Detecting Ultrasonic Harbor Porpoise Clicks with Hetrodyning and Teager–Kaiser Energy Operator

2022 7th International Conference on Frontiers of Signal Processing (ICFSP) | 978-1-6654-8158-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICFSP55781.2022.9924846

Patrick J. Donnelly
Department of Computer Science
Oregon State University
Corvallis OR, United States
patrick.donnelly@oregonstate.edu

Daniel De Leon
Department of Computer Science
Oregon State University
Corvallis OR, United States
deledani@oregonstate.edu

David K. Mellinger
Hatfield Marine Science Center
Oregon State University
Newport OR, United States
david.mellinger@oregonstate.edu

***Abstract*—This study advances research in the design of autonomous machine listening agents to identify the presence of marine mammals. We compare two common deep learning architectures in the detection of ultrasonic harbor porpoise clicks in noisy coastal soundscapes. We reduce the high dimensionality of the ultrasonic underwater recordings with a heterodyned Teager–Kaiser energy operator approach, which shifts the ultrasonic signal down to lower frequencies. We cross-validate our deep learning models by day to increase our model's ability to generalize between the ever-changing background noise environment. We demonstrate our approach over a large dataset of (n=8286) examples of harbor porpoise click steams recorded in the Bay of Fundy. We discuss our results in the context of emphasizing the need to overcome such shifts in noise to design a robust porpoise click classifier that is capable of real-time deployment and able to generalize to all real-world conditions.**

Keywords—bioacoustics, convolutional neural network, heterodyne, long short-term memory, deep learning, marine mammal detection, Teager–Kaiser energy operator

I. Introduction

Phocoena phocoena is an aquatic mammal, more commonly known as the harbor porpoise. The harbor porpoise is one of the smallest species porpoises and adults measure only around five feet in length. True to their name, they are frequently observed in coastal waters. Like other toothed cetaceans, harbor porpoises use echolocation to navigate and hunt. They emit clicks to interpret the echoes that bounce off of their prey and other objects, providing information such as size, structure, material composition and shape of objects [1], [2]. This ability is known as echolocation or biological sonar.

Compared to most dolphins, however, harbor porpoises emit clicks at a much narrower bandwidth, ranging between 6 to 26 kHz at a center frequency between 130 and 140 kHz [3]. Harbor porpoise click durations last between 44 and 115 µs with a variable interclick interval between 30 and 100 ms. These clicks are classed as narrow-bandwidth high-frequency (NBHF). Evolutionary biologists believe that species developed NBHF clicks by selection pressure to avoid predation from killer whales [4]. Porpoises produce clicks for communication and to forage for food [5]. The spectrogram shown in Fig. 1 represents a typical train of clicks emitted by a harbor porpoise.

The Bay of Fundy in Nova Scotia, Canada hosts some of the strongest tides in the world, fluctuating between 13 and 38 feet during the year [6]. Given the magnitude of these currents, the bay is targeted as an ideal location to harvest energy using tidal turbines [7]. This bay, however, also host a large population of harbor porpoises. The large tidal turbines pose a physical threat that can hit a passing porpoise with its blade. Additionally, these turbines produce a significant amount of mechanical noise that can interfere with their ultrasonic echolocation used to forage and communicate [3]. Harbor porpoise activity has been shown to decline when turbine are in operational in Bay of Fundy [8].

For these reasons, there is a strong need to create a reliable acoustic monitoring system to detect the presence of these porpoises in real time. Such a tool would provide an opportunity for an expedient intervention to switch off the tidal turbines as soon as porpoises are detected in the vicinity. This would enable humans to continue to harvest tidal energy while better managing and mitigating our impact on their environment.

