



# PROYECTO SILENCIO

ANÁLISIS DE LOS REGISTROS DE RUIDO PARA LA IDENTIFICACIÓN DE PRINCIPALES FUENTES DE RUIDO E IMPLEMENTACIÓN DE UN ALGORITMO PARA IDENTIFICAR POSIBLE PRESENCIA DE CETÁCEOS

Versión: **Final**  
Fecha: **29/09/2021**  
Responsable: **Soledad Torres Guijarro**

## Contenido

1. Introducción.....	2
2. Análisis de las grabaciones del hidrófono de la plataforma de Cortegada. ....	2
2.1. Ambiente sonoro.....	2
2.2. Contaminación por ruidos de la plataforma. ....	3
2.3. Mejora el sistema de fondeo. ....	4
3. Cálculo de indicadores de ruido ambiental.....	6
3.1. Descripción de indicadores.....	6
3.2. Método de cálculo.....	7
3.3. Código.....	7
4. Detección automática de barcos.....	8
4.1. Estado del arte.....	9
4.2. Método de detección.....	9
4.3. Código.....	10
4.4. Resultados.....	11
5. Detección automática de cetáceos.....	12
5.1. Estado del arte.....	12
5.2. Método de detección.....	13
5.3. Código.....	13
5.4. Resultados.....	15
6. Mejoras en el código para su ejecución in situ.....	17
6.1. Optimización del tiempo de ejecución.....	17
6.2. Incorporación de un codificador de audio.....	18
7. Conclusiones.....	19
REFERENCIAS.....	20

Este trabajo ha sido realizado por la Universidad de Vigo para el proyecto SILENCIO, del Centro Tecnológico del Mar, que se desarrolla con la colaboración de la Fundación Biodiversidad, del Ministerio para la Transición Ecológica y el Reto Demográfico, a través del Programa Pleamar, cofinanciado por el FEMP.

*Las opiniones y documentación aportadas en esta publicación son de exclusiva responsabilidad del autor o autores de los mismos, y no reflejan necesariamente los puntos de vista de las entidades que apoyan económicamente el proyecto*

## 1. Introducción

El presente trabajo tiene como primer objetivo desarrollar una herramienta que permita diferenciar y caracterizar las fuentes naturales y antropogénicas en los registros del hidrófono instalado en la Plataforma de Cortegada. Para conseguirlo, se plantean los objetivos secundarios siguientes:

- Analizar las grabaciones realizadas durante la duración del proyecto (10 meses de datos), identificando fuentes de sonido de origen natural y fuentes de origen antropogénico, y extrayendo la información necesaria para evaluar el ruido ambiental en la zona
- Implementar un programa que detecte automáticamente sonidos de cetáceos y de tráfico marino, programado en Python, y que pueda ejecutarse en tiempo real el hardware instalado en la propia plataforma. Evaluar sus prestaciones sobre las grabaciones de Cortegada.
- Identificar posibles mejoras en la instalación y configuración del hidrófono

## 2. Análisis de las grabaciones del hidrófono de la plataforma de Cortegada.

El sonido que capta el hidrófono instalado en la Plataforma de Cortegada es una combinación de componentes presentes en el ambiente sonoro submarino de la zona, algunas de origen natural y otras antropogénico, y de componentes contaminantes, entendiéndose como tales las que genera la propia plataforma, los equipos instalados en ella, y las producidas por la interacción del hidrófono con el medio. A continuación describimos cada una de ellas, y las mejoras introducidas en el sistema de fondeo del hidrófono para reducir las componentes contaminantes.

### 2.1. Ambiente sonoro.

Los sonidos de mayor energía presentes en las grabaciones son de origen antropogénico, concretamente los generados por embarcaciones. Durante los meses de recogida de la almeja (normalmente de octubre a mayo), entre las 6 y las 12 horas UTC aproximadamente, predomina el ruido de los motores fueraborda de las embarcaciones que trabajan en el banco de marisqueo de Carril y que pesan su recolección en la propia plataforma. Este ruido enmascara casi totalmente al resto de fuentes sonoras. La figura 1 muestra el nivel de ruido del día 5/08/2018, día de recolección de almeja; puede observarse el incremento entre las 6 y las 12 horas UTC (8 y 14h hora local), correspondiente a su jornada laboral.

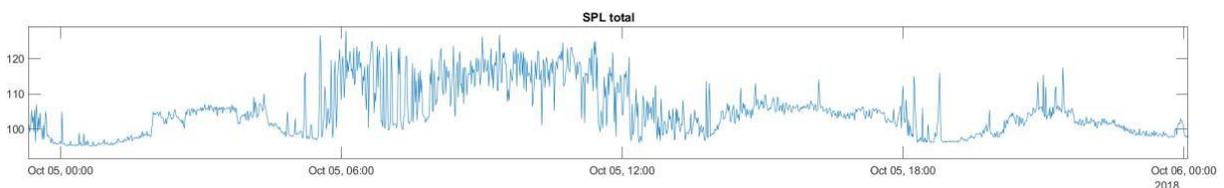


Fig.1: nivel de ruido de un día de recolección de almeja

Fuera de estos periodos de actividad extractiva, con ayuda de datos AIS (Sistema de Identificación Automática) se ha podido identificar en las grabaciones el sonido producido por

pasos aislados de barcos de pasajeros, pesqueros y cargueros. La figura 2 muestra un espectrograma de un paso aislado de un barco.

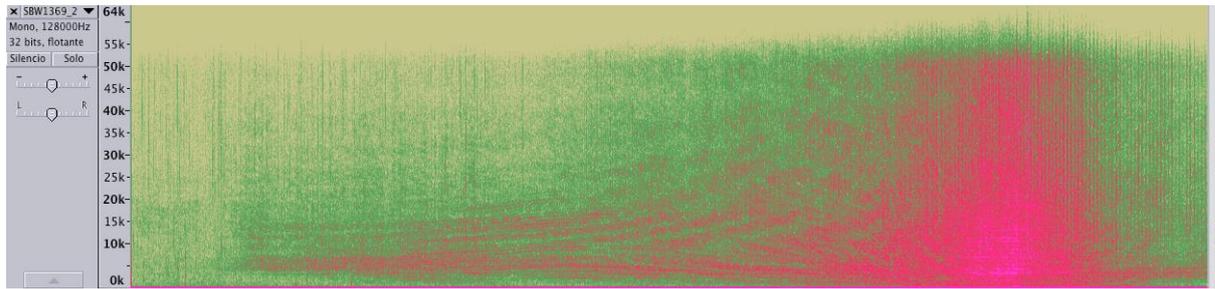


Fig. 2: Espectrograma del paso aislado de un barco

Entre los sonidos de origen natural, cabe destacar los silbidos, *bursts* y clics de ecolocalización producidos en principio por delfines mulares (*Tursiops truncatus*) que se avistan frecuentemente en las proximidades de la plataforma [1], aunque no se puede descartar la presencia de vocalizaciones de otros cetáceos avistados en ocasiones dentro de la Ría de Arousa [2].

