Scientific African 24 (2024) e02128

Contents lists available at ScienceDirect

Scientific African

journal homepage: www.elsevier.com/locate/sciaf

Linear discriminant analysis based hidden Markov model for detection of *Mysticetes'* vocalisations

O.M. Ogundile [a,*], A.A. Owoade [a], O.O. Ogundile [a], O.P. Babalola [b]

[a] Department of Computer Science, Tai Solarin University, Ogun State, Nigeria
[b] Department of Electrical, Electronics, and Computer Engineering, Cape Peninsula University of Technology, Bellville, South Africa

ARTICLE INFO

Editor name: Rashid A. Saeed

Keywords:
Feature extraction
Feature vector
HMM
LDA
Mysticetes'
Vocalisations detection

ABSTRACT

Mysticetes' produce distinctive vocalisations which are used for echolocation, communication, and other marine functions. These cryptic vocalisations are studied by marine scientist to determine the behavioural patterns and movement of this suborder of cetaceans within their ecosystem. In practice, these vocalisations are gathered using passive acoustic monitoring over days, weeks, months, and even years. Therefore, it is complex to study these sounds using traditional visual inspection techniques because the gathered datasets are huge. Machine learning (ML) tools such as Gaussian mixture models (GMMs), support vector machines (SVMs), and hidden Markov models (HMMs) have been adopted in recent times to proffer analytic solutions to automatically detect and study these cryptic vocalisations. Notwithstanding, the feature extraction techniques employed play a vital role in determining the performance of these ML tools. In most cases, the performance of the feature extraction technique is directly proportional to the performance of the ML tools. Thus, the method of linear discriminant analysis (LDA) is introduced in this article as a feature extraction technique that can be adapted with the HMMs (LDA-HMM) to seamlessly detect the vocalisations of *Mysticetes*. The performance of the proposed LDA-HMM detector is compared with other recent detectors for *Mysticetes'* vocalisations in the literature using two different species: Humpback whale songs and Bryde's whale pulses. Experimental results show that the developed LDA-HMM detector is a performance-efficient alternative in comparison to the recent detection techniques studied in this article. Besides, the LDA-HMM detector offers less computational time complexity; as such, it is more suitable for real-time applications.

Introduction

The detection and study of *Mysticetes'* vocalisations has recently gained popularity; most especially, with the ease provided by machine learning (ML) tools [1]. *Mysticetes'* are suborder of cetaceans existing in the waters of some region of the world. Their study has contributed immensely to the economic values of some countries. According to the findings in [2,3], the whale watching industry generates an annual revenue exceeding US$2 billion annually. Thus, this has contributed to government revenue while concurrently creating employment opportunities for a substantial workforce. To study the movement and behaviour of these *Mysticetes* in their ecosystem, marine biologist or ecologist obtain their vocalisations using a procedure termed passive acoustic monitoring (PAM). These sounds are accumulated over days, weeks, months, and even years. Therefore, the recorded sound datasets are huge, making it complex to detect and study their vocalisations with the traditional visual inspection technique.

* Corresponding author.
E-mail address: ogundileoo@tasued.edu.ng (O.M. Ogundile).

https://doi.org/10.1016/j.sciaf.2024.e02128
Received 14 November 2023; Received in revised form 30 December 2023; Accepted 21 February 2024
Available online 23 February 2024

However, with the advent of ML tools such as Gaussian mixture models (GMMs), support vector machines (SVMs), hidden Markov models (HMMs), and so on, ecologist can easily detect the vocalisation of these cryptic species and study their behaviour within their ecosystem. Machine learning tools have been widely employed for the automated detection of the vocalisations of Mysticetes of *Mysticetes* [2,4,5]. The performance of most of these ML tools robustly depends on the feature extraction technique(s) employed. In reality, the performance of ML tools is a function of the performance of the feature extraction techniques. For instance, different feature extraction techniques such as empirical mode decomposition (EMD) [6], dynamic mode decomposition (DMD) [7], Mel-scale frequency cepstral coefficients (MFCC) [8], linear predictive coefficient (LPC) [9] have been combined with ML tools such as the GMM and HMM to automatically detect *Mysticetes'* vocalisations in the literature [10,11].

With a focus on HMMs, this machine learning tool finds common application in diverse fields, including cloud computing, weather forecast, human speech processing, source coding, among others [12]. Also, in animal sound detection, HMM has been used to automatically detect the vocalisations of *Mysticetes*, because it is robust and flexible, and can easily be used to automatically detect sounds from a sequence of observations [13]. Nonetheless, as mentioned earlier, irrespective of the sturdiness of the HMM, their detection accuracy (sensitivity) and reliability (false discovery rate (FDR)) depend on the feature extraction technique employed. Hence, attention must be carefully given to the feature extraction technique(s) adopted with the ML tool to improve the FDR and sensitivity performances of the ML detection techniques.

In this regard, the method of linear discriminant analysis (LDA) is introduced in this article as a feature extraction technique that can be adapted with the HMM for detecting the vocalisations of *Mysticetes*. The LDA, also called Fisher's discriminant analysis is an algorithm commonly used for pattern classification and dimensionality reduction [14,15]. LDA is a method used in statistics and other fields (such as biometrics [16,17], Bioinformatics [18], and chemistry [19]), to find a linear combination of features that characterises two or more classes of objects or events. The LDA technique transforms the features in a large data matrix into a lower dimensional space without losing the information contained in the large data matrix [14]. Thus, the lower dimensional space or variables can be easily conceptualised with different machine learning tools as discussed in this article. The developed LDA-HMM detector performance will be analysed theoretically using two different *Mysticetes'* vocalisation datasets documented on a cetacean sound database.[1] Furthermore, this article compares the results of the proposed LDA-HMM detector to recent techniques found in literature such as the MFCC-HMM, LPC-HMM, DMD-HMM, and EMD-HMM detection techniques in terms of their FDR and sensitivity performances.

Therefore, the contribution and relevance of this article is highlighted as follows.

1. The method of LDA is adapted for the first time with HMM for detecting the vocalisations of *Mysticetes* (Bryde's whale and Humpback whale).
2. The LDA-HMM outperforms the traditional MFCC-HMM and LPC-HMM detection techniques and the recently developed EMD-HMM and DMD-HMM techniques in the literature
3. The introduced LDA-HMM exhibits less computational complexity in comparison to the aforementioned methods; as such, it is more suitable for real-time applications.

This article is arranged as follows. Section "Related works" presents a review of the related works. In particular, this section discusses the ML detectors used in this article for performance comparison. In Section "Linear discriminant analysis", the basic principles of the LDA are discussed. Section "Proposed LDA feature vector" explains the formulation of the LDA as a feature vector suitable for the HMM. In addition, Section "LDA feature vector: Numerical illustration" explains the LDA feature extraction process using mathematical illustration. The rudiment of the HMM is briefly discussed in Section "Hidden Markov model" while the LDA-HMM detector simulation parameters and set-up are explained in Section "Simulation parameters and set-up". Section "Results and discussion" presents the results from the simulation runs and discusses these results in detail. Lastly, the article is summarised with discernible remarks in Section "Conclusion".

Related works

As earlier mentioned, cetaceans have contributed immensely to the economic values of some countries; as such, the detection of *Mysticetes'* vocalisations have received great attention. Ogundile and Versfeld [20] developed two threshold template-based detectors for inshore Bryde's whale short pulse calls. One of the threshold template detectors is based on the dynamic time warping (DTW) machine learning tool while the other is developed based on the LPC principles. The two threshold detectors demonstrated a moderate computational time complexity; nevertheless, their performance in terms of false positive rate (FPR) and sensitivity proved inadequate under specific parameter conditions.

