Taylor & Francis
Taylor & Francis Group

Bioacoustics
The International Journal of Animal Sound and its Recording

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tbio20

Extracting dolphin whistles in complex acoustic scenarios: a case study in the Bay of Biscay

Ramón Miralles, Carles Gallardo, Guillermo Lara & Manuel Bou Cabo

To cite this article: Ramón Miralles, Carles Gallardo, Guillermo Lara & Manuel Bou Cabo (2024) Extracting dolphin whistles in complex acoustic scenarios: a case study in the Bay of Biscay, Bioacoustics, 33:3, 260-277, DOI: 10.1080/09524622.2024.2338387

To link to this article: https://doi.org/10.1080/09524622.2024.2338387

BIOACOUSTICS
2024, VOL. 33, NO. 3, 260–277
https://doi.org/10.1080/09524622.2024.2338387

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Extracting dolphin whistles in complex acoustic scenarios: a case study in the Bay of Biscay

Ramón Miralles[a], Carles Gallardo[a], Guillermo Lara[b] and Manuel Bou Cabo[b]

[a]Institute of Telecommunications and Multimedia Applications (iTEAM), Universitat Politècnica de València, Valencia, Spain; [b]Underwater Acoustics Group, Spanish Institute of Oceanography (IEO-CSIC), C.O. Murcia, Spain

ABSTRACT
Accurate whistle contour extraction is crucial in many dolphin behavioural studies. Traditionally, whistle contour extraction involves a first step of finding whistle candidates by peak-level detection in the time-frequency domain, followed by a determination of when peaks are close enough to each other to be part of the same whistle contour. In complex scenarios, such as those with a large number of individuals vocalising simultaneously or those with a sudden increase in background noise, peak-level detection may not provide a number of accurate whistle candidates that is large enough to extract the whistle contour or to disambiguate individual whistles when they cross one another. In these adverse scenarios, a different approach, based on the pyknogram representation, can produce a more accurate detection of whistle candidates and evenly distributed candidates throughout the duration of the whistle. This work compares the peak-level extraction approach of the spectrogram with the point-density extraction approach of the pyknogram. We propose a technique that combines estimates of the central frequency and bandwidth to extract whistle candidates in adverse scenarios. The method has been successfully used for the vocalisation extraction of dolphins in the Bay of Biscay (Spain) using a database of more than 2000 dolphin whistles.

ARTICLE HISTORY
Received 8 June 2023
Accepted 25 March 2024

KEYWORDS
Underwater acoustics; time–frequency analysis; passive acoustic monitoring; cetacean whistle extraction

1. Introduction

Over the last few years, several studies have worked on methods for extracting marine mammal calls (whistles, moans, and other tonal sounds) when performing passive acoustic monitoring. In the majority of the cases, this has been achieved by automatically tracking these animal sounds in the time-frequency plane following the contour ridges or by peak-level detection. Some approaches found in the literature include: detecting all local maxima and fitting a curve through the peaks at successive time slices (Mallawaarachchi et al. 2008; Mellinger et al. 2011); using image processing techniques to track the spectral ridges (Kershenbaum and Roch 2013); using particle filters to find estimates of the posterior distribution (that of the estimated contour given the spectral peaks) (White and Hadley 2008; Roch et al. 2011); using an adaptive notch filter to minimise the output by placing notches at the whistle peaks (Johansson and White 2011);

CONTACT Ramón Miralles ✉ rmiralle@dcom.upv.es

and using the probability hypothesis density filter (Gruden and White 2020) as an approximation to the optimal Bayesian filter. These approaches entail, either explicitly or implicitly, two stages: detecting the whistle candidates and then extracting the whistle by joining the candidates using different criteria. In this work, we use 'candidates' to refer to the set of peaks, pixels, high-density regions, or any other set of detected points that might be potentially part of a whistle contour. In all of the methods mentioned above, the whistle extraction stage works well when there is an accurate and even distribution of the candidates. However, these approaches behave very differently when some candidates are not properly detected. This may happen under adverse conditions such as low Signal-to-Noise Ratio (SNR) whistles (Mallawaarachchi et al. 2008), quick changes in the noise floor, or overlapping whistles from several animals vocalising simultaneously (Roch et al. 2011).

In 1996, Potamianos introduced (Potamianos and Maragos 1996) a new time-frequency representation method that was named pyknogram (from the Greek word 'pykno'= dense). This representation proved to be especially appropriate for the extraction of formants in speech signals. The technique clearly displays the formant position and bandwidth with high and low-density regions (see Figure 1 for an example of how the pyknogram compares to the spectrogram).

Conceptually speaking, the pyknogram was devised to exploit the fact that speech production can be approximated by a sum of AM-FM models representing each one of the formants. It makes use of non-linear methods, such as the Teager-Keiser energy operator, to track the instantaneous frequency of each one of the components. Somewhat more recent works have proven how the pyknogram can help in different speech-related problems: speaker verification (Vijayan et al. 2016); speaker verification in overlapped scenarios (in which it achieved a relative 20% improvement across different signal-to-interference ratios (Shokouhi and Hansen 2017)); and identification of segments of overlapping speech in co-channel recordings (Yousefi et al. 2018). These are some of the most recent works, and the pyknogram showed good consistent behaviour in challenging scenarios in all of them.

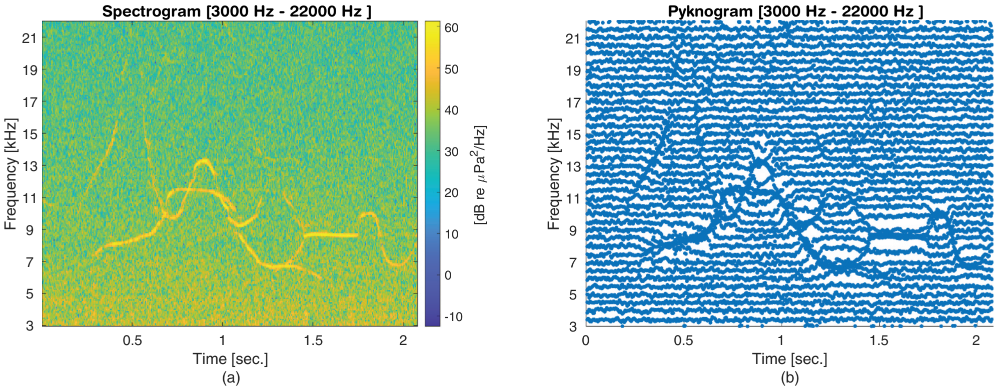

Figure 1. (a) Spectrogram computed using a hamming window of length 10.7 ms (which results in 93.5 Hz frequency bin resolution) and (b) pyknogram computed using a filterbank of 1kHz bandwidth with 50% frequencial overlapping (which results in 500 Hz frequency bin resolution) of an underwater recording containing multiple dolphin whistles.