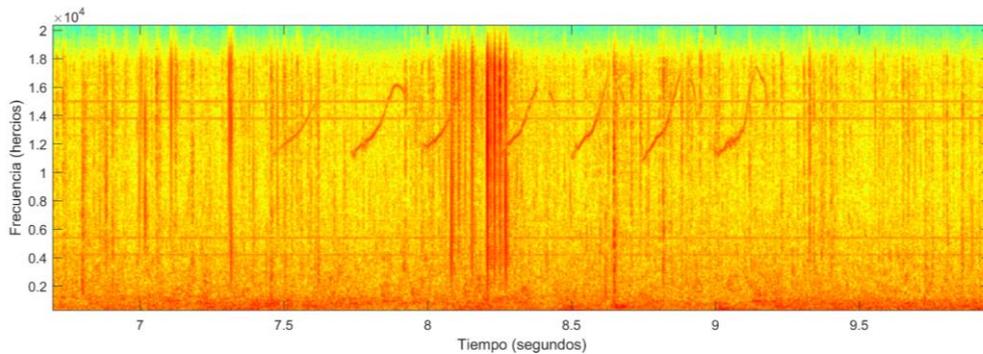


Fig. 3: Espectrograma con sonidos de cetáceos

## 2.2. Contaminación por ruidos de la plataforma.

Entendemos como tales todos los sonidos no presentes en el ambiente sonoro, sino producidos por la introducción en el medio marino de la plataforma y sus equipos, incluido el propio hidrófono. Hemos identificado los siguientes ruidos de plataforma:

- El hidrófono está inicialmente suspendido de la plataforma compartiendo vertical con otros equipos, lo cual provoca ruido de fricción entre la cubierta del hidrófono y los elementos de suspensión y los cables del resto de equipos. Estos sonidos pueden confundirse con vocalizaciones de cetáceos.
- Las olas producen movimientos verticales, flexiones y torsiones de la plataforma, lo que se traduce en un desplazamiento vertical del hidrófono dentro del agua, que produce un intenso ruido de flujo de muy baja frecuencia (por debajo de 1 Hz). Este ruido puede comprometer el rango dinámico de la grabación, de forma que los sonidos de interés queden muy atenuados.
- Las sondas CTD instaladas en la plataforma se comunican con su sistema de adquisición de datos por medio de un módem inductivo, que emplea una secuencia de tonos que contaminan la grabación del hidrófono. La figura 4 muestra la presencia de

estos tonos en un espectrograma, donde puede observarse la interrupción de los mismos para escuchar a los equipos con los que se conecta. Estos tonos falsean los valores de los indicadores de ruido ambiental, y dificultan la detección automática de vocalizaciones de cetáceos.

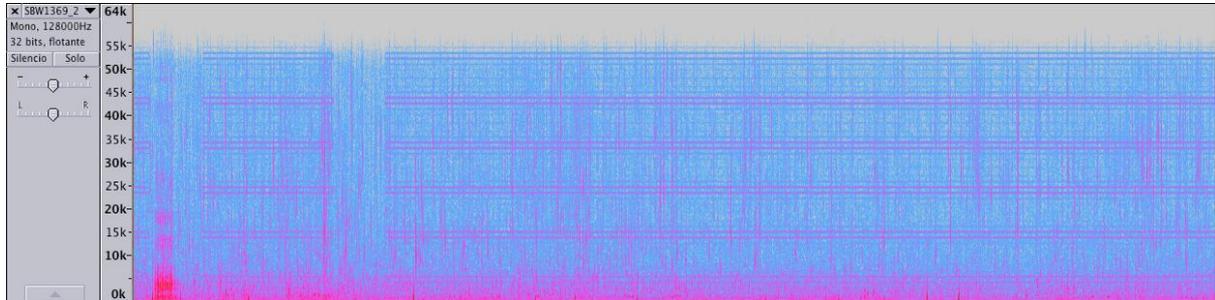


Fig. 4. Espectrograma con los tonos del módem inductivo y ruidos de impacto de sedimentos

- La plataforma está situada en una zona con fuertes corrientes de marea, que pueden superar los 3 m/s; esto provoca que aparezca en las grabaciones un elevado ruido de flujo de corriente, así como ruidos impulsivos provocados por el impacto de los sedimentos arrastrados por la corriente contra el hidrófono (aparecen como líneas verticales en la figura 4).
- El ruido de flujo de corriente falsea los indicadores de ruido ambiental, especialmente en los periodos sin ruido de embarcaciones, y los impactos se confunden con clics de ecolocalización de cetáceos. El efecto de la corriente de marea puede observarse en la figura 1: fijándose en la envolvente inferior de la curva (la que uniría los valores mínimos), se aprecian los cuatro ciclos de la marea correspondientes a las 24 horas de la grabación.
- El hidrófono capta también otros sonidos producidos por la plataforma, por ejemplo el ruido de la cadena de fondeo.

### 2.3. Mejora el sistema de fondeo.

Con el objetivo de minimizar los ruidos contaminantes, con fecha 3 de agosto de 2021 se han introducido las siguientes mejoras en la instalación del hidrófono:

1. Suspender el hidrófono en una nueva vertical separada del resto de equipos, para evitar el ruido provocado por el rozamiento del hidrófono con los elementos de suspensión y cables de comunicación de los demás equipos. La ubicación propuesta para el fondeo del hidrófono está junto a la pared norte de la segunda caseta (cruz verde en la figura 5). Se ha escogido de manera que ampliando solo 10 m el cable del hidrófono, lo separe lo más posible de las fuentes de ruido: los otros equipos que hay en el agua (en la vertical situada en la cruz roja de la figura 5), los costados de la batea donde amarran los barcos, y los flotadores donde golpean las olas. La ubicación elegida también facilita practicar una nueva trampilla en el suelo de la plataforma, porque está entre dos travesaños con orientación NE-SW de manera que se puedan cortar las tablas sin necesidad de poner refuerzos transversales.



Fig. 5: ubicación propuesta para el nuevo fondeo del hidrófono

2. Añadir un amortiguador en la línea de fondeo. Dado que el hidrófono está colgando de la plataforma, los movimientos de ésta se transmiten por la línea del fondeo y provocan ruido de flujo de agua en la grabación. Para evitarlo, lo ideal sería utilizar una boya sumergida, que mantuviera el hidrófono en posición vertical y a una altura constante sobre el fondo de la ría, en lugar de colgando de la plataforma. El problema es que eso obligaría a dejar en banda tantos metros del cable de señal como carrera de marea hay en el sitio, lo cual entrañaría un elevado riesgo de que se enganchen elementos arrastrados por la corriente que arranquen el cable. La alternativa finalmente implementada es dejar colgando el hidrófono con un muerto al final de la línea que no llega a tocar el fondo, de forma similar a como están instalados los demás equipos, pero incorporando un muelle para minimizar el movimiento que la plataforma comunica al hidrófono (ver Fig. 6). Una solución parecida a esta es la que se recomienda para despliegue de hidrófonos desde barco en la Guía de buenas prácticas de medidas submarinas del NPL [3, figura 4.1].



