NOVEMBER 04 2025

North Atlantic right whale detection range performance quantification on a bottom-mounted linear hydrophone array using a calibrated acoustic source

Vincent E. Premus; Philip A. Abbot; Eric Illich; Ted A. Abbot; John Browning; Vitaly Kmelnitsky

J. Acoust. Soc. Am. 158, 3672–3686 (2025)
https://doi.org/10.1121/10.0039669

JASA ARTICLE

North Atlantic right whale detection range performance quantification on a bottom-mounted linear hydrophone array using a calibrated acoustic source

Vincent E. Premus,[a] Philip A. Abbot, Eric Illich, Ted A. Abbot, John Browning, and Vitaly Kmelnitsky
Ocean Acoustical Services and Instrumentation Systems, Inc., a wholly-owned subsidiary of ThayerMahan, Inc., 175 Cabot Street, Suite 400, Lowell, Massachusetts 01854, USA

ABSTRACT:
Experimental results are presented which quantify hydrophone array detection performance for the case of a North Atlantic right whale upcall using a calibrated acoustic projector with GPS reconstruction in the southern New England offshore wind construction area. Measurements of detection range and *in situ* transmission loss are reconciled to produce an empirical figure of merit in the subject environment for both a 32-channel bottom-mounted hydrophone array and single hydrophone. The results reveal a 3.6-fold detection range advantage for the array in this 17log*R* spreading loss environment. A passive sonar equation treatment is also applied to validate the hypothesis that the detection of a North Atlantic right whale upcall is fundamentally a narrowband detection problem, contrary to long-held convention. This finding has important implications for the treatment of noise bandwidth in baleen whale acoustic detection performance modeling generally, and for the extrapolation of such detection performance to new noise environments.
 https://doi.org/10.1121/10.0039669
(Received 15 May 2025; revised 1 September 2025; accepted 6 October 2025; published online 4 November 2025)
[Editor: Shane Guan] Pages: 3672–3686

I. INTRODUCTION

The use of coherently beamformed hydrophone arrays for passive acoustic marine mammal monitoring (PAM) was first introduced by Clark and Gagnon in 1993 after scientists were given access to the U. S. Navy's Sound Surveillance System to investigate its application to the acoustic tracking of low-frequency baleen whales.[1] Originally deployed on the seabed to leverage the existence of the deep sound channel to detect adversary submarines, that technology was demonstrated to detect and track a singing blue whale at basin scales.[2] More than three decades since that breakthrough, high-spatial resolution hydrophone arrays hosted on autonomous maritime systems are on the threshold of becoming the benchmark for long-range passive acoustic monitoring and protection of endangered species, such as the North Atlantic right whale (NARW).[3]

The frequency band below 500Hz, where many baleen whale vocalizations occur, tends to be dominated by anthropogenic noise due to commercial shipping and, more recently, offshore wind construction activity. Such noise sources drive an ambient noise spectrum that is strongly anisotropic and significantly elevated relative to naturally occurring ambient noise mechanisms, such as wind-wave interaction and biologics. In the presence of such noise, arrays offer two distinct advantages over the single hydrophone: (1) spatial noise rejection, which increases detection sensitivity by improving signal-to-noise ratio (SNR); and (2) spatial resolution, which delivers the capacity to resolve bearing and track vocalizing individuals.[3–5] Once a cost-prohibitive and power-demanding solution deployable only from crewed vessels, high-spatial resolution passive acoustic arrays are now routinely deployed and operated from both mobile and fixed autonomous maritime platforms.[3,6,7]

The proliferation of advanced instrumentation and methods for passive acoustic monitoring during offshore wind construction has also prompted a reexamination of the treatment of performance metrics and modeling in the bioacoustics community. Such efforts are important as they can inform policy and permitting decisions implemented by government regulatory bodies such as the U.S. Bureau of Ocean Energy Management, which has administrative oversight responsibility for offshore wind leases in U.S. coastal waters, and the National Oceanic and Atmospheric Administration, which oversees compliance with the Marine Mammal Protection Act. The traditional approach to detection performance modeling is based on the passive sonar equation, informed by assumptions or measurements of the important environment variables that drive performance, e.g., source level, noise level, and transmission loss (TL), as well as system-specific variables, such as recognition differential. All these quantities, even recognition differential, exhibit statistical variability and as such should be treated as random variables. The sonar equation represents these variables in the mean. Despite this limitation, the sonar equation is an intuitive and powerful methodology that is very useful when employed in a rigorous manner. It is thus vital that performance metrics and models are based on accurate

[a]E-mail: vpremus@thayermahan.com

https://doi.org/10.1121/10.0039669

assumptions and common definitions, and validated with calibrated, ground-truthed observations of system performance whenever possible.

Recent public forums, such as the Whale Detection Technology Evaluation Workshop series organized by the Regional Wildlife Science Collaborative in 2024 and reported on by the Pacific Northwest National Laboratories in Szesciorka *et al.*,[8] have called attention to the need to formalize and standardize assumptions and metrics used to evaluate passive acoustic monitoring technologies. The Pacific Northwest National Laboratories report was particularly helpful in summarizing the challenges involved, although notably absent from the report was acknowledgement of the potential of coherently beamformed acoustic array technology to address the PAM problem. It is evident from these deliberations that improved clarity and commonality are needed regarding a number of issues, including but not limited to: (1) the characterization of NARW upcall root mean square (rms) source level; (2) the treatment of in-band noise level relevant to the detection decision; (3) the specification of system recognition differential; (4) the clear definition of probability of detection vs range (sometimes called detection function) and detection range, the range at which detection probability decays to 50%; and (5) methods for their measurement using calibrated sources with precise positional reconstruction. Further, metrics, such as recognition differential, detection function, and detection range, should be accompanied by a concurrent statement of false alarm probability or false alarm rate to be unambiguously specified. With a commitment to clearly defined metrics and the use of calibrated, ground-truthed measurements, the performance of PAM systems can be rigorously quantified.

The characterization of NARW upcall source level is perhaps the performance modeling input with the greatest need for clarification. Reported source level estimates in the open literature vary widely, ranging from 152 dB re 1 μ Pa@1 m (Ref. 9) to 172 dB re 1 μ Pa@1 m (Ref. 10). The rms designation is not always explicitly stated, but we will assume that it is always implied due to the broadband nature of the signal of interest. The source level estimate due to Palmer *et al.*[9] suffers from uncertainty associated with localization error due to receiver clock synchronization, which influences backpropagation of measured received level to source level, as well as the use of an overly favorable spreading loss range dependence, 14.5logR, that appears to underestimate TL in the New England continental shelf environment, as will be shown in Sec. III. Others, such as Trygonis *et al.*,[11] who found a source level estimate of 155 dB re 1 μ Pa@1 m, suffer from low sample size. On the other hand, Clark *et al.* reported a mean root mean square source level of 172 $\pm$ 6.8 dB re 1 μ Pa@1 m in a frequency band from 71 to 224 Hz, based on a sample set of 100 upcalls.[12] The following year, using a sample set that was significantly increased to 353 upcalls, the same group reported an average rms source level of 165 $\pm$ 3.5 dB re 1 μ Pa@1 m, but in a different (smaller) frequency band from 142 to $-$179 Hz.[13] The main takeaway is that, across all methods for source level inversion, attributes such as frequency band, sample size, range estimation, and TL model vary widely. Modeling convention at the present time leans in the direction of more conservative source level estimates for the NARW upcall. Consequently, in this study, an rms source level of 160 dB re 1 μ Pa@1 m was employed, partly due to the constraints on the calibrated acoustic projector and partly due to restrictions on active acoustic transmissions imposed by government regulatory agencies.

There have been several recent efforts aimed at the quantification of NARW upcall detection performance using single hydrophones[14,15] or multi-hydrophone systems that do not employ array gain.[16] Estabrook *et al.*[14] are notable for using a 17 $\log_{10}R$ TL model in their study of right whale distribution using single hydrophones in the southern New England lease area. This spreading loss coefficient is consistent with that measured herein in the Revolution Wind lease area in September 2024, as reported in Sec. III. One limitation of that study and system performance assessment was the omission of recognition differential, or minimum SNR, from their sonar equation treatment. This leaves open the possibility that detection range estimates may have been overestimated by failing to account for the performance cost of imposing a minimum SNR for reliable detection. Johnson *et al.*[15] have also reported on detection range estimates from both a moored buoy and Slocum glider instrumented with single hydrophones. In that work, range ground truth was developed using an experimental technique that relies on backpropagation of acoustic normal modes detected at a vertical line array using modeled normal mode group velocities. Ground truth range uncertainty in that study was estimated to be on the order of 1 km. For the work presented herein, performance metrics emphasize reliance on a calibrated U.S. Navy Underwater Sound Reference Laboratory type J-13 acoustic projector transmitting actual, prerecorded upcalls with GPS position reconstruction, rather than on a natural-occurring source of unknown source level and unknown position information. The calibrated source also supports the direct measurement of *in situ* TL, thereby minimizing the reliance on modeling of spreading loss coefficient to infer a system figure of merit (FOM). The methodology for inferring detection range will follow that of Premus *et al.*,[17] which relied on the reconciliation of a measured probability of detection vs range curve and measured TL to yield an empirically determined FOM for a desired receiver detection sensitivity.

The structure of the paper runs as follows: Section II begins with an overview of the system and sensor supporting the experiment and a description of the environment and calibrated source operations; Sec. III outlines the measurements, including (1) in-band noise level, (2) detection function, and (3) TL and FOM; Sec. IV outlines the passive sonar equation detection performance model, with emphasis on the impact of noise bandwidth on predicted performance, showing the detection contours corresponding to narrowband (12 Hz) and broadband (360 Hz) in-band noise model assumptions, respectively; finally, in Sec. V, we summarize

the key findings and present some directions for future work, highlighting the importance of correctly accounting for noise bandwidth in the projection of performance to new noise environments such as pile driving.

II. EXPERIMENT OVERVIEW

A. The sensor: Bottom-mounted horizontal line array

The passive acoustic sensing system used to support this work was the ThayerMahan SeaPicket system,[6] comprised of a 32-channel, bottom-mounted hydrophone array tethered to a surface buoy instrumented with an embedded digital signal processor (DSP), satellite communications modem, batteries, and photovoltaic cells. Figure 1(a) shows the SeaPicket surface expression after deployment. Figure 1(b) depicts the sensor, a 32-channel, low-power, hydrophone array built by Raytheon Missiles and Defense (Portsmouth, RI). Hydrophones are uniformly spaced at one half-wavelength for a design frequency of 625 Hz, or 1.2 m spacing, and 37.2 m total aperture length. The hydrophones are piezoelectric crystals with a sensitivity of −199 dB re 1 V/μ Pa and pre-amp gain of 10 dB. An analog-to-digital converter is integrated into each channel, and hydrophone response is digitized with 24-bit precision at a sample rate of 2.5 kHz. Hydrophone and pre-amp power draw is approximately 30 mW/channel. The array also includes high-precision, non-acoustic sensor modules uniformly distributed forward, mid, and aft, that measure array heading, pitch, and depth. An array receiver, or node card, converts array telemetry to Ethernet User Datagram Protocol packets for transmission to the embedded DSP. The total power draw for the acoustic sensor (array plus receiver electronics) is less than 2 W.

