Marine Environmental Research 199 (2024) 106571

Contents lists available at ScienceDirect

Marine Environmental Research

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

Machine-learning based detection of marine mammal vocalizations in snapping-shrimp dominated ambient noise

Hari Vishnu [a,*], V.R. Soorya [a], Mandar Chitre [a,b], Yuen Min Too [a], Teong Beng Koay [a], Abel Ho [a]

[a] Acoustic Research Laboratory, 12A Kent Ridge Road, National University of Singapore, 119227, Singapore
[b] Department of Electrical and Computer Engineering, National University of Singapore, Singapore

ARTICLE INFO

Keywords:
Indo-pacific humpback dolphin
Snapping shrimp
Dolphin
Impulsive noise
Singapore
Passive acoustics
Pink dolphin

ABSTRACT

Passive acoustics is an effective method for monitoring marine mammals, facilitating both detection and population estimation. In warm tropical waters, this technique encounters challenges due to the high persistent level of ambient impulsive noise originating from the snapping shrimp present throughout this region. This study presents the development and application of a neural-network based detector for marine-mammal vocalizations in long term acoustic data recorded by us at ten locations in Singapore waters. The detector's performance is observed to be impeded by the high shrimp noise activity. To counteract this, we investigate several techniques to improve detection capabilities in shrimp noise including the use of simple nonlinear denoisers and a machine-learning based denoiser. These are shown to enhance the detection performance significantly. Finally, we discuss some of the vocalizations detected over three years of our acoustic recorder deployments using the robust detectors developed.

1. Introduction

The presence of marine-mammals in a region, including apex predators such as dolphins, is indicative of the health of the marine ecosystem (Sergio et al., 2008; Alves et al., 2023). Our study focuses on the warm coastal waters around Singapore, which have documented sightings of several delphinid species, mostly the Indo-Pacific humpback dolphins (*Sousa chinensis*; IPHD) (Lee and Ooi, 2020; Tay and Ong, 2014) and bottlenose dolphins (*Tursiops aduncus*), but also the Irrawady dolphins (Chua and Lim, 2014), long-beaked common dolphins, false killer whales (Lim and Tay, 1996) and finless porpoises (Ming and Ng, 1990). This region has also documented recent visitations of dugongs (*Dugong dugon*) and a rebounding population of smooth-coated otters (*Lutrogale perspicillata*) (Ng et al., 2022; Khoo and Lee, 2020; Shivram et al., 2023).

The paucity of baseline data on marine-mammal populations due to a lack of research impedes assessment of their status and delays conservation action, particularly in developing countries. Degradation and loss of existing habitat due to human activities including construction, noise pollution, and vessel activity have been identified as growing issues for marine-mammals in some areas (Marsh and Sobtzick; Piwetz et al., 2021). For example, the IPHD, which inhabit tropical waters from the eastern Indian Ocean throughout Southeast Asia to central China (Jefferson and Rosenbaum, 2014), are deemed vulnerable in terms of extinction risk (Jefferson et al., 2017). In recent years, their global population has decreased and is severely fragmented. Dugong populations are considered endangered in many parts of Southeast Asia (Ng et al., 2022; Marsh and Sobtzick), and the populations of dugongs and smooth-coated otters are deemed vulnerable to extinction (Marsh, 2002; Khoo et al., 2021). This accentuates the need to study and estimate the populations of these animals in order to understand ecosystem health and the effect of anthropogenic activities, and inform conservation efforts.

In our study region in Singapore, there are no reliable current estimates of the local population or visitation patterns of most marine-mammal species (Ng et al., 2022; Marsh, 2002). In nearby Malaysia, IPHD numbers have been estimated in various regions (Kuit et al., 2021; Zulkifli Poh et al., 2016; Kamaruzzan and Jaaman, 2013) and they have been documented at various locations along Borneo's coast, including Sarawak, Sabah in Malaysia, Indonesia and Brunei (Minton et al., 2016). Most of the abundance estimates in the Southeast Asia region have been based on photo-identification of individuals with mark-recapture methods (Chan and Karczmarski, 2017; Kuit et al., 2021).

The first step towards estimating the visitation patterns and population density of marine-mammals is detecting them. Passive acoustics is

* Corresponding author.
E-mail address: harivishnu@gmail.com (H. Vishnu).

https://doi.org/10.1016/j.marenvres.2024.106571
Received 18 March 2024; Received in revised form 25 May 2024; Accepted 29 May 2024
Available online 31 May 2024

one of the best-available techniques to detect and monitor marine-mammals, who regularly vocalize during foraging, travelling, or communication. Bottlenose dolphins and IPHDs are known to produce vocalizations classified as whistles, click-trains and burst pulses (Sims et al., 2012; Parijs and Corkeron, 2001). The vocalizations of IPHD and another dolphin species found in neighbouring waters, the Australian humpback dolphin (*Sousa sahulensis*) (Jefferson and Rosenbaum, 2014) have been studied across Asia (Sims et al., 2012; Li et al., 2013; Stead, 2013; Wang et al., 2015; Yuan et al., 2021; Caruso et al., 2020b; Bono et al., 2022) and Australia (Parijs and Corkeron, 2001; Van Parijs and Corkeron, 2001), and their whistle features can vary significantly as seen from the differences between Wang et al. (2013), Van Parijs and Corkeron, 2001 and Seekings et al. (2010). Dugongs, too, produce sounds classified into chirp-squeaks, barks and trills (Ichikawa et al., 2006) which can be used to monitor them acoustically (Tanaka et al., 2021). Otters are also known to produce a wide repertoire of sounds (Lemasson et al., 2014).

Passive acoustic monitoring (PAM) of marine-mammals is well-studied in the literature - for a detailed review, see Fleishman et al. (2023), Au and Lammers (2016), Shiu et al. (2020), Sousa-Lima et al. (2013) and references therein. PAM has several advantages - it is a non-invasive method that can provide wide area-of-coverage monitoring over a long time duration, unhindered by bad visibility, time of the day, cloud cover or rough weather, with good temporal resolution. PAM has been integrated with standard population abundance/density estimation techniques (Hildebrand et al., 2015; Küsel et al., 2016). Sound data is relatively cheap to acquire and store. Furthermore, marine-mammals like dugongs and dolphins are often difficult to spot or monitor visually due to limited surface expression, and the turbidity of Singapore's waters makes underwater visual monitoring infeasible, but they may be detected acoustically over much larger ranges. In this sense, acoustics is a good complement to other population-assesment techniques.

In order to study the marine-mammal populations in the waters around Singapore, we collected 54.2 months of ambient noise recordings from several sites using single-hydrophone recording systems. Interpreting this data effectively involves sifting through large amounts of data to detect the marine-mammal vocalizations, which have large variability within individuals as well as species. Manual inspection of such large amounts of data is time-consuming and infeasible. This necessitates an automated method to detect a large class of marine-mammal vocalizations, for which machine-learning (ML) approaches are a great fit (Fleishman et al., 2023). Amongst ML methods, convolutional neural networks (CNNs), which are shift-invariant, are shown to perform well in detecting classes of patterns in timeseries and images with suitable training, and hence are one of the natural choices for marine-mammal vocalization detection (Shiu et al., 2020).

A large chunk of PAM-based marine-mammal detection studies have focused on waters where the ambient noise has been low enough to allow detection at high signal-to-noise ratios (SNR). The warm tropical Singapore waters offer an additional challenge in this regard in terms of the loud ambient noise levels. At frequencies above 2 kHz, the noise is dominated by snapping shrimp spread across Singapore's waters in large colonies, with the levels sometimes exceeding 70 dB re 1 μPa^2 at 10 kHz (Potter et al., 1997a, 1997b). These shrimp shut their claws, collapsing a bubble and creating a loud impulsive snap sound with a source level of around 187 dB re 1 μPa (Au and Banks, 1998) that can be heard over large distances. Snapping shrimp-generated ambient noise has been studied throughout the world (Au and Banks, 1998; Ferguson and Cleary, 2001; Readhead, 1997; Bohnenstiehl et al., 2016; Yuan et al., 2018), and in regions such as Singapore, it drastically hinders detection of underwater signals (Chitre et al., 2006).

In this work, we develop an ML-based detector to detect marine-mammal vocalizations in long-term sound recordings. In order to surmount the challenge posed by the noise, we highlight the use of existing denoising approaches (Mahmood et al., 2016) to reduce the effect of shrimp snaps on the sound. We further develop an ML-based denoiser to mitigate the effect of shrimp noise and aid detection. This jointly trained denoiser-detector system, named DEVMAN (detector for vocalizations of marine-mammals using neural networks) offers enhanced performance, and is successfully applied to long-term acoustic recordings from Singapore waters to assess the occurrence and diversity of marine-mammal vocalizations. In the future, this will pave the path towards understanding the visitation patterns of local marine-mammals.