Fig. 6: nueva línea de fondeo del hidrófono

3. Añadir una pantalla anti-flujo. Para reducir el ruido del flujo de agua, y los impactos de los sedimentos que arrastra el agua, se ha instalado un calcetín de espuma sobre el hidrófono a modo de pantalla.



Fig. 7: pantalla anti-flujo instalada sobre el hidrófono

Y con fecha 15 de septiembre de 2021:

4. Reducir la contaminación tonal. Se ha configurado el modem inductivo que se comunica con los CTD instalados en la plataforma en modo baja potencia. Tras este cambio, que no afecta a la calidad de la comunicación, desaparece casi totalmente la contaminación por tonos que presentaban las grabaciones del hidrófono.

Por otro lado, se ha decidido configurar el hidrófono para que grabe un minuto de sonido de cada 3, con una frecuencia de muestreo de 128.000 muestras por segundo, lo que proporciona un ancho de banda de 51,2 kHz. Este ciclo de grabación es suficiente para caracterizar el ruido ambiental de la zona, y el ancho de banda recoge los silbidos de los cetáceos que la frecuentan, todo ello sin producir una cantidad de datos excesivamente grande. Este aspecto es importante ya que las grabaciones deben almacenarse localmente, y su descarga periódica a los servidores en tierra firme requiere desplazar personal a la plataforma.

### 3. Cálculo de indicadores de ruido ambiental.

Uno de los objetivos fundamentales de instalar un hidrófono en la Plataforma de Cortegada es determinar el estado medioambiental de la zona en relación al ruido antropogénico. La Directiva Marco 2008/56/EC, de Estrategias Marinas (DMEM) [4] establece dos criterios –con sus indicadores asociados– para establecer un descriptor de Buen Estado Ambiental de las aguas europeas respecto al ruido submarino:

Criterio 11.1 – Distribución en el tiempo y el espacio de impulsos sonoros elevados, de frecuencia media y baja. Este criterio está relacionado con actividades como el uso de explosivos, construcción marina, exploración sísmica, etc.

Criterio 11.2 – Sonido continuo de baja frecuencia. Este criterio está relacionado principalmente con la actividad del tráfico marítimo.

Dado que, como ya se ha visto, la mayor parte del ruido antropogénico en la zona se debe al tráfico marítimo, a partir de las grabaciones del hidrófono se implementa el cálculo de los indicadores correspondientes a este segundo criterio.

#### 3.1. Descripción de indicadores.

El indicador sugerido por la DMEM para cuantificar el ruido de tráfico marítimo se define como [4] **Indicador 11.2.1:** *Tendencias en el promedio anual de la presión acústica al cuadrado asociada con el ruido ambiental en cada una de las dos bandas de tercio de octava, una*

centrada a 63 Hz y la otra a 125 Hz, expresada como un nivel en decibelios, en unidades de dB re 1  $\mu$ Pa, ya sea medido directamente en las estaciones de observación, o inferido de un modelo utilizado para interpolar o extrapolar a partir de mediciones en estaciones de observación.

Para dar satisfacción a este indicador es necesario calcular el nivel de presión sonora (SPL por sus siglas en inglés) en las bandas de tercio de octava centradas en 63 y en 125 Hz. Para completar la descripción del nivel de ruido de la zona, y siguiendo las tendencias más recientes, añadiremos a estos descriptores el SPL en la bandas de tercio de octava de 2 y 5 kHz, y el SPL total en la banda de 20 Hz a 20 kHz. Estos cinco indicadores, que denominaremos SPL63, SPL125, SPL2k, SPL5k y SPLt, se calculan cada 20 segundos, de acuerdo con la temporización del portal de datos de EMODnet [5] donde se comparten en tiempo real.

### 3.2. Método de cálculo.

El cálculo de los indicadores descritos en el apartado anterior se realiza mediante el procesado de los audios de 1 minuto, que cada 3 minutos proporciona el hidrófono sito en la plataforma de Cortegada. Cada uno de estos archivos es troceado en tramas de 1s mediante inventariado Hanning para, posteriormente, calcular la densidad espectral de potencia (PSD) de cada una de ellas a través de transformadas rápidas de Fourier o FFTs. De la integral, en cada una de las bandas de interés, de la densidad espectral de potencia se obtienen los niveles de presión sonora (SPL) de cada uno de los indicadores descritos en el apartado anterior.

Dado que la resolución temporal de resultados es de 20 s y siguiendo las recomendaciones de la DMEM, se calcula el nivel SPL promediando aritméticamente a lo largo del tiempo. Así, se promedian los niveles obtenidos, en unidades de potencia, en grupos de 20 tramas de 1s. Este promediado es conservador, pues favorece los niveles más altos de cada promediado, precisamente por lo cual es el recomendado por la normativa pues permite establecer un nivel seguro que no será superado la mayor parte del tiempo.

Los resultados de este cálculo, al igual que ocurría antes, se almacenan automáticamente en un archivo csv generado por el código que corre en la plataforma.

Los datos disponibles desde el comienzo del proyecto hasta la fecha de la corrección de los tonos (apartado 2.3 4) han sido corregidos mediante sustracción espectral y se entregan en un archivo csv "niveles1120\_0821.csv". Estos tonos provocaban diferencias de hasta 12 dB en el SPLt.

### 3.3. Código.

Para la obtención de los indicadores de SPL descritos el código python se vale de 2 funciones: *ReadWav(fn)* y *SPL\_det(data,fs,fn,qbarc)* que detallamos a continuación:

- *ReadWav(fn)*:

Entradas

-----

*fn* : str

Ruta al fichero

Salidas

-----

*data* : Array

Muestras del archivo de audio.

`fs : int`

Frecuencia de muestreo del audio.

Esta función es la encargada de leer el archivo de audio obtenido del hidrófono devolviendo un vector con las muestras correspondientes y una variable con la frecuencia de muestreo.

- `SPL_det(data,fs,fn,qbarc):`

Entradas

-----

`data : array`

Muestras del audio a procesar.

`fs : int`

Frecuencia de muestreo.

`fn : str`

Nombre del archivo a procesar.

`qbarc : mp.Queue`

Cola de comunicación de datos con el proceso principal.

Salidas

-----

A través de `qbarc`:

Niveles de `SPLt`, `SPL63`, `SPL125`, `SPL2k`, `SPL5k` cada `TimeInterval` seg

Presencia de barco cada `TimeInterval` seg; 1->D. Positiva 0->D.

Negativa

Esta función realiza la tarea combinada del cálculo de los indicadores de presión y de la detección de barco (detallada en el siguiente apartado) ya que, tal y como se detallará en el apartado 4, la detección de barcos parte de los indicadores para realizar su trabajo.

