

DETECCIÓN AUTOMÁTICA DE SONIDOS DE RORCUAL EN PRESENCIA DE EXPLOSIONES DE CAÑONES DE AIRE SÍSMICOS.

Ramón Miralles, Guillermo Lara

Instituto de Telecomunicación y Aplicaciones Multimedia (iTEAM), Univ. Politécnica de Valencia, Camino de Vera S/N, 46022, Valencia, SPAIN
rmiralle@dcom.upv.es

Resumen

Los pulsos acústicos de 20 Hz emitidos por los Rorcuales han sido ampliamente descritos por muchos autores. Sin embargo, estas ráfagas de pulsos regulares, son difíciles de detectar automáticamente dada su baja frecuencia y la facilidad con la que quedan enmascarados por las detonaciones producidas por los cañones de aire empleados en prospecciones sísmicas.

En este trabajo se propone un algoritmo de procesamiento de señal capaz de detectar y extraer de forma automática los pulsos de los Rorcuales de las grabaciones de acústica submarina en entornos ruidosos. El algoritmo propuesto explota a través de la DCT y del espectrograma las diferencias temporales y frecuenciales que existen entre los pulsos de Rorcual y las detonaciones sísmicas. Estas diferencias nos servirán para poder extraer y clasificar, tras un proceso de segmentación, los sonidos de los Rorcuales.

El algoritmo propuesto se comporta de forma robusta proporcionando buenas tasas de detección en presencia de explosiones producidas por prospecciones sísmicas.

Se presentan resultados reales de aplicación del algoritmo en sonidos de Rorcuales en el Mar Mediterráneo a su paso por las islas Columbretes en el 2011.

Palabras-clave: Procesado Tiempo frecuencia, acústica subacuática, bioacústica

Abstract

Fin Whales 20 Hz pulses have been described in detail by many authors. However, these bursts of regular pulses are difficult to be detected and extracted from underwater acoustic records. The low frequency of these sounds, alongside with the high probability of these sounds being masked in presence of underwater detonations produced by seismic airguns makes it difficult to devise automatic detection algorithms.

In this work we propose a signal-processing algorithm to automatically detect and extract Fin Whales pulses from underwater acoustic noise. The proposed algorithm takes advantage, by means of the Discrete Cosine Transform and the Spectrogram, of the temporal and frequency differences between the Fin Whale pulses and seismic airgun explosions. These differences will be valuable features to extract and classify, after a segmentation procedure, the Fin Whale sounds.

The proposed algorithm allows good detection ratio and has proven to be robust in the presence of noise from seismic prospecting.

We also present real results using the proposed algorithm, to detect Fin Whales pulses in the Mediterranean Sea as they passed close to Columbretes Islands in 2011.

Keywords: Time-frequency signal processing, Underwater Acoustics, bioacoustics

PACS no. 43.60.Hj, 43.30.Sf

1 Introducción

La protección y conservación del rorcual común (*Balaenoptera physalus*) en el Mar Mediterraneo, requiere de un completo conocimiento de las rutas y de los patrones migratorios de la especie. Tradicionalmente esta información se ha venido obteniendo a través de costosas campañas de avistamiento y de colaboraciones entre diferentes organismos. La información así obtenida esta sujeta a un gran sesgo ya que depende de que los animales suban a la superficie y la frecuencia de avistamientos de esta especie es baja a pesar de que la población total de ejemplares en el Mediterraneo podría ser superior a los 3500 individuos [1].

Por otro lado, y gracias al abaratamiento de la tecnología es cada vez mas sencillo contar con sistemas de grabación de audio submarinos que pueden ser usados para obtener información del paso de estos cetáceos. Las señales acústicas de 20 Hz emitidas por el rorcual común han sido ampliamente descritas por muchos autores y su utilidad en el estudio del comportamiento de las ballenas demostrada [2]. Estos pulsos, pese a que pueden llegar a tener una intensidad elevada [2], resultan muy difíciles de detectar de forma automática en algunas situaciones. Como ejemplo, mencionar la detección en presencia de detonaciones producidas por los cañones de aire empleados en prospecciones sísmicas. La frecuencia y la periodicidad de las detonaciones pueden confundir fácilmente a los algoritmos automáticos de detección. Además, la presencia de ruidos de motores de barcos y la atenuación acústica en situaciones en las que el paso del rorcual está alejado del hidrófono hacen aún más complicada esta tarea.

En el presente trabajo se presenta un algoritmo de detección automática de sonidos del rorcuales capaz de funcionar en situaciones con baja relación señal a ruido y con grabaciones contaminadas por ruidos de diferentes naturalezas: ruidos con una fuerte componente tonal persistente en el tiempo (como los producidos por motores de barcos, e interferencias) y ruidos impulsivos (como los producidos por las detonaciones en prospecciones sísmicas).

