Entropy-based detection and classification of Bryde's whale vocalizations

O. P. Babalola[1], O. O. Ogundile[2], A. M. Usman[3,*]

[1]French-South African Institute of Technology, Department of Electrical, Electronic, and Computer Engineering, Cape Peninsula University of Technology, Bellville, South Africa
[2]Department of Industrial Technical Education, Tai Solarin University of Education, Ogun State, Nigeria
[3]Department of Electrical and Electronics Engineering, University of Ilorin, Kwara State, Nigeria

ABSTRACT: Investigation of a cetacean species distribution, periodicity, and population is made possible by long-term monitoring of their vocalizations. In this study, a sample entropy (SampEn) method is proposed for the automatic detection of Bryde's whale calls. Additionally, the k-means approach is presented to automatically classify the whale signals to whale calls and noise depending on the signal-to-noise ratio instead of using a manual threshold approach. The performance of the proposed detection scheme is compared to the traditional dynamic time warping (DTW) algorithm. The detection performance comparison result shows that the proposed SampEn scheme effectively detects Bryde's whale calls in the presence of ambient noise with a higher accuracy and lower error rate performance compared to the template-based DTW algorithm, achieving 90.73% accuracy and 8.17% error rate.

KEYWORDS: Bryde's whale, Cetacean, Detection, DTW, SampEn, *k*-means

[Received Feb. 16, 2025; Revised Mar. 15, 2025; Accepted Mar. 22, 2025] Print ISSN: 0189-9546 | Online ISSN: 2437-2110

I. INTRODUCTION

Bryde's whales, like other marine mammals, require conservation and protection from threats brought on by a variety of factors including rising ocean temperatures from climate change, disturbances and accidents, pollution, and other damaging human activities Usman et al (2020). According to Penry et al. (2016); and Child et al. (2017), the inshore Bryde's whale species are specifically classified as vulnerable and need to be conserved to avoid extinction. Hence, there is a need to develop effective monitoring schemes to examine the existence and numbers of species, observe the potential implications of human activity, and take measures for ecological preservation.

Recently, passive acoustic monitoring (PAM) has been demonstrated as a successful method for observing whales without interfering with the animal's behaviour. The advancement of autonomous underwater sensors, which are typically incorporated into monitoring equipment for predicting long-term patterns and seasonal variation and evaluating the consequences of environmental degradation and human activity, has expanded the potential of PAM. Traditionally, whale vocalizations from recordings were manually identified and reported by either observing the whale signal on a spectrogram or having marine ecology experts listen to the sounds. However, the manual detection approach is susceptible to human errors, and thus, different automatic detection methods have been developed for the detection of whale vocalizations through sound Usman and Versfeld (2022).

In the literature, template-based spectrogram analysis and dynamic time warping (DTW) have been employed for whale vocalization detection. These methods involve the segmentation of whale vocalizations in the acoustic data into whale calls and ambient noise Ogundile and Versfeld (2020). The spectrogram-based approach, such as spectrogram cross-correlation Mellinger and Clark (2000); Gillespie (2004); Dennis et al. (2011); Roch et al. (2011) distinguishes the vocalization frequency and time variations, displayed in the spectrogram as graphically unique patterns.

Moreover, the DTW algorithm enables either time expansion or compression to improve sound signal matching. This is performed by creating a dissimilarity matrix for the sounds, which is used to compare the contours of the sound. In Buck and Tyack (1993), DTW was introduced to match the contours and offer a numerical similarity measurement between the contours in the narrow-band signature whistles obtained from bottlenose dolphin recordings. The authors in Brown and Miller (2007); and Brown et al. (2006) adapted an automated DTW algorithm for killer whale detection. Likewise, in Ogundile and Versfeld (2020), the DTW and linear predictive coding (LPC) schemes were utilized for the automatic detection of Bryde's whale vocalizations.

Besides, machine learning (ML)-based detection and classification methods such as support vector machines (SVMs) Roch et al. (2008), hidden Markov model (HMM); Ogundile et al. (2021, 2022); Babalola and Versfeld (2024); Usman and Versfeld (2022), Gaussian mixed model (GMM) Brown and Smaragdis (2009), neural networks (NNs) and its variants Jia Jiang et al. (2019); Bermant et al. (2019) have been introduced for whale vocalization detection and classification in recent years. The supervised ML schemes have been shown to be useful for automatic whale detection. However, the scheme requires that the training dataset must be manually annotated.

*Corresponding author: usman.am@unilorin.edu.ng ***doi:*** http://dx.doi.org/10.4314/njtd.v22i1.3404

Segmenting whale vocalizations may be difficult in practice due to the existence of other sources of vocalizations such as fish and snapping shrimp vocalizations coupled with the ambient noise. Additionally, each animal's species-specific vocalizations have different amplitudes, time spans, and spectral-temporal patterns, making manual annotation and supervised ML training more difficult Usman et al (2020). The continuous data of whale recordings are typically very large, resulting in a high computational complexity during the training and detection phases in machine learning. Thus, necessitating the use of feature extraction methods such as dynamic mode decomposition (DMD) Ogundile et al. (2021), principal component analysis Ogundile et al. (2022), empirical mode decomposition (EMD) Ogundile et al. (2020) were adapted with the machine learning algorithms for detection and classification purposes.

Statistical entropy measures, such as permutation entropy, approximate entropy, and sample entropy have been generally used to assess the unpredictability of time-series signals such as measuring electrical activities in the brain Richman and Moorman (2000), heart rate characterization Moorman et al. (2011); Keller et al. (2014), and other physiological activities based on electroencephalograms (EEGs). The high dynamic and nonlinear characteristics of the physiological time series are similar to whales' vocalizations in noisy auditory situations. Consequently, permutation entropy (PermEnt) and statistical complexity were effectively incorporated for the automatic detection of fish vocalizations in Siddagangaiah (2018). This method of detection achieved a high accuracy performance of detection that was not affected by the presence of the majority of the anthropophonics_and geophones in the recordings. Furthermore, Shishidhar et al. Siddagangaiah et al. (2020) expanded on the findings in Siddagangaiah (2018) for the detection of dolphin whistles and clicks using PermEnt and SampEn, respectively. However, the approach utilizes a fixed threshold approach for binary classification of the whale calls.

