

Dynamic mode decomposition: A feature extraction technique based hidden Markov model for detection of *Mysticetes*' vocalisations

O.O. Ogundile^{a,b,*}, A.M. Usman^a, O.P. Babalola^a, D.J.J. Versfeld^a

^a Department of Computer Science, Tai Solarin University of Education, Ogun State, Nigeria

^b Department of Electrical and Electronic Engineering, Stellenbosch University, Stellenbosch, South Africa

ARTICLE INFO

Keywords:

Detection
DMD
EMD
FDR
HMM
LPC
MFCC
Mysticetes
Pulse calls
Sensitivity
Songs

ABSTRACT

The detection and classification of *Mysticetes*' vocalisations have evoked the attention of researchers over the years because of their relevance to the marine ecosystem. These vocalisations are gathered employing different passive acoustic monitoring techniques. The vocalisation datasets are accumulated over a period; thus, they are large and cannot be easily analysed manually. Consequently, efficient machine learning (ML) tools such as the hidden Markov models (HMMs) are used extensively to automatically detect and classify these huge vocalisation datasets. As with most ML tools, the detection efficiency of the HMMs depend on the adopted feature extraction technique. Feature extraction techniques such as the Mel-scale frequency cepstral coefficient (MFCC), linear predictive coefficient (LPC), and empirical mode decomposition (EMD) have been adopted with the HMMs to detect different *Mysticetes*' vocalisations. This article introduces the method of dynamic mode decomposition (DMD) as a feature extraction technique that can be adopted with the HMMs for the detection of *Mysticetes*' vocalisations. The DMD has emerged as a robust tool for analysing the dynamics of non-stationary and non-linear signals. It is a completely data-driven tool that decomposes a signal over a certain period of time into relevant modes. Here, these spatial-temporal modes are reconstructed mathematically to form reliable feature vectors of the decomposed vocalisation signals. The performance of the proposed DMD-HMM detection technique is demonstrated using the acoustic datasets of two different *Mysticetes*' vocalisations: Humpback whale songs and Inshore Bryde's whale short pulse calls. In both species, the proposed DMD-HMM exhibits superior sensitivity and false discovery rate performances as compared to the MFCC-HMM, LPC-HMM, and EMD-HMM detection methods. Likewise, this proposed DMD-HMM detection method can be extended to other *Mysticetes*' that produce characteristics sounds.

1. Introduction

The automatic detection of fauna vocalisation has gained the attention of researchers, most especially in the marine ecosystem due to their economic importance and values. In particular, whales are important in maintaining the state of health of the ecosystem besides serving as guard species for the marine ecosystem Usman et al. (2020). Economically, the whale watching enterprise yearly generates over US\$2 billion, thereby enhancing the revenue of governments and engaging thousands of people Usman et al. (2020); Smith et al. (2013). Subsequently, many marine animals research groups are consistently deploying recorders in a process called passive acoustic monitoring (PAM), to gather the sound data of these mammals over weeks, months and years. In so doing, large datasets are created, which can be manually impossible to analyse by

ecologists, either by observing the spectrograms or paying attention to the sound recordings. For this reason, different automatic detection and classification techniques have been developed in past years to proffer prompt and accurate solutions. These techniques vary in performance depending on the species, experimental set-up and application requirements. Factors such as the physical environment where the recordings were taken, species, size of datasets available and features extracted influence the performance output Usman et al. (2020). Machine learning (ML) tools such as *k*-mean clustering MacQueen (1967); Forgy (1965), Gaussian mixture models (GMMs) Reynolds (2009); Duda et al. (2001), support vector machines (SVMs) Roch et al. (2008), hidden Markov models (HMMs) Jurafsky and James (2018); Rabiner (1989); Rabiner and Juang (1986) are widely used to develop efficient sound detection and classification techniques for marine mammals. These ML

* Corresponding author at: Department of Computer Science, Tai Solarin University of Education, Ogun State, Nigeria.

E-mail address: gundileoo@gmail.com (O.O. Ogundile).

<https://doi.org/10.1016/j.ecoinf.2021.101306>

Received 27 February 2021; Received in revised form 16 April 2021; Accepted 17 April 2021

Available online 28 April 2021

1574-9541/© 2021 Published by Elsevier B.V.

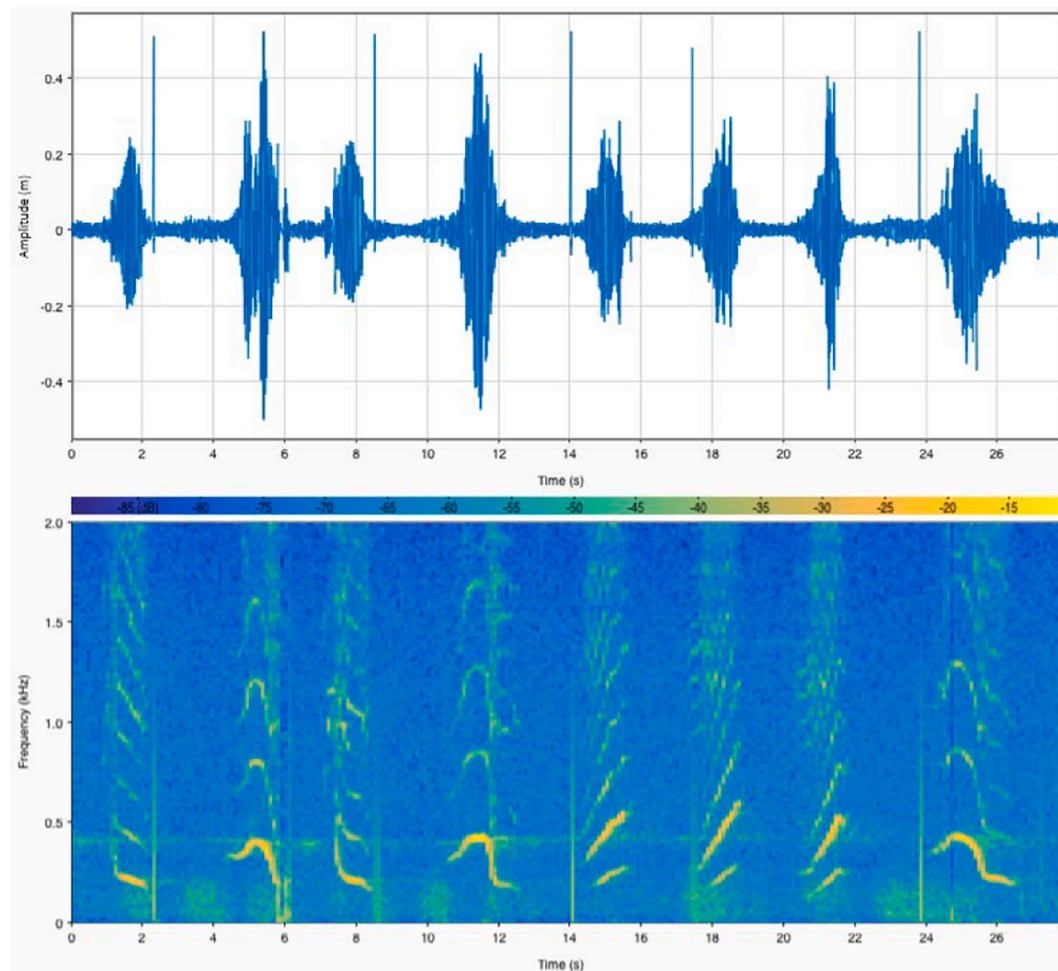


Fig. 1. Time series and spectrogram representation of Humpback whale song.

tools are somewhat reliable and achieve high sensitivity, ρ and low false discovery rate, λ .

Emphasising on HMMs, it is a flexible technique that is popularly used in marine sound detections including cryptic species. HMMs easily allow sound to be detected from a series of defined observation Yao et al. (2009). The performance of the HMMs depend on the reliability of these series of observations. In most cases, the more reliable the observations the better the sensitivity, ρ and false discovery rate, λ performances of the HMM. These series of observations are usually referred to as the feature vectors and can be computed mathematically using different feature extraction methods. In the past, the Mel-scale frequency cepstral coefficient (MFCC) Ogundile et al. (2020a, 2020b); Putland et al. (2018); Majeed et al. (2015) and linear predictive coefficient (LPC) Ogundile et al. (2020a, 2020b); Ogundile and Versfeld (2020); Makhoul (1975) methods are commonly used to derive feature vectors that are adopted with the HMM for the detection of marine mammals sound, especially *Mysticetes*' (suborder of Cetaceans) vocalisations. The empirical mode decomposition (EMD) Huang et al. (1988); Liu et al. (2019) was recently adopted with the HMM in Ogundile et al., 2020a, 2020b and was shown to be a performance efficient alternative to the MFCC-HMM and LPC-HMM methods for the detection of *Mysticetes*' vocalisations.