Para el cálculo de los indicadores, recibe las muestras de 1 minuto de audio a procesar (*data*), su frecuencia de muestreo (*fs*) y el nombre del archivo a procesar (*fn*) para realizar el procesado necesario y devolver los indicadores *SPLt*, *SPL63*, *SPL125*, *SPL2k* y *SPL5k* a través de la variable de cola *qbarc* que se utiliza para la comunicación con procesos en paralelo que será detallada en el apartado 6. El procesado depende de la variable *TimeInterval*, definida en la cabecera del archivo, y la cual determina cada cuanto tiempo se promedian las ventanas de 1 segundo dentro del archivo de audio de 1 minuto. En nuestro caso, como ya hemos explicado en el apartado anterior, obtendremos los indicadores cada 20s, por lo tanto se configura `TimeInterval=20`.

## 4. Detección automática de barcos.

En este apartado se describirán las acciones realizadas para la obtención de un detector automático de barcos a través de las grabaciones del hidrófono. Como se verá a continuación, el presente detector aprovecha los históricos de los indicadores de nivel de presión sonora (SPL) para entrenar un clasificador automático mediante Machine Learning que posteriormente se utiliza para la anotación en tiempo real de la presencia/ausencia de embarcaciones.

## 4.1. Estado del arte.

Los más recientes trabajos que tratan el tema de la detección y clasificación de embarcaciones a partir del sonido submarino suelen abordar el problema desde un punto de vista de Deep Learning mediante el uso de redes neuronales [6]. Son métodos que pueden proporcionar resultados muy acertados, pero a costa de requerir elevadas cargas computacionales y un entrenamiento costoso con una base de datos muy extensa. Otros métodos que permiten disminuir la carga computacional sin afectar considerablemente a la eficiencia utilizan clasificadores más sencillos, basados en machine learning o incluso en aplicar umbrales a ciertas características del sonido.

En cualquiera de los casos descritos, la aproximación al problema es la siguiente: obtener las características que permitan discriminar lo mejor posible la presencia de ruido de tráfico en las grabaciones, y entrenar un *clasificador* que aprenda a separar estas dos clases a partir de dichas características. Una vez entrenado el clasificador, se valida éste contra una base de datos –no contenida en el entrenamiento– para testar su eficacia. Una vez validada la eficacia del modelo, se pone en funcionamiento con los nuevos registros.

Los clasificadores obtenidos tratan de separar las distintas clases mediante diferentes técnicas que pueden ser desde umbrales unidimensionales, hiperplanos hasta complejas redes neuronales.

## 4.2. Método de detección.

Se desarrolla un método de machine learning mediante el cual se entrenan varios clasificadores con una base de datos de entrenamiento (10000 registros etiquetados de tramas de 20s de audio) y posteriormente se valida contra una base de datos de validación (3380 registros).

Los datos se obtienen del periodo entre el 15 y 29 de Septiembre de 2021 dado que los datos anteriores sufrían de los efectos perjudiciales del anterior fondeo –tal y como se detalló en el apartado 2.3– aunque estos datos previos al nuevo fondeo fueron de gran utilidad en el desarrollo de la metodología. La base de datos está desbalanceada en cuanto al número de registros por clase dado que las apariciones de barcos son mucho menos frecuentes que sus ausencias.

El algoritmo se implementa en tres fases:

- **Selección de características.** Primeramente, las bases de datos son procesadas para obtener características definitorias del evento a detectar en el audio. Se definen a continuación las características propuestas:
  - SPLt, SPL63, SPL125, SPL2k, SPL5k: Nivel de presión sonora total y en las bandas de 63 Hz, 125 Hz, 2 kHz y 5 kHz
  - E1, E2: SPL en las bandas de 1 a 5 kHz y de 5 a 10 kHz.

Los indicadores, SPL63 y SPL125, son los recomendados por la DMEM para la definición de ruido de barco y resultan de gran eficacia en entornos con bajo ruido de fondo en baja frecuencia. Dado que, como se ha explicado en el apartado 2, el entorno de Cortegada presenta bastante ruido en estas bandas de frecuencia, se decide añadir los indicadores SPLt, SPL2k y SPL5k para incluir información del espectro global y de otras bandas significativas de la

emisión de barcos. Las características E1 y E2 resultan de utilidad a la hora de separar eventos de espectro amplio como impactos de sedimentos que arrastra la corriente.

- **Entrenamiento.** Las características de cada registro se incluyen en las bases de datos etiquetadas y se testan los clasificadores. Para este proceso de entrenamiento se utiliza la técnica de *folding* la cual separa la base de datos de entrenamiento en subconjuntos aleatorios de entrenamiento/validación y entrena y valida cada uno de los clasificadores en estos subconjuntos, obteniéndose la eficacia del clasificador como la media de la eficacia en cada uno de los subconjuntos. Este método permite evaluar un gran número de clasificadores explorando exhaustivamente la base de datos de entrenamiento y manteniendo intacta la base de datos de validación. Tras este proceso se selecciona por su eficacia el clasificador *RUSBoosted Trees* [7], un método bastante popular debido a su baja complejidad y notable eficiencia, en términos de resultados y tiempo de procesado, además de mejorar resultados en los casos de las bases de datos desbalanceadas como el que nos atañe.
- **Validación.** El clasificador seleccionado se testa contra la base de datos de validación para evaluar su eficacia. El clasificador es exportado en un formato ligero que se incluye en el código python de la aplicación y que se encargará de decidir la clase del segmento de audio en función de las características solicitadas del mismo.

### 4.3. Código.

La metodología descrita se implementa en el código a través de la función `SPL_det(data,fs,fn,qbarc)`:

Entradas

-----

`data` : array

Muestras del audio a procesar.

`fs` : int

Frecuencia de muestreo.

`fn` : str

Nombre del archivo a procesar.

`qbarc` : mp.Queue

Cola de comunicación de datos con el proceso principal.

Salidas

-----

A través de `qbarc`:

Niveles de SPLt, SPL63, SPL125, SPL2k, SPL5k cada `TimeInterval` seg

Presencia de barco cada `TimeInterval` seg; 1->D. Positiva 0->D.

Negativa

Esta función recibe la muestras de audio a analizar (`data`) obtiene la características propuestas y predice su clase mediante la función `clf.predict`, la cual devuelve un '1' en caso de predecir barco y '0' en caso contrario. El objeto `clf` representa el clasificador que se desarrolló en las fases previas y que es cargado en la cabecera del código en la zona de configuración del detector de barcos.

#### 4.4. Resultados.

La siguiente tabla muestra los principales resultados obtenidos con el clasificador desarrollado tras confrontarlo a la base de datos de validación conformada por 131 etiquetas de *barco*('1') y 3249 de *no barco*('0'):

		Predicción	
		0	1
Realidad	0	95% 3209	4.8% 157
	1	9% 12	90% 119

Tabla 1. Tabla de verdad con los resultados del clasificador propuesto

Tal y como podemos ver en la tabla 1 el clasificador detecta correctamente la ausencia de barco en un 95% de los casos mientras que produce cerca de un 5% de falso positivos; de la misma manera se obtiene un 90% de detección correcta de barcos y un 9% de falsos negativos.

**Los falsos positivos** vienen dados en su mayoría por la influencia del ruido habitual de la cadena de fondeo de la plataforma y/o ruidos impulsivos similares como pueden ser impactos ocasionales en el hidrófono.

