

1  
2  
3  
4  
5 **IMAGE PROCESSING TECHNIQUE APPLIED ON UNDERWATER**  
6 **ACOUSTIC FIN WHALE AUTOMATIC DETECTOR**

7  
8 **PACS:** 43.30.Sf  
9

10 Feliu-Tena, Blanca<sup>1,3</sup>; Bou-Cabo, Manuel<sup>2,3</sup>; Lara, Guillermo<sup>2,3</sup>; Rodilla, Miguel<sup>1,3</sup>; Miralles,  
11 Ramón<sup>4,3</sup>; Belda, Eduardo J.<sup>1,3</sup>  
12

13 <sup>1</sup> Instituto de Inv. Para la Gestión Integrada de Zonas Costeras (IGIC), Universitat Politècnica  
14 de València (UPV), C/Paranimf 1, Grau de Gandía, 46730 Valencia, Spain.  
15 [blafete@epsq.upv.es](mailto:blafete@epsq.upv.es); [mrodilla@hma.upv.es](mailto:mrodilla@hma.upv.es); [ebelda@dca.upv.es](mailto:ebelda@dca.upv.es)

16 <sup>2</sup> Instituto Español de Oceanografía (CN.IEO-CSIC), C.O. Murcia, C/el Varadero 1, Lo Pagan,  
17 30740 Murcia, Spain. [manuel.bou@ieo.csic.es](mailto:manuel.bou@ieo.csic.es); [guillermo.lara@ieo.csic.es](mailto:guillermo.lara@ieo.csic.es)

18 <sup>3</sup> Unidad Mixta de Investigación IEO-UPV, Tinglados Muelle Frutero, Grau de Gandia, 46370  
19 Valencia, Spain

20 <sup>4</sup> Institute of Telecommunications and Multimedia Applications (iTEAM), Universitat Politècnica  
21 de València (UPV), Camino de Vera S/N, 46022 Valencia, Spain. [rmiralle@dcom.upv.es](mailto:rmiralle@dcom.upv.es)  
22

23 **Palabras Clave:** rorcual común, acústica pasiva, detección automática, tratamiento digital de  
24 imagen  
25

26 **ABSTRACT.**

27 The underwater acoustic monitoring has been widely used to monitor the behaviour, habitat use  
28 and many other aspects related to the cetacean's life. The fin whale is the second biggest  
29 animal on earth and some aspects of its habitat use are not well understood. The underwater  
30 acoustic monitoring of fin whales can bring the opportunity to have long term data related to a  
31 given location. However, the acoustic detection of bioacoustics signals produced by these  
32 animals represent a relevant challenge due to its low frequency characteristics. The low number  
33 of detections expected, together with the spectral overlap of low frequency underwater noise  
34 with vocalizations bandwidth, turn necessary to dispose of long-term monitoring data and  
35 detection tools to retrieve the fin whale pulses from large amount of data. Therefore, exploring  
36 techniques to detect the low frequency vocalization signals of fin whales would be a valuable  
37 tool for bio-acousticians. We present a method based on an image processing technique that  
38 applies a color-based decomposition on spectrogram images to automatically detect signals  
39 compatible with fin whales' vocalizations. This technique has been revealed as a potential tool  
40 to perform the automatic detection.  
41

