Cloud-enabled passive acoustic monitoring array for real-time detection of marine mammals

Koay Teong Beng*, Hari Vishnu*, Abel Ho*, Ng Yew Choong*, Wu Yusong*, Simen Hexeberg*, Mandar Chitre*¶, Karenne Tun+

*Acoustic Research Laboratory, Tropical Marine Science Institute, National University of Singapore

¶Department of Electrical and Communication Engineering, National University of Singapore

+National Parks Board, Singapore

***Abstract*—Passive acoustics is an effective, robust, non-invasive technique to detect and monitor vocalizing marine mammals. The tropical waters around Singapore are visited or inhabited by several species of marine megafauna. This paper presents a passive acoustic monitoring system developed for real-time detection of marine mammals in Singapore waters. The system is cloud-enabled, and contains machine-learning capable edge-computing systems with surface cameras and underwater hydrophone arrays which were deployed at Sisters' Islands in Singapore. A set of online and offline software tools have been developed to facilitate monitoring, visualization, annotation and data curation to support continual training of the detector. The paper discusses the design, architecture, development, and deployment of this system, offering a comprehensive solution to enhance the monitoring and study of marine megafauna in the region.**

***Index Terms*—PAM, acoustics, ML, marine mammal, Indo-Pacific humpback dolphin, dolphin, edge-computing**

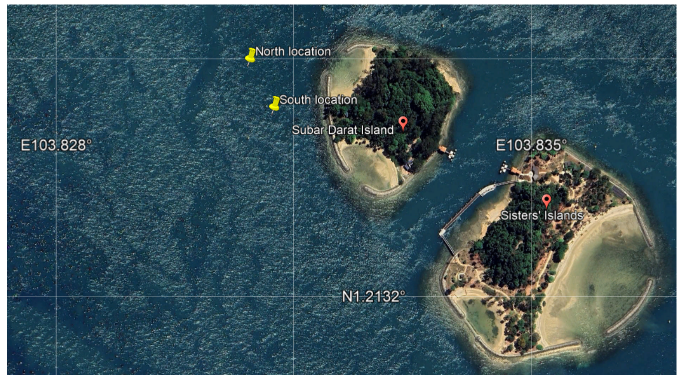


Fig. 1: Two locations near Sisters' Islands in Singapore waters (designated 'North' and 'South') where PAM unit was deployed.

I. Introduction

The shallow tropical waters around Singapore are known to be visited or inhabited by several species of marine megafauna, including Indo-Pacific humpback dolphins (*Sousa chinensis*) [1], bottlenose dolphins, Irrawaddy dolphins, finless porpoise, dugongs and otters [2]. Some of these species, including the Indo-Pacific humpback dolphins and dugongs, are considered vulnerable to extinction [3], [4] and need to be monitored closely in order to inform conservation actions. While some of the other megafauna species in the region are not classified as endangered yet, it is crucial to understand their population trends, given their role in the food chain and habitat. However, there is a lack of reliable baseline estimates of the population of most marine mammal species in local Singapore waters [4]. Most of the estimates in the Southeast Asia region rely on photo-identification of individuals with mark-recapture methods [5]. Given the low visibility in local waters, monitoring megafauna using the underwater visual approach is exceptionally challenging.

Passive acoustics is an effective means of detecting and monitoring vocalizing marine mammals. Passive acoustic monitoring (PAM) of marine mammals is well-studied [6], [7], and specifically, machine-learning (ML) based approaches to mining large acoustic data for their vocalizations has been researched in detail. While visual and photo-identification surveys provide information only on surface activities over small time periods and may be limited by visibility conditions or rough weather, PAM can be applied continuously over extended periods of time even in such conditions. It is non-invasive and can remotely sense these animals, providing information with good spatiotemporal resolution [8]. Vocalizations of marine mammal species such as Indo-Pacific humpback dolphins have been studied across Asia [9], and these creatures are known to vocalize via whistles, click-trains and barks (or burst pulses). Dugongs also produce sounds classified into chirp-squeaks, barks and trills [10]. Thus, these species can be potentially monitored using PAM.

PAM can be used for detection and classification of marine mammal vocalizations, and potentially for population estimation or behavioural assesments. Such assessments benefit highly from multi-modal data spanning other forms of sensory detection and verification, and data on environmental parameters such as water quality, time of the day, weather conditions and ambient noise conditions. Real-time detection of marine mammal visitations can also enable conservation bodies and biodiversity analysts to better understand these occurences.