This investigation explores efficient strategies to classify ultrasonic harbor porpoise vocalizations (clicks) recorded in the Bay of Fundy. Specifically, we strive to create an autonomous machine listening agent that can predict if a harbor porpoise click exists in a given clip of ultrasonic audio. This is a challenging task given the high levels of background noise often present in the signal. In this work, we detect ultrasonic harbor porpoise clicks in a comparison of two deep model architectures a convolutional neural network (CNN) and a long short-term memory (LSTM) recurrent neural network.

This work makes several contributions to the areas of marine mammal detection with deep-learning. First, we compare two common deep learning architectures on the same dataset. Second, we propose a leave-one-day-out cross-validation strategy to better generalize across our limited observation data and to overcome potential limitations caused by bursts of background noise. Third, we propose a technique of heterodyning ultrasonic marine mammal audio as a dimensionality technique prior to the input to the deep learning models. By shifting the high frequency clicks down to lower frequencies we significantly reduce the dimensionality of input data and, subsequently, the time needed to train deep learning models on raw audio signals. Furthermore, such an approach enables researchers to treat input signals directly without the need for timely preprocessing and feature extraction. This enables our approach to be deployed in real time, compared to common approaches requires spectrograms or signal processing.

978-1-6654-8158-8/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: OREGON STATE UNIV. Downloaded on November 22,2026 at 02:44:33 UTC from IEEE Xplore. Restrictions apply.

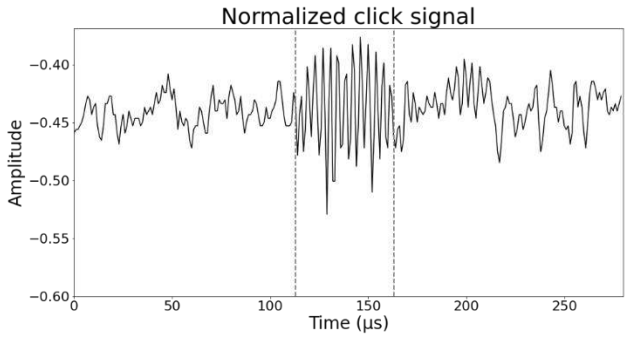


Fig. 1. Waveform centered on a single porpoise click.

II. Background and Related Work

In this section, we briefly explain the heterodyne method and the Teager-Kaiser Energy Operator that we employ in our approach. We also review related work relevant to our approach.

A. Heterodyne

Heterodyning is a signal processing method broadly used across many applications in communications. Heterodyning shifts a frequency band into another band of equal width. For example, bat bioacousticians use real-time heterodyning to shift ultrasonic frequencies down to the ranges audible by humans.

Digital heterodyning is performed in three main steps. First, a band pass filter is applied around the frequencies of interest. This removes all energy except in the frequency band to be shifted. Second, the frequencies are shifted by mixing the signal with a sine wave to produce heterodyned signals appearing at the sum and difference of the two signals. Third, a low pass filter removes the unneeded copy of the heterodyned signal.

We show the effects of heterodyning on a sample click train in Fig. 2. The heterodyne operation manages to preserve the clarity of the clicks while effectively reducing the sample rate from 512 kHz down to 100 kHz.

B. Teager-Kaiser Energy Operator

The Teager-Kaiser energy operator (TKEO) [9] is a signal conditioner that was first popularized for reducing noise when detecting onsets in electromyography (EMG) signals [10]. TKEO is a measure of changes of energy of a time-varying frequency. TKEO calculates the energy from instantaneous amplitude and frequency measures of the signal. This is a major advantage over other onset detection methods because it does not requiring the timely processing long windows of samples. TKEO is now used in various applications from mechanics to image and audio processing.

TKEO is a discrete operator a signal x defined as

$$\psi[x(n)] = x^2(n) - x(n+1)x(n-1)$$

where n is the discrete sample number in x. The TKEO conditioner measures the instantaneous changes in energy at each sample by taking the difference of the squared value of the current sample and the product of the previous and the following samples in the signal. The TKEO operation significantly reduces the noise while also increasing the amplitude of each peak.