(a)

(b)

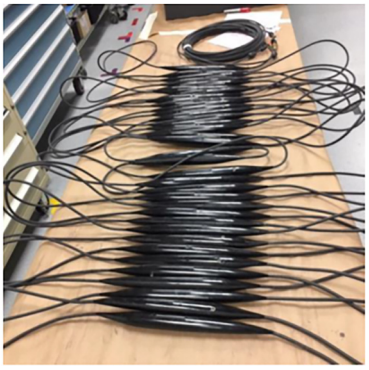

FIG. 1. (a) Surface expression of the SeaPicket system comprised of a surface buoy instrumented with an embedded DSP, satellite communications modem, batteries, and photovoltaic cells; and (b) 32-channel, bottom-mounted hydrophone array as viewed in the laboratory at Raytheon Integrated Defense Systems, Portsmouth, RI.

The embedded DSP that forms the foundation for the SeaPicket system is essentially the same as that of the ThayerMahan Outpost system,[3] a real-time, C-language, passive sonar processing architecture implemented on a Linux-based, 64-bit embedded computing platform. The embedded platform of the SeaPicket is comprised of a Toradex Viola carrier board, which hosts a Colibri iMX8 system on chip[18] featuring a quad-core Cortex A7 processor with 2 GB of DDR3L RAM that is optimized for low power consumption; the power draw of this embedded processor is about 2 W. Array element data are recorded on a solid-state hard drive with at least 1 TB of storage capacity, enough to archive nearly 100 days of continuous array element data at the 2.5 kHz sample rate.

The real-time passive sonar processing architecture of the SeaPicket system has been described in Premus *et al.*,[3] and so, will only be briefly overviewed here. The processing flow consists of data conditioning and sensor health monitoring followed by a fast Fourier transform (FFT), which transforms the element time series to the frequency domain. Spatial filtering is performed using a frequency domain conventional beamformer that samples beamspace on a uniformly spaced cosine grid in relative bearing. Beam response is normalized to ensure distortionless response to a plane wave input signal model. Element data are Hanning shaded for reduced sidelobe response. All hydrophones are calibrated, and hydrophone sensitivity and preamp gain are used to support real-time reporting and analysis of ambient noise levels and received levels of signals of interest in absolute units of received level (dB re 1 μ Pa^2 in a specified analysis frequency band). Following the beamformer, a broadband integration operation is performed to produce a detection surface that reports the distribution of acoustic energy as a function of relative bearing and time.

Detection and classification are performed using a variant of the spectrogram correlator algorithm first introduced for baleen whale classification by Mellinger and Clark,[19] that is programmed for the North Atlantic right whale upcall. The spectrogram correlator employed herein is a computationally efficient implementation based on a "binarized" spectrogram that compares a candidate spectrogram feature to members of a "kernel library" to produce a confidence or similarity score. The binarized nature of the calculation reflects the fact that both the spectrogram feature and correlation kernel have been thresholded, or detected, relative to a local background noise estimate to yield a binary image; the resulting confidence score is a count of the number of time-frequency pixels the two binary images share. The implementation follows the earlier work of Abbot *et al.*,[20,21] which successfully applied this approach

https://doi.org/10.1121/10.0039669

to the detection and classification of humpback whales. The spectrogram correlator method is particularly well suited to the detection of right whale upcalls due to their very repeatable nature. The correlation kernel library used herein was developed from training data sampled from the well-known 2013 St. Andrews Detection, Classification, Localization, and Density Estimation Workshop database,[22] a compilation of curated, ground-truthed exemplars made openly available to the bioacoustics community for classifier development and performance quantification and comparison. A set of seven kernels were selected to span the modest amount of variability observed in measured upcall characteristics, e.g., bandwidth, start frequency, end frequency, and frequency rate. Confidence thresholds were programmed for a desired false alarm rate of 1 per hour using signal present and signal absent data segmented from the training database used to develop the kernel library. Finally, it should be noted that the binarized spectrogram correlator runs in real-time, processing 32 beams concurrently at a frequency resolution of 3.9 Hz and an update rate of 64 ms.

The performance of the ThayerMahan NARW upcall spectrogram correlator has been quantified using 72 h each of signal present (2767 exemplars) and signal absent test data from the St. Andrews database: again, test data held apart from the training data used to select library members and train confidence thresholds. Figure 2 shows the performance of the spectrogram correlator in the form of a receiver operating characteristic curve that plots false alarm rate vs probability of correct classification (Pcc) compared with reported results for the best-performing machine learning model described in Shiu *et al.*,[23] recognized by many in the bioacoustics community as state-of-the-art for upcall classification at the present time.

It is clear from Fig. 2 that the spectrogram correlator performance compares favorably with that of the machine learning model, yielding an instantaneous Pcc of 0.64 at a programmed false alarm rate of 1/h. It is worthwhile noting that if a whale is assumed to call multiple times in an interval of several (e.g., 3–15) minutes, the probability of a detection event is much higher than the instantaneous Pcc would suggest. For example, if the instantaneous probability of missing a single upcall is 1 – Pcc, or 0.36, and each occurrence may be viewed as an independent event, the probability of missing three upcalls in succession is reduced to 0.046 (e.g., 0.36^3). The probability of detection of a vocalizing individual is then 1 minus the probability of a miss, 1 – 0.046, or 0.954. This observation has been noted by others,[15] and should be considered when evaluating the efficacy of an algorithm where it is reasonable to assume that multiple calls can be expected to occur within a given interval.

B. Environment description and source operations

The acoustic measurements that are the subject of this analysis were conducted at a shallow water site off the southern coast of Massachusetts on September 4, 2024. The

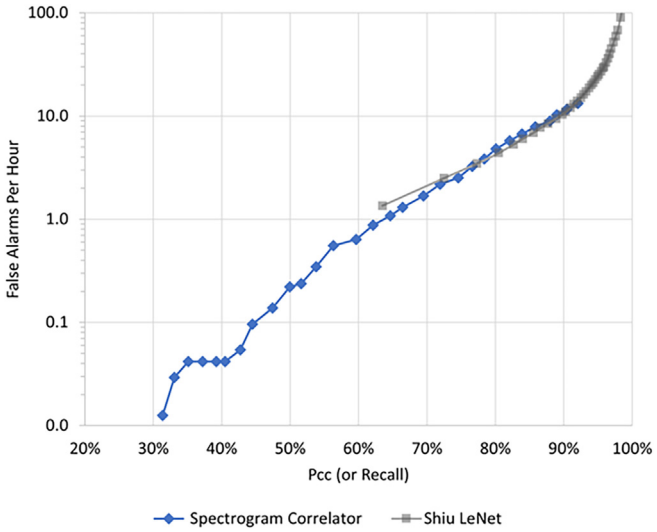

FIG. 2. Performance of the NARW upcall spectrogram correlator in the form of a receiver operating characteristic curve compared with that of the machine learning model of Shiu *et al.* (Ref. 23).

support vessel was the *R/V Blackhawk*, a 54-foot twin engine coastal vessel operated by ThayerMahan. Figure 3(a) shows a bathymetric map delineating the boundary (black) of the Revolution Wind lease area, approximately 30 km southwest of Martha's Vineyard.[24] Bathymetry in the op area is shallow and gently sloping from north to south, with water depths ranging from 30 m at the northern boundary to roughly 50 m at the south. The SeaPicket system, known internally as Goshen, was one of four bottom-mounted arrays deployed in the May-June timeframe in support of a passive acoustic monitoring program during wind farm construction. The network of four arrays performed very well during the 3-month construction period. A preliminary report of real-time results showed thousands of baleen whale detections and localizations during the 3-month deployment period.[25] In one notable example, a humpback whale was detected during active pile driving on two arrays concurrently, at ranges of 2.7 and 21 km, respectively, yielding a localization solution more than 14 km outside of the pile driving exclusion zone with sufficient precision to avoid an unnecessary shutdown of construction activity.[25,26]

The location of Goshen at the western boundary of the lease area at 41 N 7.502′, 71 W 15.156′, is denoted with a red solid circle in Fig. 3(b). The array was deployed with a NE/SW orientation of 75°/255°. Source operations were conducted on a northwest radial relative to the array phase center to preserve an approximate broadside aspect to the array, while remaining outside the designated lease area boundary. The support vessel conducted active source transmissions at range standoffs of 1, 2, 4, 6, 7, and 8 NM (or, equivalently, 1.9, 3.7, 7.4, 11.1, 13.0, and 14.8 km) using a U.S. Navy Underwater Sound Reference Laboratory calibrated projector, type J-13.[27]

The J-13, shown in Fig. 4(a), is an electro-acoustic piston source. It weighs about 150 lbs in-air and thus requires the use of a hydraulic winch, and a davit or A-frame, for

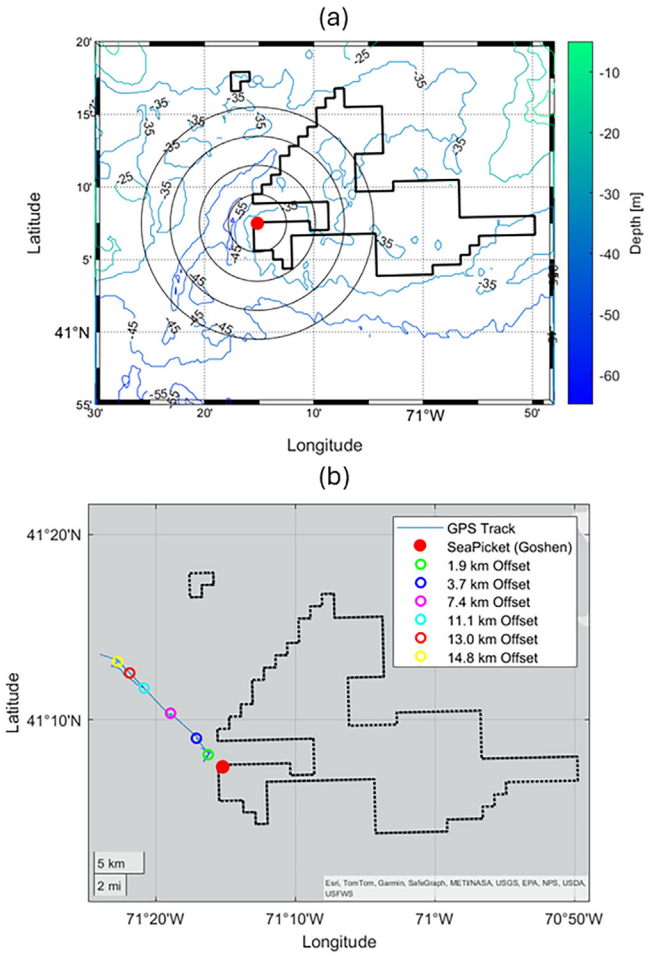

FIG. 3. (a) Revolution Wind lease area in southern New England, the site of the bottom-mounted array experiment, overlaid on bathymetric map. (b) Location of source operation range offsets. The 32-channel hydrophone array located at 41 N 7.0 71 W 15.5 is denoted by the red solid circle. Coordinates of the support vessel deploying the J-13 source at range offsets of 1.9, 3.7, 7.4, 11.1, 13.0, and 14.8 km are indicated by the green, blue, magenta, cyan, red, and yellow open circles, respectively.

