Analysis of Template-Based Detection Algorithms for Inshore Bryde’s Whale Short Pulse Calls

OLAYINKA O. OGUNDILE, (Member, IEEE), AND DANIEL J. J. VERSFELD, (Member, IEEE)

1Department of Electrical and Electronics Engineering, Stellenbosch University, Stellenbosch 7600, South Africa
2Department of Physics and Telecommunication, Tai Solarin University of Education, Ijebu Ode 2118, Nigeria

Corresponding author: Olayinka O. Ogundile (ogundileoo@gmail.com)

This work was supported in part by the National Research Foundation of South Africa (NRF) under Grant 116036.

ABSTRACT

Marine mammals use sound for communication and echolocation within their ecosystems. The detection of these sounds is an important aspect of signal processing, such that we can estimate the spatial position and direction of arrival of these mammals, and have an understanding of their ecology. Passive acoustic monitoring (PAM) is widely used to understand marine mammal movement and vocal repertoire. In PAM, datasets are accumulated over days, months or years. Thus, it is impracticable to manually analyse the datasets because it is very large. This motivated the development of automated sound detection techniques for marine mammals, which most often varies depending on the vocal duration, frequency range and call type. In this paper, continuous recordings of Bryde’s whale (Balaenoptera edeni edeni) short pulse calls (< 3.1s long) were collected on a weekly basis from December 2018 to April 2019 on sighting of the individual in a single site in the endmost South-West of South Africa. The sound, previously not documented off South Africa, was observed on visual confirmation of the presence of inshore Brydes’s whale. In addition, the paper develops and analyses two automated template-based detection algorithms for this short pulse call, employing dynamic time warping (DTW) and linear predictive coding (LPC) techniques. These proposed template-based detectors are novel, as they have not being previously used in Bryde’s whale sound detection in the literature. When applied to the continuous recordings of the short pulse calls, the DTW-based and LPC-based detection algorithms obtained a sensitivity of 96.04% and 97.14% respectively for high signal-to-noise ratio (about 10dB above the ambient sound). Otherwise, for low SNR, the DTW-based and LPC-based detection algorithms obtained a sensitivity of 94.98% and 96.00% respectively. These detection algorithms exhibit low computational time complexity and can be modified to analyse the movement of obscure but vocal marine species instead of manual identification.

INDEX TERMS

Bryde’s whale, DTW, LPC, PAM, pulse call, sound detection.

I. INTRODUCTION

Over the years, increased human marine activities such as fishery and shipping have threatened the ecosystems of marine mammals [1]–[5]. As a result of these anthropogenic impacts, it is difficult to make informative decisions about the movement and spatial position of marine species [3]. Also, since marine mammals spend most of their time below water, it is difficult to visually observe and monitor these marine mammals. Therefore, passive acoustic monitoring (PAM) provides a valuable modality for study of marine mammal movement and distribution because (1) animals are very vocal, and (2) sound propagates much further in water. Besides, PAM can be used to collect datasets in outlying areas over days, months or years. More importantly, PAM is used in unfavourable weather conditions and it is suitable for the tracking of highly mobile marine mammals such as cetaceans [3], [6].

Bryde’s whales, also referred to as Eden’s whales are species of the order Cetacea. They are currently grouped as a single species called Balaenoptera edeni (B. edeni), where two subspecies have been suggested: Balaenoptera edeni edeni (B. e. edeni) and Balaenoptera edeni brydei (B. e. brydei) The B. e. edeni is the small, coastal form
of the B. edeni while B. e. brydei refers to the large, off-shore form [6]–[9]. Several studies have been recently carried out on the existence of Bryde’s whales from different geographic location, such as the Gulf of Mexico [9]–[11], Gulf of California [12], [13], Hauraki Gulf, New Zealand [3], [14], [15], Eastern Tropical Pacific [6], [16], and Southern Brazil [17]–[20]. While some of these literatures have described the population, spatial distribution, genetic and phylogenetic features, attributed nomenclature, a few have described the potential vocal repertoire of this marine species. In [10] and [12], the vocal repertoire of Bryde’s whales in the Gulf of Mexico and Gulf of California respectively are described. It is observed that the recorded Bryde’s whale calls from this region ranges from 50Hz to 1200Hz. However, aside describing the Bryde’s whale vocal repertoire, the authors in [10] and [12], do not identify method(s) to automatically detect these established calls as proposed in this paper.

