Hidden Markov models for detection of Mysticetes vocalisations based on principal component analysis

Olayinka O. Ogundile, Oluwaseyi P. Babalola, Seun G. Odeyemi & Kazeem I. Rufai

To cite this article: Olayinka O. Ogundile, Oluwaseyi P. Babalola, Seun G. Odeyemi & Kazeem I. Rufai (2022): Hidden Markov models for detection of Mysticetes vocalisations based on principal component analysis, Bioacoustics, DOI: 10.1080/09524622.2022.2047786

To link to this article: https://doi.org/10.1080/09524622.2022.2047786

Published online: 25 Mar 2022.

Submit your article to this journal

Article views: 49

View related articles

View Crossmark data
Hidden Markov models for detection of Mysticetes vocalisations based on principal component analysis

Olayinka O. Ogundile\textsuperscript{a,b}, Oluwaseyi P. Babalola\textsuperscript{c}, Seun G. Odeyemi\textsuperscript{d} and Kazeem I. Rufai\textsuperscript{a}

\textsuperscript{a}Department of Computer Science, Tai Solarin University of Education, Nigeria; \textsuperscript{b}Department of Electrical and Electronic Engineering, Stellenbosch University, Stellenbosch, South Africa; \textsuperscript{c}Department of Electrical, Electronics and Computer Science Engineering, Cape Peninsula University of Technology, Bellville, South Africa; \textsuperscript{d}School of Computing, Electronics and Mathematics, Coventry University, UK

\textbf{ABSTRACT}

The economic relevance of Mysticetes has prompted marine ecologists and biologists to investigate this suborder of cetaceans. Mysticetes produce distinct vocal repertoires, which are recorded to analyse the behaviour of the species within its ecology. Passive acoustic monitoring (PAM) is a standard technique for tracking Mysticete movement and vocalisation. PAM collects enormous datasets over a long period, making it practically impossible to analyse with typical visual examination methods. Machine learning (ML) techniques such as hidden Markov models (HMMs) have made automatic recognition and analysis of extensive sound recordings possible. Nevertheless, the performance of ML tools is determined by the adopted feature extraction technique. Hence, this article introduces the method of principal component analysis (PCA) as a performance-efficient alternative feature extraction technique for detecting Mysticete vocalisations using HMM. Performance of the developed PCA-HMM detector is compared with state-of-the-art detectors using two different Mysticete vocalisations (Humpback whale songs and Bryde’s whale short pulses). In both species, results show that the PCA-HMM detector has the best performance and is more suitable for use in real-time application since it exhibits less computational time complexity.

1. Introduction

Automatic detection of Mysticetes vocalisations has attracted the attentions of biologists and marine ecologists, particularly since the advent of machine learning (ML) tools. Mysticete vocalisations, like those of other cetacean suborders, produce cryptic calls that cannot be easily analysed using typical visual inspection. Passive acoustic monitoring (PAM) is often used to gather these calls over days, months, and years. As a result, the collected dataset is enormous and cannot be manually analysed, either by perceiving spectrograms or carefully listening to the sound recordings. In this regard, different ML techniques such as hidden Markov models (HMMs) (Rabiner and Juang 1986; Rabiner 1989; Jurafsky and Martin 2008; Ogundile et al. 2020a, 2021b), support vector machines
(SVMs) (Roch et al. 2008), Gaussian mixture models (GMMs) (Richard et al. 2001; Reynolds 2009; Ogundile et al. 2021a) have been used extensively in the automatic detection of Mysticete vocalisations. The effectiveness of various ML algorithms vary based on the experimental conditions, types of vocalisations, species, and application requirements.

Figure 1 depicts a simple block diagram of a ML detection and classification system for Mysticete vocalisations. From the figure, the sound data is first collected using PAM under different conditions such as locations, time, and so on, depending on the application requirements. The raw sound data is preprocessed by filtering the unwanted frequency bands alongside background noise and DC components. Subsequently, some of the sound signals are identified manually from the datasets by human inspection. Distinctive features are extracted from the identified sound signal, which are used to train the system in preparation for testing and detection/classification. At the testing stage, features are obtained from the unknown datasets, which are further refined using the trained ML parameters. The refined feature vector is then passed through a detector such that the vector is compared with the estimated ML parameters to obtain a desired output.

The HMM is a ML tool that is widely employed in a variety of applications including speech processing, source coding, weather forecast, cloud computing, and so on. HMM has also been employed in the identification of animal sounds, particularly enigmatic cetacean vocalisations, since it is flexible, robust, and can be easily used to detect sounds from a given set of observations (Yao et al. 2009). Regardless of the robustness of the HMM and most ML tools, their detection accuracy and reliability depend on the feature extraction technique used in the detection process as shown in Figure 1. Several feature extraction techniques have been combined with the HMMs and other ML tools to automatically detect Mysticete vocalisations, including the time domain method (Babalola et al. 2021), Mel-scale frequency cepstral coefficients (MFCC) (Majeed et al. 2015; Putland et al. 2017; Ogundile et al. 2020a), empirical mode decomposition (EMD) (Huang et al. 1998; Liu et al. 2019; Ogundile et al. 2020b), linear predictive coefficient (LPC) (Makhoul 1975; Ogundile and Versfeld 2020; Ogundile et al. 2020a), and dynamic mode decomposition (DMD) (Ogundile et al. 2021b). The performance of HMM heavily depends on the adopted feature extraction technique used to mathematically derive features (feature vector) from the sound signal. In fact, the higher the reliability of the extracted feature vector, the better the performance of the HMMs. Thus, it is important that careful attention is given to the feature extraction methods to enhance the accuracy of the automated detection technique.

**Figure 1.** Block diagram of a ML detection and classification system for Mysticete vocalisations.
This article introduces the principal component analysis (PCA) method as a feature extraction technique that can be adopted with the HMM to detect Mysticete vocalisations. PCA is an optimization technique for displaying features in multivariate datasets (Bartholomew 2010; Abbas F.M. Alkarkhi 2019). The technique reduces the dimension of large datasets by transforming variables from a large set to a smaller set without losing important information. Hence, the reduced variable set can be easily visualised and analysed using different ML tools. The performance of the proposed PCA-HMM detection technique is theoretically demonstrated using acoustic datasets of different Mysticete vocalisations, collected from a referenced cetacean sound database (Mobysound.org1 Mellinger and Clark (2006)). In addition, the proposed PCA-HMM detection technique is compared to existing techniques in the literature such as the EMD-HMM, MFCC- HMM, LPC-HMM and DMD-HMM detectors using sensitivity (detection accuracy) and false discovery rate (FDR: detection reliability) as performance metrics.

This article is arranged as follows. Section 2 describes the Mysticetes cetacean sub-order and explains the characteristics of the vocalisations produced by two Mysticetes used for results verification in this paper. In Section 3, the different feature extraction techniques used in the literature for the detection of Mysticete vocalisations and for comparison in this study are discussed. In Section 4, the proposed PCA feature vector is discussed in detail. The concepts of HMM, k-means clustering, and GMMs are explained in Section 5. Section 6 presents the simulation set-up, datasets used for result verifications, and parameters. Section 7 discusses the results obtained from the automated PCA-HMM detection system for Mysticete vocalisations under various experimental conditions. Section 8 summarises with possible research direction.

2. Mysticete vocalisations

Mysticetes are species of the order Cetacea. Mysticetes filter food from water using a series of horny plates called ‘Baleen’ that hangs from their upper jaw; as such, they are called Baleen whales (Kenenedy 2020; Ogundile et al. 2021b). These Baleen whales produce different characteristic calls that are of interest to marine ecologists due to the economic importance and value of these whales (Usman et al. 2020). For instance, the whale watching business produces over US$2 billion yearly (Usman et al. 2020). This in return boosts the income of governments while employing thousands of people. Thus, the detection of Mysticete vocalisations has gained attention of marine ecologist and biologist over decades. Presently, there are about fourteen types of Mysticete such as Humpback whales, Fin whales, Sei whales, Gray whales, Bryde’s whales and so on (Kenenedy 2020). However, in this paper, only Bryde’s whale pulse calls and the Humpback whale songs are used to verify the PCA-HMM detection system; thus they are discussed accordingly in subsequent sections.

2.1. Brydes whales

Bryde’s whale was sighted in the far end south-west of South Africa for the first time during a survey of marine mammals in 1913 (Olsen 2009). Then, they were grouped as a single species *Balaenoptera edeni*. More recently, they are classified into two subtypes; namely, *B. e. edeni* is the small inshore form while *B. e. brydei* is the offshore form (on Mack 1998;
Constantine et al. 2018; Ogundile and Versfeld 2020; Ogundile et al. 2021b). Bryde’s whale produce characteristics call that can be used to identify them in combination with their physical appearance. Figure 2 depicts a typical example of an inshore Bryde’s whale short pulse call (sampled at 96 kHz) recorded in False bay, South-West, South Africa (Ogundile and Versfeld 2020; Ogundile et al. 2021b).