With these considerations in mind, we developed a cloud-enabled ML-capable edge-computing system with a surface camera and underwater hydrophone array. Two of these systems were deployed at the Sisters' Islands in Singapore waters, to perform real-time detection of marine mammals complemented with environmental data. The locations of the deployed PAM systems are shown in Fig. 1. A set of online and offline

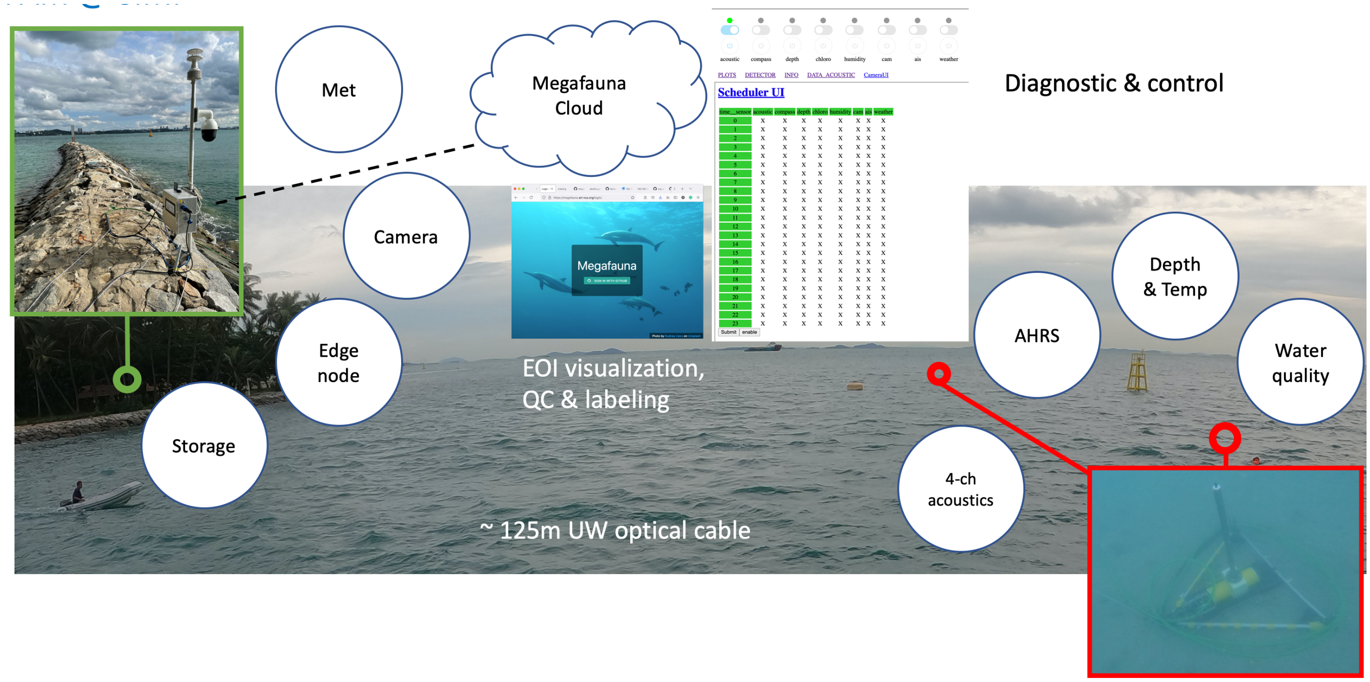


Fig. 2: Overview of the cloud-enabled passive acoustic monitoring array with the computation node surface sensors on the left, underwater acoustic array with underwater environmental sensors on the right and web interfaces for remotely controlling the setup offsite in the middle.

tools were developed to facilitate monitoring, visualization, annotation, and data curation to support continual training of the detector. This paper discusses the design, architecture, development and deployment of this system and its supporting software. Section II describes the PAM system hardware, setup and deployment. Section III details the software tools employed on the PAM and at the user end to visualize outputs from the system. Section IV discusses the system's detection and direction estimation performance, and section V concludes the paper.

