

Cancelación del Ruido de Ambiente en Grabaciones de Vocalizaciones de Manatíes Antillanos

Jorge Castro

Centro de Investigaciones en
Tecnologías de la Información
y Comunicación
Universidad de Costa Rica
jorge.castroc@ucr.ac.cr
(506) 8862-3108

Arturo Camacho

Escuela de Ciencias de la
Computación e Informática
Universidad de Costa Rica
arturo.camacho@ecci.ucr.ac.cr

Mario Rivera

Centro de Investigaciones en
Tecnologías de la Información
y Comunicación
Universidad de Costa Rica
mariorivera@gmail.com

ABSTRACT

El manatí antillano es una especie que se encuentra amenazada a lo largo de su ámbito de distribución. Para promover su conservación, es importante localizar y contar individuos. La estimación de población a través de sus vocalizaciones es una opción confiable, de bajo costo, no invasiva y novedosa en la región centroamericana. Sin embargo, para desarrollar un método de conteo automático, conviene primero cancelar el ruido de ambiente en las grabaciones. En este trabajo se propone un algoritmo para cancelar el ruido ambiental en grabaciones de campo de vocalizaciones de manatíes, basado en la transformada *wavelet*, la función de autocorrelación y técnicas de agrupamiento. El algoritmo propuesto logra una mejora promedio de la razón de señal a ruido de 28 dB para grabaciones de emisiones vocales de manatíes antillanos contaminadas con ruido ambiental de alta intensidad.

Author Keywords

Manatí; cancelación; ruido; vocalizaciones; wavelets;

INTRODUCCIÓN

El manatí antillano (*Trichechus manatus manatus*) constituye una parte representativa de la biodiversidad de los ecosistemas en los humedales centroamericanos. Desafortunadamente, pertenece a la lista de especies amenazadas de la IUCN desde hace varias décadas [2]. El problema tratado en este artículo es el de la cancelación de ruido en grabaciones de campo de vocalizaciones de manatíes. La eliminación del ruido de ambiente es necesaria para la detección automática de sus emisiones vocales y el conteo automático de individuos con base en estas. El conteo de individuos permite realizar estimaciones de población que contribuyen con la gestión y protección del manatí.

En investigaciones recientes se han propuesto diversos métodos para la cancelación de ruido de ambiente en vocalizaciones de manatíes. Se ha investigado el uso de filtros adaptativos

lineales [9, 10], métodos no lineales basados en la transformada *wavelet* [4, 7, 6] y técnicas de separación de fuentes que utilizan múltiples receptores [5]. Los resultados de los métodos basados en la transformada *wavelet* han sido superiores a los obtenidos mediante filtros adaptativos lineales.

El algoritmo de cancelación de ruido propuesto está basado en la transformada *wavelet* no decimada (UDWT, por sus siglas en inglés) y la función de autocorrelación [4]. En adición, utiliza un algoritmo de agrupamiento para seleccionar el umbral a aplicar en el dominio *wavelet* [1].

CARACTERIZACIÓN DEL RUIDO Y DE LAS VOCALIZACIONES

El ruido presente en las aguas poco profundas en que habita el manatí no es estacionario y se compone de diversas fuentes, entre ellas, el producido por el camarón de la familia *alpheidae*, el movimiento del agua, los motores de los botes y las emisiones de otras especies [9]. La cantidad de camarones aumenta conforme a la cercanía al mar y por lo tanto, también el ruido producido por estos [8]. El ruido producido por este camarón se caracteriza por ser de banda ancha (llega hasta 100 kHz), tener una duración menor a 10 ms y alcanzar niveles superiores a los 90 dB (re 1 μ Pa) [9]. Otros factores que contribuyen al ruido son el movimiento del agua, los motores de los botes y las emisiones de otras especies.

