January 22, 2025

Classification of Bryde's whale individuals using high-resolution time-frequency transforms and support vector machines

Jean Baptiste Tary[1,a], Christine Peirce[2], Richard W. Hobbs[2]

[1] *Geophysics section, School of Cosmic Physics, Dublin Institute for Advanced Studies, Dublin, Ireland*

[2] *Department of Earth Sciences, Durham University, Lower Mountjoy, South Road, Durham, DH13LE, UK*

Abstract

Whales generate vocalizations which may, deliberately or not, encode caller identity cues. In this study, we analyze calls produced by Bryde's whales and recorded by ocean-bottom arrays of hydrophones deployed close to the Costa Rica Rift in the Panama basin. These repetitive calls, consisting of two main frequency components at ~20 and ~36 Hz, have been shown to follow five coherent spatio-temporal tracks. Here, we use a high-resolution time-frequency transform, the 4th-order Fourier synchrosqueezing transform (FSST4), to extract time-frequency characteristics (ridges) from each call to appraise their suitability for identifying individuals from each other. Focusing on high-quality calls recorded less than 5 km from their source, we then cluster these ridges using a Support Vector Machine (SVM) model resulting in an average cross-validation error of ~11% and balanced accuracy of ~86 ±5%. Comparing these results with those obtained using the standard short-time Fourier transform, *k*-means clustering, and lower-quality signals, the FSST4 approach, coupled with SVM, substantially improves classification. Consequently, the Bryde's whale calls potentially contain individual-specific information, implying that individuals can be studied using ocean-bottom data.

[a] tary@cp.dias.ie

I. INTRODUCTION

As animals interact with each other, they often, intentionally or not, encode individual-specific information in their communication (e.g., *Janik, 2009*). There can be different causes of dissimilarities in acoustic signatures between individuals, such as physical characteristics, environmental and cultural conditions, and temporal changes in these characteristics and conditions (*Knight et al., 2024*). This information can then be used by the animals for various purposes (e.g., territory definition, conspecifics and offspring identification) and for their study (e.g., population size, migrations, general behavior). For the study of cetaceans, identifying specific individuals is key to better understanding species' ecology and evolution over time, their environment, and anthropogenic impacts (e.g., climate changes, water and acoustic pollution, shipping operations etc.). Various approaches to individual identification have been developed over the years, ranging from visual observations, animal tags, and acoustic tags. Alternatively, passive monitoring of animal underwater acoustic communications provides an opportunity to monitor cetaceans for longer time periods and over larger areas, and especially so if individuals can be identified. Large datasets of acoustical signatures including individual animal attribution are, however, challenging to obtain for different reasons, such as the labor-intensive nature of using animal tags, source attribution for passive monitoring studies, signal deterioration for long-range applications, and background acoustical conditions.

Identifying individuals using acoustic recordings has been applied to a wide range of animals such as the South Polar skua (*Charrier et al., 2001*) and gorillas (*Salmi et al., 2014*) in addition to cetaceans, where the latter studies include Bottlenose dolphins (*Janik & Sayigh, 2013*) and Sperm whales (*Gero et al., 2016*). Sperm whale vocalizations are characterized by complex signals and temporal patterns, which define both vocal clans at the scale of an ocean basin and individuals (*Gero et al., 2016; Oliveira et al., 2016; Bermant et al., 2019*). In the case of baleen whales, fewer studies of individual identification exist, and these are mainly of Humpback whales. However, identifying individuals is challenging for various reasons, ranging from signal source attribution (*Zeh et al., 2024*) to limited knowledge of their vocal repertoire (e.g., *White & Todd, 2024*), where characteristics are also species dependent. In the case of Humpback whales, individuals have been identified using song cepstral content together with a Support Vector Machine (SVM) model (*Mazhar et al., 2007*), the use of some signal units in songs and their combinations (*Lamoni et al., 2023*), and call temporal patterns and amplitudes (*Zeh et al., 2024*). In addition, *McDonald et al. (2001)* suggested that some specific frequency characteristics of A-B calls of Blue whales could be used to identify individuals, and that frequency features extracted from spectrograms could provide information on individual North Atlantic right whales (*McCordic et al., 2016*).

In this study, we focus on whale calls recorded by Ocean-Bottom Seismographs (OBS) and a vertical array (VA) of hydrophones deployed in the Panama Basin in January and February, 2015 (*Hobbs & Peirce, 2015*; *Tary et al., 2024*) (Figure 1). The calls under consideration are very similar, short in duration (~3-5 s-long) and consist of two main frequency components: a ~1 s-long component at ~36 Hz and a ~3 s-long component at ~20 Hz (Figure 2). These calls are identified as likely corresponding to Be1 calls attributed to Bryde's whales by *Oleson et al. (2003)*. In this region of the Eastern Tropical Pacific Ocean, Bryde's whales are common despite their low abundance (*Wade & Gerrodette, 1993*; *Palacios et al., 2012*). The location of the calls generated by whales within the OBS network have lateral uncertainties less than a few kilometers (*Tary et al., 2024*). Some of these calls occur in spatio-temporal sequences and form trajectories across the network (Figure 1).

Bryde's whales generally travel as individuals or in pairs, and rarely in larger groups. With the call localization not having the resolution to distinguish between collocated whales (i.e., whales separated by distances on the order of 10s to 100s of meters), we assume that each of these trajectories corresponds to a different individual whale and determine if the associated calls can be classified into different groups based on their time-frequency features. If so, this could indicate that individual information is encoded within these relatively simple calls.

The calls under investigation are relatively short with simple characteristics, which contrasts with other calls used for individual identification such as those of Humpback whales (*White & Todd, 2024*). For short and simple calls, general attributes would likely not provide sufficient information to distinguish between individuals, considering other causes of variability such as animal behaviors, wave propagation effects, and the impact of background noises (natural or anthropogenic). In order to capture several attributes at the same time (e.g., signal component durations, mean frequencies, frequency modulations), here we employ time-frequency representations to classify these calls. Whale calls are generally analyzed using their spectrogram (e.g., *Mellinger & Clark, 2000*). In order to improve the definition of time-frequency information extracted from each call, instead of using the full time-frequency representations, we extract time-frequency ridges from representations obtained using a variant of the short-time Fourier transform (STFT), called the high-order synchrosqueezing transform (FSSTN - *Pham & Meignen, 2017*).

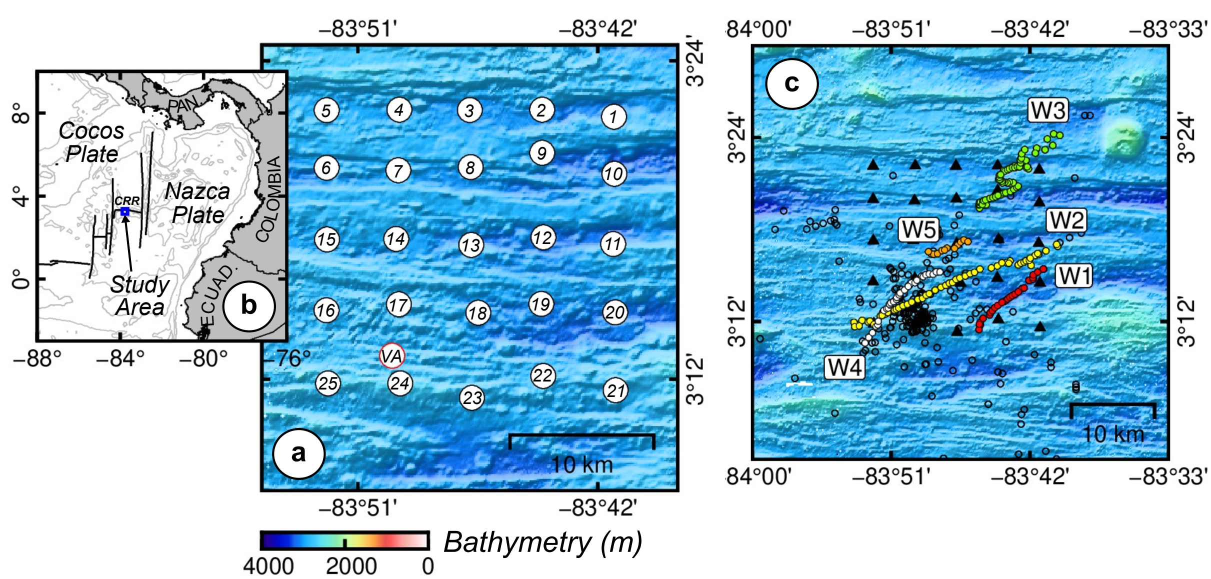

FIG. 1. (Color online). a) Bathymetry map of the survey area showing the ocean-bottom seismograph (OBS - numbered circles) and vertical hydrophone array (VA - red circle) positions. b) Location of the survey area (blue rectangle) in the Panama basin, close to the Costa Rica Rift (CRR) spreading ridge boundary between the Cocos and Nazca tectonic plates. c) Whale call locations from *Tary et al. (2024)* (open circles), with color-coded circles corresponding to the high-quality whale calls included in the support vector machine classification (whale tracks 1: red, 2: yellow, 3: green, 4: white, and 5: orange). OBS locations are indicated by the black triangles.