Historically, in the extreme South-West of South Africa, Olsen [21] described a new species of whale. He named them Bryde’s whale after Mr. Johan Bryde who was at that time the Norwegian consul to South Africa. Although, in Olsen [21], the Bryde’s whales harvested off South Africa was labelled B. brydei, it was later revealed in Best [22] that there are two allopatric forms of Bryde’s whale off South Africa. Subsequently, it was affirmed that the B. brydei described by Olsen [21] include characteristics from the inshore and offshore forms of B. edeni [23]–[25]. Much recently, different literature have described the genetic and phylogenetic features of Bryde’s whales off South Africa [26], [27]. However, no work has previously documented the Bryde’s whales calls off South Africa. In this paper, we analyse a continuous recording of inshore Bryde’s whale (B. e. edeni) short pulse calls collected on sighting of the individual in a single site in the endmost South-West of South Africa. This observed pulse call is previously undocumented for B. e. edeni off South Africa and it can serve as an important contribution to the study of Bryde’s whale vocal repertoire off South Africa. In addition, we present the characteristic of the recorded short pulse call. Similar to other studies [10], the call is identified by observation and matching the spectral and temporal features described in closely related studies as in [12], which was conducted in the Gulf of California.

The datasets containing the B. e. edeni short pulse calls was accumulated over months. Therefore, it is impracticable to manually analyse all the collected datasets. In this regard, the paper develops two automated template-based detection algorithms for the short pulse call, employing dynamic time warping (DTW) [28] and linear predictive coding (LPC) [29], [30] techniques. Template-based detectors automatically recognise unknown sound signals only when a set of the signal is manually identified by an expert. This detection technique is widely used for sound detection in digital signal processing. As such, it has generally been adapted in animal vocalisation detection [31]–[33]. However, we emphasise that the template-based detection techniques have not been used in Bryde’s whales sound detection in the literature. Thus, these proposed template-based algorithms for the observed Bryde’s whale short pulse call is innovative and it produces good detection accuracy (sensitivity). As discussed in Section VI, when the proposed template-based detectors are applied to the continuous recordings of the Bryde’s whale short pulse calls, the DTW and LPC template-based detectors obtained a sensitivity of 96.04% and 97.14% respectively for high signal-to-noise ratio (snr), depending on an empirically determined reliability value (Γ). On the other hand, the accuracy of the DTW and LPC template-based detectors decrease to 94.98% and 96.00% respectively as the background noise increase (the background noise is mostly due to bad weather conditions during recordings).

The rest of the paper is structured as follows. Section II describes the recording location, PAM set-up and the datasets preprocessing phase. In Section III, we explain the characteristics of the B. e. edeni short pulse call with standard parameters. Section IV briefly explains the two signal processing techniques used in the detection algorithms. The developed template-based detection algorithms are discussed in Section V. Section VI analyses the results of the detection algorithms for some specified parameters. The paper is concluded in Section VII.