This article presents the use of the sample entropy (SampEn) approach as a computationally efficient alternative to machine learning for the detection of Bryde's whale calls in the presence of other biophonies and anthropophonics. The proposed SampEn method estimates the variation in randomness in the time series dataset without prior knowledge of the source generating the data, making it suitable for broad-scale and long-term analysis. An additional contribution of this study is the introduction of a *k*-means approach to automatically compute the class average of the entropy measure according to the signal-to-noise ratio. This approach eliminates the need for an entropy threshold that is manually determined by the user. The performance of the proposed SampEn detection scheme is theoretically evaluated using recorded data from the MobySound.org library[1] Mellinger and Clark (2006). Furthermore, the proposed SampEn detector scheme is compared with the conventional dynamic time warping (DTW) detector using accuracy and error rate as performance indicators.

The remainder of this work is as follows. A general description of the whale vocalizations and different call types is presented in Section 2.1. The method of data collection and recordings is further discussed in Section 2.2. In Section 2.3, the proposed entropy-based detector is introduced as a basis for the detection process. Section 2.4 emphasizes Bryde's whale detection process, consisting of a dynamic time-warping (DTW) approach and the proposed SampEn method. In Section 3, experiments are performed on continuous recordings. The significance of selecting the appropriate parameter for the proposed algorithm is highlighted, and the performance of the proposed algorithm in terms of accuracy and error rate is compared with DTW. Section 4 discusses and evaluates the proposed detection and classification algorithm with reference to similar studies in the literature. The study is concluded in Section 5.

II. MATERIALS AND METHODS

A. *Bryde's Whale Calls*

Bryde's whales belong to the family of baleen whales, which are part of the group of great whales such as the blue and humpback whales. Studies classify the two genetically distinct subspecies of Bryde's whales as the smaller *Balaenoptera edeni* species, predominantly found in coastal waters, and the larger *Balaenoptera edeni brydei* species, which resides in deep waters away from the shore (Constantine, Kershaw). Bryde's whales are widely distributed globally, specifically in tropical and subtropical oceans with warm temperatures ranging between 16.1°C and 22.2°C. Some groups of Bryde's whales migrate seasonally, moving away from or towards the equator, while others are local residents.

The Bryde's whale vocalizations have been recorded in regions such as the Gulf of Mexico Rosel (2014); Rice et al. (2014); Sirovic et al. (2014), coastal areas of Southern Brazil´ Figueiredo and Simao (2014); Lodi et al. (2015); Goncalves et al. (2015); Tardin et al. (2017), Hauraki Gulf of New Zealand Constantine et al. (2015); Izadi et al. (2018), Eastern Tropical Pacific Constantine et al. (2018), and the Gulf of California Omora et al. (2015); Kerosky et al. (2012). The majority of these studies focused on population description, geographic variations, biological and evolutionary traits, and associated taxonomy. Moreover, only a few Rice et al. (2014); and Omora et al. (2015) have explored the potential acoustic properties of the whale species.

Different Bryde's whale vocal repertoires have been reported in the literature, depending on the geographical location. Bryde's whales generate varying low-frequency sounds, such as pulses, growls, pulsed moans, unpulsed moans, and moans, ranging from 20 to 950 Hz Omora et al. (2015); Oleson et al. (2003). Additionally, the length, amplitude modulation, and

harmonic content (if any harmonic exists) of the emitted sounds vary based on animal activities such as communication, obstacle avoidance, and other events. Generally, Bryde's whale vocalizations are generated as continuous signals with basic frequencies under 60 Hz, lasting for about 1 − 3 seconds.

The authors Oleson et al. (2003) recorded Bryde's whale calls in the Eastern Tropical Pacific, Southern Caribbeans, and Western North Pacific, Japan. The different whale call types were categorized based on their frequency and duration attributes. Six distinct Bryde's whale call types, known as the Be1 to Be6 calls, were discovered in the Eastern Tropical Pacific. The Be1 call was the most recorded in the area. The recorded call was divided into two parts: 1) Be1 started with an upper-frequency component of around 37 Hz with a variable duration, and 2) Be1 started at a lower frequency of about 23 Hz and downswept to about 4 Hz over the course of approximately 2 seconds. The Be2 call was discovered to have a frequency in the range of 40 Hz to 48 Hz with a duration of 1.3 seconds. The frequency range of the Be3 call was between 24.4 Hz and 26.9 Hz while the call lasted for 1.7 seconds. The Be4 call had a fundamental frequency of 59.5 Hz and an upper frequency of 60.2 Hz with a duration of about 1.4 seconds. The upsweep Be5 call had a frequency range of 26.0 Hz to about 26.8 Hz, lasting for 1.76 seconds. The Be6 call type contained a cluster of downswepts such that each sweep consisted of an upper frequency ranging from about 208 Hz to 57.1 Hz with an average call duration of 2.83 seconds. In Sirovic et al. (2014), a similar Bryde's whale call type was recorded in the Gulf of Mexico and identified as the Be9 call.

Furthermore, the Southern Caribbean was found to have just one type of Be7 call in Oleson et al. (2003). The whale call, which occurred every 2.8 minutes and lasted between 0.8 and 2.5 seconds, had a frequency component of 44.3 Hz. Besides, the authors reported the Be8a and Be8b call types in the Western North Pacific coast of Japan. The Be8a whale calls are short pulsed moans that last for less than 0.6 seconds with frequencies in the range of 43 Hz to 48 Hz. Each moan averaged 51.6 pulses per second and contained about 13 pulses. However, the non-pulsed downswept Be8b call type had a short duration with a frequency ranging from 137 Hz to 192 Hz.

B. Bryde's Whale Call Recordings

In this study, a sample of Bryde's whale calls from the MobySound.org library Mellinger and Clark (2006) is considered for the purpose of testing and analysis. The Bryde's whale calls on MobySound library were recorded in the eastern tropical Pacific using autonomous hydrophones. Each sound file has a corresponding annotated table showing the start and end times (seconds), least and upper frequencies (Hz), and signal-to-noise ratio (dB) for the whale calls. For a detailed description of the dataset, including the number of recordings, total duration of whale calls, noise levels, and noise filtering applied, please refer to Mellinger and Clark (2006); and Oleson et al. (2003). Figure 1 shows the time series and spectrogram view of the whale calls, having a fundamental frequency ranging from 24.23 ± 22.35 Hz and an upper frequency ranging from 29.85 ± 27.44 Hz. The duration of calls is between 1.47 seconds and 2.28 seconds. Table 1 provides an overview of the key features of the Bryde's whale dataset used in this study.