Cetacean whistles and moans can also be approximated using AM-FM models, and thus the pyknogram might work well for the contour extraction of these tonal sounds. In a somewhat related line of work, in (Cornel et al. 2010), C. Ioana showed how, in spite of crossing or noise interferences, extracting the instantaneous frequency and phase provided a superior accuracy when following time-frequency variations. This is just more evidence that suggests that a thorough study on the use of the pyknogram, which is also based on the instantaneous frequency, might provide some benefits when trying to extract whistle contours in complex acoustic scenarios.

In this work, we evaluate how the pyknogram compares to the spectrogram when extracting tonal information or whistle candidates that could later be used for whistle contour extraction. As a case study, this has been applied to extract the whistles of bottlenose dolphins (*Tursiups truncatus*), striped dolphins (*Stenella coeruleoalba*) and common dolphins (*Delphinus delphis*) in the Bay of Biscay (on the northern coast of Spain). The rest of this work is structured as follows. In Subsection 2.1, we formulate the pyknogram equations to be used in discrete passive acoustic monitoring recordings. We also propose a new paradigm for whistle candidates detection based on the pyknogram representation. This new technique is later compared with a well-known whistle candidate detection based on the spectrogram method, which is summarised in Subsection 2.2. In Subsection 2.3, we explain how the whistle candidates obtained using the spectrogram and pyknogram are used to extract the whistle contours using the GM-PHD method (Gruden and White 2016). The results include assessing the accuracy of both techniques under adverse simulated scenarios (Subsections 3.1 and 3.2) as well as in a set of challenging real-world recordings from an acoustic campaign done in the Bay of Biscay (Subsection 3.3). The selected recordings were taken in an area with a high density of marine mammals, and they contain multiple bottlenose, striped, and common dolphins vocalising simultaneously as well as other interfering noises. We conclude the work in Section 4 discussing the possibilities and limitations of the pyknogram as an alternative to the spectrogram for whistle extraction in passive acoustic monitoring.

2. Methods

2.1. Pyknogram-based whistle candidate detection

Let's assume a signal $x(t)$ composed of N whistles $w_k(t)$, $k = \{1, 2, \cdots, N\}$ with an arbitrary amount of additive pink noise $\eta(t)$, as shown in Equation (1):

$$x(t) = \sum_{k=1}^{N} w_k(t) + \eta(t). \tag{1}$$

We used pink noise because its power spectral density decreases proportionally to the inverse of the frequency as happens with underwater ambient noise. Each one of the whistles can be approximated with an AM-FM model, as seen in Equation (2):

$$w_k(t) = a_k(t) cos\left[2\pi\left(\int_0^t f_k(\tau)d\tau\right) + \theta_k\right], \tag{2}$$

where, $a_k(t), f_k(t)$, and θ_k are respectively the instantaneous amplitude, the instantaneous frequency, and the initial phase of the whistle k. The equations for computing the pyknogram for a discrete signal $x(n)$, of time index n, can be easily obtained by discretisation of the equations presented in (Potamianos and Maragos 1996) with a sampling frequency $f_s = 1/T_s$, T_s being the sampling period. In the following steps, we assume that the total number of samples of the discretised $x(n)$ is equal to $M \cdot Q$, where M is the number of samples of the analysis frame and Q is an integer value.

(1) Use a Gabor filter bank (Gabor 1946) to decompose the broadband signal $x(n)$ into a collection of relatively lower narrow band signals $x_i(n)$. We can create a discrete version of the Gabor filter by sampling the continuous version. The impulse response is thus a discrete Gaussian modulated sinusoid given by Equation (3):

$$h_i(n) = exp(-\alpha^2(n/f_s)^2) \cdot cos(2\pi v_i n/f_s), \tag{3}$$

where v_i, $i = [1, 2, \cdots, I]$ is the centre frequency of the filter with I being the total number of bands and α being the bandwidth parameter (effective rms bandwidth approximated by (Maragos et al. 2002) $BW_{Gabor} = \alpha/\sqrt{2\pi}$). Although there are different alternatives for separating the broadband signal into narrow band signals, Gabor's method is the simplest one (Hsu et al. 2011) and provides accurate instantaneous amplitude and frequency estimates (Delprat et al. 1992) even when compared with some other abrupt filter techniques.

In this work, a bandwidth of $BW_{Gabor} = 1kHz$ for each Gabor filter was used. The filter bank covered from 3000 to 22,000 Hz with a bandwidth overlapping factor of $\Delta_W = 50\%$. The frequencies covered by the filter bank were selected in accordance with the frequency range of the three dolphin species studied. The overlapping factor value Δ_W was empirically chosen as a trade off between computational complexity and a frequency resolution that was high enough to allow the whistle contour to be reconstructed.

(2) Estimate the Instantaneous Amplitude (IA) envelope $|a_i(n)|$ and Instantaneous Frequency (IF) $f_i(n)$ in each one of the filter bank bands. To do this, the Hilbert Transform Demodulation (HTD) was used. Even though the HTD presents a higher computational complexity than other techniques, such as the Energy Separation Algorithm (ESA), the HTD provides smoother bandwidth estimates (Potamianos and Maragos 1996) and therefore less variance. The process to obtain the IA and the IF involves computing the Hilbert transform $\mathcal{H}[\cdot]$ for each one of the narrow band signals, which is shown in Equation (4) and also in Equation (5) and Equation (6):

$$|a_i(n)| = \sqrt{x_i(n)^2 + (\mathcal{H}[x_i(n)])^2} \tag{4}$$

$$\theta_i(n) = tan^{-1}\left[\frac{\mathcal{H}[x_i(n)]}{x_i(n)}\right] \tag{5}$$

$$f_i(n) = \frac{f_s}{2\pi}\delta_1[\theta_i](n), \tag{6}$$

where $\delta_1[\theta_i](n)$ is the central difference estimate of θ_i, which is computed as follows:

$$\delta_1[\theta_i](n) = (\theta_i(n+1) - \theta_i(n-1))/2. \tag{7}$$

(3) Obtain the short-time estimates of the central frequency $F_i(n_0)$ and the bandwidth $B_i^2(n_0)$ of a whistle candidate. In order to obtain a more natural estimate, weighted moments estimates of the IF and IA were used ar (Potamianos and Maragos 1996):

$$F_i(n_0) = \frac{\sum\limits_{n=n_0}^{n_0+M-1} f_i(n)|a_i(n)|^2}{\sum\limits_{n=n_0}^{n_0+M-1} |a_i(n)|^2} \tag{8}$$

$$B_i^2(n_0) = \frac{\sum\limits_{n=n_0}^{n_0+M-1} (\delta_1[a_i](n)/2\pi)^2 + (f_i(n) - F_i(n_0))^2|a_i(n)|^2}{\sum\limits_{n=n_0}^{n_0+M-1} |a_i(n)|^2}, \tag{9}$$

where $\delta_1[a_i](n)$ is the central difference estimation of a_i, which is computed as in Equation (7). The values n_0 and M are the start sample and number of samples of the analysis frame, respectively. For example, for a 50% overlap, the time frames will start at $n_0 = [0, M/2, M, \cdots, (Q-1) \times M]$.