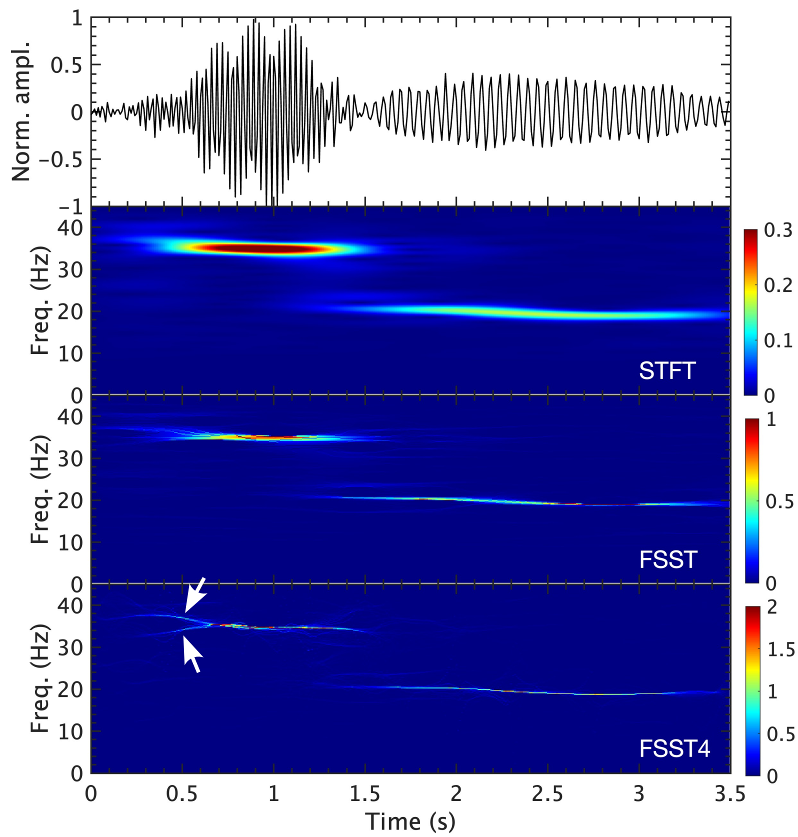

FIG. 2. (Color online). Whale call recorded by OBS 16 (see Figure 1 for location), contained within whale track 2 (Figure 1c), at 15:43:30 on January 29, 2015, and its time-frequency representations obtained using the short-time Fourier transform (STFT – Gaussian window of

~0.36 s at half-maximum with 97% overlap), the Fourier synchrosqueezing transform (FSST; 4th-order – FSST4). The white arrows locate the fork in frequency discussed in Section III.

Machine learning methods are generally applied to the detection and classification of whale calls (e.g., *Halkias et al., 2013*; *Ibrahim et al., 2021*; *Rasmussen & Širović, 2021*; *Zhong et al., 2021*; *Kather et al., 2024*), but seldom to caller identification (*Rendell & Whitehead, 2003*; *Mahzar et al., 2007*; *Bermant et al., 2019*) which is usually determined through call statistical analysis. Classifying observations in different categories using machine learning can be realized using unsupervised and supervised methods (e.g., *Bergen et al., 2019*). In our case, the whale locations can be transformed into labels to classify the signals using supervised methods. Of the different existing supervised methods, we employ support vector machines (SVM) for its demonstrated high performance in various applications (e.g., *Cervantes et al., 2020*), high generalization ability, and resistance to outliers and overfitting, even for high-dimensional data (*Kecman, 2005*). The SVM method, originally developed for binary classification, is a large margin classifier aimed at determining the optimal boundary between a subset of observations called support vectors (*Cortes & Vapnik, 1995*). To define non-linear decision boundaries between classes, the original data is often mapped to a higher-dimensional space, called the feature space, using kernels. In the present case, we demonstrate that different Bryde's whales can be distinguished using a SVM classifier using the time-frequency content of their brief calls recorded in ocean-bottom data.

II. DATA AND METHODS

Between January 26, 2015, and February 17, 2015, 25 OBSs and a VA of 12 hydrophones were deployed close to the Costa Rica Rift in the Panama basin (*Hobbs & Peirce, 2015*) (Figure 1).

This grid of instruments is approximately 20 x 20 km wide, with an instrument spacing of around 5 km. Apart from five OBSs (4, 7, 14, 17, 24) and the VA which recorded during the complete deployment, the remaining OBSs recorded from January 26, 2015 to their recovery time on February 1 or 2, 2015. The OBSs and VA were equipped with a High-Tech HTI-90-U hydrophone, while each OBS also had a 3-component short-period geophone package (Sercel L-28 4.5 Hz). Each time series was sampled at 500 Hz.

This dataset was first analyzed to study the structure of the oceanic lithosphere around the Costa Rica Rift (e.g., *Wilson et al., 2019*; *Robinson et al., 2020*). It was then re-examined to study the microseismicity (*Lowell et al., 2020*; *Tary et al., 2021*) and Bryde's whale calls (*Tary et al., 2024*) observed in this region. Focusing on the whale calls, two types of calls were observed; a repetitive call of ~4 s and a less common call consisting of brief signals of 0.5-1 s duration. The main characteristics of the most common, repetitive call, consist of two main signal components; a first wave packet of ~1 s duration centered at ~36 Hz, followed by a generally lower-amplitude, longer signal of ~3 s duration centered at ~20 Hz (Figure 2). These calls were then detected using two different methods; an energy method based on the short-term over long-term average ratio (STA/LTA), and template matching using the subspace detector applied to the calls detected by the first method. The arrival times of each call at the different instruments were then manually identified and all calls were located using a measurement-based 1D velocity model of the water column and the non-linear, probabilistic method implemented in NonLinLoc (*Lomax et al., 2001*). The calls were finally relocated relative to each other using the double-difference technique (*Waldhauser & Ellsworth, 2000*) (Figure 1c).

In order to examine the whale call characteristics and differences between individuals, we first focus on the whale calls that are well-recorded by the instruments (call-to-instrument distance < 5 km, calls at stations with manually identified start times), and observed in a single time period following a coherent spatial movement. Using these criteria, we identify five whale tracks observed at various times during the deployment (Figure 1c), whale track 1 with 27 calls recorded 71 times (January 29, 2015, some calls being recorded by more than one station), whale track 2 with 52 calls recorded 133 times (January 29, 2015), whale track 3 with 65 calls recorded 199 times (January 29, 2015), whale track 4 with 30 calls recorded 89 times (February 2, 2015), and whale track 5 with 12 calls recorded 38 times (January 26, 2015). All recorded calls for all whale tracks (i.e., 530 waveforms) are then included in the classification. The whale calls are first extracted using a longer time window of 4.5 s, and aligned using the manually identified start times. They are then re-aligned using waveform cross-correlation, relative to the event which is the most correlated to all events on average. Reviewing the call waveforms, the calls were then re-cut to 3 s duration from the call start times to both include the two main signal components and reject later signal parts with lower signal-to-noise ratio and/or other wave arrivals. The calls are finally down-sampled to a sampling rate of 100 Hz and band-pass filtered between 10 and 45 Hz.

A. Time frequency analysis: high-order synchrosqueezing transforms

The synchrosqueezing transform (SST) is a time-frequency representation which improves the readability of some time-frequency representations, such as the Continuous Wavelet Transform (CWT), by reassigning non-zero time-frequency coefficients to previously determined instantaneous frequencies (IF) (e.g., *Daubechies et al., 2011*). The main purpose of this synchrosqueezing operation is to significantly reduce frequency smearing (e.g., *Tary et al., 2014*). This has been applied to different time-frequency transforms such as the STFT (hereafter referred to as the FSST - *Thakur & Wu, 2011*) and the S-transform (*Huang et al., 2015*).