II. RECORDINGS AND PREPROCESSING

From December 2018 to April 2019, recordings were collected on a weekly basis to study B. e. edeni calls in an area of approximately 13km², close to Gordon’s bay harbour, False bay (34°08'47.3”S 18°48’10.4”E), South-West, South Africa, as shown in Fig. 1. In Fig. 1a, we show the region where the recordings is carried out in South Africa, while Fig. 1b shows the exact location where the recording is carried including the coordinates. The depth at the recording site was less than 30 meters. During these recordings, standard protocols were strictly adhered to as sanctioned by the Department of Environmental Affairs, South Africa. For instance, we kept the required minimum distance on sighting of the individual. The individual was identified each time it was sighted based on its features as discussed in [26]. Most times, recordings were carried out when a single individual is sighted. The individual was mature, but for all cases we could not identify the sex, whether it is a male or female. Note, some times, more than one individuals were sighted but no recording were carried out in such situations. The recordings is carried out using dipping hydrophones. In the set-up, a hydrophone (Aquarian Audio H2A-XLR Hydrophone with sensitivity −180dB re: 1V/µPa and frequency range from 10Hz to 100kHz) was connected to a Zoom H1N recorder, operating at 96ksps at 24 bit resolution. Dataset was saved as raw samples (.wav format) in order to preserve the phase and amplitude data as best as possible. The deployment was to dip the hydrophone from a sail boat (8m long with inboard engine), under varying conditions. That is, sailing (2-4kts/h), dropping the sails (less than 1kt/h),
heaving to (less than 2kts/h). Before recordings, the engine of the boat is shut off. Also, recordings were done when the *B. e. edeni* was the only species sighted. Thus, the raw samples were filtered with a 3rd order Butterworth bandpass filter in MATLAB, R2018b. The filter eliminate frequencies under 10Hz and above 8000Hz in order to reduce background noise and the DC component. Fig. 2 depicts a filtered sample of the short pulse call (note, we presented the time series and spectrogram representation of the pulse call before filtering using Sonic Visualiser in Fig. 3, as the spectrogram of pulse calls can be more clearly viewed in Fig. 3 in comparison to Fig. 2b). Similarly, the resulting signal was analysed using MATLAB, where different characteristics explaining the main component of the pulse calls were extracted as discussed in Section III.

**III. FEATURES OF THE BRYDE’S WHALE SHORT PULSE CALL**

During the five months recordings, one day per week of approximately two hours recordings, about fifteen different dataset were collected. A single recurrence call was observed which correspond to virtually all the vocalisation on sighting of the individual. Aside sighting the *B. e. edeni* whale physical features which corresponds to the descriptions in Penry *et al* [26], we are optimistic that this call is produced by the *B. e. edeni*. Firstly, Bryde’s whales were historically sighted in this region by Olsen [21]. Also during recordings, no other whales or calves were cited within radius of the recording site. That is, no other cetaceans were in a radius of about 3 nautical miles. Besides, the call has been observed when we visually confirmed the presence of inshore Bryde’s whales (that is, the Bryde’s whale were in radius of 1NM when the sounds were observed). To further verify the call, we carried out comparative tests when Bryde’s whales were not in the vicinity, and this “short pulse calls” were not observed. We assume that this call is probably used by the *B. e. edeni* for hunting or navigation since there are no other calves present during recordings. The pulse call has a small relative amplitude at the start, which increases to a maximum of 0.36 or minimum of −0.39, and decays rapidly as the call ends. The relative amplitude range of the Bryde’s whale pulse calls is much larger than other major sources of biological noise in bays such as the snapping shrimp sound that can be misrepresented as the Bryde’s whale sound. Also, in the frequency domain, the minimum frequencies of the pulse calls range between 0.07 ± 0.02kHz, while its highest
align the two time series by constructing an $S$ and $C$ sequences are efficiently warped in a non-linear manner to the temporal distortions between the sequences. The time possible alignment between two time sequences, utilising been used in marine mammal sound detection and classifica-
ing, and manufacturing [34]. Likewise, this technique have been used in [28] has been widely used in marine mammal vocal signals [37]. The concept of LPC is to calculate an approximated value of the current speech sample $\hat{g}(\phi)$ by a linear combination of the preceding regenerated $\theta^{th}$ samples as [29], [30], [38]:

$$\hat{g}(\phi) = \alpha_1 g(\phi - 1) + \alpha_2 g(\phi - 2) + \cdots + \alpha_{\beta} g(\phi - \theta)$$

$$= \sum_{\beta,\theta=1}^{\psi} \alpha_\beta g(\phi - \theta),$$

where $\alpha_\beta$ is the filter coefficients, $\hat{g}(\phi)$ is the approximated value of $g(\phi)$, $g(\phi - \theta)$ is the preceding $\theta^{th}$ samples, and $\psi = \beta^{th} = \theta^{th}$ is the polynomial order or the number of filter coefficients. These distinctive set of coefficients $\alpha_\beta$ can be calculated by minimising the sum of the squared differences between the linearly estimated samples and the original samples as defined in (3) [29], [30], [38]:

$$\epsilon(\phi) = g(\phi) - \hat{g}(\phi) = g(\phi) - \sum_{\beta,\theta=1}^{\psi} \alpha_\beta g(\phi - \theta),$$

where $\epsilon(\phi)$ is the error between $g(\phi)$ and $\hat{g}(\phi)$. The coefficients $\alpha_\beta$ can be determined from (3) using the autocorrelation method. Typically, the number of coefficients ranges from 10-14. In this paper, we assume $\psi = \beta^{th} = \theta^{th} = 12$. This implies that the filter coefficients is a 12 order polynomial defined as $P(\alpha_\beta)$. See [29], [30] for more discussion on the LPC technique.

### IV. DETECTION TECHNIQUES

#### A. DYNAMIC TIME WARPING

DTW introduced in Itakura [28] has been widely used in speech recognition, gesture recognition, medicine, data mining, and manufacturing [34]. Likewise, this technique have been used in marine mammal sound detection and classification [32], [35]. The DTW algorithm is used to find the best possible alignment between two time sequences, utilising the temporal distortions between the sequences. The time sequences are efficiently warped in a non-linear manner to match each other. For example, given two time series $S_1$ and $S_2$, of length $i$ and $j$ respectively, the DTW algorithm align the two time series by constructing an $i \times j$ matrix [28]:

$$D = \min \begin{pmatrix} D[i-1,j-1] & D[i-1,j] & D[i,j-1] \end{pmatrix} + |S_1_i - S_2_j|, \quad (1)$$

where each element in $D$ represent the similarities between the two time series $S_1$ and $S_2$ at positions $i$ and $j$ respectively. In this paper, we define the difference between any two time series signal as $D_{i\phi,j\phi}$. That is, the value of the element at the $i^{th}$ and $j^{th}$ position of the difference matrix $D$. For more information on the DTW technique, refer to [28], [34], [36].

#### B. LINEAR PREDICTIVE CODING

LPC, often referred to as inverse filtering have been successfully used for speech coding, synthesis and recognition. In addition, it has been used to analyse short length of marine mammal vocal signals [37]. The concept of LPC is to calculate an approximated value of the current speech sample $\hat{g}(\phi)$ by a linear combination of the preceding regenerated $\theta^{th}$ samples as [29], [30], [38]:

$$\hat{g}(\phi) = \alpha_1 g(\phi - 1) + \alpha_2 g(\phi - 2) + \cdots + \alpha_{\beta} g(\phi - \theta)$$

$$= \sum_{\beta,\theta=1}^{\psi} \alpha_\beta g(\phi - \theta),$$

where $\alpha_\beta$ is the filter coefficients, $\hat{g}(\phi)$ is the approximated value of $g(\phi)$, $g(\phi - \theta)$ is the preceding $\theta^{th}$ samples, and $\psi = \beta^{th} = \theta^{th}$ is the polynomial order or the number of filter coefficients. These distinctive set of coefficients $\alpha_\beta$ can be calculated by minimising the sum of the squared differences between the linearly estimated samples and the original samples as defined in (3) [29], [30], [38]:

$$\epsilon(\phi) = g(\phi) - \hat{g}(\phi) = g(\phi) - \sum_{\beta,\theta=1}^{\psi} \alpha_\beta g(\phi - \theta),$$

where $\epsilon(\phi)$ is the error between $g(\phi)$ and $\hat{g}(\phi)$. The coefficients $\alpha_\beta$ can be determined from (3) using the autocorrelation method. Typically, the number of coefficients ranges from 10-14. In this paper, we assume $\psi = \beta^{th} = \theta^{th} = 12$. This implies that the filter coefficients is a 12 order polynomial defined as $P(\alpha_\beta)$. See [29], [30] for more discussion on the LPC technique.