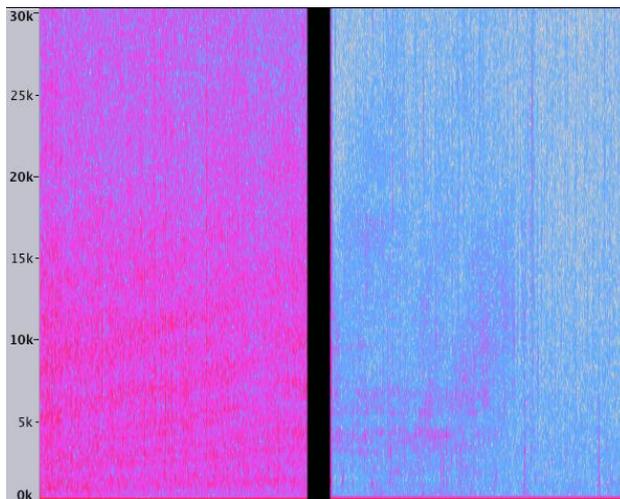


Fig. 8 Espectrograma de barco cercano (izquierda) frente a un barco lejano (derecha)

**Los falsos negativos** tienen que ver con el fenómeno del overfitting o sobreajuste que se produce cuando una base de datos está descompensada en el balance de sus clases y el clasificador tiende a sobreajustarse a los datos de la clase mayoritaria para obtener una mayor eficacia. También ha de tenerse cuenta en estos casos la etiquetación del evento barco pues, tal y como se ve en la figura 8, las diferencias energéticas (mayor cantidad de rojo implica mayor energía) pueden ser considerables entre un evento cercano y uno lejano lo cual hace patente el límite de correcto funcionamiento del detector en función de la distancia de la embarcación.

La base de datos de entrenamiento no ha podido ampliarse debido a lo tardío de la fecha en que se introdujeron las mejoras de la instalación de fondeo. Sin duda, este trabajo futuro contribuirá positivamente a la mejora en la detección.

## 5. Detección automática de cetáceos.

En esta sección se describen las tareas realizadas con el objetivo de obtener un detector automático de presencia de cetáceos a partir de las grabaciones del hidrófono. Como se describe a continuación, el detector implementado está basado en la identificación de los silbidos producidos por los cetáceos, siendo capaz de identificar silbidos entre 2 y 20 kHz. Estas frecuencias son especialmente adecuadas para detectar vocalizaciones de delfines y calderones.

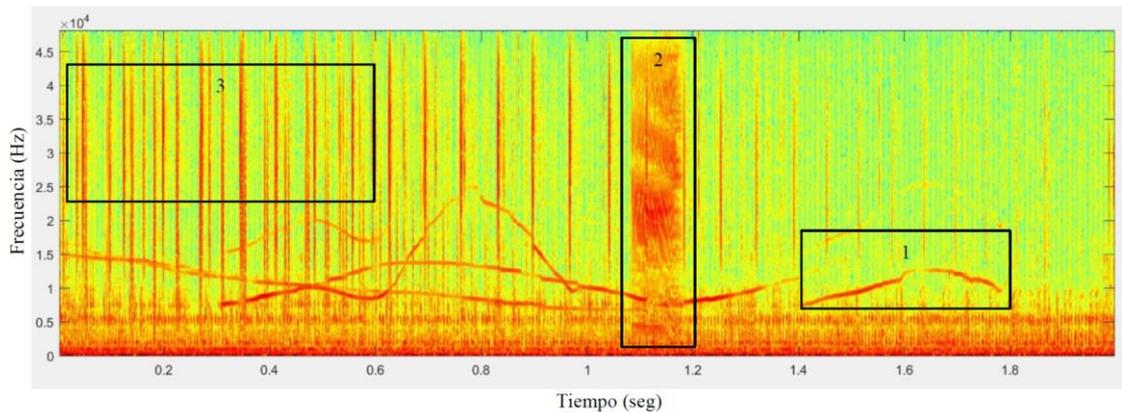


Fig. 9: Ejemplos de vocalizaciones de cetáceos: silbidos (1), bursts (2) y clics (3)

### 5.1. Estado del arte.

Los cetáceos emiten básicamente tres tipos de vocalizaciones (figura 9):

- Pulsos de ecolocalización (clics), utilizados para detección de presas y navegación para el entorno (figura 9, 3).
- Silbidos, empleados para tareas de comunicación (figura 9, 1).
- Trenes de clics, también denominados bursts, empleados principalmente para alimentación (figura 9, 2).

Los clics son sonidos impulsivos, de corta duración y espectro ancho, fácilmente identificables en condiciones de bajo ruido ambiental. Los silbidos, por su parte, son sonidos tonales, con un rango de frecuencias diferente para cada especie y con modulaciones de frecuencia de hasta varios kilohercios. Los burst son trenes de clics continuos.

Los algoritmos clásicos de detección están típicamente basados en la identificación de clics ya que gracias a su naturaleza impulsional son fáciles de extraer utilizando algoritmos sencillos [8]. En este proyecto se planteó como idea de partida utilizar esta metodología. Sin embargo, la contaminación de las grabaciones que se describe en el apartado 2.1 desaconsejó este acercamiento. Se observó que las grabaciones contenían un nivel muy alto de ruido impulsional (golpes, chasquidos, etc), que con toda seguridad confundirían al detector. Se optó en su lugar, por implementar un detector menos habitual, basado en identificación de silbidos. Como explicaremos más adelante, el detector de silbidos también se ve afectado por la contaminación de las grabaciones.

## 5.2. Método de detección.

El algoritmo de detección se basa en el denominado índice de tonalidad (TI). El índice de tonalidad es una medida que se desarrolló inicialmente para codificación de audio con el objetivo de separar componentes tonales y no tonales. El índice de tonalidad mide la predictibilidad de una componente espectral devolviendo un índice entre cero y uno. Un valor cercano a cero indica que la componente es muy tonal, mientras que un valor cercano a uno indica que la componente es no tonal (ruidosa).

Para su aplicación en el detector de cetáceos, el índice de tonalidad se calcula para todas las componentes espectrales de interés cada 0.5 ms. Cada 20 segundos, se compara el vector de índices obtenido con un umbral predeterminado. Si alguno de los valores está por debajo del umbral, se decide que hay una componente tonal en el segmento, probablemente perteneciente a una vocalización de un cetáceo.