Table 1: Summary of Bryde's Whale Dataset Characteristics

Dataset Characteristics	**Description**
Whale Species	Bryde's whale (Balaenoptera brydei)
Number of Recordings	48
Total Duration	2 hours 45 minutes
Sampling Rate	6 kHz
Fundamental frequency	24.23 ± 22.35 Hz
Upper frequency	29.85 ± 27.44 Hz
Noise Levels	Variable, with some recordings containing background noise from ships and wind
Noise Filtering	Bandpass filtering was applied to reduce background noise

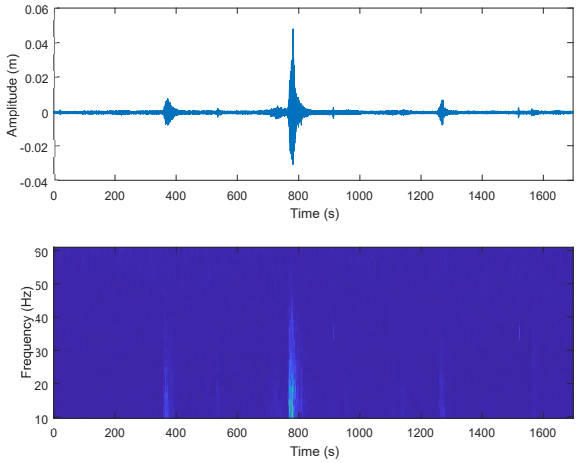


Figure 1: Time-series and spectrogram views of Bryde's whale call

C. Description of the Proposed SampEn-based Detector

SampEn is a nominal probability-based method that analyzes patterns and their regularity over specific time intervals without self-matching. Let $x(j) = \{x_i; t = 1,2, \dots, N\}$ represent a time series of N sample points, and $Y_m(i)$ be a vector with embedding dimension m, given as

$$Y_m(i) = [x_i, x_{i+1}, \dots, x_{i+m-1}],\ i = 1,2, \dots, N - m + 1. \quad (1)$$

Suppose two such vectors exist, a maximum difference between their associated scalar elements is determined using the Chebyshev distance metric as

*Corresponding author: usman.am@unilorin.edu.ng
doi: http://dx.doi.org/10.4314/njtd.v22i1.3404

$$d[Y_m(i), Y_m(j)] = \left|\left|Y_m(i) - Y_m(j)\right|\right|, \; i \neq j \\ = max(|Y_m(i+p) - Y_m(j+p)|), \quad (2)$$

where p is an integer that ranges from 0 to $m-1$.

Let $\mathcal{B}_i$ and $\mathcal{A}_i$ represent the number of vectors $Y_m(j)$ and $Y_{m+1}(j)$ within the tolerance for accepting matches r of $Y_m(i)$ and $Y_{m+1}(i)$ without self-matching, that is, i, j, respectively. For m sample points, the Heaviside step function of the distance is obtained as

$$D_m = d[Y_m(i), Y_m(j)] < r. \quad (3)$$

The likelihood that any vector $Y_m(j)$ is at a distance r from $Y_m(i)$ is represented by the correlation function

$$\mathcal{B}_i^m(r) = \frac{1}{N-m-1} \sum_{j=1}^{N-m} D_m^j. \quad (4)$$

Hence, the correlation density is determined by averaging all the vectors as

$$\mathcal{B}^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r). \quad (5)$$

This signifies the probability that two patterns will be similar for m points.

The same procedure is conducted for the $m+1$ sample space to obtain the step function of the distance as

$$D_{m+1} = d[Y_{m+1}(i), Y_{m+1}(j)] < r, \quad (6)$$

the correlation function as

$$\mathcal{A}_i^{m+1}(r) = \frac{1}{N-m-1} \sum_{j=1}^{N-m} D_{m+1}^j, \quad (7)$$

and the correlation density

$$\mathcal{A}^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \mathrm{c}A_i^m(r), \quad (8)$$

which demonstrates the probability that two patterns will be similar for $m+1$ points.

Therefore, the sample entropy of parameters $m, r,$ and, N is represented as

$$\mathrm{SampleEn}(m, r, N) = \lim_{N\to\infty} \left[-\ln\left(\frac{\mathcal{A}^m(r)}{\mathcal{B}^m(r)}\right)\right], \quad (9)$$

which is further estimated by the statistical sample entropy as

$$\mathrm{SampleEn}(m, r, N) = -\ln\left(\frac{\mathcal{A}^m(r)}{\mathcal{B}^m(r)}\right). \quad (10)$$

Given that $\mathcal{B}$ is the total number of vector matches of length m and $\mathcal{A}$ is the total number of forward matches of length $m+1$, Equation (10) can be expressed as

$$\mathrm{SampleEn}(m, r, N) = -\ln\left(\frac{\mathcal{A}}{\mathcal{B}}\right), \quad (11)$$

where

$$\mathcal{A} = \left[\frac{(N-m-1)(N-m)}{2}\right] \cdot \mathcal{A}^m(r), \quad (12)$$

and

$$\mathcal{B} = \left[\frac{(N-m-1)(N-m)}{2}\right] \cdot \mathcal{B}^m(r). \quad (13)$$

Equation (11) directly applies the correlation integral method suggested in Grassberger and Procaccia (1983) since the estimate is computed as the negative natural logarithm of the conditional probability that two patterns within a tolerance value r for m points stay within r of one another at the succeeding point exclusive of self-matching. The conditional probability of $\mathcal{A}$ and $\mathcal{B}$ is typically less than 1 since there are always fewer matches ($m+1$ points) than there are vector possibilities (m points).

D. Bryde's Whale Detection

In this study, the detection performance of the proposed SampEn is compared to the DTW detector. The detectors are evaluated using recordings of the confirmed Bryde's whale from the MobySound library.