Each one of the estimates of the central frequency $F_i(n_0)$ at time frame n_0 and band f_i is a pyknogram point. The scatter plot of $F_i(n_0)$ is known as the pyknogram representation (see Panel (b) of Figure 1).

Whistle candidates can be extracted from the pyknogram taking into account that the presence of a dolphin whistle (or any other signal resembling an AM-FM component) introduces high-density regions, i.e. the points are close to each other in the frequency domain (see Figure 1). The absence of whistles or the presence of any other broadband sounds do not alter the natural distribution of the central frequency estimates, producing a plot density that is defined by the separation among bands of the Gabor filter bank. If the pyknogram is interpreted as a 2D point cloud, we can adapt many of the point cloud denoising algorithms to extract only those points that are associated with whistle fragments. For the sake of simplicity, we will use the Parzen window technique (Parzen 1962) to obtain the probability density function of the pyknogram point separation in the frequency domain. This way, we can calculate the density of each pyknogram point using Equation (10):

$$d(F_i(n_0)) = \frac{1}{I \cdot h} \sum_{j=1}^{I} K\left(\frac{F_i(n_0) - F_j(n_0)}{h}\right). \tag{10}$$

where the function $K(\cdot)$ is a kernel function. In this work, we used a rectangular kernel $K(p) = 1,\ |p| \leq 1$ and 0 for the rest of p. Other kernel functions, such as a Gaussian kernel, can be used, providing a slightly superior number of detected whistle candidates. However, the uniform kernel is more robust to noise and provides a number of whistle candidates that is large enough to track their contours in most situations. Equation (10) computes the frequency separation among $F_i(n_0)$ and the rest of the pyknogram points $F_j(n_0)$ and normalizes it by the bandwidth h. In this work, $h = 250Hz$. This is one-half of the Gabor filter separation between consecutive bands and is obtained as:

$$h = BW_{Gabor}/2 \cdot (1 - \Delta_W/100). \tag{11}$$

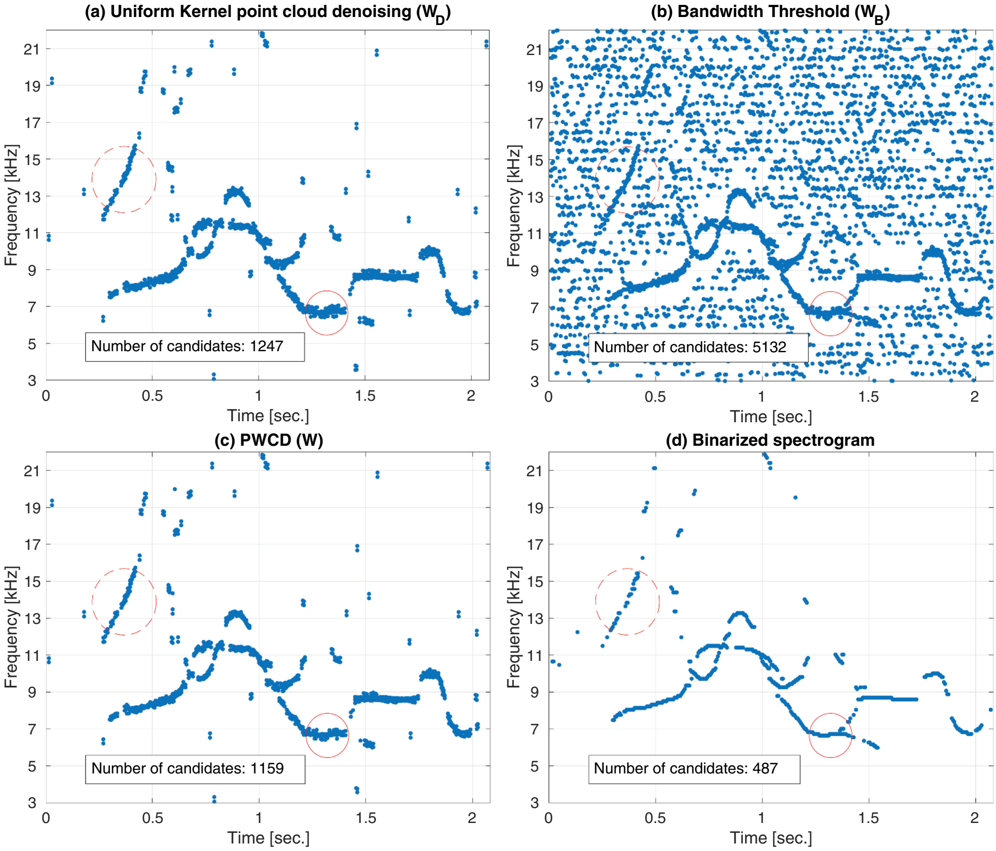

Figure 2. An example of whistle candidate extraction in real dolphin whistles by method: (a) 1247 candidates for point cloud denoising (W_D); (b) 5132 candidates for bandwidth threshold (W_B); (c) 1159 candidates for PWCD (W); and (d) 487 candidates for binarized spectrogram. The candidate count was done for the whole pyknogram representation. The red circles indicate areas to focus on in order to see the dispersion of the extracted whistle candidates.

As a result, $d(F_i(n_0))$ is proportional to the number of points that are separated in frequency less than h Hz. Therefore, if $P = \{\mathbf{p}_j \in \mathbb{R}^2\}$ with $j = \{1, 2, \cdots, I \times Q\}$ is the set of all estimated pyknogram central frequency points where each $\mathbf{p}_j = [n_0 \cdot T_s, F_i(n_0)]$, we can obtain all of the whistle density points W_D that are whistle candidates using $W_D = \{\mathbf{p}_j \in P | d(F_i(n_0)) > 1/(I \cdot h)\}$. The panel (a) of Figure 2 shows the whistle candidates W_D detected for the sound-clip analysed in Figure 1. We used a temporal window of 10.7 ms ($M = \lceil 10.7e^{-3} \cdot f_s \rceil$) and 50% overlapping in the computation frequency $F_i(n_0)$ and bandwidth $B_i^2(n_0)$.

A different way of selecting the whistle candidates from the pyknogram is by selecting only those central frequency estimates $F_i(n_0)$ that have an associated bandwidth $B_i^2(n_0)$ that is lower than a given bandwidth threshold BW_h. We call those candidates $W_B = \{\mathbf{p} \in P | B_i^2(n_0) < BW_h\}$. This technique produces slightly sharper whistle candidate estimates (compare the continuous line red circle region in Panels (a) and (b) of Figure 2. However, empirical tests on real signals have shown that in order to have a number of whistle candidates throughout the duration of the whistle that is comparable to the one obtained with the previous technique, we need to use a high bandwidth threshold. As an example, Panel (b) of Figure 2 was obtained using $BW_h = BW_{Gabor}/4 = 250Hz$, which is used in the remainder of this work. The result when using W_B is a large number of random candidates that do not belong to real whistles (see the number of candidates in Panel (b) of Figure 2 which, in this example, is equal to 5132).