42 **RESUMEN.**

43 La monitorización acústica submarina ha sido ampliamente utilizada para monitorizar el  
44 comportamiento, uso del hábitat y otros aspectos relacionados con la vida de los cetáceos. El  
45 rorcual común es el segundo animal más grande del planeta, aun así, todavía existen muchos  
46 aspectos que se desconocen. La monitorización acústica de rorcual común puede brindar la  
47 oportunidad de tener datos de largos periodos desde una localización puntual. Sin embargo, la  
48 detección de señales bio-acústicas representa un reto por la baja frecuencia de sus  
49 vocalizaciones. El bajo número de detecciones en el tiempo, junto con el elevado  
50 enmascaramiento por ruido de baja frecuencia, hace necesario disponer de una monitorización  
51 a largo plazo y de herramientas de detección que permitan extraer las vocalizaciones de  
52 grandes volúmenes de datos. Por esta razón, la exploración de técnicas de detección de  
53 vocalizaciones de rorcual común de baja frecuencia representa una herramienta valiosa para  
54 los bio-acústicos. Presentamos un método basado en la técnica de procesamiento de imagen  
55 que aplica una descomposición basada en color de las imágenes del espectrograma que  
56 detecta automáticamente señales compatibles con vocalizaciones de rorcual común. Esta  
57 técnica ha revelado tener potencial como herramienta destinada para la detección automática.  
58  
59

60 **1. INTRODUCCIÓN**

61

62 El rorcual común (*Balaenoptera physalus*) es el cetáceo de mayor tamaño que se puede  
63 encontrar en el mar Mediterráneo. No obstante, se tiene todavía muy poca información sobre su  
64 distribución, zonas de reproducción, zonas de alimentación o rutas migratorias. Esta especie en  
65 su subpoblación mediterránea es además de especial interés debido a que se encuentra  
66 catalogada como Amenazada por la IUCN desde el año 2021 [1].

67

68 Numerosos estudios se han realizado durante la última década [2] [3] en relación a la estima de  
69 densidad de población, distribución y usos del hábitat. Cabe destacar que desde 2021 y gracias  
70 al apoyo de la Fundación Biodiversidad, la Universidad Politécnica de Valencia (UPV) está  
71 desarrollando un proyecto que trata de responder algunas incógnitas acerca de la vida de los  
72 rorcuales en el Mediterráneo, el Proyecto CaboRorcual. Este proyecto se desarrolla  
73 principalmente en el Cabo de San Antonio y Cabo de la Nao (Alicante), lugar que se caracteriza  
74 por poder observarse este paso migratorio de rorcual común muy próximo a la costa. Aunque  
75 existen diferentes técnicas de muestreo para obtener información sobre las poblaciones de estos  
76 misticetos, una de las acciones del proyecto contempla la obtención de datos de presencia  
77 mediante dispositivos de acústica pasiva. Para ello se ha llevado a cabo la instalación de  
78 dispositivos experimentales desarrollados por investigadores del Instituto de  
79 Telecomunicaciones y Aplicaciones Multimedia (iTEAM) de la UPV [4] que permiten la obtención  
80 de grabaciones durante un periodo de tiempo de hasta 3 meses, con una frecuencia de muestreo  
81 de 192 kH. Este dispositivo se conoce como SAMARUC (ver Figura 1).

82



83

84

Figura 1. Dispositivo de Monitorización Acústica Pasiva (PAM). El SAMARUC está  
diseñado y desarrollado por ingenieros del iTEAM de la UPV.

85

86

87 La monitorización acústica pasiva permite abarcar extensos periodos de tiempo, así como  
88 periodos en los que la monitorización visual no es posible por las inclemencias meteorológicas o  
89 la falta de luz, lo cual brinda la posibilidad de poder obtener información temporal extensa tanto  
90 de las especies objeto de estudio como de otros indicadores bio-acústicos. Dentro del marco del  
91 Proyecto CaboRorcual se han realizado un total de 12 instalaciones del dispositivo de  
92 monitorización acústica pasiva SAMARUC en cuatro localizaciones distintas (ver Figura 2),  
93 abarcando diferentes puntos de paso del rorcual común durante la migración estival. Como  
94 resultado de todas las campañas realizadas hasta la fecha se han generado un total de 88,770  
95 archivos WAV, acumulando 7,389.17 horas de grabaciones. En la Tabla 1 se incluye un resumen  
96 de las características técnicas de las medidas realizadas a lo largo del tiempo de monitorización:  
97 identificación del punto de toma de datos (ID); fecha de instalación de los equipos de acústica  
98 pasiva en el medio (Fecha); profundidad del fondo en metros del punto donde se encuentran los  
99 equipos (Prof.); frecuencia de muestreo del equipo de monitorización acústica empleado (Frec.  
100 Muestreo); ganancia en el registro de las grabaciones (Ganancia); hidrófono utilizado  
101 componente del equipo (Hidrófono); intervalo de encendido y apagado del equipo en minutos  
102 (Ciclo ON/OFF); total de archivos de 5 minutos de duración grabados en el periodo de tiempo en  
103 que ha estado funcionando el equipo (N archivos); y días totales en los que el equipo ha estado  
104 instalado y registrando información acústica (Días grabados).