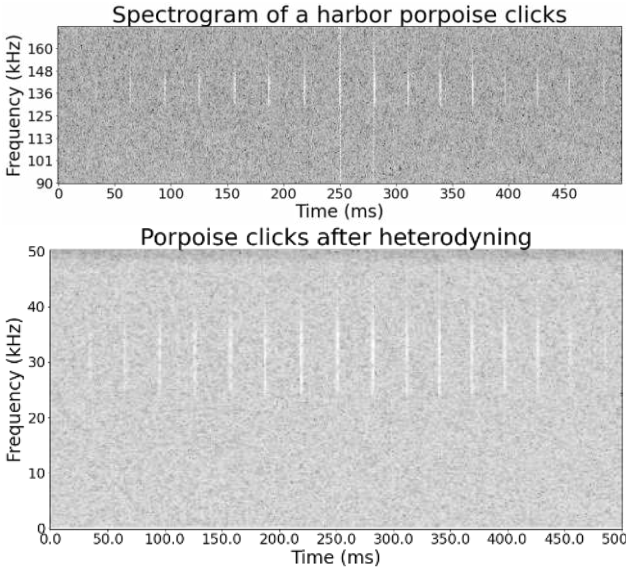


Fig. 2. Spectrograms of an example harbor porpoise click train before (top) and after hetrodyning (bottom).

C. Related Work

Heterodyning is a popular method to bring ultrasonic bat echolocation down to audible frequencies [11]. This approach enables researchers to listen to echolocation in the field. However these methods are only appropriate for narrow bands of interest, and are not well suited for monitoring of mixed acoustic environments where multiple species overlap [12].

The TKEO operator has shown promise in the identification cetacean clicks. Kandia et al. applied the TKEO operator on signals of sperm whale clicks before applying a simple forward–backward search algorithm [13]. Sperm whale clicks are produced in the 10 Hz to 30 kHz range and have interclick intervals ranging from 0.5 to 2 seconds. Compared to harbor porpoises, the sperm whale clicks are an order of magnitude lower, at the lowest end of human hearing. And this interclick interval is an order of magnitude longer than harbor porpoises.

Deep learning approaches are increasing applied to answer questions in bioacoustic research [14], including in identification of marine mammal calls [15]. Duan *et al.* compared CNN [16] and LSTM [17] models to find the LSTM slightly outperformed the CNN in the task of identifying three different marine mammals. Another researcher used CNNs to classify sperm whale clicks but used LSTMs to identify individual specimens [18]. In this work, we also compare CNNs and LSTMs to explore a novel dataset of harbor porpoise clicks.

III. Dataset

The audio analyzed in this study was collected through a deployment of a JASCO Applied Sciences Autonomous Multichannel Acoustic Recorder (AMAR) by Fund Fundy Ocean Research Center for Energy (FORCE). This study [19] assessed the operational limitations of AMAR relative to other passive acoustic monitoring (PAM) instruments when attempting to detect harbor porpoise clicks. The AMAR hydrophone was mounted on a subsea platform and deployed in Minas Passage in the Bay of Fundy, Nova Scotia Canada.

Authorized licensed use limited to: OREGON STATE UNIV. Downloaded on November 22,2026 at 02:44:33 UTC from IEEE Xplore. Restrictions apply.

A. Data Collection

The data was collected across two deployments in 2019 in which the hydrophone recorded continuously, first in early summer and again in September. This study collected a total of 47 days of ultrasonic audio. The hydrophone recorded at a sample rate of 512 kHz to fully capture the porpoise clicks that typically occur between 110 kHz and 150 kHz [19]. The continuous audio streams were segmented and stored as two-minute segments. From these segments, the study's authors identified and harvested individual clicks.