In [10], the authors developed an EMD-HMM detection technique that can be used to detect the vocalisation of inshore Bryde's whale short pulses. The EMD-HMM was shown to offer superior performance in comparison to the traditional MFCC-HMM and LPC-HMM detectors. Nevertheless, it can be noted from the result presented in their letter that the false positive rate (FPR) and sensitivity performances can still be improved. Besides, the performance of their developed detector was verified for a single *Mysticetes* specie. Therefore, the performance of their detector cannot be generalised for the different types of *Mysticetes* available. In [21], the same authors developed a selective time domain feature extraction technique-based HMM (MAP-HMM) detector for inshore Bryde's whale short pulses. Similarly, the MAP-HMM detector was compared with the LPC-HMM and MFCC-HMM detectors and it was shown to

1 http://www.mobysound.org/mysticetes.html

exhibit better FDR and sensitivity performances. However, comparing the performance of the EMD-HMM and MAP-HMM detectors, it is observed that their performances are approximately the same while the MAP-HMM offered a reduced computational time complexity. Yet, the performances of the MAP-HMM detector cannot also be generalised for the about fourteen types of *Mysticetes* available because it was documented for the inshore Bryde's whale short pulses only.

The DMD-HMM was recently developed in [11] for the detection of *Mysticetes'* vocalisations. The FDR and sensitivity performances of the DMD-HMM were seen to surpass that of the EMD-HMM, and the traditional LPC-HMM and MFCC-HMM detectors. Nonetheless, the DMD-HMM detection technique obtains this good FDR and sensitivity performances at a high computational time complexity. Thus, for real-time *Mysticetes'* vocalisations detection, the DMD-HMM detector might be impracticable.

This article proposes an LDA-HMM detector, which exhibits better FDR and sensitivity performances as compared to the DMD-HMM and DED-HMM detectors, and the traditional LPC-HMM and MFCC-HMM detectors, at a moderate computational time complexity. Moreover, the performance of this developed LDA-HMM detector is documented for two *Mysticetes'* types. Note that the LDA-HMM detector can, however, be employed to detect other *Mysticetes'* species that produce similar vocalisations.

Linear discriminant analysis

The LDA, also called Fisher's discriminant analysis is an algorithm commonly employed for pattern classification and dimensionality reduction [14,15]. LDA is a method used in statistics and other fields (such as biometrics [16,17], bioinformatics [18], and chemistry [19]), to find a linear combination of features that characterises or separates two or more classes of objects or events. The LDA technique transforms the features data matrix into a lower dimensional space, which maximises the ratio of the within-class variance to the between-class variance; thus, it guarantees maximum class separability [14]. Therefore, the lower dimensional space or variables can be easily visualised and analysed using different ML tools. The LDA achieves this transformation in three basic steps: (1) It computes the separability between different classes (between-class variance), (2) It computes the distance between the samples of each class and the mean (within-class variance), and (3) It constructs the lower dimensional space that minimises the within-class variance and maximises the between-class variance [14].

Between-class variance calculation

The between-class variance ($\mathcal{C}_{\mathcal{B}}$) also called the between-class matrix represents the distance between the mean of the ϱth class (μ_ϱ) and the total mean (μ). That is, the LDA technique finds a lower dimensional space that maximises the between-class variance [14]. For example, given a ℓ dimensional signal, $A_{J,\ell}$, defined by Eq. (1):

$$A_{J,\ell} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,\ell} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,\ell} \\ \vdots & \vdots & \ddots & \vdots \\ a_{J,1} & a_{J,2} & \cdots & a_{J,\ell} \end{pmatrix}, \tag{1}$$

where $a_{J,\ell}$ is a sample point in $A_{J,\ell}$. Note that each sample represents a point in $\mathcal{D}$-dimensional space. That is each sample can be represented by $\mathcal{D}$ features ($a_\varrho \in \mathcal{R}^{\mathcal{D}}$). The dimension, ℓ is an important requirement used to determine the performance of the LDA feature extraction technique. The dimension of the feature vector extracted using the LDA technique is determined by the value of ℓ, as discussed later.

Assuming that $A_{J,\ell}$ is partitioned into g number of classes as defined by Eq. (2):

$$\mathbb{A}_{J,\ell} = \begin{pmatrix} \varpi_1 \\ \varpi_2 \\ \vdots \\ \varpi_g \end{pmatrix}, \tag{2}$$

where ϖ_g is a matrix in $A_{J,\ell}$ ($\varpi_g \in A_{J,\ell}$). Also, assume that each class has ε samples, where ε_ϱ denotes the number of samples of the ϱth class. Thus, the total number of samples ($\hbar$) can be represented mathematically as defined by Eq. (3):

$$\hbar = \sum_{\varrho=1}^{g} \varepsilon_\varrho. \tag{3}$$

Therefore, to compute $\mathcal{C}_{\mathcal{B}}$, the separation distance between the different classes ($b_\varrho - b$) is firstly determined as defined by Eq. (4) [14]:

$$(b_\varrho - b)^2 = (\mathcal{W}^\dagger \mu_\varrho - \mathcal{W}^\dagger \mu)^2 = \mathcal{W}^\dagger (\mu_\varrho - \mu)(\mu_\varrho - \mu)^\dagger \mathcal{W}, \tag{4}$$

where b is an estimate of the total mean of all the classes given as $b = \mathcal{W}^\dagger \mu$ while b_ϱ is the estimate of the mean of the ϱth class given as $\mathcal{W}^\dagger \mu_\varrho$. The transformation matrix of the LDA is denoted as $\mathcal{W}$, $\mu_\varrho(1 \times \mathcal{D})$ denotes the mean of the ϱth class as defined by Eq. (5) and $\mu(1 \times \mathcal{D})$ is the total mean of all the classes as defined by Eq. (6) [14].

$$\mu_\varsigma = \frac{1}{\varepsilon_\varsigma} \sum_{a_\varrho \in \varpi_\varsigma} a_\varrho, \tag{5}$$

$$\mu = \frac{1}{\hbar}\sum_{\varrho=1}^{\hbar} a_{\varrho} = \sum_{\varrho=1}^{g} \frac{\varepsilon_{\varrho}}{\hbar}\mu_{\varrho}. \tag{6}$$

In Eq. (4), the separation distance between the mean of the ϱth class (μ_{ϱ}) and the total mean (μ) is denoted as $(\mu_{\varrho}-\mu)(\mu_{\varrho}-\mu)^{\dagger}$. This can also be called the between-class variance of the ϱth class ($\mathcal{C}_{\mathcal{B}_{\varrho}}$). Therefore, Eq. (4) can be written as [14]:

$$(b_{\varrho}-b)^2 = \mathcal{W}^{\dagger}\mathcal{C}_{\mathcal{B}_{\varrho}}\mathcal{W}. \tag{7}$$

Accordingly, the total between-class variance which is the summation of all the ϱth between-class variance is computed as defined by Eq. (8) [14]:

$$\mathcal{C}_{\mathcal{B}} = \sum_{\varrho=1}^{g} \varepsilon_{\varrho}\mathcal{C}_{\mathcal{B}_{\varrho}}. \tag{8}$$

Within-class variance calculation

The within-class variance ($\mathcal{C}_{\aleph}$) also called the within-class matrix represents the difference between samples and the mean of that class. That is, the LDA technique finds a lower dimensional space that minimises the within-class variance [14]. The within-class variance of each class ($\mathcal{C}_{\aleph_{\varsigma}}$) can be computed as defined by Eq. (9) [14]:

$$\sum_{a_{\varrho}\in\varpi_{\varsigma},\varsigma=1,\ldots,g} (\mathcal{W}^{\dagger}a_{\varrho}-b_{\varsigma})^2 = \sum_{a_{\varrho}\in\varpi_{\varsigma},\varsigma=1,\ldots,g} \mathcal{W}^{\dagger}\mathcal{C}_{\aleph_{\varsigma}}\mathcal{W}, \tag{9}$$

where $\mathcal{C}_{\aleph_{\varsigma}}$ is defined by Eq. (10) [14]:

$$\mathcal{C}_{\aleph_{\varsigma}} = \sum_{\varrho=1}^{\varepsilon_{\varsigma}} (a_{\varrho\varsigma}-\mu_{\varsigma})(a_{\varrho\varsigma}-\mu_{\varsigma})^{\dagger}, \tag{10}$$

and $a_{\varrho\varsigma}$ denotes the ϱth samples in the ςth class. Similarly, the total within-class variance is the sum of all the ϱth within-class variance as defined by Eq. (11) [14]:

$$\mathcal{C}_{\aleph} = \sum_{\varrho=1}^{g} \mathcal{C}_{\aleph_{\varsigma}}. \tag{11}$$