105

106

107

108

109

En la presente comunicación se va a abordar la implementación de una técnica basada en  
tratamiento digital de imagen, con el objetivo de disponer de un algoritmo de detección  
automática de vocalizaciones que permita tratar de manera adecuada la gran cantidad de datos  
obtenidos a partir de la monitorización acústica continuada.

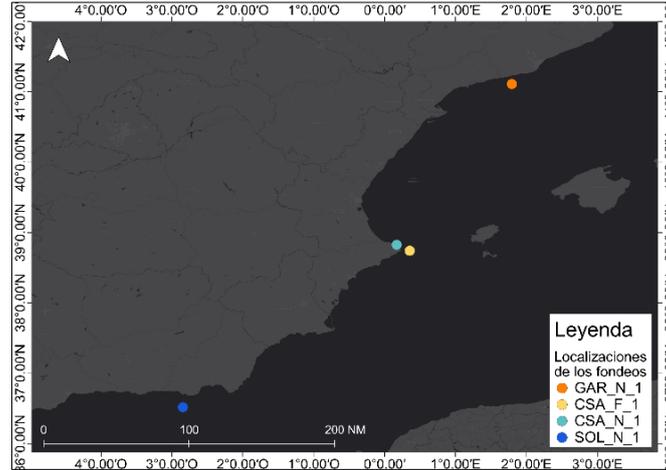


Figura 2. Localizaciones donde se ha llevado a cabo la instalación del dispositivo experimental SAMARUC en el marco del Proyecto CaboRorcual de la UPV.

Tabla 1. Características y parámetros de las campañas de monitorización mediante línea de instrumentación utilizando dispositivos acústico pasivos realizadas en el marco del Proyecto CaboRorcual de la UPV. PAM empleado: SAMARUC (iTEAM) de la UPV.

ID	Fecha	Prof. (m)	Frec. Muestreo (kHz)	Ganancia	Hidrófono	Ciclo ON/OFF	N Archivos	Días grabados
CSA_F	20210603	80	192	6	C57-754	5/5	7722	55
CSA_N	20200618	24	192	0	C-41	5/10	3059	33
CSA_N	20210326	24	192	0	C57XRS-855	5/5	3882	19
CSA_N	20210603	24	192	0	C57-754	5/5	6822	49
CSA_N	20210805	24	192	0	C57XRS-855	5/5	8686	61
CSA_N	20211025	24	192	0	C57-753	5/10	8070	85
CSA_N	20220128	24	192	0	C57-753	5/5	8686	61
CSA_N	20220509	24	192	0	C57XRS-864	5/5	5596	40
CAS_N	20220617	24	192	0	C57XRS-864	30/1	10122	36
CAS_T	20220617	29	192	0	C57XRS-855	30/1	8958	32
GAR_N	20220331		192	0	C57XRS-856	5/5	8712	62
SOL_N	20210616	84	192	6	C57XRS-858	5/10	8355	88

Las vocalizaciones emitidas por el rorcual común se pueden atendiendo a tres rangos frecuenciales: 20Hz, 40-80Hz (denominados en la literatura como Downsweeps) y 135Hz [5]. Las vocalizaciones de 135Hz se han reportado como coincidentes a las de 20Hz, por lo que no se ha realizado un detector específico. La metodología aplicada para frecuencias de 20Hz y 40-80Hz es similar, segmentando las imágenes en diferentes regiones del espacio frecuencial dependiendo de la señal a estudiar.