From this source of 47 days of audio recorded from the AMAR, the authors manually annotated eight days of harbor porpoise clicks. To find the porpoise clicks they analyzed the energy level in specific frequency bands using the Matlab bioacoustic analysis tool Ishmael [20]. The Ishmael detector detects possible porpoise clicks when the ratio of the energy of the porpoise click frequency to energy in a lower frequency band (below 110 kHz) is greater than a detection threshold chosen heuristically in order to maintain a false-negative rate of less than 5%. Domain specialists checked these automated detection to manually confirm the presence of porpoise clicks to confirm. Centered on each click, the authors created a half-second window. We use these half-second windows as a positive example of a porpoise click, given the classification label *click.*

B. Data Selection

These eight days contain 4,143 annotated clicks. The number of clicks collected varied by day. Seven days stem from the same week of recording. The eighth day was recorded in an earlier deployment. In Fig. 3 we show a long term spectral average (LTSA) plot of the sample period.

To balance our dataset with audio examples in which a click is not present, we created an equivalent number of *negative* examples, labeled as *noClick*. For consistency we followed the procedure used to identify the positive examples and applied the Ishmael detector to identify noise peaks. However, we focused on the segments that had been manually confirmed to not have clicks in them [20]. We samples segments from the same set of days as the click examples. These noClicks were harvested from segments near the click segments in an attempt to maintain consistent background noise. After compiling our final dataset, all the 8,286 click and noClick examples (see Table I) were individually normalized to a level of -0.1 dB relative to 0 dB.

TABLE I. NUMBER OF HALF SECOND EXAMPLES BY DAY

Date	*Click*	*NoClick*
07/14	924	924
09/05	337	337
09/07	769	769
09/08	334	334
09/09	831	831
09/10	213	213
09/11	407	407
09/12	328	328
Total	4143	4143

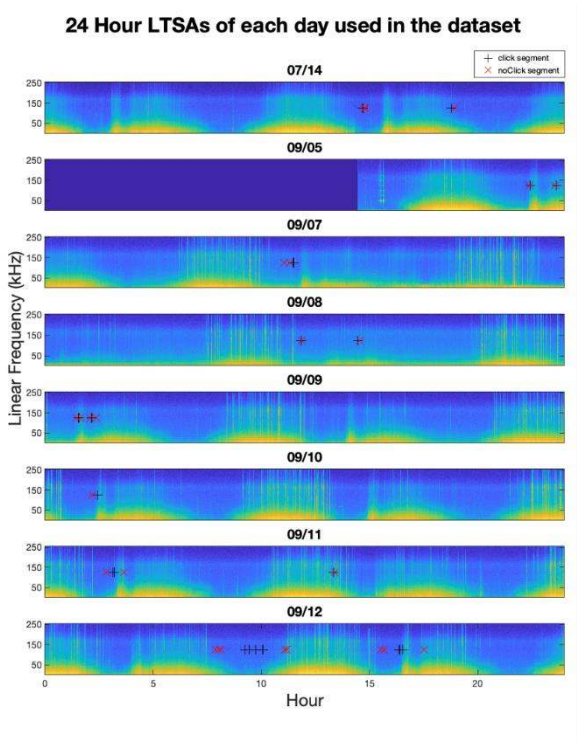


Fig. 3. LTSA of each sample day with markers noting where click (+) and noclick (x) exist. 09/05 is incomplete as it was the first day of recording session.

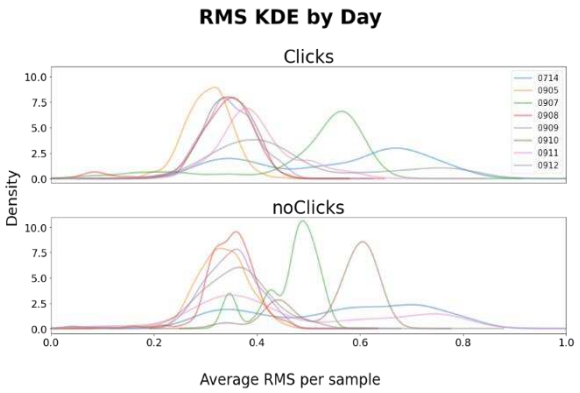


Fig. 4. Histogram of RMSs distributions of examples by day.