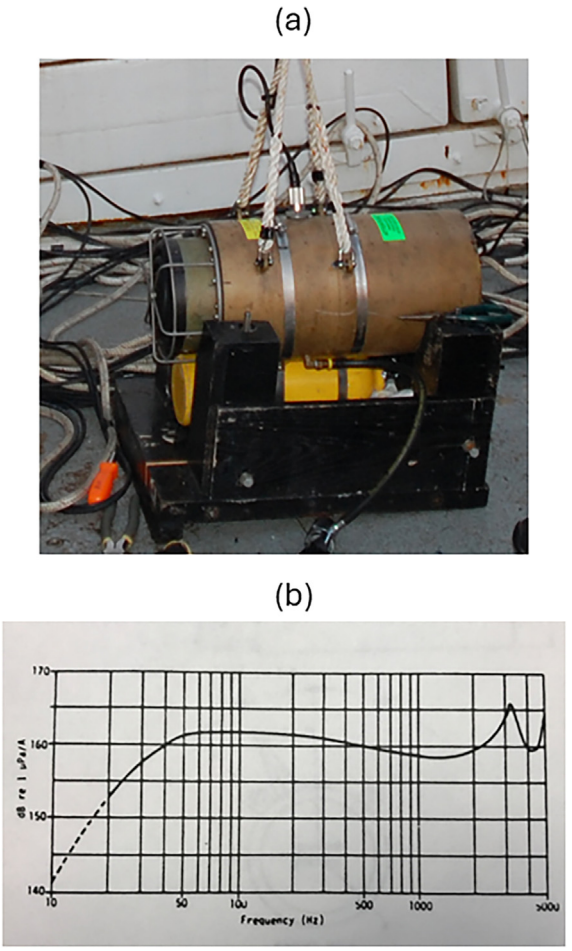

FIG. 4. (a) J-13 acoustic projector with pressure compensation. (b) J-13 transmit voltage response curve (Ref. 27). The nearly flat frequency response from 40 Hz to 2 kHz attests to the suitability of the J-13 for the transmission of North Atlantic right whale upcalls.

deployment. A bottle of compressed air is required for pressure compensation of the diaphragm at deep deployment depths. The J-13 was selected as the primary acoustic projector as it has omni-directional spatial response and excellent frequency response at low frequency. The transmit voltage response curve of the J-13, depicted in Fig. 4(b), shows flat response down to 50 Hz, which is well suited to the support band of the NARW upcall, generally understood to be approximately 40–400 Hz.[28] At each range offset, the J-13 transmitted a wave file comprised of a single representative upcall excerpted from the 2013 St. Andrews Detection, Classification, Localization, and Density Estimation database looped at a constant rep rate of 12 per minute. At each location, it was deployed mid-water column at a depth of 16 m for a period of approximately 30 min, resulting in roughly 300–400 detection opportunities per range offset. Acoustic transmissions were continuously monitored with a calibrated reference hydrophone positioned 1 m away from the sound source center of pressure, and a MATLAB script was employed to record the hydrophone telemetry and compute time domain rms source level in real time. This is essential to verify source level in support of *in situ* TL measurement and quantification of detection performance, as well as to ensure compliance with government regulations concerning active acoustics. The upcall was transmitted at a source level of 160 dBrms re 1 μ Pa@1 m.

Figure 5(a) depicts an excerpt of a subband peak energy detection surface, or bearing-time record (BTR), showing broadband acoustic energy measured at the array as a function of relative bearing and time for the 30-s interval starting at 133 910 GMT on September 4, 2024. Acoustic energy associated with the upcalls transmitted by the J-13 are evident from the sequence of dark spots observed at the expected 5-s repetition rate at a cosine relative bearing of approximately 0.6, which is consistent with the array orientation and GPS ground truth reconstruction of the support vessel. The green squares overlaid on the detection surface denote real-time detection decisions of the autonomous NARW upcall auto-detector running on the embedded processor and reported via Immarsat Broadband Global Area Network satellite during the deployment. The convention for time registration is such that the timestamp reported by the auto-detector is aligned with the onset of energy from

https://doi.org/10.1121/10.0039669

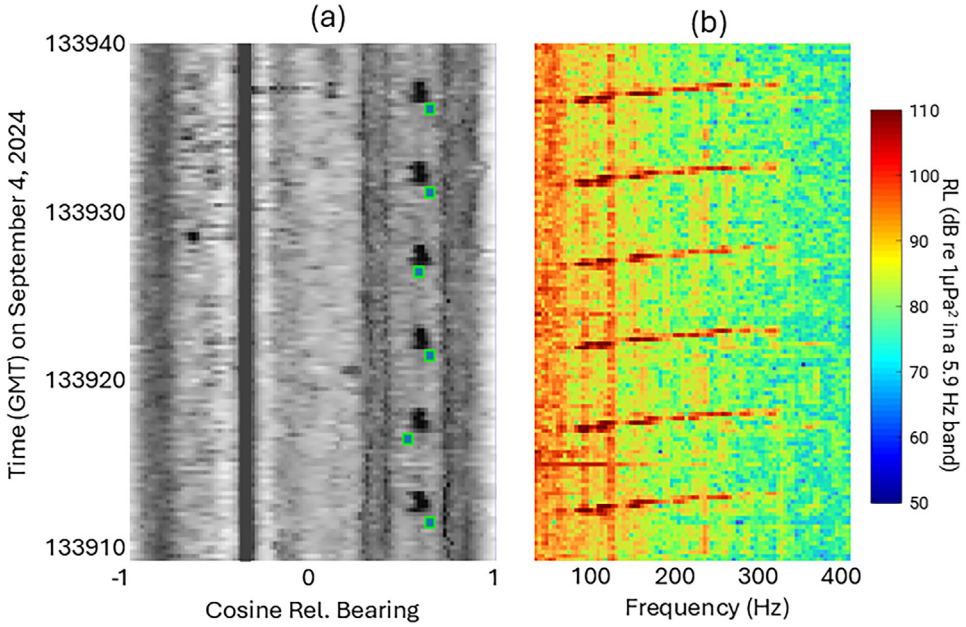

FIG. 5. (a) Subband peak energy detection surface, or BTR, depicting broadband acoustic energy measured at the array as a function of relative bearing and time for the 30-s interval starting at 133 910 GMT on September 4, 2024. The dark spots appearing at 5-s intervals at a cosine relative bearing of approximately 0.6 correspond to the energy detection of transmissions from the J-13. The green squares overlaid on the detection surface reflect the positive identification of these transmissions as NARW upcalls as reported by the real-time auto-detector running on the embedded DSP. During this period, the range to the J-13 is approximately 2 km, so detection coverage is 100% (as expected). The support vessel, R/V Black Hawk, has its engines idling at low RPM to minimize radiated noise contamination of the J-13 transmissions: steady state energy from the support vessel is faintly visible in the intervals between the upcall detections. At least five dark gray/black traces are also visible in the BTR corresponding to nearby surface vessels transiting the op area. These appear as vertical lines on the detection surface with little or no bearing rate due to the short observation interval. (b) Spectrogram of the beam response at cosine relative bearing of 0.6 illustrating the evidence for the correct classifier decisions, or NARW upcalls detected at the expected 5-s repetition rate.

the detected upcall. Notice that there is a small amount of bearing fluctuation, or bearing error, observed in the overlay of the real-time auto-detector output on the energy detection surface. This may be attributed to multiple factors, including miscalibration of array orientation, coherent multipath interaction with the beamformer beam response, and the impact of ambient noise on "best beam" selection by the auto-detector. For the array employed herein, the 3 dB down beam width at the cut frequency of 625 Hz is 3.5°.[3] Theoretically, beam resolution decreases by a factor of 2× for every octave down in frequency, so the 3 dB beam width at the mid-point of the NARW upcall support band, 150 Hz, is about 14°. However, when the array orientation has been properly calibrated, the mean beam pointing error typically measures less than 4° with a standard deviation of 2°–3°.

Radiated noise from *R/V Black Hawk*, with its engines idling at low RPM to avoid contamination of the source transmissions, is faintly detectable above the ambient background in the intervals between the transmitted upcalls. At least five dark gray/black traces are visible in the BTR corresponding to nearby surface vessels transiting the op area during this time interval. This illustrates the capacity of the array to spatially resolve transmissions from the J-13 on the support vessel from other nearby vessels in the area. This interferer density was characteristic of the noise environment encountered throughout the experiment. Last, Fig. 5(b) shows the spectrogram of the beam response at cosine relative bearing of 0.6 illustrating the evidence for the correct classifier decisions, or NARW upcalls detected at the expected 5-s repetition rate.

Meteorological conditions during the data collection were calm, with winds 0–5 knots and an estimated sea state of SS0: visual observations of the sea surface reported glass calm conditions. A conductivity-temperature vs depth cast performed during the deployment is shown in Fig. 6. The resulting sound speed stratification was typical of late summer New England: shallow 10 m isovelocity mixed-layer overlaying a strongly downward refracting thermocline from 10 to 30 m with a sound speed gradient of about 2.0 s^{-1}. With the J-13 fixed at 16 m, both the source and array were deployed within or below the thermocline.