This article introduces the method of dynamic mode decomposition (DMD) Schmid and Sesterhenn (2010) as a feature extraction technique that can be adopted with the HMMs for the detection of *Mysticetes*' vocalisations. The DMD has recently emerged as a robust tool for analysing the dynamics of non-stationary and non-linear signals. It is a completely data-driven tool that decomposes a signal, $s(t)$ over a certain

period of time, t into relevant modes. These spatial-temporal modes are reconstructed mathematically in this article to form reliable feature vectors that can be adapted with the HMM. The DMD was originally developed for applications in fluid mechanics Schmid and Sesterhenn (2010); interestingly, it is now being used in different applications such as imaging Sikha et al. (2017), forecasting Tirunagari et al. (2017); Mohan et al. (2018), and so on, but it has not been adapted as a feature extraction technique for HMMs in detecting *Mysticetes*' vocalisations as proposed in this article. The performance of the proposed DMD-HMM detection technique is demonstrated using the acoustic datasets of two different *Mysticetes*' vocalisations; that is, Humpback whale songs and inshore Bryde's whale short pulse calls. The introduced DMD-HMM detection method is flexible and can be easily modified to detect similar characteristics vocalisations of other *Mysticetes*.

The contribution and relevance of this article is as follows. First, to the best of the authors knowledge, the method of DMD is used for the first time in marine mammal sound detection in this article. Thus, this proposed DMD-HMM detection approach for *Mysticetes*' vocalisations is original. Second, the DMD-HMM approach achieves a better feature characteristic of the vocalisation when compared to the conventional MFCC-HMM and LPC-HMM detection approaches, and the recent EMD-HMM detection approach. Hence, the DMD-HMM approach can serve as a performance efficient alternative to the already established approaches used for comparison in this article. In addition, the introduced DMD-HMM reduces the computational burden of the HMM in comparison to the MFCC-HMM and the LPC-HMM, but offer similarly computational burden as the EMD-HMM method as discussed in Section 7. Hence, it can be used in real-time detection of *Mysticetes*' vocalisations.

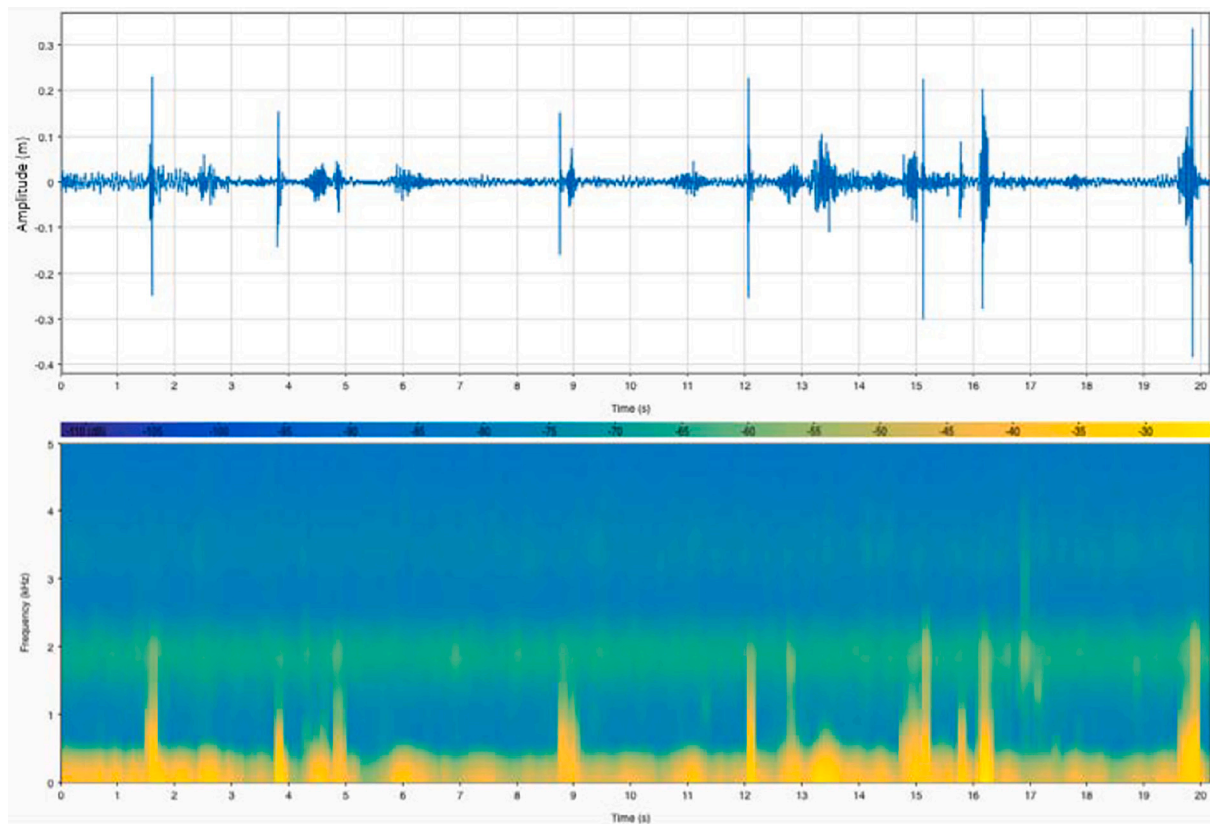


Fig. 2. Time series and spectrogram representation of inshore Bryde's whale short pulse call documented in Ogundile and Versfeld (2020).

In addition, the proposed tool in this article provides measurement data that can be applied by marine ecologist to obtain high level sound analysis. These include determining the range and seasonality of marine animal and the abundance of species. Lastly, the methodology introduced in this article can be of interest to researchers working on the detection of animal sounds in general.

The rest of the article is structured as follows. In section 2, the *Mysticetes*' vocalisations used for testing are briefly described with relevant figures. Section 3 briefly explains the HMM and how it is tailored with the *k*-mean clustering technique and the GMM. Backgrounds on the DMD are explained in Section 4. In Section 5, the existing feature extraction techniques (MFCC, LPC, and EMD) adopted with the HMMs are briefly explained. Of interest, this section introduces the method of DMD as a feature extraction technique that can be adopted with the HMMs. Section 6 explains the test procedure in details. In Section 7, the comparative parameters is explained while discussing the obtained results. The article is concluded in Section 8.

2. *Mysticetes*' vocalisations

Mysticetes, also known as Baleen whales are species of the order Cetacea. Instead of teeth, *Mysticetes* have series of horny plates termed 'Baleen' hanging from their upper jaw, which they use to filter food from water Sasaki et al. (2005); Kennedy (2020). There are currently about fourteen types of *Mysticetes* Kennedy (2020). However, in this article, Humpback whale songs and Bryde's whale pulse calls are used for results verifications. Nevertheless, as earlier mentioned, the introduced DMD-HMM technique can be modified to detect other *Mysticetes*' vocalisations.

2.1. Humpback whale

Humpback whale, also called *Megaptera novaeangliae* is one of the

largest members of the *Mysticetes*' species. Unlike other *Mysticetes*, the humpback whale has distinctive shapes, including the pectoral fin and knobby head, which makes them easy to distinguish using their physical appearance. Humpbacks produce vocalisations that are arranged in complex periodic sequences, having the characteristics of a 'song'. The song contains both tonal and pulsed sounds and can change from time to time. Humpback whales can be found in all oceans; thus, the information on their population structure is of importance to scientist. Subsequently, their vocal repertoire is gathered across different location over years. This has motivated the development of divers automated method to detect and classify Humpback songs. Most especially, the HMM approach has been used, adopting the MFCC and LPC feature extraction methods (see, Pace (2008); Tolkova et al. (2017)). As such, the use of Humpback songs is encouraged in this article to verify the performance of the introduced DMD-HMM detection technique. The song recordings of Humpback whale documented off the north coast of Kauai, Hawaii, USA¹ is used for testing in this article. The waveform of the Humpback whale songs documented on mobysound.org² as viewed from a Sonic Visualiser³ is shown in Fig. 1.

2.2. Inshore Bryde's whale

Bryde's whale was first sighted off South Africa during marine mammal surveys Olsen (1913). They are grouped as a single species known as *Balaenoptera edeni* (*B. edeni*). Presently, there are two sub-species of *B. edeni* in existence; namely, *Balaenoptera edeni edeni* (*B. e. edeni*) and *Balaenoptera edeni brydei* (*B. e. brydei*). The latter are the large

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

² The mobysound website contains different *Mysticetes*' vocalisations and other order of cetacean vocalisation.