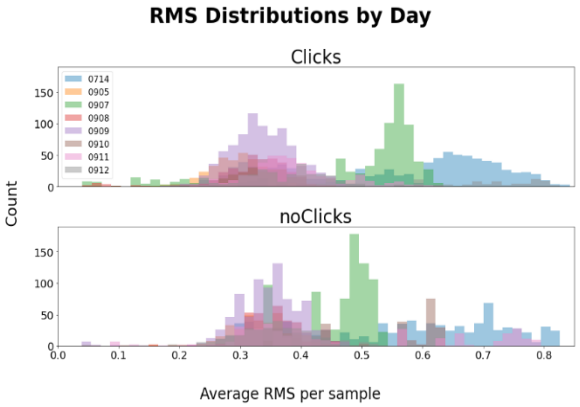


Fig. 5. KDE Gaussian representation of the RMSs distribution.

Authorized licensed use limited to: OREGON STATE UNIV. Downloaded on November 22,2026 at 02:44:33 UTC from IEEE Xplore. Restrictions apply.

C. Data Analysis

One of the challenges of acoustic recording in the Bay of Fundy is that the noise environment is highly variable hour by hour. These noise events might be caused by a passing ship, tidal turbines, or even the noise of the tides themselves.

To explore how our samples varied over time, we calculated the root-mean-square (RMS) across our samples and compared the distributions by day, as shown in Fig. 4. We observe that two days in particular, 07/14 and 09/07, exhibit very different distributions that the other five days. To better illustrate this phenomenon, we show a KDE Gaussian representation of the average RMS by day as Fig. 5. We note that 09/10 has a unique density distribution for sample of *noClick* but not for its clicks. These differences in noise profiles between days illustrate some of the challenges of this noisy environments.

D. Hetrodyne-TKEO

Each raw audio observation in our dataset consists of 256,000 samples, totally around 512 KB. We heterodyne these signals to reduce our signal to have 51,200 samples, a dimensionality reduction by a factor of five. Following the heterodyne transformation, we apply the TKEO operator which retains the size of 51,200 samples per half second window. The steps of this transformation are illustrated in Fig. 6.

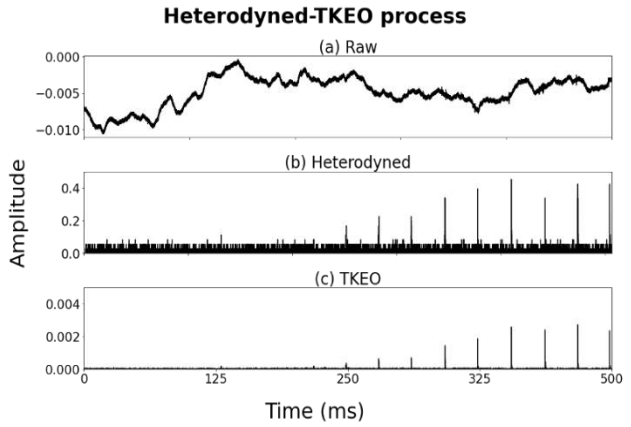


Fig. 6. Waveforms at each step of the heterodyned-TKEO process: (a) raw audio, (b) heterodyned and normalized signal, and (c) result of TKEO operator.

IV. Results

We compared two common deep learning architectures: a Convolutional neural network (CNN) and Long Short-Term Memory (LSTM) architectures. For the CNN, we used two 2D convolutional layers with MaxPooling, followed by two fully connected dense layers with dropout. We selected a similar architecture for the LSTM, containing two LSTM layers with dropout followed by two fully connected dense layers.