Section II of this paper focuses on the recording setup for collecting long-term acoustic data in Singapore waters. Section III focuses on design of an automated ML detector for marine-mammal vocalizations, and denoising techniques to improve their performance. In Section IV we discuss the results obtained using the detector, and we conclude the paper in Section V.

All spectrogram plots shown in this paper are computed using 1024 Fast Fourier transform (FFT) points with 70% overlap, and a Hann window. Each spectrogram is normalized by its 70th percentile value.

2. Experimental setup and data collection

Single-hydrophone stand-alone LS1 model recorders acquired from Loggerhead Instruments(shown in Fig. 2) were deployed across ten island sites (shown in Fig. 1) for several months intermittently between June 2019 to August 2022. Each recorder contains a hydrophone from High-Tech Inc with a sensitivity of −179.90 dB re V/μPa, and was set to a gain of 16.77 dB, sampling frequency of 96 kHz and a bit-depth of 16 bits. The system also recorded the depth, temperature and ambient light. The recorder was deployed at the ocean bottom for approximately 1 month per deployment. In total, 54.2 recorder-months of data were obtained over this period from the island sites, details of which are shown in Table 1 (column 4). The acoustic recordings, of which a sample 1-s timeseries and spectrogram are shown in Fig. 3, show significant snapping shrimp activity. Snaps from a large number of shrimp manifest as randomly occurring impulses in the recordings, and the statistics of this noise cannot be well-modelled by a Gaussian distribution (Bertilone and Killeen, 2001; Chitre et al., 2006).

3. Detector design

3.1. Detector architecture

We assessed two standard CNN-based detector architectures for detection - one based on LeNet (LeCun et al., 1995) and one which is a variant of the Visual-Geometry-group (VGG) architecture (Shiu et al., 2020). LeNet has three convolutional layers, whereas the VGG-based detector has six layers, and both are terminated with two fully-connected layers and an output layer. Our focus is specifically on frequency-modulated (FM) sounds such as whistles and chirp-squeaks, because dolphin clicks are harder to distinguish in snapping shrimp noise (Caruso et al., 2020a). With this in mind and based on the acoustics of local marine-mammals from the literature, we focus on the 2–20 kHz frequency range. The input timeseries is first normalized with respect to its median absolute deviation, so that the detector can work independently of the amplitude scale of the input. A spectrogram of the input timeseries, computed with 2048 FFT points with no overlap, and cropped to lie within the above-mentioned frequency band, is the input to the CNN detectors. The architecture of the VGG-based detector is shown in Fig. 4. The output layer yields a single detector score value which can be thresholded to yield a detection decision on whether the sound clip contains (1) marine-mammal vocalizations, i.e. a 'signal' (embedded in noise), or (0) only ambient noise and sounds that are not of marine mammal origin. Training of the network is done using Tensorflow in Python, and the Adam optimizer is used with a learning rate that decreases with the number of epochs. No pretrained models are used. During training, a sigmoid nonlinearity is added to the output layer to limit its output within the [0, 1] range, and the cost function used for training is the cross-entropy between the detector's output and

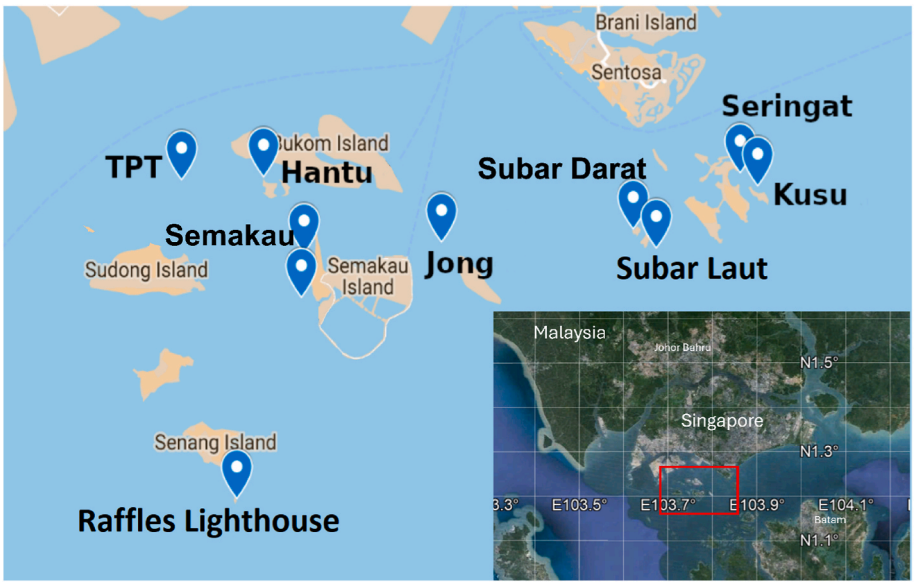

Fig. 1. Map of island sites across Singapore where single-hydrophone passive acoustic recorders were deployed (blue markers), and (inset) zoomed out map of the region around Singapore showing the area studied marked by a red box.

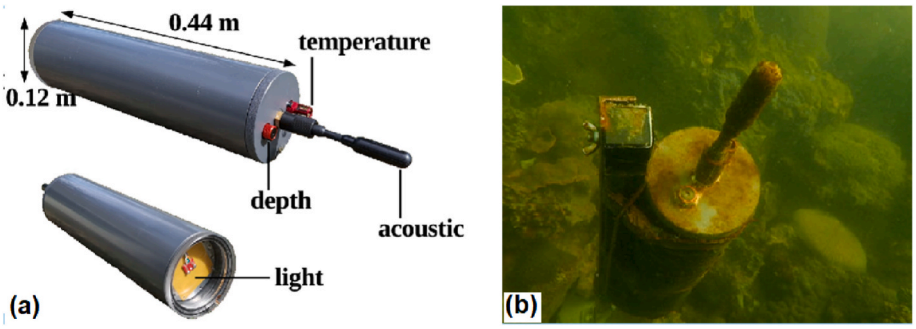

Fig. 2. (a) The recorder used for the study, and (b) the recorder shown deployed underwater.

Table 1
Island-sites at which acoustic data was recorded, with number of vocalizations detected and number of days on which these were detected.

Location	No. of vocalizations detected	No. of days with detections	Months of data recorded
Subar-Darat	860	68	6.4
Subar-Laut	255	25	3.9
Seringat	383	24	6.8
Kusu	783	49	4.6
Jong	182	41	5.2
Hantu	493	55	7.7
Terumbu Pempang Tengah	497	37	5.6
Semakau Northwest (NW)	0	0	4.4
Semakau Southwest (SW)	2	2	4.4
Raffles Lighthouse	65	15	5.2

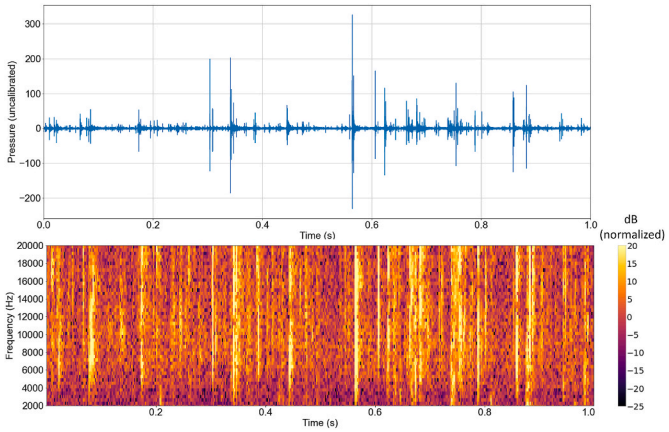

Fig. 3. (a) Sample 1 s recorded timeseries showing snapping shrimp noise background and (b) corresponding spectrogram of the recording showing broadband nature of the snaps.

a binary digit indicating whether a signal was present in the input timeseries or not. We add batch normalization after each convolutional layer, use ReLu nonlinearities, and dropouts during training to improve generalization. Through the course of detector development, it was found that VGG-based detectors yielded better performance than LeNet (as will be presented in the results in Table 2), so further assessment was performed only with the VGG-based detectors. The metric used to evaluate the trained detectors is the recall at a probability of false alarm (PFA) of 1%, following the Neyman-Pearson criteria (Kay, 1998). The intuition here is that operating at a low PFA allows us to keep the volume of data shortlisted by the detector manageably low for further

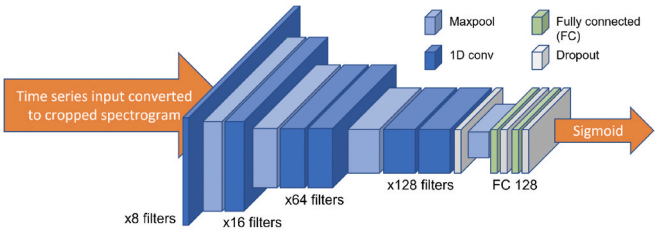

Fig. 4. Architecture of the VGG-based detector.