Source depth dependence was not studied during this test due to limitations on ship time and resources. While source depth can play a role in TL in certain environments, it is unlikely to be a significant factor in the frequency band of interest to NARW upcall detection for the sound speed stratification shown in Fig. 6. If the mixed layer in Fig. 6 was sufficiently upward refracting to support a true surface duct, the duct thickness of approximately 10 m would not be enough to cause trapping of the low frequencies associated with NARW upcall. Following Urick,[29] the relationship between cutoff frequency and surface duct thickness is given by the following:

$$H = 36\lambda^{1/2}, \tag{1}$$

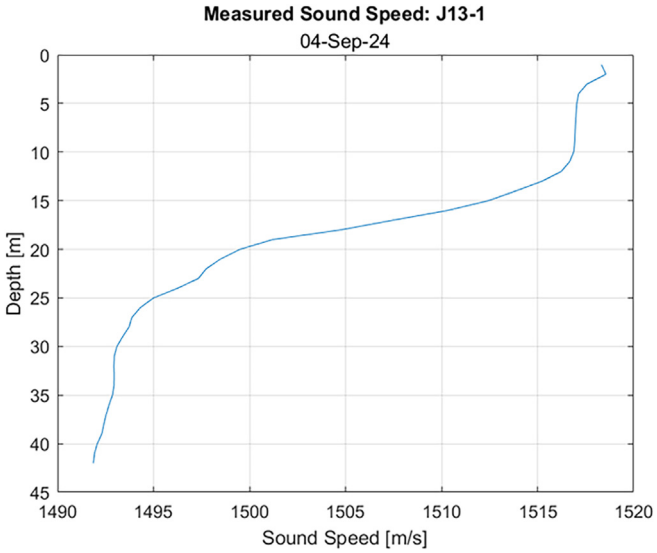

FIG. 6. Measured sound speed profile calculated from conductivity-temperature vs depth cast performed on September 4, 2024, showing downward refracting propagation environment with sound speed gradient of approximately $2\,s^{-1}$.

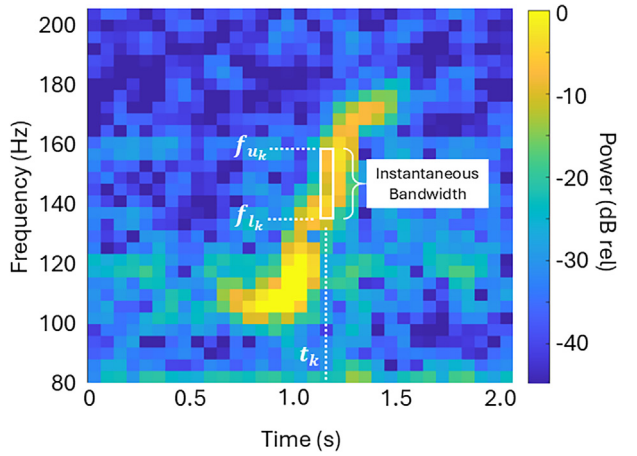

FIG. 7. Spectrogram of a North Atlantic right whale upcall exemplar from St. Andrews 2013 database processed with 0.25 s FFT, 75% overlap, and Hanning window. While the support band for this representative upcall is roughly 100–180 Hz, the instantaneous bandwidth of the upcall as a function of time is observed to vary from approximately 5 Hz (at start and finish) to almost 30 Hz (in the middle), depending on which FFT frame (vertical slice) is sampled.

where H is the surface duct thickness in feet and λ is the maximum wavelength trapped by the duct in feet. For the duct thickness of approximately 40 feet shown in Fig. 6, this corresponds to a cutoff frequency of about 4.2 kHz. Thus, frequencies below the cutoff will not be trapped by such a duct overlaying downward refracting conditions. Other experimental studies of TL measured in continental shelf environments have also observed the limited depth dependence in mean TL (<2 dB) during similar sound speed stratification conditions.[30]

III. PERFORMANCE MEASUREMENT

A. Noise bandwidth and in-band noise level

Measurements of ocean acoustic ambient noise are typically reported as broadband levels, most often in octave or third octave bands.[31–33] For example, Van Parijs *et al.*[33] recently published a thorough and comprehensive summary of 2 years of noise measurements to baseline the acoustic environment of the southern New England offshore wind lease area. Median seasonal third octave band sound pressure levels were reported at a number of sites, including two denoted COX01 and COX02 that were very close to the location of the September 2024 source op, that showed third octave band levels in the middle of the NARW upcall support band (e.g., 150 Hz) of 90–95 dB re 1 μ Pa. Broadband noise descriptions are important to the quantification of exposure levels and studies of the behavioral impact of noise from shipping and construction activity on marine mammals. Alternatively, in the classic 1962 benchmark paper, Wenz[34] summarized decades of acoustic ambient noise observations into regimens of seismic, shipping, sea state, bubbles, and molecular agitation corrected to spectrum level, or noise spectrum density, in units of dB re 1 μ Pa//Hz, a narrowband description that enables the aggregation of the many different ocean noise mechanisms into a common unified framework. As an example, Wenz[34] associates moderate-to-heavy shipping noise at a frequency of 150 Hz with a spectrum level of 80 dB re 1 μ Pa//Hz, which is consistent with the third octave band observations of Van Parijs *et al.*,[33] once the correction has been made for measurement bandwidth (i.e., 95 dB – $10\log_{10}$). It is well known that, for detection functions that involve the incoherent broadband integration of acoustic energy over some finite bandwidth, BW, narrowband noise spectrum level requires a correction of $10\log_{10}BW$ to obtain the equivalent in-band noise level (assuming the noise level is relatively constant across the integration band).

A key premise underlying the experimental investigation of NARW upcall detection performance presented herein is the hypothesis that, contrary to convention, upcall detection is fundamentally a narrowband, rather than broadband, detection problem. That is, the noise bandwidth employed in the passive sonar equation model should be linked to the instantaneous bandwidth of the upcall, not the full support bandwidth of the upcall, which can cover a frequency range of up to 40–400 Hz.[28] Figure 7 depicts a spectrogram of a representative upcall from the St. Andrews database. The frequency support for this upcall exemplar spans approximately 100–180 Hz, or 80 Hz in total bandwidth. The upcall duration is roughly 0.75 s. The digital signal processing supporting the right whale detector-classifier string employs a 640-sample (i.e., 250 ms) FFT, which yields a frequency bin spacing of 3.9 Hz, or window corrected bin resolution [e.g., effective noise bandwidth (ENBW)] of 5.9 Hz. At an overlap percentage of 75%, the corresponding FFT frame rate, dT, is 0.0634 s. Depending on the SNR, there are typically 12–16 FFT frames per upcall.

Given the bandwidth of the upcall depicted in Fig. 7, it is tempting to characterize this detection problem as one of a broadband nature. However, such an interpretation fails to differentiate the full upcall support band from the noise bandwidth that directly impacts the detection decision. In

https://doi.org/10.1121/10.0039669

the case of the spectrogram correlator, the detection of the candidate spectrum feature, and the computation of the kernel library correlation score, is performed at the beamformer output in a sliding window of duration long enough to capture the longest kernel in the library; this is done concurrently for all beams. This analysis window advances through the beamformed data one FFT frame at a time. The presence of the upcall in any given analysis window is not determined from the entire upcall support bandwidth at once, as in the case of a matched filter detecting the presence of a known *a priori* waveform replica,[35] but rather for the spectrogram pixels that support the detection event. To support a detection event, a spectrogram pixel must (1) exceed the binarization threshold of 5.5 dB and (2) fall within the boundary of one of the correlation kernels comprising the upcall library. The example in Fig. 7 shows that, for a vertical spectrum slice at time, t_k, the minimum (lower) frequency of the upcall is denoted by f_{l_k} and the maximum (upper) frequency is denoted by f_{u_k}. The difference $f_{u_k} - f_{l_k}$ defines the instantaneous bandwidth of the feature at time t_k. In this example, the number of frequency bins spanning any given spectrum slice varies from as little as 2 to as many as 8, corresponding to a range of instantaneous bandwidth of 8–32 Hz at the given FFT frequency resolution.

The signal power in each FFT frame is computed by integrating the power over all pixels that exceed the minimum SNR threshold (5.5 dB) in that frame. The average received level, L_R (and in-band noise level as discussed below) for the upcall detection event is then computed by averaging the instantaneous power over the number of FFT frames, K, spanning the upcall, as follows:

$$L_R = 10\log_{10}\left[\frac{1}{K}\sum_{1}^{K}\sum_{f_{l_k}}^{f_{u_k}}\left|S_k(f)\right|^2\right], \tag{2}$$

where $S(f)$ is the frequency domain representation of the signal, $s(t)$. It follows that the mean instantaneous upcall bandwidth (MIBW) is given by the following:

$$MIBW = \frac{1}{K}\sum_{k=1}^{K} f_{u_k} - f_{l_k}. \tag{3}$$

The MIBW may also be determined by dividing the total number of signal pixels in the upcall by the number of FFT frames spanning the upcall. It will be shown in Sec. IV that it is the MIWB that most closely defines the in-band noise bandwidth that drives the detection decision. For the example of Fig. 7, there are 60 total pixels (yellow) that exceed the SNR threshold of 5.5 dB and 14 FFT frames, resulting in a mean instantaneous bandwidth of 4.2 bins, or 16.4 Hz.

For any physical interpretation of detection performance to be accurate, it must follow that the associated in-band noise level be defined in the same exact manner as that of the detected signal. Thus, for each upcall detection, in-band noise level is computed in the same manner as received level using Eq. (2), with the difference being that the supporting time-frequency cells, or pixels, over which the calculation is performed are taken from FFT frames of spectrogram data immediately preceding and succeeding the upcall detection by ± 3 s, respectively. The actual noise level used is the minimum of the two noise values reported in order to minimize bias due to noise outliers or transients. This assertion is a departure from sonar equation performance modeling presented in the past, wherein the noise bandwidth is taken to be the full support band of the NARW upcall, e.g., 71–224 Hz in the case of Estabrook *et al.*[14] and 50–225 Hz in the case of Palmer *et al.*[16] Even the present authors themselves employed a noise bandwidth of 50–250 Hz in Premus *et al.*[3] It will be demonstrated in Sec. IV that this broadband convention for defining in-band noise level underpredicts detection performance. Only the "narrowband" detection model tied to the MIBW correctly explains the measured detection performance obtained using calibrated, GPS ground-truthed source operations performed on September 4, 2024.