II. PAM SYSTEM DEPLOYMENT AND HARDWARE

The PAM systems were deployed off Small Sisters' Island (*Subar Darat*) from February 2022 to November 2022. Each system consisted of a surface station (highlighted in green in Fig. 2, and also shown in Fig. 3) that housed an Automatic Identification System (AIS) receiver to collect ship traffic data, a weather station, a camera, and a computation node. The surface unit was installed on a pole that was bolted and cemented on a nearby water breaker (see Fig. 3). It contained a weatherproof electrical box that held all the electronics. The surface unit had a power management module that allowed us to shut down individual components in the setup for power conservation, including the two PAM arrays connected to it. It had two edge computers (NVIDIA Jetson Nano) for command and control and processing of data from each PAM array. One of the edge computers also received data from a camera mounted on the top for opportunistic ground truthing of detected vocalizations. The computation node ran an ML model locally to detect whistle-like vocalizations in real-time, and stored the collected acoustic data locally. A mobile router

Fig. 3: Photo of the PAM surface unit.

was used to provide internet connectivity that linked the PAM to our cloud server and allowed users to access the PAM remotely via virtual private network (VPN). It also periodically transmitted its status to the cloud and notified the user of any vocalization detection or system error.

Two underwater acoustic arrays with environmental sensors were deployed in the coastal waters about 100 m from the shore and cabled to the surface station (see Fig. 4, array is highlighted in red in Fig. 2). The array contained (a) four hydrophones configured into a 1-meter tetahedral shape, (b) depth sensor, (c) attitude-heading-reference system (AHRS) sensor, (d) chlorophyll sensor, (e) turbidity sensor, (f) Phycoerythrin, and (g) temperature sensors. The depth and AHRS sensor provided crucial information on the location/orientation of the array, allowing accurate estimation of the direction of the sound in the world frame. The depth sensor also provided valuable tidal cycle information. The chlorophyll sensor allowed us to correlate the chlorophyll content and biological productivity against dolphin occurrence and the soundscape of the location. The humidity sensor allowed early detection of water leakage in the system during long-term underwater deployments. Each array recorded signals at a sampling rate of 400 kHz - this allowed us to detect whistle-like frequency modulated sounds, and detect as well as localize the high-frequency sounds like dolphin clicks by utilizing the time-difference-of-arrivals of signals at the hydrophones.

The system was powered by photovoltaic systems with enough lead-acid batteries to buffer up to seven days of bad weather and support night operations. Ten solar panels were deployed to power the PAM-array system, and their energy consumption-generation ledger was monitored continuously. Since we were able to control the power of individual subsystems, the setup allowed us to keep operating the critical components when the energy reservoir was low by powering down less important sensors.

Two 120 m underwater cables were used to connect each underwater hydrophone array to the surface station's edge computers. These were hybrid optical and power cables rated for the harsh underwater environment. This cable transmitted power and network communications over long distances for the underwater unit while maintaining good throughput. Hence, the data could be transmitted to the shore for real-time processing and storage. It allowed easy data retrieval since the data was stored on shore instead of underwater.

III. Software

A comprehensive set of essential software tools were developed and integrated with the system to support data acquisition, vocalization detection, real-time event alert, event annotation, environmental monitoring, diagnostics, and visualization of primary data (acoustics and visual) and secondary data (such as environmental and shipping data). These will be detailed here. A list of software tools developed is also mentioned in Table I with a description of each one's functionality.

A. ML-based vocalization detector

The PAM system ran an ML-based detection algorithm named DEVMAN [11] onsite to detect marine mammal vocalizations in real-time and alert the marine mammal supervisors and/or biodiversity stakeholders to assess the event remotely. A data processing flow was set up to handle data collection, labelling, and quality checks (QCs) of local detections. After verification, these events could be reintroduced into the training data to revise and continuously improve the ML detector.

Software name	Functionality
DEVMAN	ML-based vocalization detection
EViz	Storage and online visualization these onsite-triggered events-of-interest with photos and summary diagnostics
Iceberg	Acquire, summarize and visualize AIS data
Prometheus	Monitoring and time-series display
ReCorD	Control and diagnostics
Baywatch	Batch labelling and online annotation of snippets of data for quality checking
Barreleye	Mass labelling of detector-flagged audio clips offline

TABLE I: Summary of software tools and apps developed to complement the monitoring system

The detector was deployed on a Jetson Nano single-board computer.