Las vocalizaciones del manatí usualmente evocan una fuerte sensación de altura musical (son vocalizaciones armónicas). Estas emisiones vocales tienen una frecuencia fundamental entre 2 y 5.5 kHz, una duración entre 200 y 500 ms y un contenido armónico extenso, que suele alcanzar frecuencias ultrasónicas. A diferencia del ruido de ambiente, las vocalizaciones armónicas presentan poca energía por debajo de los 2 kHz y tienen una función de autocorrelación que decae lentamente [4]. Ambas características son explotadas por el algoritmo de cancelación de ruido propuesto.

ALGORITMO DE CANCELACIÓN DE RUIDO

Las estrategias de cancelación de ruido con *wavelets* parten de la elección de una *wavelet* base apropiada, de manera que la energía de la señal objetivo queda concentrada en un número pequeño de coeficientes de gran amplitud, y que el ruido de ambiente queda distribuido en una gran cantidad de coeficientes de pequeña amplitud. De esta forma, la aplicación de

Se autoriza hacer copias de este trabajo o partes de él para uso académico, siempre y cuando sea sin fines de lucro. La copia debe contener esta nota y la cita completa en la primera página. La copia bajo otras circunstancias y la distribución o publicación podría requerir permisos adicionales o pago de derechos.

JoCICI 2015, Marzo 5-6, San José, Costa Rica.

Copyright 2015 CITIC-PCI

la transformada *wavelet* para el realce de una señal, es un proceso de tres etapas: descomposición en *wavelets*, umbralización de los coeficientes *wavelet* y reconstrucción de la señal.

La señal acústica del ruido submarino se suele modelar como ruido aditivo de la siguiente forma:

$$s(t) = x(t) + r(t) \quad (1)$$

donde $x(t)$ es la señal de la vocalización del manatí, $r(t)$ es la señal del ruido y $s(t)$ es la señal de la vocalización contaminada con ruido.

El algoritmo de cancelación de ruido propuesto se muestra en la figura 1. Primero, se le aplica a la señal contaminada $s(t)$ un filtro paso alto con frecuencia de corte $f_c = 2$ kHz para eliminar el ruido de ambiente que se sale del rango de frecuencias de las vocalizaciones de manatí. Luego, se divide la señal en I ventanas de 3 ms cada una y se aplica la UDWT con la *wavelet* base *Daubechies* 8 y una descomposición en $J = 4$ niveles. Este tamaño de ventana es suficientemente pequeño como para considerar el fragmento de la señal como estacionario en sentido amplio (WSS, por sus siglas en inglés), lo que aumenta la velocidad de cómputo. Además, es suficientemente grande como para cubrir al menos seis períodos de la frecuencia más baja de interés (2 kHz). Se seleccionó la UDWT por su simplicidad con respecto a otras transformadas *wavelet* y porque elimina el problema de variación ante la traslación que posee la transformada discreta de *wavelet*. Además, la familia *Daubechies* ha sido empleada exitosamente en la cancelación del ruido de vocalizaciones de manatíes [4, 6]. La descomposición en cuatro niveles, parte la señal en bandas relevantes según la distribución en frecuencia de las emisiones vocales del manatí. De esta forma, para cada ventana i se tiene una señal $w_i(j, k)$ de coeficientes *wavelet* de nivel j y traslación k . Luego, se aplica la función de autocorrelación a los coeficientes *wavelets* obtenidos en cada nivel y se calcula la raíz cuadrada de la media de los cuadrados (valor RMS) de la función de autocorrelación. Cabe resaltar que en el dominio *wavelet*, la función de autocorrelación de las vocalizaciones de manatí y del ruido de ambiente, conservan sus diferencias [4].