E. DTW-Based Detection

In Itakura (1975); Ogundile and Versfeld (2020), DTW was presented as a method for perfectly matching two time series based on their temporal differences. The method of detection entails nonlinearly modifying the time series to produce a match or similarity between them. Given two time series $x(1)$ and $x(2)$ with N and M data points respectively. The DTW technique compares similarities between points $x(1)_i : i = 1,2,\dots,N$ and $x(2)_j : j = 1,2,\dots,M$ by creating an $(N+1) \times (M+1)$ distance matrix

$$\mathrm{C_{i,j}} = \left|\mathrm{x(1)_i - x(2)_j}\right| + \mathrm{m} \begin{cases} C_{i-1,j-1}, \\ C_{i-1,j}, \\ C_{i,j-1}, \end{cases} \quad (14)$$

where $\mathrm{C_{i-1,j-1}}$represents the match, $\mathrm{C_{i-1,j}}$ is the insertion, and $\mathrm{C_{i,j-1}}$indicates deletion. To obtain the alignment path between the two sequences, a trace back through the matrix is performed from $\mathrm{C_{N,M}}$ to $\mathrm{C_{0,0}}$. Note that the value of $\mathrm{C_{N,M}}$ signifies the overall cost of aligning two sequences.

Let $\mathcal{T}_{h,t}$ and $\Gamma_{\{h,t\}}$ be matrices containing templates from the aurally and visually recognized MobySound Bryde's whale signal. The templates of the whale signal have h number of

samples with varying lengths q, depending on the call durations, t. Hence, the confirmed Bryde's whale signal template matrix is given as

$$\mathcal{T}_{a,\mathcal{b}} = \begin{pmatrix} [x(1)_1, x(1)_2, \dots, x(1)_t] \\ [x(2)_1, x(2)_2, \dots, x(2)_t] \\ \vdots \\ [\mathrm{x(h)}_1, \mathrm{x(h)}_2, \dots, \mathrm{x(h)}_\mathrm{t}] \end{pmatrix}. \quad (15)$$

The template is warped according to Equation (14) to obtain a $h \times h$ dissimilarity hollow matrix

$$\mathcal{D} = \begin{bmatrix} 0 & \mathcal{N}_{1,2} & \cdots & \mathcal{N}_{1,\hbar} \\ \mathcal{N}_{2,1} & 0 & \cdots & \mathcal{N}_{2,\hbar} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{N}_{\hbar,1} & \mathcal{N}_{\hbar,2} & \cdots & 0 \end{bmatrix}, \quad (16)$$

where each component of $\mathcal{D}$ represents the similarity of each pair of samples. Thereafter, the maximum value at each column is selected to form a $h \times h$ row vector

$$\widehat{\mathcal{D}} = \begin{bmatrix} \max(\mathcal{D}_{i,1}) & \max(\mathcal{D}_{i,2}) & \cdots & \max(\mathcal{D}_{i,\hbar}) \end{bmatrix}, \quad (17)$$

where $i = 1, 2, \dots h$. Subsequently, the DTW algorithm computes the average energy of the sample u in the template as

$$= \sum_{q=1}^{t} u_t^2. \quad (18)$$

The E is compared to a set threshold δsuch that $E \geq$ δsignifies the occurrence of Bryde's whale call within the window frame. In addition, the similarities between the signal satisfying the threshold condition and the h template samples are derived from Equation (14) as

$$\mathcal{D}(\mathfrak{s}) = \begin{pmatrix} \mathcal{D}(\mathfrak{s})_1 & \mathcal{D}(\mathfrak{s})_2 & \cdots & \mathcal{D}(\mathfrak{s})_h \end{pmatrix}. \quad (19)$$

As a result, a score $\mathcal{Z}$ is derived by counting the number of times each entry in $\mathcal{D}(\mathfrak{s})$ has values that are less than or equal to the associated column of $\widehat{\mathcal{D}}$. Similar to Ogundile and Versfeld (2020), $\mathcal{Z}$ is compared to a selected reliability parameter $0 \leq \mathcal{R} \leq 1$. Suppose that $\mathcal{Z} \geq \mathcal{R} * \mathcal{Z}$ the window size is stored as the Bryde's whale call.

F. SampleEn-Based Detection

Let $x(j)$: $\{j = 1,2, \dots, N\}$ represent the annotated dataset with Nsample points from the MobySound Bryde's whale recording. The Bryde's whale signal is divided into data frames Γ based on the window step size $\mathfrak{s}$ of the signal such that each frame contains L data points. Suppose $f_1, f_2, \dots, f_L$ are the data points in the frame $\Gamma(\mathrm{j})$: $\mathrm{j} = 1,2, \dots, \mathfrak{s}$. A sound signal matrix $\mathcal{M}(\mathrm{j})$ is formed for each Γ as

$$\mathcal{M}(j) = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{m+1} \end{pmatrix} = \begin{pmatrix} f_1, & f_2, & \dots, & f_{L-m} \\ f_2, & f_3, & \dots, & f_{L-m+1} \\ \vdots & \vdots & \dots & \vdots \\ f_{m+1}, & f_{m+2}, & \dots, & f_L \end{pmatrix}. \quad (20)$$

The distance measures between the vectors $Y_1, Y_2, \dots, Y_m$ in $\mathcal{M}(j)$are computed without self-matching for each of the scalar elements $f_1, f_2, \dots, f_L$, using Equation (2). Hence, the Heaviside functions of the distances are derived based on Equation (3). Thereafter, the likelihood that a vector Y_1 is at a distance r from another vector $Y_2, Y_3 \dots, Y_m$ is computed using the correlation function in Equation (4). Therefore, the conditional probability $\mathcal{B}^m(r)$ that two Bryde's whale calls are similar for m points, without matching any data point with itself, is obtained by Equation (5). This match between two vectors of length m is referred to as a *similarity match*.

Similarly, the distance measures and Heaviside step functions are computed between the $m + 1$ vectors for each of the scalar elements $f_1, f_2, \dots, f_L$without self matches. Subsequently, the likelihood that a vector Y_1 is at a distance r from another vector $Y_2, Y_3 \dots, Y_{m+1}$ is computed using the correlation function in Equation (7). Also, the conditional probability that two Bryde's whale signals match $\mathcal{A}^m(r)$, without self-matching, for $m + 1$ data points are obtained based on Equation (8). Here, the match between two vectors of length $m + 1$ is known as a *forward match*. Thus, the sample entropy is calculated as the negative natural logarithm of the ratio between the *forward match* and *similarity match* using Equation (11).