## 2. DESARROLLO DE LA METODOLOGÍA

El detector de vocalizaciones de rorcual se ha desarrollado utilizando el lenguaje de programación R, popularmente utilizado en aplicaciones de diversa índole (computación estadística, aplicaciones de tratamiento digital de imagen, análisis espacial, etc.) con el entorno de desarrollo RStudio [6]. En los pasos de su desarrollo (Figura 3) se utilizan diferentes librerías de este software libre. La primera es la librería *soundgen* [7] con la cual se realiza la selección de la banda de frecuencia del espectrograma obtenido a través de la señal acústica que posteriormente será analizada. Sobre esta imagen obtenida en código de pseudocolor RGB (Rojo (Red-R), Verde (Green-G), Azul (Blue-B)) generado con la paleta de color 'seewave' de la librería *soundgen*, se realiza el cálculo de la descomposición de color. Esta descomposición se implementa utilizando la librería *colordistance* [8], la cual permite segmentar las coordenadas RGB de cada píxel. Finalmente, con el objetivo de obtener una potencial detección bio-acústica,

139 el algoritmo utiliza la librería *countcolors* [9] para buscar en la imagen la presencia de píxeles en  
 140 un rango de color previamente definido a través de la calibración del método con señales de  
 141 rorcual detectadas de manera manual. En caso de detectar presencia de vocalización de rorcual  
 142 común en la imagen, ésta será guardada para su posterior revisión y verificación por un operador.  
 143 En la Figura 3 se muestra el diagrama de flujo del algoritmo desarrollado.  
 144

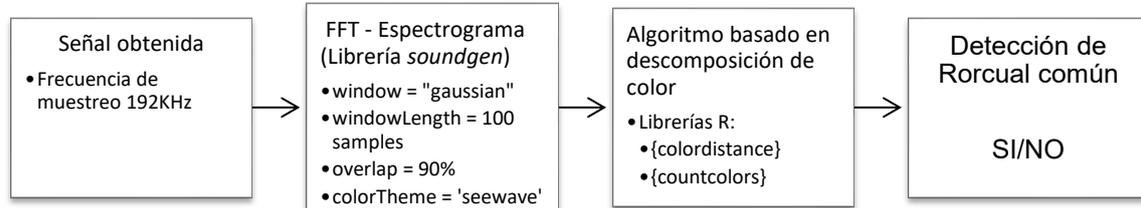


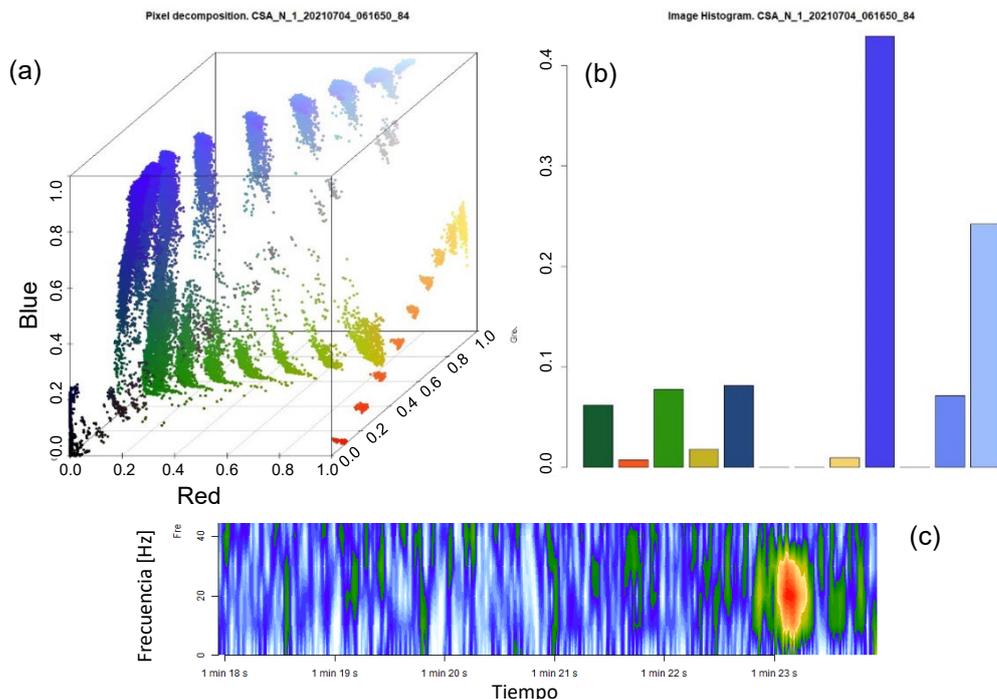
Figura 3. Diagrama de flujo del algoritmo desarrollado.