La función de autocorrelación para señales WSS se calcula mediante la fórmula

$$r_{ss}(\tau) = \frac{1}{N} \sum_{k=\tau}^N (s(k) - \bar{s})(s(k - \tau) - \bar{s}) \quad (2)$$

donde τ es el desfase, N el tamaño de la señal y \bar{s} su valor medio. Para discernir entre una función de autocorrelación con decaimiento lento (como la de una vocalización de manatí) y otra con decaimiento rápido (como la del ruido), se calcula el valor RMS sobre la función de autocorrelación, desde el desfase $\tau = 20$ hasta el desfase $\tau = 120$. Posteriormente, se concatenan en una matriz $R_{J \times I}$ los valores RMS obtenidos para las I ventanas procesadas y a cada fila de la matriz se le aplica una media móvil de 22 puntos para eliminar los transientes producidos por el camarón *alpheidae*, que pueden alcanzar valores RMS altos por instantes cortos de tiempo. A la matriz resultante se le denomina $G_{J \times I}$.

Para determinar la posible presencia de una vocalización de manatí, se inspecciona la matriz $G_{J \times I}$. A diferencia del método seguido por Gur y Niezrecki [4], en el cual se determina un umbral fijo para la evaluación, en nuestro caso se utiliza un algoritmo de agrupamiento [1] que se adapta mejor a los cambios de nivel de las vocalizaciones y el ruido. Este algoritmo toma una señal unidimensional y le asigna a cada uno de sus puntos una de dos posibles clases, de acuerdo a un criterio de maximización de separación de los centroides de las clases, evaluado en ventanas de múltiples tamaños. Estas clases se representan mediante los valores de cero y uno, e indican la ausencia o posible presencia de un manatí, respectivamente. Luego de aplicar el algoritmo de agrupamiento, se realiza una verificación de la longitud de los segmentos de ceros y unos resultantes, para que no sean inferiores a la longitud mínima esperada de un canto. El resultado final se almacena en una matriz $H_{J \times I}$, que se usa para seleccionar el umbral a aplicar a los coeficientes *wavelet* para cancelar el ruido de ambiente, como se explica a continuación. Los coeficientes $w_i(j, k)$ obtenidos anteriormente para cada ventana son encogidos según un umbral adaptativo y una regla de umbralización suave. Este umbral varía para cada ventana i y nivel j de los coeficientes dependiendo del valor en $H(j, i)$. El umbral se define de la siguiente forma [4]:

$$\eta_{i,j} = \begin{cases} \max_k \{w_i(j, k)\} & \text{si } H(j, i) = 0 \\ \sigma(2 \log(N))^{1/2} & \text{si } H(j, i) = 1 \end{cases} \quad (3)$$

donde N es el número de coeficientes, σ es el nivel del ruido estimado a través de la desviación absoluta de la mediana (MAD) de los coeficientes de alta frecuencia del primer nivel [4], y $\sigma(2 \log(N))^{1/2}$ es el umbral universal propuesto por Donoho y Johnstone [3]. De esta forma, si $H(j, i) = 0$, el umbral es fijado al valor máximo de los coeficientes evaluados, y si $H(j, i) = 1$, se usa el umbral universal. La regla de umbralización suave reduce a cero todos los coeficientes con valor absoluto menor o igual al umbral, y el resto de coeficientes son *encogidos* hacia cero por una cantidad equivalente al umbral. Este último paso evita discontinuidades en la señal (producto de la umbralización dura). Finalmente, se aplica la transformada no decimada *wavelet* inversa (IUDWT) y se obtiene la señal con el ruido cancelado $\hat{x}(t)$.

SIMULACIÓN

Para evaluar el método propuesto se usaron grabaciones de campo de vocalizaciones de manatíes antillanos tomadas en el río San San, Panamá. De la colección con que se contaba, se seleccionaron las vocalizaciones con menor ruido ambiental (21 de ellas) y se concatenaron en un solo archivo, dejando al menos un segundo entre cada vocalización. Para contaminar la grabación, se seleccionó una muestra de ruido de alta intensidad (proveniente del mismo ambiente) y se sumó a la grabación original. Este ruido está compuesto principalmente por el movimiento del agua y el producido por camarones *alpheidae*. Las grabaciones del ruido de ambiente y de vocalizaciones de manatíes fueron muestreadas a 96 kHz.