### V. TEMPLATE-BASED DETECTION ALGORITHMS

In developing the detectors, some of the short pulse calls were manually identified from the datasets, recorded on different days to form the templates. These short pulse calls are identified from a small section of the dataset while the remaining section (the larger section) of the dataset is used to verify the performance of the detector. Fig. 4 shows some of the visually identified short pulse calls from two different days. The identified short pulse calls from each day are termed Template $A$ ($T_A$) and Template $B$ ($T_B$). Two templates were chosen to verify the performance of the developed detection algorithms for change in background noise. The recordings where the samples in $T_A$ are identified has an average $snr$ of $+3.84dB$ better in comparison to $T_B$. Each of the template contains $k$ number of samples of $l$ varying lengths.
Talarities between different time series waveform. Thus, in this
As mentioned, the DTW algorithm is used to find the simi-
larities between each sample in the template. The value of
$k$ in a template. The value of $t$ is the sampling point,
where $C_d$ is chosen to be 18, 12 and 6 as shown in Section VI. The length of each
sample in the template varies depending on the value
of $C_d$. The template-based detectors is thus expatiated
as follows.

A. DTW-BASED
From Table 1, observe that the short pulse call duration ($C_d$)
is between $1.2 - 3.1s$. As such, the templates are specified
to contain samples of different length $l$ in the range of the
$C_d$. Expressing each sample in the template as a row vector,
we define the template as:

$$T_{Ai\|B} = \begin{bmatrix} [t_1, t_{1,1}, \ldots, t_{1,C_d}] \\ [t_2, t_{2,1}, \ldots, t_{2,C_d}] \\ \vdots \\ [t_k, t_{k,1}, \ldots, t_{k,C_d}] \end{bmatrix},$$ (4)

where $t$ is the sampling point, $k$ is the number of sample
in a template. The value of $k$ is chosen to be 18, 12 and 6 as shown in Section VI. The length of each
classifier in the template varies depending on the value
of $C_d$. The template-based detectors is thus expatiated
as follows.

A. DTW-BASED

As mentioned, the DTW algorithm is used to find the simi-
larities between different time series waveform. Thus, in this
detection algorithm, the template $T_{Ai\|B}$ with $k$ number
of manually identified samples is warped with each other
using (1) to form a $k \times k$ dissimilarities template matrix
defined as:

$$T = \begin{bmatrix} 0 & D_{1,2} & \ldots & D_{1,k} \\ D_{2,1} & 0 & \ldots & D_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ D_{k,1} & D_{k,2} & \ldots & 0 \end{bmatrix},$$ (5)

where each element in $T$ is the $D_{p,b}$ dissimilarities between
each sample in the template. Subsequently, the algorithm
finds the maximum entry of each column of $T$ to form a $1 \times k$
row vector defined as:

$$T_{max} = [T_{max,1}, T_{max,2}, T_{max,3}, \ldots, T_{max,k}],$$ (6)

The detection process then starts by sliding through the
recordings $R_{2D}$ with a defined window size $(w)$ and overlapping
size $(o_v)$. Having known the $C_d$ of most of the short pulse
calls, we set $w = 1s$ and $o_v = w/2$. For each selected window,
the algorithm calculates a relative energy of the waveform as
defined as:

$$E = \sum_{i=1}^{C_d} (t_{C_d})^2,$$ (7)

where $t$ is the sampling point. The $\delta$ value is empirically set.
Matching the values of $\delta$ and $E$, any selected window with an
$E$ lower than $\delta$ is not considered as the B. e. edeni pulse call.
This significantly reduces the computational time complexity
of the proposed detection algorithm. Onward, the similarities
between the window frame with $E \geq \delta$, and the $k$ samples in
the template is computed using (1) to form a $1 \times k$ row vector
defined as:

$$T_w = [T_{w1}, T_{w2}, T_{w3}, \ldots, T_{wk}],$$ (8)

where each element in $T_w$ is the $D_{p,b}$ dissimilarities between
the selected window frame and each sample in the template.
Thereafter, the algorithm finds a count score ($\gamma$) by match-
ing (8) and (6). That is, $\gamma$ is computed by counting the number
of times the value in each column of $T_w$ is less or equal to
the value in the corresponding column of $T_{max}$. The value
of $\gamma$ is therefore compared with a predetermined reliability
value $\Gamma$. This value of $\Gamma$ determines the performance of the
proposed detector. A small value of $\Gamma$ increases the sensitivity
of the detector with the price of increased false positive rate
as is subsequently shown. With this in mind, a trade-off value
should be defined for $\Gamma$ based on observations. Results are
presented in Section VI for different values of $\Gamma$. Note, $\Gamma$
ranges between $0 - 1$. If $\gamma \geq [\Gamma * \gamma]$ the window size $w$
is stored as the pulse call of the B. e. edeni whale.