El artículo se estructura en cinco partes. En la sección 2 haremos una breve descripción estadística y frecuencial de los pulsos de rorcual común. La información obtenida se explotará en las secciones siguientes para diseñar detectores robustos. Así, en la sección 3 describiremos el preprocesado de las grabaciones. Este preprocesado permite eliminar (empleando la Transformada Coseno) una gran cantidad de ruidos con persistencia temporal. En la sección 4 se mostrará como empleando técnicas de procesado de imagen aplicadas al espectrograma preprocesado se puede diferenciar entre las detonaciones y los pulsos de 20 Hz. Finalmente en la sección 5 se mostraran comparativas del algoritmo propuesto con el detector del programa XBAT.

2 Breve caracterización estadística y frecuencial de los pulsos de Rorcual Común.

Para la realización de este trabajo se han utilizado datos proporcionadas por los biólogos del Oceanogràfic de Valencia. Se trata de grabaciones realizadas mediante un dispositivo Ecological Acoustic Recorder (EAR) desarrollado por el Oceanwide Science Institute de Hawaii. Las grabaciones se realizaron en el Mar Mediterráneo, en las Islas Columbretes, provincia de Castellón en las fechas que van desde el día 21 de Septiembre del 2011 al día 22 de Diciembre del 2011. Se grabaron 4 horas al día durante 3 meses obteniendo un total de 360 horas de grabaciones. Dichas grabaciones se realizaron a una frecuencia de muestreo $f_s=250$ Hz. También se proporcionó el historial detallado de las detecciones de pulsos de rorcual etiquetadas manualmente por los biólogos del Oceanogràfic.

Es importante caracterizar a nivel estadístico y frecuencial los pulsos de rorcual para poder explotar sus características en la detección de estos. Para ello, se calcularon las frecuencias mínimas y máximas de los pulsos, la duración media, así como el tiempo medio de repetición entre pulsos. La Tabla 1 resume esta información. También se estimó el histograma del tiempo medio entre pulsos (Figura 1). Dicho histograma se utilizó para calcular la media de separación entre pulsos así como su desviación típica.

Tabla 1: Valores medios estimados de las grabaciones realizadas por el EAR en Columbretes.

	Valores medios estimados
Frecuencia inferior pulso	17.15 Hz
Frecuencia superior pulso	24.89 Hz
Duración pulso	1.8 segundos
Número medio de pulsos consecutivos de las ráfagas de pulsos de rorcual	21 pulsos

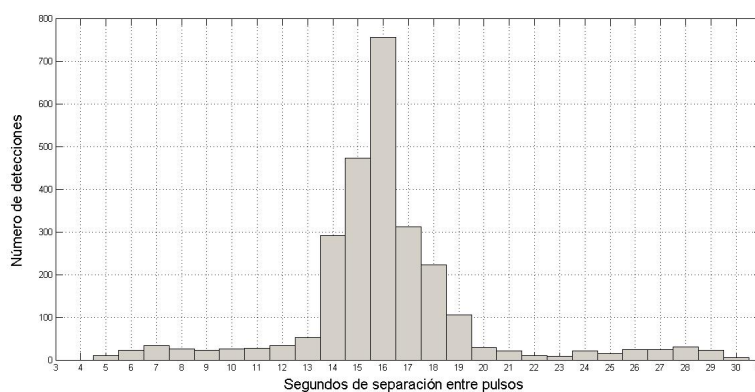


Figura 1: Histograma de las separaciones entre los pulsos de rorcual proporcionadas por los biólogos del Oceanográfico

3 Preprocesado del espectrograma: eliminación de ruido persistente en el tiempo a través de la Transformada Coseno.

Sea $x(n)$ la secuencia discreta con frecuencia de muestreo f_s , registrada por la sonoboya y susceptible de contener pulsos de 20 Hz de rorcual contaminados con ruido de detonaciones. En primer lugar obtendremos la representación tiempo frecuencia mediante el espectrograma $P(n, k)$ [3].

$$P(n, k) = \left| DFT_{N_2} [w(m) \cdot x(n - m)] \right|^2, \quad (1)$$

siendo DFT_{N_2} la transformada de Fourier de tiempo y frecuencia discreta, N_2 es el número de muestras equidistantes desde $\omega = 0$ a 2π , y $w(n)$ la ventana de análisis. El tamaño de la matriz $P(n, k)$ será $N_1 \times N_2$ donde N_1 dependerá de la longitud total de la secuencia de audio $x(n)$ (o de un fragmento de esta), del avance y del solape entre muestras. La frecuencia discreta k se obtendrá por muestreo de la circunferencia unidad con N_2 puntos ($k = 0, 1, \dots, N_2 - 1$). A lo largo de todos los

periodogramas calculados en el presente trabajo se ha utilizado un solape de un 25% y una ventana de análisis de 128 puntos. Estos valores, junto con los 250 Hz de frecuencia de muestreo, dan como resultado que un pixel de ancho en el eje temporal del espectrograma equivale aproximadamente a 0.4 segundos.