The final method proposed here for whistle candidates detection is the intersection of the candidates obtained using the two techniques, $W = \{W_D \cap W_B\}$. We have named this method Pyknogram-based Whistle Candidate Detection (PWCD). It combines the advantages of the two techniques, providing an accurate and even distribution of the whistle candidates while at the same time providing sharp whistle estimates (see Panel (c) of Figure 2).

2.2. Spectrogram-based whistle candidate detection

The proposed PWCD technique was compared with a traditional spectrogram peak-based candidate whistle detection method. Of all of the different spectrogram-based extraction techniques, we chose to compare it with the one described in Gillespie et al. (2013). This technique is included in PamGuard, a popular software developed to automatically identify vocalisations of marine mammals, which has been used many times as a benchmark. In his work, D. Gillespie proposed a six-step process for whistle detection and tracking: click removal, spectrogram calculation, spectrogram noise removal (median filter, average subtraction, and Smoothing Kernel (SK)), 2D thresholding, connection of regions, and separation of crossing whistles. We only implemented the first four steps of the process for the comparison since those are the ones that are specifically related to the detection of whistle peaks or candidates as named here. The first four steps in Gillespie et al. (2013), did not fully optimise the step of candidate detection (similarly to what happens with the PWCD technique), since subsequent whistle contour tracking stages would deal with a moderate false alarm rate of candidates at this intermediate stage. When computing the spectrogram (as we did with the PWCD), we used the same temporal window length (10.7 ms) and the 50% overlap that was used in Gillespie et al. (2013). The spectrogram was computed using a hamming window.

Using the aforementioned technique, we analysed the sound-clip used in Figure 1. The whistle peaks detected between 3 and 22 kHz are shown as a scatter plot in Panel (d) of Figure 2. One of the first things that can be observed when looking at this particular example is that when using W_D, the whistle candidates detected have less spectral resolution than the spectrogram has (compare the panels (a) and (d) of Figure 2). However, in some weak regions of the whistle with a fast sweep rate (compare the dashed line red circle), W_D gives more uniform candidates and an overall higher number than the spectrogram does. This might be due to several factors that affect the spectrogram-based candidate detection: first, the click removal step described in Gillespie et al. (2013) might also remove whistle candidates that follow a path close to a vertical slope; second, the candidate extraction in the spectrogram relies on a thresholding process that sometimes fails to extract candidates in the lower intensity parts of a whistle. Decreasing the spectrogram threshold should provide a general increase in whistle point candidates that subsequent whistle extraction and tracking stages will refine.

2.3. Spectrogram- and pyknogram-based whistle extraction

Whistle candidates detected using the PWCD and the spectrogram were used by the Gaussian mixture probability hypothesis density (GM-PHD) for whistle extraction as described in Gruden and White (2016). The aim was to compare how the different whistle candidates behaved when used by the same whistle extraction algorithm. Different metrics were computed for the PWCD and the spectrogram before and after the whistle was extracted.

The GM-PHD algorithm, which was downloaded from Gruden (2022), worked with the same settings used here: 10.7 ms window size, 50% overlap, and a time increment of 5.35 ms (see Sections 2.1, and 2.2). Similarly to what is done for extracting whistles with spectrogram candidates using the GM-PHD, the PWCD candidates were used to fit a quadratic polynomial and the maximum was obtained by using the fitted polynomial.

In this work, whistle extraction metrics were calculated only for real whistles (Section 3.3) and not for simulated whistles (Sections 3.1 and 3.2).

2.4. Metrics for comparing the performance of the pyknogram and the spectrogram approaches

In order to systematically study the number of whistle candidates that each technique succeeds in recovering (recall) when compared with the theoretical instantaneous frequencies as well as the errors committed in the process (precision), we need to define how these metrics are computed. The main idea is summarised in Figure 3.

The frequency resolution of the PWCD is connected to the bandwidth of the Gabor filterbank BW_{Gabor} and its overlap Δ_W, whereas the frequency resolution of the spectrogram is connected to time window length. As a result, for the same sampling frequency, the two techniques do not provide the same number of whistle candidates per Hertz. Additionally, the number of candidates in both techniques is very likely to vary (and not always in a similar way) due to many factors such as whistle slope, noise, bandwidth, etc.

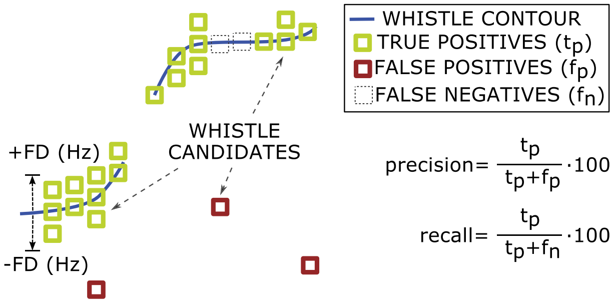

Figure 3. Precision and recall of whistle candidates can be obtained by looking for the true positive, the false positives, and the false negatives. FD is the maximum frequency deviation from the ground truth whistle contour.

To achieve a fair comparison, we merged adjacent candidates within each time frame and counted that merged group of candidates as 1. We considered the group of candidates to be a valid whistle match if it failed within a frequency deviation (FD) of $\pm$ 350 Hz of the theoretical instantaneous frequency (the whistle contour) (Roch et al. 2011).

3. Analysis and results

3.1. Whistle candidate detection in simulated overlapping whistles

One of the advantages of the pyknogram, as is the case for the proposed PWCD method, is its ability to perform well in different overlapping sound scenarios. In order to study this, we performed simulations with two synthetic whistles crossing with different slopes. Consider the sum of two noisy whistles modelled as described in Equation (1) and Equation (2). The two whistles have instantaneous frequencies that vary linearly as given by $f_1(t) = f_0 - \Delta_f \frac{T-2t}{2T}$ and $f_2(t) = f_0 + \Delta_f \frac{T-2t}{2T}$ with $T = 0.25$ sec. The instantaneous amplitude of both whistles is constant and equal to one ($a_1(t) = a_2(t) = 1$), and the initial phase is randomly distributed in the range $[0, 2\pi]$. The two synthetic whistles have a SNR of −6 dB in the whistle band due to the added pink noise, $n(t)$. In this study, the SNR was computed as the ratio of whistle power with respect to additive noise power in the bandwidth of interest (3–22 kHz). Figure 4 shows an example of a spectrogram for simulated crossing whistles, with $f_0 = 12\,kHz$ and $\Delta_f = 9\,kHz$. In this situation, the

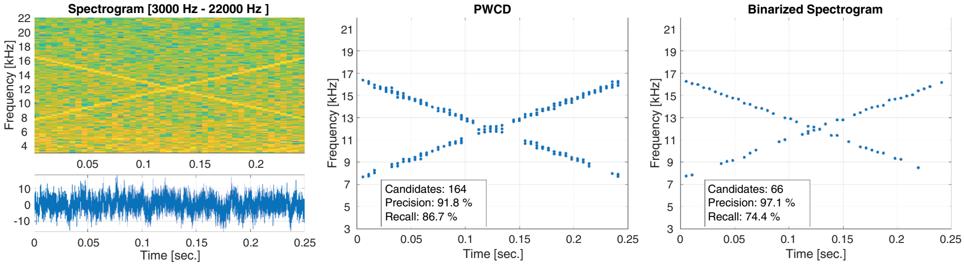

Figure 4. An example of detecting whistle candidates in simulated crossing whistles using the Spectrogram and the PWCD techniques. Top panels $\Delta_f = 2\,kHz$. Bottom panels $\Delta_f = 10\,kHz$.