2.2. Humpback whales

The Humpback whale is one of the biggest species of Mysticete. The Humpback whale (*Megaptera novaeangliae*) has a distinctive physical appearance (such as its knobby head and pectoral fin) in comparison to other Mysticete species; thus, they can be easily distinguished (Ogundile et al. 2021b). Humpback whales produce vocalisations that have the characteristics of a ‘song’ because they are organised in complex periodic sequences. The song comprises both pulsed and tonal sound that changes from time to time (Ogundile et al. 2021b). Figure 3 shows a typical example of a Humpback whale song (sampled at 4 kHz), recorded off the north coast of Kauai, Hawaii, USA (Ogundile et al. 2021b). The information about the population of *Megaptera novaeangliae* is of importance to marine biologist because they are found in all oceans. Therefore, their vocal repertoire is collected from different locations over days, months and years, which has encouraged the development of various ML detection techniques for Humpback whale songs.
3. Feature extraction techniques

The employed feature extraction technique determines the effectiveness of all ML detection methods for cetaceans. The extracted attributes should be sufficiently reliable to represent the signals of interest. In Mysticete vocalisation, the extracted feature from the vocalisation is used to create a feature vector in formats suited for the adopted ML approach. For instance, given a Mysticete vocalisation of the form defined by:

\[ X_i = (x_1 \ x_2 \ \ldots \ x_i), \]  

where \( x_i \) is a sampling point in \( X_i \) and \( i \) is the number of sampling point. Different feature extraction techniques can be applied to the sound signal, \( X_i \) to obtain a feature vector of the form defined by:

\[ Z_j = (z_1 \ z_2 \ \ldots \ z_j), \]  

where \( z_j \) is an element in \( Z \) and \( j \) is the dimension of the feature vector (note that \( i > j \)). Importantly, the feature extraction techniques can be applied to \( X_i \) in smaller snippets called frames, depending on the value of \( i \). For example, suppose \( i=500 \) and a frame size of \( f=100 \) is chosen, the feature extraction technique is applied to \( X_i \) for each frame to produce the following matrix:

\begin{figure}
\centering
\includegraphics[width=\textwidth]{humpback_whale_song.png}
\caption{Time series and spectrogram representation of a Humpback whale song.}
\end{figure}
where \( k \) is the number of frames in \( X_i \) (\( k=5 \)). In the literature, various feature extraction algorithms have been used to automatically detect Mysticete vocalisations. In most cases, the more reliable the feature vector, the higher the ML technique’s performance. The following are descriptions of some of the comparison approaches used in this article.

### 3.1. Mel-scale frequency cepstral coefficient

The MFCC is a feature extraction technique that is frequently utilised in signal processing applications such as speech recognition, drone sound detection, image identification, gesture recognition, and recognition of cetacean, among others (Majeed et al. 2015). The MFCC technique utilises two types of filter; linear spaced filter and logarithm spaced
filter to extract reliable feature from a time varying signal, which can be used to represent the entire signal. The filters transform the signal from the time domain to the Mel-scale frequency (frequency domain). The process of MFCC feature extraction is in seven sequential steps; (i) Pre-emphasis (ii) Framing, (iii) Windowing, (iv) Fast Fourier transform, (v) Mel-scale filter bank, (vi) Logarithm operation, and (vii) Discrete cosine transform (Majeed et al. 2015; Ogundile et al. 2021b).

Consider a time varying signal \( X_i \), MFCC estimates the cepstral coefficient \( \gamma \) from \( X_i \) as (Zheng et al. 2001; Ogundile et al. 2021b):

\[
y_u = \sum_{\nu=1}^{\alpha} \phi_{\nu} \cos[u(\nu - 0.5) \frac{\pi}{\alpha}], \quad u = 1, 2, \ldots, \alpha, \tag{4}\]

where \( \phi_{\nu} \) is the logarithmic energy of the \( \nu^{th} \) Mel-spectrum band. The total number of cepstral coefficients is represented as \( \alpha \), which typically ranges from 10–14 depending on the application requirements. The value of \( \alpha \) determines the dimension, \( j \) of the feature vector (that is, \( \alpha = j \)). The dimension, \( j \) is essential for determining the computational time complexity of the detection mechanism in several ML tools, such as the HMMs. This implies that the higher the value of \( j \), the more complex the detection system. Similar to Equation 3, the MFCC estimate the feature vector as:

\[
Z_{\text{MFCC}} = \begin{pmatrix}
y_{1,1} & y_{1,2} & \cdots & y_{1,j} \\
y_{2,1} & y_{2,2} & \cdots & y_{2,j} \\
\vdots & \vdots & \ddots & \vdots \\
y_{k,1} & y_{k,2} & \cdots & y_{k,j}
\end{pmatrix}\tag{5}
\]

### 3.2. Linear predictive coefficient

LPC is an inverse filtering feature extraction technique used in signal processing applications such as speaker recognition and verification, speech coding, cetacean vocalisation and so on (Pace et al. 2009; Min and Tewfik 2010). The LPC technique employs a linear combination of the preceding regenerated signal, \( \alpha^{th} \) to predict the value of the current signal, \( P(\alpha) \) given by (Makhoul 1975; Min and Tewfik 2010):

\[
\hat{P}(\alpha) = \varphi_1 P(\alpha - 1) + \varphi_2 P(\alpha - 2) + \cdots + \varphi_m P(\alpha - \rho) = \sum_{m=1}^{\nu} \varphi_m P(\alpha - \rho), \tag{6}\]

where \( \hat{P}(\alpha) \) is the estimated value of \( P(\alpha) \), \( \varphi_m \) is the filter coefficient, \( P(\alpha - \rho) \) is the previous \( \alpha^{th} \) signal and \( \nu \) is the number of filter coefficients. The LPC coefficient is therefore computed by minimising the sum of the squared differences between the original signal and the linearly estimated signal as (Makhoul 1975; Min and Tewfik 2010):

\[
\psi(\alpha) = P(\alpha) - \sum_{m=1}^{\nu} \varphi_m P(\alpha - \rho). \tag{7}\]
In Equation 7, the error between $P(\alpha)$ and $\hat{P}(\alpha)$ is represented as $\psi(\alpha)$. Thus, the filter coefficient, $\varphi_m$ can be estimated from Equation 7 using different mathematical methods such as the autocorrelation method. Also, the total number of filter coefficient, $\nu$ typically range from 10–14 depending on the application requirements. Similar to Equation 3, the LPC estimate the feature vector as:

$$
Z_{LPC} = \begin{pmatrix}
\psi_{1,1} & \psi_{1,2} & \cdots & \psi_{1,j} \\
\psi_{2,1} & \psi_{2,2} & \cdots & \psi_{2,j} \\
\vdots & \vdots & \ddots & \vdots \\
\psi_{k,1} & \psi_{k,2} & \cdots & \psi_{k,j}
\end{pmatrix}.
$$

(8)

### 3.3. Empirical mode decomposition

EMD is a signal processing approach for feature extraction in a variety of applications such as radar analysis and signal emitter recognition (Liu et al. 2019; Ogundile et al. 2020b, 2021b). The EMD was recently applied as a feature extraction technique for Mysticete vocalisations in Ogundile et al. (2020b). This feature extraction technique is a completely data driven tool for decomposing a signal, $X_i$ into zero-mean functions called intrinsic mode functions (IMFs). The IMFs are calculated in an iterative process known as sifting, and used to represent the feature of the decomposed signal. During sifting, the IMFs are arranged in descending order by frequency components. The output IMFs from the final iteration can be reconstructed to recover the initial decomposed signal, $X_i$ as defined by (Ogundile et al. 2021b):

$$
\hat{X}_i = \sum_{h=1}^{j} I_h + R_j,
$$

(9)

where $\hat{X}_i$ is an approximation of the initial decomposed signal, $X_i$, $I_h$ is the value of the IMF at the $h$ iteration and $R_j$ is the residual signal at the final iteration, $j$. The estimated IMFs are always in the form of a matrix such that the number of column in the matrix, $T$ represents the dimension, $j$ of the feature vector. Similar to Equation 3, the EMD derived the feature vector from the calculated IMFs in matrix form as:

$$
Z_{EMD} = \begin{pmatrix}
\Gamma_{1,1} & \Gamma_{1,2} & \cdots & \Gamma_{1,j} \\
\Gamma_{2,1} & \Gamma_{2,2} & \cdots & \Gamma_{2,j} \\
\vdots & \vdots & \ddots & \vdots \\
\Gamma_{k,1} & \Gamma_{k,2} & \cdots & \Gamma_{k,j}
\end{pmatrix}.
$$