Table 2
Performance comparison of ML vocalization detector using different denoising approaches on test data.

Detector type	Recall at PFA of 0.01
LeNet detector using clipper denoiser	0.537
VGG detector with no nonlinear preprocessing	0.710
VGG detector using clipper denoiser	0.763
VGG detector using ISC-denoiser	0.772
VGG detector using separately trained ML-based denoiser	0.818
VGG detector trained end-to-end with ML-based denoiser	0.830

downstream tasks, such as abundance estimation.

3.2. Training data and augmentation

The data used for training, validation and testing are generated synthetically by mixing available recordings of marine-mammal vocalizations with that of ambient noise recorded at different locations around Singapore waters. The motivation for this approach is that very few recordings of vocalizations typical of Singapore waters were available beforehand. Hence, the data is compiled by mixing the available vocalization recordings and noise to develop a detector capable of handling a shrimp-dominated noise background representative of local waters.

In the initial stage, the vocalization sounds were obtained (1) from online open repositories (Sayigh et al., 2016; NOAA [National Oceanic and Atmospheric Administration]. NOAA Fisheries and N.N.F.S.C, 2023) containing clips spanning 17 dolphin species and 2 porpoise species, and (2) from previously obtained tank recordings of IPHD vocalizations in Singapore (Seekings et al., 2010). Later, as vocalizations of interest were detected in the initial set of underwater recordings, they were incorporated into the training dataset so that the detector could be retrained and fine-tuned to search for vocalizations from local marine-mammal species. Each such iteration of the detector involved first manually sorting through the clips scored the highest by the detector from amongst a subset of the newly obtained recordings spread across time and geographical locations, and annotating the true positives and false alarms. The top-scored 100 clips per day were considered for the additional annotation. Annotators also listened to the recordings in the vicinity of true positives to pick up possible missed detections. These additional annotated sound clips were added into the existing training/validation dataset. During retraining of the detector, the previously trained weights were used to initialize the neural network so that the training would be faster, i.e., the detector would not need to be trained from scratch, but just fine-tuned. These iterations were done once every 3–4 months when the recorders were retrieved from the recording sites.

The data consists of 1-s segments of "signal + noise" and "noise-only" samples. At the beginning of the detector development, the noise-only sound samples for training, validation and testing, were selected from the only available ambient noise recordings from four locations in local waters, namely, Subar-Darat, Subar-Laut, Raffles Lighthouse, and St. John's island which constituted 8.9 h of data (corresponding to 32,040 noise-only samples). Later on, as more data was collected, noise-only data from the other sites were also incorporated into the training data after manual verification.

To generate the signal + noise training data, we employ a mix-and-match approach in order to efficiently use the available data and obtain maximum diversity in training, which will consequently improve the generalization of the detector. Sound clips containing vocalizations are broken into 1-s segments, and randomly mixed in different ratios with 1-s ambient-noise segments selected from the hours of recording, to generate signal + noise samples of varying SNRs within the [−10, 0] dB range. Different signal segments are randomly mixed with different noise segments in each iteration. As an indicator of the diversity of data - we use about 200 million unique permutations of the signal + noise samples for training. In addition to the above approach, the data is augmented by applying random flipping and shifting operations to the signals to further increase the training data diversity.

The validation and test datasets both consist of ten thousand 1-s sound segment samples each, of which five thousand samples contain noise, and five thousand contain vocalizations at SNRs within [−5, 0] dB synthetically generated by mixing vocalizations with ambient noise. The validation and test datasets are non-overlapping with the training set. The validation data is used to determine the criterion for stopping training to avoid overfitting - the training is stopped when the recall evaluated on the validation set begins to reduce consistently, or is saturated (with an increase of less than 0.001 over two thousand training samples).

3.3. Nonlinearity based denoising

Snapping shrimp noise hinders the detection of vocalizations, as demonstrated visually in the first column of Fig. 5 which are spectrograms of dolphin whistles in snapping shrimp noise. The whistles are corrupted by the broadband snap noise, and in some cases (such as example 1), the whistle is not easy to spot in the noisy spectrogram.

Linear processing with spectral bandpass filtering is not effective in denoising snapping shrimp noise because it spans a very large bandwidth. Some earlier approaches used techniques such as wavelet denoising (Seekings et al., 2010) and spectrogram-denoising (Mallawaarachchi et al., 2008) to tackle shrimp noise. Nonlinear zero-memory nonlinear processors such as sign and stochastic-resonance based detectors (Chitre et al., 2006; Hari et al., 2012) have been shown to improve signal detection in impulsive non-Gaussian noise. We adopt the clipper nonlinearity, a simple and easy-to-implement function parameterized by a threshold T, applied on the time-domain waveform of the recorded data. The output of a clipper for an input time-domain data sample $x(n)$ at sample n is

$$y(n) = \begin{cases} -T & \text{for } x(n) < -T \\ x(n) & \text{for } -T \le x(n) \le T \\ T & \text{for } x(n) > T \end{cases} \quad (1)$$

The clipper output is bounded at large values of the input and linear at small values. Thus, it limits the effect of large outliers. Inappropriate selection of T may affect detection performance, and hence this parameter needs careful selection. Based on an analytical study of the SNR improvement obtained by using a clipper in impulsive noise (Appendix A), an effective threshold for clipping would increase with increase in the noise scale σ. Based on this, the threshold is selected as a multiple of the value σ estimated for each sound clip. It is effective in suppressing the effect of shrimp noise to some degree, as shown in the improved visual quality of spectrograms in Fig. 5 column 2 as compared to column 1.

While the clipper is effective at limiting large outliers, it is a zero-memory nonlinearity which does not take into account the information that shrimp snaps often occur in bursts, each consisting of a bunch of consecutive outlier samples in the timeseries. In other words, the noise is not independent and identically distributed in time, but has some temporal memory. Later work exploited knowledge on the

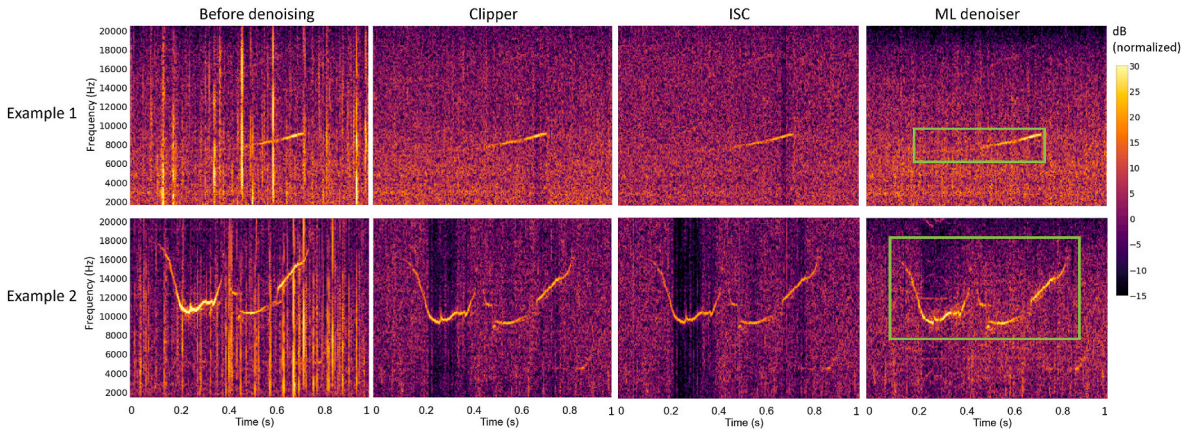

Fig. 5. Examples showing effectiveness of nonlinear denoising techniques on a signal corrupted by shrimp noise: columns show spectrograms of (1) noisy sound segment, and denoised sound segments using (2) clipper, (3) ISC denoiser and (3) ML-based denoiser.

memory of the shrimp noise to design memory-based nonlinear denoisers (Mahmood et al., 2017) of which the Isotropic Sign Correlator (ISC) is the simplest one. Using the ISC on a noisy sound segment also improves the spectrogram quality, as seen in Fig. 5 column 3 in comparison to column 1.