Given that the Bryde's whale signals are obtained in the presence of other undesired sounds such as fish vocalization, snapping shrimps, and so on, an automated threshold of call detection is presented to classify the SampEn measure into whale calls X_c and any other wide-band signal as noise X_s.

The class assignment is performed iteratively based on the assignment function

$$A[\mathcal{S}] = \begin{cases} 1, \text{ SampEn} \in X_c \\ 0, \text{ SampEn} \in X_s, \end{cases}. \quad (21)$$

Suppose β_c is the whale signal mean and β_s is the noise signal mean, the assignment function allocates the class with the closest mean as

$$\mathcal{S} = \begin{cases} 1, \text{ SampEn} \leq \beta_c + \alpha(\beta_s - \beta_c) \\ 0, \text{ otherwise} \end{cases}. \quad (22)$$

Note that α is chosen as a very small value since the presence of a call lowers the SampEn measurement. Thus, the class means are updated at each iterative step till the difference in the mean values equals zero.

*Corresponding author: usman.am@unilorin.edu.ng
doi: http://dx.doi.org/10.4314/njtd.v22i1.3404

A summary of the SampEn detection process is presented in Algorithm 1:

Algorithm 1 SampEn Detection Process

1: **Input**: $[x(j): \{j = 1,2, \dots, N\}, Fs, m, r]$

2: **Output**: $\mathcal{S}$
3: Select $\mathfrak{s}$ based on the duration of each call in x

4: Predefine t using $\mathfrak{s}$ intervals

5: **for** $j \leftarrow 1$ to length$(t) - 1$ do

6: Split x into data frames
7: $\Gamma(j) \leftarrow x\big((t(j) * Fs)$ to $(t(j+1) * Fs)\big)$
8: Create data matrix

9: **for** $k \leftarrow 1$ to $m + 1$ do
10: $\mathcal{M}(j) \leftarrow xk$ to $-m + k - 1$

11: **end for**

12: Compute distance measures $\mathcal{C}_m$ and $\mathcal{C}_{m+1}$ for m and $m + 1$ based on Equation (2). Exclude self-matching
13: Calculate Heaviside function D_m and D_{m+1} for m and $m + 1$ using Equation (3)

14: Compute correlation function $\mathcal{B}_i^m(r)$ and $\mathcal{A}^{m+1}(r)$
15: Obtain the correlation density $\mathcal{B}_i^m(r)$ and $\mathcal{A}^{m+1}(r)$
16: SampleEn$(m, r, N) = -\ln\left(\frac{\mathcal{A}}{\mathcal{B}}\right)$

17: **end for**
18: Perform class assignment $A[\mathcal{S}]$
19: Update β_c and β_s
20: Check convergence

G. Performance metrics

The ability of the proposed SampEn to detect Bryde's whale vocalization is demonstrated based on the accuracy and error rate performance metrics. Let *TP*, *TN*, *FP*, and *FN* denote the numbers of true positives, true negatives, false positives, and false negatives from predictions, respectively. Details of the four parameters used for the metrics evaluation are explained in Usman and Versfeld (2022). Accuracy is computed as

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}, \tag{23}$$

while the error rate is given as

$$\mathcal{E} = \frac{FN + FP}{TP + FP + TN + FN}. \tag{24}$$

Additionally, the performance of the proposed SampEn detector is compared to the DTW using the same performance metrics: Acc and $\mathcal{E}$.

III. EXPERIMENTS

A. Dependence of SampEn Detection on m

It is important to select the appropriate parameters m and r to define the length of the vectors and the tolerance of the points to one another, respectively. From the literature, m is generally selected as 2, which shows more of the data dynamics and r is chosen as a factor of the standard deviation σ of the data, ranging between 0.2 and 0.5 Richman and Moorman (2000). Therefore, the dependence of the proposed SampEn detector on m is analyzed in this section.

Figures 2 and 3 depict the spectrogram views of Bryde's whale recording from the MobySound library, as well as the performance of the proposed SampEn on the whale signal at $r = 0.2 \times \sigma$, a value commonly used in the literature for sample entropy analysis, and embedding dimensions $m = 1, 2, 3,$ and 4, which were evaluated for their impact on the sample entropy estimates for detecting Bryde's whale calls. SampleEn values below the *k*-means threshold value $\mathcal{H}$ indicate a proper correlation with the whale calls. For $m = 1$, the embedding dimension is too small, such that the short templates lack sufficient predictive information. Consequently, the probability of the forward match is underestimated, and the SampEn decoder performs poorly. Moreover, for $m = 2$, the SampleEn detector successfully tracks all calls that appeared three times in the spectrogram view despite continuous noise and snapping shrimps.

Typically, the conditional probability of the forward match should increase by conditioning on a longer vector and consequently, improve the performance of the detector. Nevertheless, Figure 3 shows that increasing the value of m to 3 does not significantly improve the performance of the detector. Rather, the computational complexity of the SampEn detector increases by incrementing the value of m. Furthermore, if the value of m increases to 4, the SampEn detector tracked more noise signals as the desired calls, which is depicted by more SampEn values dip below $\mathcal{H} = 1.0$. The false detections are due to the noise from measurement defects, which reduces the number of signal vector matches as m increases. If m continues to increase, the templates will be mostly determined by chance instead of dynamic matching. Hence, the value of $m = 2$ is maintained in this study.