B. Validation of frequency-domain measured received level

The capacity of the frequency domain representation of the detected upcall to accurately represent the true received level of the signal of interest was calibrated through comparison with measurements of rms source level made at a 1-m

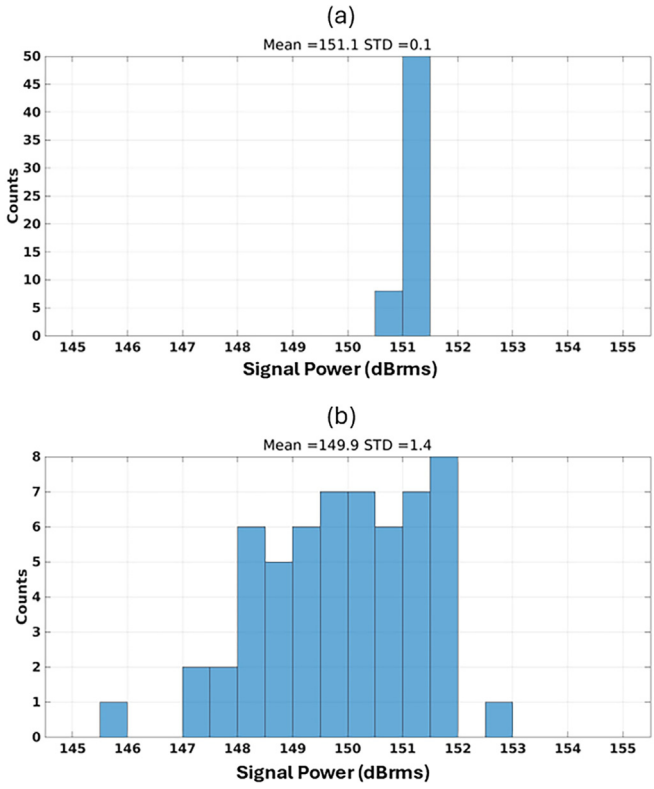

FIG. 8. Histograms of (a) rms source level measured in the time domain as defined in Eq. (3) and (b) rms received level measured in the frequency domain as defined in Eq. (2), for a sample size of 59 transmitted upcalls as measured on the J-9 1-m reference hydrophone in Gloucester Harbor, MA in June 2024.

reference hydrophone using a Underwater Sound Reference Laboratory J-9 acoustic projector in Gloucester Harbor in June 2024. In that experiment, the J-9 transmitted the same representative upcall at a source level of 150 dBrms re 1 μ Pa@1 m. The ground truth reference for rms source level, L_S, is computed from the 1-m reference hydrophone using the following time domain definition:

$$L_s = 10\log_{10}\left[\frac{1}{T}\sum_{1}^{K} s^2(k)\right], \tag{4}$$

where $s(k)$ is the discretized signal of interest and T is the signal duration.[34] Figure 8 depicts two histograms compiled for (1) rms source level measured in the time domain as defined in Eq. (3) and (2) rms received level measured in the frequency domain as defined in Eq. (2), for a sample size of 59 transmitted upcalls. The time domain calculation of source level measured at the 1-m reference hydrophone is very precise, reporting a mean source level of 151.1 dBrms re 1 μ Pa@1 m and a standard deviation, σ, equal to 0.1 dB. The frequency domain calculation compared well although with somewhat more spread, reporting a mean source level of 149.9 dBrms re 1 μ Pa@1 m at a σ equal to 1.4 dB. The greater spread is attributed to the fact that the transmitted upcall does not agree precisely with any of the individual kernel library members (this is by design as it will almost certainly be the case in practice for naturally occurring upcalls), resulting in a minor degree of signal model mismatch at the output of the spectrogram correlator operation. The key takeaway, however, is that the rms received level measured in the frequency domain can be said to be calibrated against the rms received level calculated in the time domain at the output of the 1-m reference hydrophone. It follows that the pixels in the spectrogram determined to support the upcall detection event, i.e., those exceeding the binarization threshold and matching at least one kernel library member, characterize the MIBW of the detected upcall, and by extension the noise bandwidth driving the detection decision.

C. Mean instantaneous upcall bandwidth

Analysis of the J-13 measurements of September 4, 2024, as a function of range reveals that, consistent with expectations, the MIBW observed at the receiver is strongly dependent on SNR. Table I summarizes the statistics of upcall bandwidth measured at the output of the beamformer as a function of range over the sample size of 1500 upcall exemplars. With two minor exceptions, the average number of spectrogram pixels and average number of FFT frames per detection event decreases monotonically with range, and thus SNR. At a bin resolution of 3.9 Hz, the MIBW is observed to decrease from 6.8 bins (26.5 Hz) at 1.9 km to 1.9 bins (7.4 Hz) at 13 km.

TABLE I. Measured NARW upcall MIBW vs range measured from J-13 (1500 transmissions) on September 4, 2024.[a]

	1.9 km	3.7 km	7.4 km	11.1 km	13 km
Average signal pixels per detection	92.0	85.1	48.7	51.5	15.6
Average FFT frames per upcall	13.6	14.5	13.1	12.2	7.2
MIBW (bins)	**6.8**	**5.9**	**4.2**	**4.2**	**1.9**
SD instantaneous bandwidth (bins)	0.5	0.6	0.6	0.6	0.8

[a]MIBW is determined by dividing the average total pixels by the average number of FFT frames taken over all detection events at each range offset. As expected, the MIBW is observed to decrease with range consistent with the reduction in SNR with range. For the ranges bracketing the measured R_{50} for the array (11 and 13 km, highlighted in green), the MIBW is observed to decrease from 4.2 to 1.9 bins, further reinforcing the finding that an MIBW of 3 bins or 12 Hz corresponds to an NRD of 5.5 dB.

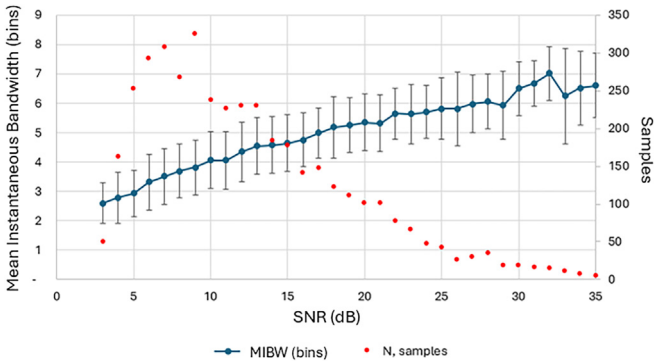

FIG. 9. North Atlantic right whale upcall MIBW vs SNR (blue) with $\pm 1\sigma$ confidence intervals for a sample of 4090 upcalls ranging in SNR from 3 to 35 dB. Sample size at each SNR is denoted in red. Bandwidth is reported in bins of 3.9 Hz width. Observe that the MIBW corresponding to a SNR of 5.5 dB, the NRD of the spectrogram correlator, is 3 bins or 12 Hz. This is the noise bandwidth used in sonar equation detection performance modeling for the NARW upcall in Sec. IV.

Measurement of instantaneous bandwidth taken from 4090 exemplars of the 2013 St. Andrews database shows a similar SNR dependence. Figure 9 depicts MIBW vs SNR (blue) with $\pm 1\sigma$ confidence intervals for a sample of 4090 upcalls ranging in SNR from 3 to 35 dB. Sample size at each SNR is denoted in red. MIBW was observed to vary from 2.5 bins (9.75 Hz) at 3 dB SNR to 6.5 bins (25.35 Hz) at 30 dB SNR. At the recognition differential of the spectrogram correlator algorithm, 5.5 dB, the MIBW is 3 bins (11.7 Hz). Based on this finding, the noise bandwidth driving the detection decision, and thus adopted in the sonar equation model projections that follow, is 12 Hz.

D. Observations of in-band noise level and upcall received level

Figure 10 shows histograms of in-band noise level (red) and upcall received level (blue) compiled over all 1500 J-13 detection events broken down by range offset. The spread between the red and blue histograms is indicative of the SNR at each range. Note that in-band noise level for each detection event is reported in units of dB re 1 μ Pa but is not associated with a fixed noise bandwidth; as explained above, this is because the exact noise bandwidth associated with each detection event is data-dependent. The in-band noise level was observed to change throughout the 7-h experiment, from a low of 79.2 dB re 1 μ Pa at 1432 GMT to a high of 89.1 dB re 1 μ Pa at 1940 GMT. This variability is not surprising as there was a high degree of shipping activity in the lease area on that day; in addition to normal vessel

https://doi.org/10.1121/10.0039669

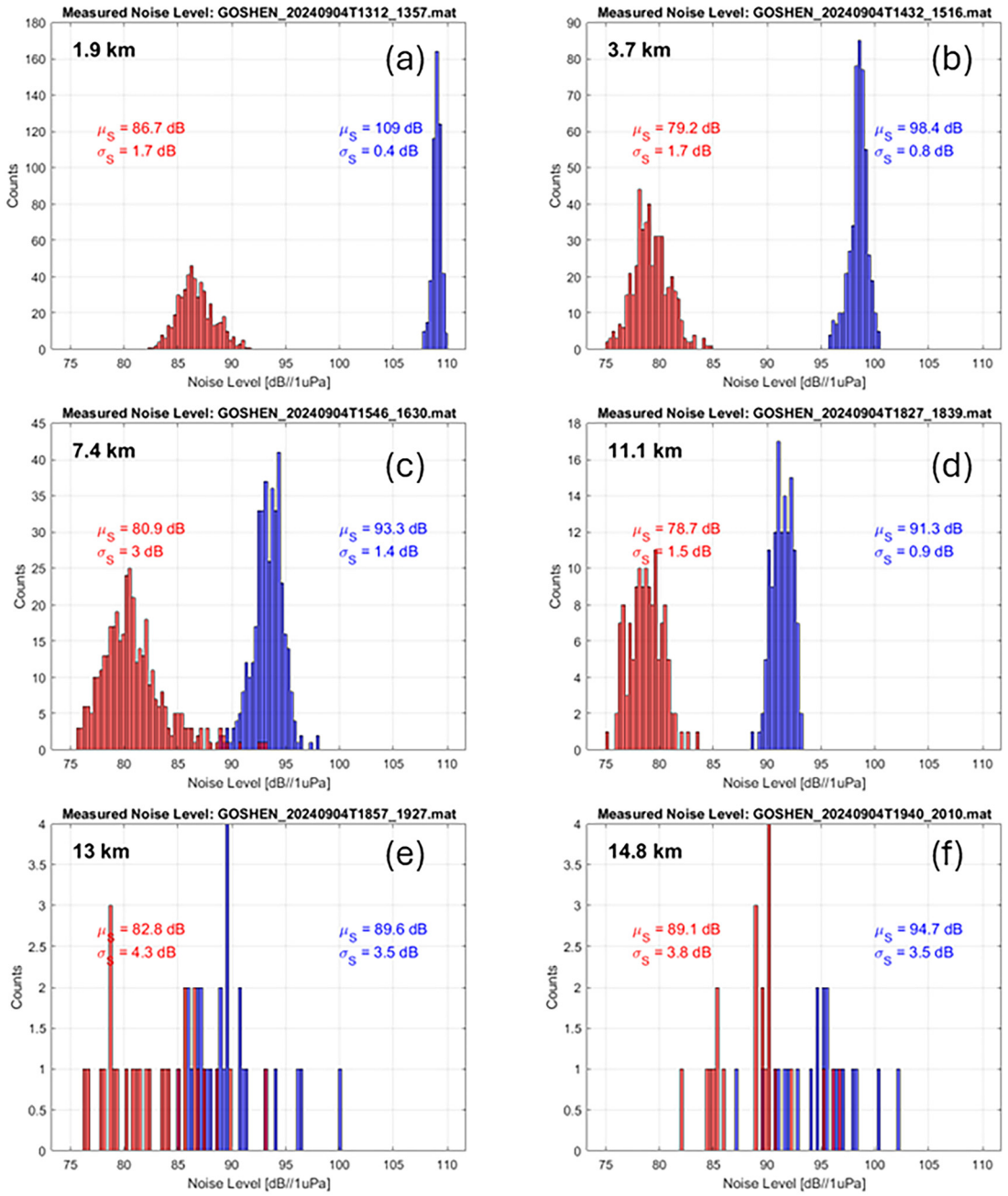

FIG. 10. Histograms of rms received signal level (blue) and in-band beam noise level (red), broken out by range, for all 1500 upcall transmissions from the J-13 measured at the output of the beamformer during the 7-h deployment. Received level and in-band beam noise level are computed in the same manner, integrated over frequency and averaged over time, for those spectrogram pixels exceeding the binarization threshold of 5.5 dB and falling within the upcall kernel contour. In-band noise level reported for each detection event reflects the average instantaneous bandwidth of each detected upcall. In-band beam noise level varied by as much as 10 dB, from 79 to 89 dB re 1 μ Pa during the test due to the large number of surface vessels operating in the lease area. Note also that the sample sizes were smaller at the longer ranges due to the smaller number of detections at those ranges.