³ <http://www.sonicvisualiser.com>

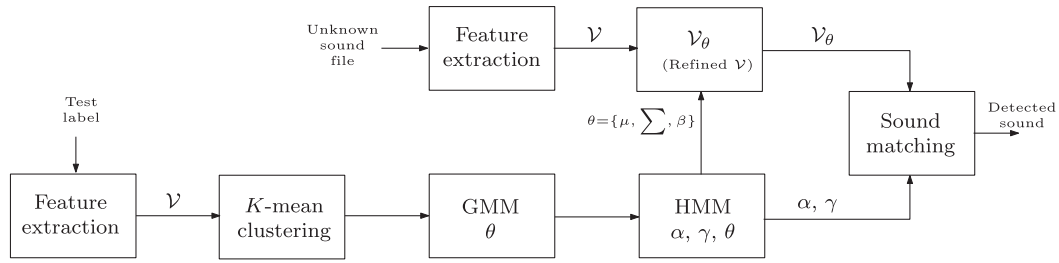


Fig. 3. *Mysticetes'* vocalisations detection system based on the HMM.

offshore form of *B. edeni* while the former are the small coastal or inshore form of *B. edeni* Ogundile and Versfeld (2020); Committee on Taxonomy (2017); Constantine et al. (2018). It is not easy to distinguish Bryde's whales from other *Mysticetes* using only their physical appearance. However, Bryde's whales produce characteristic calls which can be used for identification in support with their physical appearance Ogundile et al. (2020a, 2020b). Many research contributions have described the geographical location and population, genetic and phylogenetic features, nomenclature, and vocalisations of Bryde's whales but a few have described the methods that can be used to automatically detect their produced vocalisations. The use of Bryde's short pulse calls is therefore motivated in this article to verify the performance of the introduced DMD-HMM detection technique. The pulse call recordings of inshore Bryde's whales documented in Ogundile and Versfeld (2020) in an area of approximately 13km², close to Gordon's bay harbour, False bay, South-West, South Africa is used for testing in this article. The sound recording is a short pulse call with a small relative amplitude at the start, which increases to its peak and swiftly decays as the call ends. The waveform of the inshore Bryde's whale short pulse call viewed from a Sonic Visualiser as presented in Ogundile and Versfeld (2020) is depicted in Fig. 2.

3. Hidden Markov models

The HMMs approach has gained popularity for the analysis and recognition of bioacoustics vocalisation because of their flexible nature. HMM is a probabilistic ranking classifier that allots a tag to every fragment in a series of observation. Hence, it calculate the probability distribution over the series of observations and outputs the most likely observation sequence Usman et al. (2020). As such, it can seamlessly model and classify the series of observations obtained from the bioacoustics vocalisation. The operations of the HMM can be divided into two sequential stages: training stage and detection stage. In the training stage, the HMM computes three key parameters from the series of observation or feature vector, $\mathcal{V}$ obtained from the test samples: (1) start probability, α (2) transition probability matrix, γ , and (3) Gaussian parameters, $\theta = \{\mu, \Sigma, \beta\}$. The transition probability matrix, γ represents the feature vector extracted from the signal as a series of states across time. The matrix consists of probability values used to switch between states. The γ can be compiled by using the maximum-likelihood estimation, where methods such as the Baum-Welch algorithm (BM-alg) Baum et al. (1970) can be employed. The BM-alg is modelled with the Gaussian emission distribution parameter, θ ; that is, the mean μ , covariance matrix Σ , and mixture weight β , which assume a flat start or random values at the beginning of the BM-alg process. Interestingly, the HMM process is sensitive to these random start values of μ , Σ , and β , as it hampers the detection performance of the model. Consequently, in this article, the start values for the μ , Σ , and β are initialised using the k -mean clustering technique and GMM as depicted in Fig. 3.

During detection, the trained HMM parameters, α , γ , and θ are used to match the feature vector, $\mathcal{V}$ from the unknown sound file. The extracted feature, $\mathcal{V}$ is then modelled with the Gaussian parameters, θ to obtain a refined feature vector $\mathcal{V}_\theta$. Afterwards, the α and θ are used to

match this refined feature vector $\mathcal{V}_\theta$ employing the Viterbi algorithm (V-alg) Viterbi (1967). The V-alg stochastically produces closely matched hidden sequence from the $\mathcal{V}_\theta$, α and θ . This simple HMM approach employing the BM-alg and V-alg have been used for the detection of *Mysticetes'* vocalisation and other order of cetacean vocalisation. Regardless, the extracted feature vector is pivotal to the performance of the HMM. A reliable feature vector enhances the performance of the HMM. That is, the reliability of the feature vector is directly proportional to the detection performance of the HMM. Therefore, this study strives to improve the reliability of the feature vector by introducing the method DMD.

4. Dynamic mode decomposition: Background

Dynamic mode decomposition introduced in Schmid and Sesterhenn (2010) is a completely data-driven matrix decomposition technique that is developed using the concept of linear Koopman operator. It is a decomposition model that is primarily used for analysing the dynamics of non-stationary and non-linear signals or systems. DMD was developed for application in fluid mechanics but has been used in different applications over the years as earlier mentioned. An important feature of the DMD algorithm is the ability to extract spatial-temporal modes from the decomposed signal. DMD combines the properties of both the principal component analysis (PCA) and Fourier transform in decomposing a signal into spatial-temporal modes. It performs PCA for computations in the spatial domain while Fourier transform is used for analysing the signal in the frequency domain.

Given a signal or system with an observation matrix X having n number of points collected at uniformly spaced time gaps, X can be regrouped into two separate matrices X_1 and X_2 as defined in (1)–(3).

$$X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]. \quad (1)$$

$$X_1 = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n-1}] \in X, \quad (2)$$

$$X_2 = [\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n] \in X. \quad (3)$$

The column length of X_1 and X_2 is $m = n - 1$ and the two observation matrices X_1 and X_2 are overlapping with time. Assuming that the signal evolve slowly, the n^{th} observation can be represented as a linear combination of the preceding $n - 1$ observations with a residual error, R as defined in (4) Schmid and Sesterhenn (2010).

$$x_m = a_1 x_1 + a_2 x_2 + \dots + a_{m-1} x_{m-1} + R. \quad (4)$$

Accordingly, the goal of the DMD is to calculate the eigendecomposition of a mapping matrix, B using the Koopman operator as defined in (5) Schmid and Sesterhenn (2010),

$$BX_1 \approx X_2 \Rightarrow B = X_2 X_1^\dagger, \quad (5)$$

where $X_1^\dagger$ is the pseudoinverse of X_1 . Eq. (5) shows that the DMD of the pair (X_1, X_2) can be represented by the eigendecomposition of B . That is, the DMD spatial-temporal modes and eigenvalues are the eigenvector, Ψ of the mapping matrix, B with distinctive eigenvalues, Ψ . For most DMD applications, the matrix dimension of B is large; thus, the complexity of

the eigendecomposition becomes impracticable. Eventually, a matrix $\tilde{B}$ that is rank-reduced and has identical non-zero eigenvalues as B is introduced to ease the computational complexity issues. In this regard, (5) can be represented as defined in (6).

$$X_2 \approx X_1 \tilde{B}. \quad (6)$$

In summary, given the observations X_1 and X_2 , the DMD algorithm computes the spatial-temporal modes from the eigenvector as follows.

1. Calculates the reduced singular value decomposition (SVD) of X_1 as defined in (7)

$$X_1 = U \Sigma V^*, \quad (7)$$

where $U \in m \times r$, Σ is a diagonal $\in r \times r$, $V \in n \times r$, V^* is the transpose of V , and r is the rank of X_1 .

2. Calculates the matrix $\tilde{B}$ as defined in (8) and (9)

$$X_2 \approx X_1 \tilde{B} = U \Sigma V^* \tilde{B}, \quad (8)$$

$$\tilde{B} = U^* \Sigma^{-1} V X_2. \quad (9)$$

3. Calculates the eigenvectors and eigenvalues from $\tilde{B}$, so that

$$\tilde{B} \Upsilon = \Psi \Upsilon, \quad (10)$$

where Ψ is a diagonal matrix of eigenvalues and the column of Υ are the corresponding eigenvectors.

4. The DMD modes $\mathcal{M}$ can be calculated from Υ as defined in (11)

$$\mathcal{M} = U \Upsilon = X_2 V \Sigma^{-1} \Upsilon. \quad (11)$$

The DMD modes, $\mathcal{M}$ are in matrix form with rank r , which the equal to the number of column, m ($r = m$) of the observations X_1 and X_2 . These modes, $\mathcal{M}$ are modified in this article to form feature vectors that can be combined with the HMM for the detection of *Mysticetes*' vocalisations as discussed in Section 5.4.