La Figura 2 muestra el espectrograma de un fragmento de señal captada por un hidrófono en la que se pueden observar ruido concentrado en frecuencias o tonal (en rojo) y detonaciones (en azul). Es importante observar, para entender el funcionamiento del preprocesado, que el ruido tonal distribuye gran parte de su energía en el tiempo horizontalmente, mientras que el ruido impulsivo la distribuye verticalmente. La distribución y manipulación de las componentes horizontales y verticales de la energía resulta una tarea sencilla empleando la transformada coseno [4].

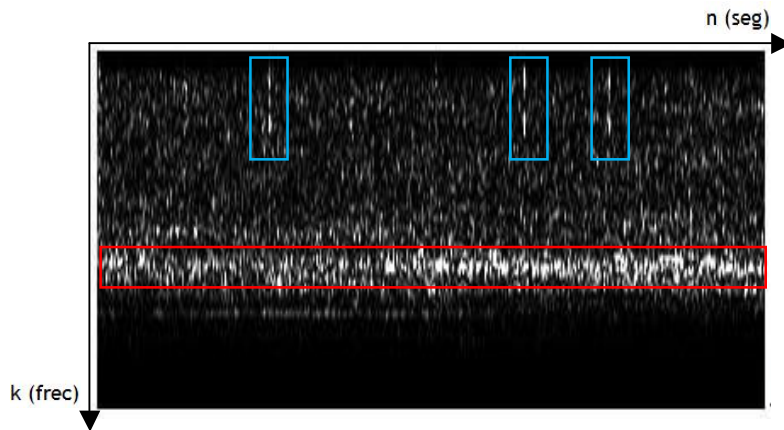


Figura 2: Espectrograma de un archivo de sonido.

Si procesamos la matriz que contiene el espectrograma $P(n, k)$ como una imagen en escala de grises y calculamos la Transformada Discreta Coseno Bidimensional (DCT_{2D}) podemos eliminando selectivamente algunos coeficientes de la Transformada Discreta Coseno (DCT) eliminar a voluntad la energía que se encuentre distribuida en las componentes horizontales así como las derivas temporales lentas que pudiera haber sufrido la grabación de audio submarina.

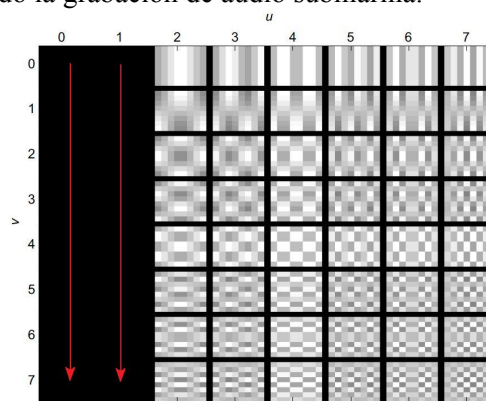


Figura 3: Funciones base de una transformada coseno de tamaño de 8x8 donde se ilustra, a modo de ejemplo, como las dos primeras columnas se han eliminado. En estas dos columnas se encuentra la información de las variaciones en la dirección vertical y variaciones suaves en la dirección horizontal ($BW_h = 2$).

Sean $F(u, v) = DCT_{2D} \{P(n, k)\}$ los coeficientes de la DCT del Espectrograma. La eliminación selectiva de coeficientes podría realizarse de forma directa haciendo (ver Figura 3),

$$F'(u, v) = \begin{cases} 0 & u = 0, 1, \dots, BW_h \\ F(u, v) & u > BW_h \end{cases}, \quad (2)$$

y recuperando $P_{denosied}(n, k) = DCT_{2D}^{-1} \{F'(u, v)\}$. En la ecuación (2), BW_h indica el número de columnas de la DCT_{2D} a eliminar y estaría relacionado con la frecuencia de corte en la dirección horizontal de un supuesto filtro paso alto equivalente. En la mayoría de las situaciones, y dado que el número de columnas a eliminar no será muy elevado una alternativa computacionalmente menos costosa será restar una imagen con las componentes a eliminar ($P_r(n, k)$):

$$P_{denosied}(n, k) = P(n, k) - P_r(n, k), \quad (3)$$

donde $P_r(n, k) = DCT_{2D}^{-1} \{F(u, v) \cdot W(u)\}$ y $W(u)$ es una función que selecciona sólo algunas columnas de $F(u, v)$, como se indica en la ecuación (4).