145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175

### 2.1. Definición de Coordenadas RGB

Las imágenes obtenidas del análisis frecuencial de las señales grabadas se segmentan en fragmentos de 6 segundos y se descomponen en los tres colores primarios RGB. Para que el método pueda seleccionar de forma automática donde hay y donde no hay una vocalización de rorcual, se ha definido previamente con qué coordenadas RGB coincide esta señal. Las coordenadas RGB son los tres valores en el rango (0-1) que definen el color específico de nuestra señal. Para ello se han seleccionado varios fragmentos en los que previamente se habían detectado vocalizaciones de rorcual (véase ejemplo en Figura 4c) y se ha realizado una descomposición por píxeles (Figura 4a) en un espacio tridimensional en el que los píxeles se agrupan según la coordenada RGB a la que corresponden. De la misma imagen se obtiene el histograma (Figura 4b) que definen la proporción de cada coordenada RGB, realizando la agrupación por subdivisiones de los ejes del gráfico 3D. Se muestra también en la Figura 5 un ejemplo del resultado de la descomposición por color de un fragmento en el que no se identifica ninguna vocalización de rorcual.

Se puede observar en el primer ejemplo (Figura 4) cómo la vocalización de rorcual queda definida por unas coordenadas RGB (0.8, 0.2, 0.2), observándose una clara rampa ascendente que mantendría el R en 0.8, aumentando el G hasta casi 1 y el B hasta 0.5, mientras que en la descomposición por píxeles de la imagen que no contenía vocalización de rorcual común no se puede observar ningún píxel que coincida con esas coordenadas RGB. Además, comparando ambos histogramas (Figura 4b y Figura 5b) se aprecia como la distribución es muy diferente, encontrándose completamente ausente en el segundo histograma (Figura 5b) el color correspondiente al rojo y siendo mucho más elevada la presencia del color verde.