From Table 1, the $C_d$ ranges from 1.2-3.2s; thus, the algo-
rithm synchronises every stored $w$ in the range of the $C_d$. The
algorithm achieves this by comparing the previously detected
pulse call with the current one as:

$$O = e_{w_{p}} - b_{w_{i}},$$ (9)
where $w_p$ and $w_c$ is the previous and current detected pulse call respectively, $e_{wp}$ is last sampling point of $w_p$, and $b_{wc}$ is the first sampling point of $w_c$. If $O \neq N_{o_v} - 1$, $w_p$ is not synchronised with $w_c$ ($N_{o_v}$ is number of sampling points in $o_v$). Otherwise, it is synchronised as:

$$w_c = w_p + o_v.$$  (10)

The template-based DTW detector is hereby summarised in algorithm 1.

**Algorithm 1** DTW Detection Algorithm

**Input:** $T_{A!|B}$, $k$, $R_D$, $w$, $o_v$, $\delta$, $\Gamma$, $w_p = 0$, $e_{wp} = 0$

**Output:** $w_c$ as detected pulse call

1: build $k \times k$ $T$ matrix based on (5)
2: find $T_{max}$ based on (6)
3: choose window size, $w$, $o_v$: Slide through $R_D$
4: calculate $E$ for each $w$
5: if $E < \delta$
   return to 3
6: else
7: compute $T_w$ based on (8)
8: match $T_{max}$ and $T_w$ to find $\gamma$
9: if $\gamma < \lfloor \psi \gamma \rfloor$
   return to 3
10: else
11: store $w_c$, and $b_{wc}$
12: end if
13: end if
14: if $e_{wp} - b_{wc} \neq N_{o_v} - 1$
15: Detect $w_c$
16: else
17: Detect $w_c = w_p + o_v$
18: store $w_p = w_c$, $e_{wp} = b_{wc}$
19: end if

Of note, in some cases, the duration of the manually identified calls and the automatic detected call differs in duration, such that the automatically detected call is slightly longer. In such situation, the automatic detected waveform duration can be synchronised to approximately match the manually identified waveform duration. The detected waveform can be divided into smaller windows $s_w$ ($s_w \ll w$). Subsequently, (7) can be computed for these $s_w$ waveforms. The result can be matched with a small $\delta$ ($s_\delta$) ($s_\delta \ll \delta$), where $s_\delta$ is determined empirically. Doing this, the two ends of the detected waveform can be synchronised to fit an approximate of the manually identified call duration.

**B. LPC-BASED**

The LPC-based detector is developed from the filter coefficients $\alpha_\beta$ produced using (3). The detector first find the roots of the filter coefficient polynomial $P(\alpha_\beta)$ to obtain a $1 \times \psi - 1$ row vector as:

$$\mathcal{R} = [\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \ldots, \mathcal{R}_{\psi-1}],$$  (11)

where the elements of $\mathcal{R}$ are often complex numbers. Note that the roots of $P(\alpha_\beta)$ can be derived using different mathematical methods as presented in [39]. Therefore, $\mathcal{R}$ is computed for the template $T_{A!|B}$ to obtain a $k \times \psi - 1$ matrix defined as:

$$\mathcal{R}_{T_{A!|B}} = \begin{bmatrix} \mathcal{R}_{1,1} & \mathcal{R}_{1,2} & \ldots & \mathcal{R}_{1,\psi-1} \\ \mathcal{R}_{2,1} & \mathcal{R}_{2,2} & \ldots & \mathcal{R}_{2,\psi-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{R}_{k,1} & \mathcal{R}_{k,2} & \ldots & \mathcal{R}_{k,\psi-1} \end{bmatrix}. $$  (12)