El algoritmo se implementa en tres fases:

- **Generación del espectrograma** - Se parte de segmentos de 20 segundos de duración que se diezman usando un factor de 3 para reducir el ancho de banda y se separan en tramas de 5 ms solapadas el 90%. Se obtiene el espectro de cada trama empleando una FFT (Fast Fourier Transform).
- **Cálculo del índice de tonalidad** - A partir del espectrograma anterior, se seleccionan las componentes espectrales de interés (2 a 20 kHz en la implementación actual). Se calcula la predictibilidad de cada componente a partir de las componentes espectrales de las dos tramas anteriores. Se obtiene así una matriz de índices de tonalidad con dimensiones tiempo por frecuencia. La matriz resultante se normaliza usando la media y varianza calculadas en la dimensión temporal.
- **Promediado de la matriz** - En este punto la matriz obtenida contiene un vector de muestras espectrales cada 0.5 ms. Esta precisión es necesaria para seguir las variaciones rápidas de los silbidos de los cetáceos, pero para obtener una única decisión cada 20 segundos es posible y conveniente reducir las dimensiones. Para esto se promedian bloques de 20x20 puntos de la matriz y se calcula el mínimo en el eje frecuencial. Se obtiene así un índice de tonalidad cada 10 ms, es decir, un vector de 2000 valores para un segmento de 20 segundos.
- **Decisión** - Finalmente se comprueba si los valores del vector son inferiores a un determinado umbral. En el algoritmo implementado es posible indicar el número de veces que el vector debe descender por debajo del umbral para tomar una decisión. Si se cumplen ambas condiciones, es decir, si desciende por debajo del umbral el número indicado de veces, se decide que el segmento contiene una componente muy tonal, presumiblemente un silbido de cetáceo.

## 5.3. Código.

El algoritmo se ejecuta en la subrutina del fichero python *Deteccion(datain,fsin,i,fn)*:

Entradas

-----

`datain : array`

```
vector 20 segundos de muestras de audio a la frecuencia original
fsin : int
    Frecuencia de muestreo.
i : int
    Índice del bloque (hay tres bloques de 20 segundos por cada
    fichero de un minuto).
fn : str
    Nombre del archivo de audio.
Salidas
-----
det : int
    Resultado de la detección, 1->D. positiva 0->D. negativa.
```

Los dos últimos parámetros de entrada se emplean tan solo para generar mensajes informativos.

Los parámetros de configuración se encuentran en la cabecera del fichero y son los siguientes:

- *umbral* y *Numbral* - Umbral de detección y número de veces que debe ser superado para una detección positiva. Están fijados a -5 y 1 respectivamente. Incrementando los valores de *umbral* o de *Numbral* la detección será menos sensible, es decir, con menos probabilidad de falsas alarmas, pero con mayor probabilidad de silbidos no detectados.
- *fmin* y *fmax* - Frecuencias mínimas y máximas de detección. Fijadas a 2kHz y 20 kHz respectivamente. El rango es óptimo para la detección de delfines y calderones, pero debería ser ampliado en su rango inferior para la detección de cetáceos de mayor tamaño (ballenas) y en su rango superior para la detección de marsopas.
- *decfactor* - factor de diezmado, fijado a un valor de 3, de manera que la frecuencia de muestreo real empleada es de  $128000/3=42.2667$  kHz. Si se modifica la frecuencia de muestreo del hidrófono, es necesario cambiar este factor para mantener un ancho de banda compatible con el rango de detección fijado por *fmin* y *fmax*. Debe cumplirse que  $fs/2 > fmax$ .
- *Winlen*, *Overlap* y *OverSamp* - Longitud en segundos de la ventana de análisis, solapamiento entre ventanas y número de puntos de la FFT usada en el espectrograma. Reduciendo *Winlen*, u *Overlap* se puede disminuir el tiempo de cálculo del algoritmo aunque degradando la precisión de la detección. Una modificación de estos parámetros requiere reajustar el umbral de detección, por lo que no es conveniente su modificación. Los valores por defecto son 5 ms para *Winlen*, 0.9 para *Overlap* (solapamiento del 90%) y 1 para *OverSamp* (no se realiza ningún sobremuestreo espectral).

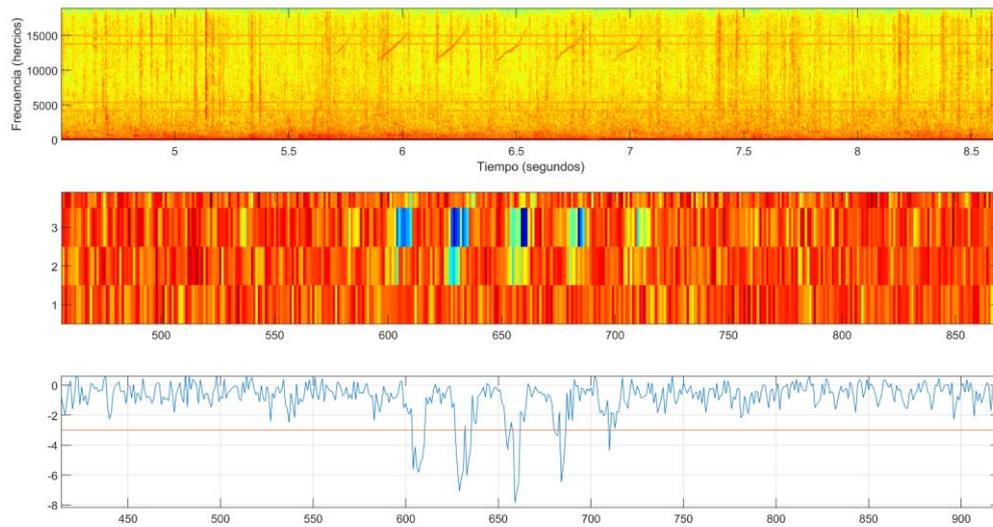


Fig. 10: Ejemplo de funcionamiento del detector de silbidos. Gráfica superior: espectrograma del segmento de audio conteniendo una secuencia de silbidos. Gráfica intermedia: matriz del índice de tonalidad tras la reducción de dimensionalidad. Gráfica inferior: vector de índices mínimos y umbral de detección.

#### 5.4. Resultados.

Los experimentos realizados sobre la base de datos de grabaciones disponible durante el desarrollo del proyecto, llevaron a las siguientes conclusiones:

- El detector basado en el índice de tonalidad funciona correctamente. En la figura 10 se muestra un ejemplo de detección de un segmento de grabación. Se puede observar que el vector de índices desciende bajo el umbral (gráfica inferior), por lo que el silbido que aparece en la grabación se detecta correctamente.
- Los silbidos a frecuencias cercanas al límite inferior de análisis, se detectan con más dificultad, debido al ruido de baja frecuencia.
- Los ruidos tonales generados por el modem (comentados anteriormente) no influyen en la detección. Esto se debe a que los tonos contienen una modulación en frecuencia con variaciones muy rápidas en comparación con el tamaño de trama usado, de manera que no son detectados como tonos estables y predecibles.
- Sin embargo, en los periodos en que el modem deja de transmitir o en que inicia la conexión, se emiten tonos estables sin modular que engañan al detector. Esto sucede de manera periódica produciendo un nivel muy alto de falsas alarmas.

Este último efecto puede observarse en la figura 11. Los períodos en que el modem emite un tono estable (presuntamente antes de negociar la conexión), producen detecciones erróneas que se repiten periódicamente. Los resultados para esta base de datos contenían por tanto un número muy alto de falsas alarmas (detecciones erróneas). Viendo estos resultados, se abandonaron los experimentos con la base de datos de partida, a la espera de trabajar con grabaciones obtenidas con el fondeo mejorado.