Dado que el objetivo final del algoritmo de cancelación de ruido propuesto es facilitar el reconocimiento de vocalizaciones y la identificación de individuos, es conveniente realzar

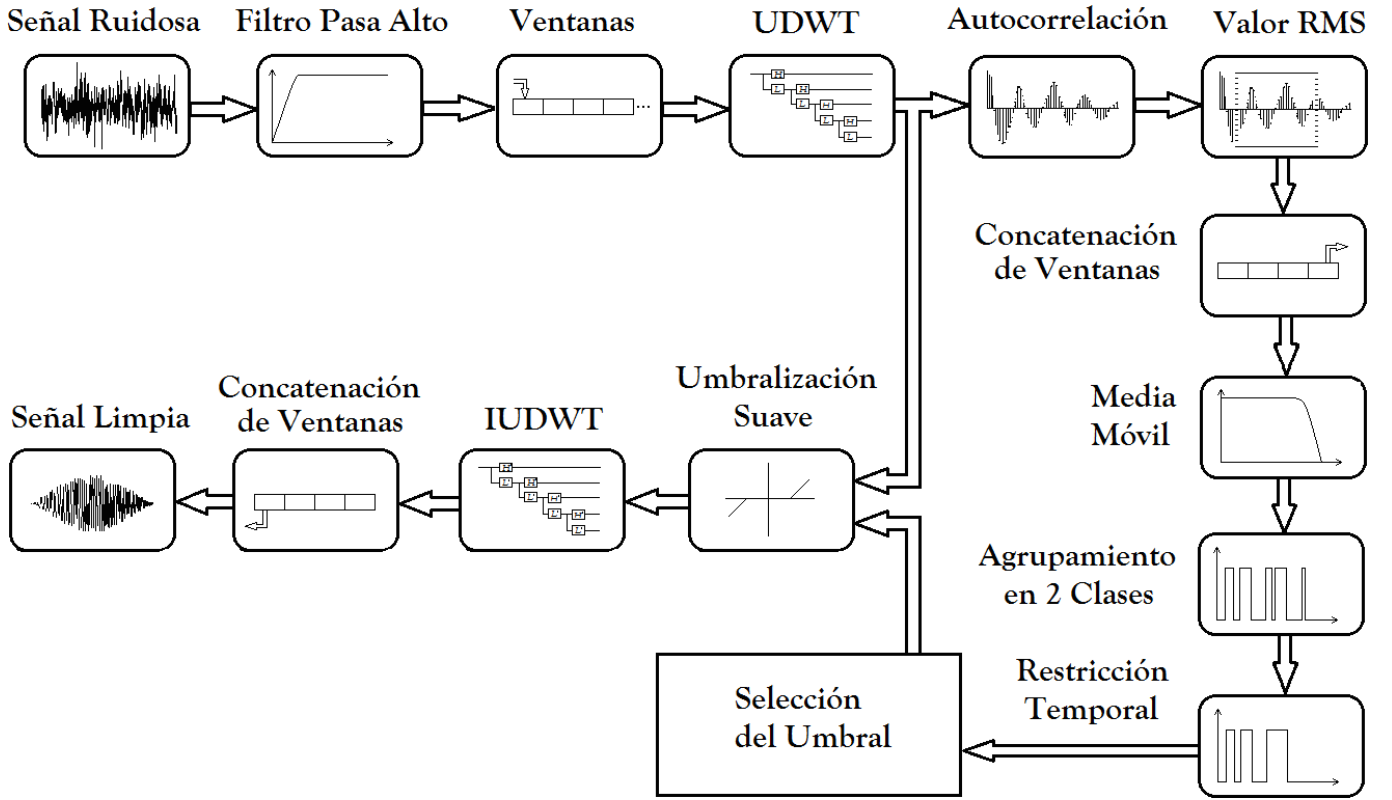


Figura 1. Diagrama de bloques del algoritmo de cancelación de ruido

las vocalizaciones y preservar su forma de onda original. Para cuantificar el realce alcanzado, se usa la razón de señal a ruido (SNR, por sus siglas en inglés), y para medir la preservación de la forma de onda, se usa la razón de señal a distorsión (SDR, por sus siglas en inglés) [5].