(10)

Note that the number of IMFs is a function of the length of the whale calls. That is, the bigger the length of the signal, the more is the value of $j$. 
3.4. **Dynamic mode decomposition**

Similar to the EMD, dynamic mode decomposition is a data driven matrix decomposition tool based on the knowledge of linear Koopman operator (Schmid and Sesterhenn 2010; Ogundile et al. 2021b). The DMD is a decomposition model used for studying the behaviour of non-linear and non-stationary systems or signals. The DMD has been used in different applications such as fluid mechanics (Schmid and Sesterhenn 2010), indoor localisation and positioning (Babalola and Balyan 2021), and detection of cetacean vocalisations (Ogundile et al. 2021b). The strength of the DMD algorithm is in its ability to decompose a signal into spatial temporal modes (STMs). It achieves this by combining the PCA and Fourier transform properties. The Fourier transform is used to study the signal in the frequency domain while the PCA is used for calculation in the spatial domain (Ogundile et al. 2021b). Given an observation matrix $Q$ with $y$ number of points obtained at specified time gaps, $Q$ can be restructured into two different matrices $Q_1$ and $Q_2$ as defined by (11)-(13):

$$Q = (q_1, q_2, ..., q_y), \quad (11)$$

$$Q_1 = (q_1, q_2, ..., q_{y-1}) \in Q, \quad (12)$$

$$Q_2 = (q_2, q_3, ..., q_y) \in Q. \quad (13)$$

Note that the column length $l$ of $Q_1$ and $Q_2$ is the same ($l=y-1$) and $Q_1$ and $Q_2$ are overlapping with increase time gaps. Thus, the STMs can be computed from the observation matrices $Q_1$ and $Q_2$ in four steps (Ogundile et al. 2021b):

1. Computes the reduced singular value decomposition of $Q_1$ as defined by Equation 14,

$$Q_1 = HCG\dagger, \quad (14)$$

where $H \in l \times r$, $C$ is a diagonal $\in r \times r$, $G \in l \times r$, the rank of $Q_1$ is given as $r$ and $\dagger$ is the transpose operator.

2. Computes the matrix $V$ as defined by Equation 15 and 16

$$Q_2 \approx Q_1V = HCG^*V, \quad (15)$$

$$V = H^*C^{-1}GQ_2. \quad (16)$$

3. Computes the eigenvalues and eigenvectors from $V$ as defined by Equation (17),

$$V\mathcal{F} = \mathcal{P}\mathcal{F}, \quad (17)$$

where $\mathcal{P}$ is a diagonal matrix of eigenvalues and the corresponding eigenvectors are the column of $\mathcal{F}$.

4. Thus, the DMD modes $M$ is calculated from $\mathcal{F}$ as defined by Equation (18).

$$M = H\mathcal{F} = Q_2G^{-1}\mathcal{F}. \quad (18)$$
The DMD modes obtained in Equation 18 are in matrix form with a rank $r$, where $r = j$, is the column length of the observation $Q_1$ and $Q_2$. These DMD modes, $M$ are reconstructed in (Ogundile et al. 2021b) to form feature vectors of the form of Equation 19, that can be combined with the HMM to detect Mysticete vocalisations.

$$Z_{DMD} = \begin{pmatrix} \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,j} \\ \beta_{2,1} & \beta_{2,2} & \cdots & \beta_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k,1} & \beta_{k,2} & \cdots & \beta_{k,j} \end{pmatrix}.$$ (19)

3.5. Principal component analysis

The PCA is a multivariate technique that can be used to reduce the dimension of a large dataset, by converting the variables in the large set into a smaller set without losing the information contained in the large dataset (Bartholomew 2010; Abbas and Alkarkhi 2019). Therefore, the smaller variable set can be quickly visualised and analysed using different ML tools while conserving as much information as possible. PCA makes use of simple equations constructed from the initial variables, which are referred to as components. The PCA is a very useful feature extraction technique, most especially when the variables to be analysed contains both positive and negative points (Abbas and Alkarkhi 2019). The mathematical operation of the PCA can be summarised into three sequential steps: (1) Standardisation, (2) Covariance matrix computation, and (3) Computation of the Eigenvector and Eigenvalues to obtain the principal components.

**Step 1: Standardisation.** The process of standardisation enables each sampling point or initial variables in a continuous signal, $X_i$ to equally contribute to the feature analysis. Basically, if there are huge differences between the variables in the dataset, the variables having large value will dominate the variables with small value, which can lead to a bias in the output. Therefore, scaling the dataset to comparable level can prevent such occurrence. For a given $X_i = (x_1, x_2, \ldots, x_i)$, the standardisation is implemented as:

$$\hat{X}_i = \begin{pmatrix} x_1 - \bar{M}(X_i) \\ \frac{x_2 - \bar{M}(X_i)}{\bar{D}(X_i)} \\ \frac{x_3 - \bar{M}(X_i)}{\bar{D}(X_i)} \end{pmatrix},$$ (20)

In Equation 20, $\hat{X}_i$ is the transform signal, $X_i$, where all the variables or sampling points in $\hat{X}_i$ are of the same scale, $\bar{M}$ is the mean of $X_i$ and $\bar{D}$ is the standard deviation of $X_i$.

**Step 2: Covariance Matrix Computation.** The next step in the PCA procedure is to compute the covariance matrix, $\mathbf{K}$. The $\mathbf{K}$ is computed to identify the correlation between the variables in $\hat{X}_i$, that is, the variation between each point and the mean of $\hat{X}_i$. This step is important since the variables are highly correlated such that they contain redundant information. The $\mathbf{K}$ is a $g \times g$ symmetric matrix, where $g$ is the dimension of the matrix. The entries of $\mathbf{K}$ are covariances associated with all likely pairs of the initial variables. Suppose a $d$-dimensional matrix ($d = 3$) of the form:

$$C = \begin{pmatrix} C_1 & C_2 & C_3 \end{pmatrix},$$ (21)
C_1, C_2 and C_3 are defined as:

\[
C_1 = \begin{pmatrix}
  c_{11} \\
  c_{12} \\
  \vdots \\
  c_{1b}
\end{pmatrix}, \quad C_2 = \begin{pmatrix}
  c_{21} \\
  c_{22} \\
  \vdots \\
  c_{2b}
\end{pmatrix}, \quad C_3 = \begin{pmatrix}
  c_{31} \\
  c_{32} \\
  \vdots \\
  c_{3b}
\end{pmatrix},
\]

the covariance of matrix C is given as:

\[
\hat{C} = \text{cov}(C) = \begin{pmatrix}
  \sum \frac{c_1^2}{b} & \sum \frac{c_1c_2}{b} & \sum \frac{c_1c_3}{b} \\
  \sum \frac{c_2c_1}{b} & \sum \frac{c_2^2}{b} & \sum \frac{c_2c_3}{b} \\
  \sum \frac{c_3c_1}{b} & \sum \frac{c_3c_2}{b} & \sum \frac{c_3^2}{b}
\end{pmatrix},
\]

(22)

The diagonal elements of Equation 22 are the variance of C_1, C_2 and C_3, and the covariance of two vectors (for example, cov(C_1, C_2)) is given by:

\[
cov(C_1, C_2) = \frac{\sum (C_1 - \mathcal{M}(C_1))(C_2 - \mathcal{M}(C_2))}{b - 1} = \sum \frac{C_1C_2}{b}.
\]

(23)

Note that cov(C_1, C_2)=cov(C_2, C_1); thus, the entries of the \( \mathbf{K} \) are symmetric with respect to the main diagonal.

**Step 3: Eigenvector and Eigenvalues Computation.** The eigenvectors, \( \mathcal{F} \) and eigenvalues, \( \mathcal{P} \) are computed from the \( \mathbf{K} \) to obtain the PC of the signal. The PC is determined by solving:

\[
\mathbf{K}\mathcal{F} = \mathcal{P}\mathcal{F},
\]

(24)

where \( \mathcal{F} \) and \( \mathcal{P} \) are the eigenvectors and eigenvalues of the covariance matrix. The \( \mathcal{P} \) are scalar values and \( \mathcal{F} \) are non-zero vectors representing the PC. This implies that each \( \mathcal{F} \) represent a principal component. The \( \mathcal{F} \) shows the direction of the PCA space and the corresponding \( \mathcal{P} \) represents the length, magnitude or scaling factor of the \( \mathcal{F} \) (Tharwat 2016). The \( \mathcal{F} \) with the highest eigenvalue has the \( b \) highest variance and it is the first principal component. That is, it carries the maximum information possible. Thus, the PC are arranged in descending order with respect to the information carried by the eigenvector. Arranging the PC in this manner reduces the feature dimension without necessarily losing much information (eliminating the components with little or no information). These PC can be constructed as a linear combination of the main signal and the number of PC depends on the \( d \)-dimension of the the signal to be analysed. For example, the 3-dimensional matrix defined by Equation 21 will yield 3 principal components after decomposition.