5. Feature extraction methods for HMMs

The feature extraction techniques used for comparison in this article are explain briefly. Thereafter, the method of DMD as a feature extraction technique is introduced.

5.1. MFCC

The MFCC has been used extensively for feature extraction in different signal processing application. Examples of such application includes, image identification, speech recognition, cetacean vocalisation, gesture recognition, drone sound recognition, and so on [Majeed et al. \(2015\)](#). MFCC are used to extract distinctive feature from a time varying signal. Its uses two types of filters (linear spaced filters and logarithm spaced filters) to extract features from the signal, by transforming the signal from the time domain into the Mel-scale frequency [Majeed et al. \(2015\)](#). Summarily, feature extraction using the MFCC involves seven sequential steps; (1) Pre-emphasis (2) Framing (3) Windowing (4) Fast Fourier transform (5) Mel-scale filter bank (6) Logarithm operation, and (7) Discrete cosine transform [Majeed et al. \(2015\)](#); [Ogundile et al. \(2020a, 2020b\)](#). Thus, given a signal, the MFCC computes the cepstral coefficient as defined in (12) [Zheng et al. \(2001\)](#); [Ogundile et al. \(2020a, 2020b\)](#).

$$\nu_l = \sum_{j=1}^{\delta} \Gamma_j \cos \left[l(j-0.5) \frac{\pi}{\delta} \right], \quad l = 1, 2, \dots, \delta. \quad (12)$$

In (12), Γ_j is the logarithmic energy of the j^{th} Mel-spectrum band. The number of cepstral coefficient is defined as δ , which typically varies between 10 and 14 depending on the application. In this article, a value of $\delta = 12$ is selected for the detection of the Humpback whale songs and the Bryde's whale short pulse calls. That is, the MFCC-HMM detection detector is 12-dimensional for both *Mysticetes*' vocalisations. Accordingly, if a sample, S of the sound signal depicted in [Fig. 1](#) and/or [Fig. 2](#) is divided into k number of windows with roughly τ numbers of sampling points, the feature vector can be derived in form of a matrix as defined in (13).

$$\mathcal{V}_{MFCC} = \begin{bmatrix} \nu_{1,1} & \nu_{1,2} & \nu_{1,3} & \dots & \nu_{1,\delta} \\ \nu_{2,1} & \nu_{2,2} & \nu_{2,3} & \dots & \nu_{2,\delta} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \nu_{k,1} & \nu_{k,2} & \nu_{k,3} & \dots & \nu_{k,\delta} \end{bmatrix} \quad (13)$$

5.2. LPC

Similar to the MFCC, the LPC has been used extensively in different signal processing applications such as speech coding, synthesis and recognition, speaker recognition and verification, marine mammals vocal repertoire, and so on (see, ref. [Pace \(2008\)](#); [Min and Tewfik \(2010\)](#)). The LPC predicts an approximate value for the current sample signal $f(i)$ by combining the previous regenerated, u^{th} samples linearly as defined in (14) [Makhoul \(1975\)](#); [Min and Tewfik \(2010\)](#).

$$\bar{f}(i) = \epsilon_1 f(i-1) + \epsilon_2 f(i-2) + \dots + \epsilon_v f(i-u) = \sum_{v,u=1}^{\psi} \epsilon_v f(i-u) \quad (14)$$

In (14), $\bar{f}(i)$ is the approximated valued of $f(i)$ and $f(i-u)$ is the previous u^{th} sample. The filter coefficient is represented as ϵ_v and $\psi = u^{\text{th}} = v^{\text{th}}$ is the number of filter coefficients. These LPC coefficients ϵ_v is computed by minimising the sum of the squared differences between the linearly estimated samples and the primary samples as defined in (15) [Makhoul \(1975\)](#); [Min and Tewfik \(2010\)](#).

$$E(i) = f(i) - \sum_{v,u=1}^{\psi} \epsilon_v f(i-u) \quad (15)$$

As shown in (15), $E(i)$ is the error between $\bar{f}(i)$ and $f(i)$. Therefore, the LPC coefficients can be estimated from (15) using different computational technique such as the autocorrelation method. The value of $\psi = u^{\text{th}} = v^{\text{th}}$ varies between 10 and 14. So, for a fair comparison with the MFCC method, a fixed value of $\psi = u^{\text{th}} = v^{\text{th}} = 12$ is assumed in this article to detect the Humpback whale songs and Bryde's whale short pulse call. This also implies that the LPC-HMM detector is 12-dimensional. Similarly, if a sample, S of the sound signal depicted in [Fig. 1](#) and/or [Fig. 2](#) is divided into k number of windows with roughly τ numbers of sampling points, the feature vector can be derived in form of a matrix as defined in (16).

$$\mathcal{V}_{LPC} = \begin{bmatrix} \epsilon_{1,1} & \epsilon_{1,2} & \epsilon_{1,3} & \dots & \epsilon_{1,\psi} \\ \epsilon_{2,1} & \epsilon_{2,2} & \epsilon_{2,3} & \dots & \epsilon_{2,\psi} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \epsilon_{k,1} & \epsilon_{k,2} & \epsilon_{k,3} & \dots & \epsilon_{k,\psi} \end{bmatrix} \quad (16)$$

5.3. EMD

The EMD was recently put forward as a feature extraction technique that can be combined with the HMM for the detection of *Mysticetes*' vocalisations. Aside from marine sound detection, the EMD has been used for different signal processing applications such as radar signal emitter recognition, speech analysis, and so on [Ogundile et al. \(2020a, 2020b\)](#). The EMD introduced in [Huang et al. \(1988\)](#) is an entirely data-

driven tool used to decompose a signal, $s(t)$ over a random duration τ into intrinsic mode functions (IMFs). These IMFs are the zero-mean functions that represents the features of the signal that is decomposed. The IMFs are computed using an iterative process called sifting, where they are arranged by frequency content (Hz) in descending order. The resulting IMFs from the sifting process can be reconstructed to obtain the original decomposed signal, $s(t)$ as defined in (17).

$$\bar{s}(t) = \sum_{q=1}^Q I_q + R_Q. \quad (17)$$

In (17), $\bar{s}(t)$ is an approximation of the original decomposed signal $s(t)$, I_q is the IMF value at q number of iteration and R_Q is the residual signal at Q number of IMFs. The derived IMFs are usually in matrix form, where its column length represents the dimension of the feature vector. In Ogundile et al. (2020a, 2020b), this IMF matrix is reconstructed in six steps to form a feature vector that can be combined with the HMM. Given the sound signal depicted in Fig. 1 and/or Fig. 2, if a sample, S is divided into k number of windows with roughly τ numbers of sampling points, the feature vector can be derived from the computed IMF to form a matrix as defined in (18).

$$\mathcal{V}_{EMD} = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \dots & \sigma_{1,Q} \\ \sigma_{2,1} & \sigma_{2,2} & \dots & \sigma_{2,Q} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k,1} & \sigma_{k,2} & \dots & \sigma_{k,Q} \end{bmatrix} \quad (18)$$

The number of specified IMFs, Q varies depending on the length of the signal to be detected. In this article, $Q=6$ is specified for the Bryde's whale short pulse call, as there are no more performance improvement in the detection model with a further increase in the number of IMFs. Whereas, $Q=9$ is specified for the Humpback whale song because its waveform is longer than that of the Bryde's whale. More so, the performance of the detection model flattens at $Q=9$. As such, a further increase in the value Q only increase the computational burden on the HMM without improving the detection performance of the model.

5.4. Proposed DMD

The derived DMD modes, $\mathcal{M}$ in (11) is modified in this section to form feature vectors that can be combined with the HMM. First, a key parameter that needs to be addressed in forming this feature vector is the rank, r of the modes. The rank, r determines the dimension of the proposed DMD feature vector. It is important to keep the value of r as small as possible because the larger the value of r the more the computational burden in performing the HMM process. More so, there is a trade-off between the performance and the computational complexity of this proposed detection model when selecting the value of r . In this article, the value of r is determined by using simulation runs as shown in Section 7.1.

Second, the structure of the of *Mysticetes*' vocalisations can be represented in the form of a column or row vector. It is paramount to carefully restructure the sound vector with respect to the specified r (which determines the dimension of the feature vector) to form the matrix, X . Thus, given a window k of the sound vocalisation to be decomposed, with τ number of sampling points, as defined by the vocalisation column vector of (19),

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_\tau \end{bmatrix}, \quad (19)$$

where w_τ represents the τ^{th} sampling point. The, matrix, X can be formed by reshaping W with respect to r (that is, the column of X is $r+1$) as defined in (20).