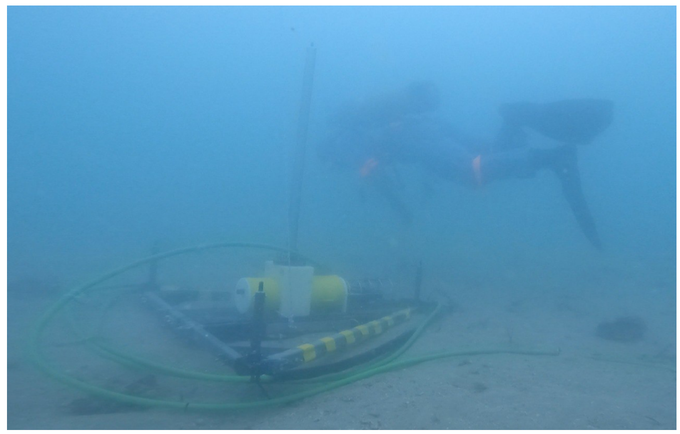


Fig. 4: Underwater photo of the installed 4-channel PAM hydrophone array for real-time monitoring.

The ML-based detector on the system is trained in a supervised fashion, which needs large amounts of labelled data. Due to the lack of sample vocalizations from local waters, we initially started the training with a curated set of whistles from different dolphin species found in online sound libraries [12], [13]. We augmented this data by applying random flipping, shifting and adding locally recorded ambient noise data, to increase the diversity of training and validation data and improve the generalization of the detector. Later on, data was collected using single-hydrophone recorders deployed across several locations in Singapore waters [11]. The training and validation dataset were iteratively improved by adding newly detected signals from local waters into these datasets, and regularly revamping the dataset to remove ambiguous or less useful samples. They were then used to retrain the detector and evaluate its performance on a control dataset. The high levels of impulsive noise generated by snapping shrimp in local waters makes detection challenging [14]. Thus, an ML based denoiser and detector was trained to be robust to this and pick up frequency-modulated (whistle-like) marine mammal vocalizations within this background noise [11]. The broadband nature of the echolocation clicks of dolphins makes it hard to distinguish it from snapping shrimp snaps unless detailed characteristics of the pulses are analyzed [15]. On the other hand, whistles are narrowband and hence are easier to pick up from the spectrogram as compared to snapping shrimp noise.

B. Visualization and annotation tools

An online browser-based visualization app called EViz was developed to store and visualize onsite-triggered events-of-interest (EOI) (see Fig. 5(a)). Using this, researchers or supervisors could choose to investigate the event and label/annotate the snippet online. The user could also access Prometheus, a monitoring tool with time-series plots of sensor states and some summarized sensors values obtained from the PAM array for diagnosis (Fig. 5(b)). A conservative detection threshold was established for the onsite operation of the edge detector (more details are given in Section IV). When the edge detector flagged a detection, subscribed users were notified of these events via a message on the Telegram app. The acoustic clip that triggered the detection was uploaded online, along with a photo acquired at the time of the detection from the camera, and the recorded environmental parameters of interest during the event.

A companion software named Iceberg was also developed to visualize the information available on local shipping activity from the AIS (Figs. 5(c) and (d)).

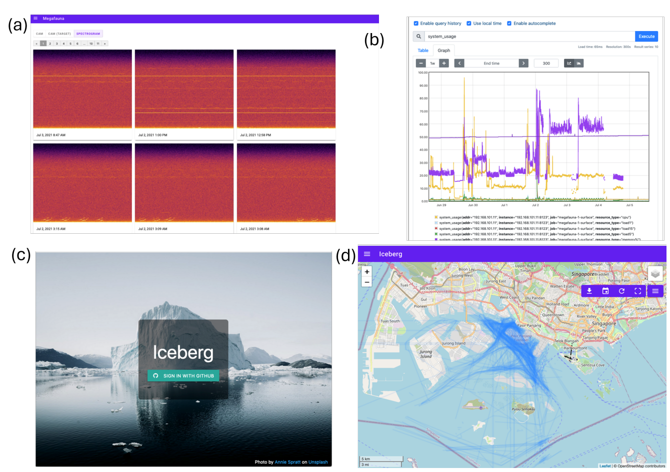


Fig. 5: (a) EViz, a cloud service to store events and provide visualizations of onsite-detected events of interest, (b) Prometheus, a remote system diagnostics software for the PAM-array, (c) login screen and (d) information screen with ship vessel tracks from Iceberg, a tool to store and visualize AIS information.

A minimalist online annotation tool named Baywatch was developed to support batch labelling and QC of the data recorded and flagged by the detector. This data could undergo several versions of annotations before being categorized into a final form suitable for interpretation. Users wishing to solicit feedback on the data acquired by them, can curate and upload a subset of snippets into this app and invite peers to review them via an online link. The app displays the time-series and spectrogram of the audio snippet with a set of audio playback control buttons (play, pause and stop), a list of predefined labels the user can choose from, and a text box for adding in additional comments (shown in Fig. 6). In addition to quality-checking of a small number of snippets, it is also required to annotate a large amount of ground truth data for the purpose of further training and validation of the ML detector. With this in mind, we also developed Barreleye, an offline application to help developers swiftly review, annotate, and cluster snippets en-masse.