The singular vector decomposition (SVD) is one of the most popular linear algebra principle used to solve for the \( \mathcal{F} \) and \( \mathcal{P} \) in Equation 24. The SVD aims to decompose a matrix \( \hat{C} \) into three matrices such that:

\[
\hat{C} = \mathcal{LYZ}^\dagger,
\]

(25)
where $\mathcal{L}$ is a $b \times d$ matrix referred to as the left singular vector, $\mathcal{Y}$ is a $b \times b$ diagonal matrix, and $\mathcal{Z}$ is a $d \times d$ matrix called the right singular vector. The entries on the diagonal of $\mathcal{Y}$, which are arranged in descending order are the $\mathcal{P}$, while the columns of the right singular vector, $\mathcal{Z}$ are the $\mathcal{F}$ or the PC.

4. Proposed PCA feature vector

The derived PC from Equation 25 is restructured to form a feature vector suitable for use with the HMM to detect Mysticete vocalisations. It is however important to first determine the dimension $d$ of the PC, which serves as the feature vector. The value of $d$ must be kept as small as possible in order to reduce the computational time complexity the PCA technique will impose on the HMM. The larger the value of $d$, the more is the computational time complexity imposed on the HMM. Therefore, there is a trade-off in the sensitivity performance and computational time complexity of the PCA-HMM in selecting the value of $d$. As such, the value of $d$ is selected using simulation runs.

The sound waveforms are usually represented in the form of a row or column vector. Hence, it is essential to efficiently reconstruct this vocalisation vector with respect to the selected $d$ to form the matrix $\mathcal{C}$. Given the standardised vocalisation vector,

$$\mathcal{X}_i = \left( \frac{x_1 - \mathcal{M}(x_i)}{\mathcal{D}(x_i)} \frac{x_2 - \mathcal{M}(x_i)}{\mathcal{D}(x_i)} \ldots \frac{x_d - \mathcal{M}(x_i)}{\mathcal{D}(x_i)} \right) = (\hat{x}_1 \ \hat{x}_2 \ \ldots \ \hat{x}_i), \quad (26)$$

matrix $\mathcal{C}$ is formed from $\mathcal{X}_i$ as:

$$\mathcal{C} = \begin{pmatrix} c_{1,1} & c_{1,2} & \ldots & c_{1,d} \\ c_{2,1} & c_{2,2} & \ldots & c_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ c_{b,1} & c_{b,2} & \ldots & c_{b,d} \end{pmatrix} \quad (27)$$

The first element in row one of $\mathcal{C}$ ($c_{1,1}$) is the first element in $\mathcal{X}_i$; that is, $c_{1,1}=\hat{x}_1$. Similarly, the last element in row $b$ of $\mathcal{C}$ ($c_{b,d}$) is the last element in $\mathcal{X}_i$ ($c_{b,d}=\hat{x}_i$).

Thereafter, the PC can be obtained from $\mathcal{C}$ by sequentially computing Equation 22 ($\mathcal{X}$ computation) and Equation 25 ($\mathcal{P}$ and $\mathcal{F}$ computation) to obtain:

$$\mathcal{Z} = \begin{pmatrix} z_{1,1} & z_{1,2} & \ldots & z_{1,d} \\ z_{2,1} & z_{2,2} & \ldots & z_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ z_{d,1} & z_{d,2} & \ldots & z_{d,d} \end{pmatrix} \quad (28)$$

Thus, the feature vector is derived as:

$$\kappa_d = [(\frac{1}{d} \sum_{o=1}^{d} |z_o|)_1, (\frac{1}{d} \sum_{o=1}^{d} |z_o|)_2, \ldots, (\frac{1}{d} \sum_{o=1}^{d} |z_o|)_d] = [\nu_1, \nu_2, \ldots, \nu_d]. \quad (29)$$
Note that the dimension of the obtained feature vector using the PCA is \( d \), where \( d=j \). Accordingly, given that the sound signal is divided into \( k \) number of frames (where each frame has approximately \( f \) number of sampling points), the feature can be derived in matrix form as:

\[
Z_{\text{PCA}} = \begin{pmatrix}
\nu_{1,1} & \nu_{1,2} & \ldots & \nu_{1,j} \\
\nu_{2,1} & \nu_{2,2} & \ldots & \nu_{2,j} \\
\vdots & \vdots & \ddots & \vdots \\
\nu_{k,1} & \nu_{k,2} & \ldots & \nu_{k,j}
\end{pmatrix}
\]

(30)

This developed PCA feature extraction technique for Mysticete vocalisations is hereby summarised as follows:

**Algorithm 1: PCA Feature Vector.**

**Input:** \( X_i, f \)

**Output:** Feature vector \( Z_{\text{PCA}} \)

1: Standardise \( X_i \) as in Equation 20
2: Determine the dimension, \( d \) of the PC
3: Restructured \( X_i \) wrt \( d \) as in Equation 27
4: Compute \( \mathbf{R} \) as in Equation 22
5: Compute the eigenvector, \( \mathbf{F} \) and eigenvalues, \( \mathbf{P} \) as in Equation 25
6: Derive the PC, \( \kappa_d \) as in Equation 29 and Equation 30

### 4.1. PCA feature vector: numerical example

Consider the random signal shown in Figure 4 with points defined by Equation 31:

\[
X_i = (0.7094, 0.7547, 0.2760, -0.6797, 0.6551, -0.1626, 0.1190, 0.4984, \\
-0.9597, 0.3404, 0.5853, -0.2238, 0.7513, -0.2551, 0.5060, -0.6991, \\
0.8909, -0.9593, 0.5472, 0.1386, -0.1493, 0.2575, -0.8407, 0.2543, \\
0.8143, -0.2435, 0.9293, 0.3500, -0.1966, 0.2511, -0.6160, 0.4733, \\
0.3517, -0.8308, 0.5853, 0.5497, -0.9172, 0.2858, -0.7572, 0.7537),
\]

(31)

the PC can be extracted from \( X_i \) as follows. Firstly, \( X_i \) is standardised using Equation 20 to obtain Equation 32:

\[
\hat{X}_i = (1.0286, 1.1055, 0.2929, -1.3294, 0.9364, -0.4516, 0.0264, 0.6704, \\
-1.8047, 0.4022, 0.8180, -0.5555, 1.0997, -0.6086, 0.6833, -1.3623, \\
1.3367, -1.8040, 0.7533, 0.0597, -0.4290, 0.2615, -1.6027, 0.2561,
\]
1.2067, −0.5889, 1.4019, 0.4185, −0.5093, 0.2507, −1.2213, 0.6278, 0.4214, −1.5859, 0.8180, 0.7575, −1.7326, 0.3096, −1.4610, 1.1038. (32)

Next, the standardised signal \( \tilde{X}_i \) is carefully restructured to form a matrix \( C \) defined by Equation 33:

\[
C = \begin{pmatrix}
1.0286 & 1.1055 & 0.2929 & -1.3294 & 0.9364 \\
-0.4516 & 0.0264 & 0.6704 & -1.8047 & 0.4022 \\
0.8180 & -0.5555 & 1.0997 & -0.6086 & 0.6833 \\
-1.3623 & 1.3367 & -1.8040 & 0.7533 & 0.0597 \\
-0.4290 & 0.2615 & -1.6027 & 0.2561 & 1.2067 \\
-0.5889 & 1.4019 & 0.4185 & -0.5093 & 0.2507 \\
-1.2213 & 0.6278 & 0.4214 & -1.5859 & 0.8180 \\
0.7575 & -1.7326 & 0.3096 & -1.4610 & 1.1038
\end{pmatrix}.
\] (33)

Note that dimension \( d=5 \) is assumed in this restructuring; however, as earlier mentioned, the value of \( d \) is determined in this study using simulation runs. The \( R \) is subsequently computed from \( C \) using Equation 22 to obtain Equation 34:

\[
\hat{C} = \begin{pmatrix}
0.8744 & -0.5313 & 0.4974 & -0.3136 & 0.2029 \\
-0.5313 & 1.1406 & -0.3893 & 0.3799 & -0.2390 \\
0.4974 & -0.3893 & 1.1445 & -0.7892 & 0.0254 \\
-0.3136 & 0.3799 & -0.7892 & 0.8560 & -0.1110 \\
0.2029 & -0.2390 & 0.0254 & -0.1110 & 0.1697
\end{pmatrix}.
\] (34)