The SST was originally developed for slowly-varying, well-separated frequency components (*Daubechies et al., 2011*). However, in the case of the STFT, higher-order SSTs were developed, improving the time-frequency concentration and mode reconstruction of the FSST for strongly amplitude-modulated and frequency-modulated multi-component signals (*Oberlin et al., 2015*; *Pham & Meignen, 2017*). In this case, we first consider a signal $s(t)$ that can be decomposed into a series of K frequency components as

$$s(t) = \sum_{k=1}^{K} A_k(t)\, e^{i2\pi\theta_k(t)} + \varepsilon(t), \tag{1}$$

where $A_k(t)$ and $\theta_k(t)$ are the time-varying instantaneous amplitude and phase of the kth component, respectively, and $\varepsilon(t)$ is time-varying random noise. Instantaneous frequencies are then estimated using

$$\widehat{\omega}(t,\eta) = \frac{1}{2\pi}\frac{\partial \arg S_F(t,\eta)}{\partial t}, \tag{2}$$

where $\arg S_F(t,\eta)$ is the argument of the complex valued STFT representation $S_F(t,\eta)$ at time t and frequency η. In order to limit the reassignment of noise components, only non-zero frequency components above a pre-defined threshold ζ are reassigned on $\left(t, \widehat{\omega}(t,\eta)\right)$ positions following

$$T_F^{\zeta}(t,\omega) = \frac{1}{g^*(0)} \int_{\{\eta, |S_F(t,\eta)| > \zeta\}} S_F(t,\eta)\,\delta\left(\omega - \widehat{\omega}(t,\eta)\right) \mathrm{d}\eta, \tag{3}$$

where δ is the Dirac distribution and g^* the complex conjugate of the window function g. Here, we focus on the main aspects of the method presented by *Pham & Meignen (2017)*. Focusing on signal modes of s having non-negligible phase derivatives $\theta^{(n)}(t)$ for $n \geq 3$, using a high-order Taylor expansion of eq. 1 in τ close to t for a mode amplitude and phase gives

$$s(\tau) = \exp\left(\sum_{n=0}^{N} \frac{1}{n!}\left([\log(A)]^{(n)}(t) + i2\pi\theta^{(n)}(t)\right)(\tau - t)^n\right), \tag{4}$$

where $Z^{(n)}(t)$ is the n^{th} derivative of Z at t, and N is the order of the Taylor expansion of phase $\theta(\tau)$. Modifying the STFT representation $S_F(t,\eta)$ as well as the IF $\widehat{\omega}(t,\eta)$ accordingly, this requires the estimation of a frequency modulation operator $\tilde{q}^{[n,N]}$ and leads to the following definition of a N$^{\text{th}}$ order IF at time t and frequency η

$$\widetilde{\omega}_\eta^{[N]}(t,\eta) = \begin{cases} \widetilde{\omega}(t,\eta) + \sum_{n=2}^{N} \tilde{q}_\eta^{[n,N]}(t,\eta)\left(-x_{n,1}(t,\eta)\right), \text{if } S_F(t,\eta) \neq 0 \text{ and } \partial_\eta x_{j,j-1}(t,\eta) \neq 0, 2 \leq j \leq N \\ \widetilde{\omega}(t,\eta) \text{ otherwise} \end{cases}, \quad (5)$$

with

$$\begin{cases} x_{n,1}(t,\eta) = \dfrac{S_F^{t^{n-1}}(t,\eta)}{S_F(t,\eta)} \text{ for } 1 \leq n \leq N \\ x_{n,j}(t,\eta) = \dfrac{\partial_\eta x_{n,j-1}(t,\eta)}{\partial_\eta x_{j,j-1}(t,\eta)} \text{ for } 2 \leq j \leq N \text{ and } j \leq n \leq N \end{cases}. \quad (6)$$

The real part of $\widetilde{\omega}^{[N]}(t,\eta)$ is incorporated into eq. 3 for $T_F^\zeta(t,\omega)$ to obtain the N$^{\text{th}}$ order FSST. Frequency ridges are then extracted from the resulting time-frequency representations by iteratively searching for energy maxima in the time-frequency plane (*Meignen et al., 2017*). The threshold parameter ζ on the STFT representation is set to a small value of 0.001 in order to avoid removing any signal components. The window function used to calculate the STFT is a Gaussian function $g = \sigma^{-1} e^{-\frac{\pi t^2}{\sigma^2}}$. The parameter σ controls the width of the Gaussian window and, hence, the time and frequency localizations of the STFT (e.g., *Tary et al., 2014*). Its value in our case is 0.11, corresponding to the minimum of the Rényi entropy of STFT representations of the different whale calls presented in this study (*Stanković, 2001*).

To train the SVM model, two ridges are extracted from the 4$^{\text{th}}$ order SST representations (FSST4), corresponding to the two primary IF components of the whale calls (Figure 2). To transform each whale call into a one-dimensional vector, we only keep the IF of maximum amplitude for all time samples. As the numbers of training examples per whale is relatively limited, we also reduce the number of training attributes by down-sampling this ridge frequency vector by three (i.e., one point every 0.03 s) resulting in 100 IF measures per whale call (Figure 3). Finally, the IF data is normalized to obtain values ranging between -1 and 1.

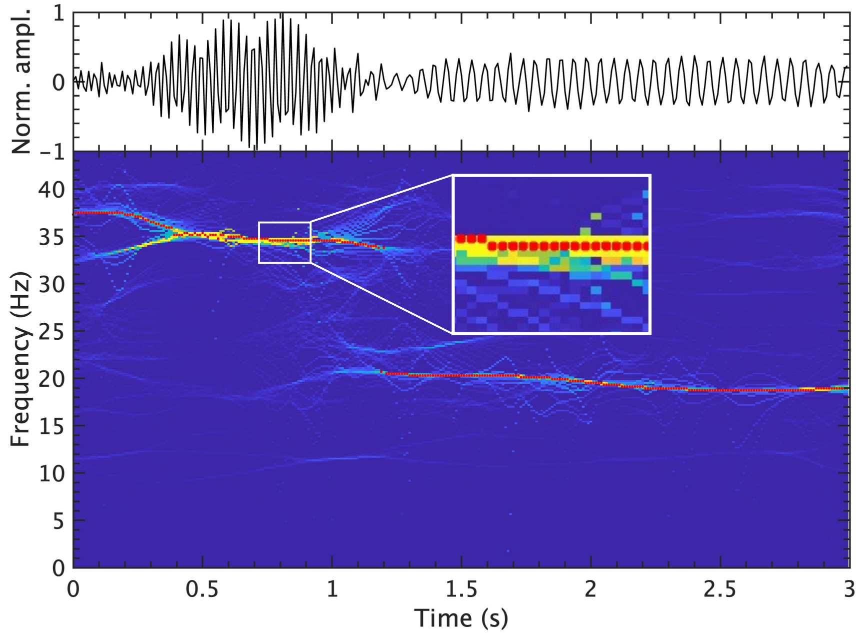

FIG. 3. (Color online). (top) Whale call recorded by OBS 24 at 15:43:30 on January 29, 2015, and contained within whale track 2. (bottom) The time-frequency representation obtained using the FSST4, together with the extracted ridge values (red dots), of which one in three values were used for SVM classification.

B. Classification support vector machine model

The SVM method is a supervised learning method that can be used for multi-class classification and regression (*Boser et al., 1992*; *Vapnik, 1995*). In general terms, the SVM method seeks to separate training examples based on their features, or these features after (non-)linear mapping, in a number of classes using the largest margin between some of the training examples located close to this margin. These margin-defining training examples are called support vectors and they define two optimal hyperplane positions separating the two classes. A multi-class classification using SVM is generally obtained using a series of two-class SVMs and combining their classification results. An SVM model can then be used to "predict" which class a new training example would belong to.