$$W(u) = \begin{cases} 1 & u = 0, 1, \dots, BW_h \\ 0 & \text{resto } u \end{cases}. \quad (4)$$

Aprovechando las propiedades de la Transformada Discreta Coseno Multidimensional $P_r(n, k)$ puede calcularse empleando únicamente transformadas directas e inversas unidimensionales. Supongamos definimos los operadores $DCT_{nu}[\cdot]$ y $DCT_{kv}[\cdot]$ como los operadores que calculan la DCT en la variable n y k respectivamente. Igualmente definimos sus inversas $DCT_{un}^{-1}[\cdot]$ y $DCT_{vk}^{-1}[\cdot]$. Podemos escribir:

$$DCT_{kv} [DCT_{nu} [P(n, k)]] = F(u, v). \quad (5)$$

Para obtener $P_r(n, k)$ seleccionamos sólo las componentes deseadas del término de la derecha de la ecuación (5) y tomamos transformadas inversas.

$$DCT_{kv} [DCT_{nu} [P(n, k)]] \cdot W(u) = F(u, v) \cdot W(u). \quad (6)$$

$$P_r(n, k) = DCT_{un}^{-1} [DCT_{vk}^{-1} [F(u, v) \cdot W(u)]]. \quad (7)$$

Sustituyendo la ecuación (6) en (7) tenemos:

$$P_r(n, k) = DCT_{un}^{-1} [DCT_{nu} [P(n, k)] \cdot W(u)]. \quad (8)$$

O también desarrollando las expresiones de las DCT en forma de sumatorios:

$$\begin{aligned}
 P_r(n, k) = & \frac{1}{2} \sum_{n=0}^{N_1-1} P(n, k) + \\
 & + \sum_{u=1}^{BW_n-1} \left(\sum_{n=0}^{N_1-1} P(n, k) \cdot \cos \left[\frac{\pi}{N_1} \left(n + \frac{1}{2} \right) u \right] \right) \cdot \cos \left[\frac{\pi}{N_1} \left(n + \frac{1}{2} \right) u \right].
 \end{aligned} \tag{9}$$

La ecuación (9) nos permite calcular, con un coste computacional bajo, un espectrograma que contiene sólo las componentes indeseadas con variación temporal lenta y que restadas al periodograma original facilitan la detección de los pulsos de rorcual así como la diferenciación de los ruidos producidos por detonaciones sísmicas. En la Figura 4 se muestra un ejemplo de un periodograma simulado con componentes horizontales y verticales y el efecto del preprocesado propuesto para eliminar las componentes indeseadas. Tal y como se esperaba se eliminan las componentes de ruido con dirección horizontal, preservándose las componentes verticales.

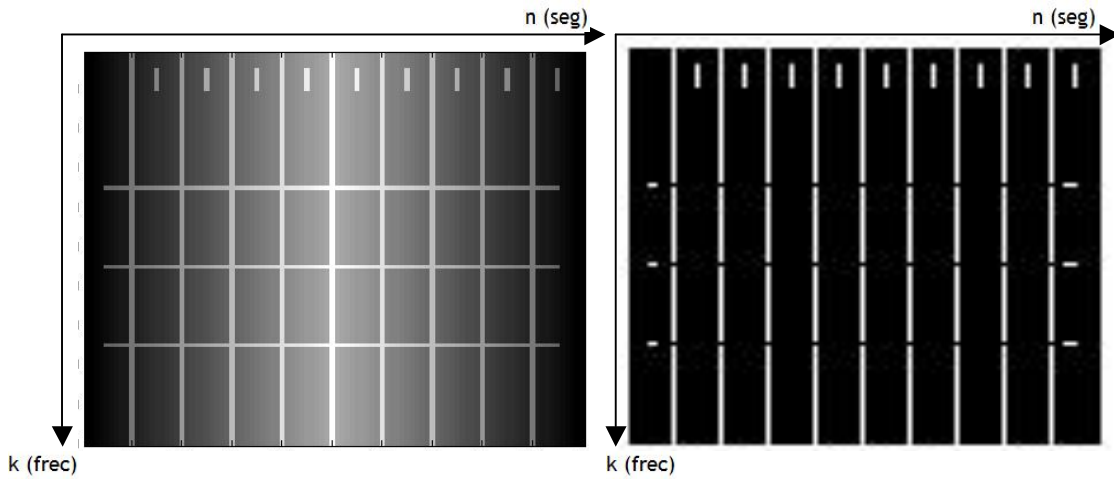


Figura 4: Ejemplo de espectrograma simulado. Izquierda: espectrograma original $P(n, k)$. Derecha espectrograma con el preprocesado propuesto $P_{denoised}(n, k)$.