We trained both the CNN and the LSTM using the heterodyne-TKEO transformed signals as input. We trained our models over the eight days using a leave-one-day-out cross-validation strategy (8-LODO) to ensure that signals collected together within a same short time period are all used for either training or testing to help ensure our models generalize between different days and noise conditions. The results of this experiment are shown in TABLE II.

TABLE II. Classification Results for CNN and LSTM Models

	CNN	LSTM
Precision	.78	.83
Recall	.92	.88
F_1 Score	.83	.85
Accuracy	.81	.84

The LSTM marginally outperformed the CNN, similar to results exploring the identification of different marine mammals [16], [17]. Confusion matrices by day tested are shown in Fig. 7 for both models. We observe that on 09/09 the CNN has a concerning false positive (FP) rate of 0.60, indicating it misclassified more *noClicks* than it correctly classified. This date was not a day we identified as an outlier in our analysis of RMS level. We also observe that the LSTM did not suffer from similar difficulties with this day. Conversely on 09/05, the LSTM suffers a FP rate of 0.51, indicating that it misclassified more *noClicks* than it classified correctly. The CNN also had difficulty on this day, but still outperformed the LSTM model.

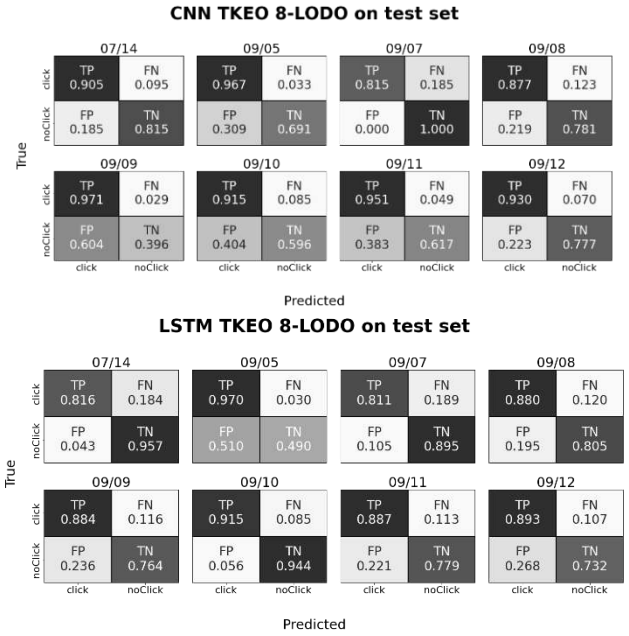


Fig. 7. Confusion matrices for CNN (top) and LSTM (bottom) tested on the day held out for testing.

To further explore the differences between the two models, we examined the types of misclassifications by day. We observed that the two models each struggled on different days. For example, the CNN misclassified significantly more on 09/09 than the LSTM. Conversely, the LSTM struggled on 09/05 significantly more than the CNN.

The set differences shown in Fig. 9 illustrate the unique misclassifications of each architecture while the intersections show misclassifications that both architectures had in common. The set differences of the misclassifications are all higher than their respective intersections. For example 09/10 has a significantly low intersection rate meaning that the errors the CNN and LSTM make are mostly on different examples. This implies that together, these two models could work in tandem to

Authorized licensed use limited to: OREGON STATE UNIV. Downloaded on November 22,2026 at 02:44:33 UTC from IEEE Xplore. Restrictions apply.

improve prediction accuracy. As future work we plan to explore ensemble approaches to leverage the strengths the CNN and LSTM across the varying noise conditions while minimizing the effect of their individual weaknesses.