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,r+1} \\ x_{2,1} & x_{2,2} & \dots & x_{2,r+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{c,1} & x_{c,2} & \dots & x_{c,r+1} \end{bmatrix} \quad (20)$$

The elements in the first row of X are extracted from the 1^{st} point to the $r+1$ points of W while the c^{th} row of X represents the last $r+1$ points of W . That is, the element $x_{1,1}$ in X is the element w_1 in W ($x_{1,1}=w_1$). Likewise, the element $x_{c,r+1}$ in X is the element w_τ in W ($x_{c,r+1}=w_\tau$). Next, the observations X_1 and X_2 can therefore be extracted from X as defined in (21) and (22) respectively.

$$X_1 = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,r} \\ x_{2,1} & x_{2,2} & \dots & x_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ x_{c,1} & x_{c,2} & \dots & x_{c,r} \end{bmatrix} \in X, \quad (21)$$

$$X_2 = \begin{bmatrix} x_{1,2} & x_{1,3} & \dots & x_{1,r+1} \\ x_{2,2} & x_{2,3} & \dots & x_{2,r+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{c,2} & x_{c,3} & \dots & x_{c,r+1} \end{bmatrix} \in X. \quad (22)$$

Thereafter, the DMD modes are derived according to the four steps highlighted in Section 4 of the decomposition process. The modes are represented in matrix form as defined in (23).

$$\mathcal{M} = \begin{bmatrix} \mathcal{M}_{1,1} & \mathcal{M}_{1,2} & \dots & \mathcal{M}_{1,r} \\ \mathcal{M}_{2,1} & \mathcal{M}_{2,2} & \dots & \mathcal{M}_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{M}_{c,1} & \mathcal{M}_{c,2} & \dots & \mathcal{M}_{c,r} \end{bmatrix} \quad (23)$$

Therefore, feature vector can be derived from (23) as defined in (24).

$$\mathcal{V}_t = \left[\left(\frac{1}{c} \sum_{a=1}^c |\mathcal{M}_a| \right)_1, \left(\frac{1}{c} \sum_{a=1}^c |\mathcal{M}_a| \right)_2, \dots, \left(\frac{1}{c} \sum_{a=1}^c |\mathcal{M}_a| \right)_r \right] \\ = [l_1, l_2, \dots, l_r] \quad (24)$$

For a vocalisation sample, S , divided into k number of windows with roughly τ numbers of sampling points, the feature vector can be derived in form of a matrix as defined in (25).

$$\mathcal{V}_{DMD} = \begin{bmatrix} l_{1,1} & l_{1,2} & \dots & l_{1,r} \\ l_{2,1} & l_{2,2} & \dots & l_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ l_{k,1} & l_{k,2} & \dots & l_{k,r} \end{bmatrix} \quad (25)$$

Note that the DMD modes matrix has r columns, which means that the DMD-HMM detection technique is r -dimensional. Using simulations runs as shown in Section 7.1, the value of r is selected to be nine ($r=9$) for both the Humpback whale songs and the inshore Bryde's whale short pulse call. The proposed DMD feature extraction method is hereby summarised in Algorithm 1.

Algorithm 1. DMD Feature Vector.

Algorithm 1: DMD Feature Vector.**Input:** W, k, τ **Output:** $\mathcal{V}_{DMD}$ as the extracted feature**1:** Determine the rank, r of the modes**2:** Restructured W wrt r as defined in (20)**3:** Obtain X_1 and X_2 , where $X_1, X_2 \in X$ **4:** Computes the SVD of X_1 as in (7)**5:** Computes matrix $\tilde{B}$ as in (9)**6:** Computes the eigenvector and eigenvalues from $\tilde{B}$ as in (10)**7:** Computes the DMD modes M from the eigenvector, Υ as in (11)**8:** Derive $\mathcal{V}_{DMD}$ (24) and (25)**6. Sample training**

The sound recordings for the Humpback whale songs dataset and the Bryde's whale short pulse calls dataset were visually and aurally analysed with the aid of Sonic Visualiser. Both recordings were divided into two unequal sections, where the small section (about 25% of the whole sound file) is used to train the DMD-HMM detector. These training samples are annotated in two categories. In the case of the Humpback whale songs dataset, the two categories are (1) Humpback whale songs, and (2) Noise. While the annotations for the Bryde's whale short pulse calls dataset, are categorised into (1) Bryde's whale short pulse calls, and (2) Noise. During the annotations, in each case, the sound that cannot be recognised visually or aurally are taken as noise. Also note that the large sound sections are annotated in order to verify the performance of the proposed DMD-HMM detector.

Table 1
DMD-HMM comparison for different r : $\tau=500$.

Humpback whale songs						
r	ρ (%)			λ (%)		
	$\eta=30$	$\eta=24$	$\eta=18$	$\eta=30$	$\eta=24$	$\eta=18$
7	77.76	75.65	72.44	3.80	5.22	6.21
8	78.87	76.98	74.13	3.17	4.44	5.55
9	80.88	78.77	76.03	2.62	3.99	5.00
10	80.88	78.77	76.04	2.62	3.99	5.00
11	80.89	78.77	76.04	2.63	3.99	5.00

Bryde's whale short pulse calls						
r	ρ (%)			λ (%)		
	$\eta=18$	$\eta=12$	$\eta=6$	$\eta=18$	$\eta=12$	$\eta=6$
7	99.13	98.12	96.99	0.07	0.08	0.10
8	99.55	98.38	97.83	0.06	0.07	0.09
9	99.70	98.91	98.21	0.05	0.06	0.08
10	99.70	98.91	98.21	0.05	0.06	0.08
11	99.70	98.92	98.21	0.05	0.06	0.07

Having categorised the training samples into two, a total of two HMMs were formed in each case, where one is used to model the desired sound (Humpback whale songs or Bryde's whale short pulse calls) and the other is used to model the unwanted sound (noise). In the case of the Bryde's whale short pulse calls, the number of test samples, η is varied between 6 and 18 samples; that is, $\eta=6$, $\eta=12$, and $\eta=18$. On the other hand, the test samples for the Humpback whale songs is varied between 18 and 30 ($\eta=18$, $\eta=24$, and $\eta=30$). Larger number of test samples are used to train the Humpback whale songs as compared to the Bryde's whale short pulse calls because the waveform duration of the Humpback whale songs is longer. As such, it requires more test samples to train the HMM so as to improve the reliability of the detector. Furthermore, the training samples are divided into τ number of sampling point, whereby performance evaluations are carried out for different value of τ ($\tau=100$, $\tau=300$, and $\tau=500$).

An ergodic type HMM with four states and two mixture is used in this article. The desired sound and the unwanted sound are trained separately with the four states and two mixture HMM. The dimension of the

Table 2
DMD-HMM comparison for different r : $\tau=300$.

Humpback whale songs						
r	ρ (%)			λ (%)		
	$\eta=30$	$\eta=24$	$\eta=18$	$\eta=30$	$\eta=24$	$\eta=18$
7	79.02	76.13	74.44	3.40	4.11	5.65
8	81.33	79.22	76.56	2.77	3.64	5.01
9	84.08	81.18	79.69	1.56	2.35	3.89
10	84.08	81.18	79.69	1.56	2.35	3.90
11	84.08	81.18	79.69	1.56	2.35	3.90

Bryde's whale short pulse calls						
r	ρ (%)			λ (%)		
	$\eta=18$	$\eta=12$	$\eta=6$	$\eta=18$	$\eta=12$	$\eta=6$
7	99.22	98.89	97.99	0.05	0.07	0.09
8	99.61	99.02	98.17	0.05	0.06	0.08
9	99.80	99.20	98.29	0.04	0.05	0.06
10	99.80	99.20	98.29	0.04	0.05	0.06
11	99.80	99.20	98.30	0.04	0.05	0.06

Table 3
DMD-HMM comparison for different r : $\tau=100$.