Lower dimensional space construction

The between-class variance ($\mathcal{C}_{\mathcal{B}}$) and within-class variance ($\mathcal{C}_{\aleph}$) are subsequently used to estimate the transformation matrix ($\mathcal{W}$) of the LDA technique as defined by Eq. (12) [14]:

$$\arg\max_{\mathcal{W}} \frac{\mathcal{W}^{\dagger}\mathcal{C}_{\mathcal{B}}\mathcal{W}}{\mathcal{W}^{\dagger}\mathcal{C}_{\aleph}\mathcal{W}}. \tag{12}$$

where Eq. (12) is referred to as the Fisher's criterion. As such, the LDA technique is sometimes called the Fisher's discriminant analysis. Eq. (12) can be reformulated as defined by Eq. (13) [14]:

$$\mathcal{C}_{\aleph}\mathcal{W} = \Omega\mathcal{C}_{\mathcal{B}}\mathcal{W}. \tag{13}$$

In Eq. (13), Ω is the eigenvalues of the transformation matrix $\mathcal{W}$. Thus, the solution to the $\ell\times\ell$ transformation matrix ($\mathcal{W} = \mathcal{C}_{\aleph}^{-1}\mathcal{C}_{\mathcal{B}}$) is obtained by computing the eigenvalues ($\Omega = [\Omega_1, \Omega_2, \ldots, \Omega_{\mathcal{D}}]$) and the eigenvectors ($\mathcal{P} = [\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_{\mathcal{D}}]$). The Ω are scalar values and $\mathcal{P}$ are non-zero vectors representing the directions of the new space. This implies that each $\mathcal{P}$ represents one axis of the LDA space and the corresponding Ω represents the magnitude of the $\mathcal{P}$ [14]. The first LDA variable is the $\mathcal{P}$ that has the highest eigenvalue, which in turn has the highest variance. This means it contains the maximum possible message. Thus, the eigenvector is arrayed according to the message contained in the eigenvector in descending order. Arraying the LDA variables in this way enables dimensionality reduction with little or no loss in information. These LDA variables can be constructed as a linear combination of the signal to be analysed, where the dimension, ℓ specifies the number of LDA variables.

The singular vector decomposition (SVD) is a common linear algebra method that can be employed to solve for $\mathcal{P}$ and Ω. The SVD will decompose the transformation matrix $\mathcal{W}$ into three matrices such that:

$$\mathcal{W} = \mathcal{L}\mathcal{F}\mathcal{O}^{\dagger}, \tag{14}$$

where $\mathcal{L}$ is termed the left singular vector, $\mathcal{F}$ is a diagonal matrix, and $\mathcal{O}$ is termed the right singular vector with a $\ell\times\ell$ dimension. The Ω are the entries of the diagonal of $\mathcal{F}$ arrayed in descending order while the $\mathcal{O}$ are the $\mathcal{P}$ or the LDA variables.

Proposed LDA feature vector

In this section, the LDA variables derived in Eq. (14) are reconstructed to form a suitable feature vector that can be adapted with the HMM to automatically detect the vocalisations of *Mysticetes*. Firstly, it is key to show how the dimension, ℓ of the LDA feature vector is determined. The dimension, ℓ determines the computational time complexity the LDA will impose on the HMM training process. As such, dimension, ℓ must be as small as possible to maintain a moderate computational time complexity. This means there is a trade-off between the FDR and sensitivity performances, and the computational time complexity in choosing a value for ℓ. Therefore, the dimension, ℓ is chosen in this article using simulation runs as discussed in Section "Results and discussion".

Essentially, the *Mysticetes'* vocalisations are in the form of a column or row vector. Thus, this column or row vector should be restructured according to the chosen dimension, ℓ to form the matrix $A_{j,\ell}$ defined in Eq. (1). Hence, given that the *Mysticetes'* vocalisations signal is defined in the form of a row vector as:

$$A = \begin{pmatrix} a_1 & a_2 & \dots & a_i \end{pmatrix}, \tag{15}$$

the matrix $A_{j,\ell}$ can be formed from A_i to obtain Eq. (1):

$$A_{j,\ell} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,\ell} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,\ell} \\ \vdots & \vdots & \ddots & \vdots \\ a_{j,1} & a_{j,2} & \cdots & a_{j,\ell} \end{pmatrix}.$$

The elements in the 1st sampling point to the ℓ sampling points of A_i are the points in the first row of $A_{j,\ell}$. Whereas, the last row (j^{th}) of $A_{j,\ell}$ represents the last ℓ sampling points of A_i. This implies that the 1st entry in row one of $A_{j,\ell}$ ($a_{1,1}$) is the 1st entry in A_i; that is, $a_{1,1} = a_1$. Similarly, the last entry in row j of $A_{j,\ell}$ ($c_{j,\ell}$) is the last entry in A_i ($c_{j,\ell} = a_i$).

Therefore, the between-class variance ($C_{\mathcal{B}}$) and within-class variance ($C_{\aleph}$) is computed sequentially from $A_{j,\ell}$ using Eqs. (8) and (11) respectively. Subsequently, the Fisher's criterion (Eq. (13)) is employed to compute the transformation matrix ($\mathcal{W}$), which is decomposed to obtain O as defined by Eq. (16):

$$O = \begin{pmatrix} o_{1,1} & o_{1,2} & \cdots & o_{1,\ell} \\ o_{2,1} & o_{2,2} & \cdots & o_{2,\ell} \\ \vdots & \vdots & \ddots & \vdots \\ o_{\ell,1} & o_{\ell,2} & \cdots & o_{\ell,\ell} \end{pmatrix}. \tag{16}$$

Thus, the feature vector is derived in this study from Eq. (16) as defined by Eq. (17):

$$\begin{aligned} \Bbbk_\ell &= \left[\left(\frac{1}{\ell} \sum_{\kappa=1}^{\ell} |o_\kappa| \right)_1, \left(\frac{1}{\ell} \sum_{\kappa=1}^{\ell} |o_\kappa| \right)_2, \dots, \left(\frac{1}{\ell} \sum_{\kappa=1}^{\ell} |o_\kappa| \right)_\ell \right] \\ &= [v_1, v_2, \dots, v_\ell]. \end{aligned} \tag{17}$$

Note again that the feature vector produced using the LDA is of dimension, ℓ. Subsequently, if the sound signal to be analysed is divided into k number of frames and all the frames, k have approximately f equal number of sampling points, the feature vector can be computed in the form of a matrix as defined by Eq. (18):

$$Z_{LDA} = \begin{pmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,\ell} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,\ell} \\ \vdots & \vdots & \ddots & \vdots \\ v_{k,1} & v_{k,2} & \cdots & v_{k,\ell} \end{pmatrix}. \tag{18}$$

LDA feature vector: Numerical illustration

Fig. 1 depicts a random signal with sampling points defined by Eq. (19):

$$\begin{aligned} A_i = \big(& -1.2884, -0.1609, -1.1714, 1.0470, \\ & -1.4158, 1.0212, -0.8964, -1.0870, -0.3712, 0.4093, \\ & -1.3853, -0.2269, 0.0596, -0.8740, -1.2033, -1.4264, \\ & -0.7578, -0.9526, 0.3105, -0.1625, -0.4113, 0.4147, \\ & 1.0378, -1.0145, -0.5640, 0.3173, -0.2495, 0.6901, \\ & -0.3680, 0.3484, -0.8459, -0.2133, 0.5551, 0.0780, \\ & 0.5037, 0.5558, -1.3610, 0.3493, -0.1729, -0.3253 \big). \end{aligned} \tag{19}$$