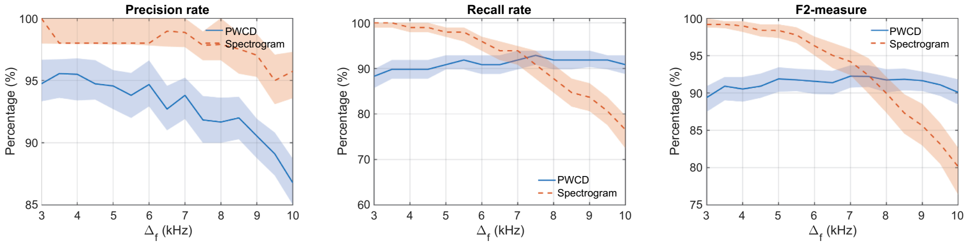

Figure 5. Evolution of the precision, recall, and F2-measure for the PWCD and the spectrogram-based whistle candidate detection in crossing whistles as Δ_f increases. The results were obtained for 500 Monte Carlo runs with a SNR = −6 dB.

Table 1. F2-measure computed over 500 Monte Carlo runs for simulated crossing whistles when the Signal to Noise Ratio (SNR) changes. The simulations are obtained for three different Δ_f values.

	F2-measure					
	$\Delta_f = 2kHz$		$\Delta_f = 5kHz$		$\Delta_f = 9kHz$	
SNR/dB	Spectrogram	PWCD	Spectrogram	PWCD	Spectrogram	PWCD
−3	99.6%	97.7%	99.5%	98.2%	88.0%	97%
−4	98.9%	96.1%	98%	97.6%	74.6%	96.7%
−5	97.3%	92.4%	93%	96.4%	56.6%	95.7%
−6	92.7%	86.5%	80.5%	93.9%	38.8%	92.9%

PWCD achieves lower precision than the spectrogram. However, it is capable of retrieving more whistle candidates (higher recall) than the spectrogram does.

With the aim to study how both techniques perform for different Δ_f, we performed 500 Monte Carlo runs when $f_0 = 12\,kHz$ and Δ_f varies from 2 kHz to 9 kHz. This value of maximum sweep frequency $9kHz/0.25s = 36kHz/s$ is a realistic range of what it is measured for some species, such as the common dolphin ($33.5kHz/s$) according to (Gannier et al. 2010). The SNR was kept constant at −6 dB. The precision and recall rates of the detected candidates were computed as described within $\pm 350Hz$ of the theoretical instantaneous frequency. The left panel of Figure 5 shows how the spectrogram achieves higher precision than the PWCD does. The recall rate of both techniques is shown in the middle panel of Figure 5. Although there is a considerable difference in the recall metric for small Δ_f (almost 100% for the binarised spectrogram and around 87% for the PWCD), this difference is reduced as Δ_f increases. Thus, for high Δ_f, the PWCD shows higher recall than the spectrogram does (76% for the binarised spectrogram vs. 91% for the PWCD). With respect to recall, the PWCD shows a more stable behaviour when Δ_f changes than the spectrogram does.

We used the Fβ-measure, computed as $F_\beta = (1 + \beta^2) \cdot \frac{precision \cdot recall}{(\beta^2 \cdot precision) + recall}$, as a single-score metric summarising both precision and recall (Christen et al. 2023). We specifically used the F2-measure, which gives more weight to recall and less to precision. The selection of this measure was decided based on the fact that some of the false positives can be easily reduced at a later stage by looking for candidates that can be connected with previous or posterior candidates. However, there is always an extra difficulty in recovering the whistle contour if the number of false negatives becomes too high. The F2-

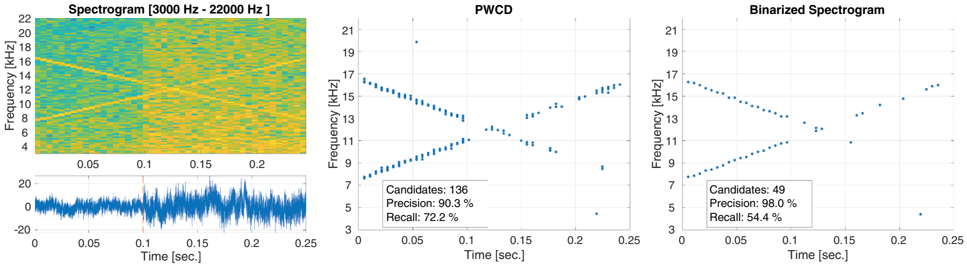

Figure 6. Example of the spectrogram and the PWCD technique to extract whistle candidates when a sudden change of $\Delta SNR = 1.5$ dB in the noise floor occurs. The sudden change occurs at 0.1 sec. And is marked with a vertical red-dashed line in the temporal representation of the signal.

measure as Δ_f increases is shown in the right panel of Figure 5. The figure illustrates that the behaviour for large Δ_f is better for the PWCD than it is for the binarised spectrogram. However, for small Δ_f, the spectrogram gives better results.

We also computed the F2-measure of the spectrogram and the PWCD methods for different SNR and different Δ_f. The results are shown in Table 1. The table shows that, for $\Delta_f = 2kHz$, the spectrogram produces candidates with higher F2-measures than the PWCD. However, for $\Delta_f = 9kHz$, the PWCD produces higher F2-measures than the spectrogram. For $\Delta_f = 5kHz$ for very low SNR (SNR = −5 dB and SNR = −6 dB), the PWCD outperforms the spectrogram. For SNR = −4 dB and SNR = −3 dB, the situation changes and the spectrogram produces better candidates than the PWCD. It is important to highlight that these results are obtained before doing any type of whistle extraction or tracking stage that will reduce the false positives in both methods.

3.2. Whistle candidate detection in simulated sudden increases of the noise floor

Cetacean recordings often contain unexpected acoustic events that may lead to sudden rises in the noise floor: increases in wind velocity, rainfall, and anthropogenic sources are some examples (Roch et al. 2011). These noise floor changes produce a considerable number of false positives in many of the spectrogram peak-based whistle extraction techniques. In order to see how the proposed PWCD works in this situation, we performed a variation of the previously described simulation. As before, whistles were simulated using AM-FM components, but this time the whistle register was divided into two parts at a random time instant. The sudden changes were obtained by increasing/decreasing the SNR of the first and second parts by a factor of $\pm\Delta SNR$. The left panel in Figure 6 shows an example where the SNR is increased by 1.5 dB at $t = 0.1$ sec for crossing whistles with $\Delta_f = 9kHz$. The middle and right panels show the whistle candidates detected by the PWCD and the binarised spectrogram, respectively. In Figure 6, the PWCD provided a larger number of candidates than the spectrogram did and considerable good precision.