La SNR se obtiene mediante la fórmula

$$\text{SNR} = 10 \log_{10} \frac{\sum_{i=1}^{21} P_i}{\sum_{i=1}^{21} Q_i} \quad (4)$$

donde P_i es el valor RMS de la i ésima vocalización alterada y Q_i es el valor RMS de la sección de ruido, de igual longitud, precedente a la i ésima vocalización. La mejora en SNR se calcula como la diferencia entre la SNR de las vocalizaciones con el ruido cancelado $\hat{x}(t)$ y las vocalizaciones contaminadas $s(t)$. Para calcular la SNR de las vocalizaciones contaminadas, se usan las vocalizaciones originales $x(t)$, para desprestigiar el ruido concurrente sumado a cada vocalización.

La SDR se define como la razón entre la potencia del valor RMS de la vocalización original y la potencia del valor RMS de la señal de error:

$$\text{SDR} = 10 \log_{10} \frac{\sum_{i=1}^{21} D_i^2}{\sum_{i=1}^{21} E_i^2}, \quad (5)$$

donde D_i es el valor RMS de la i ésima vocalización original y E_i es el valor RMS del error presente en la i ésima vocalización alterada:

$$E_i = \sqrt{\frac{1}{N_i} \sum_{j=1}^{N_i} |\hat{x}_i(j) - x_i(j)|^2}, \quad (6)$$

donde N_i corresponde a la duración de la vocalización i ésima. La mejora en SDR se calcula como la diferencia entre la SDR de $\hat{x}(t)$ y $s(t)$. La SDR se escogió por encima del comúnmente utilizado error cuadrático medio normalizado (NMSE) porque brinda resultados en dB, que son comparables con la SNR. Dado que para calcular la SDR es necesario conocer la señal de vocalizaciones sin ruido, no se usaron grabaciones de vocalizaciones originalmente contaminadas con ruido de ambiente.

RESULTADOS

Los resultados obtenidos para la SNR y SDR al aplicar el algoritmo propuesto se muestran en el cuadro 1. La mejora promedio en SNR es de 26.93 dB y en SDR es de 1.27 dB. En la figura 2 se muestra la forma de onda de la señal de vocalizaciones original, contaminada con ruido y luego de aplicar el algoritmo de cancelación de ruido. El panel *c*) evidencia el buen desempeño obtenido, en cuanto a SNR, respecto a la señal del panel *b*).

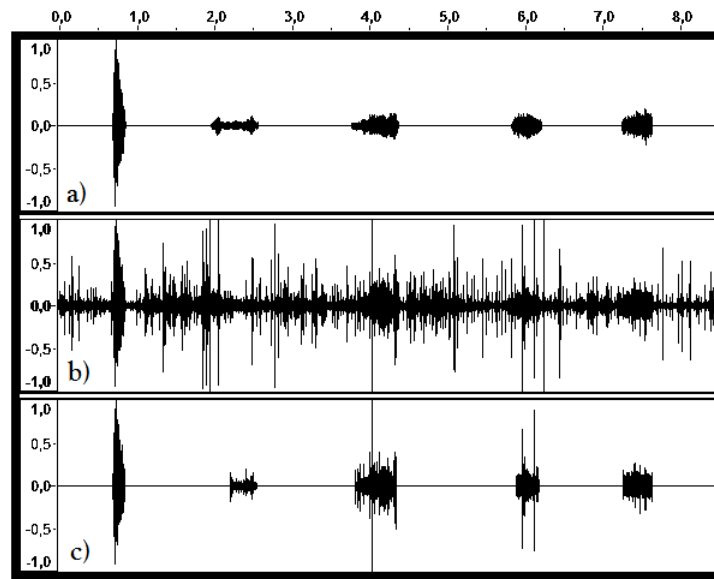


Figura 2. Forma de onda de a) las vocalizaciones originales, b) las vocalizaciones contaminadas con ruido ambiental y c) las vocalizaciones luego de aplicar el algoritmo de cancelación de ruido