4 Segmentación del espectrograma y detección de eventos.

Dividiremos esta sección en dos subsecciones. En la primera mostraremos como binarizar el espectrograma aplicando técnicas de segmentación de imagen. En la segunda indicaremos como se ha llevado a cabo el etiquetado y diferenciación entre los pulsos de rorcual y detonaciones sísmicas.

4.1 Umbralización del espectrograma mediante el método Otsu modificado.

Una de las técnicas de la segmentación de imágenes es la umbralización. Se emplea cuando hay una clara diferencia entre los objetos a extraer respecto del fondo de la escena, en nuestro caso, los pulsos de rorcuales y detonaciones ya que los ruidos tonales habrán desaparecido con el preprocesado.

Si procesamos $P_{denosied}(n, k)$ como si una imagen en escala de grises se tratara, el resultado será una imagen binaria en la que los píxeles etiquetados con '1' corresponderán a los eventos a detectar y los píxeles etiquetados con '0' al ruido de fondo.

La elección del umbral T en el proceso de binarización es sumamente importante para asegurarnos una correcta detección y un nivel de falsas alarmas controlado. El umbral puede depender de forma global de $P_{denosied}(n, k)$, de alguna propiedad local de los píxeles vecinos, y hasta dinámicamente de la posición (n, k) del píxel. De entre todas las técnicas de umbralización disponibles en la bibliografía [6] [7] se ha seleccionado el método Otsu [5], como una alternativa sencilla y automática que garantiza la máxima separabilidad de las clases resultantes.

Dado que los eventos a detectar ocurren en unas frecuencias determinadas, modificaremos el método de Otsu, obteniendo N_2 umbrales óptimos, uno para cada fila de píxeles (frecuencias) del espectrograma, con ello obtendremos una mayor independencia a la hora de umbralizar toda la imagen conjunta adaptando el umbral según la frecuencia a la que estemos trabajando.

Supongamos que la matriz $P_{denosied}(n, k)$ presenta unos niveles de intensidad que se encuentran entre 1 y L . Para la umbralización en dos niveles el vector de umbrales óptimo $T^*(k)$ se obtiene maximizando cada $\sigma_{B_k}^2$ de acuerdo con las fórmulas de Otsu [5],

$$T^*(k) = \max_{1 \leq T(k) \leq L} \left\{ \sigma_{B_k}^2(T(k)) \right\} \quad (10)$$

donde $\sigma_{B_k}^2$ es la varianza entre clases para la fila k -ésima [5]. Las ventajas de emplear una umbralización Otsu por filas se hacen evidentes en el ejemplo que se muestra en la Figura 5, donde se aprecia fácilmente como utilizando la umbralización propuesta se permite distinguir los pulsos de rorcual y de detonaciones. La umbralización global, por el contrario, no permite detectar los pulsos de rorcual. La explicación de este comportamiento esta en la curva de respuesta en frecuencia del hidrófono que presenta diferentes sensibilidades a diferentes frecuencias.

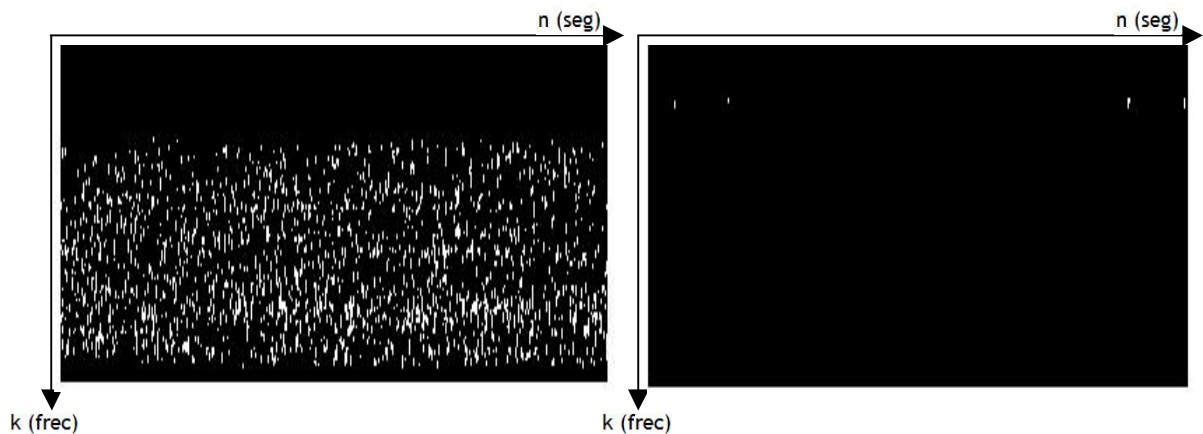


Figura 5: Ejemplo de binarización del espectrograma. Izquierda con el método Otsu original. Derecha con el método Otsu por filas.