traffic, there was construction activity to support the installation of nacelles on piles that had been driven during the previous 4 months. The mean in-band noise level during the 7-h period from 1300 to 2000 GMT was 83 dB re 1 μ Pa. Converting to noise spectrum level at an omni-directional hydrophone, by adding back the mean array gain ($\sim$10 dB) and correcting for a noise bandwidth of 12 Hz (i.e., subtracting 10 $\log_{10}$12), yields a noise spectrum level of 82 dB re 1 μ Pa//Hz. The noise levels observed herein agree reasonably well with third octave band ambient noise measurements (95 dB re 1 μ Pa, or spectrum level 80 dB re 1 μ Pa@150 Hz) reported by Van Parijs *et al.* in 2022 for the Cox's Ledge site (measurement units COX01 and COX02) in closest proximity to the Revolution Wind lease area.[33]

E. Probability of detection vs range (detection function)

The metric that most completely summarizes the detection performance of a passive acoustic monitoring system in a particular TL and noise environment is the detection

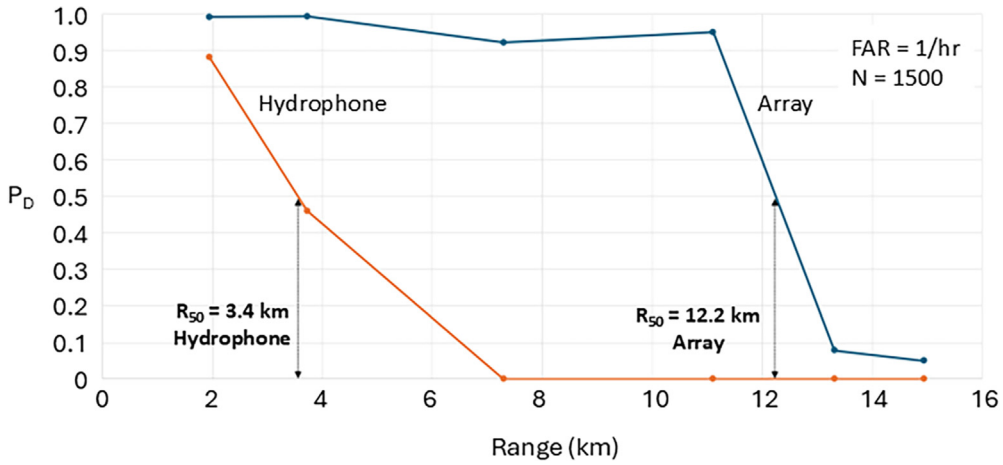

FIG. 11. Measured real-time probability of detection vs range for 32-channel hydrophone array (blue) and single hydrophone (orange) over 1500 NARW upcall transmissions from the J-13 on September 4, 2024. Auto-detector operating point was programmed for a false alarm rate of 1 per hour. Source level as measured by 1-m reference hydrophone was 160 dBrms re 1 μ Pa@1 m. Detection range, R_{50}, is defined as the range at which detection probability decays below 50%. The detection range for the single hydrophone (channel 16 of the array) was measured to be 3.4 km, while that of the 32-channel array was 12.2 km, which constitutes a 3.6-fold detection range advantage in this 17logR TL environment.

function, or probability of detection and classification vs range curve. Probability of detection, P_d, in this instance implies correct classification as we are concerned not just with the detection of acoustic energy, but the detection of a particular signal of interest. It is defined as that fraction of the time that the target is positively identified, when available to the sensor, at a given range.

Figure 11 depicts the measured probability of upcall detection and classification, for a source level of 160 dBrms re 1 μ Pa@1 m, as a function of range for the 32-channel hydrophone array (blue) under test and single hydrophone (orange); channel 16 of the array is used as the representative data point for the single hydrophone. There were approximately 1500 detection opportunities over the 7-h event of September 4, with sample size numbering about 300 at each of the range offsets from 1.9 to 11.1 km, and somewhat fewer at the higher range offsets of 13 and 14.8 km. As stated above, the receiver sensitivity was tuned for a false alarm rate of 1 per hour; to be complete, any statement of detection sensitivity or detection range must be accompanied by an associated false alarm rate; otherwise, the receiver operating point is ambiguous. It should be noted that, due to time constraints on this 1-d deployment (each range offset required a roughly 1-h evolution to reposition the vessel, deploy and recover the J-13 projector), only range dependence was considered in this experiment. Future experiments will examine both range and azimuthal dependence. Detection probability for the array was consistently between 90% and 100% for ranges of 1.9–11.1 km—unsurprising given that SNR varied from 22 to 12 dB in this range interval—before dropping off rapidly at 13 km. The single hydrophone response peaked at probability of detection of 90% at a range of 1.9 km and decayed rapidly thereafter.

If it is desirable to reduce the detection function to a single scalar value, then the metric detection range, R_{50}, is often used. Detection range is defined as the range at which the probability of detection in a given environment has decayed to 50%. Linear interpolation of the array detection function between 11.1 and 13 km shows that the 50% crossing occurs at a range of 12.2 km for the 32-channel array and 3.4 km for the single hydrophone, yielding a 3.6-fold detection range advantage for the array over the hydrophone. This is consistent with theoretical expectations for a sensor exhibiting on the order of 10 dB of array gain in a 17 $\log_{10}R$ continental shelf TL environment (e.g., $10^{10/17} = 3.8$).

While the measured detection range for the array in this experiment exhibited the expected range multiplier advantage over the single hydrophone, the absolute detection range for the array was somewhat less than expected. Previous model predictions by the authors for detection performance of a 32-channel array in the Revolution Wind lease area against a 160 dBrms re 1 μ Pa@1 m source level, assuming an isotropic noise distribution characterized by a noise spectrum level for Wenz moderate to heavy shipping of 78 dB re 1 μ Pa//Hz, projected 20 km.[3] As mentioned above, the mean noise spectrum level on September 4 was measured to be 82 dB re 1 μ Pa//Hz, or about 4 dB re 1 μ Pa//Hz higher than the modeled noise assumption. In this 17 $\log_{10}R$ spreading loss environment, a difference of 4 dB translates to a factor of 1.7× in range (i.e., $10^{4/17} = 1.7$), which maps the 12.2 km detection range measured on September 4 to a predicted detection range of 20.7 km. If the in-band beam noise levels had been 4–5 dB lower on September 4, the 32-channel array would have likely demonstrated a detection range in excess of 20 km.

F. TL and FOM

Figure 12 shows measured and modeled TL results for the Revolution Wind lease area along the radial to the support vessel *R/V Black Hawk* and J-13 source depicted in Fig. 3(b). *In situ* TL measurement was accomplished by simply comparing the measured frequency domain rms received level (from Fig. 10) to the ground-truth time domain rms

https://doi.org/10.1121/10.0039669

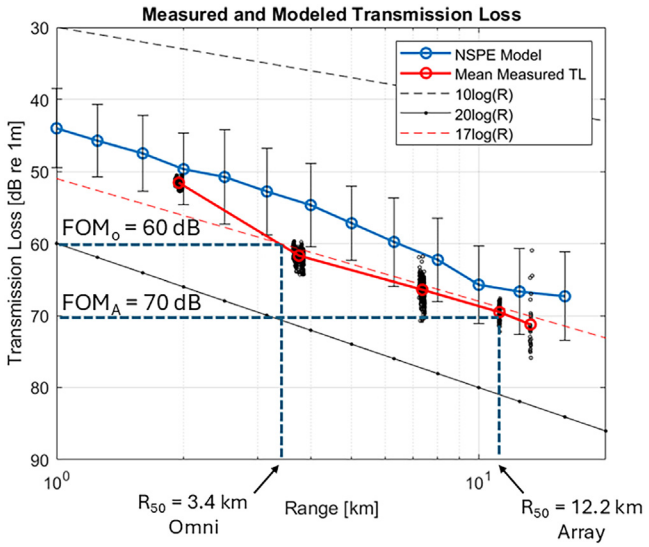

FIG. 12. Measured and modeled TL in the Revolution Wind lease area on September 4, 2024: Measured TL at J-13 range offsets of 1.9, 3.7, 7.4, 11.1, and 14.8 km (black circles) computed from the difference between measured rms received level and measured rms source level for every NARW upcall detection event; $17\log_{10}R$ spreading loss (red dashed); NSPE modeled TL at 150 Hz in third octave range bins with $\pm 1\sigma$ error bars (blue). Array FOM corresponding to measured R_{50} of 12.2 km is 70 dB. Single hydrophone FOM corresponding to measured R_{50} of 3.4 km is 60 dB.

source level at the 1-m reference hydrophone. The black circles in Fig. 12 denote *in situ* TL observations corresponding to upcall detection events at each of the range offsets. Clearly, there is variability in the measured TL at each range offset, on the order of 1.5-3.5 dB from the standard deviation of RL reported in Fig. 10. The variability is higher at the longer ranges, but not unexpectedly high.[36] The solid red line connects the mean TL values at each range. A least squares fit of a scalar spreading coefficient to the measured TL observations is denoted by the dashed red line, indicating a $17\log_{10}R$ spreading loss environment (for ranges >3 km). It should be noted that the measured spreading coefficient of 17 is generally much higher than that assumed in recent studies of upcall source level estimation from measured observations[9] and the evaluation of a passive acoustic coastal buoy for cetacean detection.[16] The validation of TL models using measured data is vital for the accurate physical interpretation of *in situ* system detection performance. Overly favorable TL models can have undesirable consequences, including the underestimation of source level in the inversion of upcall source level from data and the overstatement of detection range performance projections to new environments.