Cuadro 1. Resultados del algoritmo propuesto, respecto a la SNR y la SDR, para grabaciones contaminadas con ruido de ambiente de alta intensidad.

Métrica	$s(t)$	$\hat{x}(t)$	Mejora
SNR	1.14	28.07	26.93
SDR	-1.61	-0.35	1.27

CONCLUSIONES Y TRABAJO A FUTURO

La notable mejora de la SNR indica que el algoritmo de cancelación de ruido propuesto logra realzar las vocalizaciones, ante ruido de ambiente de alta intensidad. Sin embargo, el valor negativo del SDR informa que la distorsión respecto a la forma de onda original es considerable. Esto revela que parte de la mejora reportada en SNR está asociada a ruido residual presente en las vocalizaciones.

Para mejorar la SDR se propone el uso de la transformada *wavelet packet*, la cual permitiría obtener una mejor resolución en frecuencia y aislar más efectivamente los armónicos de las emisiones vocales respecto al ruido de ambiente. Por otra parte, el algoritmo de cancelación de ruido funciona como un primer peldaño en el desarrollo de un método para el reconocimiento automático de vocalizaciones y el conteo automático de manatíes. Se pretende usar el método desarrollado en el instituto Smithsonian para la gestión de población de manatíes en Bocas del Toro, Panamá.

REFERENCIAS

1. Camacho, A. Detection of pitched/unpitched sound using pitch strength clustering. In *Proceedings of the Ninth International Conference on Music Information Proceedings of the Ninth International Conference on Music Information Retrieval* (2008), 533–537.
2. Deutsch, C., Self-Sullivan, C., and Mignucci-Giannoni, A. *Trichechus manatus*. Tech. rep., IUCN Red List of Threatened Species, 2008.

3. Donoho, D. L., and Johnstone, J. M. Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 81, 3 (1994), 425–455.
4. Gur, B. M., and Niezrecki, C. Autocorrelation based denoising of manatee vocalizations using the undecimated discrete wavelet transform. *Acoustical Society of America* 122 (2007), 188–199.
5. Gur, B. M., and Niezrecki, C. A source separation approach to enhancing marine mammal vocalizations. *The Journal of the Acoustical Society of America* 126, 6 (2009), 3062–3070.
6. Gur, B. M., and Niezrecki, C. A wavelet packet adaptive filtering algorithm for enhancing manatee vocalizations. *Journal of the Acoustical Society of America* 129(4) (2011), 2059–2067.
7. Ren, Y., Johnson, M. T., and Tao, J. Perceptually motivated wavelet packet transform for bioacoustic signal enhancement. *Acoustical Society of America* 124 (2008), 316–327.
8. Rivera, M., and Guzman, H. Consideraciones etológicas para el manejo del manatí antillano (*trichechus manatus*) en los humedales costeros del caribe norte de panamá y caribe sur de costa rica. preprint.
9. Yan, Z., Niezrecki, C., and Beusse, D. O. Background noise cancellation for improved acoustic detection of manatee vocalizations. *Acoustical Society of America* 117, 6 (2005), 3566–3573.
10. Yan, Z., Niezrecki, C., Cataffesta, Louis N., I., and Beusse, D. O. Background noise cancellation for manatee vocalizations using an adaptive line enhancer. *Journal of Acoustical Society of America* 120 (2006), 145–152.