Using a training dataset consisting of M training examples $\{\mathbf{x}_i, y_i\}, i = 1, \dots, M$, each training example $\mathbf{x}_i$ is a vector of attributes and has a corresponding class label y_i ($y_i \in \{-1,1\}$). For linearly-separable data, a vector $\mathbf{w}$ and a scalar b exist so that

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \qquad i = 1, \dots, M, \tag{7}$$

is defining an optimal hyperplane defined by

$$\mathbf{w} \cdot \mathbf{x} + b = 0, \tag{8}$$

that separates the data points into two classes of y_i equal to 1 and -1 with the widest margin (*Cortes & Vapnik, 1995*). The geometrical distance of the training examples to the hyperplane is given by

$$\Delta_\mathrm{i} = \frac{y_i(\mathbf{w} \cdot \mathbf{x}_i + b)}{\|\mathbf{w}\|} \geq \frac{1}{\|\mathbf{w}\|}, \tag{9}$$

where $\|\mathbf{w}\|$ is the ℓ_2-norm of $\mathbf{w}$. Finding the optimum hyperplane corresponds to maximizing Δ_i for training examples close to the hyperplane or, equivalently, minimizing $\|\mathbf{w}\|$. The primal optimization problem can then be expressed as

$$\begin{gathered} \min_{\gamma,\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{s.t.}\ \ y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1,\ \ i = 1, \dots, M. \end{gathered} \tag{10}$$

Using the Lagrangian of this optimization problem we obtain

$$\mathcal{L}(\mathbf{w}, b, \alpha) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{M} \alpha_i [y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1], \tag{11}$$

where α_i are the Lagrange multipliers corresponding to each training example (*Cortes & Vapnik, 1995*). Minimizing $\mathcal{L}(\mathbf{w}, b, \alpha)$ implies that $\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i$ and results in the following dual optimization problem

$$\begin{gathered} \max_{\alpha} W(\alpha) = \sum_{i=1}^{M} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{M} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{s.t. } \alpha_i \geq 0, \;\; i = 1, \dots, M \\ \sum_{i=1}^{M} \alpha_i y_i = 0. \end{gathered} \tag{12}$$

Instead of directly using the training example attributes in $\mathbf{x}_i$, they can be transformed to a higher dimensional feature space using function $\phi(\mathbf{x}_i)$, modifying $\mathbf{w} = \sum_i \alpha_i y_i \phi(\mathbf{x}_i)$ and replacing the inner product $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$ in eq. 12 by $\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ which corresponds to the definition of a kernel (*Vapnik, 1995*). Most common are the linear kernel, the radial basis function (RBF) or Gaussian kernel, and the polynomial kernel. In addition, for variables that are not linearly separable, or in the case of data errors, it might not be possible or desirable to obtain a hyperplane that takes into account all training examples equally.

To overcome these limitations, the optimization problem can be modified to use soft margins controlled by a boundary parameter. In this case, using ℓ_1 regularization, the dual optimization problem of eq. 12 becomes

$$\begin{gathered} \max_{\alpha} W(\alpha) = \sum_{i=1}^{M} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{M} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{s.t. } 0 \leq \alpha_i \leq C, \;\; i = 1, \dots, M \\ \sum_{i=1}^{M} \alpha_i y_i = 0, \end{gathered} \tag{13}$$

where the upper bound parameter C defines the maximum penalty on training examples close to the margin boundaries. Besides SVM classifier optimization, we then have different hyperparameters to define to improve the results, namely the choice of kernel function, the upper bound parameter C, the kernel scaling parameter γ for the Gaussian kernel, and the polynomial order for the polynomial kernel function.

In this study, all data examples are first randomly shuffled and then partitioned into a training and a test set using a 90% - 10% split. To limit the influence of a particular split on the final results, we run 50 instances of this procedure using different randomizations and splits. The classification error is then estimated using the average balanced accuracy and its standard deviation, taking into account class number imbalance, and average cross-validation classification error (average of 4-fold, 5-fold and 10-fold cross-validation for the 50 runs). The statistical significance of all model classifications is assessed using permutation tests (*Ojala & Garriga, 2010*; *Combrisson & Jerbi, 2015*), which compare classifier performances using the original data to its performance using randomly permuted class labels (i.e., whale track numbers). The model is trained on the training dataset in the same way as the original model, and its performance measured by its balanced accuracy on the test set. This procedure is repeated 1000 times to determine the 99.9% percentile threshold of the balanced accuracy distribution, obtaining a significance level at $p < 0.001$.

III. RESULTS

The time-frequency content of whale calls is a key element for whale species identification and, hence, is useful for other purposes such as for their detection. The time-frequency representation generally in use is the STFT or the spectrogram. Figure 2 shows a Bryde's whale call example with high signal-to-noise ratio, together with its STFT, FSST and FSST4 representations. As visible in this example, the well-defined, quasi-harmonic components of these whale calls are highly suited for analysis using the SST (e.g., *Daubechies et al., 2011*). The frequency reassignment sharpens the two main time-frequency components of the signal and, thus, improves the readability of the representation. Comparing the FSST4 results with those of the other two transforms, the frequency resolution is higher and lower-amplitude frequency modulations are better delineated using this method. For example, in Figure 2 the fork in frequency around time 0.5 s is only clearly visible in the FSST4 representation. This likely arises from the ability of the FSST4 to better handle frequency modulated signals relative to the FSST (*Pham & Meignen, 2017*), which translates into a better estimation of the time-frequency ridges (Figure 3). The ridges compiled from all calls recorded by the OBSs located less than 5 km away from the whale location show that their time-frequency attributes exhibit only slight variations (Figure 4).

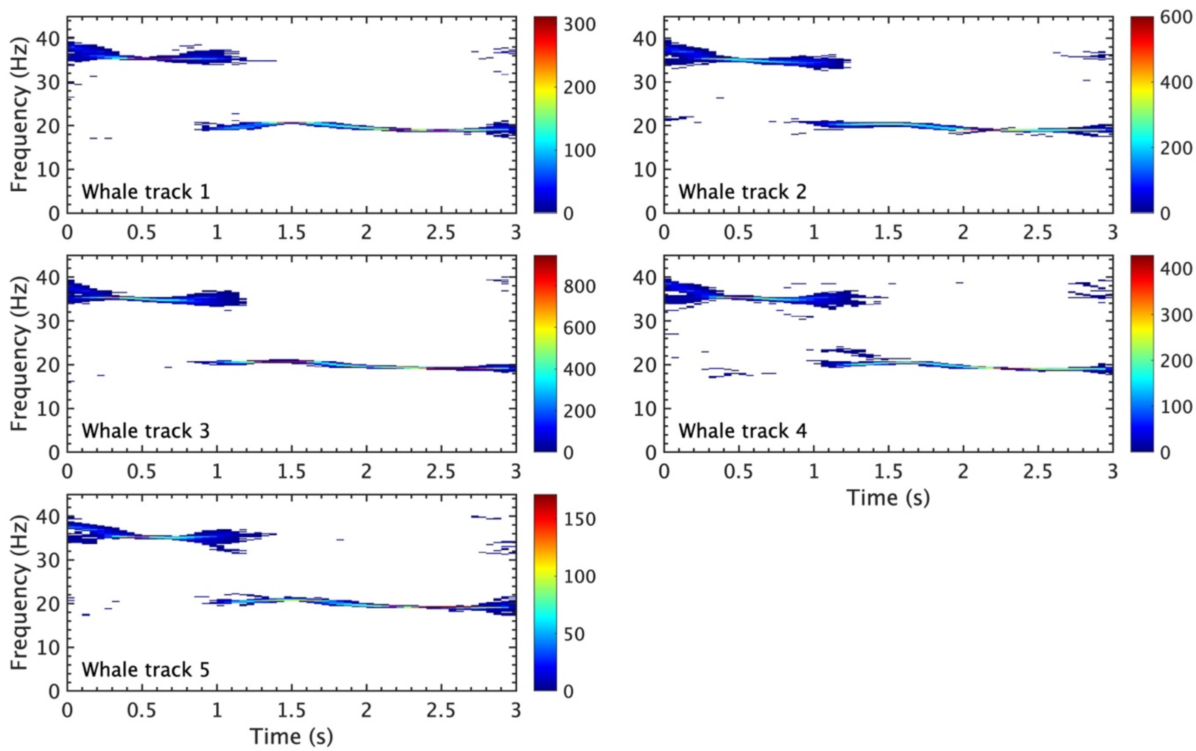

FIG. 4. (Color online). Density plots showing the ridges extracted from FSST4 representations for all high-quality whale calls contained in the five whale tracks (i.e., calls recorded within 5 km of the OBSs for periods showing coherent movements). These representations show

the two main signal components (i.e., at ~36 and ~20 Hz), and the similarities and potential differences of the signals between whale tracks.

For the first signal component at ~36 Hz, more variability is present for the lower amplitude parts of the signal at the beginning and end of the component. For most of the calls, the higher amplitude branch at the beginning of the signal shows increasing frequency between ~36 Hz and ~40 Hz. Less variability is observed for the second signal component. A number of time-frequency ridges are present at ~36 Hz for the second signal component, which likely correspond to calls with noisy components of lower amplitudes. Another source of signal variability is the timing of the end and beginning of the two signal parts between ~1 and ~1.5 s. Each of these variabilities may play a role in the ability of the classification to distinguish between different whale tracks.