Applying the clipper and ISC as preprocessors before the VGG-based detector yields a significant improvement in the detector's recall (Table 2). Using the clipper with a preset data-adaptive threshold of $T = 5\sigma$ yields a recall of 0.763, an improvement of 7.5% over using no denoising before detection. Furthermore, a comparison between the LeNet and VGG architectures with the clipper showed that the latter performed 42% better. Using the ISC yields a minor improvement over the clipper.

3.4. ML-based denoising

The success of the clipper and ISC indicates that incorporating prior knowledge of the ambient noise can yield detection performance improvement. We argue that a more sophisticated denoiser that effectively incorporates the environment-specific information on noise to remove snaps may yield further improvement. With this in mind, we explore an ML-based denoising approach, inspired by the approach of convolutional denoising autoencoders (Zhou and Yang, 2020; Li et al., 2020). Denoising CNNs that operate directly on time series data (Li et al., 2020) or spectrogram data (Zhou and Yang, 2020; Testolin and Diamant, 2020) have been explored. Since shrimp impulses are temporally concentrated and hence easier to isolate in the time-domain, in this work, a one-dimensional (1D) CNN denoiser is used, which denoises the noisy timeseries data provided as input.

The architecture of the denoiser used is described in Fig. 6(a). It consists of five 1D convolutional layers with a bottleneck in the middle. The denoiser takes a 1 s sample of input audio signal at a time. The numbers within the blocks in Fig. 6(a) indicate the size of each convolutional layer, starting from 96000 samples corresponding to 1 s of input. The kernel size used in each layer is seven and stride is four, leading to gradual reduction in each layer's size as one moves from the input towards the bottleneck while there is an increase in the depth of each layer. ReLu nonlinearities are used after each layer. After the bottleneck, there is a layer-by-layer increase in layer size and a corresponding decrease in depth, until at the output there is a denoised 1-s timeseries. Fig. 6(b) illustrates how training data is generated for the denoiser. A clean 1-s signal $s(n)$ (chosen from the available library of high SNR sounds) is used at the input to generate a clean signal spectrogram $S(f, n)$. It is then mixed with an ambient noise timeseries $w(n)$ to generate a noisy timeseries, which is the input to the denoiser. For each training sample, the SNR of the noisy input segment is randomly chosen

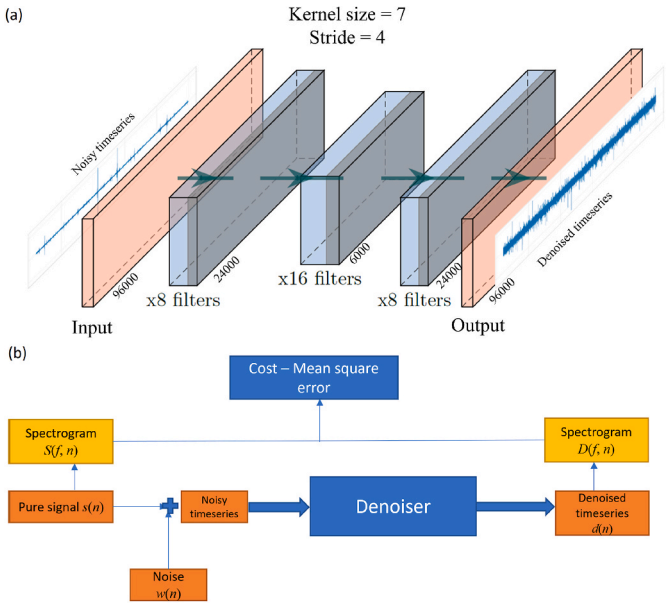

Fig. 6. (a) Architecture of ML-based denoiser with 5 convolutional layers with a bottleneck in the middle. The numbers in front of the blocks indicate the convolutional layer sizes. (b) Block diagram illustrating how training data is generated for the denoiser. A clean 1 s signal of high SNR is used at the input to generate a clean signal spectrogram. It is then mixed with an ambient noise clip to generate a noisy timeseries, which is the input to the denoiser. The cost function is the mean squared error between the spectrograms of the denoiser's output and the clean signal.

from the interval $[-10, 5]$ dB. Furthermore, the data is augmented by applying flipping and random shifting to signals. The cost function is the mean squared error between the spectrograms of the denoiser's output and the clean signal. The intuition behind this is that the denoiser's designed objective is to generate a denoised timeseries with as clean a spectrogram as possible, which is expected to improve the performance of the detection downstream.

As in the case of the detector, the validation dataset is used here to determine the stopping criterion for training. The training is stopped when the cost function evaluated on the validation dataset begins to increase, or is determined to be saturated. The trained denoiser is found to be effective in generating denoised signals with clean spectrograms as shown in Fig. 5, column 4.

The trained ML-based denoiser can be used as a preprocessor to denoise sound segments before the VGG-based detector is applied to

them. We can go one step further in this approach, by training the combined denoiser-detector system end-to-end to maximize detection performance on the validation dataset. This joint system approach, referred to as DEVMAN henceforth, fine-tunes the denoiser and detector components to work better together, and ensures that the denoiser is tuned more towards improving signal quality in a way that aids detection, rather than the quality of the spectrogram alone. Both these approaches are evaluated with the test dataset for detection.

4. Results and discussion

When the ML-denoiser is applied to the sound segments as a preprocessor before applying the VGG-based detector, its performance is 6% better than the clipper and ISC (Table 2). The end-to-end trained DEVMAN system performs better than all other detectors considered, yielding a recall of 0.83. This advantage is owing to the fact that the ML-based denoiser of this system is tuned to the specific characteristics of shrimp noise and working towards improving detector performance rather than just spectrogram quality or SNR.

The 17% of missed detections from the test dataset were analyzed to assess the reason they were missed by the detector. In most cases, these samples were found to be of very low-SNR and too challenging for the detectors to pick up, examples of which are shown in Fig. 7.

The DEVMAN detector was applied to the long-term acoustic recordings obtained from the 10 island sites. The detected sounds were manually examined by experts to weed out the false alarms, and trimmed down to a smaller subset of valid vocalization sounds. In total, about 3520 detections of biological origin were picked up by the system over 54 recorder-months (the breakup for individual sites is shown in Table 1 column 2). Marine-mammal vocalizations were detected in the frequency band of interest on 316 days added across all locations. An additional limited evaluation of DEVMAN's performance was conducted on the datasets obtained from the recorders. This was done by picking nine days of deployments spread over four different geographic locations (Subar-Darat, Seringat, Subar-Laut and Hantu), annotating this data fully, and comparing it against the detections by DEVMAN to evaluate its performance. The evaluation showed that the detector yielded a recall of 0.892 at a PFA of 1%, which is higher than the recall observed on the test dataset.

While there is very little published literature on the distribution of IPHD and other dolphins in Singapore waters to compare these detections against (except for Lee and Ooi, 2020 at Jurong Island, where we did not place recorders, and Tay and Ong, 2014 at Seringat), some of the detections picked up by the recorders roughly correspond to locations where dolphins and other marine mammals have been reported in the anecdotal literature and news media previously, eg. near Seringat (Ee, 2024; Straits Times, 2024) where we detected vocalizations on 24 days. Furthermore, smooth-coated otters have been spotted in many parts of Singapore in recent years (Khoo and Lee, 2020; Shivram et al., 2023)

We categorize the detected vocalizations into 6 classes labelled a to f, whose spectrograms are shown in Fig. 8. Sound class (a) is a family of FM sounds with centre frequencies between 4 and 7 kHz. Class (b) encompasses high-FM sounds with centre frequencies of at least 18 kHz, and usually heavily modulated with bandwidths of up to 3 kHz. Class (c) consists of FM sounds with a fundamental frequency usually around 2–3 kHz, and several (about 8–15) harmonics, which is similar to the sound of small-clawed otters found around Asia (Lemasson et al., 2014). Class (d) and (e) are both series of broadband transients with a constant repetition rate. The timeseries of class (e) reveals a set of closely-spaced broadband impulses with a nearly constant inter-impulse interval of about 2.5–4 ms which exhibit numerous harmonics in the spectral content spanning between 1 and 20 kHz, closely matching descriptions of dolphin burst click pulses (also referred to as barks or buzzes) as reported in the literature (Van Parijs and Corkeron, 2001; Sims et al., 2012; Wang et al., 2015). Barks are reported mostly during IPHD socializing and foraging. Class (f) is a broadband dolphin click-train with an inter-click-interval of between 2.5 and 4 ms, as revealed in the timeseries plotted in Fig. 8(h). Low-FM sounds of class type (a) were also sometimes observed along with these click-trains.

