método considera el espectro típico del ruido que afecta la señal que es proporcionado por el usuario y lo utiliza para "substraerlo" de la señal observada en el dominio espectral. En esta aplicación este método llega a eliminar gran parte del ruido ambiental.

2.1.2 Detección de señal basado en energía y criterio de duración

Las señales de manatíes tienen una duración entre 70ms y 800 ms como máximo, aunque algunos estudios proponen que han obtenido un máximo de 900ms expuesto en [1, 2].

Para poder obtener la duración de la señal libre de ruido se requiere ahora eliminar los silencios y esto se logra escaneando la señal filtrada para que las zonas de silencio sean removidas por medio del cálculo de energía en cortos periodos de tiempo, matemáticamente se puede calcular como en la ecuación (1) y la señal resultante se muestra en la figura 1.

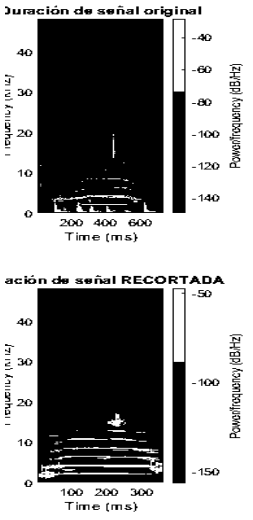

Figura 1. Antes y después de Filtrar y eliminar silencios de vocalizaciones. Segmentos de 10ms se escogieron para este propósito.

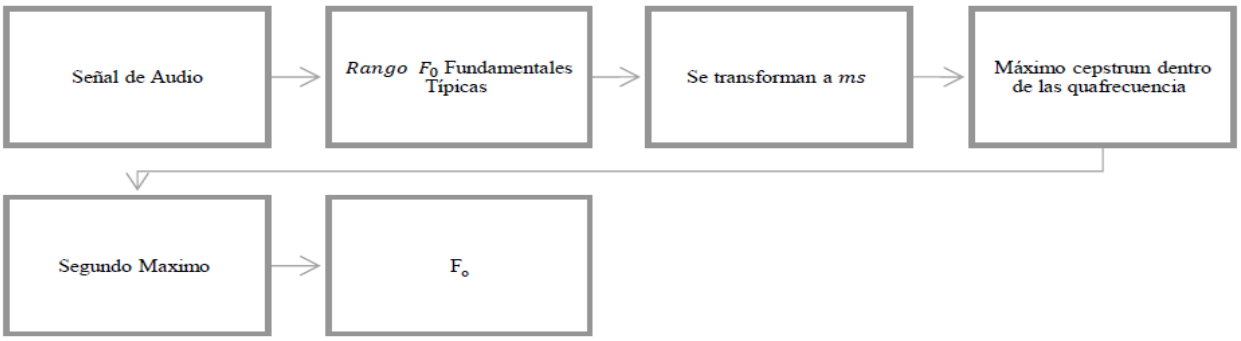

Figura 2. Diagrama de pasos estipulados para obtener la Frecuencia Fundamental por método cepstrum.

Si en un segmento la energía promedio es menor que un valor umbral proporcional a la energía promedio de la señal entera presentada en la ecuación (2), entonces será descartado.

$$E_n = \sum_{k=1}^{w_n} |x[k]|^2 w[n-k] \quad (1)$$

$$E_{avg} = \frac{1}{N} \sum_{k=1}^{N} |x[k]|^2 \quad (2)$$

Donde $x[n]$ es la señal bajo estudio, $w[n]$ corresponde a la ventana de análisis, W_n es la duración de la ventana, N es el número en la señal bajo estudio.

2.2 Frecuencia Fundamental y Número de Armónicos

Para esta aplicación se utilizaron dos técnicas para el cálculo de la frecuencia fundamental, el método CEPSTRUM y la función THD o distorsión armónica total.

2.2.1 Cepstrum

El análisis de cepstrum es una técnica de procesamiento de señal, que se utilizará para poder calcular la frecuencia fundamental. El cepstrum de una señal se define como el resultado de calcular la transformada de Fourier inversa del espectro de la señal estudiada en escala logarítmica (dB).

El cepstrum de una señal es la transformada de Fourier del logaritmo (con fase instantánea o no envolvente) del espectro de la señal estudiada. A veces es llamado el espectro del espectro.

$$Cc = F^{-1}\{Log(F\{f(t)\})\}V \quad (3)$$

Algorítmicamente se pueden calcular los coeficientes Cesptrales utilizando los pasos mostrados en la figura 2.

El estudio de cepstrum nos indica que para que este algoritmo funcione y evalué de forma correcta los coeficientes, la señal debe estar procesada y se deben conocer los rangos de frecuencia en donde se encuentra la fundamental, estas frecuencias típicas de los manatíes corresponden a (2.5 a 4kHz). Una vez conocidos estos parámetros, se procede a calcular los coeficientes. Los rangos de frecuencia fundamental deben transformarse a milisegundos para que el cepstrum pueda identificar la posición.