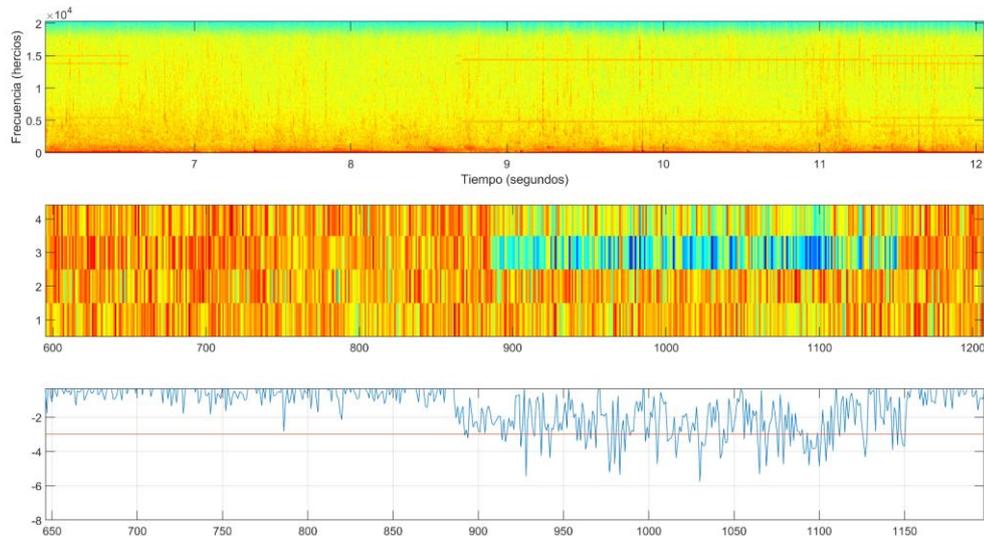


Fig. 11: Ejemplo de detección errónea debida a los tonos interferentes del modem. Obsérvese que el índice de tonalidad detecta solo las zonas en que los tonos son estables.

Una vez disponibles estas grabaciones, se realizó un nuevo experimento que describimos a continuación. Los resultados que se presentan corresponden a grabaciones realizadas entre el 15 y el 17 de septiembre de 2021 y capturadas con el fondeo mejorado. Hay que indicar que al disponer de esta parte de la base de datos al final del proyecto, no fue posible realizar un etiquetado detallado. Los resultados deben ser entendidos como aproximaciones.

Se utilizaron 1010 ficheros de un minuto de duración, correspondientes a cerca de 17 horas de grabación. En total se clasificaron 3030 segmentos de 20 segundos. Tras un etiquetado no exhaustivo, se encontraron y marcaron 35 segmentos que contenían silbidos de diferentes intensidades y duraciones. Para realizar el etiquetado se partió del propio detector, analizando las grabaciones que presentaban los índices de tonalidad de menor valor. Los resultados se muestran para diferentes umbrales en la tabla 2. Las columnas TP (true positive) y TN (true negative) indican el número de segmentos conteniendo silbidos (TP) y no conteniendo silbidos (TN) que han sido correctamente detectados. La columna FP (false positive) indica el número de segmentos identificados erróneamente como segmentos que contienen silbidos, mientras que FN (false negative), indica el número de segmentos etiquetados erróneamente como no segmentos que no contienen silbidos.

Umbral	TP	TN	FP	FN
-3	6	2993	2	29
-2.2	22	2983	12	13
-2.1	24	2976	19	11

Tabla 2. Resultados para el experimento sobre las grabaciones en el fondeo mejorado. Umbral - umbral de detección empleado. TP - segmentos con sonidos de cetáceos correctamente detectados (True Positive). TN - segmentos sin sonidos de cetáceos correctamente detectados (True Negative). FP - Segmentos incorrectamente identificados como positivos (False Positive). FN - Segmentos incorrectamente detectados como negativos (False Negative).

Se puede observar que el umbral de -3 que se ajustó con la base de datos inicial no es el más adecuado en el nuevo fondeo, produciendo una tasa muy alta de falsos negativos. Sin embargo, la desaparición de los tonos del modem reduce de manera drástica la tasa de falsos positivos. Se recomienda por tanto fijar el umbral a -2.2. Un umbral de -2.1 produce un nivel de falsos positivos demasiado alto.

De todas formas, y recalcando las observaciones sobre la validez del experimento explicadas anteriormente, del análisis detallado de los resultados pueden extraerse las siguientes conclusiones:

- La mayor parte de los falsos positivos (estimamos que más del 50%) son debidos a ruidos producidos por barcos y por golpes metálicos en el nuevo fondeo. Los falsos positivos debidos a los barcos pueden ser reducidos de manera drástica combinando la salida del detector de barcos con el detector de cetáceos. El resto de los falsos positivos se pueden reducir reacondicionando el fondeo.
- Gran parte de los falsos negativos se deben a silbidos débiles o muy cortos, en cualquier caso difíciles de distinguir en el espectrograma. Dependiendo de la aplicación deseada para el detector, el umbral de -3 puede ser el más indicado, detectando los silbidos más claros y limitando de forma drástica el número de detecciones positivas erróneas.

## 6. Mejoras en el código para su ejecución in situ.

La plataforma de Cortegada dispone de una RaspBerry Pi desde la cual se ejecuta un código en Python que calcula los niveles de SPL, genera informes y descarga los archivos del hidrófono. Se propone añadir los códigos de detección a este código así como optimizar los tiempos de ejecución –y por tanto el consumo energético– y añadir la funcionalidad de exportación de los audios en formato mp3 para estudiar la posibilidad del envío al servidor en tiempo real. A continuación detallamos las tareas realizadas.

### 6.1. Optimización del tiempo de ejecución.

En el caso que nos implica, la optimización del tiempo de ejecución puede ser de gran importancia al estar estrechamente relacionado con el tiempo que la RaspBerry Pi necesita estar encendida para realizar el trabajo y por lo tanto con el consumo energético de la plataforma. A continuación se resumen las acciones realizadas sobre el código inicial con este objetivo:

- Portabilidad a Python 3. El código actual trabaja bajo python3 lo cual supone una mejora no sólo de prestaciones sino también un leve mejora (1.2–1.3x) en los tiempos de ejecución.
- Lectura y procesado de niveles SPL. El método utilizado para la lectura de archivos de audio (librería *fsread*) mejora la eficacia de lectura de los audios y la obtención de los datos en un formato óptimo para su posterior procesado. Se simplifica el código de obtención de niveles gracias, en parte, al nuevo método de lectura, y se vectoriza todo el procesado de FFTs obteniéndose en conjunto mejoras de hasta 4x.

- Procesado en paralelo. El análisis de los procesos implicados en el código muestra que el proceso “cuello de botella” es la descarga ftp de los audios del hidrófono, tardando alrededor de 40s por archivo. Por lo tanto, en lugar de mantener la estrategia de descargar todo los audios y luego procesar en paralelo se propone descargar y una vez tengamos el primer audio comenzar a procesar en paralelo a la descarga de los siguientes.