As before, we performed 500 Monte Carlo runs. However, this time we changed the ΔSNR from 0 to 3 dB in order to study the behaviour of the two techniques. In each one of the runs, the slope of the two crossing whistles was randomly changed (Δ_f was

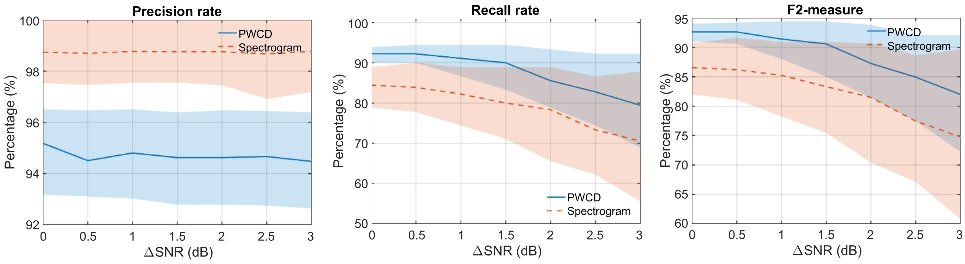

Figure 7. Evolution of the precision, recall, and F2-measure for the PWCD and the spectrogram-based whistle candidate detection when ΔSNR changes. The results were obtained for 500 Monte Carlo runs with a SNR = −6 dB and Δ_f randomly changing between 3000–9000 Hz.

uniformly distributed between 3000 and 9000 Hz). The results are shown in Figure 7. It can be concluded that although the number of properly detected candidates (precision rate) was higher for the spectrogram technique than it was for the PWCD, the number of possible candidates extracted (recall rate) was lower. The overall behaviour of the PWCD was slightly better than that of the spectrogram.

With regard to the number of candidates obtained for simulated signals, as can be observed, the PWCD has some advantages with respect to whistle slope and changes in noise floor levels over the binarised spectrogram. It is important to remember that all of the candidates from the PWCD and the spectrogram-based technique will be fed into a tracking algorithm, at a posterior stage. It is after this stage, where real precision and recall curves should be evaluated, as done in Subsection 3.3. Nevertheless, taking into account that the exact same tracking algorithm will be used later, a prior study of the precision and recall helps to determine the scenarios where one of the techniques might potentially work better than the other.

3.3. *Whistle candidate detection and whistle extraction in real scenarios*

With the aim of testing the performance of the proposed PWCD technique, different complex acoustic scenarios were selected. All of the scenarios come from the recordings of an acoustic campaign that was done in the Bay of Biscay on the 20th of June, 2019 as part of the RAGES EU project. The location corresponds to an area of high marine

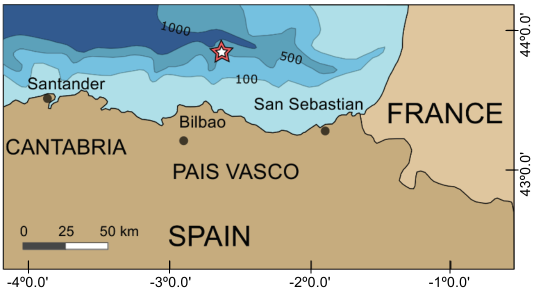

Figure 8. Approximate location of the RAGES deployment and location of the recordings used (marked with a star on the map).

mammal density, which is shown in Figure 8. The recording site is within close range of a gas platform (6 km), named Gaviota, where sudden noises occur during its operation. The signals were acquired with the SAMARUC passive acoustic monitoring device (Universitat Politècnica de València) ar:lar19 (Lara et al. 2019, 2020), equipped with a Cetacean Research hydrophone (C57) and a sampling frequency $f_s = 192$ kHz. The hydrophone depth was 414 meters. Although there was no visual confirmation, habitat-based density models of cetacean species (Camilo et al. 2018) along with signature whistles allowed us to identify bottlenose dolphins, striped dolphins, and common dolphins as the main species vocalising in the recordings. The selected scenarios shown in Figure 9 contain vocalisations of the aforementioned species and can be described as: overlapped whistles coming from many striped and common dolphins vocalising simultaneously (1); isolated whistles with low SNR mainly due to a sudden increase in background noise: (2a) for anthropogenic noise (2b) for ambient noise; and, a combination of isolated and multiple overlapped whistles in the presence of echolocation clicks (3).

Figure 9 shows an example of what whistle extraction looks like in the three different scenarios selected (previously described) using the spectrogram and the PWCD. The figure shows that while maintaining good precision, the number of frequencial whistle components extracted is higher for the PWCD than it is for the binarised spectrogram (higher recall). Visual comparison shows that the components are more uniformly distributed over the whistle contour in the PWCD than they are in the binarised spectrogram. At posterior stages, this should benefit the process of tracking the whistle contour or disambiguating individual whistles when they cross one another.

3.3.1. Ground truth and results

In order to establish how well the two methods compare when extracting the whistle candidates, we need to compare the output of the two methods with the whistle contours extracted by a trained analyst (ground truth information). For that purpose and similarly to what was done in (Roch et al. 2011), we created a custom software in MATLAB to allow the bioacoustics data analyst to interactively specify the whistle contours by clicking on a few whistle points. Cubic spline data interpolation was shown to the analyst to check that the manual annotated whistle matched the spectrogram contour (instantaneous frequency). This process was replicated for each and every one of the whistles in the scenario dataset. Even though a huge effort was made to record accurate ground truth information, there are always some errors and missed whistle fragments. However, the metrics previously used in the simulations were designed so that these errors affect both the spectrogram technique and the PWCD technique in a very similar way. We analysed over two thousand whistles in the three proposed scenarios. The metrics obtained in each one of the scenarios are the same ones already used for the simulations (precision, recall, and F2-measure). We computed the metrics for the spectrogram-based and pyknogram-based candidates (Table 2). Taking into account that the candidates were also used for whistle extraction using the GM-PHD, Table 2 shows the metrics with the SK and without the SK ($\neg SK$). Be aware that the peak candidate extraction using Gillespie's method, as implemented in the GM-PHD (Gruden 2022), did not use the SK step. Finally, we computed the metrics after the whistle extraction using the GM-PHD from the candidates without the SK (Table 3).