To better demonstrate the advantages of the FSST4 and SVM, we consider the following four cases:

i) SVM classification using ridges extracted from FSST4 representations,
ii) SVM classification using ridges extracted from STFT representations,
iii) *k*-means classification using ridges extracted from FSST4 representations, and
iv) SVM classification using ridges extracted from FSST4 representations for lower-quality calls recorded at distances between 5 and 10 km from the whale location, enabling evaluation of the influence of signal quality on the classification performance.

The *a priori* clustering of high-dimensional data can be visualized using the t-Distributed Stochastic Neighbor Embedding (t-SNE) method (*van der Maaten & Hinton, 2008*), which performs a non-linear mapping of the high-dimensional data to lower dimensions. The t-SNE method is applied to the aforementioned cases using ℓ_1 distance as similarity metric and a perplexity of 30 (Figure 5). The t-SNE visualization shows that the best cluster separation is obtained using the FSST4 together with the high-quality signals (calls observed less than 5 km away from its source). No clear clusters are visible in the case of the lower-quality signals (calls observed 5 to 10 km away from its source). Interestingly, while calls from whale tracks 3 and 5 are well separated from the other calls, whale track 1 is slightly separated from whale tracks 2 and 4, and some mixing occurs for whale tracks 2 and 4 and for whale tracks 3 and 5.

The SVM classification models all use Gaussian kernels, the other hyperparameters being defined through 500 iterations of Bayesian optimization instead of grid search to reduce training time. Allowed values for the upper bound parameter C and the kernel scaling parameter γ range between 0.0001 and 10000. In the case of the *k*-means algorithm, the number of clusters is set to 5 and their initial centroids are set using the *k*-means++ algorithm (*Arthur & Vassilvitskii, 2007*). The final cluster centroids are then obtained by repeating five times the minimization of the sum of absolute distances between data points and centroids (ℓ_1 distance) using different initializations, keeping the centroids corresponding to the minimum total distance. This procedure reduces the probability of obtaining centroids corresponding to a local minimum far from the global minimum.

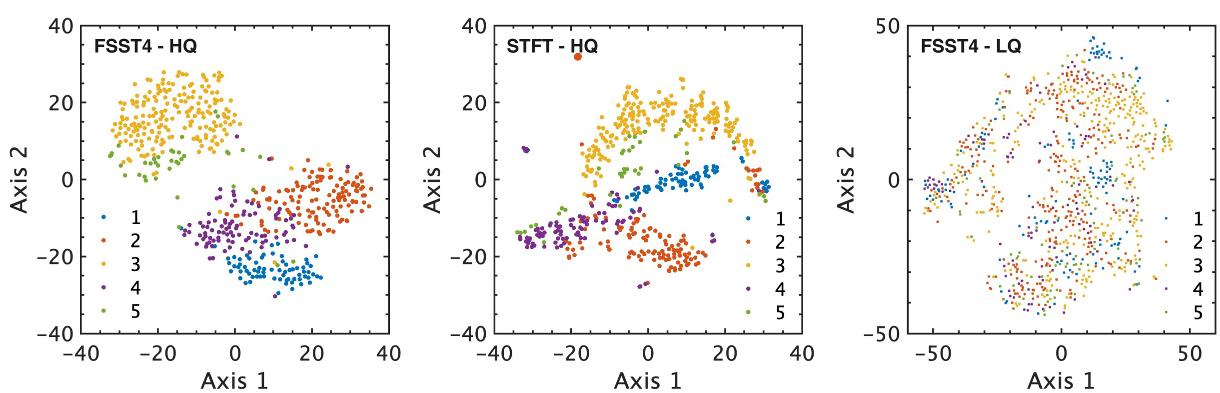

FIG. 5. (Color online). t-SNE visualization of the time-frequency ridge data, color-coded by whale track number (see Figure 1c), for ridges extracted using the FSST4 and high-quality signals (left), the STFT and high-quality signals (middle), and the FSST4 and lower-quality signals (right).

All three visualizations use a perplexity of 30 and ℓ_1 distance, and are similar to those obtained using higher perplexity values.

The different classification results are presented in Table I. For SVM using the FSST4 and high-quality calls observed at less than 5 km from the source (Gaussian kernel, upper bound C of ~38.1, kernel scale γ of ~25.0), we obtain a training error of ~0.4% for the model with the highest balanced accuracy on the test set (2 calls misclassified out of the 477 calls in the training set), a training average cross-validation error of ~11%, and a balanced accuracy of ~86 $\pm$5% on the test set (chance ~25% at p < 0.001). The confusion chart corresponding to this model shows that whale tracks 1 and 3 are associated with the least misclassification errors (i.e., ~8% and ~5%, respectively), whereas whale track 5 is associated with the highest misclassification error (~35%) (Figure 6). Using the STFT instead of the FSST4 slightly increases both the number of incorrectly classified whale calls during training and the training average cross-validation error to 14% (Gaussian kernel, upper bound C of ~51.5, kernel scale γ of ~10.2), and decreases the average balanced accuracy to ~78 $\pm$8% on the test set (chance ~25% at p < 0.001). The lowest misclassification errors are obtained for whale tracks 2 and 3, the highest classification error being associated with whale track 5 (~46%). Using lower quality signals with SVM and the FSST4 (Gaussian kernel, upper bound C of ~21.6, kernel scale γ of ~9.2), the classification error is ~1% with a higher average cross-validation error of ~38% and an average balanced accuracy of ~53 $\pm$5% on the test set (chance ~27% at p < 0.001). This might indicate that more overfitting is present in this model. In this case, the misclassification error is greater than 25% for all whale tracks, with whale track 4 having a misclassification error reaching ~78%. Lastly, using *k*-means and ridges extracted from high-quality calls using FSST4, we obtain a training error of ~39% and a balanced accuracy of ~64% (chance ~64% at p < 0.001). The misclassification error is consistently over 30% for all whale tracks, with a maximum of ~49% for whale track 3 (Figure 6).

Table I. Classification results for the different cases using either SVM or *k*-means, FSST4 or STFT, and high-quality (HQ, calls observed at less than 5 km from its source) or lower-quality (LQ, calls observed at distances between 5 and 10 km from its source) signals. For each combination, the training classification error for the model with the highest balanced accuracy on the test set is shown, together with the average training cross-validation error (C-V error, average of 4-fold, 5-fold and 10-fold cross-validation for the 50 runs) for SVM, and the balanced accuracy (Bacc, average and standard deviation over the 50 runs for SVM).

	Training error (%)	C-V error (%)	Bacc (%)
SVM + FSST4 + HQ signals	0.4	11	86 $\pm$5
SVM + STFT + HQ signals	2	14	78 $\pm$8
SVM + FSST4 + LQ signals	1	38	53 $\pm$5
k-means + FSST4 + HQ signals	39		64