The signal A_i is carefully restructured to form matrix $A_{j,\ell}$ as defined by Eq. (20):

$$A_{j,\ell} = \begin{pmatrix} -1.2884 & -0.1609 & -1.1714 & 1.0470 & -1.4158 \\ 1.0212 & -0.8964 & -1.0870 & -0.3712 & 0.4093 \\ -1.3853 & -0.2269 & 0.0596 & -0.8740 & -1.2033 \\ -1.4264 & -0.7578 & -0.9526 & 0.3105 & -0.1625 \\ -0.4113 & 0.4147 & 1.0378 & -1.0145 & -0.5640 \\ 0.3173 & -0.2495 & 0.6901 & -0.3680 & 0.3484 \\ -0.8459 & -0.2133 & 0.5551 & 0.0780 & 0.5037 \\ 0.5558 & -1.3610 & 0.3493 & -0.1729 & -0.3253 \end{pmatrix}. \tag{20}$$

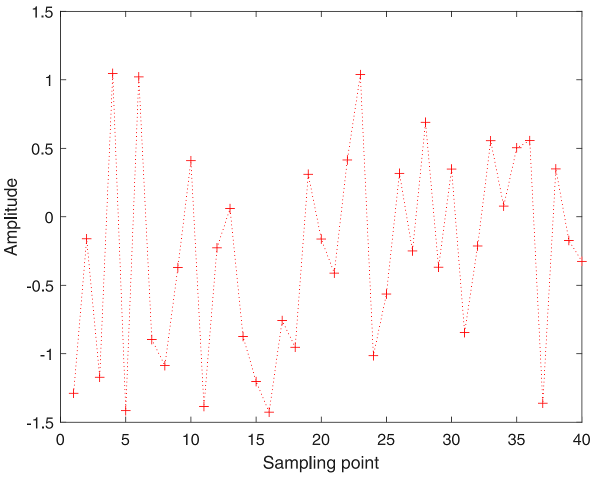

Fig. 1. Random signal.

Note, ℓ=5 is used to restructure A_i. However, as mentioned earlier, this article finds the value of ℓ using simulation runs. Next, $A_{J,\ell}$ is partitioned into two classes (that is, g=2) as shown in Eq. (21):

$$A_{J,\ell} = \left(\begin{array}{ccccc} -1.2884 & -0.1609 & -1.1714 & 1.0470 & -1.4158 \\ 1.0212 & -0.8964 & -1.0870 & -0.3712 & 0.4093 \\ -1.3853 & -0.2269 & 0.0596 & -0.8740 & -1.2033 \\ -1.4264 & -0.7578 & -0.9526 & 0.3105 & -0.1625 \\ \hline -0.4113 & 0.4147 & 1.0378 & -1.0145 & -0.5640 \\ 0.3173 & -0.2495 & 0.6901 & -0.3680 & 0.3484 \\ -0.8459 & -0.2133 & 0.5551 & 0.0780 & 0.5037 \\ 0.5558 & -1.3610 & 0.3493 & -0.1729 & -0.3253 \end{array}\right). \tag{21}$$

Therefore, the between-class variance ($C_{\mathcal{B}}$) is computed as in Eq. (8) to obtain:

$$C_{\mathcal{B}} = \begin{pmatrix} 1.3860 & -0.4614 & -0.2452 & -0.2193 & 0.5117 \\ -0.4614 & 0.5121 & 0.2135 & -0.1115 & -0.2382 \\ -0.2452 & 0.2135 & 1.3620 & -0.6237 & 0.2794 \\ -0.2193 & -0.1115 & -0.6237 & 0.8237 & -0.0234 \\ 0.5117 & -0.2382 & 0.2794 & -0.0234 & 1.1113 \end{pmatrix}. \tag{22}$$

Similarly, the within-class variance ($C_{\aleph}$) is computed as in Eq. (11) to obtain:

$$C_{\aleph} = \begin{pmatrix} 0.4539 & 0.1066 & 0.9741 & -0.2677 & 0.3933 \\ 0.1066 & 0.0250 & 0.2288 & -0.0629 & 0.0924 \\ 0.9741 & 0.2288 & 2.0907 & -0.5746 & 0.8441 \\ -0.2677 & -0.0629 & -0.5746 & 0.1579 & -0.2320 \\ 0.3933 & 0.0924 & 0.8441 & -0.2320 & 0.3408 \end{pmatrix} \tag{23}$$

The transformation matrix, $\mathcal{W}$ is subsequently computed as defined in Eq. (13):

$$\begin{aligned} \mathcal{W} &= C_{\aleph}^{-1} C_{\mathcal{B}} \\ &= \begin{pmatrix} 1.0962 & 0.2574 & 2.3527 & -0.6467 & 0.9499 \\ 0.7168 & 0.1684 & 1.5385 & -0.4229 & 0.6211 \\ 1.3628 & 0.3201 & 2.9249 & -0.8039 & 1.1809 \\ 1.0867 & 0.2552 & 2.3322 & -0.6410 & 0.9416 \\ -0.3170 & -0.0745 & -0.6804 & 0.1870 & -0.2747 \end{pmatrix}. \end{aligned} \tag{24}$$

The LDA variables can therefore be obtained by finding the SVD of $\mathcal{W}$ such that

$$\mathcal{W} = \mathcal{L}\mathcal{F}\mathcal{O}^{\dagger},$$

where,

$$\mathcal{L} = \begin{pmatrix} -0.4975 & 0.7154 & 0.0306 & 0.1502 & -0.4659 \\ -0.3254 & -0.1913 & 0.3612 & 0.7859 & 0.3307 \\ -0.6186 & 0.0783 & -0.1868 & -0.4179 & 0.6338 \\ -0.4932 & -0.6272 & 0.2234 & -0.2484 & -0.5017 \\ 0.1439 & 0.2282 & 0.8853 & -0.3514 & 0.1416 \end{pmatrix}, \tag{25}$$

$$\mathcal{F} = \begin{pmatrix} 5.7284 & 0 & 0 & 0 & 0 \\ 0 & 0.0000 & 0 & 0 & 0 \\ 0 & 0 & 0.0000 & 0 & 0 \\ 0 & 0 & 0 & 0.0000 & 0 \\ 0 & 0 & 0 & 0 & 0.0000 \end{pmatrix}, \tag{26}$$

$$\mathcal{O} = \begin{pmatrix} -0.3846 & -0.4449 & -0.5373 & -0.6000 & 0.0733 \\ -0.0903 & 0.0119 & 0.0277 & 0.1446 & 0.9849 \\ -0.8255 & 0.5415 & 0.1344 & -0.0033 & -0.0855 \\ 0.2269 & 0.2279 & 0.5308 & -0.7754 & 0.1170 \\ -0.3333 & -0.6758 & 0.6409 & 0.1335 & -0.060 \end{pmatrix}. \tag{27}$$

As mentioned, the eigenvectors, $\mathcal{P}$ are the entries in the columns of $\mathcal{O}$, which acts as the decomposed LDA variables. Hence, $\mathcal{O}$ is reconstructed using Eq. (17) to form the LDA feature vector as defined by Eq. (28):

$$\Bbbk_{\ell} = \begin{pmatrix} -1.4068 & -0.3395 & 0.7965 & -1.1006 & 1.0296 \end{pmatrix}. \tag{28}$$

Algorithm 1 summarises the proposed LDA feature extraction method for *Mysticetes'* vocalisations.

Algorithm 1: LDA Feature Vector.