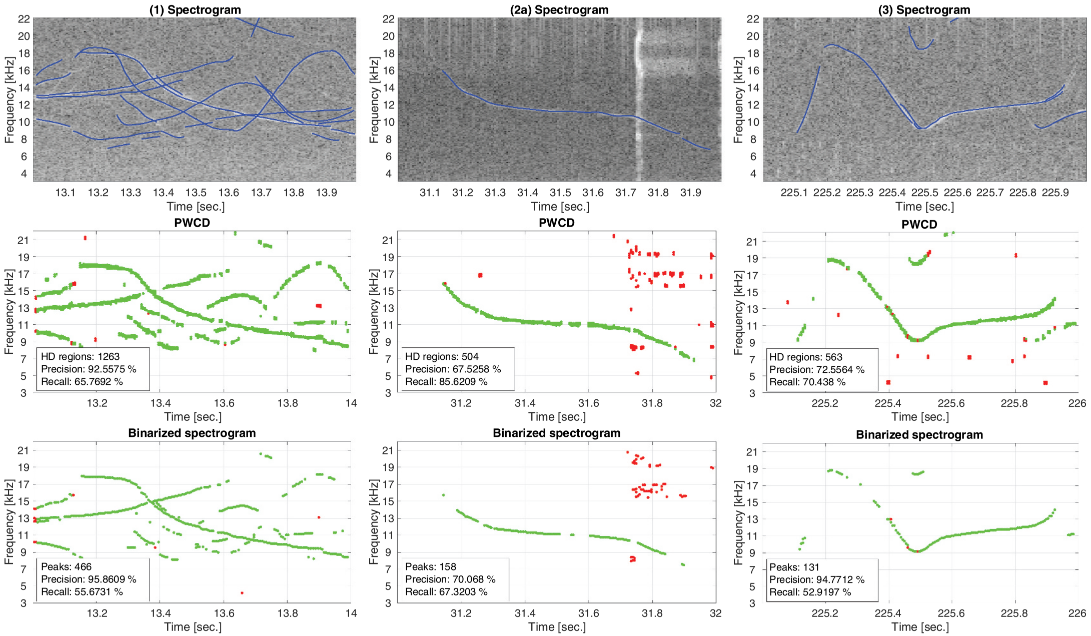

Figure 9. Three examples of how the proposed technique behaves in the three described scenarios: left-column scenario (1), middle-column scenario (2a), and right-column scenario (3). The green points correspond to extracted whistle candidates that match the expert annotation (true positives); the red points indicate extracted whistle candidates that do not match the expert annotation (false positives).

Table 2. Precision, recall, and F2-measure metrics for whistle candidate detection using the spectrogram (with and without the smoothing kernel: *SK* and ¬*SK*, respectively) and the pyknogram in all the scenarios.

		Spectrogram						PWCD		
		Precision		Recall		F2-measure				
Whistle scenarios	Whistles #	% *SK*	% ¬*SK*	% *SK*	% ¬*SK*	% *SK*	% ¬*SK*	Precision %	Recall %	F2-measure %
(1) Multiple overlapping	966	96.0	67.3	53.8	77.8	58.9	74.6	91.6	62.9	67
(2a) Isolated with anthropogenic noise	340	48.9	20.9	51.1	79.9	50.3	47.6	41.2	66.7	56
(2b) Isolated with ambient noise	83	83.9	17.3	43.5	77.5	48.7	43.4	53.1	56.7	54.5
(3) Combined with echolocation clicks	634	92.2	35.7	61.8	83.5	65.2	61.3	76	72.1	72.1
Total	2023	80.2	35.3	52.6	79.6	55.8	56.7	65.5	64.6	62.4

3.3.2. Discussion on candidate detection and whistle extraction in real scenarios

The analysis of the candidate detection metrics (Table 2) shows that, with the SK, the PWCD achieved better recall metrics than the spectrogram. The combined F2-measure was also higher for the PWCD compared to the spectrogram. However, the precision was always higher for the spectrogram than it was for the PWCD. This is in agreement with the results obtained in the simulations where the pyknogram provided better recall and F2-measure when the noise floor increased suddenly and whistles overlapped with high Δ_f.

When the SK was not used (¬*SK* column), recall was higher for the spectrogram than it was for the PWCD, and precision was lower for the spectrogram compared to that of the PWCD. The PWCD achieved a better F2-measure than the spectrogram for all of the scenarios, except the scenario of multiple overlapping whistles (1).

In summary, the overall percentage of whistle candidate detection (F2-measure) in the Bay of Biscay recordings increased by 6.6% with the SK and by 5.7% without the SK when using the PWCD when compared to the spectrogram (62.4% vs 55.8% or vs 56.7%). Although this is far below the 20% improvement that some authors claim the pyknogram improves the extraction of tonal components in speech, this small improvement might be worth it in some challenging scenarios.

Once the whistle extraction using the GM-PHD method was done (Table 3), the results completely changed. The GM-PHD was able to discard spectrogram candidates that did

Table 3. Precision, recall, and F2-measure results after the GM-PHD whistle contour extraction in the three scenarios.

	Spectrogram			PWCD		
Whistle scenarios	Precision %	Recall %	F2-measure %	Precision %	Recall %	F2-measure %
(1) Multiple overlapping	88.0	66.0	69.6	92.8	53.2	57.6
(2a) Isolated with anthropogenic noise	21.7	64.8	46.3	34.9	53.2	48.2
(2b) Isolated with ambient noise	90.9	63.7	67.8	94.0	48.6	53.8
(3) Combined with echolocation clicks	79	75	76	85	67	70
Total	69.9	67.4	64.9	76.7	55.5	57.4

not belong to real whistles, increasing the precision at the cost of a reduction in the recall. Something similar happened in some scenarios for the PWCD: an increase in the precision and a decrease of the recall. However, the GM-PHD did the extraction task better for the spectrogram than it did it for the PWCD. The F2-measure, after whistle extraction, was higher for the spectrogram than it was for the PWCD in all the scenarios, except for the scenario (2a). This makes sense if we take into account that the variances of the GM-PHD and system noise covariance matrix was optimised to work with the settings of the spectrogram frequency resolution (which was 93.75 Hz for 10.7 ms). The PWCD, on the other hand, had a frequency resolution given by the Gabor filterbank of 500 Hz.

4. Conclusions

We have presented an alternative method for whistle candidate detection based on the pyknogram. The technique, named PWCD, has been shown to have better combined F2-measure than the spectrogram in some challenging scenarios such as multiple overlapping whistles and regions with anthropogenic noise. This behaviour is due to the fact that the density distribution of the pyknogram points is less affected by the presence of broadband noise and sudden increases in the noise floor. Monte Carlo simulations were done to illustrate this behaviour (Subsections 3.1 and 3.2).

The PWCD has some additional advantages over the spectrogram that may be attractive in some situations. First, time and frequency resolution in the PWCD can be controlled separately by the window length and the bandwidth of the Gabor filters, respectively. This can be useful for the analysis of certain short cetacean calls. Second, the PWCD is capable of extracting a larger number of whistle regions in high slope crossing whistles than the spectrogram does.

The application of the proposed PWCD in a real dataset containing more than 2000 ground-truth annotated whistle sounds demonstrated that, before the whistle extraction stage, the candidates obtained with the PWCD technique outperformed the candidates obtained with the spectrogram. In the best of the scenarios, the PWCD technique obtained an accuracy of 67% compared to 58.9% (measured using the F2-measure), which is an increase of slightly over 8%. The overall accuracy result in the combined scenarios was also increased by approximately 6.6% when using the PWCD with respect to the spectrogram.