4.2 Segmentación y etiquetado de pulsos de rorcual.

Una vez binarizado el espectrograma, deberemos seleccionar tan solo aquellas regiones de éste que provengan de pulsos de rorcual diferenciándolas de otras regiones que pudieran aparecer como resultado de una detonación. Para realizar esta diferenciación se empleará información frecuencial y estadística así como un etiquetado por área tal y como se explica a continuación.

Supondremos en primer lugar que gracias al preprocesado y la umbralización los únicos píxeles etiquetados como “1” son aquellos que provienen pulsos y detonaciones. Definimos un evento como un conjunto de píxeles detectados con una continuidad, es decir que estén unidos entre ellos.

Sobre el periodograma binarizado seleccionamos los eventos que tengan un área mayor o igual a 3 píxeles. A continuación, y para diferenciar entre pulso y detonación nos basaremos en las propiedades frecuenciales y estadísticas de los eventos aplicando las reglas detalladas en la Tabla 2.

Tabla 2: Reglas para diferenciar eventos de pulsos de rorcual de detonaciones sísmicas.

1.- Si duración (eje n) > 5 píxeles se elimina el evento.	Éste ancho en píxeles se obtiene* como umbral superior de la duración media del pulso de rorcual obtenida de la Tabla 1.
2.- Si frecuencias del evento (eje k del espectrograma) $\notin [17, 25]$ Hz se elimina el evento.	Con ello eliminaremos las detonaciones y nos quedaremos solo con los pulsos de rorcual. Las detonaciones tienen una naturaleza impulsiva muy parecida a los pulsos de rorcual, con la diferencia de que tienen un ancho de banda de 60.35 Hz mayor que los 8 Hz que poseen los pulsos de rorcual.
3.- Si separación entre eventos (eje n) < 36 píxeles se elimina el evento.	Este valor se obtiene de convertir* la tasa mínima de repetición entre pulsos (Figura 1).

Este proceso de etiquetado, selección por área y posterior filtro según las reglas definidas se repetirá para eventos de área 2 píxeles, con la salvedad que en este caso no será necesario eliminar los eventos que posean un ancho mayor de 5 píxeles, ya que habrán sido eliminados anteriormente. Todos los eventos que queden después de aplicar las reglas de filtrado serán considerados como eventos candidatos a provenir de pulsos de rorcual.

Dado que cada vez que aparecen pulsos de rorcual estos aparecen en ráfagas de en media 20 pulsos y siempre de más de 3 pulsos consecutivos utilizaremos esta información para refinar las detecciones y buscar posibles detecciones que pudieran haber escapado al algoritmo. Si el número de eventos

* En función de las variables implicadas en el cálculo del espectrograma (f_s , tamaño de ventana de análisis y solape entre ventanas).

candidatos es menor que 3, consideraremos que no son pulsos de rorcual y serán descartados (ver histograma en la Figura 1). Por el contrario, si el número de eventos candidatos es mayor o igual a 3, serán considerados pulsos de rorcual y procederemos a buscar otros posibles pulsos que pudieran no haberse detectado y que se encontraran antes o después de los detectados. Para ello, procederemos volviendo a examinar el espectrograma binarizado pero en este caso buscando eventos de área igual a 1 píxel. Para evitar aumentar el número de falsas detecciones sólo realizaremos esta búsqueda en una pequeña región del periodograma. Esta región en la que se buscarán posibles pulsos de rorcual se obtendrá con una máscara en las posiciones que mas probablemente ocuparían los pulsos de rorcual no detectados. Esta máscara se crea a partir del vector de separaciones (*sep*) entre los pulsos de rorcual que ya tenemos detectados tal y que se detalla en la Tabla 3.

La justificación de este algoritmo se basa en que tal y como se ha presentado en la sección 2, la mayoría de las separaciones entre los pulsos están entre 14 y 18 segundos (entre 36 y 46 píxeles en el periodograma). Si obtenemos dos grupos de pulsos de rorcual con una separación entre ellos de más de 46 píxeles existe una amplia probabilidad de que entre ellos existan más pulsos que no se hayan detectado. Si bien no podemos establecer un patrón totalmente regular de separaciones a la hora de realizar una búsqueda más exhaustiva si que es posible buscar, de entre todas las combinaciones posibles, un patrón de posiciones con separaciones variables entre 36 y 46 píxeles, en los que hay una alta probabilidad de que se encuentre un pulso (o varios pulsos) de rorcual.

Cabe destacar que en ocasiones habrá varias posibles combinaciones válidas. En estos casos el algoritmo seleccionara de forma automática el vector para crear la máscara en el que las separaciones entre pulsos estén lo más cercanas posibles a 16 segundos (40 píxeles) que es la media entre pulsos de rorcual (ver Figura 1).