Fig. 9 shows 6 example types of FM sounds picked up by the DEVMAN detector over our deployment period, which form a subset of the sounds described in Fig. 8 (a). The spectrograms are computed after denoising the sounds. Fig. 9(a) describes FM sounds with an ascending span over 1 kHz, followed by a relatively flat zone and then a descending span, similar to the IPHD whistle class (d) described by Wang et al. (2013). Fig. 9(b) encompasses relatively flat sounds with less than 1 kHz fluctuation, sometimes with an upward inflection at the end. These are comparable to the "flat"-type IPHD whistles described by Wang et al. (2013). Fig. 9(c) covers FMs with a slight downward slope with the frequency falling by about 1 kHz over the duration of the signal, comparable to class (b) described by Wang et al. (2013). Fig. 9(d) is similar to (c) but with a short upward and then downward inflection at the end. Fig. 9(e) consists of FMs with a short fast upward rise in frequency, followed by a zone of slow upward slope, comparable to class (f) in Wang et al. (2013) and some of the classes described by Sims et al. (2012), but with lower frequency variation. Fig. 9(f) are "U-shape" FMs followed by a zone of downward slope, also exhibiting similarities to some classes in Wang et al. (2013) and Sims et al. (2012).

Most of the vocalizations were detected in the first half of each year (January to May) as compared to the later parts, and the detections were mostly picked up in the Sisters island area around Subar-Darat (50%), Kusu (20%) and Subar-Laut (16%). Relatively fewer detections were picked up at Hantu (30%), TPT (30.3%) and Jong (11.1%), whereas there were only two detections picked up near Semakau island, where an anchorage is located.

5. Conclusion

ML-based detectors facilitate effective monitoring of marine-mammals via passive acoustics. Snapping-shrimp generated noise in tropical waters such as those in Singapore impedes the detection of

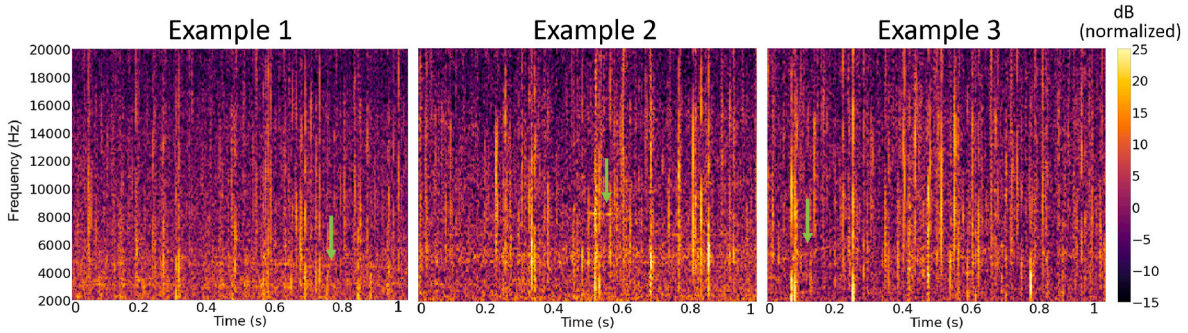

Fig. 7. Spectrograms of test samples missed by the VGG-detector, which are of very low SNR.

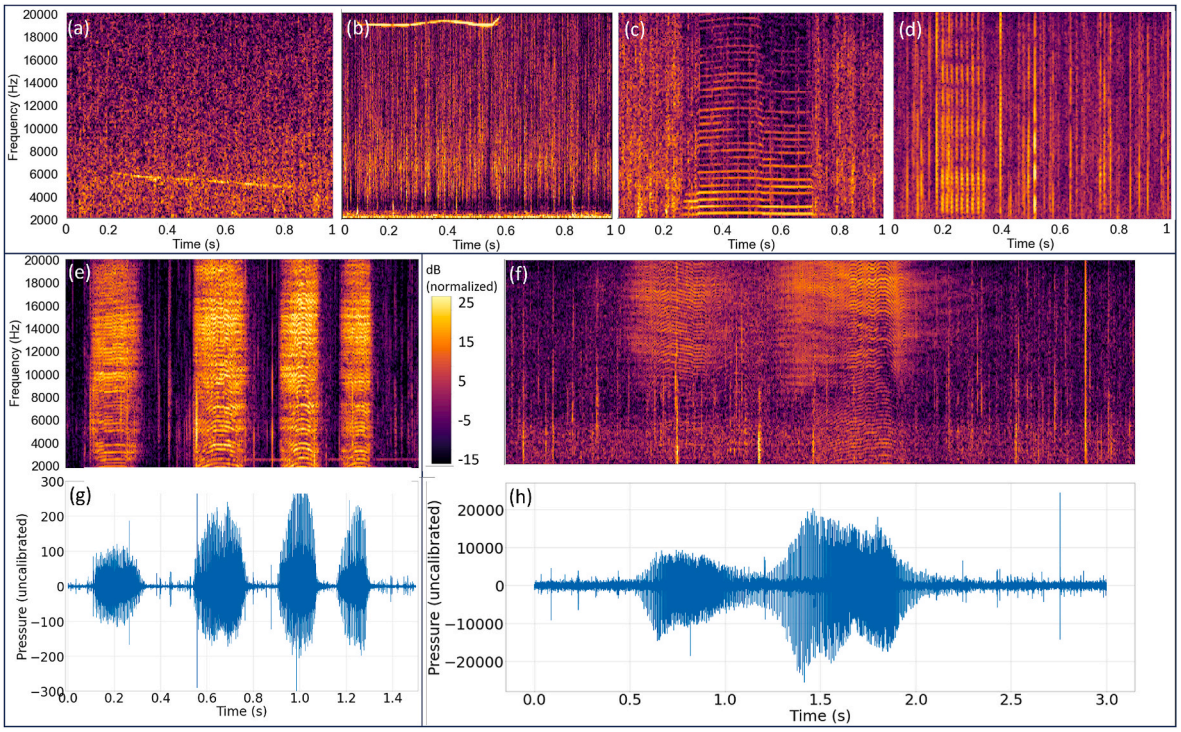

Fig. 8. (a)–(f) Spectrograms of 6 classes of sounds detected, and (g) and (h) sound timeseries corresponding to classes (e) and (f).

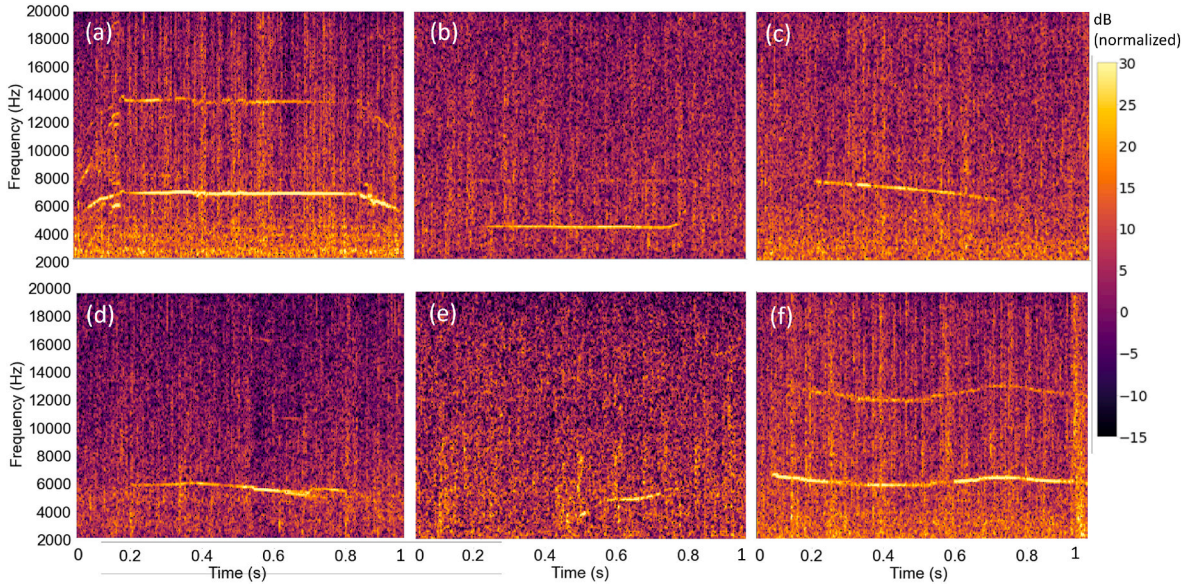

Fig. 9. Different types of FM sounds picked up by the detector.

marine-mammal vocalizations. We have explored three approaches to mitigate this noise: (1) simple, fixed nonlinear denoising methods, (2) ML-based denoising trained independently from the detector, and (3) ML-based denoiser trained end-to-end with the detector (DEVMAN). The DEVMAN yielded the best performance amongst the approaches explored, and can facilitate downstream tasks for assessing behavioral patterns or population density. Some ongoing efforts are leveraging DEVMAN for assessing the presence of multiple vocalizing individuals using a recently deployed hydrophone-array system (Hexeberg et al., 2023), and integrating DEVMAN into stand-alone PAM systems as an edge-detector for real-time monitoring.