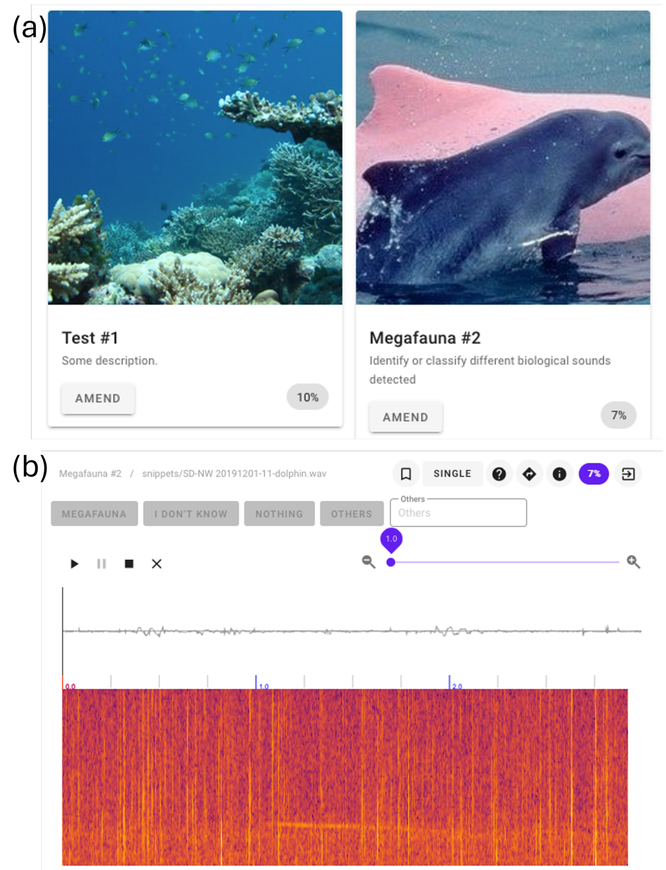
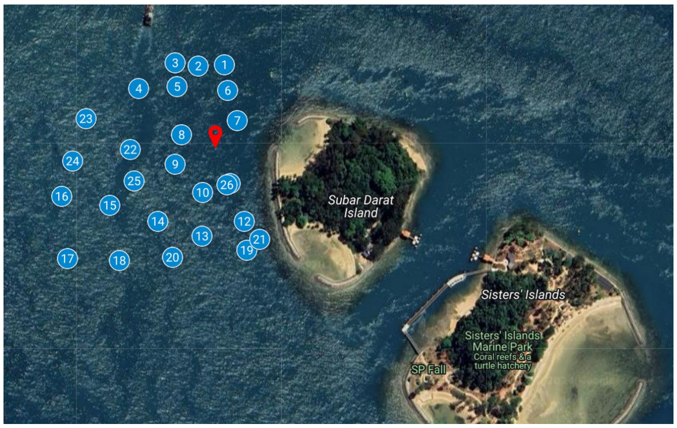


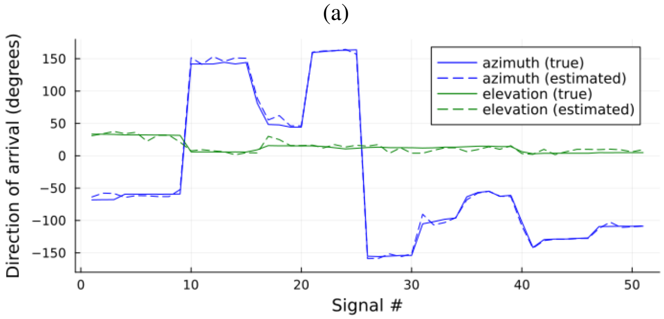
Fig. 6: Baywatch, a cloud service focuses on over-the-cloud annotations of audio clips with predefined labels: (a) the summary panel and (b) annotation screen of the software.