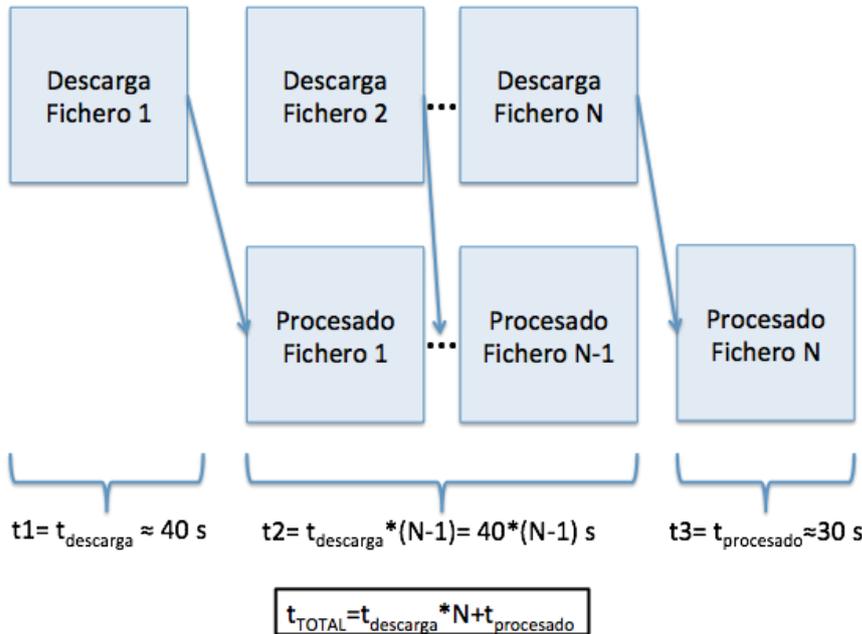


Fig. 12: Esquema propuesto de procesado en paralelo de N archivos de audio

Tal y como podemos ver en la figura 12 el esquema propuesto obtiene un tiempo total de ejecución  $T = t_{descarga} * N + t_{procesado}$  frente al esquema anterior que obtenía un tiempo de  $T_{old} = N * t_{descarga} + (N/4) * t_{procesado\_old}$ . Se mejoran notablemente los tiempos y por tanto el consumo energético; un ejemplo:

- Suponiendo  $N=12$  y  $t_{procesado} \approx t_{procesado\_old} = 30$
- $T = 40 * 12 + 30 = 510 \text{ s}$      $T_{old} = 40 * 12 + (12/4) * 30 = 570 \text{ s}$

El procesado de cada fichero se realiza en paralelo, utilizando un núcleo de la RaspBerri Pi para el cálculo del SPL y la detección de barcos, otro núcleo para la detección de cetáceos y otro para la codificación mp3. De esta forma, a pesar de incluir mejoras sustanciales en el código (detectores, codificación mp3), se consigue optimizar el nuevo procesado incluso ligeramente por debajo de los tiempos del procesado previo.

## 6.2. Incorporación de un codificador de audio.

Los archivos de audio del hidrófono son demasiado pesados para transmitirse a través del enlace de comunicaciones de la plataforma (enlace móvil con tecnología 3G). Sin embargo, una versión reducida mediante un codificador mp3 podría permitir realizar esta transmisión.

Se programa un codificador mp3 que realiza las tareas necesarias para el objetivo planteado a través de la función `compMP3(data,fs,file_name,qmp3)`:

Entradas

```
-----  
data : array  
    Muestras del archivo wav.  
fs : int  
    Frecuencia de muestreo del wav.  
file_name : str  
    Nombre del archivo wav.  
qmp3 : mp.Queue  
    Cola de comunicación de datos con el proceso principal
```

La función *compMP3* recibe un vector de datos de entrada que normaliza, comprime y separa en archivos cada 20 segundos (según la trama de análisis configurada en *TimeInterval*), generando así 3 archivos mp3 por audio analizado listos para transmitirse y reproducirse.

La función está programada para exportar las 3 tramas del audio analizado para probar la capacidad de transmisión del enlace de Cortegada aunque podría fácilmente ser optimizada para sólo mantener las tramas en las que se haya producido algún tipo de detección.

## 7. Conclusiones.

A lo largo del proyecto se ha realizado un estudio en profundidad del sonido submarino captado por el hidrófono de Cortegada. El estudio de las componentes contaminantes provocadas por la plataforma ha permitido implementar mejoras muy importantes en la instalación del hidrófono, que han reducido en gran medida o eliminado totalmente estos sonidos contaminantes. Del estudio del ambiente sonoro se desprende la presencia predominante de ruido de tráfico marítimo, y vocalizaciones de cetáceos.

Además del cálculo de los indicadores de ruido submarino de la DMEM, se ha desarrollado un detector automático de barcos y un detector automático de vocalizaciones de cetáceos, con elevadas tasas de acierto. Son susceptibles de mejora cuando se disponga de más grabaciones.

Tanto el cálculo de indicadores de ruido como los detectores de barcos y cetáceos se han codificado en lenguaje Python y funcionan en tiempo real en el procesador situado en la plataforma. A mayores se ha añadido un codificador MP3 que permite enviar a tierra firme y en tiempo real una versión comprimida de las grabaciones, todo ello sin incrementar el consumo de energía.

## REFERENCIAS

- [1] Pierce, Graham J., et al. "Trends in cetacean sightings along the Galician coast, north-west Spain, 2003–2007, and inferences about cetacean habitat preferences." *Journal of the Marine Biological Association of the United Kingdom* 90.8 (2010): 1547-1560.
- [2] Tamara Narganes Homfeldt, Séverine Methion, Bruno Díaz López, Identifying Suitable Habitats for Harbour Porpoise (*Phocoena phocoena*) Conservation in Galician Waters, NW Spain, World Marine Mammal Conference, December 2019, DOI:[10.13140/RG.2.2.14669.79842](https://doi.org/10.13140/RG.2.2.14669.79842)
- [3] Robinson, Stephen P., Paul A. Lepper, and Richard A. Hazelwood. "Good practice guide for underwater noise measurement." (2014).
- [4] European community, "2008/56/EC european directive, Establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive)" 2008.
- [5] European Marine Observation and Data Network (EMODnet), [www.emodnet.eu](http://www.emodnet.eu)
- [6] Sheng Shen, Honghui Yang, Junhao Li, Guanghui Xu and Meiping Sheng. "Auditory Inspired Convolutional Neural Networks for Ship Type Classification with Raw Hydrophone Data" (2018) DOI:[10.3390/e20120990](https://doi.org/10.3390/e20120990)
- [7] Chris Seiffert, Taghi M. Khoshgoftaar, Member, IEEE, Jason Van Hulse, Member, IEEE, and Amri Napolitano. "RUSBoost: A Hybrid Approach to Alleviating Class Imbalance" (2010)
- [8] A. M. Usman, O. O. Ogundile and D. J. J. Versfeld, "Review of Automatic Detection and Classification Techniques for Cetacean Vocalization," in *IEEE Access*, vol. 8, pp. 105181-105206, 2020, DOI: [10.1109/ACCESS.2020.3000477](https://doi.org/10.1109/ACCESS.2020.3000477).