Each entry in $\mathcal{R}_{T_{A!|B}}$ is subsequently compared with a defined complex reference point $r_p$ ($r_p = 0 + i0$) to find the euclidean distance. This forms a corresponding $k \times \psi - 1$ distance template matrix defined as:

$$T = \begin{bmatrix} D_{\mathcal{R}_{1,1}} & D_{\mathcal{R}_{1,2}} & \ldots & D_{\mathcal{R}_{1,\psi-1}} \\ D_{\mathcal{R}_{2,1}} & D_{\mathcal{R}_{2,2}} & \ldots & D_{\mathcal{R}_{2,\psi-1}} \\ \vdots & \vdots & \ddots & \vdots \\ D_{\mathcal{R}_{k,1}} & D_{\mathcal{R}_{k,2}} & \ldots & D_{\mathcal{R}_{k,\psi-1}} \end{bmatrix}. $$  (13)

Thus, the algorithm finds the maximum entry of each column of $T$ to form a $1 \times \psi - 1$ row vector defined as:

$$T_{max} = [T_{max,1} \ T_{max,2} \ T_{max,3} \ldots \ T_{max,\psi-1}].$$  (14)

The detection process continues in a similar way as in the DTW detector using the same set of parameters ($w$, $o_v$, $\delta$, $\Gamma$). However, (8) used in the DTW detector is computed in this case by comparing the roots of the selected sample of window $w$ to $r_p$. In this way, we obtain a $1 \times \psi - 1$ distance sample row vector defined as:

$$T_w = [T_{w_1} \ T_{w_2} \ T_{w_3} \ldots \ T_{w_{\psi-1}}].$$  (15)

The template-based LPC detector is summarised in Algorithm 2.

**Algorithm 2** LPC Detection Algorithm

**Input:** $T_{A!|B}$, $k$, $R_D$, $w$, $o_v$, $\delta$, $\Gamma$, $w_p = 0$, $e_{wp} = 0$, $r_p$, $\psi$

**Output:** $w_c$ as detected pulse call

1: compute $\mathcal{R}$ using (2), (3) and (11)
2: compute $\mathcal{R}_{T_{A!|B}}$
3: Build $k \times \psi - 1$ $T$ matrix based on $r_p$
4: find $T_{max}$ based on (14)
5: similar process as in Algorithm 1, steps 3-18

**VI. TEST RESULTS AND DISCUSSION**

In this section, the proposed template-based detectors were applied to recognise continuous recordings of the short pulse calls. In the results presented, we verified the performance of the detectors for different values of $k$. The reliability $\Gamma$ is also a factor we used in the result comparisons. In addition, we evaluate the quality of detection algorithms by evaluating the detection sensitivities $S$, false positive rates $F_p$, and failure rates $F$ of both methods as defined in (16) [40]:

$$S = \frac{\eta}{\eta + \rho}, \quad F_p = \frac{\tau}{\eta + \tau} \quad \text{and} \quad F = 1 - S.$$  (16)
where $\eta$ is the number of times the manually detected short pulse call matches the output of the automatic detectors, $\rho$ is the number of times the proposed detectors missed the manually detected pulse calls, and $\tau$ is the number of times the proposed detectors wrongly recognised a signal as the pulse call. In all cases, a high value of $S$ is desirable to rate the accuracy of any sound detection technique. This in turn indicates a small value of $F$ as shown in (16). More so, the smaller the value of $F_p$, the more dependable is the detection algorithm.

As earlier mentioned, we verified the performance of the detectors for varying weather conditions (background noise). Template $A$ $(T_A)$ comprises of less noisy samples while $T_B$ contain samples with more background noise. Table 2-7, shows the performance of the detectors as a function of $k$ for different empirically selected values of $\Gamma$. In Table 2 and 3, a $\Gamma = \frac{4}{6}$ was used in both detection algorithms. Firstly, as $k$ increases, the performance of the detectors improve linearly. The performance of the detectors for the two templates differ with $T_A$ (Table 2) obtaining a superior performance in comparison to $T_B$ (Table 3). This implies that the lower the noise in the identified samples used in the template, the better the performance of the detectors.