Input: A_i, f, k
Output: *Feature vector,* Z_{LDA}
1: *Find the dimension,* ℓ *of the LDA components*
2: *Restructured* A_i *wrt* ℓ *as in Eq.* (1)
3: *Calculate* $\mathcal{C}_{\mathcal{B}}$ *as in Eq.* (8)
4: *Calculate* $\mathcal{C}_{\aleph}$ *as in Eq.* (11)
5: *Calculate* $\mathcal{W}$ *as in Eq.* (13)
6: *Calculate the eigenvector,* $\mathcal{P}$ *and eigenvalues,* Ω *as in Eq.* (14)
7: *Calculate the LDA components,* $\Bbbk_g$ *as in Eq.* (17) *and Eq.* (18)

Hidden Markov model

The HMM is a machine learning tool that has been deployed in a variety of applications, such as pattern recognition, data compression, speech recognition, and marine sound detection, among others [2,22,23]. HMM is a robust and flexible stochastic state machine, in which any change in the hidden sequence result in the emission of a symbol. The HMM operation can be perceived in two sequential steps, (1) training steps and (2) detection steps.

In the training step, crucial variables—namely, start probability (π), transition matrix ($\mathcal{T}$), and Gaussian emission distribution parameters ($\mathcal{E}$)—are estimated from the obtained feature vector, Z_{LDA}. The transition matrix, $\mathcal{T}$, contains probability values dictating state transitions, representing the feature vector as a sequence of states over time. The emission distribution parameters, $\mathcal{E}$, include mean (μ), covariance matrix ($\sum$), and mixture weight (σ). These parameters are computed using the expectation–maximisation (EM) algorithm, where the Baum–Welch (BM) algorithm [24] serves as a notable example. The $\mathcal{E}$ parameters start with a flat value during the EM process, and their flat values can significantly impact HMM's performance. To address this, the k-mean clustering algorithm (KMC) [25] and Gaussian Mixture Model (GMM) are sequentially employed to initialise the $\mathcal{E}$ parameters as depicted in Fig. 2.

During the detection step, the trained $\mathcal{E}$ parameters are matched with the sound waveform to be recognised, utilising the Viterbi algorithm (V-alg) [26]. The V-alg processes the estimated start probability, π, transition matrix, $\mathcal{T}$, and the extracted feature vector from the sound to be recognised, stochastically producing the closely matched hidden sequence.

Simulation parameters and set-up

Mysticetes also called Baleen whales are a class of the suborder of cetacean. *Mysticetes* use their horny plates called Baleen which hang from their upper jaws to filter food from water [11,27]. As documented in [27], there are about fourteen different types of Baleen whales in existence. Examples include Sei whales, Bryde's whales, Humpback whales, Fin whales, and so on. However, this article uses two different *Mysticetes'* vocalisation datasets obtain from *Mobysound.org* to theoretically document the performance of the developed LDA-HMM detector. The first sound dataset is a Humpback whale song gathered off the north coast of Kaui, Hawaii, USA while the other is a Bryde's whale pulse gathered from Eastern Tropical Pacific as detailed on *Mobysound.org*. The two sound datasets are visually and aurally analysed using software called Sonic Visualiser. This visualiser also aided the annotation of the

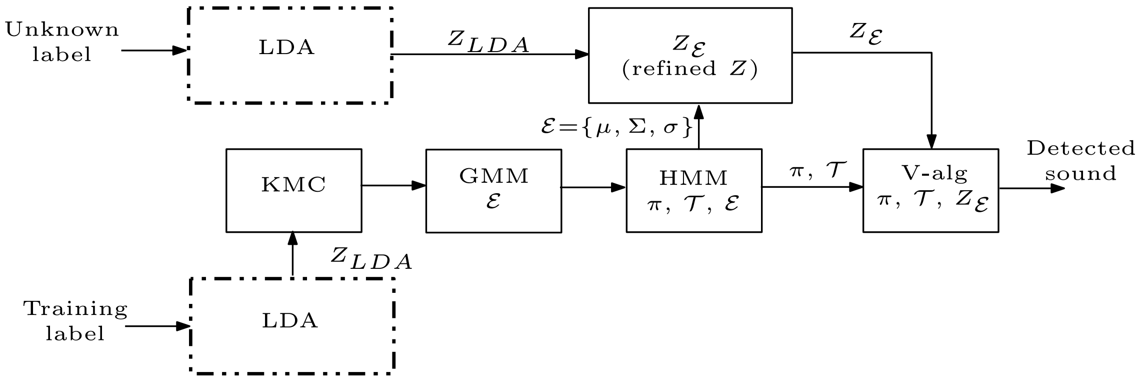

Fig. 2. LDA-HMM detector for *Mysticetes'* vocalisations.

sound dataset, which were categorised as either *Mysticetes'* vocalisations or noise (note that any sound which is not the *Mysticetes'* vocalisations is taken as noise). In addition, these datasets are split into two unequal portions. A small portion of about 25% is used in training the HMM and a large portion is used to verify the performance of the LDA-HMM detector. The number of the training sample, t extracted from the small portion is varied to document the LDA-HMM detector performance for different t. Likewise, each training sample is approximately divided into f number of sampling points. This implies that f and t are the two important parameters that are used to validate the developed LDA-HMM detector performance.

Fig. 2 presents the schematic diagram of the developed LDA-HMM detector. As indicated in the figure, the derived feature vector from the training label (the small portion) is first fed through the KMC block, which starts the process of initialising the Gaussian emission distribution parameters, $\mathcal{E} = \{\mu, \sum, \sigma\}$. Subsequently, these parameters are further refined in the GMM block before being fed into the HMM block. The set-up adopts a *4 states* and *2 mixture weight* ergodic type HMM. Two HMMs are used to model the *Mysticetes'* vocalisations and noise. That is, one HMM for the noise signal (HMM_1) and the other for the *Mysticetes'* vocalisations waveform (HMM_2). HMM_1 and HMM_2 are both *4 states* and *2 mixture weight* ergodic HMM. After training the two HMMs separately, HMM_1 and HMM_2 are combined to form a *8 states* and *4 mixture weights* ergodic type HMM. The first 1–4 states represent one of the HMM while the next 5–8 states represent the other HMM. In this case, states 1–4 represent the noise signal while states 5–8 represent the *Mysticetes'* vocalisations waveform. The trained Gaussian parameters are then used to refine the extracted feature vector from the unknown sound label (the large sound portion) to obtain a refined feature vector. This refined feature vector, the transition matrix, $\mathcal{T}$ and the start probability, π are used in the V-alg block to recognise the unknown label as either the *Mysticetes'* vocalisations or noise. It is important to mention that the V-alg jumps from states 1–4 to states 5–8 with an equal transition probability.

Results and discussion

LDA-HMM performance for different ℓ

The results of the developed LDA-HMM detector for *Mysticetes'* vocalisations is presented in Tables 1–5 for different dimension, ℓ in terms of the sensitivity and the FDR performances. As defined in [11,21], the sensitivity is the detection accuracy of the detector while the FDR is a measure of the reliability of the detector. A high value of the sensitivity indicates a good detection accuracy. On the other hand, a low value of the LDA is desirable as it indicates that the output of the detector can be trusted. As shown in the tables, the performance of the LDA-HMM detector is documented for ℓ= 6, 7, 8, 9, and 10 for both the Bryde's whale pulses and Humpback whale songs. In addition, the number of the training samples, t is varied from t= 6, 12, and 18 for the Bryde's whale pulses and t= 18, 24, and 30 for the Humpback whale song. Note that the number of training samples for the Humpback whale song is more as compared to the Bryde's whale pulses because the Humpback whale waveform is longer than that of the Bryde's whale pulses. As such, more training samples are used to reliably train the HMM. Also, note that the number of sampling points, f for each training sample is varied from f=100, 200, 300, 400, and 500.