Furthermore, it was illustrated that employing simple preprocessor nonlinearities such as the clipper and ISC also yield a significant performance benefit in comparison to no denoising. These provide a good low-complexity alternative when the computational complexity of the ML-based denoiser may be a constraint, such as in low-cost edge computation devices. A comparative analysis of two architectures - VGG and LeNet - highlighted the superior performance of the VGG-based architecture in terms of detections in local waters.

Future work in this direction will pivot towards the development of classifiers for differentiating the vocalizations picked by the detector, a comprehensive evaluation of the detector across different marine mammal vocalization classes, and confirmation of the identities of the species corresponding to each of the sound classes. Furthermore, an in-

depth spatio-temporal analysis of the detections is planned. This analysis can uncover potential temporal patterns (eg. diel/seasonal/annual, and correlations of these with known environmental events, for example). It can also identify spatial patterns and possible correlations with environmental variates, distributions of known local species and anthropogenic and biological ambient noise. This can help elucidate the behavioral factors leading to these variations. The work in this paper lays the groundwork for such studies in the future, which will help us better understand the data in the context of the regional biodiversity, and provide important information for policy makers in decision making and for aiding conservation efforts.

CRediT authorship contribution statement

Hari Vishnu: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **V.R. Soorya:** Software, Methodology, Investigation, Formal analysis, Data curation. **Mandar Chitre:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Yuen Min Too:** Software, Resources, Investigation, Data curation. **Teong Beng Koay:** Validation, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Abel Ho:** Software, Resources, Methodology, Data curation.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

Acknowledgements

We acknowledge funding support from the National Parks Board, Singapore, and from the National Research Foundation, Prime Minister's Office, Singapore under its Marine Science Research and Development Program MSRDP P-20. This support was used for collection of the data, and to fund manpower for its analysis and interpretation.

Appendix A

Snapping shrimp noise cannot be adequately modelled by a Gaussian distribution, and heavy-tailed distributions such as the symmetric α-stable family of distributions have been shown to better model this noise instead (Chitre et al., 2006). This family is parameterized by an impulsiveness parameter α and a scale parameter σ, and the probability density functions (pdf) do not have a closed form expression except for two cases, one of which is the Cauchy distribution (when $\alpha = 1$). The Cauchy distribution represents highly impulsive data, so a clipper robust in Cauchy-distributed ambient noise may be expected to be robust in snapping-shrimp dominated environments. Here, we use the example of Cauchy-distributed ambient noise to highlight the dependence of an effective value of the clip nonlinearity's threshold on the scale σ of the noise.

Consider the clipper function defined in (1). Assume the nth sample of the noisy signal $x(n)$ consists of the clean signal $s(n)$ embedded in ambient noise $w(n)$, described as

$$x(n) = s(n) + w(n) \tag{A.1}$$

where $w(n)$ is distributed with pdf f_σ and distribution function F_σ.

Based on the insights from the signal detection literature (Kay, 1998; Hari et al., 2012), we can obtain intuition on the suitable threshold T to be used by considering the optimal threshold T' that maximizes the SNR of the clipper output $y(n)$, given by

$$S_{y(n)} = \frac{(E[y(n)])^2}{E[y^2(n)] - (E[y(n)])^2} \tag{A.2}$$

where $E[.]$ denotes the expectation operator. In (A.2), the denominator signifies the variance of $y(n)$. From (1), the pdf of $y(n)$ can be obtained as

$$\widetilde{f_y} = \begin{cases} F_\sigma(-T - s(n)), & y(n) = -T \\ f_\sigma(y(n)), & -T < y(n) < T \\ F_\sigma(T - s(n)), & y(n) = T \end{cases} \tag{A.3}$$

For the case when $w(n)$ is Cauchy distributed, the expressions for f_σ and F_σ are known in closed form, hence the expression (A.2) is tractable in closed form and can be used to analytically obtain T'. It can be shown that for a weak signal ($s(n) \to 0$), T' is equal to the value that maximizes a cost function given by

$$T' = \mathtt{argmax}_T \left[\frac{\tan^{-1}\left(\frac{T}{\sigma}\right)^2}{\frac{T}{\sigma} + \frac{0.5\pi T^2}{\sigma^2} - \tan^{-1}\left(\frac{T}{\sigma}\right)\left(1 + \frac{T^2}{\sigma^2}\right)} \right]. \tag{A.4}$$

By finding the maximum of the expression in (A.4), we obtain $T' = 0.4\sigma$ for Cauchy-distributed noise. Note that this is independent of the signal amplitude, but depends on σ. Though T' is not necessarily equal to the optimal threshold that maximizes the detection of vocalizations in spectrograms, it provides a rule of thumb that the selection of the threshold must scale with the scale parameter σ of the noise. In practise, it is time-consuming to estimate σ for a large number of sound segments during training and inference, so we use a robust proxy of the scale in impulsive noise which is the median absolute deviation of the timeseries (Pham-Gia and Hung, 2001).

References

Alves, F., Queiroz, N., Jodice, P.G., 2023. Editorial: ecological and behavioral traits of apex predators in oceanic insular ecosystems: advances and challenges in research and conservation. Front. Mar. Sci. 10, 1–3. https://doi.org/10.3389/fmars.2023.1252360.

Au, W.W.L., Banks, K., 1998. The acoustics of the snapping shrimp Synalpheus parneomeris in Kaneohe Bay. J. Acoust. Soc. Am. 103, 41. https://doi.org/10.1121/1.423234.

Au, W.W.L., Lammers, M.O. (Eds.), 2016. Listening in the Ocean. Modern Acoustics and Signal Processing. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4939-3176-7. URL: http://link.springer.com/10.1007/978-1-4939-3176-7.

Bertilone, D., Killeen, D., 2001. Statistics of biological noise and performance of generalized energy detectors for passive detection. IEEE J. Ocean. Eng. 26, 285–294. https://doi.org/10.1109/48.922794.

Bohnenstiehl, D.R., Lillis, A., Eggleston, D.B., 2016. The curious acoustic behavior of estuarine snapping shrimp: temporal patterns of snapping shrimp sound in sub-tidal oyster reef habitat. PLoS One 11, 1–21. https://doi.org/10.1371/journal.pone.0143691.

Bono, S., Kimura, S.S., Teoh, Z.Y., Ng, j.E., Ichikawa, K., Ponnampalam, L.S., 2022. Whistle variation in indo-pacific humpback dolphins (Sousa chinensis) in relation to behavioural and environmental parameters in northwestern peninsular Malaysia. Acoust Aust. 50, 315–329. https://doi.org/10.1007/s40857-022-00273-6 doi: 10.1007/s40857-022-00273-6.

Caruso, F., Dong, L., Lin, M., Liu, M., Gong, Z., Xu, W., Alonge, G., Li, S., 2020a. Monitoring of a nearshore small dolphin species using passive acoustic platforms and supervised machine learning techniques. Front. Mar. Sci. 7 https://doi.org/10.3389/fmars.2020.00267.

Caruso, F., Dong, L., Lin, M., Liu, M., Xu, W., Li, S., 2020b. Influence of acoustic habitat variation on Indo-Pacific humpback dolphin (Sousa chinensis) in shallow waters of Hainan Island, China. J. Acoust. Soc. Am. 147, 3871–3882. https://doi.org/10.1121/10.0001384.