The PC can therefore be obtained by finding the SVD of \( \hat{C} \) such that

\[
\hat{C} = \mathcal{L}_Y \mathcal{Z}^\dagger
\]

where,

\[
\mathcal{L} = \begin{pmatrix}
-0.4383 & -0.3052 & -0.7969 & 0.1330 & -0.2491 \\
0.4858 & 0.6661 & -0.5025 & 0.2507 & 0.0704 \\
-0.5754 & 0.5125 & -0.0810 & -0.5129 & 0.3696 \\
0.4767 & -0.3932 & -0.3244 & -0.6270 & 0.3461 \\
-0.1164 & -0.2143 & -0.0254 & 0.5131 & 0.8226
\end{pmatrix},
\] (35)

\[
\mathcal{Y} = \begin{pmatrix}
2.5111 & 0 & 0 & 0 & 0 \\
0 & 0.9372 & 0 & 0 & 0 \\
0 & 0 & 0.4686 & 0 & 0 \\
0 & 0 & 0 & 0.2158 & 0 \\
0 & 0 & 0 & 0 & 0.0525
\end{pmatrix},
\] (36)

\[
\mathcal{Z} = \begin{pmatrix}
-0.4383 & -0.3052 & -0.7969 & 0.1330 & -0.2491 \\
0.4858 & 0.6661 & -0.5025 & 0.2507 & 0.0704 \\
-0.5754 & 0.5125 & -0.0810 & -0.5129 & 0.3696 \\
0.4767 & -0.3932 & -0.3244 & -0.6270 & 0.3461 \\
-0.1164 & -0.2143 & -0.0254 & 0.5131 & 0.8226
\end{pmatrix}.
\] (37)
As mentioned, the columns of $Z$ are the eigenvectors, $F$ which serves as the decompose PC. Thus, $Z$ is restructured using Equation (29) to form the PCA feature vector defined by Equation (38):

$$
\kappa_d = (0.4185 \ 0.4183 \ 0.3460 \ 0.4073 \ 0.3716).
$$

### 5. Hidden Markov models

The HMM is a ML tool popularly used in different signal processing applications such as speech recognition and analysis, pattern recognition, data compression, and so on, because of its flexibility and robustness in modelling non-stationary random processes (Usman et al. 2020). Likewise, it has been widely used for the detection of different bioacoustic sounds. HMM can be viewed as a ranking classifier that probabilistically allocates a label to each fragment in a sequence of observations. Subsequently, it computes a probability distribution over the sequence of observations and return the most probable sequence (Usman et al. 2020; Ogundile et al. 2021b). This makes HMM suitable for classifying and modelling the sequence of observations extracted from bioacoustic signals. Note that the ergodic HMM type is adopted in this study, where transition is possible between any state. Thus, the transition probabilities are non-zero (Usman et al. 2020). HMM can be described with the following components:

1. The number of transition states $s$ in the model.

2. The transition probability, $T$. Given two states $s_1$ and $s_2$, the transition probability is defined as $T_{s_1s_2}$, which is the probability of moving from state $s_1$ to state $s_2$. Note that the probability of transiting to state $s_2$ from state $s_1$ only depends on the current state $s_1$ and not on any other state. Thus, the $T$ is of the form of a $s \times s$ matrix as defined by:

$$
T = \begin{pmatrix}
T_{1,1} & T_{1,2} & \cdots & T_{1,s} \\
T_{2,1} & T_{2,2} & \cdots & T_{2,s} \\
\vdots & \vdots & \ddots & \vdots \\
T_{s,1} & T_{s,2} & \cdots & T_{s,s}
\end{pmatrix}.
$$

(3) A sequence of observation, $W=\{w_1, w_2, \ldots, w_M\}$ that equals the output of the modelled system, where $M$ is the number of observations.

4. The emission distribution probability, $\mathcal{E}$, which indicates the likelihood of emitting an observation from a given state $s$. The $\mathcal{E}$ is usually in the form of a $s \times M$ matrix as defined by:

$$
\mathcal{E} = \begin{pmatrix}
\mathcal{E}_{1,1} & \mathcal{E}_{1,2} & \cdots & \mathcal{E}_{1,M} \\
\mathcal{E}_{2,1} & \mathcal{E}_{2,2} & \cdots & \mathcal{E}_{2,M} \\
\vdots & \vdots & \ddots & \vdots \\
\mathcal{E}_{s,1} & \mathcal{E}_{s,2} & \cdots & \mathcal{E}_{s,M}
\end{pmatrix}.
$$

(40)

Note, the $\mathcal{E}$ is a function of three key parameters; namely, the mean $\mu$, covariance matrix, $\Sigma$ and mixture weight, $\sigma$. It can be represented as $\mathcal{E}=\{\mu, \Sigma, \sigma\}$.

5. The start probability, $\pi=\{\pi_1, \pi_2, \ldots, \pi_M\}$, which indicates the starting point of the distribution process. The $\pi$ at every point of the process always sums up to 1; that is,
Therefore, a given HMM can be represented with these three parameters as:

\[ \chi = \{\pi, T, \mathcal{E}\}. \]  

Applying the HMM in modelling a signal, the HMM strives to solve these three key problems in order to determine the best sequence of observations:

1. Calculation of the observation probability: Given an observation sequence, \( W = \{w_1, w_2, \ldots, w_M\} \) and a model \( \chi = \{\pi, T, \mathcal{E}\} \), the HMM attempts to estimate the likelihood that this known model will generate a particular sequence of observations. This process can be achieved using the forward-backward algorithm called the BM-alg (Baum et al. 1970). The BM-alg is a maximum-likelihood expectation iterative process that optimises the HMM parameters, \( \chi = \{\pi, T, \mathcal{E}\} \), using the emission distribution parameters, \( \mathcal{E} = \{\mu, \Sigma, \sigma\} \) until the criteria are met.

2. Learning/Training (determining the model parameters): This is the most important and difficult part of the HMM process. In this phase, the HMM finds the model that will best fit the given data. That is, the HMM parameters \( \chi = \{\pi, T, \mathcal{E}\} \) are re-estimated continuously to maximise the probability of a given observation sequence. This process can be achieved using the forward-backward algorithm called the BM-alg (Baum et al. 1970). The BM-alg is a maximum-likelihood expectation iterative process that optimises the HMM parameters, \( \chi = \{\pi, T, \mathcal{E}\} \), using the emission distribution parameters, \( \mathcal{E} = \{\mu, \Sigma, \sigma\} \) until the criteria are met.

3. Decoding (detecting the optimal sequence): The HMM attempts to find the hidden states or paths of the HMM from the observation \( W \). Basically, this step is achieved using the V-alg (Viterbi 1967). The V-alg is a forward error correction decoding algorithm that calculates or finds a new path based on previous path knowledge. It calculates all possible hidden paths in a given sequence and outputs the path with the best probability.

As mentioned in Ogundile et al. (2020a, 2020b, 2021b), HMM is sensitive to flat or random values of the emission distribution parameters, \( \mathcal{E} = \{\mu, \Sigma, \sigma\} \). Thus, the HMM is often combined sequentially with two ML tools (the K-means Clustering Algorithm (K-MC) technique (Forgy 1965; MacQueen 1967) and GMM) to initialise the emission distribution parameters. Similarly, the HMM will be combined with these two ML tools in this article.

6. Simulation set-up and parameters

6.1. **PCA-HMM detection system**

Figure 5 depicts the block diagram of the proposed PCA-HMM. Here, reliable features are extracted from the annotated test file using the PCA technique. As earlier mentioned, the HMM is sensitive to random values of the Gaussian parameters; as such, the extracted feature vector, \( Z \) are initially fed into the K-MC block to start the training process. The output from the K-MC block is fed into the GMM block, which further refines the
Gaussian parameters, $E = \{\mu, \Sigma, \sigma\}$ before it is fed to the HMM block. The HMM refines the Gaussian parameters based on the supplied feature vector, $Z$ to generate new Gaussian parameters (the generated $E$ is more reliable than that generated from the GMM block), the start probability, $\pi$ and the transition probability matrix, $T$. Hence, the output of the HMM is represented as:

$$x = \{\pi, T, E\}.$$  

To perform detection, features are also extracted from the unknown sound dataset as shown in Figure 5. These extracted features are subsequently refined using the Gaussian parameters, $E = \{\mu, \Sigma, \sigma\}$ to produce a modified or refined feature vector expressed as $Z_E$. Thus, the modified feature vector, $Z_E$, the start probability, $\pi$ and the transition probability matrix, $T$ are fed into the V-alg as depicted in Figure 5. Hence, the input to the V-alg is defined by:

$$V_{\text{alg}_{\text{in}}} = \{\pi, T, Z_E\}. \quad (43)$$

The V-alg calculates all possible hidden paths in the given sequence, $Z_E$ using the $\pi$ and $T$, thereby outputting the path with the best probability.