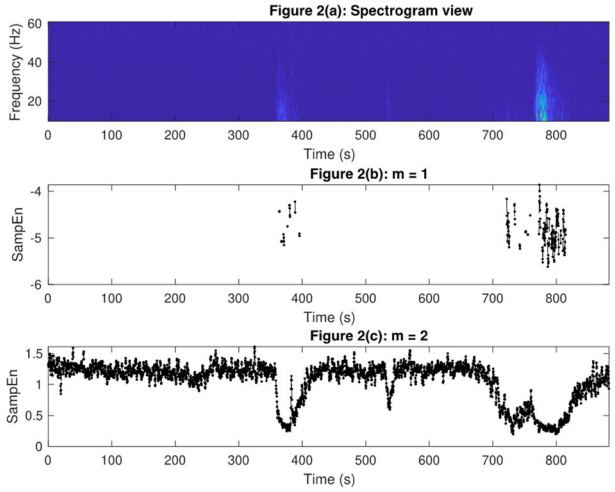


Figure 2: SampEn Detection Performance, m = 1,2

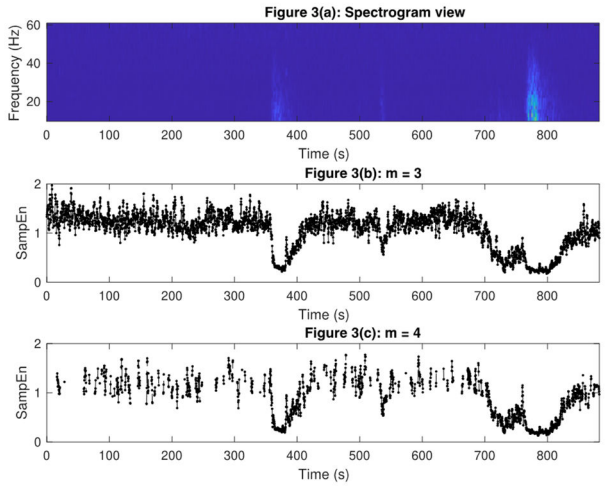


Figure 3: SampEn Detection Performance, m = 3,4

B. Detection Performance Comparison

Table 2 shows the accuracy and error rate performance of the SampEn and DTW algorithms for the detection of Bryde's whale calls. The process of whale vocalization detection begins by defining a moving window step size $\mathfrak{s}$ for the calls. Based on the call duration, the continuous data measured for the calls are empirically selected. From Table 2, it is observed that the DTW detector exhibits the highest *Acc* and the lowest $\mathcal{E}$ performance at $\mathfrak{s}$ = 2.2830. More so, the SampEn detector yields the highest *Acc* and least $\mathcal{E}$ at $\mathfrak{s}$ = 0.4714.

The performance of the proposed SampEn and the DTW detectors on Bryde's whale vocalization is shown in Figure 4. The basis for evaluating the performance of the DTW detector and the proposed SampEn detector is the manually annotated signal in Figure 4(b). The result indicates that the DTW detector has more false positive detection compared to the SampEn detector since DTW incorrectly indicates the presence of whale calls at 8.8 and 25.19 minutes where the calls were not present in the annotated signal.

Table 2: Accuracy and Error Rate Detector's performance at different values of s

MobySound Bryde's whale calls

DTW			SampleEn		
$\mathfrak{s}$	*Acc* (%)	$\mathcal{E}$ (%)	$\mathfrak{s}$	*Acc* (%)	$\mathcal{E}$ (%)
0.1478	87.14	11.71	0.0148	89.91	9,11
1.4780	88.92	9.85	0.4714	90.73	8.17
2.2830	88.90	9.72	0.2283	90.02	8.35

IV. DISCUSSION

The proposed SampEn detection method is evaluated with existing DTW. Both detection techniques are experimented on continuous recordings of Bryde's whale vocalizations obtained from the MobySound.org library. The recording includes the whale calls as well as background ocean noise, anthropophonics from boat engines, and biophonicsp from shrimps that snap at the same frequency range as the whale's vocalization. As opposed to the proposed SampEn with *k*-means, the template-based DTW shows more tendency to match background noise and snapping shrimps that occur in the same frequency range with the whale sounds instead of the desired signal. Thus, increases the likelihood of detection and classification errors. This is a typical defect of the template-based matching algorithms.

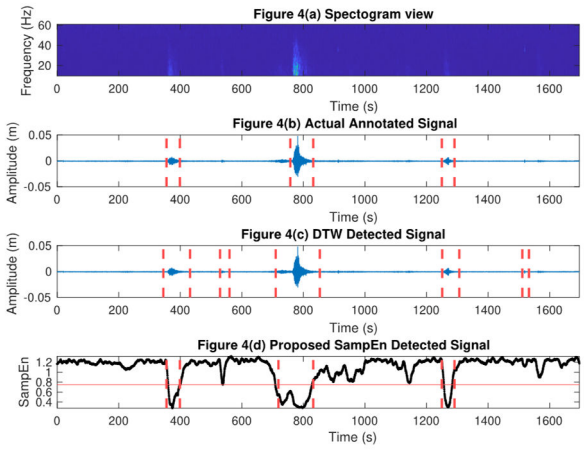


Figure 4: Detection Performance Comparison of SampEn and DTW on the confirmed MobySound whale vocalization

Supervised machine learning algorithms involve some pre-processing procedures, including noise filtering and data smoothing, to remove unwanted signals at a certain amplitude and frequency to improve whale vocalization. The preprocessing processes may remove some important vocalization, thus, leading to a false negative. Although bandpass filtering in Bryde's whale frequency range would be advised to attenuate frequencies outside the whale's frequency range to provide the best results, the proposed SampEn

*Corresponding author: usman.am@unilorin.edu.ng
doi: http://dx.doi.org/10.4314/njtd.v22i1.3404

detection technique does not necessitate significant pre-processing. The proposed detection approach is not machine learning based and therefore, it may not be the best choice to evaluate the results with machine learning methods. Hence, the findings in this study are compared with the similarity-based DTW technique.

To the best of the author's knowledge, this is the only study that discusses the detection of Bryde's whale calls in the presence of snapping shrimps. Other studies such as Ogundile et al. (2021, 2022) categorized snapping shrimps and anthropogenic noise as unwanted noise, which could bias the models. The results presented in this study show that the proposed SampEn is robust to most of the noise sources, including snapping shrimps, and fish vocalizations. According to the result in Figure 4, the Bryde's whale calls detected by SampEn had a high accuracy of 9.85% and a low error rate performance of 8.17% in the presence of snapping shrimps.

The SampEn detector does not need any dimensionality reduction technique and also does not require vocalization training as required by supervised machine learning. Thus, reducing the computational complexity of the algorithm. The time computational requirements for processing the 30-minute Bryde's whale data recording at a sampling frequency of 249 Hz was approximately 1.3 seconds on MATLAB. As shown in Table 2, the moving window step $\mathfrak{s}$ of the algorithm was empirically chosen as 1.4780 seconds for detecting the Bryde's whale calls.