176

177

178

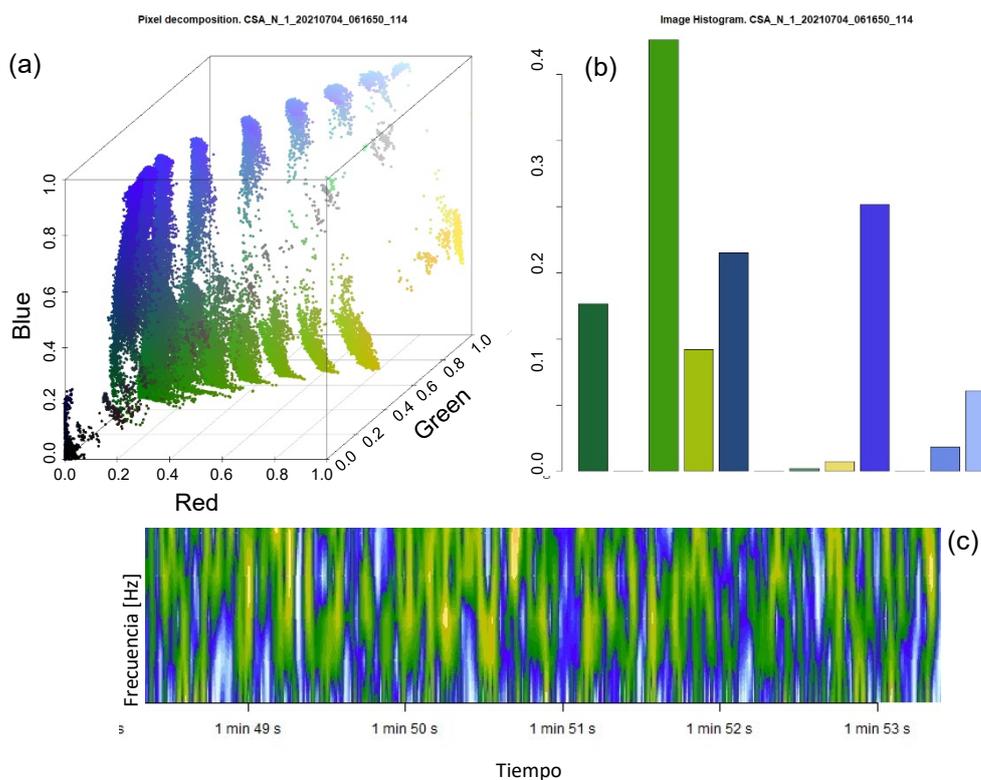
179

180

181

182

Figura 4. (a) Descomposición de la imagen en píxeles, agrupándolos según coordenadas RGB. (b) Histograma de la proporción de píxeles de cada grupo RGB. (c) Imagen RGB recortada del espectrograma con un pulso de 20Hz de rorcual común a partir de la cual se han obtenido los gráficos (a) y (b) de esta figura.



183

184

185

186

187

188

189

Figura 5. (a) Descomposición de la imagen en píxeles, agrupándolos según coordenadas RGB. (b) Histograma de la proporción de píxeles de cada grupo RGB. (c) Imagen RGB recortada del espectrograma sin ningún tipo de vocalización de rorcual común a partir de la cual se han obtenido los gráficos (a) y (b) de esta figura.

190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208

### 2.3. Rendimiento de los Detectores

El estudio del rendimiento de los detectores se ha realizado sobre un total de 144 archivos de 5 minutos de duración cada uno, que corresponden a un día completo de obtención de datos en el cual había vocalizaciones a las diferentes frecuencias descritas anteriormente. Las métricas utilizadas para el cálculo del rendimiento del detector se describen a continuación:

- Falso positivo (FP): El detector selecciona un fragmento como positivo, pero realmente no lo es.
- Falso negativo (FN): El detector no selecciona un fragmento en el que realmente sí existe una vocalización de rorcual.
- Positivo real (TP): El detector selecciona un fragmento como positivo y sí lo es.
- Tasa de Falsos Positivos (FPR): Proporción de fragmentos seleccionados por el método como positivos que en realidad no contienen una vocalización de rorcual.

$$FPR = \frac{FP}{TP + FP} \quad (1)$$

209  
210  
211  
212

- Tasa de Positivo Real (TPR): Proporción de fragmentos seleccionados por el método como positivos que, en efecto, sí contienen una vocalización de rorcual.

$$TPR = \frac{TP}{TP + FN} \quad (2)$$

213  
214  
215  
216  
217

- Tasa de Error (ERR): Proporción normalizada, de error del método. Es la relación entre los falsos positivos y los falsos negativos frente al total de fragmentos seleccionados como positivos.

$$ERR = \frac{FP + FN}{TP + FP + FN} \quad (3)$$

218  
219  
220  
221  
222

- Precisión (PREC): El cálculo de la precisión tiene en cuenta los positivos verdaderos frente al total de fragmentos seleccionados como positivos, tanto los positivos verdaderos como los falsos positivos.

$$PREC = \frac{TP}{TP + FP} \quad (4)$$

223  
224  
225

### 3. RESULTADOS Y DISCUSIÓN

226  
227  
228  
229  
230

Los 144 archivos analizados para el cálculo del rendimiento del método contenían un total de 73 pulsos de 20Hz y 26 pulsos de 40-80Hz. En la Tabla 2 se muestran los resultados obtenidos para cada una de las métricas.