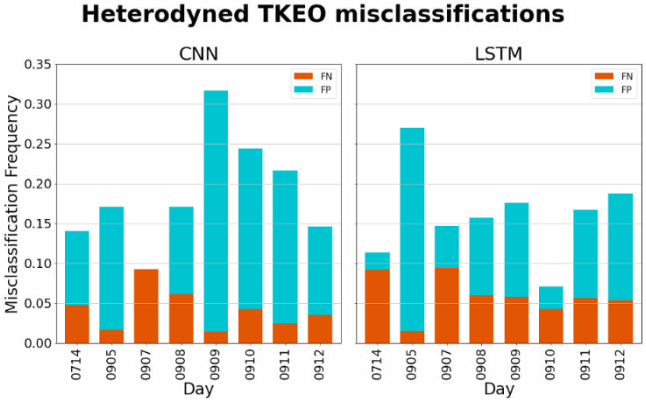


Fig. 8. Misclassifications for the CNN (left) and LSTM (right).

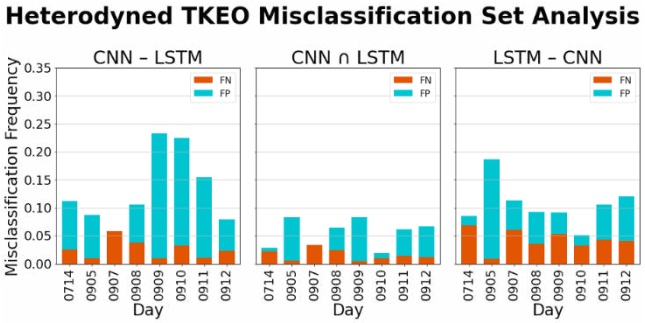


Fig. 9. Set differences (left and right) and intersections (middle) of the TKEO misclassifications from the CNN and LSTM.

V. Conclusion

In this work we propose a novel approach to classifying harbor porpoise clicks in a high noise environment. We train deep learning models directly on a minimally processed raw audio signal, without the timely process of feature extraction or generation of spectrograms. We heterodyne the high dimensional signal to reduce the sample size of ultrasonic signal and we applied the TKEO operator to reduce the noise of the signal. We compare two different deep learning approaches to demonstrate the feasibility of a heterodyned-TKEO approach to identify ultrasonic marine mammal clicks. Despite these encouraging results, more work remains to be done to validate this approach on more data, with other marine mammals with echolocations, and across other geographic locations.

References

[1] W. W. L. AU, "Echolocation in Dolphins with a Dolphin-Bat Comparison," *Bioacoustics*, vol. 8, no. 1–2, pp. 137–162, Jan. 1997, doi: 10.1080/09524622.1997.9753357.

[2] W. W. L. Au, "Biosonar Discrimination, Recognition, and Classification," in *The Sonar of Dolphins*, W. W. L. Au, Ed. New York, NY: Springer, 1993, pp. 177–215. doi: 10.1007/978-1-4612-4356-4_9.

[3] L. A. Miller and M. Wahlberg, "Echolocation by the harbour porpoise: life in coastal waters," *Front Physiol*, vol. 4, p. 52, Apr. 2013, doi: 10.3389/fphys.2013.00052.

[4] P. T. Madsen, D. Carder, K. Bedholm, and S. Ridgway, "Porpoise clicks from a sperm whale nose—Convergent evolution of 130 kHz pulses in toothed whale sonars?," *Bioacoustics*, vol. 15, no. 2, pp. 195–206, 2005.

[5] P. M. Sørensen, D. M. Wisniewska, F. H. Jensen, M. Johnson, J. Teilmann, and P. T. Madsen, "Click communication in wild harbour porpoises (Phocoena phocoena)," *Sci Rep*, vol. 8, no. 1, Art. no. 1, Jun. 2018, doi: 10.1038/s41598-018-28022-8.

[6] A. W. Archer, "World's highest tides: Hypertidal coastal systems in North America, South America and Europe," *Sedimentary Geology*, vol. 284–285, pp. 1–25, Feb. 2013, doi: 10.1016/j.sedgeo.2012.12.007.