Observed from Tables 1–5, as the value of ℓ increase from 6 to 7, the performance of the LDA-HMM detector improves for both *Mysticetes'* species. In Table 5 (Bryde's whale pulses) for example, for t=18 and f =100 the LDA-HMM detector offers a sensitivity performance gain of 0.61, and a reduction of 0.02 in the FDR performance. Likewise, in the same table, for f =100 and t=30 for the Humpback whale songs, the LDA-HMM detector offers a sensitivity performance gain of 4.31, and a reduction of 0.9 in the FDR performance. Nevertheless, it can be noted that a further increase in the value of ℓ; that is, from ℓ = 8, 9 and 10, the LDA-HMM detector shows no significant improvement in the FDR and sensitivity performances. This will only increase the computational time complexity of the HMM training process. In this regard, a dimension ℓ=7 is chosen as the close to optimal ℓ for the LDA-HMM detector because it strikes a balance between the detectors' computational time complexity and its performance gain.

Also, observe from the tables that the performance of the detector falls as the value of the sampling points, f increases from 100 to 500. For instance, as f increases from 100 to 500, the LDA-HMM detector sensitivity performance falls from 98.61% to 91.99% for

Table 1
LDA-HMM performance for different ℓ: f=500.

Bryde's whale pulses						
ℓ	Sensitivity (%)			FDR(%)		
	t=18	t=12	t=6	t=18	t=12	t=6
6	91.36	90.01	88.99	0.12	0.13	0.15
7	91.99	90.74	90.21	0.11	0.12	0.13
8	91.99	90.74	90.21	0.11	0.12	0.13
9	91.99	90.74	90.21	0.11	0.12	0.13
10	91.99	90.75	90.21	0.11	0.12	0.13
Humpback whale songs						
ℓ	Sensitivity (%)			FDR(%)		
	t=30	t=24	t=18	t=30	t=24	t=18
6	78.75	76.65	73.51	3.00	4.18	5.23
7	81.83	79.91	77.22	2.01	3.09	3.97
8	81.84	79.91	77.22	2.01	3.09	3.98
9	81.83	79.92	77.22	2.00	3.09	3.98
10	81.84	79.91	77.22	2.01	3.09	3.98

Table 2
LDA-HMM performance for different ℓ: f=400.

Bryde's whale pulses						
ℓ	Sensitivity (%)			FDR(%)		
	t=18	t=12	t=6	t=18	t=12	t=6
6	92.92	91.96	91.13	0.10	0.11	0.13
7	93.81	92.38	91.42	0.08	0.09	0.11
8	93.82	92.38	91.42	0.08	0.09	0.11
9	93.81	92.38	91.42	0.08	0.09	0.11
10	93.81	92.38	91.42	0.08	0.09	0.11
Humpback whale songs						
ℓ	Sensitivity (%)			FDR(%)		
	t=30	t=24	t=18	t=30	t=24	t=18
6	79.52	77.00	73.66	2.70	3.96	5.01
7	82.16	80.01	77.29	1.42	2.49	3.70
8	82.16	80.01	77.29	1.42	2.49	3.70
9	82.16	80.01	77.29	1.42	2.50	3.70
10	82.16	80.01	77.29	1.42	2.49	3.70

Table 3
LDA-HMM performance for different ℓ: f=300.

Bryde's whale pulses						
ℓ	Sensitivity (%)			FDR(%)		
	t=18	t=12	t=6	t=18	t=12	t=6
6	94.10	93.08	92.46	0.10	0.11	0.12
7	95.00	94.11	93.38	0.08	0.09	0.10
8	95.00	94.11	93.38	0.08	0.09	0.10
9	95.00	94.11	93.38	0.08	0.09	0.10
10	95.00	94.11	93.38	0.08	0.09	0.10
Humpback whale songs						
ℓ	Sensitivity (%)			FDR(%)		
	t=30	t=24	t=18	t=30	t=24	t=18
6	80.03	77.06	75.51	2.33	3.21	4.18
7	85.11	82.22	80.68	1.11	2.00	3.18
8	85.11	82.22	80.68	1.11	2.00	3.18
9	85.12	82.22	80.68	1.11	2.00	3.18
10	85.11	82.22	80.68	1.11	2.00	3.18

ℓ=7 and t=18 (Bryde's whale pulses) while its FDR performance rises from 0.05% to 0.11%. Likewise, as f increases from 100 to 500, the LDA-HMM detector sensitivity performance falls from 89.88% to 81.84% for ℓ=7 and t=30 (Humpback whale songs) while its FDR performance rises from 1.10% to 2.01%. This fall in the LDA-HMM detector performance is expected when f increases.

Table 4
LDA-HMM performance for different ℓ: f=200.

Bryde's whale pulses						
ℓ	Sensitivity (%)			FDR(%)		
	t=18	t=12	t=6	t=18	t=12	t=6
6	96.22	94.99	94.87	0.09	0.10	0.11
7	97.10	95.88	94.87	0.07	0.08	0.09
8	97.11	95.89	94.87	0.07	0.08	0.09
9	97.11	95.88	94.87	0.07	0.08	0.09
10	97.11	95.88	94.87	0.07	0.08	0.09
Humpback whale songs						
ℓ	Sensitivity (%)			FDR(%)		
	t=30	t=24	t=18	t=30	t=24	t=18
6	85.34	80.69	78.62	1.96	2.40	3.01
7	88.38	84.79	81.81	1.03	1.78	2.33
8	88.37	84.80	81.81	1.03	1.78	2.33
9	88.38	84.79	81.82	1.03	1.78	2.33
10	88.37	84.79	81.80	1.03	1.78	2.33

Table 5
LDA-HMM performance for different ℓ: f=100.

Bryde's whale pulses						
ℓ	Sensitivity (%)			FDR(%)		
	t=18	t=12	t=6	t=18	t=12	t=6
6	98.00	97.04	96.21	0.07	0.08	0.09
7	98.61	97.69	96.58	0.05	0.06	0.07
8	98.61	97.69	96.58	0.05	0.06	0.07
9	98.61	97.69	96.58	0.05	0.06	0.07
10	98.61	97.69	96.58	0.05	0.06	0.07
Humpback whale songs						
ℓ	Sensitivity (%)			FDR(%)		
	t=30	t=24	t=18	t=30	t=24	t=18
6	85.57	81.00	81.20	2.00	2.36	2.99
7	89.88	85.70	82.29	1.10	1.45	1.98
8	89.88	85.70	82.29	1.10	1.45	1.98
9	89.87	85.71	82.29	1.10	1.45	1.98
10	89.88	85.70	82.29	1.10	1.45	1.98

This is typical with most detectors because more reliable features can be extracted from smaller frames. Nonetheless, in order not to increase the computational time complexity, a moderately small frame size should be used; most especially, when the signal duration is long.

Additionally, note from the tables that the performance of the LDA-HMM detector is enhanced with an increase in the value of t for both species. In Table 5 (Bryde's whale pulses) for example, for ℓ=7 and f=100 the LDA-HMM detector offers a sensitivity performance gain of 2.03% and a reduction of 0.02 in the FDR performance as t increases from 6 to 18. Similarly, in the same table (Humpback whale songs), for ℓ=7 and f=100 the LDA-HMM detector offers a sensitivity performance gain of 7.59% and a reduction of 0.88% in the FDR performance as t increases from 18 to 30. Therefore, the performance of the LDA-HMM detector can be enhanced by increasing the value of t for each species. It is however important to mention that the performance of the detector will eventually saturates when t reaches a particular value.

LDA-HMM comparison with existing methods

Tables 6–10 present the simulation runs of the developed LDA-HMM detector with other recent methods in the literature. The tables compare the LDA-HMM detector for ℓ =7 with the MFCC-HMM [10,11], LPC-HMM [10,11], EMD-HMM [10] and DMD-HMM [11] detectors. Observe from the tables that the number of filter coefficients for the LPC (υ) and MFCC (α) is 12 ($\upsilon = \alpha$=12). Hence, they are both 12-dimensional models. The DMD-HMM detector assumes a rank r=9 as deployed in [11]. That is, the DMD-HMM detector is a 9-dimensional model. In the case of the EMD-HMM detector, the intrinsic mode functions (IMFs), $\mathcal{I}$, which specifies the dimension of the model varies for the two *Mysticetes'* species. Therefore, $\mathcal{I}$=6 is assumed for the Bryde's whale pulses while $\mathcal{I}$=9 is used for the Humpback whale songs because it produces longer waveform. This means that the Humpback whale songs use a 9-dimensional model and the Bryde's whale pulses use a 6-dimensional model as deployed in [10].