The results changed when the candidates were used to extract the contour using the GM-PHD method and the spectrogram outperformed the PWCD in most of the scenarios. The final accuracy (F2-measure) of the whistle extraction was equal to 57.4% for the PWCD and 64.9% for the spectrogram. Even though the GM-PHD implementation used was specifically trained to work with the spectrogram settings, in the scenario of isolated whistles with low SNR, the PWCD was able to obtain slightly better accuracy than the spectrogram (48.2% vs 46.3%). Candidates obtained using the PWCD might have some potential to achieve better whistle tracking in adverse scenarios when paired with the appropriate whistle extraction techniques. Being able to extract whistle contours in noisy scenarios is an important research line that may help to develop automatic alert systems. As a result, a thorough study of the different whistle extraction techniques and their adaptation for working with the proposed PWCD technique is a future line of work.

Acknowledgements

This work was partially supported by the following: LIFE PortSounds (grant agreement LIFE20-ENV/ES/000387), the European Commission – DG ENV (grant agreement 110661/2018/794607/SUB/ENV.C2), the Spanish Institute of Oceanography (IEO), AZTI, and the Oceanografic of Valencia. We would like to thank the crew of the ships who participated in the deployment and recovery of the acoustic campaign.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the Directorate General for Environment (DG ENV) [110661/2018/794607/SUB/ENV.C2]; European Climate, Infrastructure and Environment Executive Agency (CINEA) [LIFE20-ENV/ES/000387].

ORCID

Ramón Miralles http://orcid.org/0000-0003-0039-2553

References

Camilo S, Gerrodette T, Louzao M, Valeiras J, García S, Cerviño S, Pierce GJ, Santos MB. 2018. Assessing the environmental status of the short-beaked common dolphin (Delphinus delphis) in north-western Spanish waters using abundance trends and safe removal limits. Prog Oceanogr. 166:66–75. doi: 10.1016/j.pocean.2017.08.006.

Christen P, Hand DJ, Kirielle N. 2023. A review of the f-measure: its history, properties, criticism, and alternatives. ACM Comput Surv. 56(3):1–24. doi: 10.1145/3606367.

Cornel I, Cédric G, Yann S, Jerome IM. 2010. Analysis of underwater mammal vocalisations using time–frequency-phase tracker. Appl Acoust. 71(11):1070–1080. doi: 10.1016/j.apacoust.2010.04.009.

Delprat N, Escudie B, Guillemain P, Kronland-Martinet R, Tchamitchian P, Torresani B. 1992. Asymptotic wavelet and Gabor analysis: extraction of instantaneous frequencies. IEEE Trans Inf Theory. 38(2):644–664. doi: 10.1109/18.119728.

Gabor D. 1946. Theory of communication. J Electr Eng. 93(3):429–457. doi: 10.1049/ji-3-2.1946.0076.

Gannier A, Fuchs S, Quebre P, Oswald JN. 2010. Performance of a contour-based classification method for whistles of mediterranean delphinids. Appl Acoust. 71(11):1063–1069. doi: 10.1016/j.apacoust.2010.05.019.

Gillespie D, Caillat M, Gordon J, White P. 2013. Automatic detection and classification of odontocete whistles. J Acoust Soc Am. 134(3):2427. doi: 10.1121/1.4816555.

Gruden P. 2016. GM-PHD whistle detector [software]. GitHub repository (commit: 4e105cd); [accessed 2022 Jul 16]. https://github.com/PinaGruden/GMPHD_whistle_contour_tracking

Gruden P, White PR. 2016. Automated tracking of dolphin whistles using gaussian mixture probability hypothesis density filters. J Acoust Soc Am. 140(3):1981–1991. doi: 10.1121/1.4962980.

Gruden P, White PR. 2020. Automated extraction of dolphin whistles - a sequential Monte Carlo probability hypothesis density approach. J Acoust Soc Am. 148(5):3014–3026. doi: 10.1121/10.0002257.

Hsu MK, Sheu JC, Hsue C. 2011. Overcoming the negative frequencies- instantaneous frequency and amplitude estimation using osculating circle method. J Mar Sci Technol. 19(5): doi: 10.51400/2709-6998.2165.

Johansson AT, White PR. 2011. An adaptive filter-based method for robust, automatic detection and frequency estimation of whistles. J Acoust Soc Am. 130(2):893–903. doi: 10.1121/1.3609117.

Kershenbaum A, Roch MA. 2013. An image processing based paradigm for the extraction of tonal sounds in cetacean communications. J Acoust Soc Am. 134(6):4435. doi: 10.1121/1.4828821.

Lara G, Bou-Cabo M, Esteban JA, Espinosa V, Miralles R. 2019. Design and application of a passive acoustic monitoring system in the Spanish implementation of the marine strategy framework directive. 6th International Electronic Conference on Sensors and Applications; Nov 15–30; MDPI. p. 1–7.

Lara G, Bou-Cabo M, Esteban JA, Espinosa V, Miralles R. 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(5353):1–13. doi: 10.3390/s20185353.

Mallawaarachchi A, Ong SH, Chitre M, Taylor E. 2008. Spectrogram denoising and automated extraction of the fundamental frequency variation of dolphin whistles. J Acoust Soc Am. 124(2):1159–1170. doi: 10.1121/1.2945711.

Maragos P, Loupas T, Pitsikalis V. 2002. On improving doppler ultrasound spectroscopy with multiband instantaneous energy separation. Proceedings Int'l Conf. DSP-2002; Jul 1-3; Santorini, Greece

Mellinger DK, Martin SW, Morrissey RP, Thomas L, Yosco JJ. 2011. A method for detecting whistles, moans, and other frequency contour sounds. J Acoust Soc Am. 129(6):4055–4061. doi: 10.1121/1.3531926.

Parzen E. 1962. On estimation of a probability density function and mode. Ann Math Stat. 33(3):1065–1076. doi: 10.1214/aoms/1177704472.

Potamianos A, Maragos P. 1996. Speech formant frequency and bandwidth tracking using multiband energy demodulation. J Acoust Soc Am. 99(6):3795–3806. doi: 10.1121/1.414997.

Roch MA, Scott Brandes T, Patel B, Barkley Y, Baumann-Pickering S, Soldevilla MS. 2011. Automated extraction of odontocete whistle contours. J Acoust Soc Am. 130(4):2212–2223. doi: 10.1121/1.3624821.

Shokouhi N, Hansen JHL. 2017. Teager–Kaiser energy operators for overlapped speech detection. IEEE/ACM Trans Audio, Speech, Language Process. 25(5):1035–1047. doi: 10.1109/TASLP.2017.2678684.

Vijayan K, Raghavendra Reddy P, Sri Rama Murty K. 2016. Significance of analytic phase of speech signals in speaker verification. Speech Commun. 81:54–71. doi: 10.1016/j.specom.2016.02.005.

White P, Hadley M. 2008. Introduction to particle filters for tracking applications in the passive acoustic monitoring of cetaceans. Canadian Acoustics. 36(1):146–152. https://jcaa.caa-aca.ca/index.php/jcaa/article/view/2004

Yousefi M, Shokouhi N, Hansen JHL. 2018. Assessing speaker engagement in 2- person debates: overlap detection in United States presidential debates. Proceedings of the Interspeech 2018; Hyderabad, India, p. 2117–2121.