For different call types occurring at specific times and frequencies in the recording, the approach of Siddagangaiah et al. (2020) could not determine or categorize the precise number of co-occurring calls or call types since the entropy detection threshold was manually obtained. However, the *k*-means scheme was introduced in this study to automatically determine the threshold of entropy measures. This enables the automatic classifications of calls during long-time monitoring. The proposed detection and classification algorithm will ensure a proper understanding of the seasonal occurrence of the Bryde's whale, leading to the enhancement of conservation measures.

V. CONCLUSION

In this study, a simple and effective sample entropy approach was presented as an alternative to machine learning techniques for the automatic detection of Bryde's whale calls. This is due to the high computational complexity that the machine learning techniques exhibited in the literature. The proposed SampEn algorithm analysed patterns and their regularity over specific time intervals without self-matching. Thus, the proposed algorithm is not prone to errors from background noise and snapping shrimps. Additionally, the *k*-means algorithm was introduced to avoid manually setting thresholds on the entropy measures. The *k*-means method automatically classifies the whale signals as calls and no calls based on the signal-to-noise ratio. The whale vocalization detection performance of the proposed SampEn with the *k*-mean approach was verified using data from the MobySound library. Experimental results showed that the proposed SampEn with the *k*-means scheme detected Bryde's whale call from MobySound more accurately with less error rate compared to the template-based DTW algorithm. The proposed algorithm utilized a minimal embedded dimension of $m = 2$, which yielded an accuracy performance of 90.73% and an error rate performance of 8.17% at a moving window step of $\mathfrak{s}$ = 1.4780 while achieving a reduced computational complexity. Hence, the proposed SampEn detection technique with *k*-means classification is a suitable performance and cost-efficient alternative for the practical detection and classification of Bryde's whale vocalization. Future work will focus on evaluating the robustness of the proposed algorithm to high noise conditions and exploring its applicability to other bioacoustics signals. Additionally, investigating the use of other clustering algorithms and comparing their performance with the k-means approach will be of interest.

AUTHORS CONTRIBUTIONS

O. P. Babalola: Conceptualization; Data curation; Analysis; Investigation; Methodology; Project administration; Software; Validation; Visualization; Writing – original draft; Writing review & editing. **O. O. Ogundile:** Conceptualization; Supervision; Analysis; Investigation; Methodology; Software; Validation; Writing - review & editing. **A. M. Usman:** Analysis; Investigation; Methodology; Project administration; Writing review & editing.

ACKNOWLEDGEMENTS

This work was supported in part by research funding from the National Research Foundation (NRF) of South Africa under Grant Numbers 116036 and 138764. The opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of NRF. The authors also gratefully acknowledge the support received from Stellenbosch University through a subcommittee research award granted during the curation phase of the project.

REFERENCES

Babalola, O.P. and Versfeld D. J. J. (2024). Wavelet-based feature extraction with hidden Markov model classification of antarctic blue whale sounds. Ecological Informatics 80, 102468.

Bermant, P.; Bronstein, M.; Wood, R.; Gero, S. and Gruber D. (2019). Deep machine learning techniques for the detection and classification of sperm whale bioacoustics. Scientific Reports 9, 1–10.

Brown, J.; Hodgins-Davis, A. and Miller, P. (2006). Classification of vocalizations of killer whales using dynamic time warping. The Journal of the Acoustical Society of America 119, EL34–40.

Brown, J. and Miller, P. (2007). Automatic classification of killer whale vocalizations using dynamic time

warping. The Journal of the Acoustical Society of America 122, 1201–7.

Brown, J. C. and Smaragdis P. (2009). Hidden Markov and Gaussian mixture models for automatic call classification. The Journal of the Acoustical Society of America 125, EL221–EL224.

Buck, J. R. and Tyack, P. L. (1993). A quantitative measure of similarity for Tursiops truncatus signature whistles. The Journal of the Acoustical Society of America 94, 2497–2506.

Child, M.; Roxburgh, L.; Do Linh San, E. and Raimondo, D., Davies-Mostert, H. (2017). The Red List of Mammals of South Africa, Swaziland and Lesotho 2016.

Constantine, R.; Iwata, T.; Nieukirk, S. and Penry, G. (2018). Future directions in research on Bryde's whales. Frontiers in Marine Science 5, 333.

Constantine, R.; Johnson, M.; Riekkola, L.; Jervis, S.; Kozmian-Ledward, L.; Dennis, T.; Torres, L. G. and Aguilar de Soto, N. (2015). Mitigation of vesselstrike mortality of endangered Bryde's whales in the Hauraki Gulf, New Zealand. Biological Conservation 186, 149–157.

Dennis, J.; Tran, H.D. and Li, H. (2011). Spectrogram image feature for sound event classification in mismatched conditions. IEEE Signal Processing Letters 18, 130–133.

Figueiredo, L. and Simao, S. (2014). Bryde's whale (Balaenoptera edeni) vocalizations from southeast Brazil. Aquatic Mammals 40, 225–231225.

Gillespie, D. (2004). Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram. Canadian Acoustics 32, 39–47.

Goncalves, L.; Augustowski, M. and Andriolo, A. (2015). Occurrence, distribution and behaviour of Bryde's whales (cetacea: Mysticeti) off south-east Brazil. Journal of the Marine Biological Association of the United Kingdom -1, 1–12.

Grassberger, P. and Procaccia, I., (1983). Estimation of the Kolmogorov entropy from a chaotic signal. Phys. Rev. A 28, 2591–2593. URL:

Itakura, F. (1975). Minimum prediction residual principle applied to speech recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing 23, 67–72.

Izadi, S.; Johnson, M.; Aguilar de Soto, N. and Constantine, R., (2018). Night-life of Bryde's whales: ecological implications of resting in a baleen whale. Behavioral Ecology and Sociobiology 72.

jia Jiang, J.; ran Bu, L.; jie Duan, F.; quan Wang, X.; Liu, W.; bo Sun, Z. and yue Li, C. (2019). Whistle detection and classification for whales based on convolutional neural networks. Applied Acoustics 150, 169–178.

Keller, K.; Unakafov, A. and Unakafova, V. (2014). Ordinal patterns, entropy, and EEG. Entropy 16, 6212–6239.