Humpback whale songs						
r	ρ (%)			λ (%)		
	$\eta=30$	$\eta=24$	$\eta=18$	$\eta=30$	$\eta=24$	$\eta=18$
7	84.56	79.89	77.99	2.97	3.14	4.01
8	86.00	82.11	80.09	2.09	2.43	2.99
9	88.91	84.67	81.33	1.23	1.67	2.15
10	88.92	84.67	81.33	1.23	1.67	2.15
11	88.92	84.67	81.33	1.23	1.67	2.15

Bryde's whale short pulse calls						
r	ρ (%)			λ (%)		
	$\eta=18$	$\eta=12$	$\eta=6$	$\eta=18$	$\eta=12$	$\eta=6$
7	99.23	99.14	98.11	0.05	0.06	0.08
8	99.44	99.38	98.42	0.04	0.05	0.07
9	99.89	99.62	98.61	0.03	0.04	0.05
10	99.89	99.62	98.61	0.03	0.04	0.05
11	99.89	99.62	98.61	0.03	0.04	0.05

HMMs is determined by the size of the feature extraction vector as discussed in Section 5. After training the HMMs separately, the noise HMM model and the desired sound HMM model are combined to form an eight states and four mixture HMM model. Thus, the first 1–4 states represent the desired sound and the remaining 5–8 states represent the noise with equal switching or transition probabilities. To evaluate the detector, the combined HMM and the refined feature vector $\mathcal{V}_\theta$ of the unknown sound file to be detected (the large sound section) are fed into the V-alg. The V-alg uses the α and γ parameters to predict the sequence of states, thereby classifying the unknown signal as either the desired sound or the unwanted sound. Note, for emphasis, the Humpback whale songs dataset and the Bryde's whale short pulse calls dataset are analysed separately. Suitably, their results are presented separately as shown in subsequent sections.

7. Test results and discussion

7.1. Test results for different rank, r

The proposed DMD-HMM detector is firstly evaluated for different r in terms of the sensitivity, ρ and false discovery rate, λ . The sensitivity, ρ is the ability of the automatic detector to correctly detect or identify the sound as defined by (26):

$$\rho = \frac{TP}{TP + FN}. \quad (26)$$

On the other hand, the false discovery rate, λ is used to measure the reliability of the developed automatic detector as defined by (27):

$$\lambda = \frac{FP}{FP + TP}. \quad (27)$$

In (26) and (27), TP (true positive) is the number of times the output of automated detector tallies with the manually identified sound, FN (false negative) is the number of times the automated detector missed the manually identified sound and FP (false positive) is the number of times the detector erroneously recognise an unwanted sound as the desired sound.

Tables 1–3 show the results of the proposed DMD-HMM detector for different r . In the tables, the performance of the the DMD-HMM detector is shown for $r=7, 8, 9, 10$, and 11 for both the Humpback whale songs and the Bryde's whale short pulse calls. As can be noted, the ρ performance in the three tables improves as r increases from 7 to 9. However, the ρ performance of the detector flattens as r increase from 9 to 10. Likewise, there is no significant change in the ρ performance as r increase from 10 to 11. Thus, a further increase in the value of r do not necessarily improves the performance of the detector; rather, it complicates the computational burden of the HMM because the dimension of the model will increase. This is also evident in the λ performance of the DMD-HMM detector. To buttress this points, Figs. 4–7 present the performance of the detector for $r = 1 - 15$. From the figures, the performance of the detector flattens when r increases from 9 at the expense of

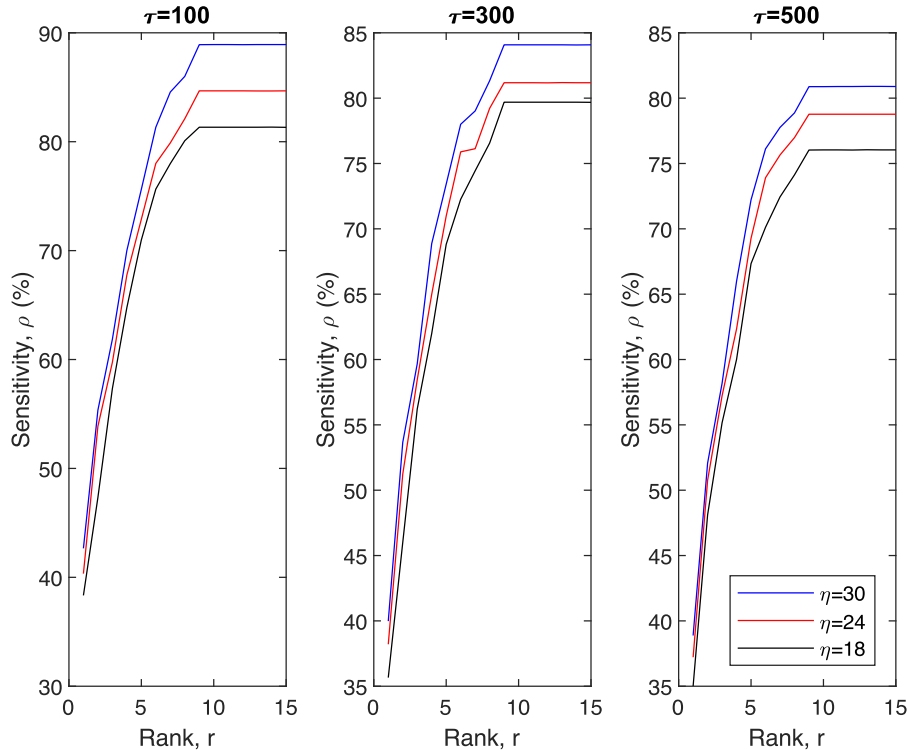


Fig. 4. Humpback whale: DMD-HMM detector ρ performance for different r .

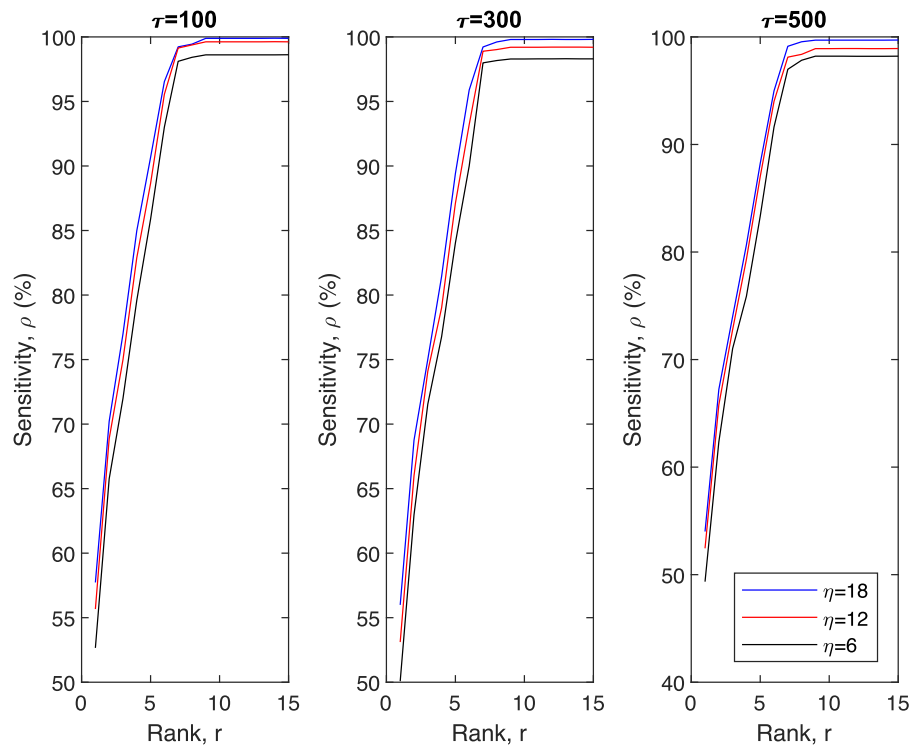


Fig. 5. Bryde's whale: DMD-HMM detector ρ performance for different r .

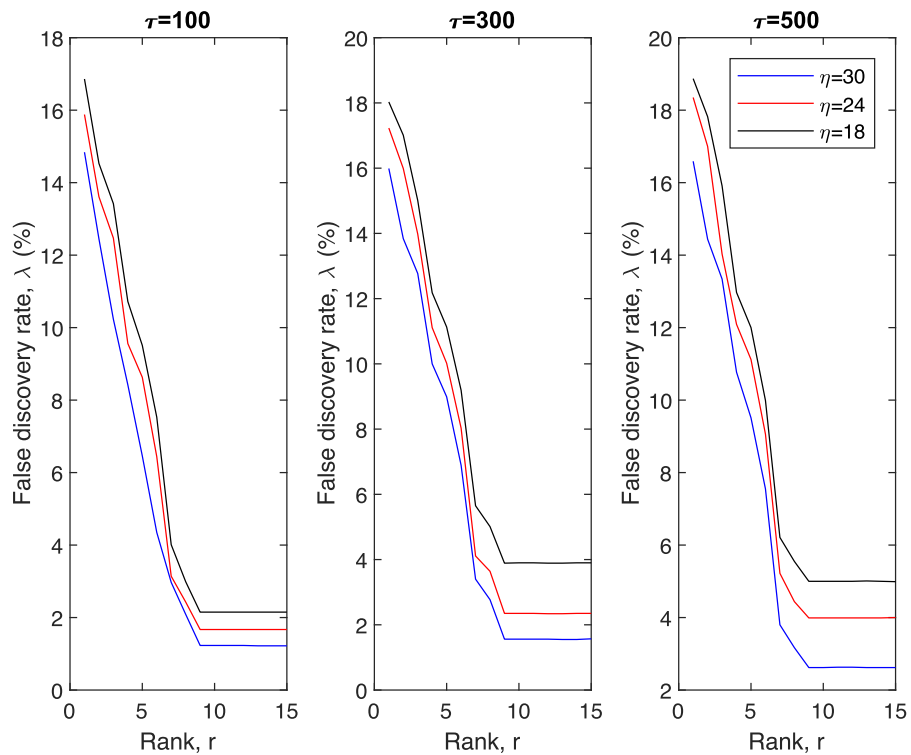


Fig. 6. Humpback whale: DMD-HMM detector λ performance for different r .

increased computational complexity. Therefore, in this article, $r=9$ is selected as the close to optimal value for the DMD-HMM detector, as it exhibits a good trade-off between the performance gain and complexity of the detector. (See [Tables 4 and 5](#).)