IV. PAM ARRAY DETECTION AND DIRECTION ESTIMATION

The tetrahedral constellation of the PAM allows direction-of-arrival (DOA) estimation of source signals, which can aid in localization tasks. Besides characteristics of the source signal such as SNR, bandwidth and frequency, the estimation accuracy is affected by the accuracy of the position and orientation of the PAM – neither of which was known precisely. Therefore, a calibration exercise was undertaken to correct the positions and orientations of the array, using a controlled acoustic source. A series of communication signals, comprising a Linear Frequency Modulated (LFM) header followed by a pseudo-random sequence with the timestamp and GPS position encoded, was transimitted approximately 0.5 m below the surface from 26 locations in the neighbourhood of the PAM at the North location (Fig. 7a). Signals with low received SNR were discarded, while DOA was estimated for the remaining signals using the DETSAC algorithm developed in [16]. DOA estimates with high uncertainty were ignored and the remaining 51 DOAs were used for calibration. These shortlisted signals were used to adjust the PAM's position (longitude, latitude, depth) and orientation (roll, pitch, yaw) by minimizing the mean absolute error between the estimated DOAs and ground-truth DOAs calculated using the signals' GPS locations. The resulting DOA estimates are shown in Fig. 7b and achieved a median absolute error of 4.9°, which is within the error margin imposed by GPS uncertainty.



(a)



(b)

Fig. 7: (a): The PAM at the North location (red) and the approximate location for each series of signal transmissions during the PAM calibration process. The same signal was transmitted multiple times at each location but not all transmissions were used for calibration. (b): DOA estimates and corresponding GPS-based ground truth directions for the 51 shortlisted signals after calibration.

In order to (1) test practical issues with the in-field performance of the detector and the complete PAM processing chain including event detector, cloud upload and online event-reporting, and (2) evaluate performance variability with different detector thresholds and signal types, a controlled experiment was done with the PAM array in the coastal waters near Sisters' islands, Singapore. We deployed the array from a boat onto the seabed and tied it to a mooring buoy. We then deployed a transmitter in the vicinity of the array and played back ten types of frequency modulated signals (dolphin whistles)– this consisted of 1 baseline high-SNR signal obtained from an online source, and 9 signals recorded locally using the single-hydrophone system (with relatively lower SNR), denoted as signals 1-9. The signals were transmitted at different source levels. Due to this controlled variation in source levels and variability in the local ambient noise, the PAM received the vocalizations at different SNRs, allowing us to assess how the detector would perform with variation in SNR.

First, we plot the histogram of the detector scores for clips containing only ambient noise (Fig. 8(c)). The plot demonstrates that the scores show a high amount of variability due to ambient noise, which is expected in the challengingly noisy and variable Singapore waters with high biological and shipping activity. This histogram guides an appropriate selection of detector score to obtain a preferred operating false-alarm rate. For example, the histogram shows that setting the threshold to a value of 2.5 yields a false-alarm rate of about 0.1%. We choose this score to be the operating point for the detector's real-time operation which would alert stakeholders via the cloud app, with the intuition that we want to minimize the number of false alerts. For post-processing and interpretation of the data, however, we set a lower threshold of 1 which yields $\tilde{1}$% false-alarm rate, as this yields more detection probability, and the irrelevant data can be further filtered out later in post-processing as necessary.

The baseline signal was transmitted at different SNRs within the range (-15, 15) dB, with repetitions at some SNR values, to test the variability in detector score. It can be seen from Fig. 8(b) that even for the same signal at the same received SNR, there can be considerable variability in the detector scores. This is because detectability of the signal (and thus the detector score) is not only a function of the SNR, but also depends on multiple other factors such as whether the noise obscures important signal features that aid detection.

We fit a third-order polynomial curve to the individual scores variation obtained with the baseline signal in Fig. 8(b), and this is plotted in Fig. 8(a) alongside the detector scores for other signals. The detector thresholds that yielding false-alarm rates of 0.1% and 1% are plotted in the figure too, along with the median score corresponding to ambient noise. It is seen that at an operating detector threshold corresponding to 1% false-alarm rate, the baseline signal is detected at an SNR of -7.5 dB at least. Fig. 8(a) shows that the different signals exhibit variability in detector scores. This is because the scores also depend on the frequency variability or shape of the signal itself that is considered to be a convincing evidence of the signal being of marine mammal origin by the detector (based on the training data). This signal shape varies between the different signals considered.