5 Comparativas

El algoritmo propuesto se ha comparado con los resultados de detecciones realizadas por los biólogos del Oceanogràfic de Valencia. Dichas detecciones fueron llevadas a cabo por expertos escuchando y visualizando los espectrogramas de las grabaciones de audio. Las grabaciones empleadas han sido las mismas que se han descrito en la sección 2. También se ha comparado el algoritmo propuesto con una plantilla de detección del XBAT (Programa de código abierto para MATLAB R2006a desarrollado por Harold Figueroa y habitualmente empleado en numerosos trabajos de bioacústica). Los algoritmos de detección empleados en este programa están basados en máscaras y correlaciones. El empleo de este tipo de detectores en condiciones de baja relación señal a ruido no es una técnica apropiada llegándose a obtener probabilidades de detección muy bajas tal y como se verá a continuación.

Los resultados obtenidos de la comparación de las detecciones se muestran en la Tabla 4. La tabla muestra la probabilidad de detección y la probabilidad de falsa alarma de los dos detectores comparados. Se observa, tal y como habíamos mencionado, la baja capacidad de detección basada en correlaciones (XBAT), que no supera el 9%. Es interesante resaltar con respecto al algoritmo de detección propuesto, que si bien la tasa de detecciones no es muy elevada (en torno al 50%) se consigue una baja probabilidad de falsa alarma (cercana al 1 %). Se ha comprobado que en situaciones en las que es tolerable subir la probabilidad de falsa alarma hasta el 10% se puede llegar a alcanzar probabilidades de detección cercanas al 80%.

Tabla 3: Algoritmo de extracción del patrón más probable de separaciones entre pulsos.

<p>Inputs vector <i>sep</i> //separación de los pulsos detectados.</p> <p>Outputs vector <i>patron</i> //máscara donde se buscarán los posibles pulsos de rorcual.</p> <p>Inicializamos $i = 1$, $d = 40$ //número medio de píxeles de separación entre pulsos, $max = 46$ //máxima separación entre pulsos, $min = 36$ //mínima separación entre pulsos.</p> <ol style="list-style-type: none"> 1. Si $sep(i) < max$, añadimos $sep(i)$ a <i>patron</i> y vamos al paso 11 2. $cociente_1 = sep(i)/d$ 3. $c_1 = round(cociente_1)$ 4. Si $sep(i)/d - c_1 > 0$ //el entero más cercano es el inferior <ul style="list-style-type: none"> ▪ $r_1 = sep(i) - d \cdot c_1$ 5. Si $sep(i)/d - c_1 \leq 0$ //el entero más cercano es el superior <ul style="list-style-type: none"> • $r_1 = d \cdot c_1 - sep(i)$ 6. $cociente_2 = r_1/c_1$ 7. $c_2 = floor(cociente_2)$ 8. $r_2 = r_1 - c_1 \cdot c_2$ 9. Si $sep(i)/d - c_1 > 0$ //el entero más cercano es el inferior <ul style="list-style-type: none"> ▪ Si $c_2 < c_1$ <ul style="list-style-type: none"> • Añadimos r_2 veces el valor $(d + c_2 + 1)$ a <i>patron</i> • Añadimos $c_1 - r_2$ veces el valor $(d + c_2)$ a <i>patron</i> ▪ En caso contrario añadimos $sep(i)$ a <i>patron</i> 10. Si $sep(i)/d - c_1 \leq 0$ //el entero más cercano es el superior <ul style="list-style-type: none"> ▪ Si $c_2 < c_1$ <ul style="list-style-type: none"> • Añadimos r_2 veces el valor $(d - c_2 - 1)$ a <i>patron</i> • Añadimos $c_1 - r_2$ veces el valor $(d - c_2)$ a <i>patron</i> ▪ En caso contrario añadimos $sep(i)$ a <i>patron</i> 11. $i = i + 1$ y vamos al paso 1 //hasta el final de <i>sep</i>
--

Tabla 4: Comparativa de las detecciones y estadísticas de probabilidades de detección y falsa alarma.

	Detecciones reales	Detector propuesto	Detecciones XBAT
Número de detecciones	2101	1111	186
Probabilidad de detección	-	52,88%	8,85%
Probabilidad de falsa alarma	-	1,38%	2,28%

Para ilustrar el funcionamiento en el tiempo del detector propuesto se muestra la Figura 6. En ella se muestran sólo las detecciones del primer mes y medio (por claridad en la gráfica). Se observa en la figura el buen comportamiento y como las pérdidas de detección se reparten de forma aproximadamente uniforme en todo el periodo mostrado.