Kerosky, S.M.; Sirovič, A.; Roche, L.K.; Baumann-Pickering, S.; Wiggins,´ S.M. and Hildebrand, J.A. (2012). Bryde's whale seasonal range expansion and increasing presence in the southern California bight from 2000 to 2010. Deep Sea Research Part I: Oceanographic Research Papers 65, 125–132. doi:https://doi.org/10.1016/j.dsr.2012.03.013.

Kershaw, F.; Leslie, M.; Collins, T.; Mansur, R.; Smith, B.; Minton, G.; Baldwin, R.; Leduc, R.; Anderson, R. C.; Brownell, R. and Rosenbaum, H. (2013). Population differentiation of 2 forms of Bryde's whales in the Indian and Pacific oceans. Journal of Heredity doi:10.1093/jhered/est057.

Lodi, L.; Tardin, R.; Hetzel, B.; Maciel, I.; Figueiredo, L. and Simao, S. (2015). Bryde's whale (cetartiodactyla: Balaenopteridae) occurrence and movements in coastal areas of southeastern Brazil. Zoologia (Curitiba Impresso) 32, 171–175. doi:10.1590/S1984-46702015000200009.

Mellinger, D. K. and Clark, C. W. (2000). Recognizing transient low-frequency whale sounds by spectrogram correlation. The Journal of the Acoustical Society of America 107, 3518–3529. doi:10.1121/1.429434.

Mellinger, D. K. and Clark, C.W. (2006). MobySound: A reference archive for studying automatic recognition of marine mammal sounds. Applied Acoustics 67, 1226–1242.

Moorman, J.; Carlo, W.; Kattwinkel, J.; Schelonka, R., Porcelli, P.; Navarrete, C.; Bancalari, E.; Aschner, J.; Walker, M.; Perez, J.; Palmer, C.; Stukenborg, G.; Lake, D. and O'Shea, T. (2011). Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: A randomized trial. The Journal of Pediatrics 159, 900–6.e1. doi:10.1016/j.jpeds.2011.06.044.

Ogundile, O. O.; Usman, A. M.; Babalola, O. P. and Versfeld, D. J. J. (2021). Dynamic mode decomposition: A feature extraction technique based hidden Markov model for detection of mysticetes' vocalisations. Ecological Informatics 63, 101306. doi:https://doi.org/10.1016/j.ecoinf.2021.101306.

Ogundile, O. O. and Versfeld, D. J. J. (2020). Analysis of template-based detection algorithms for inshore Bryde's whale short pulse calls. IEEE Access PP, 1–1. doi:10.1109/ACCESS.2020.2966254.

Ogundile, O. O.; Babalola, O. P.; Odeyemi, S. G. and Rufai, K. I. (2022). Hidden Markov models for detection of mysticetes vocalisations based on principal component analysis. Bioacoustics 0, 1–29.

Ogundile, O. O.; Usman, A.M. and Versfeld, D. J. J. (2020). An empirical mode decomposition-based hidden Markov model approach for detecting Bryde's whale pulsed calls. The Journal of the Acoustical Society of America 147, EL125–EL131.

Oleson, E.; Barlow, J.; Gordon, J.; Rankin, S. and Hildebrand, J., 2(003). Low-frequency calls of Bryde's whales. Marine Mammal Science - MAR MAMMAL SCI 19, 407–419.

Omora, L.; Romero-Vivas, E. and Urb, J. (2015). Calls of Bryde's whale (balaenoptera edeni) recorded in the Gulf of California. The Journal of the Acoustical Society of America 138.

Penry, G.; Findlay, K. and Best, P. (2016). A conservation assessment of Balaenoptera edeni. In Child MF, Roxburgh L, Do Linh San E, Raimondo D, Davies-Mostert HT, editors. The Red List of Mammals of South Africa, Swaziland and Lesotho. South African National Biodiversity

*Corresponding author: usman.am@unilorin.edu.ng

doi: http://dx.doi.org/10.4314/njtd.v22i1.3404

Institute and Endangered Wildlife Trust, South Africa. p. 8pp.

Rice, A.; Palmer, K.; Tielens, J.; Muirhead, C. and Clark, C. (2014). Potential Bryde's whale (Balaenoptera edeni) calls recorded in the northern Gulf of Mexico. The Journal of the Acoustical Society of America 135, 3066–76.

Richman, J. and Moorman, J., (2000). Physiological time-series analysis using approximate entropy and sample entropy. American journal of physiology. Heart and circulatory physiology 278, H2039–49.

Roch, M. A.; Scott Brandes, T.; Patel, B.; Barkley, Y.; Baumann-Pickering, S. and Soldevilla, M. S. (2011). Automated extraction of odontocete whistle contours. The Journal of the Acoustical Society of America 130, 2212–2223.

Roch, M.A.; Soldevilla, M. S.; Hoenigman, R., Wiggins, S. M. and Hildebrand, J.A., (2008). Comparison of machine learning techniques for the classification of echolocation clicks from three species of odontocetes. Canadian Acoustics 36, 41–47.

Rosel, P. E. and Wilcox, L.A., (2014). Genetic evidence reveals a unique lineage of Bryde's whales in the northern Gulf of Mexico. Endangered Species Research 25, 19–34.

Siddagangaiah, S. (2018). Complexity-entropy based approach for detection of fish choruses. The Journal of the Acoustical Society of America 144, 1692– 1693.

Siddagangaiah, S.; Chen, C. F.; Hu, W. C., Akamatsu, T.; McElligott, M.; Lammers, M. and Pieretti, N. (2020). Automatic detection of dolphin whistles and clicks based on entropy approach. Ecological Indicators 117, 106559.

Tardin, R.; Chun, Y.; Simao, S. and Alves, M. (2017). Modeling habitat use by bryde's whale balaenoptera edeni off southeastern brazil. Marine Ecology Progress Series 576.

Usman, A. M.; Ogundile, O. O. and Versfeld, D. J. J. (2020). Review of automatic detection and classification techniques for cetacean vocalization. IEEE Access 8, 105181–105206.

Usman, A.M. and Versfeld, D.J.J. (2022). Detection of baleen whale species using kernel dynamic mode decomposition-based feature extraction with a hidden Markov model. Ecological Informatics 71, 1–16.

Sirovic, A.; Bassett, H.; Johnson, S.; Wiggins, S. and Hildebrand, J. (2014). Bryde's´ whale calls recorded in the gulf of mexico. Marine Mammal Science 30.