In addition, observe from the tables, the performance of the DMD-

HMM detector improves as the number of samples, τ reduce from 500 to 100. This performance gain is expected as in the case with most detection models that exhibit better ρ and λ performances for a reasonable small value of τ . Moreover, from the tables, the performance of the DMD-HMM detector offers better ρ and λ performances when used

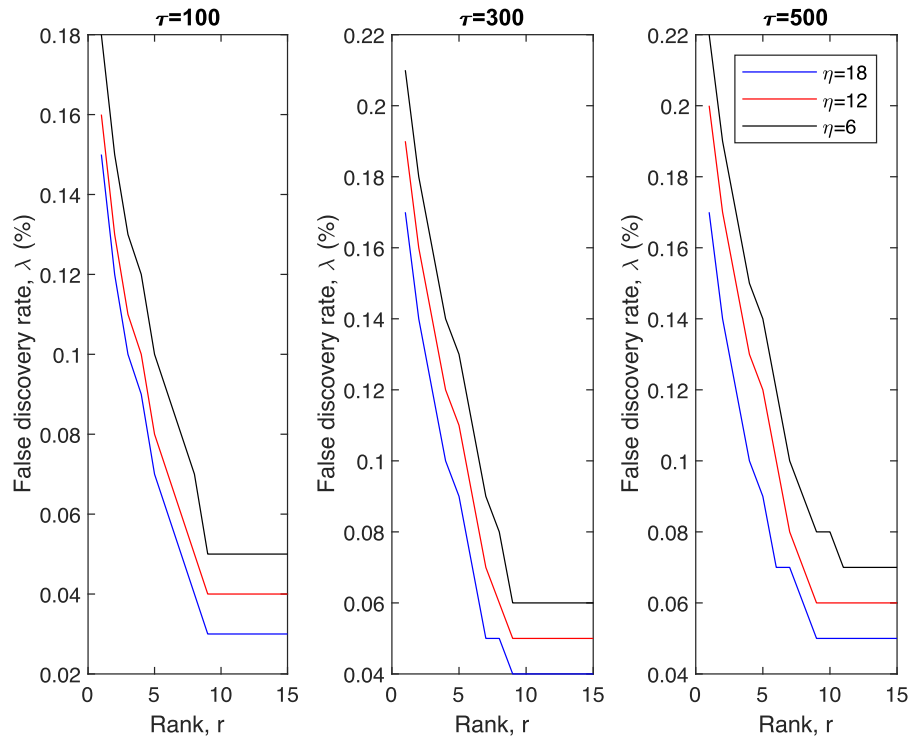
Fig. 7. Bryde's whale: DMD-HMM detector λ performance for different r .

Table 4

Comparison of different feature extraction methods: $\tau=500$.

Humpback whale songs								
$r=9, Q=9, \delta=12, \psi=12$								
η	ρ (%)				λ (%)			
	DMD	EMD	MFCC	LPC	DMD	EMD	MFCC	LPC
30	80.88	77.91	75.06	72.09	2.62	3.63	4.28	5.28
24	78.77	75.01	73.08	70.07	3.99	4.83	5.34	6.16
18	76.03	73.32	71.05	6.63	5.00	5.97	6.23	7.07

Bryde's whale short pulse calls								
$r=9, Q=6, \delta=12, \psi=12$								
η	ρ (%)				λ (%)			
	DMD	EMD	MFCC	LPC	DMD	EMD	MFCC	LPC
18	99.70	99.46	96.58	91.23	0.05	0.09	0.13	1.27
12	98.91	98.78	95.65	97.63	0.05	0.09	0.16	1.42
6	98.21	97.31	94.57	83.00	0.06	0.11	0.24	1.55

to detect the Bryde's whale short pulse calls as compared to the Humpback whale songs. This is because the Humpback whale songs waveform have longer duration; thus, they are more difficult to track. Hence, the reason why the number of test samples, η for the Humpback whale songs are more than that of the Bryde's whale short pulse calls. Notwithstanding, in Table 3, at $\tau=100$, $\eta=30$, and $r=9$, the detector achieves a ρ performance of %88.91 ($\rho=\%88.91$) and a λ performance of %1.23 ($\lambda=\%1.23$), which is acceptable for real-time sound detection applications. Interestingly, these performance values are better than what is obtainable with the existing methods as presented in subsequent section. Besides, it is envisaged that the obtained ρ and λ performances can be improved if the number of test samples, η increase, but will start to saturates at a particular value of η .

Table 5

Comparison of different feature extraction methods: $\tau=300$.

Humpback whale songs								
$r=9, Q=9, \delta=12, \psi=12$								
η	ρ (%)				λ (%)			
	DMD	EMD	MFCC	LPC	DMD	EMD	MFCC	LPC
30	84.08	81.76	78.79	75.03	1.56	1.99	2.46	3.33
24	81.18	77.98	75.11	72.34	2.35	3.01	3.62	4.04
18	79.69	75.66	72.06	70.01	3.89	4.22	4.99	5.87

Bryde's whale short pulse calls								
$r=9, Q=6, \delta=12, \psi=12$								
η	ρ (%)				λ (%)			
	DMD	EMD	MFCC	LPC	DMD	EMD	MFCC	LPC
18	99.80	99.67	97.71	92.59	0.04	0.08	0.13	1.23
12	99.20	98.98	96.77	88.12	0.04	0.08	0.14	1.37
6	98.29	97.89	95.65	83.99	0.08	0.11	0.22	1.44

7.2. Test results with other feature Extration methods

In this section, the performance of the proposed DMD-HMM detector at $r=9$ is compared with the MFCC-HMM detector, LPC-HMM detector and the recent EMD-HMM detector. The ρ and λ performances are presented in Tables 4–6. First, as shown in the tables, the DMD-HMM detector offers better ρ and λ performances at all values of τ . More so, the ρ and λ performances of all the detectors plummet with increase in the value of τ as anticipated. Second, as explained in the case of the DMD-HMM detector, the MFCC-HMM, LPC-HMM and EMD-HMM detectors obtained better ρ and λ performances with the Bryde's whale short pulse calls in comparison with the Humpback whale songs. Nevertheless, the DMD-HMM detector offered superior ρ and λ performances in comparison to the other detectors. For emphasis, in Table 6, at $\tau=100$, and $\eta=30$ (Humpback whale songs) the proposed DMD-HMM detector obtained a ρ

Table 6Comparison of different feature extraction methods: $\tau=100$.

Humpback whale songs								
$r=9, Q=9, \delta=12, \psi=12$								
η	ρ (%)				λ (%)			
	DMD	EMD	MFCC	LPC	DMD	EMD	MFCC	LPC
30	88.91	84.23	80.02	77.77	1.23	1.77	2.10	2.99
24	84.67	80.11	77.02	73.88	1.67	2.05	2.78	3.50
18	81.33	79.40	76.77	71.98	2.15	2.63	3.01	4.01

Bryde's whale short pulse calls								
$r=9, Q=6, \delta=12, \psi=12$								
η	ρ (%)				λ (%)			
	DMD	EMD	MFCC	LPC	DMD	EMD	MFCC	LPC
18	99.89	99.70	98.44	97.31	0.03	0.06	0.10	1.00
12	99.62	99.11	97.30	93.14	0.04	0.06	0.11	1.17
6	98.61	98.21	96.49	92.44	0.06	0.08	0.21	1.36

performance gain of %4.68, %8.89 and %11.14 in comparison to the EMD-HMM, MFCC-HMM, and LPC-HMM detectors respectively. Likewise, the λ performance of the DMD-HMM detector reduces by %0.54, %0.87 and %1.76 as compared to the EMD-HMM, MFCC-HMM, and LPC-HMM detectors respectively. This portrays the DMD-HMM detector as a performance efficient alternative to the EMD-HMM, MFCC-HMM, and LPC-HMM detectors.