Luego se encuentra el máximo cepstrum dentro de la quafrecuencia, se marca el segundo pico en donde ocurrió esto y finalmente este valor pertenecerá a la frecuencia fundamental.

2.2.2 Distorsión Armónica Total (THD)

Es la relación entre el contenido armónico de la señal y la primera armónica o fundamental.

La distorsión armónica es un parámetro técnico utilizado para definir la señal de audio que sale de un sistema. La distorsión armónica se produce cuando la señal en estudio posee más de un componente de frecuencia. Puesto que son armónicos, es decir múltiplos de la señal fundamental esta distorsión no es tan disonante y es más difícil de detectar. La ecuación (4) define la forma en que trabaja este método. Donde $P0$ es la potencia de la frecuencia fundamental y Pi con i > 0 son las potencias de todos los demás armónicos que contiene la señal. De esta manera se observa en la figura 4 el periodo grama que muestra en donde están ubicados tanto la frecuencia fundamental como los armónicos múltiplos de esta.

$$THD = \Sigma \frac{Potencia\ de\ los\ armónicos}{Potencia\ de\ la\ frecuencia\ fundamental} \quad (4)$$

2.3 Clasificador

Se implementó un clasificador de tipo supervisado, en la cual el usuario debe conocer a priori las clasificaciones reales de las vocalizaciones. Es necesario contar con conjuntos de entrenamiento para todas las clases predefinidas, que se ha establecido como madre y cría. Se requieren muestras con un mínimo de vocalizaciones para cada clase. Una vez que se tengan una variable representativa de la característica esencial de las vocalizaciones, en este caso las frecuencias, se procede a la clasificación de las vocalizaciones haciendo uso de un clasificador de vecino más cercano.

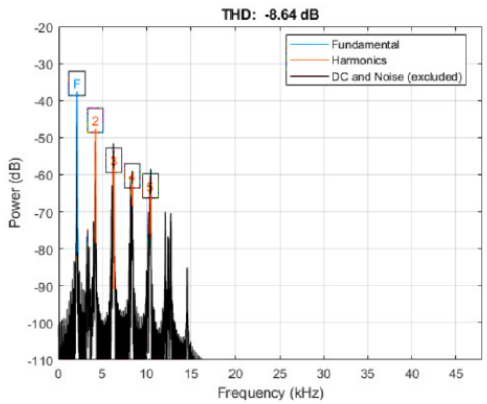

Figura 3. Periodograma obtenido con la función THD de una vocalización de manatí.

El clasificador de mínima distancia (vecino más cercano) es el criterio más sencillo, y consiste en asignar cada vocalización a la clase más cercana. Para ello, se mide la distancia euclidiana entre el vector de características de las vocalizaciones (frecuencias) en cuestión y la media del agrupamiento (frecuencia promedio) de las vocalizaciones para cada clase.

2.3.1 Preclasificación

En la etapa inicial previa al clasificador, se analizaron diferentes audios con duraciones aproximadas de 8 a 20 minutos, resultados de recopilaciones realizadas por hidrófonos en diferentes puntos críticos, donde se ha tenido la presencia confirmada de manatís. Para calcular la diferencia entre la frecuencia promedio obtenida para adultos y para crías, y la frecuencia que estudiada. Se calcula la distancia euclidiana definida en la ecuación 5.

$$d_E = \sqrt{(x2 - x1)^2 + (y2 - y1)^2} \quad (5)$$

2.3.2 Post clasificación

Una vez obtenidas las distancias euclidianas, se busca la distancia mínima entre la frecuencia que se analiza contenida en el vector y la frecuencia promedio, de esta manera se asigna a la clase que posea la distancia más pequeña. El proceso de asignar las clases a un nuevo vector es conocido como el mapa de clasificación, el cual contiene la información resultante del algoritmo. Junto con el mapa de clasificación y la máscara se crea a la matriz de confusión similar a la expuesta en la figura 5. La matriz de confusión contiene varias métricas que determinan el rendimiento del clasificador y cada una se calcula de manera distinta.