[7] Z. Zhou, M. Benbouzid, J.-F. Charpentier, F. Scuiller, and T. Tang, "Developments in large marine current turbine technologies – A review," *Renewable and Sustainable Energy Reviews*, vol. 71, pp. 852–858, May 2017, doi: 10.1016/j.rser.2016.12.113.

[8] D. Tollit, R. Joy, J. Wood, A. Redden, and C. Booth, "BASELINE PRESENCE OF AND EFFECTS OF TIDAL TURBINE INSTALLATION AND OPERATIONS ON HARBOUR PORPOISE IN MINAS PASSAGE, BAY OF FUNDY, CANADA," *Journal of Ocean Technology*, vol. 14, p. 24, Aug. 2019.

[9] J. F. Kaiser, "On a simple algorithm to calculate the 'energy' of a signal," in *International Conference on Acoustics, Speech, and Signal Processing*, Apr. 1990, pp. 381–384 vol.1. doi: 10.1109/ICASSP.1990.115702.

[10] S. Solnik, P. Rider, K. Steinweg, P. DeVita, and T. Hortobágyi, "Teager–Kaiser energy operator signal conditioning improves EMG onset detection," *Eur J Appl Physiol*, vol. 110, no. 3, pp. 489–498, 2010, doi: 10.1007/s00421-010-1521-8.

[11] R. Brigham, E. Kalko, G. Jones, S. Parsons, and H. Limpens, *Bat Echolocation Research. Tools, Techniques and Analysis*. Edited by Bat Conservation International, Austin, TX, 2004.

[12] C. L. Walters, A. Collen, T. Lucas, K. Mroz, C. A. Sayer, and K. E. Jones, "Challenges of using bioacoustics to globally monitor bats," in *Bat evolution, ecology, and conservation*, Springer, 2013, pp. 479–499.

[13] V. Kandia and Y. Stylianou, "Detection of sperm whale clicks based on the Teager–Kaiser energy operator," *Applied Acoustics*, vol. 67, no. 11, pp. 1144–1163, Nov. 2006, doi: 10.1016/j.apacoust.2006.05.007.

[14] D. Stowell, "Computational bioacoustics with deep learning: a review and roadmap," *PeerJ*, vol. 10, p. e13152, 2022.

[15] M. Bittle and A. Duncan, "A review of current marine mammal detection and classification algorithms for use in automated passive acoustic monitoring," *Annual Conference of the Australian Acoustical Society 2013, Acoustics 2013: Science, Technology and Amenity*, pp. 208–215, Jan. 2013.

[16] D. Duan *et al.*, "Real-time identification of marine mammal calls based on convolutional neural networks," *Applied Acoustics*, vol. 192, p. 108755, Apr. 2022, doi: 10.1016/j.apacoust.2022.108755.

[17] D. Duan, "Detection Method for Echolocation Clicks Based on LSTM Networks," *Mobile Information Systems*, vol. 2022, p. e4466037, Mar. 2022, doi: 10.1155/2022/4466037.

[18] P. C. Bermant, M. M. Bronstein, R. J. Wood, S. Gero, and D. F. Gruber, "Deep Machine Learning Techniques for the Detection and Classification of Sperm Whale Bioacoustics," *Scientific Reports*, vol. 9, no. 1, Art. no. 1, Aug. 2019, doi: 10.1038/s41598-019-48909-4.

[19] D. J. Hasselman *et al.*, "2020 State of the Science Report, Chapter 10: Environmental Monitoring Technologies and Techniques for Detecting Interactions of Marine Animals with Turbines," 2020.

[20] D. K. Mellinger, "Acoustic feature extraction and classification in Ishmael," *The Journal of the Acoustical Society of America*, vol. 134, no. 5, pp. 3986–3986, Nov. 2013, doi: 10.1121/1.4830526.

Authorized licensed use limited to: OREGON STATE UNIV. Downloaded on November 22,2026 at 02:44:33 UTC from IEEE Xplore. Restrictions apply.