Moreover, the MFCC-HMM and LPC-HMM detectors exhibits higher computational complexity in comparison to the proposed DMD-HMM detector. The DMD-HMM detector is a 9-dimensional model as compared to the 12-dimensional model of both the MFCC-HMM and LPC-HMM detectors. As earlier explained, the higher the dimension of the model, the more the computational burden in performing the HMM process. This is an added advantage to the DMD-HMM detector over the MFCC-HMM and LPC-HMM detectors. In comparison to the EMD-HMM detector, the two detectors offer the same computational burden to the HMM process in detecting the Humpback whale songs because they are both 9-dimensional. On the other hand, in detecting the Bryde's whale short pulse calls, the DMD-HMM detector exhibit a higher computational burden in comparison to the EMD-HMM detector. Regardless, the DMD-HMM detector offered superior ρ and λ performance gains over the EMD-HMM detector. This makes it a suitable alternative over the EMD-HMM detector in terms of the model detection accuracies. Lastly, just as in the case of the existing detectors used for comparison in this article, this proposed DMD-HMM detector can be easily modified to detect other *Mysticetes*' vocalisations depending on the application requirements.

8. Conclusion

This article has introduced the method of DMD as a feature extraction technique that can be adopted with the HMMs for the detection of *Mysticetes*' vocalisations. The output (spatial-temporal modes) of the DMD process is modified to form reliable feature vectors that can be combined with the HMM. The sensitivity and false discovery rate performances of this proposed detection model was verified using two different *Mysticetes*' vocalisations. Furthermore, the performance of the proposed DMD-HMM detector was compared with three existing methods (MFCC-HMM, LPC-HMM and EMD-HMM) in order to test its superiority. Interestingly, this proposed DMD-HMM detector proved to be very effective, as it offered good sensitivity and false discovery rate performance gains in comparison to these existing methods. More so, this proposed detector model achieved this feat at a reduced computational time and delay complexity. Therefore, we propound that the DMD-HMM detector can be a performance efficient alternative to these existing methods in studying *Mysticetes*' that produces characteristics

vocalisation. More so, the introduced methodology as depicted in Fig. 3 can be adapted by researchers to detect other marine animal sounds. In some cases, the DMD feature extraction process can be modified or restructured to accommodate the call type.

Declaration of Competing Interest

None.

Acknowledgements

The financial assistance of the National Research Foundation (NRF) of South Africa towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF (Grant number: 116036).

References

- Baum, L.E., Petrie, T., Soules, G., Weiss, N., 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.* 41, 164–171.
- Committee on Taxonomy (2017). List of marine mammal species and subspecies. Society of Marine Mammalogy, . www.marinemammalscience.org (download: April, 2019).
- Constantine, R., Iwata, T., Nieuwkirk, S.L., Penry, G.S., 2018. Future directions in research on bryde's whales. *Front. Mar. Sci.* 5, 333.
- Duda, R., Hart, P., Stork, G.D., 2001. *Pattern Classification (volume 2nd ed)*.
- Forgy, E.W., 1965. Cluster analysis of multivariate data: efficiency versus interpretability of classification. *Biometrics* 21 (3), 768–769.
- Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H., 1988. 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), 903–995.
- Jurafsky, D., James, H.M., 2018–2019. *Speech and Language Processing*. Downloaded: 15th April, pp. 1–17.
- Kennedy, J., 2020. 14 Types of Baleen Whales. ThoughtCo. [thoughtco.com/types-of-baleen-whales-2291520](https://www.thoughtco.com/types-of-baleen-whales-2291520).
- Liu, B., Feng, Y., Yin, Z., Fan, X., 2019. Radar signal emitter recognition based on combined ensemble empirical mode decomposition and the generalized S-transform. *Math. Probl. Eng.* 2019, 1–15.
- MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1. University of California Press, pp. 281–297.
- Majeed, S., Husain, H., Abdul Samad, S., Idbeaa, T., 2015. Mel frequency cepstral coefficients (MFCC) feature extraction enhancement in the application of speech recognition: a comparison study. *J. Theor. Appl. Inf. Technol.* 79, 38–56.
- Makhoul, J., 1975. Linear prediction: a tutorial review. *Proc. IEEE* 63, 561–580.
- Min, C.-H., Tewfik, A., 2010. Automatic characterization and detection of behavioral patterns using linear predictive coding of accelerometer sensor data. In: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 220–223.
- Mohan, N., Soman, K.P., Kumar, S.S., 2018. A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model. *Appl. Energy* 232, 224–229.
- Ogundile, O.O., Versfeld, D.J.J., 2020. Analysis of template-based detection algorithms for inshore Bryde's whale short pulse calls. *IEEE Access* 8, 14377–14385.
- Ogundile, O.O., Usman, A.M., Babalola, O.P., Versfeld, D.J.J., 2020a. A hidden Markov model with selective time domain feature extraction to detect inshore Bryde's whale short pulse calls. *Ecol. Informat.* 57 (Art. no. 101087).
- Ogundile, O.O., Usman, A.M., Versfeld, D.J.J., 2020b. An empirical mode decomposition based hidden Markov model approach for detection of Bryde's whale pulse calls. *J. Acoust. Soc. Amer.* 147 (2), EL125–EL131.
- Olsen, Ø., 1913. On the external characters and biology of Bryde's whale (*Balœnoptera brydei*), a new Rorqual from the coast of South Africa. *Proc. Zool. Soc. London* 1073–1089.
- Pace, F., 2008. *Comparison of Feature Sets for Humpback Whale Song Classification*. Master dissertation University of Southampton.
- Putland, R., Ranjard, L., Constantine, R., Radford, C., 2018. A hidden Markov model approach to indicate Bryde's whale acoustics. *Ecol. Indic.* 84, 479–487.
- Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* 77, 257–286.
- Rabiner, L.R., Juang, B.H., 1986. An introduction to hidden Markov models. *IEEE ASSP Mag.* 3, 4–16.
- Reynolds, D., 2009. Gaussian mixture models. In: Li, S.Z., Jain, A. (Eds.), *Encyclopedia of Biometrics*. Springer US, Boston, MA, pp. 659–663.
- Roch, M.A., Soldevilla, M.S., Hoenigman, R., Wiggins, S.M., Hildebrand, J.A., 2008. Comparison of machine learning techniques for the classification of echolocation clicks from three species of odontocetes. *Can. Acoust.* 36 (1), 41–47.

- Sasaki, T., Nikaido, M., Hamilton, H., Goto, M., Kato, H., Kanda, N., Pastene, L.A., Fordyce, R.W., Hasegawa, M., Okada, N., & Liu, H. H., 2005. Mitochondrial Phylogenetics and evolution of Mysticete whales. *Syst. Biol.* 54 (1), 77–90.
- Schmid, P.J., Sesterhenn, J., 2010. Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.* 656, 5–28.
- Sikha, O.K., Kumar, S.S., Soman, K.P., 2017. Salient region detection and object segmentation in color images using dynamic mode decomposition. *J. Comput. Sci.* 25, 351–366.
- Smith, B.D., Kershaw, F., Minton, G., Rosenbaum, H.C., Leslie, M.S., Anderson, R.C., LeDuc, R.G., Baldwin, R., Brownell, J., Robert, L., Mansur, R.M., Collins, T., 2013. Population differentiation of 2 forms of Bryde's whales in the Indian and Pacific oceans. *J. Hered.* 104, 755–764.
- Tirunagari, S., Kouchaki, S., Poh, N., Bober, M., Windridge, D., 2017. Dynamic mode decomposition for univariate time series: analysing trends and forecasting. <https://hal.archives-ouvertes.fr/hal-01463744>.
- Tolkova, I., Bauer, L., Wilby, A., Kastner, R., Seger, K., 2017. Automatic classification of humpback whale social calls. *J. Acoust. Soc. Amer.* 141 (5), 3605.
- Usman, A.M., Ogundile, O.O., Versfeld, D.J.J., 2020. Review of automatic detection and classification techniques for cetacean vocalization. *IEEE Access* 8, 105181–105206.
- Viterbi, A., 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inf. Theory* 13, 260–269.
- Yao, R., Johnson, M., Clemens, P., Michael, D., Glaeser, S., Osiejuk, T., Ebenezer, O.-N., 2009. A framework for bioacoustic vocalization analysis using hidden Markov models. *Algorithms* 2, 1410–1428.
- Zheng, F., Zhang, G., Song, Z., 2001. Comparison of different implementations of MFCC. *J. Comput. Sci. Technol.* 16, 582–589.