From the tables, consider the performance of LDA-HMM detector in comparison to the DMD-HMM detector. The LDA-HMM detector outperforms the DMD-HMM detector both in terms of the FDR and sensitivity performances for different values of t and

Table 6
LDA-HMM comparison with existing techniques: f=500.

Bryde's whale LDA_ℓ=7, DMD_r=9, EMD_I=6, $MFCC_\alpha$=12, LPC_υ=12										
t	Sensitivity (%)					FDR (%)				
	LDA	DMD	EMD	MFCC	LPC	LDA	DMD	EMD	MFCC	LPC
18	91.99	90.91	87.32	85.61	82.33	0.11	0.14	0.17	0.20	1.31
12	90.74	90.00	86.79	84.43	81.51	0.12	0.15	0.20	0.23	1.43
6	90.21	89.34	86.22	84.01	80.95	0.13	0.16	0.22	0.34	1.60
Humpback whale LDA_ℓ=7, DMD_r=9, EMD_I=6, $MFCC_\alpha$=12, LPC_υ=12										
t	Sensitivity (%)					FDR (%)				
	LDA	DMD	EMD	MFCC	LPC	LDA	DMD	EMD	MFCC	LPC
30	81.83	80.90	77.97	75.10	72.18	2.01	2.63	3.64	4.24	5.29
24	79.91	78.82	75.23	73.11	70.00	3.09	4.01	4.83	5.33	6.20
18	77.22	76.11	73.41	71.14	68.59	3.97	5.00	5.96	6.19	7.11

Table 7
LDA-HMM comparison with existing techniques: f=400.

Bryde's whale LDA_ℓ=7, DMD_r=9, EMD_I=6, $MFCC_\alpha$=12, LPC_υ=12										
t	Sensitivity (%)					FDR (%)				
	LDA	DMD	EMD	MFCC	LPC	LDA	DMD	EMD	MFCC	LPC
18	93.81	92.81	89.39	87.49	84.38	0.08	0.12	0.15	0.17	1.29
12	92.38	91.40	88.16	86.60	83.38	0.09	0.13	0.18	0.22	1.42
6	91.42	90.99	87.73	85.80	82.59	0.11	0.14	0.19	0.31	1.60
Humpback whale LDA_ℓ=7, DMD_r=9, EMD_I=6, $MFCC_\alpha$=12, LPC_υ=12										
t	Sensitivity (%)					FDR (%)				
	LDA	DMD	EMD	MFCC	LPC	LDA	DMD	EMD	MFCC	LPC
30	82.16	81.11	78.69	76.17	73.71	1.42	2.40	3.30	4.00	5.00
24	80.01	77.27	76.06	74.17	71.10	2.49	3.60	4.41	5.00	5.93
18	77.29	76.35	74.21	72.12	69.92	3.70	4.65	5.60	6.00	6.74

Table 8
LDA-HMM comparison with existing techniques: f=300.

Bryde's whale LDA_ℓ=7, DMD_r=9, EMD_I=6, $MFCC_\alpha$=12, LPC_υ=12										
t	Sensitivity (%)					FDR (%)				
	LDA	DMD	EMD	MFCC	LPC	LDA	DMD	EMD	MFCC	LPC
18	95.00	94.01	90.71	88.49	85.61	0.08	0.11	0.14	0.16	1.30
12	94.11	93.22	90.00	87.70	84.27	0.09	0.12	0.16	0.22	1.35
6	93.38	92.38	88.80	86.54	83.19	0.10	0.13	0.17	0.29	1.49
Humpback whale LDA_ℓ=7, DMD_r=9, EMD_I=6, $MFCC_\alpha$=12, LPC_υ=12										
t	Sensitivity (%)					FDR (%)				
	LDA	DMD	EMD	MFCC	LPC	LDA	DMD	EMD	MFCC	LPC
30	85.11	84.11	81.74	78.83	75.23	1.11	2.03	3.00	3.59	4.29
24	82.22	81.21	77.91	75.25	72.55	2.00	3.03	4.00	4.79	5.31
18	80.68	79.73	75.92	72.17	70.00	3.18	4.24	5.00	5.46	6.15

f. Moreover, the LDA-HMM detector will be more suitable for real-time detection of *Mysticetes'* vocalisations because it is a 7-dimensional model while the DMD-HMM detector is a 9-dimensional model. This means that the DMD feature extraction method will increase the computational time complexity of the HMM training process in comparison to the LDA method; thus, it will slow down the detection process when adopted for real-time applications.

Furthermore, consider the performance of the LDA-HMM detector in comparison to the EMD-HMM detector. As documented, for different values of t and f, the LDA-HMM detector offers better performance as compared to the EMD-HMM detector. In Table 10 (Bryde's whale pulses) for example, for t=18 and f=100, the LDA-HMM detector exhibits a sensitivity performance gain of 4.11% and a reduction of 0.05% in the FDR performance in comparison to the EMD-HMM detector. In the same table, for t=30 and f=100 for the Humpback whale songs, the LDA-HMM detector exhibits a sensitivity performance gain of 1.20% and a reduction of 1.06%

Table 9
LDA-HMM comparison with existing techniques: f=200.

Bryde's whale LDA_ℓ=7, DMD_r=9, EMD_I=6, $MFCC_\alpha$=12, LPC_υ=12										
t	Sensitivity (%)					FDR (%)				
	LDA	DMD	EMD	MFCC	LPC	LDA	DMD	EMD	MFCC	LPC
18	97.10	96.11	93.15	91.70	88.77	0.07	0.10	0.12	0.14	1.26
12	95.88	94.95	92.71	90.10	87.50	0.08	0.11	0.14	0.19	1.31
6	94.87	94.00	91.61	88.82	86.19	0.09	0.12	0.15	0.25	1.44
Humpback whale LDA_ℓ=7, DMD_r=9, EMD_I=6, $MFCC_\alpha$=12, LPC_υ=12										
t	Sensitivity (%)					FDR (%)				
	LDA	DMD	EMD	MFCC	LPC	LDA	DMD	EMD	MFCC	LPC
30	88.38	87.20	85.05	82.35	79.90	1.03	1.79	2.47	3.02	3.92
24	84.79	83.84	80.35	78.22	76.30	1.78	2.23	3.20	3.90	5.01
18	81.81	80.80	76.81	74.46	72.20	2.33	2.80	4.00	4.60	5.79

Table 10
LDA-HMM comparison with existing techniques: f=100.

Bryde's whale LDA_ℓ=7, DMD_r=9, EMD_I=6, $MFCC_\alpha$=12, LPC_υ=12										
t	Sensitivity (%)					FDR (%)				
	LDA	DMD	EMD	MFCC	LPC	LDA	DMD	EMD	MFCC	LPC
18	98.61	97.73	94.50	92.72	90.32	0.05	0.08	0.10	0.12	1.21
12	97.69	96.90	93.59	91.26	89.13	0.06	0.09	0.13	0.17	1.25
6	96.58	95.10	92.71	89.93	87.75	0.07	0.11	0.14	0.23	1.38
Humpback whale LDA_ℓ=7, DMD_r=9, EMD_I=6, $MFCC_\alpha$=12, LPC_υ=12										
t	Sensitivity (%)					FDR (%)				
	LDA	DMD	EMD	MFCC	LPC	LDA	DMD	EMD	MFCC	LPC
30	88.90	88.56	87.70	83.85	81.24	1.10	1.58	2.16	2.53	3.35
24	84.68	84.37	81.48	79.92	77.85	1.45	2.01	2.80	3.22	4.39
18	81.31	81.00	77.93	75.80	73.48	1.98	2.44	3.20	4.03	5.13

in FDR performance in comparison to the EMD-HMM detector. However, the EMD-HMM detector offers less computational time complexity as compared to the LDA-HMM detector when used for Bryde's whale pulse detection. This is because the EMD-HMM detector is a 6-dimensional model while the LDA-HMM detector is a 7-dimensional model. This means that there is a trade-off between the performance gain and computational time complexity in using the EMD-HMM detector or LDA-HMM detector for detecting the Bryde's whale pulses. Yet, considering the big difference in performance gain between the LDA-HMM and EMD-HMM detectors, the LDA-HMM detector will most likely be employed. It is however important to emphasise that such decision criteria mostly depend on the application requirement. On the other hand, to detect the Humpback whale songs, the LDA-HMM detector offers less computational time complexity in comparison to the EMD-HMM detector. Note that the LDA-HMM detector is a 7-dimensional model while the EMD-HMM detector is a 9-dimensional model. Therefore, the LDA-HMM is more suitable for the real-time detection of Humpback whale songs in terms of performance gain and computational time complexity.