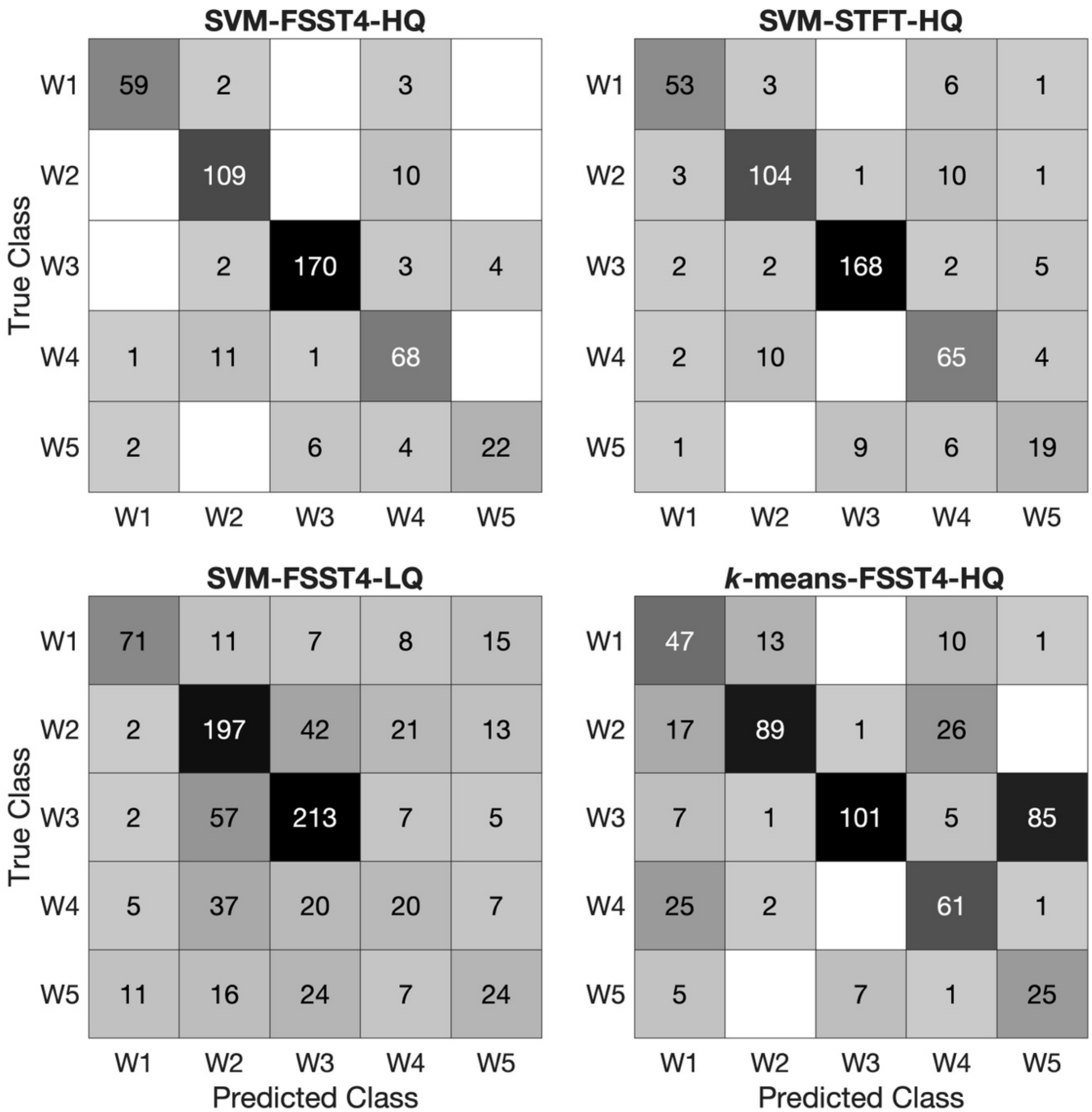

FIG. 6. (Color online). Multiclass confusion matrices obtained from cross-validating the SVM models having the highest balanced accuracy on the test set and the *k*-means results for the cases listed in Table I. Rows and columns of each matrix contain the number of calls in their actual class and the number of calls that were classified in each class by the model, respectively. W1, W2, W3, W4 and W5 correspond to whale tracks 1, 2, 3, 4 and 5, respectively.

Focusing on the SVM model using FSST4 and high-quality calls, we calculate SHAP (SHapley Additive exPlanations) values (*Lundberg & Lee, 2017*) in order to determine which features (i.e., frequency values at different time indexes) have the most effect on the final model classification (Figure 7). Shapley values quantify the changes in model classification due to a feature, for a given training example. Repeating this process for all features considering all training examples, the mean absolute Shapley values can be used to calculate SHAP feature importance; larger values corresponding to features more significant for the model. SHAP values can, however, be biased due to feature correlation, feature interaction and the small size of the training dataset (*Molnar, 2022*). Overall, the most significant features for the model classification are similar for the different whale tracks (Figure 7). Greater SHAP values are obtained at times from the beginning to ~0.3 s, ~0.6-0.7 s, ~0.8-0.9 s, ~1-1.3 s and most of the 2nd signal component from ~1.5 to ~2.6 s. The final part of the signal from ~2.6 to 3 s shows lower feature effects for all whale tracks.

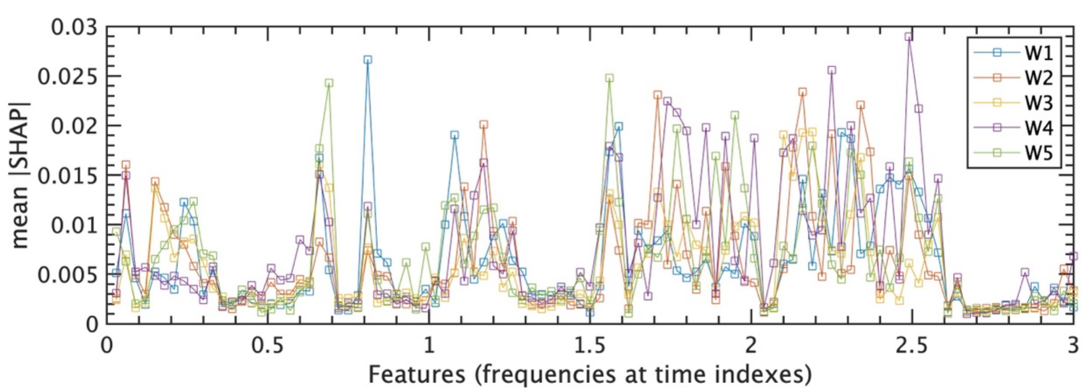

FIG. 7. (Color online). SHAP feature importance corresponding to mean absolute Shapley values for each class (i.e., whale tracks), computed using the training set for the SVM model, FSST4 transform, and high-quality signals.

IV. DISCUSSION

Time-frequency transforms are important techniques for the study of non-stationary features of signals emitted by whales. The most commonly used techniques to analyze whale calls are the STFT and the spectrogram. However, when the signals are constituted by narrow-band time-frequency components, they are well-suited for analysis by the SST (e.g., *Daubechies et al., 2011*). The examples presented in Figures 2 and 3 show that the FSST and FSST4 provide time-frequency representations of whale calls with better resolution than the STFT, which then aids the precise extraction of their characteristics and their interpretation. These SSTs are reversible which means that signal modes can be extracted and reconstructed. When signals are strongly frequency-modulated, which is often the case for whale sounds (e.g., for Humpback whales and Blue whales – *White & Todd, 2024*), high-order SST can be applied to the signal to better delineate the time-frequency features and avoid some mode mixing (*Pham & Meignen, 2017*). In this study, the whale calls consist mostly of two frequency components. As such, in principle, more time-frequency information could be included in the clustering, for example, more ridges or the full time-frequency representation. This would result in more feature parameters to be included in the clustering and would also require more training examples. If the full time-frequency representation of any of the SSTs is used, the thresholding parameter ζ would need to be better adjusted to remove noise components.

Using the FSST4 instead of the STFT seems to slightly improve the SVM classification results. This relatively small improvement might be due to the simple characteristics of the Bryde's whale calls. On the contrary, using high-quality calls instead of lower-quality calls, and SVM instead of *k*-means, appears to substantially improve the classification results. The large difference in average cross-validation error and average balanced accuracy, that depends on the signal quality, suggests that signals recorded close to the whales are needed for their identification. In the present case, this requires the signal to be observed by passive acoustic monitoring a few kilometers away from the calling whale. These conditions might, however, change depending on the method of analysis and classification, the recording conditions (i.e., sea bottom *vs.* sea surface, noise levels), and the type of calls (e.g., using temporal or frequency information, different frequency ranges).

For whales, caller identification is actually usually carried out using hydrophones deployed at the sea surface, close enough to the whales to enable them to be visually identified as well (e.g., *Gero et al., 2016*; *Lamoni et al., 2023*), and/or using dedicated instruments such as acoustic tags (*McCordic et al., 2016*; *Oliveira et al., 2016*; *Zeh et al., 2024*). Contrary to other studies performing classification using frequency measures extracted from spectra or time-frequency representations (e.g., *McDonald et al., 2001*; *McCordic et al., 2016*), the time-frequency ridges included in the classification of the present study implicitly incorporate various spectral measures such as mean frequencies of the different signal components, component durations, maximum and minimum frequencies, and frequency modulations over time. Still, other measures could be combined in the classification such as other types of calls, call temporal patterns (e.g., codas rhythms for Sperm whales), call amplitudes or data from other instruments (e.g., from geophones in the present case). Feature selection or grouping, through dimensionality reduction for instance, could also be applied to the time-frequency ridge values to decrease the classification model complexity and improve its interpretability.