Overall, Fig. 8 gives us a summary and evaluation of the system's performance variability in realistic operating conditions. The score variation for a single signal roughly follows an S-shaped (or sigmoidal) curve, as seen in the plot for the baseline signal. This variation is on somewhat expected lines and is explained as follows. At very low SNRs, the signal is almost undetectable and flattens out at a score matching the ambient noise, and its performance does not further degrade with decrease in SNR. At very high SNRs, the signal is clearly detected, and further increase in signal strength doesn't improve detectability further. Within these two extremes, the

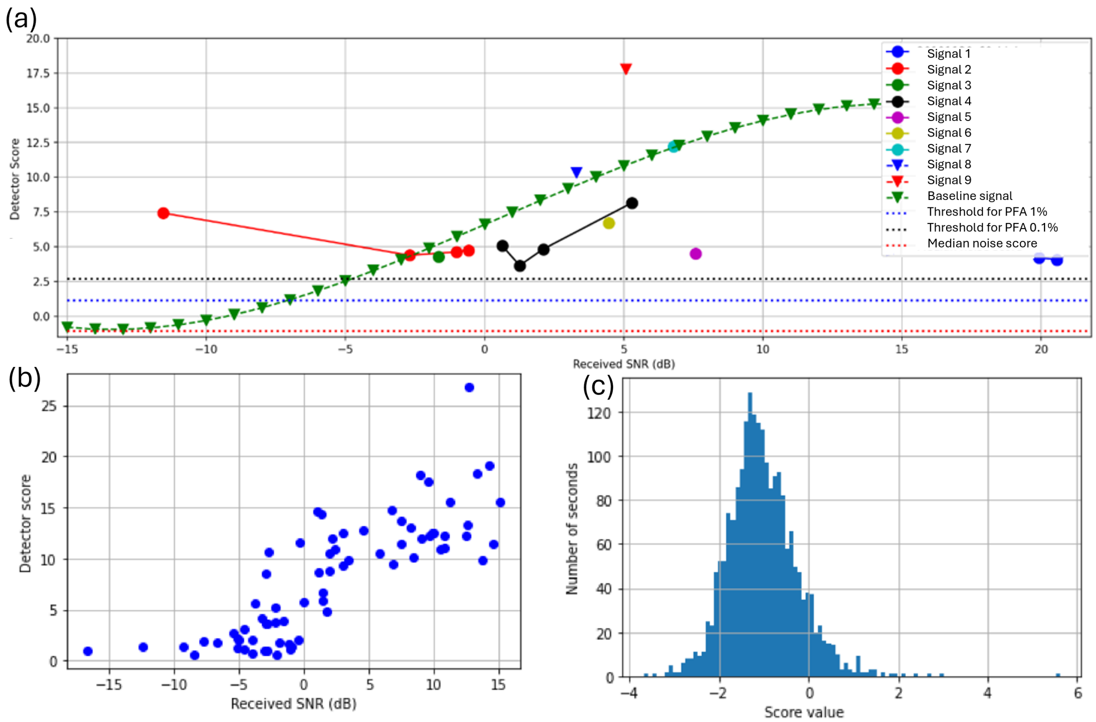


Fig. 8: Variation of detection score versus received signal SNR (in dB) of signals played back during the field-test of the PAM, (a) for all 10 dolphin signals considered, and (b) for only the baseline signal transmitted, and (c) histogram of detector scores with ambient noise (no transmissions) at the test site. The green line with triangle markers corresponding to the baseline signal in (a) is a third-order polynomial fit of the points in (b).

SNR does affect the detectability of the signal.

V. Conclusion and Future Work

This paper describes a comprehensive PAM monitoring system that combines an acoustic array for vocalization detection, an optical sensor for opportunistic ground truth, atmospheric and underwater environmental sensors to correlate the detections with environmental factors, and a machine-learning capable edge computation platform to support real-time detection, including its design, architecture, development and deployment. A comprehensive set of essential software tools was developed and integrated with the system to support data acquisition, vocalization detection, real-time event alert, event annotation, environmental monitoring, diagnostics, and visualization of primary data (acoustics and visual) and secondary data (such as environmental and shipping data). The calibration of the position and orientation of the PAM array via field trials, as well as performance analysis of the variation in the detector's performance with variability in signal types and SNRs is detailed. These provide a summary evaluation of the system's performance in realistic operating conditions.

Subsequent to calibration and performance verification, the PAM was successfully deployed for continuous long-term operation. The data from this system was used to estimate the direction-of-arrival of detected sounds, and the likelihood that multiple marine mammals were vocalizing at the same time, detailed in [16].

References

[1] B. P. Y.-H. Lee and M. Ooi, "Indo-Pacific humpback dolphins near Jurong Island," *Singapore Biodiversity Records*, pp. 84–85, 2020.