### 6.2. HMM sampling training

This study uses two different Mysticete vocalisation datasets (Humpback whale songs and Bryde’s whale pulses) obtained from Mobysound.org to theoretically demonstrate the developed PCA-HMM detection system. As documented on Mobysound.org, each of the Humpback whale songs were recorded for about 15 minutes off the north coast of Kauai, Hawaii, USA. Also, each of the Bryde’s whale pulses were recorded for approximately 15 minutes from Eastern Tropical Pacific. The results in this article are therefore presented with respect to the Humpback whale songs and Bryde’s whale pulses.

For ease of annotation, the datasets for each whale recording were visually and aurally analysed using Sonic Visualiser. Subsequently, the datasets were annotated into two categories: (1) Mysticete vocalisations, and (2) noise. After annotations, the datasets were divided into two parts. One part, which is about 25% of each whale recording was used to train the PCA-HMM detector system while the larger portion was used for testing. It is important to emphasise that the large portion of the dataset was also
annotated to validate the performance of the developed PCA-HMM detector. Since there are two categories of signal (Mysticete vocalisations, and Noise) in the dataset, a total of two HMMs are formed. That is, one model is used for the Mysticete vocalisations and the other model for the Noise.

The number of training samples, \( t \) used to train the model is varied to verify the performance of the PCA-HMM detector for changes in \( t \). In addition, the training sample is divided to approximately \( f \) number of sampling points in order to extract reliable features from each training sample. The number of frames in each training sample is represented as \( k \). Accordingly, the performance of the PCA-HMM detector is validated for different values of \( f \). Thus, the \( t \) and \( f \) are the two key variable parameters used to evaluate the performance of the PCA-HMM detector system.

An ergodic type HMM is used in this study. As such, a four states and two mixture ergodic type HMM is adopted. This implies that the two HMMs (\( HMM_1 \) – Mysticete vocalisations and \( HMM_2 \) – Noise) are trained or modelled separately with a 4 states and 2 mixture ergodic HMM. Therefore, \( HMM_1 \) and \( HMM_2 \) can be represented by Equation 44 and Equation 45 respectively.

\[
\chi_1 = \{ \pi_1, T_1, E_1 \}, \tag{44}
\]

\[
\chi_2 = \{ \pi_2, T_2, E_2 \}. \tag{45}
\]

Based on the number of states, \( s \) and mixture weight, \( \sigma \), the start probabilities, \( \pi_1 \) and \( \pi_2 \) are \( 1 \times s \) row vectors or \( s \times 1 \) column vectors, and \( T_1 \) and \( T_2 \) are \( s \times s \) matrices.

After training the Mysticete vocalisations and noise separately, the two HMMs are combined to form a eight states and four mixture ergodic type HMM, where the first 1–4 states represent the Mysticete vocalisations model while states 5–8 represent the noise model or vice versa. Thus, the combined HMM is represented as:

\[
\chi_{1,2} = \{ (\pi_1, \pi_2), (T_1, T_2), (E_1, E_2) \}. \tag{46}
\]

The combined Gaussian parameters, \( \chi_{1,2}=(E_1, E_2) \) is used to refine the feature vector \( Z \) extracted from the unknown dataset. Therefore, the refined feature vector, \( Z_{\chi}=(Z_{E_1}, Z_{E_2}) \), \( \pi=(\pi_1, \pi_2) \) and \( T=(T_1, T_2) \) are fed into the V-alg, which predicts the states and classify the signal as either noise or the Mysticete vocalisations. The input to V-alg is thereby defined by:

\[
V - \text{alg}_{in} = \{ (\pi_1, \pi_2), (T_1, T_2), (Z_{E_1}, Z_{E_2}) \}. \tag{47}
\]

Note that the V-alg transits from one whole state (1–4) to the other state (5–8) using equal switching probabilities. Lastly, the dimension \( j \) of the extracted feature vector determines the dimension of the HMM. Whereas, the larger the dimension \( j \), the more is the computational cost of the HMM process. Hence, it is paramount to keep \( j \) as small as possible without necessarily trading-off the detection performance of the automated detector.

### 6.3. Comparative parameters

The performances of existing automatic detectors are commonly measured in terms of their sensitivity and FDR.
(1) Sensitivity, $S$: the detection accuracy of the automatic detector, given as (Yao et al. 2009):

$$S = \frac{TP}{TP + FN},$$

(48)

where TP is true positives, representing the number of vocalisations that are accurately detected and FN is false negative, indicating the number of times the detector missed the vocalisation that is manually identified by the human expert. A high value of the TP indicates a high value of $S$; that is, $TP \propto S$.

(2) False discovery rate (FDR): It is the measure of the reliability of the detector, defined as (Yao et al. 2009):

$$FDR = \frac{FP}{TP + FP},$$

(49)

where FP is false positive, corresponding to the number of vocalisations incorrectly detected by the detector. A small value of the FDR indicates that the detector is reliable.

### 7. Results and discussion

#### 7.1. PCA-HMM performance for different dimension $d$

Tables 1–5 show the results of the proposed PCA-HMM detection algorithm with different dimension $d$ for Mysticete vocalisations. Recall that, $d$ determines the size of the PC. The larger the value of $d$, the more is the computational cost of the HMM training process. From the tables, the PCA-HMM detector performance is shown for $d=7, 8, 9, 10,$ and $11$. Also, the value of the training samples, $t$ increases in each table. In the case of

<table>
<thead>
<tr>
<th>Table 1. PCA-HMM comparison for different dimension $d$: $f=500$.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bryde’s whale pulses</strong></td>
</tr>
<tr>
<td>$d$</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td><strong>Humpback whale songs</strong></td>
</tr>
<tr>
<td>$d$</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
</tbody>
</table>
the Bryde’s whale pulses, the training samples are varied as \( t = 6, 12 \) and 18. However, the training samples are varied as \( t = 18, 24 \) and 30 for the Humpback whale songs since the signals are longer in duration in comparison to the Bryde’s whale pulses. Therefore, the Humpback whale songs require more training samples to efficiently train the HMM. In addition, the duration in time for each frame size \( f \) or number of sampling points in each frame is varied from \( f = 100, 200, 300, 400 \) and 500.

As evident from Tables 1–5, the performance of the PCA-HMM detector improves as \( d \) increases from 7 to 8 for both Mysticete species. For example, for \( f = 100 \) and \( t = 18 \) (Table 5: Bryde’s whale pulses), the PCA-HMM detector exhibited a sensitivity performance gain of 0.6, and 0.2 reduction in the FDR performance as \( d \) increases.
from 7 to 8. Similarly, for $f = 100$ and $t = 30$ (Table 5: Humpback whale songs), the PCA-HMM detector exhibited a sensitivity performance gain of 4.24, and 0.89 reduction in the FDR performance as $d$ increases from 7 to 8. However, it is observed that as $d$ increases further to 9, 10 and 11, the PCA-HMM detector offers no significant sensitivity performance gain or reduction in the FDR performance; rather, it increases the computational burden imposed on the HMM. Therefore, $d = 8$ is selected as the best fit dimension or optimal $d$ for the PCA-HMM detector since it offers a good trade-off between the computational complexity and performance of the detector.
Table 6. Comparison of different feature extraction methods: \( f=500 \).

<table>
<thead>
<tr>
<th>Methods</th>
<th>( t=6 )</th>
<th>( t=12 )</th>
<th>( t=18 )</th>
<th>( t=6 )</th>
<th>( t=12 )</th>
<th>( t=18 )</th>
<th>( t=6 )</th>
<th>( t=12 )</th>
<th>( t=18 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPC-HMM</td>
<td>89.97</td>
<td>89.99</td>
<td>89.23</td>
<td>0.13</td>
<td>0.14</td>
<td>0.15</td>
<td>3.109</td>
<td>2.160</td>
<td>1.782</td>
</tr>
<tr>
<td>DMD-HMM</td>
<td>89.97</td>
<td>89.98</td>
<td>89.23</td>
<td>0.13</td>
<td>0.14</td>
<td>0.15</td>
<td>4.690</td>
<td>3.884</td>
<td>2.804</td>
</tr>
<tr>
<td>EMD-HMM</td>
<td>84.44</td>
<td>86.88</td>
<td>86.12</td>
<td>0.16</td>
<td>0.19</td>
<td>0.21</td>
<td>3.002</td>
<td>2.095</td>
<td>1.704</td>
</tr>
<tr>
<td>MFCC-HMM</td>
<td>85.43</td>
<td>84.54</td>
<td>83.99</td>
<td>0.20</td>
<td>0.24</td>
<td>0.35</td>
<td>5.305</td>
<td>4.893</td>
<td>3.682</td>
</tr>
<tr>
<td>LPC-HMM</td>
<td>82.24</td>
<td>81.46</td>
<td>80.89</td>
<td>1.34</td>
<td>1.48</td>
<td>1.67</td>
<td>5.304</td>
<td>4.893</td>
<td>3.682</td>
</tr>
</tbody>
</table>

Furthermore, it can be noted that as \( f \) increases from 100 to 500, the performance of the PCA-HMM detector decreases. For example, for \( d=8 \) and \( t=18 \) (Bryde’s whale pulses), the sensitivity performance of the PCA-HMM detector drops from 97.71\% to 90.97\% as \( f \) increases from 100 to 500 while the FDR performance increases from 0.07\% to 0.13\%. Similarly, for \( d=8 \) and \( t=30 \) (Humpback whale songs), the sensitivity performance of the PCA-HMM detector drops from 88.90\% to 80.87\% as \( f \) increases from 100 to 500 while the FDR performance increases from 1.56\% to 2.61\%. This reduction in the performance of the PCA-HMM detector as \( f \) increases from 100 to 500 is expected since

Table 7. Comparison of different feature extraction methods: \( f=400 \).