The large difference between the results of SVM compared with those of *k*-means could indicate that models using non-linear decision boundaries are more suitable to correctly classify high-dimensional representations of whale calls. A disadvantage of SVM relative to *k*-means is, however, the difficulty in setting the model hyperparameters (e.g., kernel function, upper bound parameter). The classification results are mainly limited by the number of training examples available, especially for whale tracks 1, 4 and 5. The training dataset could be expanded in different ways including collecting more calls, using data from other instruments, and using data generation and augmentation strategies (e.g., *Zhu et al., 2020*). These strategies could correspond to using the same call several times with different noise types (real or synthetic), or making a synthetic call using a statistical description of the call properties (e.g., *Socheleau et al., 2015*). Another common limitation is the short observation window for each whale, due to both the temporary nature of most instrument deployments and the migratory behavior of whales, which often question the representativeness of the calls recorded.

While the present methodology and SVM model reach an average cross-validation error of ~11% and an average balanced accuracy of ~86 ±5%, more observations and further study would be needed to test its generalization to a larger population of Bryde's whales. Whale calls from whale tracks 1, 2 and 3 were all recorded during the same period (January 29, 2015). These calls can be separated in three separate tracks based on their locations and amplitudes. Bryde's whales generally travel as individuals or in pairs, and seldom in larger groups. In the classification presented in this study, we assume that each track is generated by a unique vocal whale. The unsupervised t-SNE clustering seems to show that the observed calls of the different whale tracks are different enough to define individual clusters. However, the t-SNE clustering also indicates that some calls that are known to correspond to different whales (i.e., whale tracks 1 and 2 both recorded at the same time on January, 29, 2015, but localized at different positions) can exhibit similar call features using our processing. Hence, having more than one vocal whale per whale track is still an open question, especially for whale track 3 which has the largest number of calls. The close proximity between whale tracks 1 and 2, shown by their joining tracks (Figure 1c) and the t-SNE visualization, could also be indicative of a closer connection between these two whales.

Regarding whale tracks 4 and 5 recorded on different days (i.e., February 2 and January 26, 2015, respectively), their changes in signal characteristics resulting in different clusters and classifications could be interpreted in various ways such as having five different vocal whales, or returns of whales also recorded on other days which would involve temporal changes in signal characteristics due to spatiotemporal changes in underwater signal propagation. More generally, the observed call differences resulting in their successful classification could arise from morphological differences between whales, whales belonging to different populations, and spatiotemporal changes in environmental conditions impacting signal propagation (e.g., *Knight et al., 2024*). Finally, the SVM model using subtle differences in call time-frequency characteristics to distinguish between whales, does not necessarily imply that Bryde's whales are using this information to identify themselves to conspecifics (e.g., *Gero et al., 2016*).

V. CONCLUSION

The identification of specific whale callers is important information for a range of applications such as studying whale movements and their change over time, and any external influence on their behavior. In the present study, we use highly similar low-frequency calls generated by five Bryde's whales recorded by ocean-bottom hydrophones and compute time-frequency ridges using the 4^{th}-order SST (FSST4) to extract the main frequency content of each call (i.e., time-frequency ridges). An SVM model is then trained using these time-frequency ridges to classify the whale calls. Using calls recorded less than 5 km away from the instruments, the average cross-validation error associated with the SVM model (Gaussian kernel) is ~11% with an average balanced accuracy of ~86 ±5%. Comparing these results with those using either STFT, lower-quality signals, or *k*-means clustering, shows that both the FSST4 and the SVM method improve the final results, with an increased error when using lower-quality signals and *k*-means clustering.

These classification results suggest that the short calls produced by Bryde's whales, and the time-frequency characteristics embedded in the extracted ridges, contain caller identity cues. They also seem to indicate that caller identity can be determined using ocean-bottom data, albeit using recordings less than a few kilometers away from the source. A larger number of training examples, coming from a larger number of well-identified whales and observed over a longer time period, would be needed to confirm these classification results. However, the methodology presented in this study does, nevertheless, show promising results and could be applied to other call types, improving the general understanding of whale vocalizations and ecology.

ACKNOWLEDGMENTS

This work was supported by the Natural Environment Research Council (NERC) via grant No. NE/1027010/1. We thank the people who participated in the data acquisition and processing used in this study (OSCAR research cruise JC114). The data were recorded by the UK's Natural Environment Research Council Ocean-Bottom Instrumentation Facility (*Minshull et al., 2005*). The FSSTn toolbox was used to compute the FSST and FSST4 representations (*Pham & Meignen, 2017*). The generic mapping tools (GMT) were used for some figures (*Wessel & Smith, 1991*). For the purposes of open access, the authors have applied a Creative Commons Attribution (CC-BY) license to any author accepted manuscript arising from this work.

AUTHOR DECLARATIONS

Conflict of Interest

The authors declare no conflict of interest.

DATA AVAILABILITY

Continuous data of Ocean Bottom Seismometers from cruise JC114 are archived at the NERC's British Oceanographic Data Centre and are available following: www.bodc.ac.uk/resources/inventories/cruise_inventory/report/15036/.

REFERENCES

Arthur, D., and Vassilvitskii, S. (**2007**). "K-means++: the advantages of careful seeding. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms (SODA '07)", Society for Industrial and Applied Mathematics, USA, 1027-1035.

Bermant, P. C., Bronstein, M. M., Wood, R. J., Gero, S., and Gruber, D. F. (**2019**). "Deep machine learning techniques for the detection and classification of sperm whale bioacoustics", Scientific reports, **9**(1), 12588.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (**1992**). "A training algorithm for optimal margin classifiers", In Proceedings of the fifth annual workshop on Computational learning theory, 144-152.

Charrier, I., Jouventin, P., Mathevon, N., and Aubin, T. (**2001**). "Individual identity coding depends on call type in the South Polar skua Catharacta maccormicki", Polar Biology, **24**, 378-382.

Combrisson, E., and Jerbi, K. (**2015**). "Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy", Journal of neuroscience methods, **250**, 126-136.

Cortes, C., and Vapnik, V. (**1995**). "Support-vector networks", Machine learning, **20**, 273-297.

Daubechies, I., Lu, J., and Wu, H. T. (**2011**). "Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool", Applied and computational harmonic analysis, **30**(2), 243-261.

Dunlop, R. A., Noad, M. J., Cato, D. H., and Stokes, D. (**2007**). “The social vocalization repertoire of east Australian migrating humpback whales (Megaptera novaeangliae)”, The Journal of the Acoustical Society of America, **122**(5), 2893-2905.

Gero, S., Whitehead, H., and Rendell, L. (**2016**). “Individual, unit and vocal clan level identity cues in sperm whale codas”, Royal Society Open Science, **3**(1), 150372.

Halkias, X. C., Paris, S., and Glotin, H. (**2013**). “Classification of mysticete sounds using machine learning techniques”, The Journal of the Acoustical Society of America, **134**(5), 3496-3505.

Hobbs, R., and Peirce, C. (**2015**). RRS James Cook JC114 Cruise Report. Online Report. https://www.bodc.ac.uk/resources/inventories/cruise_inventory/reports/jc114.pdf.

Huang, Z. L., Zhang, J., Zhao, T. H., and Sun, Y. (**2015**). “Synchrosqueezing S-transform and its application in seismic spectral decomposition”, IEEE Transactions on Geoscience and Remote Sensing, **54**(2), 817-825.

Ibrahim, A. K., Zhuang, H., Chérubin, L. M., Erdol, N., O'Corry-Crowe, G., and Ali, A. M. (**2021**). “A multimodel deep learning algorithm to detect North Atlantic right whale up-calls”, The Journal of the Acoustical Society of America, **150**(2), 1264-1272.

Janik, V. M. (**2009**). “Acoustic communication in delphinids”, Advances in the Study of Behavior, **40**, 123-157.

Janik, V. M., and Sayigh, L. S. (**2013**). “Communication in bottlenose dolphins: 50 years of signature whistle research”, Journal of Comparative Physiology A, **199**, 479-489.

Kather, V., Seipel, F., Berges, B., Davis, G., Gibson, C., Harvey, M., ... and Risch, D. (**2024**). “Development of a machine learning detector for North Atlantic humpback whale song”, The Journal of the Acoustical Society of America, **155**(3), 2050-2064.

Kecman, V. (**2005**). “Support vector machines–an introduction”, in *Support vector machines: theory and applications* (pp. 1-47). Berlin, Heidelberg: Springer Berlin Heidelberg.

Knight, E., Rhinehart, T., de Zwaan, D. R., Weldy, M. J., Cartwright, M., Hawley, S. H., ... and Kitzes, J. (**2024**). "Individual identification in acoustic recordings", Trends in Ecology & Evolution, **39**(10), 947-960.