The TL model results of Fig. 12 denoted by the solid blue curve were computed using U.S. Navy Standard Parabolic Equation (NSPE)[37] based on the sound speed profile of Fig. 6 and geoacoustic description of the seabed summarized in Table II, which is in reasonably good agreement with that employed by Lin *et al*. in their propagation model of offshore wind noise at the nearby Block Island wind farm off the southern coast of Rhode Island.[38] The $\pm 1\sigma$ confidence intervals capture the degree of TL spread due to binning TL in third octave range bins. Bathymetric data in the lease area were extracted from the GEBCO 2023 database, a bathymetry data set developed through the Nippon Foundation-GEBCO Seabed 2030 Project.[39] The GEBCO database provides bathymetric data, in meters, on a 15-arc-second interval grid. Note that range dependence in the NSPE TL model is only represented via the bathymetric data. The sound speed profile and seabed geoacoustic model parameters are assumed to be range independent. While the NSPE model appears to underpredict TL between 3.7 and 7.4 km, it appears to converge to the measured data at ranges of 11 km or more. While this observation suggests that the geoacoustic model inputs leave room for improvement, detection performance projections based on the NSPE TL model at ranges of 11 km or more are likely to be representative of the actual environment.

TABLE II. Seabed geoacoustic parameter model for the Revolution Wind lease area.

Parameter	Depth (m)	Value
Compressional wave speed, c_p	0.0	1719.0 m/s
	3.0	1800.0 m/s
Density	0.0	1.946 g/cm^3
	3.0	2.03 g/cms^3
Compressional wave attenuation, α_p	0.0	0.708 dB/km
	3.0	0.057 dB/km

The greatest utility of the *in situ* measured TL from Fig. 12 lies in its reconciliation with the measured detection ranges, R_{50}, in Fig. 11 for the array and single hydrophone, respectively. Through this process, one can determine a measured FOM for array and single hydrophone in the subject environment that is strictly informed by empirical observation without having to invoke any simplifying assumptions about nature of the data or the environment, e.g., Gaussianity, stationarity, etc., that may or may not be true.[17] The FOM constitutes the maximum TL that the system can tolerate for a given source level and noise level and still meet the desired receiver operating point. As such, it is the only absolute measure of system efficacy that supports the fair and objective comparison of acoustic detection performance between competing systems (notwithstanding factors such as endurance, latency, and cost). The measured figures of merit for the 32-channel array and omnidirectional hydrophone in the Revolution Wind environment are delineated by the y-intercepts of the two dashed black lines in Fig. 12 labeled FOM_A and FOM_O, respectively. In this $17\log_{10}R$ spreading loss environment, the array detection range, R_{50_A}, of 12.2 km maps to an array FOM of 70 dB, while the omni-hydrophone detection range, R_{50_O}, of 3.4 km maps to a hydrophone FOM of 60 dB. The difference between the two FOMs may be attributed to the mean array gain for the 32-channel array, which was measured to be

roughly 10 dB during the September 4 experiment. This is consistent with theoretical expectations for an array operating against a signal of interest that is 1–2 octaves below its cut frequency of 625 Hz (i.e., array gain decreases by 3 dB/octave below the array cut frequency).[29]

IV. MODELING OF DETECTION PERFORMANCE: PASSIVE SONAR EQUATION

In this section, we return to the link between noise bandwidth and instantaneous right whale upcall bandwidth and the associated implications for performance prediction based on the passive sonar equation. The passive sonar equation is usually written in terms of the FOM, introduced above, as follows:[40]

$$FOM \equiv TL = SL - (NL - AG) - NRD. \quad (5)$$

where TL represents the transmission loss, SL is the source level (in the case of the NARW upcall, rms source level), NL denotes the in-band ambient noise level measured at the hydrophone, AG is array gain such that $(NL-AG)$ is the in-band noise level at the beamformer output, and NRD is the system recognition differential, or SNR at the beamformer output required for the real-time spectrogram correlator algorithm to yield a desired receiver operating point. Analysis of the 2013 St. Andrews test data has shown that, for an operating point of $P_d = 0.5$ at a false alarm rate of 1 per hour, NRD for the current ThayerMahan system is equal to 5.5 dB. As mentioned earlier, the FOM constitutes the TL that the system can tolerate for a given SL and NL and still meet the level of performance represented by the system NRD.

To test the hypothesis of a narrowband noise model for the upcall detection problem, two sonar equation models are compared to the measured array detection range, R_{50_A}, of 12.2 km. The first employs the mean, narrowband (e.g., 12 Hz), in-band noise level of 83 dB re 1 μ Pa measured at the beamformer output in Sec. III. Recall that this noise definition inherently accounts for the exact noise bandwidth that influences the instantaneous spectrogram correlator algorithm detection decision. The second is based on the long-held convention of a broadband noise bandwidth, in this case tied to a worst case NARW upcall support band of 40–400 Hz, or a noise bandwidth of 360 Hz. The in-band noise level for that model uses the same mean in-band noise measured at the beamformer output, corrected for the difference in bandwidth, i.e., $10\log_{10}360/12$, or 14.8 dB, resulting in an in-band noise level of 97.8 dB re 1 μ Pa. Remaining consistent with the J-13 experiment, an upcall source level of 160 dBrms re 1 μ Pa@1 m was assumed, along with an NRD of 5.5 dB. TL was computed using the NSPE TL model for a frequency of 150 Hz at the center of the NARW support band: in this shallow, downward-refracting environment, TL is only weakly dependent on frequency. As in Fig. 12, the model employed range dependent bathymetry from the GEBCO database, the geoacoustic model of Table II, and the September 4 measured sound speed profile of Fig. 6.

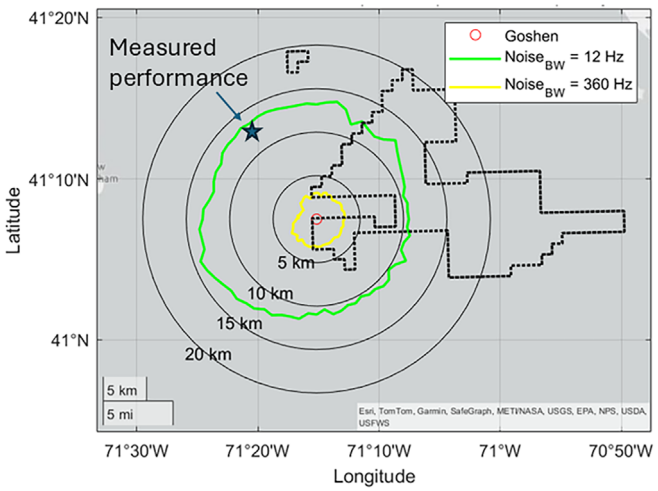

FIG. 13. Modeled detection contours for NARW upcall source level of 160 dBrms re 1 μ Pa@1 m corresponding to ENBW = 12 Hz (green) and ENBW = 360 Hz (yellow). SeaPicket array Goshen denoted by red circle. Boundary of Revolution Wind lease area is identified by the dotted black line. The sonar equation model that uses a narrowband ENBW of 12 Hz yields a predicted detection contour that agrees well with at-sea observations, while the model that employs a broadband ENBW of 360 Hz to represent the in-band ambient noise level significantly underpredicts detection performance measured during the J-13 experiment.

Figure 13 depicts detection contours corresponding to probability of detection of 50% for the narrowband (green) and broadband (yellow) noise models, respectively, overlaid on the outline of the Revolution Wind lease area (black dotted). The solid black concentric circles denote radii of constant detection range relative to the array phase center in 5 km increments. The measured array detection range, R_{50_A}, of 12.2 km, along the radial to the support vessel *R/V Black Hawk* is indicated by the blue star to the northwest of the array (red circle). The measured data point is observed to be in good agreement with the narrowband detection contour, falling short of the green contour by a couple kilometers, or about 20% of range. This modest offset could be attributed to the small amount of signal model mismatch in the spectrogram correlator due to the imperfect alignment of the transmitted upcall with the members of the kernel library; it could also be attributable to a slightly optimistic TL calculation from NSPE. In this $17\log_{10}R$ spreading loss environment, 1.5 dB of signal model mismatch translates to a factor of 1.2× in range, e.g., $10^{\frac{1.5}{17}}$. However, the broadband detection contour (yellow) underpredicts the measured detection performance, by a factor of roughly 7×. This model error is attributed to the fact that in this spreading loss environment, a 15 dB error in assumed in-band noise level costs roughly 7.4× in detection range, e.g., $10^{\frac{14.8}{17}}$. The takeaway from the results of this sonar equation model comparison is that of the two noise models, only the narrowband model is capable of explaining the measured detection performance of September 4, 2024, obtained using the calibrated, ground-truthed source.

https://doi.org/10.1121/10.0039669

V. CONCLUSION

For the first time, the detection performance of a 32-channel hydrophone array, and that of a single hydrophone, has been experimentally quantified for the case of a North Atlantic right whale upcall using a calibrated acoustic projector in the southern New England offshore wind construction area. Measurements of detection range and *in situ* TL were reconciled to yield a measured FOM against a source level of 160 dBrms re 1 μ Pa@1 m in the subject environment of 70 dB for the 32-channel hydrophone array, versus 60 dB for the single hydrophone. These results corroborated earlier findings of a 3.6-fold detection range advantage, for the array relative to the single hydrophone in this 17log$R$ spreading loss environment.[3] As such, autonomous, array-based PAM systems have redefined the state-of-the-art in passive acoustic marine mammal monitoring.

Physical interpretation of the measured detection results using a passive sonar equation treatment was then used to validate the hypothesis that the detection of a North Atlantic right whale upcall is fundamentally a narrowband detection problem. This is a significant departure from long-held convention wherein the noise bandwidth is taken to be the full support band of the NARW upcall for which reports in the open literature vary from 50–225,[9] to 71–224,[14] to 50–400 Hz.[28] The results demonstrate that the broadband noise definition significantly underpredicts measured detection range, while the narrowband noise model accurately explains the observed detection performance. Analysis of upcall transmissions from the J-13, as well as data from the St. Andrews 2013 database, shows that the MIBW of the NARW upcall is estimated to be approximately 12 Hz at a SNR of 5.5 dB, which corresponds to the auto-detector NRD for a receiver operating point of $P_d = 0.5$ and false alarm rate = 1/h associated with the spectrogram correlator classifier referenced herein. Note that the MIBW is likely to be classifier and operating point-dependent, so in general it could range from 10–20 Hz for similar classifiers. This finding has important implications for the treatment of noise bandwidth in baleen whale acoustic detection performance modeling generally, and for the extrapolation of right whale detection performance to new noise environments dominated by offshore wind construction activity and pile driving. The experimental results, metrics, and methodology presented herein provide a framework for rigorous and objective PAM detection performance quantification based on calibrated acoustic source transmissions accompanied by precise positional ground truth reconstruction. The process addresses many of the limitations that have hampered efforts on the part of offshore wind construction contractors, government regulatory agencies, academia, and other interested parties to evaluate competing PAM technologies on an objective and consistent basis through the incorporation of the following elements: (1) the clear definition of detection range, R_{50}, and its direct link to a specified false alarm rate operating point; (2) the explicit accounting for recognition differential in the sonar equation reconciliation; (3) the use of *in situ* measured TL to interpret detection performance; and most importantly, and (4) the employment of calibrated source ops with ground truth reconstruction to enable accurate quantification of detection performance. Finally, this experiment was conducted in the presence of vessels supporting construction activity, namely, the installation of nacelles post monopile installation. In the future, experiments like the one outlined herein must be expanded to validate detection performance projections in noise conditions with more complicated spatial and temporal dependencies, such as those encountered during pile driving operations.[41]

ACKNOWLEDGMENTS

The authors would like to thank science crew members on *R/V Black Hawk* supporting the source operations experiment, including V. Kmelnitsky, D. Morton, J. Williams, A. Salazar, D. Coates, and C. Glander. V.E.P. would like to acknowledge the many insightful questions and discussions during this work with scientists at the Bureau of Ocean Energy Management, in particular, M. Martin, A. Conrad, and S. Denes, as well as with D. Ireland of LGL Ecological Research Associates. V.E.P. and P.A. are particularly indebted to Dr. Chris Clark of Cornell University, and his colleague, U.S. Navy CDR (Ret.) Chuck Gagnon, for many enlightening discussions concerning the history and role of acoustic arrays in marine mammal monitoring and the preservation of endangered species marine habitat. The authors also gratefully acknowledge the support of IR&D investment from ThayerMahan Inc., at the direction of VADM (Ret.) M. Connor, CEO, and R. Hine, ThayerMahan co-founder and President, ThayerMahan Offshore.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study may be available on request from the corresponding author. The data are not publicly available due to export licensing restrictions.