Lastly, consider the performance of the LDA-HMM detector in comparison to the MFCC-HMM and LPC-HMM detectors. As documented, for different values of t and f, the LDA-HMM detector exhibits better performance as compared to MFCC-HMM and LPC-HMM detectors. More so, the MFCC-HMM and LPC-HMM detectors increase the computational time complexity of the HMM, as they are both 12-dimensional models. Thus, the LDA-HMM is more suitable for the real-time detection of *Mysticetes'* vocalisations in terms of the computational time complexity and FDR and sensitivity performances as compared to the MFCC-HMM and LPC-HMM detectors.

Conclusion

The method of LDA was introduced in this article as a feature extraction technique that can be deployed with the HMM to automatically detect the vocalisations of *Mysticetes*. The LDA technique transforms the information in a large data matrix into a lower dimensional space without losing the information contained in the large data matrix. Thus, the feature contained in this lower dimensional space is reconstructed in this article to serve as a suitable feature vector for the HMM. The LDA-HMM detector was shown to be effective and reliable in comparison to the DMD-HMM and EMD-HMM detectors, and the traditional LPC-HMM and MFCC-HMM detection techniques. The LDA-HMM detector offered a better trade-off between the FDR and sensitivity performances and computational time complexity as compared to these existing methods. The performance of this developed LDA-HMM method was documented using two different *Mysticetes'* vocalisations. However, there are about 14 types of *Mysticetes'* that produce similar vocalisations; including, Sei whales, Fin whales, Gray whales, and so on. Therefore, in the future, the performance of this proposed LDA-HMM technique can be verified for the other *Mysticetes'* vocalisations.

CRediT authorship contribution statement

O.M. Ogundile: Software, Data curation, Investigation, Writing – original draft. **A.A. Owoade:** Visualization, Supervision, Conceptualization. **O.O. Ogundile:** Methodology, Software, Editing, Writing – reviewing. **O.P. Babalola:** Software, Validation, Writing – reviewing & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. Van Wyk, J. Versfeld, J. du Preez, Detection of cetacean and fish sounds using normalized summation of harmonics and spectrogram masking, JASA Express Lett. 3 (9) (2023) 096002.
[2] A.M. Usman, O.O. Ogundile, D.J.J. Versfeld, Review of automatic detection and classification techniques for Cetacean vocalization, IEEE Access 8 (2020) 105181–105206.
[3] J.S. Smith, L.K. Hill, R.M. Gonzalez, Whale watching and preservation of the environment in central Baja California, Mexico. FOCUS Geogr. 62 (1) (2019).
[4] M.W. Rademan, D.J.J. Versfeld, J. du Preez, Soft-output signal detection for cetacean vocalizations using spectral entropy, k-means clustering and the continuous wavelet transform, Ecol. Inf. 74 (2023) 101990.
[5] O.O. Ogundile, O.P. Babalola, S.G. Odeyemi, K.I. Rufai, Hidden Markov models for detection of Mysticetes vocalisations based on principal component analysis, Bioacoustics 31 (6) (2022) 710–738.
[6] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis, Proc. R. Soc. London, Ser. A: Math. Phys. Eng. Sci. 454 (1971) (1988) 903–995.
[7] P.J. Schmid, J. Sesterhenn, Dynamic mode decomposition of numerical and experimental data, J. Fluid Mech. 656 (2010) 5–28.
[8] S. Majeed, H. Husain, S. Abdul Samad, T. Idbeaa, Mel frequency cepstral coefficients (MFCC) feature extraction enhancement in the application of speech recognition: A comparison study, J. Theoret. Appl. Inf. Technol. 79 (2015) 38–56.
[9] J. Makhoul, Linear predicion: A tutorial review, Proc. IEEE 63 (1975) 561–580.
[10] O.O. Ogundile, A.M. Usman, D.J.J. Versfeld, An empirical mode decomposition based hidden Markov model approach for detection of Bryde's whale pulse calls, J. Acoust. Soc. Am. 147 (2) (2020) EL125–EL131.
[11] O.O. Ogundile, A.M. Usman, O.P. Babalola, D.J.J. Versfeld, Dynamic mode decomposition: A feature extraction technique based hidden Markov model for detection of Mysticetes' vocalisations, Ecol. Informat. 63 (2021) 101306.
[12] R. Alghamdi, Hidden Markov models (HMMs) and security applications, Int. J. Adv. Comput. Sci. Appl. 17 (2) (2016) 39–47.
[13] R. Yao, M. Johnson, P. Clemins, D. Michael, S. Glaeser, T. Osiejuk, O.-N. Ebenezer, A framework for bioacoustic vocalization analysis using hidden Markov models, Algorithms 2 (2009) 1410–1428.
[14] A. Tharwat, T. Gaber, A. Ibrahim, A.E. Hassanien, Linear discriminant analysis: A detailed tutorial, AI Commun. 30 (2) (2017) 169–190.
[15] R. Duda, P. Hart, D. G.Stork, Pattern Classification, second ed., 2001.
[16] J. Lu, K.N. Plataniotis, S. Venetsanopoulos, T. Idbeaa, Face recognition using LDA-based algorithms, IEEE Trans. Neural Netw. 14 (1) (2003) 195–200.
[17] J. Cui, Y. Xu, Three dimensional palmprint recognition using linear discriminant analysis method, in: Proceedings of the second International Conference on Innovations in Bioinspired Computing and Applications, IBICA, IEEE, 2011, pp. 107–111.
[18] M.C. Wu, L. Zhang, Z. Wang, D.C. Christiani, X. Lin, Sparse linear discriminant analysis for simultaneous testing for the significance of a gene set/pathway and gene selection, Bioinformatics 25 (9) (2009) 1145–1151.
[19] D. Coomans, D. Massart, L. Kaufman, Optimization by statistical linear discriminant analysis in analytical chemistry, Anal. Chim. Acta 112 (2) (1979) 97–122.
[20] O.O. Ogundile, D.J.J. Versfeld, Analysis of template-based detection algorithms for inshore Bryde's whale short pulse calls, IEEE Access 8 (2020) 14377–14385.
[21] O.O. Ogundile, A.M. Usman, O.P. Babalola, D.J.J. Versfeld, A hidden Markov model with selective time domain feature extraction to detect inshore Bryde's whale short pulse calls, Ecol. Inform. 57 (2020) 101087.
[22] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE 77 (1989) 257–286.
[23] L.R. Rabiner, B.H. Juang, An introduction to hidden Markov models, IEEE ASSP Mag. 3 (1986) 4–16.
[24] L.E. Baum, T. Petrie, G. Soules, N. Weiss, A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, Ann. Math. Stat. 41 (1970) 164–171.
[25] E.W. Forgy, Cluster Analysis of Multivariate Data: Efficiency versus Interpretability of Classification, 1965.
[26] A. Viterbi, Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, IEEE Trans. Inform. Theory 13 (1967) 260–269.
[27] J. Kennedy, 14 Types of Baleen whales, 2020, ThoughtCo, https://www.thoughtco.com/types-of-baleen-whales-2291520.