[2] M. D. Khoo and B. P. Lee, "The urban Smooth-coated otters Lutrogale perspicillata of Singapore: a review of the reasons for success," *International Zoo Yearbook*, vol. 54, no. 1, pp. 60–71, 2020.

[3] T. Jefferson, B. Smith, G. Braulik, and W. Perrin, "Sousa chinensis (errata version published in 2018)," *The IUCN Red List of Threatened Species 2017: e.T82031425A123794774*, 2017.

[4] H. Marsh, "United Nations Environment Programme eds: Dugong: Status Report and Action PLANS for Countries and Territories," United Nations Environment Programme, Nairobi, Tech. Rep., 2002.

[5] S. C. Y. Chan and L. Karczmarski, "Indo-Pacific humpback dolphins (Sousa chinensis) in Hong Kong: Modelling demographic parameters with mark-recapture techniques," *PLOS ONE*, vol. 12, no. 3, p. e0174029, mar 2017. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0174029

[6] E. Fleishman, D. Cholewiak, D. Gillespie, T. Helble, H. Klinck, E. Nosal, and M. A. Roch, "Ecological inferences about marine mammals from passive acoustic data," *Biological Reviews*, vol. 98, no. 5, pp. 1633–1647, oct 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/brv.12969

[7] Y. Shiu, K. J. Palmer, M. A. Roch, E. Fleishman, X. Liu, E. M. Nosal, T. Helble, D. Cholewiak, D. Gillespie, and H. Klinck, "Deep neural networks for automated detection of marine mammal species," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.

[8] F. Caruso, G. Alonge, G. Bellia, E. De Domenico, R. Grammauta, G. Larosa, S. Mazzola, G. Riccobene, G. Pavan, E. Papale, C. Pellegrino, S. Pulvirenti, V. Sciacca, F. Simeone, F. Speziale, S. Viola, and G. Buscaino, "Long-Term Monitoring of Dolphin Biosonar Activity in Deep Pelagic Waters of the Mediterranean Sea," *Scientific Reports*, vol. 7, no. 1, pp. 1–12, 2017.

[9] P. Q. Sims, R. Vaughn, S. K. Hung, and B. Würsig, "Sounds of Indo-Pacific humpback dolphins (Sousa chinensis) in West Hong Kong: A preliminary description," *The Journal of the Acoustical Society*

of America, vol. 131, no. 1, pp. EL48–EL53, 2012. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.3663281

[10] K. Ichikawa, C. Tsutsumi, N. Arai, T. Akamatsu, T. Shinke, T. Hara, and K. Adulyanukosol, " Dugong (Dugong dugon) vocalization patterns recorded by automatic underwater sound monitoring systems ," *The Journal of the Acoustical Society of America*, vol. 119, no. 6, pp. 3726–3733, 2006.

[11] H. Vishnu, V. R. Soorya, M. Chitre, Y. M. Too, T. B. Koay, and A. Ho, "Machine-learning based detection of marine mammal vocalizations in snapping-shrimp dominated ambient noise," *Marine Environmental Research*, vol. 199, 2024.

[12] "Watkins marine mammal sound database, woods hole oceanographic institution and the new bedford whaling museum," http://https://cis.whoi.edu/science/B/whalesounds/index.cfm, accessed: 2019.

[13] N. N. F. S. C. NOAA [National Oceanic and Atmospheric Administration]. NOAA Fisheries, "Passive Acoustic Research Group," 2023. [Online]. Available: https://www.fisheries.noaa.gov/national/science-data/sounds-ocean-mammals

[14] A. Mahmood, M. Chitre, and H. Vishnu, "Locally optimal inspired detection in snapping shrimp noise," *IEEE Journal of Oceanic Engineering*, vol. 42, no. 4, 2017.

[15] F. Caruso, L. Dong, M. Lin, M. Liu, Z. Gong, W. Xu, G. Alonge, and S. Li, "Monitoring of a Nearshore Small Dolphin Species Using Passive Acoustic Platforms and Supervised Machine Learning Techniques," *Frontiers in Marine Science*, vol. 7, no. April, 2020.

[16] S. Hexeberg, H. Vishnu, K. T. Beng, A. Ho, W. Yusong, M. Chitre, K. Tun, and K. Lim, "Acoustic detector for multiple vocalizing marine mammal individuals," in *OCEANS 2023 - Limerick*. IEEE, jun 2023, pp. 1–8. [Online]. Available: https://ieeexplore.ieee.org/document/10244432/