<table>
<thead>
<tr>
<th>Methods</th>
<th>( t=6 )</th>
<th>( t=12 )</th>
<th>( t=18 )</th>
<th>( t=6 )</th>
<th>( t=12 )</th>
<th>( t=18 )</th>
<th>( t=6 )</th>
<th>( t=12 )</th>
<th>( t=18 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPC-HMM</td>
<td>80.87</td>
<td>78.78</td>
<td>76.10</td>
<td>2.61</td>
<td>4.00</td>
<td>4.99</td>
<td>5.282</td>
<td>4.219</td>
<td>3.100</td>
</tr>
<tr>
<td>DMD-HMM</td>
<td>80.87</td>
<td>78.78</td>
<td>76.09</td>
<td>2.61</td>
<td>4.00</td>
<td>4.99</td>
<td>6.073</td>
<td>5.128</td>
<td>3.998</td>
</tr>
<tr>
<td>EMD-HMM</td>
<td>77.93</td>
<td>75.10</td>
<td>73.28</td>
<td>3.62</td>
<td>4.81</td>
<td>5.96</td>
<td>5.711</td>
<td>4.351</td>
<td>3.426</td>
</tr>
<tr>
<td>MFCC-HMM</td>
<td>75.03</td>
<td>73.06</td>
<td>71.15</td>
<td>4.26</td>
<td>5.32</td>
<td>6.20</td>
<td>8.000</td>
<td>7.197</td>
<td>5.310</td>
</tr>
<tr>
<td>LPC-HMM</td>
<td>72.11</td>
<td>70.03</td>
<td>68.66</td>
<td>5.26</td>
<td>6.15</td>
<td>7.01</td>
<td>7.999</td>
<td>7.196</td>
<td>5.309</td>
</tr>
</tbody>
</table>

...
Table 8. Comparison of different feature extraction methods: $f=300$.

**Bryde’s whale pulses**

$d=8, r=9, I=6, a=12, u=12$

<table>
<thead>
<tr>
<th>Methods</th>
<th>Sensitivity (%)</th>
<th>FDR (%)</th>
<th>Computational Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t=18$</td>
<td>$t=12$</td>
<td>$t=6$</td>
</tr>
<tr>
<td>PCA-HMM</td>
<td>93.99</td>
<td>93.01</td>
<td>92.44</td>
</tr>
<tr>
<td>DMD-HMM</td>
<td>93.99</td>
<td>93.01</td>
<td>92.44</td>
</tr>
<tr>
<td>EMD-HMM</td>
<td>90.67</td>
<td>89.99</td>
<td>88.78</td>
</tr>
<tr>
<td>MFCC-HMM</td>
<td>88.42</td>
<td>87.67</td>
<td>86.51</td>
</tr>
<tr>
<td>LPC-HMM</td>
<td>85.66</td>
<td>84.22</td>
<td>83.21</td>
</tr>
</tbody>
</table>

**Humpback whale songs**

$d=8, r=9, I=9, a=12, u=12$

<table>
<thead>
<tr>
<th>Methods</th>
<th>Sensitivity (%)</th>
<th>FDR (%)</th>
<th>Computational Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t=30$</td>
<td>$t=24$</td>
<td>$t=18$</td>
</tr>
<tr>
<td>PCA-HMM</td>
<td>84.07</td>
<td>81.17</td>
<td>79.70</td>
</tr>
<tr>
<td>DMD-HMM</td>
<td>84.07</td>
<td>81.17</td>
<td>79.69</td>
</tr>
<tr>
<td>EMD-HMM</td>
<td>81.71</td>
<td>77.88</td>
<td>75.68</td>
</tr>
<tr>
<td>MFCC-HMM</td>
<td>78.81</td>
<td>75.23</td>
<td>72.12</td>
</tr>
<tr>
<td>LPC-HMM</td>
<td>75.21</td>
<td>72.52</td>
<td>70.04</td>
</tr>
</tbody>
</table>

Table 9. Comparison of different feature extraction methods: $f=200$.

**Bryde’s whale pulses**

$d=8, r=9, I=6, a=12, u=12$

<table>
<thead>
<tr>
<th>Methods</th>
<th>Sensitivity (%)</th>
<th>FDR (%)</th>
<th>Computational Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t=18$</td>
<td>$t=12$</td>
<td>$t=6$</td>
</tr>
<tr>
<td>PCA-HMM</td>
<td>96.06</td>
<td>94.91</td>
<td>93.99</td>
</tr>
<tr>
<td>DMD-HMM</td>
<td>96.06</td>
<td>94.90</td>
<td>93.99</td>
</tr>
<tr>
<td>EMD-HMM</td>
<td>93.13</td>
<td>92.69</td>
<td>91.56</td>
</tr>
<tr>
<td>MFCC-HMM</td>
<td>91.66</td>
<td>90.13</td>
<td>88.78</td>
</tr>
<tr>
<td>LPC-HMM</td>
<td>88.75</td>
<td>87.45</td>
<td>86.12</td>
</tr>
</tbody>
</table>

**Humpback whale songs**

$d=8, r=9, I=9, a=12, u=12$

<table>
<thead>
<tr>
<th>Methods</th>
<th>Sensitivity (%)</th>
<th>FDR (%)</th>
<th>Computational Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t=30$</td>
<td>$t=24$</td>
<td>$t=18$</td>
</tr>
<tr>
<td>PCA-HMM</td>
<td>87.22</td>
<td>83.88</td>
<td>80.77</td>
</tr>
<tr>
<td>DMD-HMM</td>
<td>87.22</td>
<td>83.87</td>
<td>80.77</td>
</tr>
<tr>
<td>EMD-HMM</td>
<td>85.01</td>
<td>80.32</td>
<td>76.77</td>
</tr>
<tr>
<td>MFCC-HMM</td>
<td>82.33</td>
<td>78.18</td>
<td>74.45</td>
</tr>
<tr>
<td>LPC-HMM</td>
<td>79.87</td>
<td>76.20</td>
<td>72.12</td>
</tr>
</tbody>
</table>

it easier to extract reliable features from smaller frames. However, the frame size should not be too small to avoid increasing the detector’s computational complexity; this is especially true for signals with extended durations.

Moreover, observe closely that the performance of the PCA-HMM detector improves as the value of $t$ increases. For example, for $d=8$ and $f=100$ (Table 5: Bryde’s whale pulses), the sensitivity performance of the PCA-HMM detector increases by 2.59% as $t$ increases from 6 to 18 while the FDR performance reduces by 0.03%. On the other hand, for $d=8$ and $f=100$ (Table 5: Humpback whale songs), the sensitivity performance of the
Table 10. Comparison of different feature extraction methods: f=100.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Sensitivity (%)</th>
<th>FDR (%)</th>
<th>Computational Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t = 18</td>
<td>t = 12</td>
<td>t = 6</td>
</tr>
<tr>
<td>PCA-HMM</td>
<td>97.71</td>
<td>96.88</td>
<td>95.12</td>
</tr>
<tr>
<td>DMD-HMM</td>
<td>97.71</td>
<td>96.87</td>
<td>95.12</td>
</tr>
<tr>
<td>EMD-HMM</td>
<td>94.44</td>
<td>93.53</td>
<td>92.67</td>
</tr>
<tr>
<td>MFCC-HMM</td>
<td>92.68</td>
<td>91.24</td>
<td>89.89</td>
</tr>
<tr>
<td>LPC-HMM</td>
<td>90.29</td>
<td>89.09</td>
<td>87.78</td>
</tr>
</tbody>
</table>

PCA-HMM detector increases by 7.59% as t increases from 18 to 30 while the FDR performance reduces by 0.86%. Thus, to further improve the performance of the PCA-HMM detector, the value of t can be increased for each species. But, it is important to emphasise that the performance of the detector saturates at a particular value of t. Additionally, it can be noted that the PCA-HMM detector offers better performance when used to detect the Bryde’s whale pulses in comparison to the Humpback whale songs. This is because the Humpback whale songs waveform are longer in duration, therefore, yielding feature vectors with unreasonably large value of k. Whereas, an unreasonable large value of k quickly saturates the HMM training process. Thus, more training samples are used for the Humpback whale songs to improve the output of the HMM training process.