231  
232

Tabla 2. Resultados del cálculo del rendimiento de los detectores de rorcual común de 20Hz y de 40-80Hz.

Detector	TP	FP	FN	FPR	TPR	ERR	PREC
20Hz	38	771	35	0.95	0.52	0.95	0.05
40-80Hz	25	1536	1	0.98	0.96	0.98	0.02

233  
234  
235  
236  
237

Los resultados obtenidos muestran una diferencia de casi del 50% en la TPR para el caso del detector de pulsos de 20Hz. Como se puede observar en la tabla de resultados, maximizar el valor de la variable TPR conlleva un mayor número de FP. Esto se ha observado en otros detectores disponibles en la bibliografía [10] que utilizan otra metodología de detección.

238

239

240

241

242

243

244

245

246

247

248

249

#### 4. CONCLUSIONES

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

#### AGRADECIMIENTOS

265

266

267

268

269

270

#### REFERENCIAS

271

272

273

274

275

276

[1] Panigada, S., Gauffier, P. & Notarbartolo di Sciara, G. 2021. Balaenoptera physalus (Mediterranean subpopulation). The IUCN Red List of Threatened Species 2021: e.T16208224A50387979. <https://dx.doi.org/10.2305/IUCN.UK.2021-3.RLTS.T16208224A50387979.en>. Accessed on 12 September 2022.

277

278

[2] Castellote, M., Clark, C. W., & Lammers, M. O. (2012). Fin whale (*Balaenoptera physalus*) population identity in the western Mediterranean Sea. *Marine Mammal Science*, 28(2), 325-344.

279

280

281

[3] Sciacca V, Caruso F, Beranzoli I, Chierici F, De Domenico E, Embriaco D, et al. (2015) Annual Acoustic Presence of Fin Whale (*Balaenoptera physalus*) Offshore Eastern Sicily, Central Mediterranean Sea. *PLoS ONE* 10(11): e0141838. doi:10.1371/journal.pone.0141838

282

283

284

[4] Lara, G., Miralles, R., Bou-Cabo, M., Esteban, J. A., & Espinosa, V. (2020). New Insights into the Design and Application of a Passive Acoustic Monitoring System for the Assessment of the Good Environmental Status in Spanish Marine Waters. *Sensors*, 20(18), 5353.

285

286

287

[5] Garcia, H. A., Couture, T., Galor, A., Topple, J. M., Huang, W., Tiwari, D., & Ratilal, P. (2020). Comparing performances of five distinct automatic classifiers for fin whale vocalizations in beamformed spectrograms of coherent hydrophone array. *Remote Sensing*, 12(2), 326.

288

289

[6] R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

290

291

[7] Anikin, A. (2019). Soundgen: an open-source tool for synthesizing nonverbal vocalizations. *Behavior Research Methods*, 51(2), 778-792.

- 292 [8] Hannah Weller (2021). colordistance: Distance Metrics for Image Color Similarity. R package  
293 version 1.1.2. <https://CRAN.R-project.org/package=colordistance>
- 294 [9] Hannah Weller (2019). countcolors: Locates and Counts Pixels Within Color Range(s) in  
295 Images. R package version 0.9.1. <https://CRAN.R-project.org/package=countcolors>
- 296 [10] Pereira, A. (2018). Investigation of fin whales using ocean-bottom recordings.
- 297 [11] Wenz, G. M. (1962). Acoustic ambient noise in the ocean: Spectra and sources. *The Journal*  
298 *of the Acoustical Society of America*, 34(12), 1936-1956.