[1]C. W. Clark, "Examining the threats to the North Atlantic right whale," in Testimony before the House Natural Resource Committee, Subcommittee on Water, Oceans, and Wildlife, March 7, 2019. https://www.congress.gov/116/meeting/house/109022/witnesses/HHRG-116-II13-Wstate-ClarkC-20190307.pdf (Last viewed 4/13/2010).

[2]C. W. Clark and G. C. Gagnon, "Low-frequency vocal behaviors of baleen whales in the North Atlantic: Insights from IUSS detections, locations and tracking from 1992 to 1996," J. Underwater Acoust. **52**, 609–640 (2004).

[3]V. Premus, P. Abbot, V. Kmelnitsky, C. Gedney, and T. Abbot, "A wave glider-based, towed hydrophone array system for autonomous, real-time, passive acoustic marine mammal monitoring," J. Acoust. Soc. Am. **152**, 1814–1828 (2022).

[4]D. Wang, H. Garcia, W. Huang, D. D. Tran, A. D. Jain, D. H. Yi, Z. Gong, J. M. Jech, O. R. Godø, N. C. Makris, and P. Ratilal, "Vast

assembly of vocal marine mammals from diverse species on fish spawning ground," Nature **531**, 366–370 (2016).
[5]C. Gervaise, Y. Simard, F. Aulanier, and N. Roy, "Optimizing passive acoustic systems for marine mammal detection and localization: Application to real-time monitoring north Atlantic right whales in Gulf of St. Lawrence," Appl. Acoust. **178**, 107949 (2021).
[6]M. Connor and R. Hine, "Continuous unmanned airborne and underwater monitoring platform," U.S. patent 11,105,662 B2 (2021).
[7]M. Connor, "Systems and methods for autonomous towing of an underwater sensor array," U.S. patent 9,778,388 B1 (2017).
[8]A. Szesciorka, M. Severy, K. Ampela, C. Hein, M. Richlen, J. Haxel, and J. Clerc, "Evaluating tools and technologies for monitoring baleen whales during offshore wind foundation installation," in Pacific Northwest National Laboratories Technical Report PNNL-37249 (2025).
[9]K. Palmer, G.-M. Wu, C. Clark, and H. Klinck, "Accounting for the Lombard effect in estimating the probability of detection in passive acoustic surveys: Applications for single sensor mitigation and monitoring," J. Acoust. Soc. Am. **151**, 67–79 (2022).
[10]L. T. Hatch, C. W. Clark, S. M. Van Parijs, A. Frankel, and D. Ponirakis, "Quantifying loss of communication space for right whales in and around a U. S. National Marine Sanctuary," Conserv. Biol. **26**, 983–994 (2012).
[11]V. Trygonis, E. Gerstein, J. Moir, and S. McCulloch, "Vocalization characteristics of North Atlantic right whale surface active groups in the calving habitat, southeastern United States," J. Acoust. Soc. Am **134**, 4518–4531 (2013).
[12]C. Clark, W. Ellison, L. Hatch, R. Merrick, S. Van Parijs, and D. Wiley, "An ocean observing system for large-scale monitoring and mapping of noise throughout the Stellwagen Bank Nation Marine Sanctuary," in *Reports to the National Oceanographic Partnership Program, Stellwagen Project*, Award N00014-07-1-1029 (2010).
[13]C. Clark, W. Ellison, L. Hatch, R. Merrick, S. Van Parijs, and D. Wiley, "An ocean observing system for large-scale monitoring and mapping of noise throughout the Stellwagen Bank Nation Marine Sanctuary," in *Reports to the National Oceanographic Partnership Program, Stellwagen Project*, Award N00014-07-1-1029 (2011).
[14]B. J. Estabrook, J. T. Tielens, A. R. Rahaman, D. W. Ponirakis, C. W. Clark, and A. N. Rice, "Dynamic spatiotemporal acoustic occurrence of North Atlantic right whales in the offshore Rhode Island and Massachusetts wind energy areas," Endang. Species Res. **49**, 115–133 (2022).
[15]H. D. Johnson, C. T. Taggart, A. E. Newhall, Y.-T. Lin, and M. F. Baumgartner, "Acoustic detection range of right whale upcalls identified in near-real time from a moored buoy and a Slocum glider," J. Acoust. Soc. Am. **151**, 2558–2575 (2022).
[16]K. J. Palmer, S. Tabbutt, D. Gillespie, J. Turner, P. King, D. Tollit, J. Thompson, and J. Wood, "Evaluation of a coastal acoustic buoy for cetacean detections, bearing accuracy and exclusion zone monitoring," Methods Ecol. Evol. **13**, 2491–2502 (2022).
[17]V. Premus, P. Abbot, M. Helfrick, C. Emerson, and T. Paluskiewicz, "Passive sonar performance characterization and transmission loss measurement using a calibrated mobile acoustic source," in *Proceedings of the 2nd International Conference and Exhibition on Underwater Acoustics*, Kos, Greece (June 2014).
[18]Colibri iMX7 System. https://docs.toradex.com/103125-colibri-arm-som-imx7-datasheet.pdf (Last viewed 10/23/2021).
[19]D. K. Mellinger and C. W. Clark, "Recognizing transient low-frequency whale sounds by spectrogram correlation," J. Acoust. Soc. Am. **107**, 3518–3529 (2000).
[20]T. Abbot, V. Premus, and P. Abbot, "A real-time method for autonomous passive acoustic detection-classification of humpback whales," J. Acoust. Soc. Am. **127**, 2894–2903 (2010).
[21]T. Abbot, V. Premus, P. Abbot, and O. Mayer, "Receiver operating characteristic for a spectrogram correlator-based humpback whale detector classifier," J. Acoust. Soc. Am. **132**, 1502–1510 (2012).
[22]D. Gillespie, *DCLDE 2013 Workshop Dataset*, University of St Andrews Research Portal (St Andrews, Scotland, 2019).
[23]Y. Shiu, K. 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," Sci. Rep. **10**, 607 (2020).
[24]Revolution Wind L078 Layout (May 2023).
[25]T. Abbot, V. Premus, and P. Abbot, "Observations of baleen whale vocalizations in the new England offshore wind lease area from August 2023 to July 2024 using multiple high resolution, bottom-mounted hydrophone arrays," in *Proceedings of OCEANS 2024*, Halifax, Nova Scotia, Canada (2024).
[26]V. Premus, "Observations on the use of hydrophone arrays for passive acoustic monitoring of vocalizing baleen whales," in *NYSERDA State of the Science Workshop*, (Stony Brook, NY, 2024).
[27]Underwater Electroacoustic Standard Transducers Catalogue, Naval Research Laboratory Underwater Sound Reference Detachment (May 1982).
[28]S. E. Parks, M. Johnson, D. Nowacek, and P. L. Tyack, "Individual right whales call louder in increased environmental noise," Biol. Lett. **7**, 33–35 (2011).
[29]R. J. Urick, *Sound Propagation in the Sea* (Defense Advanced Research Projects Agency, Washington, DC, 1979).
[30]P. Abbot, I. Dyer, and C. Emerson, "Acoustic propagation uncertainty in the shallow East China Sea," IEEE J. Oceanic Eng. **31**(2), 368–383 (2006).
[31]I. Dyer, "Ocean ambient noise," in *Handbook of Acoustics* (Wiley, New York, 1998).
[32]P. Arveson and D. Vendittis, "Radiated noise characteristics of a modern cargo ship," J. Acoust. Soc. Am. **107**, 118–129 (2000).
[33]S. M. Van Parijs, A. I. DeAngelis, T. Aldrich, R. Gordon, A. Holdman, J. A. McCordic, X. Mouy, T. J. Rowell, S. Tennant, A. Westell, and G. E. Davis, "Establishing baselines for predicting change in ambient sound metrics, marine mammal, and vessel occurrence within a US offshore wind energy area," ICES J. Mar. Sci. **82**, 1–14 (2025).
[34]G. M. Wenz, "Acoustic ambient noise in the ocean: Spectra and sources," J. Acoust. Soc. Am. **34**, 1936–1956 (1962).
[35]L. Jackson, *Linear Systems Theory and Application* (Addison-Wesley, Reading, MA, 1991).
[36]I. Dyer, "Statistics of sound propagation in the ocean," J. Acoust. Soc. Am. **48**, 337–345 (1970).
[37]M. Collins, "A split-step Pade solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[38]Y. T. Lin, A. E. Newhall, J. H. Miller, G. R. Potty, and K. J. Vigness-Raposa, "A three-dimensional underwater sound propagation model for offshore wind farm noise prediction," J. Acoust. Soc. Am. **145**, EL335–EL340 (2019).
[39]Nippon Foundation-GEBCO Seabed 2030 Project. https://www.gebco.net/data_and_products/historical_data_sets/ (Last viewed 4/6/2025).
[40]A. W. Cox, *Sonar and Underwater Sound*, 2nd ed. (Lexington Books, Lexington, MA, 1982).
[41]V. Premus, P. Abbot, T. Abbot, E. Illich, J. Browning, V. Kmelnitsky, J. Freise, and A. Logan, "The impact of pile driving noise on North Atlantic right whale detection using coherently beamformed hydrophone arrays," in *Eighth Underwater Acoustics Conference and Exhibition*, Halkidiki, Greece (2025).