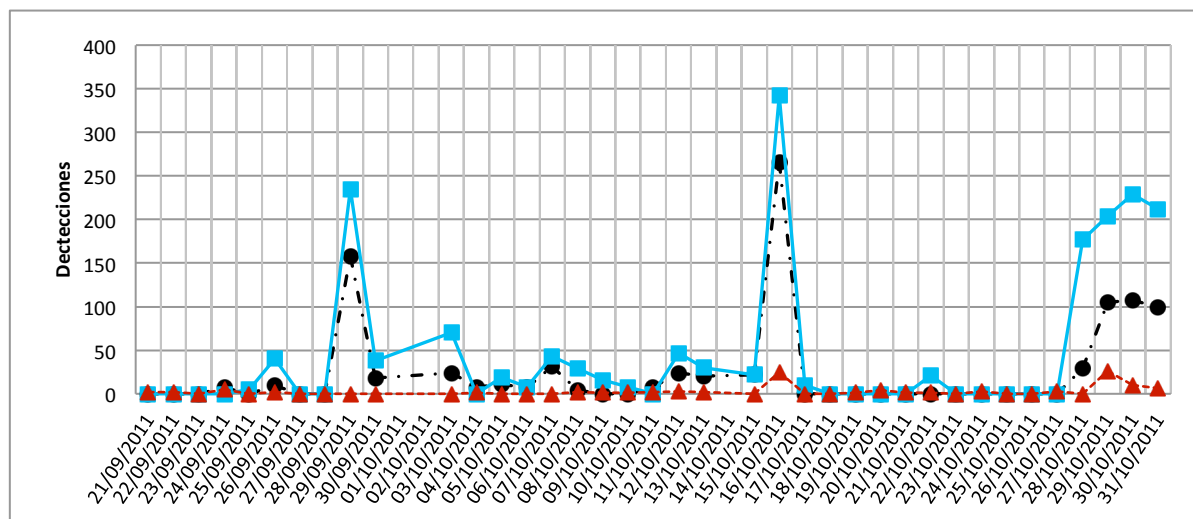


Figura 6: Comparativa del funcionamiento de los detectores en las grabaciones proporcionadas. Línea continua azul y “■”: detecciones proporcionadas por los biólogos. Línea negra a trazos y “●”: detector propuesto. Línea roja a trazos y “▲”: detector XBAT.

6 Conclusiones

En el presente trabajo se ha propuesto un algoritmo capaz de detectar pulsos de rorcual común en grabaciones de audio submarinas. En el desarrollo de este algoritmo se han empleado técnicas típicas de procesado y morfología de imagen aplicadas a la detección de pulsos sobre el espectrograma. El procesado del espectrograma como si de una imagen se tratase junto con la inclusión de información estadística de los patrones de repetición de pulsos, ha demostrado ser una técnica eficaz de obtener altas tasas de detección en situaciones de baja relación señal a ruido así como en presencia de detonaciones consiguiendo altas tasas de detección con unas falsas alarmas muy moderadas. Comparativas del detector propuesto con otro detector habitualmente empleado muestran la gran superioridad de la solución que se describe en el presente trabajo.

Agradecimientos

Este trabajo ha sido financiado por el MICINN, Plan Nacional de I+D+i con referencia TEC2011-23403 (España), y la Generalitat Valenciana PROMETEO 2010/040. Los autores agradecen a los biólogos del Oceanogràfic por el tiempo dedicado y por proporcionar las grabaciones que han podido dar lugar a estos resultados.

Referencias

- [1] G. Notarbartolo-Di-Sciara, M. Zanardelli, M. Jahoda, S. Panigada, and S. Airoidi, “The fin whale *Balaenoptera physalus* (L. 1758) in the Mediterranean Sea,” *Mammal Review*, vol. 33, no. 2, pp. 105–150, Jun. 2003.
- [2] W. A. Watkins, “The 20-Hz signals of finback whales (*Balaenoptera physalus*),” *The Journal of the Acoustical Society of America*, vol. 82, no. 6, p. 1901, 1987.
- [3] S. Haykin, editor, *Advances in Spectrum Analysis and Array Processing*, vol.1, Prentice-Hall, 1991.
- [4] Miralles R, Lara G, Esteban JA, Rodriguez A., The pulsed to tonal strength parameter and its importance in characterizing and classifying Beluga whale sounds, *J Acoust Soc Am. Mar*;131(3), pp. 2173-9, 2012.
- [5] Nobuyuki Otsu, “A threshold selection method from gray-level histogram”, *IEEE Transactions on System Man Cybernetics*, Vol. SMC-9, No. 1, pp. 62-66, 1979.
- [6] Ping-Sung Liao, Tse-Sheng Chen and Pau-Choo Chung, “A Fast Algorithm for Multilevel Thresholding”, *Journal of Information Science and Engineering* 17, 713-727, (2001).
- [7] A.S. Abutaleb “Automatic thresholding of gray level pictures using two-dimensional entropy”, *Comput. Vision Graphics Image Process.*, 47 (1989), pp. 22–32