7.2. PCA-HMM performance comparison with other methods

Tables 6–10 show the performance of the proposed PCA-HMM detector with existing methods in the literature. In the tables, the PCA-HMM detector for d = 8 (8-dimensional model) is compared with the DMD-HMM (Ogundile et al. 2021b), EMD-HMM (Ogundile et al. 2020b), MFCC-HMM (Ogundile et al. 2020b, 2021b) and LPC-HMM (Ogundile et al. 2020b, 2021b). A rank r=9 is assumed for the DMD-HMM detector as adopted in Ogundile et al. (2021b); thus, it is a 9-dimensional model. For EMD-HMM, the dimension of the model differs for the two Mysticete species. The number of intrinsic mode functions I, which determines the dimension of the model varies depending on the duration of the waveform. Thus, I=6 and I=9 are used for the Bryde’s whale pulses and Humpback whale songs respectively. This implies that the Bryde’s whale pulses employ a 6-dimensional model while the Humpback whale songs
(which are longer in duration) employ a 9-dimensional model as adopted in Ogundile et al. (2020b). The number of filter coefficients for the MFCC ($\alpha$) and LPC ($\upsilon$) is 12; thus, they are both 12-dimensional models.

First, the performance of the PCA-HMM detector is compared with the DMD-HMM detector. As shown in Tables 6–10, the PCA-HMM detector and DMD-HMM detector offer approximately the same sensitivity and FDR performances for changes in $t$ and $f$. However, the PCA-HMM detector is more suitable for use in real-time detection of Mysticete vocalisations for two reasons: (1) the PCA-HMM detector is a 8-dimensional model while the DMD-HMM detector is 9-dimensional. This implies that the DMD feature extraction method will impose more computational burden on the HMM as compared to the PCA method, which might not be encouraging for real-time detection, and (2) the DMD achieves signal decomposition by combining the PCA and Fourier transform properties. The Fourier transform is used to study the signal in the frequency domain while the PCA is used for calculation in the spatial domain (Ogundile et al. 2021b). Thus, aside performing the simple PCA step, the DMD also performs Fourier transformation which increases the processing time of the DMD algorithm. Hence, the PCA approach is more suitable for real-time detection of Mysticete vocalisations.

Second, the performance of the PCA-HMM detector is compared with the EMD-HMM detector. The PCA-HMM detector outperforms the EMD-HMM detector for different values of $t$ and $f$. For example, for $t=18$ and $f=100$ (Table 10: Bryde’s whale pulses), the PCA-HMM detector offers a sensitivity performance gain of 3.27% and a FDR reduction of 0.04% over the EMD-HMM detector. Likewise, for $t=30$ and $f=100$ (Table 10: Humpback whale songs), the PCA-HMM detector offers a sensitivity performance gain of 1.24% and a FDR reduction of 0.56% over the EMD-HMM detector. Nevertheless, for the Bryde’s whale pulses detection, the EMD-HMM detector exhibits less computational time complexity in comparison to the PCA-HMM detector because it is a 6-dimensional model. Thus, there is a trade-off between the sensitivity and FDR performances, and computational time complexity in employing either the PCA-HMM detector or the EMD-HMM detector for the detection of this Bryde’s whale pulses. Moreover, preference is given to the PCA-HMM detector since it offers a better sensitivity and FDR performance gain. The selection criteria is mostly application and/or requirement dependent. Tables 6–10 further indicate that the PCA-HMM detector exhibits less computational time complexity when employed to detect the Humpback whale songs in comparison to the EMD-HMM detector. In this case, observe that the EMD-HMM detector is a 9-dimensional model while the PCA-HMM detector is 8-dimensional. Thus, the PCA-HMM detector is more suitable as compared to the EMD-HMM detector when adopted for the detection of these Humpback whale songs.

Also, the performance of the PCA-HMM detector is compared with the MFCC-HMM and LPC-HMM detectors. The PCA-HMM detector offers superior performance for different values of $t$ and $f$ in comparison to these methods. Besides, the MFCC-HMM and LPC-HMM detectors imposes more computational burden as shown in Tables 6–10 on the HMM since they are 12-dimensional models while the PCA-HMM detector is a 8-dimensional model. This implies that the PCA-HMM detector is more
suitable for real-time detection in comparison to the MFCC-HMM and LPC-HMM detectors in terms of the sensitivity and FDR performances, and the computational time complexity.

8. Conclusion

This study introduced the method of PCA as a feature extraction technique, which was adopted with the HMM for the detection of Mysticete vocalisations. The PCA reduces the dimension of large datasets by converting the variable in the large set into principal components without losing the information contained in the large dataset. The principal components were mathematically reconstructed to form reliable feature vector, which was combined with the HMM to automatically detect Mysticete vocalisations. The performance of the developed PCA-HMM detector was compared for different dimensions. Also, the PCA-HMM detector was compared with the existing EMD-HMM, MFCC-HMM, LPC-HMM and DMD-HMM detectors using two different Mysticete vocalisations (Humpback whale songs and Bryde’s whale short pulses). For both species, results showed that the PCA-HMM detector offered good sensitivity value with high reliability and reasonable computational time complexity in comparison to these methods. Therefore, it is suitable for real-time detection of Mysticete vocalisations. The results obtained in this study are relevant to researchers and practitioners in this field as it will aid insightful decisions in selecting a detector for the detection of Mysticete vocalisations based on their application requirements.

Note


Acknowledgements

The authors acknowledge Dr. Leonid Smalov and Dr. Kamal Bentahar of the School of Computing, Electronics and Mathematics, Coventry University, United Kingdom for their insightful contributions and support throughout the research work.

Disclosure statement

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

Notes on contributors

OLAYINKA O. OGUNDILE 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 Senior Lecturer with the Department of Computer Science, Tai Solarin University of Education, Nigeria. He is also an extraordinary Senior Lecturer with the Department of Electrical and Information Engineering, Stellenbosch University, South Africa. He was a post-doctoral research fellow in the School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South
Africa, 2016, the Department of Electrical, Electronic and Computer Engineering, University of Pretoria, South Africa, 2016–2017 and the Department of Electrical and Information Engineering, Stellenbosch University, South Africa, 2019-2021. His research interests include digital communication, digital transmission techniques, digital signal processing/machine learning, channel estimation, forward error correction, and wireless sensor networks.

**OLUWASEYI P. BABALOLA** received the BSc. degree in mathematics from the University of Ibadan, Nigeria, in 2009, the MSc. degree in electrical engineering from University of the Witwatersrand, Johannesburg, South Africa, in 2017, and the PhD. degree in electronic engineering from Stellenbosch University, South Africa, in 2020. He is currently a Postdoctoral Research Fellow with the department of electrical, electronics, and computer science engineering, Cape Peninsula University of Technology, Bellville, South Africa. His research interests include advanced error-correcting coding, IoT, machine learning/artificial intelligence, visible light communication, digital signal processing techniques with applications to wireless communications, acoustic communications, and storage systems.

**SEUN G. ODEYEMI** received the B.Sc. degree in Espam Formation University, Cotonou, Benin Republic, in 2014 and his Master’s degree in Data sciences and computational intelligence, in Coventry University, U.K., in 2021. His research interest involves applying machine learning tools in sound detections.

**KAZEEM I. RUFAI** is a Nigerian academic with over two decades experience. He possesses two master degrees in Computer Science and Educational Technology from the Nigeria’s premier University, University of Ibadan. He earned his Ph.D. degree from the National University of Malaysia specializing in optimization of Intrusion Detection System using Membrane Computing paradigm. Rufai is an Associate Professor of Computer Science and a member of the Computer Professional Registration Council of Nigeria (CPRN) which is the highest ICT regulatory body in Nigeria.

**ORCID**

Olayinka O. Ogundile [http://orcid.org/0000-0002-2053-6086](http://orcid.org/0000-0002-2053-6086)

Oluwaseyi P. Babalola [http://orcid.org/0000-0001-9681-8437](http://orcid.org/0000-0001-9681-8437)

**References**