• **Exactitud:** calcula la cercanía que está el resultado de una medición del valor verdadero y se puede obtener a través de la ecuación 6.

$$Ex = \frac{TP+TN}{Total\ de\ Muestras} \quad (6)$$

• **Precisión:** Se refiere a lo cerca que está el resultado de una predicción del valor verdadero, se calcula al aplicar la ecuación 7.

$$P = \frac{TP}{TP+FP} \quad (7)$$

• **Sensibilidad o recall:** Es la proporción de casos positivos que fueron correctamente identificadas por el algoritmo, se encuentra al usar la ecuación 8.

$$S = \frac{TP}{TP+FN} \quad (8)$$

• **Especificidad:** se enfoca en los verdaderos negativos (ecuación 9).

$$Es = \frac{TN}{TN+FP} \quad (9)$$

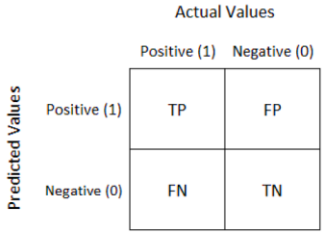

Figura 4. Modelo estándar de una matriz de confusión

3. Resultados y discusión

Para la prueba final del clasificador adulto– cría se creó un nuevo set de señales. Los mismos fueron clasificados en tres clases: 1/ adultos (madres), 2/ crías y 3/otros. La tercera categoría corresponde a aquellas señales cuyos valores estimados se salen de rangos normales para las vocalizaciones de manatíes, debido a algún error de estimación. En esta sección se utilizaron 49 señales o vocalizaciones distintas a las incluidas en la sección de prueba.

3.1 Distorsión Armónica Total (THD)

A continuación, en la figura 6 y la tabla 1 se presentan los resultados obtenidos utilizando la frecuencia fundamental calculada con la función THD.

La matriz de confusión observada en la figura 6 indica que las madres son mejor clasificadas que las crías; aun así, hubo 9 madres clasificadas como crías y 7 crías clasificadas como madres. Además, 5 vocalizaciones fueron clasificadas como 3 lo cual indica que no se cumplía con el parámetro del tiempo adecuado o se clasificó como un audio corrupto desde el principio.

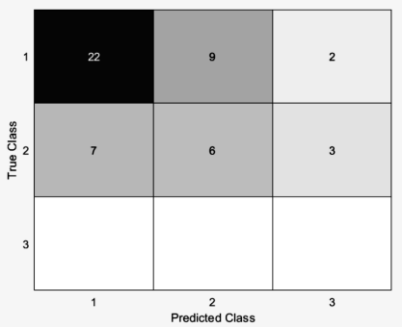

Figura 5. Matriz de confusión caso THD con data de prueba.

Tabla 1. Métricas obtenidas con el método de THD y data de prueba

#	Verdaderos positivos	22 (Actual positive =33)
1	Falsos positivos	9
2	Falsos negativos	7
3	Precisión	0.647059
4	Sensibilidad	0.758621
5	Exactitud	0.571429
6	Especificad	0.333333
7	F1	0.698413

Estos resultados demostraron que la calidad de los audios obtenidos en términos de energía de la señal y ruido residual afectaba el proceso de estimación de los parámetros. De la base de pruebas se seleccionó un conjunto de señales con mejor calidad de 16 muestras. Los resultados de este conjunto se presentan en la figura 7 y la tabla 2.

Como se muestra en la figura 7, de siete vocalizaciones de madres, seis se han clasificado de forma correcta. La única vocalización en la sección de clase de madre que no se clasificó de forma correcta, quedó en la clase tres, que está destinada para todos aquellos audios que no cumplen con los requisitos del rango del tiempo que se ha establecido como estándar.

Estos resultados corresponden a una serie de data seleccionada de buena calidad para obtener resultados precisos.

El producto obtenido en este caso es esperado ya que el clasificador rechaza los audios que están corruptos o no tienen el formato adecuado para su análisis, colocando cero automáticamente en los campos de duración y frecuencia.

Por otro lado, de nueve vocalizaciones reales de crías, ocho se clasificaron de forma correcta y solo una quedó en la clase tres, donde se repite el caso anterior. En el cálculo de métricas de la matriz de confusión el factor F1 indica que tan eficiente es el clasificador, por tanto, para la función de THD resulta bastante eficiente.

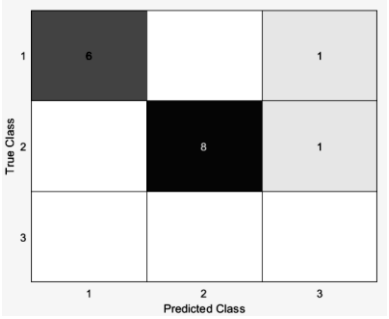

Figura 6. Resultados del método THD usando audios seleccionados de mejor calidad.

Tabla 2. Métricas obtenidas con el método de THD con audios seleccionados de mejor calidad

#	Verdaderos positivos	6 (actual positive =7)
1	Falsos positivos	1
2	Falsos negativos	0
3	Precisión	0.857143
4	Sensibilidad	1.0000
5	Exactitud	0.875000
6	Especificad	0.888889
7	F1	0.923077

3.1.1 Cepstrum

En esta matriz de confusión se presenta un comportamiento que no se había presentado antes y es el traslape de las clases de madre y crías entre sí. De treinta y tres vocalizaciones reales de madres, veinte se clasificaron como madres, y once se clasificaron como cría, por otro lado, de dieciséis crías reales se tienen que seis han sido con madres, siete se han clasificado de forma correcta y tres se ha clasificado como otro. El caso de las clasificaciones en la clase 3 son igualmente esperados como en los resultados anteriores. En un análisis previo se puede observar que el rendimiento es inferior al rendimiento presentado con frecuencias THD.