Filtering can be a better way of reducing the noise as done during preprocessing but it cannot eliminate all noise components. Moreover, as shown in Table 2 and 3, the LPC-based detectors is a more robust recogniser as compared to the DTW-based detector as it offers better performance in terms of $S$, $F_p$ and $F$.

In Table 4 and 5, the value of $\Gamma$ $(\Gamma = \frac{3}{5})$ was reduced in both detection algorithms. As shown, the $S$ of the algorithms increase with corresponding reduction in $F$. However, $F_p$ increase with a reduction in the value of $\Gamma$. Likewise in Table 6 and 7, as $\Gamma$ $(\Gamma = \frac{2}{5})$ reduces further, the $S$ increases while $F_p$ also increase in both algorithms. An increase in $F_p$ implies that the value of $\tau$ will increase as a result of a decrease in $\Gamma$ (that is, $\Gamma \propto \frac{1}{k}$). Although $\eta$ also increase as the value of $\Gamma$ decreases, this increase is not as significant in comparison to the increase in the false positive ($\tau$) calls detected. In real time, the false positive rate $F_p$ of the detector must be as low as possible while maintaining a high level of detection accuracy (sensitivity). This means that there is a trade-off between $S$ and $F_p$ in empirically determining the value of $\Gamma$. Thus, in real time, $\Gamma$ should be chosen depending on application requirements. Summarily, irrespective of the value of $\Gamma$, $T_A$ (Table 2, 4 and 6) performs better than $T_B$ (Table 3, 5 and 7) respectively, and the LPC-based detector offered superior performance in comparison to the DTW-based detector. Both detection algorithms exhibit low computational time complexity of order $O(2^k)$ and can be used in real time to analyse the movement of obscure but vocal marine species instead of using traditional methods. In addition, the developed detectors can be modified to recognise different marine mammal sounds, where parameters such as $w$, $o_\gamma$, $\delta$, and $\Gamma$ can be selected based on observation of the sound waveforms and application requirements.

### VII. CONCLUSION

The paper identified a short pulse call of a *B. e. edeni* whale off South-West South Africa. The behavioural patterns of the *B. e. edeni* whale is quite difficult to obtain. Thus, the recognition of this call is a noteworthy contribution to the knowledge of this species off South Africa and the world at large. In addition, two template-based detection algorithms were developed for this identified short pulse call, employing DTW and LPC techniques. Both algorithms were shown to demonstrate high sensitivity with reduced false positive rate. But, the LPC-based detector is a more robust recogniser as compared to the DTW-based detector. Besides, the developed
detection algorithms can be used in real time vocalisation detection because they both offer low computational time complexity. Moreover, the algorithms can be modified to analyse the movement of different obscure but vocal marine species instead of manual identification.

REFERENCES


OLAYINKA O. OGUNDILE (Member, IEEE) received the B.Eng. degree in electrical engineering from the University of Ilorin, Kwara, Nigeria, in 2007, the M.Sc. degree in communication engineering from The University of Manchester, U.K., in 2010, and the Ph.D. degree from the University of the Witwatersrand, Johannesburg, South Africa, in 2016. He is currently a Lecturer with the Department of Physics and Telecommunication, Tai Solarin University of Education, Nigeria. He is also a Postdoctoral Research Fellow with the Department of Electrical and Information Engineering, Stellenbosch University, South Africa. His research interests include digital communication, digital transmission techniques, digital signal processing/machine learning, channel estimation, forward error correction, and wireless sensor networks.

DANIEL J. J. VERSFELD (Member, IEEE) received the B.Eng. degree in electronic engineering and the M.Eng. degree in electronic engineering from North-West University, South Africa, in 1999 and 2001, respectively, and the Ph.D. degree from the University of Johannesburg, South Africa, in 2011. He is currently an Associate Professor with the Department of Electrical and Electronics Engineering, Stellenbosch University, Stellenbosch, South Africa. His research interests include algebraic coding for digital communications and signal processing applied to communications.