Chan, S.C.Y., Karczmarski, L., 2017. Indo-Pacific humpback dolphins (Sousa chinensis) in Hong Kong: modelling demographic parameters with mark-recapture techniques. PLoS One 12, e0174029. https://doi.org/10.1371/journal.pone.0174029. URL: https://dx.plos.org/10.1371/journal.pone.0174029.

Chitre, M.A., Potter, J., Ong, S.H., 2006. Optimal and near-optimal signal detection in snapping shrimp dominated ambient noise. IEEE J. Ocean. Eng. 31, 497–503. https://doi.org/10.1109/JOE.2006.875272.

Chua, M., Lim, K.K.P., 2014. Irrawaddy dolphin carcass at East Coast beach, Orcaella brevirostris. Singapore Biodiversity Records 201–202.

Ee, D., 2024. Dolphins frolicking in Singapore's backyard. URL: https://www.straitstimes.com/singapore/environment/dolphins-frolicking-in-singapores-backyard.

Ferguson, B.G., Cleary, J.L., 2001. In situ source level and source position estimates of biological transient signals produced by snapping shrimp in an underwater environment. J. Acoust. Soc. Am. 109, 3031. https://doi.org/10.1121/1.1339823. URL: http://link.aip.org/link/JASMAN/v109/i6/p3031/s1&Agg=doi.

Fleishman, E., Cholewiak, D., Gillespie, D., Helble, T., Klinck, H., Nosal, E., Roch, M.A., 2023. Ecological inferences about marine mammals from passive acoustic data. Biol. Rev. 98, 1633–1647. https://doi.org/10.1111/brv.12969. URL: https://onlinelibrary.wiley.com/doi/10.1111/brv.12969.

Hari, V., Anand, G., Premkumar, A., Madhukumar, A., 2012. Design and performance analysis of a signal detector based on suprathreshold stochastic resonance. Signal Process. 92, 1745–1757. https://doi.org/10.1016/j.sigpro.2012.01.013. URL: http://linkinghub.elsevier.com/retrieve/pii/S0165168412000278.

Hexeberg, S., Vishnu, H., Beng, K.T., Ho, A., Yusong, W., Chitre, M., Tun, K., Lim, K., 2023. Acoustic detector for multiple vocalizing marine mammal individuals. In: OCEANS 2023 - Limerick. IEEE, pp. 1–8. URL: https://ieeexplore.ieee.org/document/10244432/. doi:10.1109/OCEANSLimerick52467.2023.10244432.

High-Tech Inc, 2022. HTI-96-Min. URL: http://www.hightechincusa.com/products/hydrophones/hti96min.html.

Hildebrand, J.A., Baumann-Pickering, S., Frasier, K.E., Trickey, J.S., Merkens, K.P., Wiggins, S.M., McDonald, M.A., Garrison, L.P., Harris, D., Marques, T.A., Thomas, L., 2015. Passive acoustic monitoring of beaked whale densities in the Gulf of Mexico. Sci. Rep. 5, 1–15. https://doi.org/10.1038/srep16343.

Ichikawa, K., Tsutsumi, C., Arai, N., Akamatsu, T., Shinke, T., Hara, T., Adulyanukosol, K., 2006. Dugong (Dugong dugon) vocalization patterns recorded by automatic underwater sound monitoring systems. J. Acoust. Soc. Am. 119, 3726–3733. https://doi.org/10.1121/1.2201468.

Jefferson, T., Smith, B., Braulik, G., Perrin, W., 2017. Sousa chinensis (errata version published in 2018). The IUCN Red List of Threatened Species 2017: e. T82031425A123794774. https://doi.org/10.2305/IUCN.UK.2017-3.RLTS.T82031425A50372332.en.

Jefferson, T.A., Rosenbaum, H.C., 2014. Taxonomic revision of the humpback dolphins (Sousa spp.), and description of a new species from Australia. Mar. Mamm. Sci. 30, 1494–1541. https://doi.org/10.1111/mms.12152. URL: https://onlinelibrary.wiley.com/doi/10.1111/mms.12152.

Kamaruzzan, A.S., Jaaman, S.A., 2013. Interactions between indo-pacific humpback and irrawaddy dolphins in Cowie Bay, Sabah, Malaysia. Malay. Nat. J. 64, 185–191.

Kay, S.M., 1998. Fundamentals of Statistical Signal Processing, Vol.II: Detection Theory. Prentice-Hall, Upper Saddle River, New Jersey.

Khoo, M., Basak, S., Sivasothi, N., de Silva, P., Reza Lubis, I., 2021. Lutrogale Perspicillata. The IUCN Red List of Threatened Species.

Khoo, M.D., Lee, B.P., 2020. The urban Smooth-coated otters Lutrogale perspicillata of Singapore: a review of the reasons for success. Int. Zoo Yearbk. 54, 60–71. https://doi.org/10.1111/izy.12262.

Kuit, S.H., Ponnampalam, L.S., Hammond, P.S., Chong, V.C., Then, A.Y., 2021. Abundance estimates of three cetacean species in the coastal waters of Matang, Perak, Peninsular Malaysia. Aquat. Conserv. Mar. Freshw. Ecosyst. 31, 3120–3132. https://doi.org/10.1002/aqc.3699. URL: https://onlinelibrary.wiley.com/doi/10.1002/aqc.3699.

Küsel, E.T., Siderius, M., Mellinger, D.K., 2016. Single-sensor, cue-counting population density estimation: average probability of detection of broadband clicks. J. Acoust. Soc. Am. 140, 1894–1903. https://doi.org/10.1121/1.4962753. URL: https://pubs.aip.org/jasa/article/140/3/1894/649573/Single-sensor-cue-counting-population-density.

LeCun, Y., Bengio, Y., et al., 1995. Convolutional networks for images, speech, and time series. The Handbook of Brain Theory and Neural Networks 3361, 1995.

Lee, B.P.Y.H., Ooi, M., 2020. Indo-Pacific humpback dolphins near Jurong Island. Singapore Biodiversity Records 84–85.

Lemasson, A., Mikus, M.A., Blois-Heulin, C., Lodé, T., 2014. Vocal repertoire, individual acoustic distinctiveness, and social networks in a group of captive Asian small-clawed otters (Aonyx cinerea). J. Mammal. 95, 128–139. https://doi.org/10.1644/12-MAMM-A-313.1.

Li, S., Wang, D., Wang, K., Hoffmann-Kuhnt, M., Fernando, N., Taylor, E.A., Lin, W., Chen, J., Ng, T., 2013. Possible age-related hearing loss (presbycusis) and corresponding change in echolocation parameters in a stranded Indo-Pacific humpback dolphin. J. Exp. Biol. 216, 4144–4153. https://doi.org/10.1242/jeb.091504.

Li, Y., Wang, B., Shao, G., Shao, S., Pei, X., 2020. Blind detection of underwater acoustic communication signals based on deep learning. IEEE Access 8, 204114–204131. https://doi.org/10.1109/ACCESS.2020.3036883.

Lim, K.K.P., Tay, J.B., 1996. Nature society (Singapore). The Pangolin 7, 33–34.

Loggerhead Instruments, . Underwater Acoustic Recorders: LS1 and LS2 Family of Recorders. URL: https://www.loggerhead.com/ls1-ls2-recorders.

Mahmood, A., Chitre, M., Vishnu, H., 2017. Locally optimal inspired detection in snapping shrimp noise. IEEE J. Ocean. Eng. 42 https://doi.org/10.1109/JOE.2017.2731058.

Mahmood, A., Vishnu, H., Chitre, M., 2016. Model-based signal detection in snapping shrimp noise. In: 3rd Underwater Communications and Networking Conference, Ucomms 2016. https://doi.org/10.1109/UComms.2016.7583459. Lerici, Italy.

Mallawaarachchi, A., Ong, S.H., Chitre, M., Taylor, E., 2008. Spectrogram denoising and automated extraction of the fundamental frequency variation of dolphin whistles. J. Acoust. Soc. Am. 124, 1159–1170. https://doi.org/10.1121/1.2945711. URL: http://asa.scitation.org/doi/10.1121/1.2945711.

Marsh, H., 2002. United Nations Environment Programme Eds: Dugong: Status Report and Action PLANS for Countries and Territories. United Nations Environment Programme, Nairobi. Technical Report.

Marsh, H., Sobtzick, S., . Dugong dugon (amended version of 2015 assessment). URL: https://www.iucnredlist.org/species/6909/160756767, doi:https://dx.doi.org/10.2305/IUCN.UK.2015-4.RLTS.T6909A160756767.en.

Ming, C.L., Ng, P.K., 1990. Essays in Zoology : Papers Commemorating the 40th Anniversary of the Department of Zoology. National University of Singapore, Singapore. Department of Zoology, National University of Singapore.