Lamoni, L., Garland, E. C., Allen, J. A., Coxon, J., Noad, M. J., and Rendell, L. (**2023**). "Variability in humpback whale songs reveals how individuals can be distinctive when sharing a complex vocal display", The Journal of the Acoustical Society of America, **153**(4), 2238-2238.

Lomax, A., Zollo, A., Capuano, P., and Virieux, J. (**2001**). "Precise, absolute earthquake location under Somma–Vesuvius volcano using a new three-dimensional velocity model", Geophysical Journal International, **146**(2), 313-331.

Lowell, R. P., Zhang, L., Maqueda, M. A. M., Banyte, D., Tong, V. C. H., Johnston, R. E. R., ... and Kolandaivelu, K. (**2020**). "Magma-hydrothermal interactions at the Costa Rica Rift from data collected in 1994 and 2015", Earth and Planetary Science Letters, **531**, 115991.

Lundberg, S. M., and Su-In, L. (**2017**). "A unified approach to interpreting model predictions", Advances in neural information processing systems, **30**, 4765-4774.

McCordic, J. A., Root-Gutteridge, H., Cusano, D. A., Denes, S. L., and Parks, S. E. (**2016**). "Calls of North Atlantic right whales Eubalaena glacialis contain information on individual identity and age class", Endangered Species Research, **30**, 157-169.

McDonald, M. A., Calambokidis, J., Teranishi, A. M., and Hildebrand, J. A. (**2001**). "The acoustic calls of blue whales off California with gender data", The Journal of the Acoustical Society of America, **109**(4), 1728-1735.

Mazhar, S., Ura, T., and Bahl, R. (**2007**). "Vocalization based individual classification of humpback whales using support vector machine", in OCEANS 2007 (pp. 1-9). IEEE.

Meignen, S., Pham, D. H., and McLaughlin, S. (**2017**). "On demodulation, ridge detection, and synchrosqueezing for multicomponent signals", IEEE Transactions on Signal Processing, **65**(8), 2093-2103.

Mellinger, D. K., and Clark, C. W. (**2000**). "Recognizing transient low-frequency whale sounds by spectrogram correlation", The Journal of the Acoustical Society of America, **107**(6), 3518-3529.

Minshull, T. A., Sinha, M. C. and Peirce, C. (**2005**). "Multi-disciplinary subseabed geophysical imaging - a new pool of 28 seafloor instruments in use by the United Kingdom Ocean Bottom Instrument Consortium", Sea Technology, **46**, 27-31.

Molnar, C. (2022). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (2nd ed.).

Oberlin, T., Meignen, S., and Perrier, V. (**2015**). "Second-order synchrosqueezing transform or invertible reassignment? Towards ideal time-frequency representations", IEEE Transactions on Signal Processing, **63**(5), 1335-1344.

Ojala, M., and Garriga, G. C. (**2010**). "Permutation tests for studying classifier performance", Journal of machine learning research, **11**(6), 1833-1863.

Oleson, E. M., Barlow, J., Gordon, J., Rankin, S., and Hildebrand, J. A. (**2003**). "Low frequency calls of Bryde's whales", Mar. Mammal Sci., **19**(2), 407-419.

Oliveira, C., Wahlberg, M., Silva, M. A., Johnson, M., Antunes, R., Wisniewska, D. M., ... and Madsen, P. T. (**2016**). "Sperm whale codas may encode individuality as well as clan identity", The Journal of the Acoustical Society of America, **139**(5), 2860-2869.

Palacios, D.M., Herrera, J.C., Gerrodette, T., Garcia, C., Soler, G.A., Avila, I.C., Bessudo, S., Hernandez, E., Trujillo, F., Forez-Gonzalez, L., and Kerr, I. (**2012**). "Cetacean distribution and relative abundance in Colombia's Pacific EEZ from survey cruises and platforms of opportunity", J. Cetacean Res. Manage. **12**(1), 45-60.

Pham, D. H., and Meignen, S. (**2017**). "High-order synchrosqueezing transform for multicomponent signals analysis—With an application to gravitational-wave signal", IEEE Transactions on Signal Processing, **65**(12), 3168-3178.

Rasmussen, J. H., and Širović, A. (**2021**). "Automatic detection and classification of baleen whale social calls using convolutional neural networks", The Journal of the Acoustical Society of America, **149**(5), 3635-3644.

Rendell, L. E., and Whitehead, H. (**2003**). "Vocal clans in sperm whales (Physeter macrocephalus)", Proceedings of the Royal Society of London. Series B: Biological Sciences, **270**(1512), 225-231.

Robinson, A. H., Zhang, L., Hobbs, R. W., Peirce, C., and Tong, V. C. H. (**2020**). "Magmatic and tectonic segmentation of the intermediate-spreading Costa Rica Rift—a fine balance between magma supply rate, faulting and hydrothermal circulation", Geophysical Journal International, **222**(1), 132-152.

Salmi, R., Hammerschmidt, K., and Doran-Sheehy, D. M. (**2014**). "Individual distinctiveness in call types of wild western female gorillas", Plos One, **9**(7), e101940.

Stanković, L. (**2001**). "A measure of some time–frequency distributions concentration", Signal Processing, **81**(3), 621-631.

Socheleau, F. X., Leroy, E., Carvallo Pecci, A., Samaran, F., Bonnel, J., and Royer, J. Y. (**2015**). "Automated detection of Antarctic blue whale calls", The Journal of the Acoustical Society of America, **138**(5), 3105-3117.

Tary, J. B., Herrera, R. H., Han, J., and van der Baan, M. (**2014**). "Spectral estimation—What is new? What is next?". Reviews of Geophysics, **52**(4), 723-749.

Tary, J. B., Hobbs, R. W., Peirce, C., Lesmes Lesmes, C., and Funnell, M. J. (**2021**). "Local rift and intraplate seismicity reveal shallow crustal fluid-related activity and sub-crustal faulting", Earth and Planetary Science Letters, **562**, 116857.

Tary, J. B., Peirce, C., Hobbs, R. W., Bonilla Walker, F, De La Hoz, C., Bird, A., and Vargas, C. A. (**2024**). "Application of a seismic network to baleen whale call detection and localization in the Panama basin – a Bryde's whale example", Journal of the Acoustical Society of America, **155**(3), 2075-2086.

Thakur, G., and Wu, H. T. (**2011**). "Synchrosqueezing-based recovery of instantaneous frequency from nonuniform samples", SIAM Journal on Mathematical Analysis, **43**(5), 2078-2095.

Van der Maaten, L., and Hinton, G. (**2008**). "Visualizing data using t-SNE", Journal of machine learning research, **9**(11).

Vapnik, V. (**1995**). *The Nature of Statistical Learning Theory*, Springer-Verlag, New York.

Wade, P.R., and Gerrodette, T. (**1993**). "Estimates of cetacean abundance and distribution in the eastern tropical Pacific", Report of the International Whaling Commission **43**, 477-493.

Waldhauser, F., and Ellsworth, W. L. (**2000**). "A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California", Bulletin of the seismological society of America, **90**(6), 1353-1368.

Wessel, P., and Smith, W. H. (**1991**). "Free software helps map and display data", Eos, Transactions American Geophysical Union, **72**(41), 441-446.

White, P. R., and Todd, V. L. (**2024**). "Baleen Whale Vocalizations", in Noisy Oceans: Monitoring Seismic and Acoustic Signals in the Marine Environment, 183-207.

Wilson, D. J., Robinson, A. H., Hobbs, R. W., Peirce, C., and Funnell, M. J. (**2019**). "Does intermediate spreading-rate oceanic crust result from episodic transition between magmatic and magma-dominated, faulting-enhanced spreading?—The Costa Rica Rift example", Geophysical Journal International, **218**(3), 1617-1641.

Zeh, J. M., Perez-Marrufo, V., Adcock, D. L., Jensen, F. H., Knapp, K. J., Robbins, J., ... and Parks, S. E. (**2024**). "Caller identification and characterization of individual humpback whale acoustic behaviour", Royal Society Open Science, **11**(3), 231608.

Zhong, M., Torterotot, M., Branch, T. A., Stafford, K. M., Royer, J. Y., Dodhia, R., and Lavista Ferres, J. (**2021**). "Detecting, classifying, and counting blue whale calls with Siamese neural networks", The Journal of the Acoustical Society of America, **149**(5), 3086-3094.

Zhu, W., Mousavi, S. M., and Beroza, G. C. (**2020**). "Seismic signal augmentation to improve generalization of deep neural networks", in *Advances in geophysics* **61**, 151-177. Elsevier.