En base a los resultados obtenidos tanto para el método de cepstrum observados en la tabla 2, como la distorsión armónica total en la tabla 1, se puede deducir que el clasificador planteado no es perfecto, puede mejorarse, ya sea incluyendo más parámetros de clasificación o mejorando el proceso de estimación.

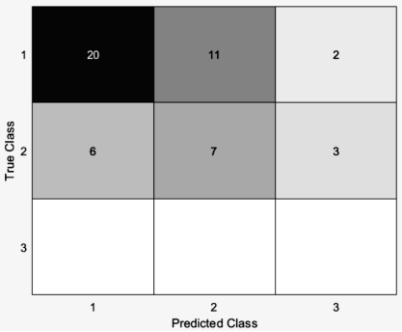

Figura 7. Matriz de confusión caso Cepstrum con data de prueba.

Tabla 3. Métricas obtenidas con el método de cepstrum usando audios de prueba

#	**Verdaderos positivos**	**20 (actual positive =33)**
1	Falsos positivos	11
2	Falsos negativos	6
3	Precisión	0.588235
4	Sensibilidad	0.769231
5	Exactitud	0.551020
6	Especificad	0.333333
7	F1	0.666667

4. Conclusiones

Uno de los principales parámetros distintivos de los individuos del manatí antillano es la frecuencia fundamental de sus vocalizaciones.

La diferencia en el rendimiento del clasificador con la data de entrenamiento y la data de prueba se considera que es debido a que algunos de los audios extraídos no cuentan con una buena calidad de sonido debido a que en algunos casos la potencia de esta señal era muy tenue.

Los resultados muestran que una buena cantidad de vocalizaciones se clasifican de forma correcta, lo que nos indica que el clasificador tiene una buena estructura inicial, pero necesita ser más robusto y riguroso con las condiciones de clasificación. Esto podría mejorarse con la inclusión de la duración promedio de vocalización por cada clase.

De los dos modelos evaluados los mejores resultados de clasificación es el que utiliza la función de THD (distorsión armónica total), lo cual es muy curioso ya que, en comparaciones realizadas, las frecuencias calculadas por el método de cepstrum son mucho más coherentes que las calculadas con THD. Sin embargo, se cree que dicha anomalía se debe a que dentro de los audios examinados de adultos y crías reales existen frecuencias inciertas y la hora de obtener el promedio en el clasificador existe un sesgo marcado.

AGRADECIMIENTOS

Expresamos nuestro agradecimiento al Dr. Héctor Guzmán del Instituto Smithsonian de Investigaciones Tropicales en Panamá por facilitación de audios de pruebas. Cabe destacar que para la adquisición de estos audios se emplearon métodos aprobados o en acuerdo por lo establecido por el Comité de Cuidado y Uso de Animales del Instituto Smithsonian de Investigaciones Tropicales (IACUC), el cual cumple las regulaciones correspondientes a nivel internacional, y con todos los requerimientos de bioética.

CONFLICTO DE INTERESES

Los autores declaran no tener algún conflicto de interés.

REFERENCIAS

[1] F. Merchan, G. Echevers, H. Poveda, J. Sanchez-Galan, H.M. Guzman, “Detection and identification of manatee individual vocalizations in Panamanian wetlands using spectrogram clustering”, J. of the Acoustical Soc. of America, 146, 2019.

[2] T. J. O’Shea and L. B. Poché Jr., “Aspects of underwater sound communication in Florida manatees (Trichechus manatus latirostris),” J. Mammal., vol. 87, no. 6, pp. 1061–1071, 2006.

[3] Sousa-Lima, Renata & Paglia, Adriano & Fonseca, Gustavo. (2008). Gender, Age, and Identity in the Isolation Calls of Antillean Manatees (Trichechus manatus manatus). Aquatic Mammals. 34. 109-122. 10.1578/AM.34.1.2008.109.

[4] J. Galić and T. Pešić-Brđanin, "The voice fundamental frequency statistical parameters under noisy conditions with the cepstrum method," 2011 10th International Conference on Telecommunication in Modern Satellite Cable and Broadcasting Services (TELSIKS), 2011, pp. 769-772, doi: 10.1109/TELSKS.2011.6143224.

[5] S. Boll, "A spectral subtraction algorithm for suppression of acoustic noise in speech," ICASSP '79. IEEE International Conference on Acoustics, Speech, and Signal Processing, 1979, pp. 200-203, doi: 10.1109/ICASSP.1979.1170696.