Minton, G., Zulkifli Poh, A.N., Peter, C., Porter, L., Kreb, D., 2016. Indo-Pacific Humpback Dolphins in Borneo, pp. 141–156. https://doi.org/10.1016/bs.amb.2015.07.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0065288115000085.

Ng, Sirius, Z.H., Ow, Yan, Xiang, Jaafar, Zeehan, 2022. Dugongs (Dugong dugon) along hyper-urbanized coastlines. Front. Mar. Sci 9, 1–12. https://doi.org/10.3389/fmars.2022.947700.

NOAA [National Oceanic and Atmospheric Administration]. NOAA Fisheries, N.N.F.S.C, 2023. Passive acoustic research group. URL: https://www.fisheries.noaa.gov/national/science-data/sounds-ocean-mammals.

Parijs, S.M., Corkeron, P.J., 2001. Evidence for signature whistle production by a pacific humpback dolphin, Sousa chinesis. Mar. Mamm. Sci. 17, 944–949. https://doi.org/10.1111/j.1748-7692.2001.tb01308.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1748-7692.2001.tb01308.x.

Pham-Gia, T., Hung, T.L., 2001. The mean and median absolute deviations. Math. Comput. Model. 34, 921–936. https://doi.org/10.1016/S0895-7177(01)00109-1.

Piwetz, S., Jefferson, T.A., Würsig, B., 2021. Effects of coastal construction on indo-pacific humpback dolphin (Sousa chinensis) behavior and habitat-use off Hong Kong. Front. Mar. Sci. 8 https://doi.org/10.3389/fmars.2021.572535. URL: https://www.frontiersin.org/articles/10.3389/fmars.2021.572535/full.

Potter, J., Lim, T., Chitre, M., 1997a. Ambient noise environments in shallow tropical seas and the implications for acoustic sensing. Oceanol. Int.

Potter, J.R., Wei, T., Chitre, M., 1997b. Acoustic imaging & the natural soundscape in Singapore waters. In: Proceedings of Mindef-NUS Joint Seminar, 1997, Singapore.

Readhead, M.L., 1997. Snapping shrimp noise near Gladstone, Queensland. J. Acoust. Soc. Am. 101, 1718. https://doi.org/10.1121/1.418153. URL: http://link.aip.org/link/?JAS/101/1718/1&Agg=doi.

Sayigh, L., Daher, M.A., Allen, J., Gordon, H., Joyce, K., Stuhlmann, C., Tyack, P., 2016. The watkins marine mammal sound database: an online, freely accessible resource. URL: https://pubs.aip.org/asa/poma/article/836962.

Seekings, J., Yeo, K.P., Chen, Z.P., Nanayakkara, S.C., Tan, J., Tay, P., Taylor, E., 2010. Classification of a large collection of whistles from Indo-Pacific humpback dolphins (Sousa chinensis). OCEANS'10 IEEE Sydney, OCEANSSYD 2010, 3–7. https://doi.org/10.1109/OCEANSSYD.2010.5603596.

Sergio, F., Caro, T., Brown, D., Clucas, B., Hunter, J., Ketchum, J., McHugh, K., Hiraldo, F., 2008. Top predators as conservation tools: ecological rationale, assumptions, and efficacy. Annu. Rev. Ecol. Evol. Systemat. 39, 1–19. https://doi.org/10.1146/annurev.ecolsys.39.110707.173545.

Shiu, Y., Palmer, K.J., Roch, M.A., Fleishman, E., Liu, X., Nosal, E.M., Helble, T., Cholewiak, D., Gillespie, D., Klinck, H., 2020. Deep neural networks for automated detection of marine mammal species. Sci. Rep. 10, 1–12. https://doi.org/10.1038/s41598-020-57549-y.

Shivram, A., Sivasothi, N., Hsu, C.D., Hodges, K.E., 2023. Population distribution and causes of mortality of smooth-coated otters, Lutrogale perspicillata, in Singapore. J. Mammal. 104, 496–508. https://doi.org/10.1093/jmammal/gyad007.

Sims, P.Q., Vaughn, R., Hung, S.K., Würsig, B., 2012. Sounds of Indo-Pacific humpback dolphins (Sousa chinensis) in West Hong Kong: a preliminary description. J. Acoust. Soc. Am. 131, EL48–EL53. https://doi.org/10.1121/1.3663281. URL: http://asa.scitation.org/doi/10.1121/1.3663281.

Sousa-Lima, R.S., Fernandes, D.P., Norris, T.F., Oswald, J.N., 2013. A review and inventory of fixed autonomous recorders for passive acoustic monitoring of marine mammals: 2013 state-of-the-industry. In: 2013 IEEE/OES Acoustics in Underwater Geosciences Symposium, RIO Acoustics 2013. IEEE. https://doi.org/10.1109/RIOAcoustics.2013.6683984.

Stead, N., 2013. Pardon me? Hearing loss in wild dolphins. J. Exp. Biol. 216 https://doi.org/10.1242/jeb.096933 iii–iii.

Straits Times, 2024. Don't be surprised if you see a dolphin swimming in Singapore's waters. URL: https://www.straitstimes.com/singapore/environment/dont-be-surprised-if-you-see-a-dolphin-swimming-in-singapores-waters.

Tanaka, K., Ichikawa, K., Kittiwattanawong, K., Arai, N., Mitamura, H., 2021. Automated classification of dugong calls and tonal noise by combining contour and MFCC features. Acoust Aust. 49, 385–394. https://doi.org/10.1007/s40857-021-00234-5 doi:10.1007/s40857-021-00234-5.

Tay, T.S., Ong, R., 2014. Indo-Pacific Humpback Dolphins off Lazarus Island. Singapore Biodiversity Records, p. 275.

Testolin, A., Diamant, R., 2020. Combining denoising autoencoders and dynamic programming for acoustic detection and tracking of underwater moving targets. Sensors 20. https://doi.org/10.3390/s20102945.

Van Parijs, S.M., Corkeron, P.J., 2001. Vocalizations and behaviour of pacific humpback dolphins Sousa chinensis. Ethology 107, 701–716. https://doi.org/10.1046/j.1439-0310.2001.00714.x. URL:

Wang, Z., Fang, L., Shi, W., Wang, K., Wang, D., 2013. Whistle characteristics of free-ranging Indo-Pacific humpback dolphins (Sousa chinensis) in Sanniang Bay, China. J. Acoust. Soc. Am. 133, 2479–2489. https://doi.org/10.1121/1.4794390. URL: https://pubs.aip.org/jasa/article/133/4/2479/917190/Whistle-characteristics-of-free-ranging-Indo.

Wang, Z.T., Nachtigall, P.E., Akamatsu, T., Wang, K.X., Wu, Y.P., Liu, J.C., Duan, G.Q., Cao, H.J., Wang, D., 2015. Passive acoustic monitoring the diel, lunar, seasonal and tidal patterns in the biosonar activity of the Indo-Pacific humpback dolphins (Sousa chinensis) in the Pearl River Estuary, China. PLoS One 10, 1–24. https://doi.org/10.1371/journal.pone.0141807.

Yuan, J., Wang, Z., Duan, P., Xiao, Y., Zhang, H., Huang, Z., Zhou, R., Wen, H., Wang, K., Wang, D., 2021. Whistle signal variations among three Indo-Pacific humpback dolphin populations in the South China Sea: a combined effect of the Qiongzhou Strait's geographical barrier function and local ambient noise? Integr. Zool. 16, 499–511. https://doi.org/10.1111/1749-4877.12531.

Yuan, Z., Richards, E.L., Song, H.C., Hodgkiss, W.S., 2018. Estimation of array tilt using snapping shrimp noise. In: 2018 OCEANS - MTS/IEEE Kobe Techno-Oceans, OCEANS - Kobe 2018, vol. 1, pp. 1–4. https://doi.org/10.1109/OCEANSKOBE.2018.8559391.

Zhou, X., Yang, K., 2020. A denoising representation framework for underwater acoustic signal recognition. J. Acoust. Soc. Am. 147, EL377–EL383. https://doi.org/10.1121/10.0001130.

Zulkifli Poh, A.N., Peter, C., Ngeian, J., Tuen, A.A., Minton, G., 2016. Abundance estimates of indo-pacific humpback dolphins (Sousa chinensis) in Kuching Bay, East Malaysia. Aquat. Mamm. 42, 462–465. https://doi.org/10.1578/AM.42.4.2016.462.